Weak Instrumental Variables Models for Longitudinal Data[*]

Zongwu Cai[a,b],   Ying Fang[b],   and   Henong Li[a]

[a]Department of Mathematics & Statistics, University of North Carolina at Charlotte,
Charlotte, NC 28223, USA
[b]Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian 361005, China

This Version: October 19, 2008

In this paper, we study a weak instrumental variables model for longitudinal data. A two stage least-squares estimator (the instrumental variables estimator) is presented. We show that the asymptotic property for the proposed estimator is different from that for cross-sectional data. Also, similar to Hahn and Kuersteiner (2002), we extend a local-to-zero assumption as in Staiger and Stock (1997) on the coefficients of the instruments in the first stage equation to a more general setting by allowing for different degrees of weakness. Moreover, the consistency and limiting distribution of the proposed estimators are established and the explicit expressions for the asymptotic bias are given. Further, we show that the discontinuity phenomenon observed in Hahn and Kuersteiner (2002) still exists for the longitudinal data case. Finally, we examine the finite sample properties of the proposed estimator by Monte Carlo simulations.

KEY WORDS: Endogenous variables; Discontinuity; Instrumental variables estimator; Longitudinal data; Panel data; Weak instrumental variables.

---

1 Introduction

During recent years, there have been a lot of attentions to weak instrumental variables (IV) models in the literature, since they have been applied to various applications such as economics and finance.[1] An instrumental variable is called to be weak instrumental variable if it is weakly correlated with the endogenous explanatory variables. As pointed out by Bound, Jaeger, and Baker (1995), the weak instrument is not a small sample problem. Indeed, they provided an empirical example of weak instruments with 329,000 observations, while Nelson and Startz (1990) and Maddala and Jeong (1992) examined the behavior of the two-stage least squares (TSLS) estimator and showed that the approximation of sampling distributions of TSLS estimator by normal distributions can be quite poor. Since then there have been a lot of research focusing on nonstandard approximations to sampling distributions.

To characterize weak instrumentals, Staiger and Stock (1997) considered initially a linear simultaneous equations model and proposed the so-called *local-to-zero* parameterization of the coefficients of the instruments in the reduced form equation. Furthermore, they showed that, under this local-to-zero framework with the number of instruments fixed, both the TSLS and limited information maximum likelihood estimators are inconsistent but instead converge to nonstandard distributions. To echo this finding, Hahn and Kuersteiner (2002) considered the same type model as in Staiger and Stock (1997) by generalizing Staiger and Stock's (1997) specification to a continuum of parametrizations under which degrees of weakness are classified into three scenarios: (I) the *weak* case, exactly considered by Staiger and Stock (1997), (II) the *nearly weak* case, in which the instruments are stronger than the case considered by Staiger and Stock (1997), and (III) the *nearly non-identified* case, in which the instruments are weaker than the case considered by Staiger and Stock (1997). Further, Hahn and Kuersteiner (2002) showed that, for both the nearly non-identified and Staiger and Stock's (1997) weak cases, the TSLS estimators are inconsistent although they have a non-normal limiting distribution. However, for the nearly weak case, the TSLS estimator is consistent and its limiting distribution exists and it is normal, but it does not reflect the type of finite sample moments usually associated with the TSLS estimator, while it was shown by Chao and Swanson (2007) that the weak instrument limit of Staiger and Stock (1997) preserves the exact finite sample moments of TSLS under some regularity conditions.

Although there has been a vast of literature on the weak instrumental variables models

---

[1]See, for example, the papers by Angrist and Krueger (1991), Cai and Li (2008), Campbell (2003), Hahn, Chao and Swanson (2007), Hahn, Hausman and Kuersteiner (2004), Hahn and Kuersteiner (2002), Hall (1978), Hausman, Stock and Yogo (2005), Li (2006), Mavroeidis (2004), Nason and Smith (2005), Neeley, Roy and Whiteman (2001), Staiger and Stock (1997), Stock (2002), Stock and Wright (2000), Stock, Wright and Yogo (2002), Woodford (2003), and Yogo (2004).

for cross-sectional data, as mentioned above, little or no attention has been paid to weak instrumental variables model for longitudinal (panel) data. Longitudinal data models have become increasingly popular among applied researchers due to their heightened capacity for capturing the complexity of human behavior as compared to cross-sectional or time-series data models. Also, more and richer longitudinal data sets have become increasingly available. Indeed, there is a rich literature on linear and nonlinear longitudinal data models for using instrumental variable approach.[2] For example, Li and Stengos (1996), Li and Ullah (1998), Baltagi and Li (2002), Cai and Li (2008) considered instrumental variable estimators for semiparametric dynamic panel data model. The detailed statistical inferences and economic interpretations on panel estimation of IV models, can be found in the books by Arellano (2003), Hsiao (2003), and Baltagi (2005).

In this paper, we consider the following parametric model for longitudinal data

$$y_{it} = \boldsymbol{\beta}^\top \mathbf{x}_{it} + \alpha_i + u_{it}, \ i = 1, \ldots, N, \ t = 1, \ldots, T$$

where $\mathbf{x}_{it}$ is a vector of explanatory variables, $\boldsymbol{\beta}$ is a vector of unknown parameters of interest, $\mathbf{A}^\top$ denotes the transpose of a matrix or vector $\mathbf{A}$, the error term u_{it} represents the effect of omitted variables that are peculiar to both individual units and time, and α_i represents the effect of those variables peculiar to the i-th individual in more or less the same fashion over time. If $\mathbf{x}_{it}$ is uncorrelated with u_{it}, the least-squares dummy-variable (LSDV) estimator (covariance estimator or within-group estimator) of $\boldsymbol{\beta}$ is consistent whether individual effects $\{\alpha_i\}$ are fixed or random. But if $\mathbf{x}_{it}$ contains some endogenous regressors, strict exogeneity of regressors no longer holds, so that the LSDV estimator is inconsistent whether individual effects $\{\alpha_i\}$ are fixed or random. Therefore, to obtain a consistent estimator, it is well documented that instrumental variables are needed. In this paper, we consider the following instrumental variables model for longitudinal data

$$y_{it} = \boldsymbol{\beta}^\top \mathbf{x}_{it} + \boldsymbol{\gamma}^\top \mathbf{w}_{it} + \alpha_i + u_{it}, \qquad \mathbf{x}_{it} = \boldsymbol{\Pi}^\top \mathbf{z}_{it} + \boldsymbol{\Phi}^\top \mathbf{w}_{it} + \mathbf{v}_{it},$$

where $\mathbf{x}_{it}$ is a vector of endogenous variables, $\mathbf{w}_{it}$ is a vector of exogenous variables, $\mathbf{z}_{it}$ is a vector of instrumental variables correlated with $\mathbf{x}_{it}$. Our focus is on the IV estimator and discussing its asymptotic properties for various scenarios, described as follows.

In various applications, longitudinal data usually have the distinguishing feature that a large number of individuals are observed over a relatively short period of time. Therefore,

---

[2]See, for example, the papers by Baltagi and Li (2002), Cai and Li (2008), Horowitz and Markatou (1996), Kao (1999), Li and Hsiao (1998), Li and Stengos (1996), and Li and Ullah (1998), Phillips and Moon (1999), and the references therein.

the theoretical and empirical analysis from the large N and small T panel data sets have been the traditional object of study in panel data analysis. However, some of longitudinal data sets, such as the Penn-World tables and the National Longitudinal Survey, cover different individuals, industries, and countries over long time periods and have been useful in assessing and comparing growth characteristics, like real per capita GDP growth. One of the important features of these data sets is that they sometimes have an appreciable time series dimension T as well as a large cross-sectional dimension N. Therefore, we need to consider the asymptotic theory of the resulting estimator for both cases: large N and small T and large N and large T. We will show surprisedly that the asymptotic theories are different for two cases under the weak instruments setting.

Our main contributions of this paper are as follows. First, for large N and fixed T, similar to Hahn and Kuersteiner (2002), we show that the IV estimator for $\boldsymbol{\gamma}$ is always consistent with the same rate of convergence at $N^{1/2}$ for all three cases of weakness and the IV estimator of $\boldsymbol{\beta}$ is only consistent with the rate of convergence at $N^{1/2-\alpha}$ for $0 < \alpha < 1/2$ (the nearly weak case, defined in Section 2) and both asymptotic distributions are normal. However, for $\alpha \geq 1/2$ (the weak and nearly non-identified cases), the IV estimator of $\boldsymbol{\beta}$ is inconsistent although it has some limiting distribution which is not normal. In particular, the explicit expression for the bias in the limiting distribution is provided and it is shown to be proportional to T^{-1} for $\alpha = 1/2$ (the weak case), which can be ignored if T is large (even fixed). These results seem to be novel in the literature. Further, similar to Hahn and Kuersteiner (2002), we show that the so-called *discontinuity* still exists for the longitudinal data. Moreover, similar to Hahn and Kuersteiner (2002) and Chao and Swanson (2007), the explicit expressions for the asymptotic bias are given. Finally, we show that when both N and T tend to infinity and the instrument variables are weak for all three cases, the IV estimator for $\boldsymbol{\gamma}$ is consistent and asymptotically normal distributed with the rate of convergence at $(NT)^{1/2}$, and the IV estimator of $\boldsymbol{\beta}$ is consistent and asymptotically normally distributed with the rate of convergence at $T^{1/2}N^{1/2-\alpha}$. Therefore, the discontinuity does not exist any more for both large N and large T.

The rest of the paper is organized as follows. In the next section, we introduce the model, extend a local-to-zero assumption as in Staiger and Stock (1997) on the coefficients of the instruments in the first stage equation to a more general setting, and give the IV estimator. In Section 3, the asymptotic distribution of the IV estimator is given. First, for a fixed T, the limiting distributions are established for all three cases with some discussions. Second, for an infinite T, the asymptotic normality of the IV estimator is presented. In Section 4, we examine the finite sample properties of the IV estimator by Monte Carlo simulations.

Appendix contains the proofs of certain lemmas needed in the proofs of the main results in Section 3 and the proofs of the main theorems.

2   Setups

Our focus in this paper is on the following longitudinal data instrumental variables model

$$y_{it} = \boldsymbol{\beta}^\top \mathbf{x}_{it} + \boldsymbol{\gamma}^\top \mathbf{w}_{it} + \alpha_i + u_{it}, \qquad \mathbf{x}_{it} = \boldsymbol{\Pi}^\top \mathbf{z}_{it} + \boldsymbol{\Phi}^\top \mathbf{w}_{it} + \mathbf{v}_{it}. \tag{1}$$

where y_{it} is a scalar dependent variable, $\mathbf{x}_{it}$ is a $p \times 1$ vector of endogenous variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, $\mathbf{w}_{it}$ is a $k \times 1$ vector of exogenous variables, $\boldsymbol{\gamma}$ is a $k \times 1$ vector of unknown parameters, $\boldsymbol{\Pi}$ is a $q \times p$ matrix of unknown parameters, $\boldsymbol{\Phi}$ is a $k \times p$ matrix of unknown parameters, $\mathbf{z}_{it}$ is a $q \times 1$ $(q \geq p)$ vector of instrumental variables correlated with $\mathbf{x}_{it}$. Here, we assume that $\mathbf{w}_{it}$ and $\mathbf{z}_{it}$ are uncorrelated with u_{it} and $\mathbf{v}_{it}$ so that $\mathbf{w}_{it}$ is a vector of exogenous variables and $\mathbf{z}_{it}$ is a vector of instrumental variables. Finally, we assume that $\{\alpha_i\}$ are independent across individuals if they are random. Denote by $\mathbf{x}_{it} = (x_{1it}, x_{2it}, \cdots, x_{pit})^\top$, $\mathbf{w}_{it} = (w_{1it}, w_{2it}, \cdots, w_{kit})^\top$, $\mathbf{z}_{it} = (z_{1it}, z_{2it}, \cdots, z_{qit})^\top$, and $\mathbf{v}_{it} = (v_{1it}, v_{2it}, \cdots, v_{pit})^\top$.

We reexpress (1) in a matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\alpha} + \mathbf{U},, \quad \mathbf{X} = \mathbf{Z}\boldsymbol{\Pi} + \mathbf{W}\boldsymbol{\Phi} + \mathbf{V}, \tag{2}$$

where $\mathbf{Y} = (\mathbf{y}_1^\top, \mathbf{y}_2^\top, \cdots, \mathbf{y}_N^\top)^\top$ and for each $1 \leq i \leq N$, $\mathbf{y}_i = (y_{i1}, y_{i2}, \cdots, y_{iT})^\top$. Here, the definitions of $\mathbf{x}_i$, $\mathbf{w}_i$, $\mathbf{z}_i$, $\mathbf{u}_i$, and $\mathbf{v}_i$ are similar to that for $\mathbf{y}_i$ and the definitions of $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{W}$, $\mathbf{U}$, and $\mathbf{V}$ are in the same fashion as that for $\mathbf{Y}$, as well as $\boldsymbol{\alpha} = (\mathbf{e}^\top \alpha_1, \mathbf{e}^\top \alpha_2, \cdots, \mathbf{e}^\top \alpha_N)^\top = (\alpha_1, \alpha_2, \cdots, \alpha_N)^\top \otimes \mathbf{e}$ with $\mathbf{e} = (1, 1, \cdots, 1)_{T \times 1}^\top$ and $\otimes$ denoting the Kronecker product. The presence of $\{\alpha_i\}$ produces a correlation among residuals of the same cross-sectional unit if they are random, and characterizes the individual effect if they are fixed.

Our main interest is to make statistical inferences on $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ under weak instruments setting. Similar to a local-to-zero assumption as in Staiger and Stock (1997) and Hahn and Kuersteiner (2002) on the coefficients of the instruments in the first stage equation, we make the following assumptions about $\boldsymbol{\Pi}$ and $\boldsymbol{\Phi}$.

Assumption 1: $\boldsymbol{\Pi} = K_N^{-1} \times \mathbf{C}$ and $\boldsymbol{\Phi} = K_N^{-1} \times \mathbf{C}_1$, where $\mathbf{C}$ and $\mathbf{C}_1$ are fixed $q \times p$ matrix and fixed $k \times p$ matrix respectively. The K_N is a scalar, and $K_N \to \infty$ as $N \to \infty$.

Assumption 2: $\{(\mathbf{w}_{it}, \mathbf{z}_{it})\}$ and $\{u_{it}\}$ are independent across both individuals and time, $\{(\mathbf{w}_{it}, \mathbf{z}_{it})\}$ and $\{\mathbf{v}_{it}\}$ are independent across both individuals and time, and $\{(\mathbf{w}_{it}, \mathbf{z}_{it})\}$ are iid across both individuals and time with the covariance matrix $\boldsymbol{\Sigma}_0 = \begin{pmatrix} \boldsymbol{\Sigma}_{ww} & \boldsymbol{\Sigma}_{wz} \\ \boldsymbol{\Sigma}_{zw} & \boldsymbol{\Sigma}_{zz} \end{pmatrix}$.

Assumption 3: We assume that $\{(u_{it}, \mathbf{v}_{it})\}$ are iid across individuals and time with the mean zero and the covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{uu} & \boldsymbol{\Sigma}_{uv} \\ \boldsymbol{\Sigma}_{vu} & \boldsymbol{\Sigma}_{vv} \end{pmatrix}$.

Assumption 1 generalizes the local-to-zero parameterization of the coefficients of the instruments in the first-stage regression (2) as in Staiger and Stock (1997). When $K_N = N^\alpha$, $0 < \alpha < 1/2$ corresponds to the nearly weak case considered by Hahn and Kuersteiner (2002), $\alpha = 1/2$ becomes the weak case considered by Staiger and Stock (1997), and $\alpha > 1/2$ reduces to the nearly non-identified case considered by Hahn and Kuersteiner (2002). Therefore, similar to Hahn and Kuersteiner (2002), in what follows, we consider three cases: the case that $N^{1/2}/K_N \to 1$ is regarded as weak in the sense of Staiger and Stock (1997), the case that $N^{1/2}/K_N \to \infty$ is considered as the nearly weak as in Hahn and Kuersteiner (2002), and the case that $N^{1/2}/K_N \to 0$ is treated as the nearly non-identified defined in Hahn and Kuersteiner (2002).

To drive the estimator of parameters, we follow the convention to remove $\{\alpha_i\}$. To this end, we transform (1) into the following form

$$
\begin{aligned}
y_{it} - \bar{y}_{i.} &= \boldsymbol{\beta}^\top(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.}) + \boldsymbol{\gamma}^\top(\mathbf{w}_{it} - \bar{\mathbf{w}}_{i.}) + (u_{it} - \bar{u}_{i.}), \\
\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.} &= \boldsymbol{\Pi}^\top(\mathbf{z}_{it} - \bar{\mathbf{z}}_{i.}) + \boldsymbol{\Phi}^\top(\mathbf{w}_{it} - \bar{\mathbf{w}}_{i.}) + (\mathbf{v}_{it} - \bar{\mathbf{v}}_{i.}),
\end{aligned}
$$

where $\bar{y}_{i.}$ is the average of $\{y_{it}\}$ over index t; that is $\bar{y}_{i.} = T^{-1}\sum_{t=1}^{T} y_{it}$, and the definitions of $\bar{\mathbf{x}}_{i.}$, $\bar{\mathbf{w}}_{i.}$, $\bar{\mathbf{z}}_{i.}$, $\bar{u}_{i.}$, and $\bar{\mathbf{v}}_{i.}$ are same as that for $\bar{y}_{i.}$. Let $\widetilde{y}_{it} = y_{it} - \bar{y}_{i.}$. We define $\widetilde{\mathbf{x}}_{it}$, $\widetilde{\mathbf{w}}_{it}$, $\widetilde{\mathbf{z}}_{it}$, and $\widetilde{\mathbf{v}}_{it}$ in the same fashion as $\widetilde{y}_{it}$. Then, (2) becomes

$$
\widetilde{\mathbf{Y}} = \widetilde{\mathbf{X}}\boldsymbol{\beta} + \widetilde{\mathbf{W}}\boldsymbol{\gamma} + \widetilde{\mathbf{U}} \equiv \widetilde{\mathbf{X}}^*\boldsymbol{\theta} + \widetilde{\mathbf{U}}, \quad \widetilde{\mathbf{X}} = \widetilde{\mathbf{Z}}\boldsymbol{\Pi} + \widetilde{\mathbf{W}}\boldsymbol{\Phi} + \widetilde{\mathbf{V}}, \tag{3}
$$

where $\widetilde{\mathbf{X}}^* = (\widetilde{\mathbf{X}}, \widetilde{\mathbf{W}})$ and $\boldsymbol{\theta}^\top = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)$. Here, the definition of $\widetilde{\mathbf{Y}}$ is similar to that for $\mathbf{Y}$ and so are $\widetilde{\mathbf{y}}_i$, $\widetilde{\mathbf{X}}$, $\widetilde{\mathbf{x}}_i$, $\widetilde{\mathbf{W}}$, $\widetilde{\mathbf{w}}_i$, $\widetilde{\mathbf{Z}}$, $\widetilde{\mathbf{z}}_i$, $\widetilde{\mathbf{U}}$, $\widetilde{\mathbf{u}}_i$, $\widetilde{\mathbf{V}}$, and $\widetilde{\mathbf{v}}_i$. In fact, $\widetilde{\mathbf{z}}_i = \mathbf{Q}\mathbf{z}_i$, where $\mathbf{Q} = \mathbf{I}_T - T^{-1}\mathbf{e}\mathbf{e}^\top$, and $\mathbf{I}_d$ denotes the $d \times d$ identify matrix. The presence of $\{\alpha_i\}$ produces a correlation among residuals of the same cross-sectional unit, although residuals from different cross-sectional units are independent. However, regardless of whether $\{\alpha_i\}$ are treated as fixed or random, the individual-specific effects for a given sample can be swept out by the idempotent transformation matrix $\mathbf{Q}$.

For simplicity of presentation, we provide some additional definitions and notations. If $\mathbf{W}$ is a $p \times q$ matrix, $\text{Vec}(\mathbf{W})$ denotes a $pq \times 1$ vector formed by stacking the columns of $\mathbf{W}$ under each; that is, if $\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_q)$, where $\mathbf{W}_i$ is a $p \times 1$ vector for $i = 1, \cdots, q$, then $\text{Vec}(\mathbf{W}) = (\mathbf{W}_1^\top, \mathbf{W}_2^\top, \cdots, \mathbf{W}_q^\top)^\top$. In the sequel, the symbols " $\Rightarrow$ " and " $\to^p$ " denote the convergence in distribution and in probability, respectively. Finally,

let $\mathbf{P}_z = \widetilde{\mathbf{Z}}^* \left(\widetilde{\mathbf{Z}}^{*\top} \widetilde{\mathbf{Z}}^* \right)^{-1} \widetilde{\mathbf{Z}}^{*\top}$, which is the orthogonal project matrix generated by $\widetilde{\mathbf{Z}}^*$ and $\widetilde{\mathbf{Z}}^* = (\widetilde{\mathbf{W}}, \widetilde{\mathbf{Z}})$. Therefore, the IV estimator of $\boldsymbol{\theta}$ is given by

$$\widehat{\boldsymbol{\theta}} = \left(\widetilde{\mathbf{X}}^{*\top} \mathbf{P}_z \widetilde{\mathbf{X}}^* \right)^{-1} \widetilde{\mathbf{X}}^{*\top} \mathbf{P}_z \widetilde{\mathbf{Y}}. \tag{4}$$

For details, see Hsiao (2003).

3 Asymptotic Theory

To derive the asymptotic properties of the resulting estimator, we consider two cases: large N and small T presented in Subsection 3.1, and large N and large T discussed in Subsection 3.2, since the asymptotic behaviors of the resulting estimator for two cases are different. To present the asymptotic distribution, we need the following additional notations. Let $\mathbf{Z}_u^\top = (\mathbf{Z}_{wu}^\top, \mathbf{Z}_{zu}^\top)$, $\mathbf{Z}_v^\top = (\mathbf{Z}_{wv}^\top, \mathbf{Z}_{zv}^\top)$, and $(\mathbf{Z}_u^\top, (\mathrm{Vec}(\mathbf{Z}_v))^\top)^\top \sim N(\mathbf{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}_0)$, where $\mathbf{Z}_{wu}$ and $\mathbf{Z}_{zu}$ are $k \times 1$ and $q \times 1$ random vectors, respectively, and $\mathbf{Z}_{wv}$ and $\mathbf{Z}_{zv}$ are $k \times p$ and $q \times p$ random matrices, respectively. Note that the distribution of a random matrix $\mathbf{Z}_v$ is defined as the distribution of $\mathrm{Vec}(\mathbf{Z}_v)$. Define $\mathbf{H}_1 = \mathrm{diag}\{\mathbf{I}_p, N^{1/2}\mathbf{I}_k\}$, $\mathbf{H}_2 = N^{1/2} \mathrm{diag}\{K_N^{-1}\mathbf{I}_p, \mathbf{I}_k\}$, and $\mathbf{H}_3 = \mathbf{H}_1$. Let $\boldsymbol{\Omega} = \boldsymbol{\Sigma}_{zz} - \boldsymbol{\Sigma}_{zw}\boldsymbol{\Sigma}_{ww}^{-1}\boldsymbol{\Sigma}_{wz}$ and $\boldsymbol{\Delta} = \boldsymbol{\Sigma}_0 \begin{pmatrix} \mathbf{C}_1 & \mathbf{I} \\ \mathbf{C} & \mathbf{0} \end{pmatrix}$. Next, we present the asymptotic results with their proofs relegated to the Appendix.

3.1 Large N and Small T

For simplicity of notation, we define $\boldsymbol{\lambda} = (T-1)^{1/2}\boldsymbol{\Omega}^{1/2}\mathbf{C}\boldsymbol{\Sigma}_{vv}^{-1/2}$, and $c(\alpha)$ to indicate the degree of weakness as $c(\alpha) = 1$ for weak case, $c(\alpha) = 2$ for nearly weak case, and $c(\alpha) = 3$ for nearly non-identified case. Set,

$$\boldsymbol{\Delta}_{c(\alpha)} = \sqrt{T-1} \left[I(c(\alpha) \neq 3) \left(\boldsymbol{\Sigma}_0 \begin{pmatrix} \mathbf{C}_1 \\ \mathbf{C} \end{pmatrix} \quad \mathbf{0} \right) + I(c(\alpha) \neq 2) \left(\frac{\mathbf{Z}_v}{\sqrt{T-1}} \quad \mathbf{0} \right) + \boldsymbol{\Sigma}_0 \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \right],$$

where $I(A)$ is the indicator function of event A. Now, the asymptotic distribution of the estimator $\widehat{\boldsymbol{\theta}}$ is stated in Theorem 1, together with its associated corollaries. All technical proofs in this subsection are given in the Appendix.

Theorem 1: *Suppose that Assumptions 1, 2 and 3 hold for a fixed T, as $N \to \infty$, we have,*

$$\mathbf{H}_{c(\alpha)} \left[\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right] \Rightarrow (\boldsymbol{\Delta}_{c(\alpha)}^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Delta}_{c(\alpha)})^{-1} \boldsymbol{\Delta}_{c(\alpha)}^\top \boldsymbol{\Sigma}_0^{-1} \mathbf{Z}_u,$$

where $\mathbf{Z}_u$ is a multivariate normal random vector with mean zero and covariance matrix $\sigma_{uu}\boldsymbol{\Sigma}_0$.

From Theorem 1, one can obtain easily the consistency and inconsistency of the IV estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively, stated in Corollary 1 for $\widehat{\boldsymbol{\beta}}$ and Corollary 2 for $\widehat{\boldsymbol{\gamma}}$ with their proofs given in the Appendix. First, define the random vector $\mathbf{Z}_1 = \boldsymbol{\Omega}^{-1/2}(\mathbf{Z}_{zu} - \boldsymbol{\Sigma}_{zw}\boldsymbol{\Sigma}_{ww}^{-1}\mathbf{Z}_{wu})\sigma_{uu}^{-1/2}$ and the random matrix $\mathbf{Z}_2 = \boldsymbol{\Omega}^{-1/2}(\mathbf{Z}_{zv} - \boldsymbol{\Sigma}_{zw}\boldsymbol{\Sigma}_{ww}^{-1}\mathbf{Z}_{wv})\boldsymbol{\Sigma}_{vv}^{-1/2}$. It is easy to check that $\mathbf{Z}_{12} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\rho \otimes \mathbf{I}_q)$, where $\mathbf{Z}_{12} = \begin{pmatrix} \mathbf{Z}_1 \\ \text{Vec}(\mathbf{Z}_2) \end{pmatrix}$ and $\boldsymbol{\Sigma}_\rho = \begin{pmatrix} 1 & \boldsymbol{\rho}^\top \\ \boldsymbol{\rho} & \mathbf{I}_p \end{pmatrix}$ with $\boldsymbol{\rho} = \boldsymbol{\Sigma}_{vv}^{-1/2}\boldsymbol{\Sigma}_{vu}\sigma_{uu}^{-1/2}$; see the proof of Lemma 2 in the Appendix. Also, set $\mathbf{C}_2 = \mathbf{C}_1 + \boldsymbol{\Sigma}_{ww}^{-1}\boldsymbol{\Sigma}_{wz}\mathbf{C}$.

Corollary 1: *Suppose that Assumptions 1, 2 and 3 hold for a fixed T, we have,*

$$\mathbf{H}_{c(\alpha),11} \left[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right] \Rightarrow \boldsymbol{\Lambda}_{c(\alpha)},$$

where $\mathbf{H}_{c(\alpha),11}$ is the upper left corner sub-matrix of $\mathbf{H}_{c(\alpha)}$, $\boldsymbol{\Lambda}_{c(\alpha)} = \sigma_{uu}^{1/2}\boldsymbol{\Sigma}_{vv}^{-1/2} \left[\mathbf{A}_{c(\alpha)}^\top \mathbf{A}_{c(\alpha)} \right]^{-1} \mathbf{A}_{c(\alpha)}^\top \mathbf{Z}_1$, and $\mathbf{A}_{c(\alpha)} = I(c(\alpha) \neq 3)\boldsymbol{\lambda} + I(c(\alpha) \neq 2)\mathbf{Z}_2$.

Corollary 2: *Under Assumptions 1, 2 and 3, for a fixed T, we have*

$$N^{1/2} \left[\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \right] \Rightarrow \boldsymbol{\Gamma}_{c(\alpha)},$$

where $\boldsymbol{\Gamma}_{c(\alpha)} = (T-1)^{-1/2}\boldsymbol{\Sigma}_{ww}^{-1} \left[\mathbf{Z}_{wu} - I(c(\alpha) \neq 2)\mathbf{Z}_{wv}\boldsymbol{\Lambda}_{c(\alpha)} \right] - I(c(\alpha) \neq 3)\mathbf{C}_2\boldsymbol{\Lambda}_{c(\alpha)}$.

Remark 1: It follows from Corollary 2 that when T is fixed, the IV estimator for $\boldsymbol{\gamma}$ is always consistent with the same convergence rate at $N^{1/2}$ for all three cases. But the conclusion for $\widehat{\boldsymbol{\beta}}$ varies. From Corollary 1, the consistency holds only for the nearly weak case but not for other two cases. For both weak and nearly non-identified cases, $\widehat{\boldsymbol{\beta}}$ is inconsistent although it has a limiting distribution, which is not normal. Therefore, this observation is similar to that in Hahn and Kuersteiner (2002) and Chao and Swanson (2007) for cross-sectional data. However, Corollary 3 below shows that for a fixed T, the IV estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are asymptotically distributed as normal for the nearly weak case.

Corollary 3: *Suppose that Assumptions 1, 2 and 3 hold for a fixed T, we have*

$$K_N^{-1}N^{1/2} \left[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right] \Rightarrow N(\mathbf{0}, \boldsymbol{\Sigma}_{\beta,1}), \qquad and \qquad N^{1/2} \left[\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \right] \Rightarrow N(\mathbf{0}, \boldsymbol{\Sigma}_{\gamma,1}),$$

where $\boldsymbol{\Sigma}_{\beta,1} = (T-1)^{-1}\sigma_{uu}(\mathbf{C}^\top\boldsymbol{\Omega}\mathbf{C})^{-1}$ and $\boldsymbol{\Sigma}_{\gamma,1} = (T-1)^{-1}\sigma_{uu} \left[\boldsymbol{\Sigma}_{ww}^{-1} + \mathbf{C}_2 (\mathbf{C}^\top\boldsymbol{\Omega}\mathbf{C})^{-1} \mathbf{C}_2^\top \right]$.

To get insights about the asymptotic bias terms for the weak and nearly non-identified cases, next we compute the means of $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_3$ for univariate case ($p = 1$) as in Theorem

1, stated in Theorem 2 below. When $p = 1$, $\boldsymbol{\Lambda}_1$, $\boldsymbol{\Sigma}_{vv}$, and $\boldsymbol{\rho}$ become scalar. Then, we use Λ_1, Λ_3, Σ_{vv} and ρ instead of $\boldsymbol{\Lambda}_1$, $\boldsymbol{\Lambda}_3$, $\boldsymbol{\Sigma}_{vv}$, and $\boldsymbol{\rho}$.

Theorem 2: *If $p = 1$, then,*

$$E[\boldsymbol{\Lambda}_1] = q\,\rho\,(T-1)^{-1}\sigma_{uu}^{1/2}\Sigma_{vv}^{1/2}\left(\mathbf{C}^\top\boldsymbol{\Omega}\mathbf{C}\right)^{-1} + o(T^{-1}), \quad and \quad E[\boldsymbol{\Lambda}_3] = \sigma_{uu}^{1/2}\Sigma_{vv}^{-1/2}\rho,$$

where $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_3$ are defined in Corollary 1.

Remark 2: Note that a result similar to Theorem 2 was obtained by Hahn and Kuersteiner (2002) and Chao and Swanson (2007) for cross-sectional data. Under some regularity conditions, $E(\Lambda_1)$ and $E(\Lambda_3)$ can be regarded as the *asymptotic bias* of the IV estimator for $\boldsymbol{\beta}$. Also, we can see from Corollary 3 that the the asymptotic bias in $E[\Lambda_1]$ is proportional to T^{-1} and it becomes smaller for larger T although fixed. Finally, we conjecture that the result in Theorem 2 would be true for $p > 1$. Of course, it deserves a further investigation in a future study.

3.2 Large N and Large T

Now we consider the asymptotic normality in the joint limit theory in which both N and T go to infinity simultaneously. The asymptotic distribution of the estimator $\widehat{\boldsymbol{\theta}}$ for both N and $T \to \infty$ is stated in the following theorem with its proof presented in the Appendix.

Theorem 3: *Under Assumptions 1, 2 and 3, for both the nearly weak and weak cases, we have*

$$T^{1/2}\mathbf{H}_1\left[\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right] \;\Rightarrow\; (\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}_0^{-1}\mathbf{Z}_u = N(\mathbf{0},\,\boldsymbol{\Sigma}_{\theta,2}), \tag{5}$$

where $\boldsymbol{\Sigma}_{\theta,2} = \sigma_{uu}(\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Delta})^{-1}$. Further, for the nearly non-identified case, if $K_N/\sqrt{NT} \to 0$, as N and $T \to \infty$, then, (5) holds.

It is clear that Theorem 3 gives the asymptotic normality of the IV estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively, stated in Corollary 4 below.

Corollary 4: *Under Assumptions 1, 2 and 3, for both the weak and nearly weak cases, we have*

$$T^{1/2}N^{1/2}K_N^{-1}\left[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right] \;\Rightarrow\; N(\mathbf{0},\boldsymbol{\Sigma}_{\beta,2}), \quad and \quad T^{1/2}N^{1/2}\left[\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right] \;\Rightarrow\; N(\mathbf{0},\boldsymbol{\Sigma}_{\gamma,2}) \tag{6}$$

where $\boldsymbol{\Sigma}_{\beta,2} = \sigma_{uu}(\mathbf{C}^\top\boldsymbol{\Omega}\mathbf{C})^{-1}$ and $\boldsymbol{\Sigma}_{\gamma,2} = \sigma_{uu}\left[\boldsymbol{\Sigma}_{ww}^{-1} + \mathbf{C}_2\left(\mathbf{C}^\top\boldsymbol{\Omega}\mathbf{C}\right)^{-1}\mathbf{C}_2^\top\right]$. Further, for the nearly non-identified case, if $K_N/\sqrt{NT} \to 0$, as N and $T \to \infty$, then, (6) holds.

The consequences of Corollary 3 are as follows. When both N and T go to infinity simultaneously, both $\widehat{\beta}$ and $\widehat{\gamma}$ are always consistent and asymptotically normally distributed. The convergence rates are same for $\widehat{\gamma}$ with a conventional rate at $N^{1/2}T^{1/2}$ for all three cases but the convergence rates are different for $\widehat{\beta}$ and change with the degree of weakness.

4 Monte Carlo Simulations

In this section we report the results based on Monte Carlo simulations to examine the finite sample performances of the IV estimators.

We consider the following data generating model:

$$y_{it} = -9x_{it} + 9w_{it} + \alpha_i + u_{it}, \quad x_{it} = -2K_N^{-1}z_{it} + 2K_N^{-1}w_{it} + v_{it},$$

where exogenous variable w_{it} is generated from uniform distribution $U(2,8)$, excluded instrument variable z_{it} is generated from uniform distribution $U(2,10)$, and α_i is generated from normal distribution $N(0,1)$. Finally, $u_{it} \sim N(0,1)$ and $v_{it} \sim N(0,1)$ are generated jointly from a bivariate normal with the correlation coefficient $\rho = 0.7$. Clearly, u_{it} and v_{it} are independent of z_{it} and w_{it}. But x_{it} is correlated with u_{it}, since u_{it} and v_{it} are correlated. For the choice of K_N, we consider three cases. For each K_N, fixed T and changed T are considered. The IV estimators for β ($\beta = -9$) and γ ($\gamma = 9$) are computed and 1000 replications are performed for each pair of T and N. We compute the absolute bias for both $\widehat{\beta}$ and $\widehat{\gamma}$ for each pair of T and N. For each pair of T and N, the 1000 values of absolute bias for β are plotted in Figures 1(a), 1(c), 2(a), 3(a), 3(c), 4(a), 4(d) in the form of boxplots, and the 1000 values of absolute bias for γ are plotted in Figures 1(b), 1(d), 2(b), 3(b), 3(d), 4(a), 4(d) in the form of boxplots.

Case I. The weak case: $K_N = N^{1/2}$, or $K_N^{-1}N^{1/2} = 1$. First, we fix T as $T = 50$. We consider five sample sizes: $N = 50, 150, 250, 350$, and 450. Therefore, we consider five pairs of (N, T): $(50, 50)$, $(150, 50)$, $(250, 50)$, $(350, 50)$, and $(450, 50)$. For each pair of them, the 1000 values of absolute bias for $\widehat{\beta}$ are plotted in Figure 1(a) in the form of boxplots, and the 1000 values of absolute bias for $\widehat{\gamma}$ are plotted in Figure 1(b) in the form of boxplots. We can observe that the bias for $\widehat{\beta}$ keeps almost same even N increases but the bias for $\widehat{\gamma}$ deceases even N increases. This implies that the IV estimator for β is not consistent and the IV estimator for γ is consistent.

Secondly, we consider the case that both N and T become larger, which can be regarded as the case that both N and T go to infinity simultaneously. We consider five pairs of (N, T): $N = 2T = 40, 80, 120, 160, 200$. Boxplots of the 1000 values of the absolute bias
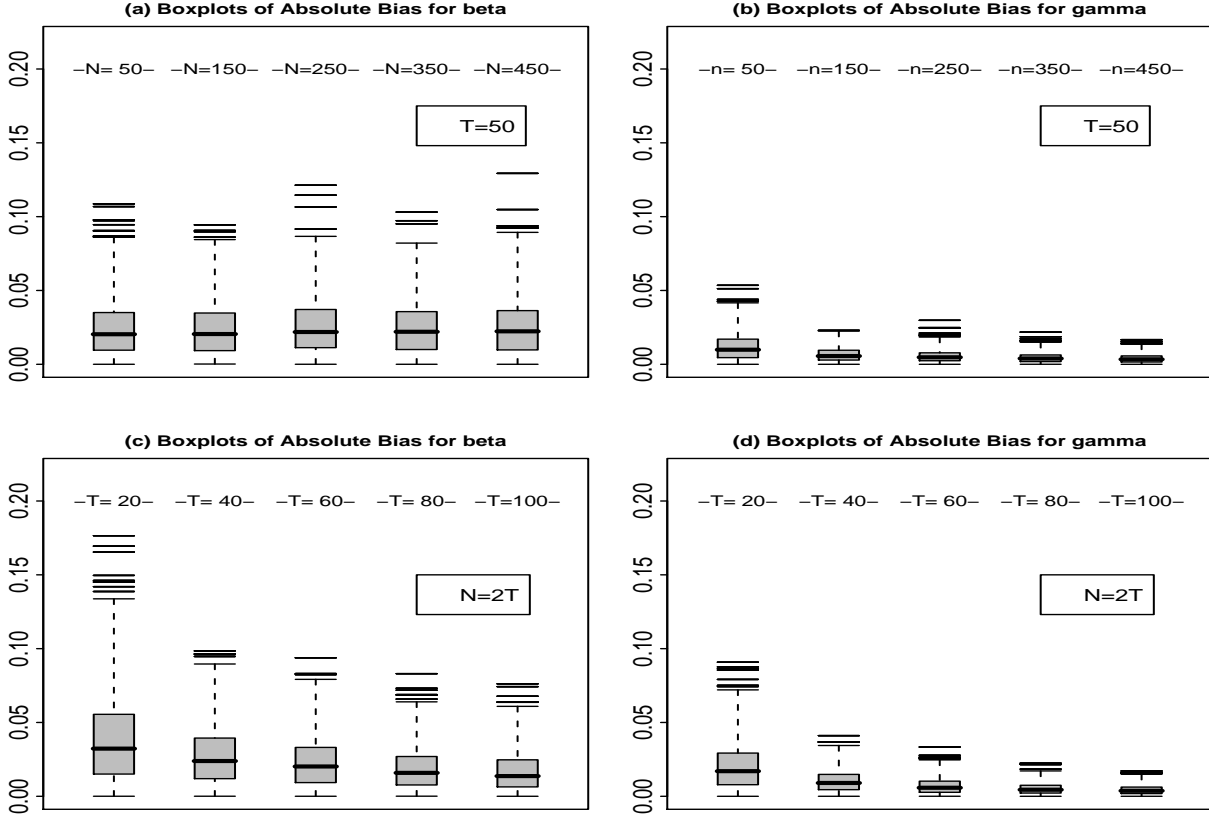
9

Figure 1: *Simulation results for Case I ($K_N = N^{1/2}$): Top panel: $T = 50$ and $N = 50$, 150, 250, 350, and 450. Bottom panel: $N = 2T = 40$, 80, 120, 160 and 200. Left panel: Boxplots of the absolute bias of the IV estimator for β. Right panel: Boxplots of the absolute bias of the IV estimator for γ.*

are displayed in Figure 1(c) for $\widehat{\beta}$ and in Figure 1(d) for $\widehat{\gamma}$. We can observe clearly from Figures 1(c) and 1(d) that both biases $\widehat{\beta}$ and $\widehat{\gamma}$ decrease as T increases. This is in the line with our asymptotic theory.

Finally, we fix a large N as $N = 250$ and then consider five values for T as 25, 50, 100, 150 and 200. Figure 2 summarizes boxplots of the absolute bias for $\widehat{\beta}$ in Figure 2(a) and for $\widehat{\gamma}$ in Figure 2(b). We can observe particularly from Figure 2(a) that the bias $\widehat{\beta}$ decreases as T increases. This observation supports the theory that when T increases, the asymptotic bias should become smaller.

Case II. The nearly weak case: $K_N = N^{0.2}$, or $K_N^{-1}N^{1/2} = N^{0.3}$. The settings are same as those in Case I. Figure 3 summarizes boxplots of the absolute bias for $\widehat{\beta}$ and $\widehat{\gamma}$. We can observe from Figure 3 that all biases for $\widehat{\beta}$ and $\widehat{\gamma}$ decrease as N increases no matter what T is small or large. Therefore, the IV estimators for β and for γ are always consistent regardless of small T or large T.
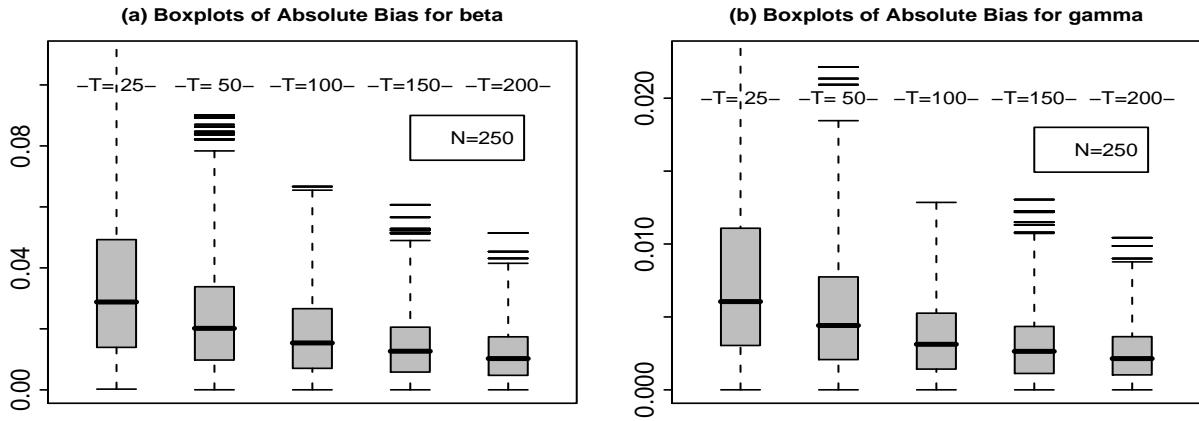
Figure 2: Simulation results for Case I with $N = 250$ and $T = 25, 50, 100, 150,$ and 200. (a) Boxplots of the absolute bias of the IV estimator for β. (b) Boxplots of the absolute bias of the IV estimator for γ.
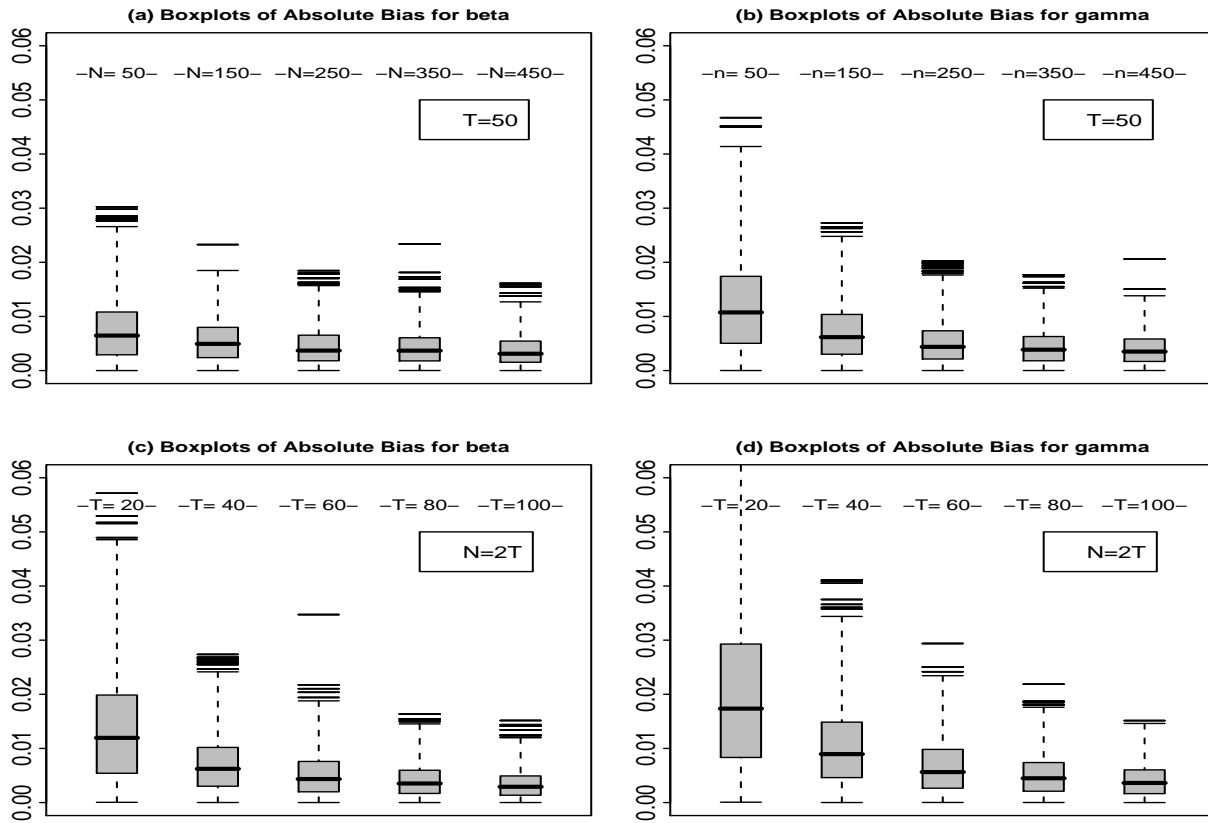

Figure 3: *Simulation results for Case II ($K_N = N^{0.2}$). Caption is the same as in Figure 1.*
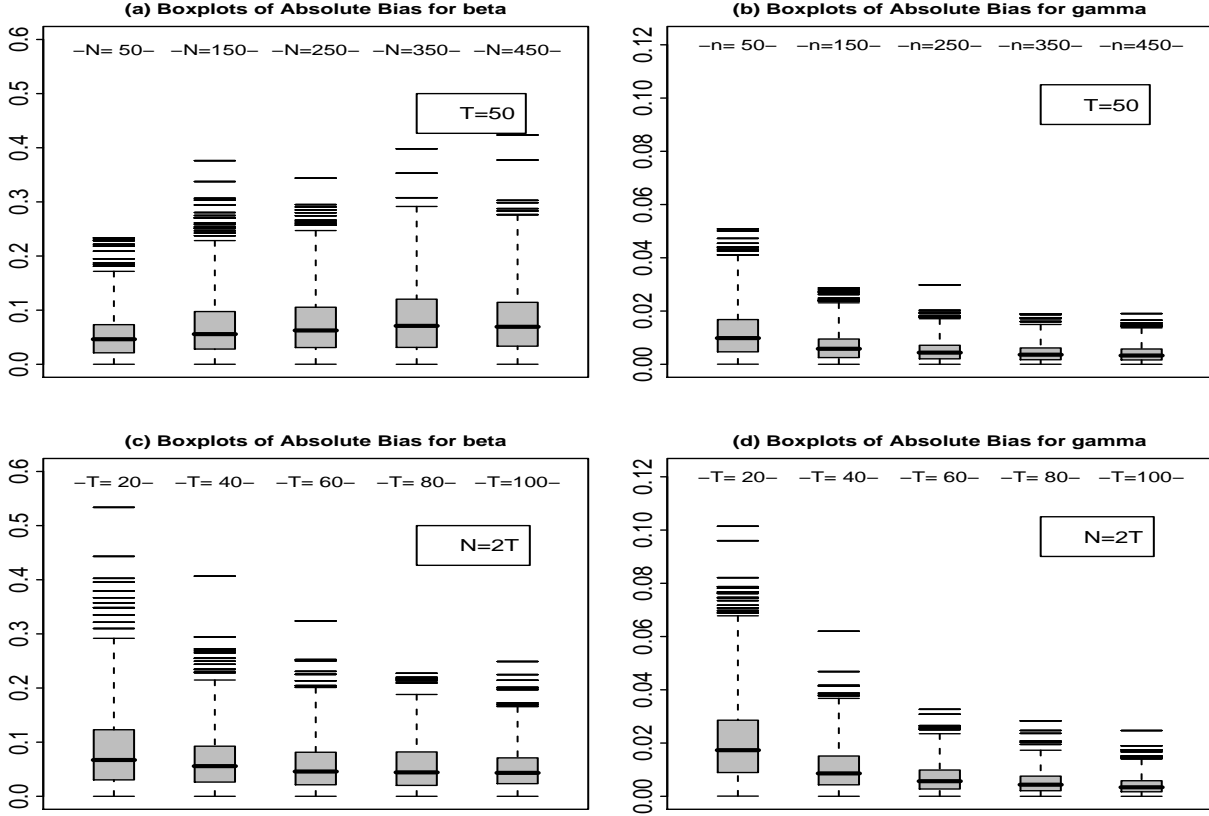
Figure 4: Simulation results for Case III ($K_N = N^{0.7}$). Caption is the same as in Figure 1.

Case III. The nearly non-identified case: $K_N = N^{0.7}$, or $K_N^{-1}N^{1/2} = N^{-0.2}$. The settings are same as those in Case I. Figure 4 gives boxplots of the absolute bias for $\widehat{\beta}$ and $\widehat{\gamma}$. It is clear from Figures 4(a) and 4(b) [large N and small T] that the IV estimator for β is not consistent and the IV estimator for γ is consistent. We can conclude from Figures 4(c) and 4(d) [large N and large T] that both $\widehat{\beta}$ and $\widehat{\gamma}$ are consistent.

Appendix

We use the same notations as introduced in Sections 2 and 3. It follows from (4) that

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \left[\widetilde{\mathbf{X}}^{*\top} \widetilde{\mathbf{Z}}^* \left(\widetilde{\mathbf{Z}}^{*\top} \widetilde{\mathbf{Z}}^* \right)^{-1} \widetilde{\mathbf{Z}}^{*\top} \widetilde{\mathbf{X}}^* \right]^{-1} \widetilde{\mathbf{X}}^{*\top} \widetilde{\mathbf{Z}}^* \left(\widetilde{\mathbf{Z}}^{*\top} \widetilde{\mathbf{Z}}^* \right)^{-1} \widetilde{\mathbf{Z}}^{*\top} \widetilde{\mathbf{U}}.$$

By observing the above form, as T is fixed, to analyze the asymptotic distribution of $\boldsymbol{\theta}$, we need to consider three components: $N^{-1}\widetilde{\mathbf{X}}^{*\top} \widetilde{\mathbf{Z}}^*$, $N^{-1}\widetilde{\mathbf{Z}}^{*\top} \widetilde{\mathbf{Z}}^*$, and $N^{-1/2}\widetilde{\mathbf{Z}}^{*\top} \widetilde{\mathbf{U}}$. To this end, some preliminary asymptotic results are given in the following two lemmas.

Lemma 1: *Suppose that Assumption 2 holds for a fixed T, then $N^{-1}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{Z}}$, $N^{-1}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{Z}}$, and*

$N^{-1}\widetilde{\mathbf{W}}^\top\widetilde{\mathbf{W}}$ *converge to* $(T-1)\boldsymbol{\Sigma}_{zz}$, $(T-1)\boldsymbol{\Sigma}_{wz}$, *and* $(T-1)\boldsymbol{\Sigma}_{ww}$ *in probability, respectively.*

Proof of Lemma 1: Since $\{\mathbf{z}_{it}\}$ are iid across both individuals and time, so are $\{\mathbf{z}_i^\top\mathbf{Q}\mathbf{z}_i\}$. By the law of large numbers,

$$
\begin{aligned}
N^{-1}\widetilde{\mathbf{Z}}^\top\widetilde{\mathbf{Z}} &= N^{-1}\sum_{i=1}^{N}\mathbf{z}_i^\top\mathbf{Q}^\top\mathbf{Q}\mathbf{z}_i \to^p E(\mathbf{z}_i^\top\mathbf{Q}\mathbf{z}_i) \\
&= TE(\mathbf{z}_{it}\mathbf{z}_{it}^\top) - \left[E(\mathbf{z}_{it}\mathbf{z}_{it}^\top) + (T-1)E(\mathbf{z}_{it})E(\mathbf{z}_{it}^\top)\right] = (T-1)\boldsymbol{\Sigma}_{zz}.
\end{aligned}
$$

Similarly, $N^{-1}\widetilde{\mathbf{W}}^\top\widetilde{\mathbf{Z}} \to^p (T-1)\boldsymbol{\Sigma}_{wz}$ and $N^{-1}\widetilde{\mathbf{W}}^\top\widetilde{\mathbf{W}} \to^p (T-1)\boldsymbol{\Sigma}_{ww}$. The proof of Lemma 1 is complete. Q.E.D

Lemma 2: *Under Assumptions 2 and 3, for a fixed T,*

$$
N^{-1/2}(\widetilde{\mathbf{W}}^\top\widetilde{\mathbf{U}}, \widetilde{\mathbf{Z}}^\top\widetilde{\mathbf{U}}, \widetilde{\mathbf{W}}^\top\widetilde{\mathbf{V}}, \widetilde{\mathbf{Z}}^\top\widetilde{\mathbf{V}}) \Rightarrow (T-1)^{1/2}(\mathbf{Z}_{wu}, \mathbf{Z}_{zu}, \mathbf{Z}_{wv}, \mathbf{Z}_{zv}). \tag{A.1}
$$

and $\mathbf{Z}_{12} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\rho \otimes \mathbf{I}_q)$.

Proof of Lemma 2: We reexpress the left hand side of (A.1) into a vector form as

$$
\mathrm{Vec}\left(N^{-1/2}(\widetilde{\mathbf{W}}^\top\widetilde{\mathbf{U}}, \widetilde{\mathbf{Z}}^\top\widetilde{\mathbf{U}}, \widetilde{\mathbf{W}}^\top\widetilde{\mathbf{V}}, \widetilde{\mathbf{Z}}^\top\widetilde{\mathbf{V}})\right) = N^{-1/2}\sum_{i=1}^{N}\boldsymbol{\xi}_i,
$$

where

$$
\boldsymbol{\xi}_i = \begin{pmatrix}\boldsymbol{\xi}_{i1}\\\boldsymbol{\xi}_{i2}\end{pmatrix} \quad \text{with} \quad \boldsymbol{\xi}_{i1} = \begin{pmatrix}\mathbf{w}_i^\top\mathbf{Q}\mathbf{u}_i\\\mathbf{z}_i^\top\mathbf{Q}\mathbf{u}_i\end{pmatrix} \quad \text{and} \quad \boldsymbol{\xi}_{i2} = \mathrm{Vec}\begin{pmatrix}\mathbf{w}_i^\top\mathbf{Q}\mathbf{v}_i\\\mathbf{z}_i^\top\mathbf{Q}\mathbf{v}_i\end{pmatrix}.
$$

Clearly, it follows from Assumptions 2 and 3 that $E[\boldsymbol{\xi}_i] = \mathbf{0}$. Now, we need to calculate the covariance matrix of $\boldsymbol{\xi}_i$. Denote by

$$
\mathrm{Var}(\boldsymbol{\xi}_i) = \begin{pmatrix}\mathbf{A}_1 & \mathbf{A}_2\\\mathbf{A}_3 & \mathbf{A}_4\end{pmatrix},
$$

where $\mathbf{A}_1 = \mathrm{Var}(\boldsymbol{\xi}_{i1})$, $\mathbf{A}_2 = \mathrm{Cov}(\boldsymbol{\xi}_{i1}, \boldsymbol{\xi}_{i2}) = \mathbf{A}_3^\top$ and $\mathbf{A}_4 = \mathrm{Var}(\boldsymbol{\xi}_{i2})$. First, we calculate $\mathbf{A}_1$.

$$
\mathbf{A}_1 = \begin{pmatrix}\mathbf{B}_1 & \mathbf{B}_2\\\mathbf{B}_3 & \mathbf{B}_4\end{pmatrix},
$$

where $\mathbf{B}_1 = \mathrm{Var}(\mathbf{w}_i^\top\mathbf{Q}\mathbf{u}_i)$, $\mathbf{B}_2 = \mathrm{Cov}(\mathbf{w}_i^\top\mathbf{Q}\mathbf{u}_i, \mathbf{z}_i^\top\mathbf{Q}\mathbf{u}_i)$, $\mathbf{B}_3 = \mathbf{B}_2^\top$, and $\mathbf{B}_4 = \mathrm{Var}(\mathbf{z}_i^\top\mathbf{Q}\mathbf{u}_i)$. By Assumptions 2 and 3,

$$
\mathbf{B}_4 = \mathrm{Var}\left[\sum_{t=1}^{T}\left(\mathbf{z}_{it} - T^{-1}\sum_{s=1}^{T}\mathbf{z}_{is}\right)u_{it}\right] = \sigma_{uu}(T-1)\boldsymbol{\Sigma}_{zz}.
$$

Similarly, $\mathbf{B}_1 = \sigma_{uu}(T-1)\boldsymbol{\Sigma}_{ww}$, and $\mathbf{B}_2 = \sigma_{uu}(T-1)\boldsymbol{\Sigma}_{wz}$. Therefore, $\mathbf{A}_1 = (T-1)\sigma_{uu} \otimes \boldsymbol{\Sigma}_0$, $\mathbf{A}_4 = (T-1)\boldsymbol{\Sigma}_{vv} \otimes \boldsymbol{\Sigma}_0$, and $\mathbf{A}_2 = (T-1)\boldsymbol{\Sigma}_{uv} \otimes \boldsymbol{\Sigma}_0$. Thus, $\mathrm{Var}(\boldsymbol{\xi}_i) = (T-1)\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}_0$. Since $\{(\mathbf{w}_{it}^\top, \mathbf{z}_{it}^\top)^\top\}$ and $\{(u_{it}, v_{it})\}$ are iid across both individuals and time, then so are $\{\boldsymbol{\xi}_i\}$. It follows from the central limit theorem that

$$N^{-1/2}\sum_{i=1}^N \boldsymbol{\xi}_i \;\Rightarrow\; \begin{pmatrix} \mathbf{Z}_{wu,T} \\ \mathbf{Z}_{zu,T} \\ \mathrm{Vec}\begin{pmatrix} \mathbf{Z}_{wv,T} \\ \mathbf{Z}_{zv,T} \end{pmatrix} \end{pmatrix} \sim N(\mathbf{0},\; (T-1)\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}_0).$$

Then,

$$N^{-1/2}(\widetilde{\mathbf{W}}^\top\widetilde{\mathbf{U}}, \widetilde{\mathbf{Z}}^\top\widetilde{\mathbf{U}}, \widetilde{\mathbf{W}}^\top\widetilde{\mathbf{V}}, \widetilde{\mathbf{Z}}^\top\widetilde{\mathbf{V}}) \;\Rightarrow\; (\mathbf{Z}_{wu,T}, \mathbf{Z}_{zu,T}, \mathbf{Z}_{wv,T}, \mathbf{Z}_{zv,T}).$$

Now we use notations $\mathbf{Z}_{wu} = (T-1)^{-1/2}\mathbf{Z}_{wu,T}$, $\mathbf{Z}_{zu} = (T-1)^{-1/2}\mathbf{Z}_{zu,T}$, $\mathbf{Z}_{wv} = (T-1)^{-1/2}\mathbf{Z}_{wv,T}$, and $\mathbf{Z}_{zv} = (T-1)^{-1/2}\mathbf{Z}_{zv,T}$ to conclude that $(\mathbf{Z}_u^\top, (\mathrm{Vec}(\mathbf{Z}_v))^\top)^\top \sim N(\mathbf{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}_0)$. Therefore, (A.1) holds. Next, we show that $(\mathbf{Z}_1^\top, (\mathrm{Vec}(\mathbf{Z}_2))^\top)^\top$ is distributed $N(\mathbf{0}, \boldsymbol{\Sigma}_\rho \otimes \mathbf{I}_q)$. For this purpose, define $\mathbf{A}_0 = \boldsymbol{\Omega}^{-1/2}\left(-\boldsymbol{\Sigma}_{zw}\boldsymbol{\Sigma}_{ww}^{-1} \quad \mathbf{I}_q\right)$. Then, $\mathbf{Z}_1 = \mathbf{A}_0\mathbf{Z}_u\sigma_{uu}^{-1/2}$, $\mathbf{Z}_2 = \mathbf{A}_0\mathbf{Z}_v\boldsymbol{\Sigma}_{vv}^{-1/2}$, $\mathrm{Vec}(\mathbf{Z}_2) = \boldsymbol{\Sigma}_{vv}^{-1/2} \otimes \mathbf{A}_0\mathrm{Vec}(\mathbf{Z}_v)$, and $\mathbf{A}_0\boldsymbol{\Sigma}_0\mathbf{A}_0^\top = \mathbf{I}_q$. Now we calculate the mean and covariance matrix of $\mathbf{Z}_{12}$. Clearly,

$$E[\mathbf{Z}_{12}] = \begin{pmatrix} \mathbf{A}_0 E(\mathbf{Z}_u)\sigma_{uu}^{-1/2} \\ \boldsymbol{\Sigma}_{vv}^{-1/2} \otimes \mathbf{A}_0 E(\mathrm{Vec}(\mathbf{Z}_v)) \end{pmatrix} = \mathbf{0}, \quad \text{and} \quad \mathrm{Var}(\mathbf{Z}_{12}) = \begin{pmatrix} \mathbf{D}_1 & \mathbf{D}_2 \\ \mathbf{D}_3 & \mathbf{D}_4 \end{pmatrix},$$

where $\mathbf{D}_1 = \mathrm{Var}(\mathbf{Z}_1)$, $\mathbf{D}_3 = \mathbf{D}_2^\top = \mathrm{Cov}(\mathrm{Vec}(\mathbf{Z}_2), \mathbf{Z}_1)$, and $\mathbf{D}_4 = \mathrm{Cov}(\mathrm{Vec}(\mathbf{Z}_2))$. Since

$$\mathbf{D}_1 = \mathrm{Cov}(\mathbf{A}_0\mathbf{Z}_u\sigma_{uu}^{-1/2}) = \sigma_{uu}^{-1}\mathbf{A}_0\mathrm{Var}(\mathbf{Z}_u)\mathbf{A}_0^\top = \sigma_{uu}^{-1}\mathbf{A}_0\sigma_{uu}\boldsymbol{\Sigma}_0\mathbf{A}_0^\top = \mathbf{A}_0\boldsymbol{\Sigma}_0\mathbf{A}_0^\top = \mathbf{I}_q,$$

$$\begin{aligned} \mathbf{D}_3 &= E((\boldsymbol{\Sigma}_{vv}^{-1/2} \otimes \mathbf{A}_0\mathrm{Vec}(\mathbf{Z}_v))(\mathbf{A}_0\mathbf{Z}_u\sigma_{uu}^{-1/2})^\top) = (\boldsymbol{\Sigma}_{vv}^{-1/2} \otimes \mathbf{A}_0)\sigma_{uu}^{-1/2}E((\mathrm{Vec}(\mathbf{Z}_v))\mathbf{Z}_u^\top)\mathbf{A}_0^\top \\ &= (\boldsymbol{\Sigma}_{vv}^{-1/2} \otimes \mathbf{A}_0)\sigma_{uu}^{-1/2}(\boldsymbol{\Sigma}_{vu} \otimes \boldsymbol{\Sigma}_0)\mathbf{A}_0^\top = \boldsymbol{\Sigma}_{vv}^{-1/2}\sigma_{uu}^{-1/2}\boldsymbol{\Sigma}_{vu} \otimes \mathbf{A}_0\boldsymbol{\Sigma}_0\mathbf{A}_0^\top \\ &= \boldsymbol{\Sigma}_{vv}^{-1/2}\sigma_{uu}^{-1/2}\boldsymbol{\Sigma}_{vu} \otimes \mathbf{I}_q = \boldsymbol{\rho} \otimes \mathbf{I}_q, \end{aligned}$$

and

$$\begin{aligned} \mathbf{D}_4 &= \boldsymbol{\Sigma}_{vv}^{-1/2} \otimes \mathbf{A}_0\mathrm{Var}(\mathrm{Vec}(\mathbf{Z}_v))(\boldsymbol{\Sigma}_{vv}^{-1/2} \otimes \mathbf{A}_0)^\top = (\boldsymbol{\Sigma}_{vv}^{-1/2} \otimes \mathbf{A}_0)(\boldsymbol{\Sigma}_{vv} \otimes \boldsymbol{\Sigma}_0)\boldsymbol{\Sigma}_{vv}^{-1/2} \otimes \mathbf{A}_0^\top \\ &= \boldsymbol{\Sigma}_{vv}^{-1/2}\boldsymbol{\Sigma}_{vv}\boldsymbol{\Sigma}_{vv}^{-1/2} \otimes \mathbf{A}_0\boldsymbol{\Sigma}_0\mathbf{A}_0^\top = \mathbf{I}_p \otimes \mathbf{I}_q, \end{aligned}$$

then, $\mathrm{Var}(\mathbf{Z}_{12}) = \boldsymbol{\Sigma}_\rho \otimes \mathbf{I}_q$. Thus, $\mathbf{Z}_{12}$ is distributed as $N(\mathbf{0}, \boldsymbol{\Sigma}_\rho \otimes \mathbf{I}_q)$. This proves Lemma 2. Q.E.D

Proof of Theorem 1: By Lemma 1, one has

$$N^{-1}\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{Z}}^* = \begin{pmatrix} N^{-1}\widetilde{\mathbf{W}}^\top\widetilde{\mathbf{W}} & N^{-1}\widetilde{\mathbf{W}}^\top\widetilde{\mathbf{Z}} \\ N^{-1}\widetilde{\mathbf{Z}}^\top\widetilde{\mathbf{W}} & N^{-1}\widetilde{\mathbf{Z}}^\top\widetilde{\mathbf{Z}} \end{pmatrix} \to^p (T-1)\boldsymbol{\Sigma}_0.$$

By Assumption 1,

$$\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*} = \begin{pmatrix} K_N^{-1}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{Z}}\mathbf{C} + K_N^{-1}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{W}}\mathbf{C}_1 + \widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{V}} & \widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{W}} \\ K_N^{-1}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{Z}}\mathbf{C} + K_N^{-1}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{W}}\mathbf{C}_1 + \widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{V}} & \widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{W}} \end{pmatrix},$$

so that

$$\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_1^{-1}N^{-1/2} = \begin{pmatrix} K_N^{-1}N^{-1/2}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{Z}}\mathbf{C} + K_N^{-1}N^{-1/2}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{W}}\mathbf{C}_1 + N^{-1/2}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{V}} & N^{-1}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{W}} \\ K_N^{-1}N^{-1/2}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{Z}}\mathbf{C} + K_N^{-1}N^{-1/2}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{W}}\mathbf{C}_1 + N^{-1/2}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{V}} & N^{-1}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{W}} \end{pmatrix},$$

and

$$\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_2^{-1}N^{-1/2} = \begin{pmatrix} N^{-1}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{Z}}\mathbf{C} + N^{-1}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{W}}\mathbf{C}_1 + K_N N^{-1}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{V}} & N^{-1}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{W}} \\ N^{-1}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{Z}}\mathbf{C} + N^{-1}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{W}}\mathbf{C}_1 + K_N N^{-1}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{V}} & N^{-1}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{W}} \end{pmatrix}.$$

First, we prove the assertion in (a). Since $K_N^{-1}N^{1/2} \to 1$, as $N \to \infty$, it follows from Lemmas 1 and 2 that

$$\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_1^{-1}N^{-1/2} \Rightarrow (T-1)^{1/2}\boldsymbol{\Delta}_1,$$

so that

$$\begin{aligned} \mathbf{H}_1(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) &= \left[(\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_1^{-1}N^{-1/2})^{\top}(N^{-1}\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{Z}}^{*})^{-1}(\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_1^{-1}N^{-1/2})\right]^{-1} \\ &\quad (\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_1^{-1}N^{-1/2})^{\top}(N^{-1}\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{Z}}^{*})^{-1}(N^{-1/2}\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{U}}) \\ &\Rightarrow (\boldsymbol{\Delta}_1^{\top}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Delta}_1)^{-1}\boldsymbol{\Delta}_1^{\top}\boldsymbol{\Sigma}_0^{-1}\mathbf{Z}_u. \end{aligned}$$

Second, we establish (b). Since $K_N^{-1}N^{1/2} \to \infty$, as $N \to \infty$, one can show easily from Lemmas 1 and 2 that

$$\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_2^{-1}N^{-1/2} \Rightarrow (T-1)^{1/2}\boldsymbol{\Delta}_2,$$

which implies that

$$\begin{aligned} \mathbf{H}_2(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) &= \left[(\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_2^{-1}N^{-1/2})^{\top}(N^{-1}\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{Z}}^{*})^{-1}(\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_2^{-1}N^{-1/2})\right]^{-1} \\ &\quad (\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_2^{-1}N^{-1/2})^{\top}(N^{-1}\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{Z}}^{*})^{-1}(N^{-1/2}\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{U}}) \\ &\Rightarrow (\boldsymbol{\Delta}_2^{\top}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Delta}_2)^{-1}\boldsymbol{\Delta}_2^{\top}\boldsymbol{\Sigma}_0^{-1}\mathbf{Z}_u. \end{aligned}$$

Finally, it suffices to show that (c) holds. Since $K_N^{-1}N^{1/2} \to 0$, as $N \to \infty$, by Lemmas 1 and 2, one has

$$\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_1^{-1}N^{-1/2} \Rightarrow (T-1)^{1/2}\boldsymbol{\Delta}_2.$$

Then,

$$\mathbf{H}_1(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \Rightarrow (\boldsymbol{\Delta}_3^{\top}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Delta}_3)^{-1}\boldsymbol{\Delta}_3^{\top}\boldsymbol{\Sigma}_0^{-1}\mathbf{Z}_u.$$

Theorem 1 is proved. Q.E.D

Proof of Corollaries 1 and 2: First, we prove the conclusion in part (a). It is clear from Theorem 1(a) that we need to calculate the two diagonal sub-matrices in $(\boldsymbol{\Delta}_1^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Delta}_1)^{-1}$. By the inverse of a partitioned matrix,

$$\boldsymbol{\Sigma}_0^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{ww}^{-1} + \boldsymbol{\Sigma}_{ww}^{-1}\boldsymbol{\Sigma}_{wz}\boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}_{zw}\boldsymbol{\Sigma}_{ww}^{-1} & -\boldsymbol{\Sigma}_{ww}^{-1}\boldsymbol{\Sigma}_{wz}\boldsymbol{\Omega}^{-1} \\ -\boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}_{zw}\boldsymbol{\Sigma}_{ww}^{-1} & \boldsymbol{\Omega}^{-1} \end{pmatrix},$$

then $\boldsymbol{\Delta}_1^\top \boldsymbol{\Sigma}_0^{-1}$ can be written as

$$\boldsymbol{\Delta}_1^\top \boldsymbol{\Sigma}_0^{-1} = (T-1)^{1/2} \begin{pmatrix} \mathbf{G}_1 & \mathbf{G}_2 \\ \mathbf{I}_k & \mathbf{0} \end{pmatrix},$$

where $\mathbf{G}_1 = \mathbf{C}_1^\top + (T-1)^{-1/2}\mathbf{Z}_{wv}^\top \boldsymbol{\Sigma}_{ww}^{-1} + (T-1)^{-1/2}\mathbf{Z}_{wv}^\top \boldsymbol{\Sigma}_{ww}^{-1}\boldsymbol{\Sigma}_{wz}\boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}_{zw}\boldsymbol{\Sigma}_{ww}^{-1} - (T-1)^{-1/2}\mathbf{Z}_{zv}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}_{zw}\boldsymbol{\Sigma}_{ww}^{-1}$, and $\mathbf{G}_2 = \mathbf{C}^\top + (T-1)^{-1/2}\mathbf{Z}_{zv}^\top \boldsymbol{\Omega}^{-1} - (T-1)^{-1/2}\mathbf{Z}_{wv}^\top \boldsymbol{\Sigma}_{ww}^{-1}\boldsymbol{\Sigma}_{wz}\boldsymbol{\Omega}^{-1}$. Thus, $\boldsymbol{\Delta}_1^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Delta}_1$ can be expressed as

$$(T-1) \begin{pmatrix} \mathbf{G}_3 & \mathbf{C}_1^\top \boldsymbol{\Sigma}_{ww} + \mathbf{C}^\top \boldsymbol{\Sigma}_{zw} + (T-1)^{-1/2}\mathbf{Z}_{wv}^\top \\ \boldsymbol{\Sigma}_{ww}\mathbf{C}_1 + \boldsymbol{\Sigma}_{wz}\mathbf{C} + (T-1)^{-1/2}\mathbf{Z}_{wv} & \boldsymbol{\Sigma}_{ww} \end{pmatrix},$$

where $\mathbf{G}_3 = (T-1)^{-1/2}\mathbf{Z}_{zv}^\top \mathbf{C} + (T-1)^{-1/2}\mathbf{C}^\top \mathbf{Z}_{zv} + (T-1)^{-1/2}\mathbf{Z}_{wv}^\top \mathbf{C}_1 + (T-1)^{-1/2}\mathbf{C}_1^\top \mathbf{Z}_{wv} + \mathbf{C}_1^\top \boldsymbol{\Sigma}_{wz}\mathbf{C} + \mathbf{C}^\top \boldsymbol{\Sigma}_{zw}\mathbf{C}_1 + \mathbf{C}_1^\top \boldsymbol{\Sigma}_{ww}\mathbf{C}_1 + \mathbf{C}^\top \boldsymbol{\Sigma}_{zz}\mathbf{C} + (T-1)^{-1}\mathbf{Z}_{wv}^\top \boldsymbol{\Sigma}_{ww}^{-1}\mathbf{Z}_{wv} + (T-1)^{-1}\mathbf{Z}_{wv}^\top \boldsymbol{\Sigma}_{ww}^{-1}\boldsymbol{\Sigma}_{wz}\boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}_{zw}\boldsymbol{\Sigma}_{ww}^{-1}\mathbf{Z}_{wv} - (T-1)^{-1}\mathbf{Z}_{zv}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}_{zw}\boldsymbol{\Sigma}_{ww}^{-1}\mathbf{Z}_{wv} - (T-1)^{-1}\mathbf{Z}_{wv}^\top \boldsymbol{\Sigma}_{ww}^{-1}\boldsymbol{\Sigma}_{wz}\boldsymbol{\Omega}^{-1}\mathbf{Z}_{zv} + (T-1)^{-1}\mathbf{Z}_{zv}^\top \boldsymbol{\Omega}^{-1}\mathbf{Z}_{zv}$. Hence,

$$(\boldsymbol{\Delta}_1^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Delta}_1)^{-1} \equiv (T-1)^{-1} \begin{pmatrix} \mathbf{G}_4 & \mathbf{G}_5 \\ \mathbf{G}_5^\top & \mathbf{G}_6 \end{pmatrix},$$

where $\mathbf{G}_4 = (T-1)\boldsymbol{\Sigma}_{vv}^{-1/2}((\boldsymbol{\lambda}+\mathbf{Z}_2)^\top(\boldsymbol{\lambda}+\mathbf{Z}_2))^{-1}\boldsymbol{\Sigma}_{vv}^{-1/2}$, $\mathbf{G}_5 = -\mathbf{G}_4(\mathbf{C}_1^\top \boldsymbol{\Sigma}_{ww} + \mathbf{C}^\top \boldsymbol{\Sigma}_{zw} + (T-1)^{-1/2}\mathbf{Z}_{wv}^\top)\boldsymbol{\Sigma}_{ww}^{-1}$, and $\mathbf{G}_6 = \boldsymbol{\Sigma}_{ww}^{-1} + \boldsymbol{\Sigma}_{ww}^{-1}(\boldsymbol{\Sigma}_{ww}\mathbf{C}_1 + \boldsymbol{\Sigma}_{wz}\mathbf{C} + (T-1)^{-1/2}\mathbf{Z}_{wv}) \mathbf{G}_4(\mathbf{C}_1^\top \boldsymbol{\Sigma}_{ww} + \mathbf{C}^\top \boldsymbol{\Sigma}_{zw} + (T-1)^{-1/2}\mathbf{Z}_{wv}^\top)\boldsymbol{\Sigma}_{ww}^{-1}$. Hence, we have

$$(\boldsymbol{\Delta}_1^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Delta}_1)^{-1}\boldsymbol{\Delta}_1^\top \boldsymbol{\Sigma}_0^{-1}\mathbf{Z}_u = \begin{pmatrix} \boldsymbol{\Lambda}_1 \\ \boldsymbol{\Gamma}_1 \end{pmatrix}.$$

Similarly, we can prove the conclusions in parts (b) and (c). The proof of Corollaries 1 and 2 is finished. Q.E.D

Proof of Corollary 3: It follows from Lemma 2 that $\mathbf{Z}_1 \sim N(0, \mathbf{I}_q)$. Then, it is easy to verify from Corollary 1 (b) that $K_N^{-1}N^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow N(\mathbf{0}, \boldsymbol{\Sigma}_{\beta,1})$, where

$$\boldsymbol{\Sigma}_{\beta,1} = \sigma_{uu}^{1/2}\boldsymbol{\Sigma}_{vv}^{-1/2}(\boldsymbol{\lambda}^\top \boldsymbol{\lambda})^{-1}\boldsymbol{\lambda}^\top \mathbf{I}_q \boldsymbol{\lambda}(\boldsymbol{\lambda}^\top \boldsymbol{\lambda})^{-1}\boldsymbol{\Sigma}_{vv}^{-1/2}\sigma_{uu}^{1/2} = (T-1)^{-1}\sigma_{uu}(\mathbf{C}^\top \boldsymbol{\Omega}\mathbf{C})^{-1}.$$

Now, we show that $N^{1/2}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \Rightarrow N(\mathbf{0}, \boldsymbol{\Sigma}_{\gamma,1})$. Applying Corollary 2 (b) leads to

$$(T-1)^{1/2}\boldsymbol{\Gamma}_2 = \boldsymbol{\Sigma}_{ww}^{-1}\mathbf{Z}_{wu} - (T-1)^{1/2}\,\mathbf{C}_2\,\boldsymbol{\Lambda}_2 \equiv \mathbf{A}_1\,\mathbf{Z}_u,$$

where $\mathbf{A}_1 = \left(\boldsymbol{\Sigma}_{ww}^{-1} + \mathbf{C}_2\,(\mathbf{C}^{\top}\boldsymbol{\Omega}\mathbf{C})^{-1}\mathbf{C}^{\top}\boldsymbol{\Sigma}_{zw}\boldsymbol{\Sigma}_{ww}^{-1} \quad -\mathbf{C}_2\,(\mathbf{C}^{\top}\boldsymbol{\Omega}\mathbf{C})^{-1}\mathbf{C}^{\top}\right)$. Next, we calculate the covariance matrix of $\boldsymbol{\Gamma}_2$. By Lemma 2, $\mathbf{Z}_u \sim N(\mathbf{0}, \sigma_{uu}\boldsymbol{\Sigma}_0)$, so that

$$(T-1)\mathrm{Cov}[\boldsymbol{\Gamma}_2] = \sigma_{uu}\,\mathbf{A}_1\boldsymbol{\Sigma}_0\,\mathbf{A}_1^{\top} = (T-1)\boldsymbol{\Sigma}_{\gamma,1}.$$

Thus, $N^{1/2}\,[\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}] \Rightarrow N(\mathbf{0}, \boldsymbol{\Sigma}_{\gamma,1})$. Hence, this proves Corollary 3. Q.E.D

In order to prove Theorem 2, the following three lemmas are needed. Note that the proofs of Lemmas 4 and 5 can be found in Ullah (1974) and Lebedev (1972, pp.268-271), respectively, and omitted.

Lemma 3: *If $p = 1$, then*

$$E\left[\left\{(\boldsymbol{\lambda} + \mathbf{Z}_2)^{\top}(\boldsymbol{\lambda} + \mathbf{Z}_2)\right\}^{-1}(\boldsymbol{\lambda} + \mathbf{Z}_2)^{\top}\right] = \boldsymbol{\lambda}^{\top}E\left[\left\{\chi_{q+2}^2(\delta)\right\}^{-1}\right],$$

where $\delta = \boldsymbol{\lambda}^{\top}\boldsymbol{\lambda}/2$.

Proof of Lemma 3: It is easy to see from Lemma 2 that $\boldsymbol{\lambda} + \mathbf{Z}_2 \sim N(\boldsymbol{\lambda}, \mathbf{I}_q)$. We re-write $\boldsymbol{\lambda} \equiv (\lambda_1, \ldots, \lambda_q)^{\top}$ and $\boldsymbol{\lambda} + \mathbf{Z}_2 \equiv (\zeta_1, \ldots, \zeta_q)^{\top}$. Then, $\zeta_i \sim N(\lambda_i, 1)$ and $\{\zeta_i\}$ are independent as well as $(\boldsymbol{\lambda} + \mathbf{Z}_2)^{\top}(\boldsymbol{\lambda} + \mathbf{Z}_2) = \sum_{i=1}^{q}\zeta_i^2 \sim \chi_q^2(\delta)$. Now, by applying Lemma 2 of Appendix B.1 of Judge and Bock (1978), for each $1 \leq i \leq q$, we have

$$E\left[\zeta_i/\sum_{j=1}^{q}\zeta_j^2\right] = \lambda_i\,E\left[E\left\{\left(\sum_{j\neq i}^{q}\zeta_j^2 + \chi_3^2(\lambda_i^2/2)\right)^{-1}\bigg|\zeta_j, j\neq i\right\}\right]$$

$$= \lambda_i\,E\left[\left\{\chi_{q+2}^2(\delta)\right\}^{-1}\right].$$

Then, we prove the lemma. Q.E.D

Lemma 4:
$$E\left[\left\{\chi_{q+2}^2(\delta)\right\}^{-1}\right] = 2^{-1}e^{-\delta}\frac{\Gamma(q/2)}{\Gamma(q/2+1)}H(q/2; q/2+1; \delta),$$

where $H(q/2; q/2+1; \delta)$ is the hypergeometric function defined as

$$H(a; b; z) = \sum_{k=0}^{\infty}(a)_k\,z^k(b)_k/[k!(b)_k],$$

and $(a)_k = a(a+1)...(a+k-1)$.

Lemma 5: *If $x > 0$, and a, $c > 0$, then as $x \to \infty$*

$$H(a; c; x) = \frac{\Gamma(c)}{\Gamma(a)} e^x x^{-(c-a)} \left[\sum_{j=0}^{k-1} \frac{(c-a)_j (1-a)_j}{j!} x^{-j} + O(|x|^{-k}) \right]$$

for any $k \geq 1$.

Proof of Theorem 2: From Lemma 2, $\mathbf{Z}_{12}$ is distributed as $N(\mathbf{0}, \mathbf{\Sigma}_\rho \otimes \mathbf{I}_q)$. Then, using the project of $\mathbf{Z}_1$ onto $\mathbf{Z}_2$ gives $\mathbf{Z}_1 = \boldsymbol{\rho} \mathbf{Z}_2 + \mathbf{Z}^*$, where $\mathbf{Z}^*$ is distributed as $N(\mathbf{0}, (1 - \boldsymbol{\rho}^2)\mathbf{I}_q)$ and is independent of $\mathbf{Z}_2$. Thus,

$$
\begin{aligned}
E[\mathbf{\Lambda}_1] &= \sigma_{uu}^{1/2} \mathbf{\Sigma}_{vv}^{-1/2} E \left[E((\boldsymbol{\lambda} + \mathbf{Z}_2)^\top (\boldsymbol{\lambda} + \mathbf{Z}_2))^{-1} (\boldsymbol{\lambda} + \mathbf{Z}_2)^\top (\boldsymbol{\rho} \mathbf{Z}_2 + \mathbf{Z}^*) \mid \mathbf{Z}_2) \right] \\
&= \sigma_{uu}^{1/2} \mathbf{\Sigma}_{vv}^{-1/2} E \left[((\boldsymbol{\lambda} + \mathbf{Z}_2)^\top (\boldsymbol{\lambda} + \mathbf{Z}_2))^{-1} (\boldsymbol{\lambda} + \mathbf{Z}_2)^\top \boldsymbol{\rho} \mathbf{Z}_2 \right] \\
&= \sigma_{uu}^{1/2} \mathbf{\Sigma}_{vv}^{-1/2} E \left[\boldsymbol{\rho} - ((\boldsymbol{\lambda} + \mathbf{Z}_2)^\top (\boldsymbol{\lambda} + \mathbf{Z}_2))^{-1} (\boldsymbol{\lambda} + \mathbf{Z}_2)^\top \boldsymbol{\rho} \boldsymbol{\lambda} \right],
\end{aligned}
$$

which, in conjunction with Lemmas 3 and 4, implies that

$$
\begin{aligned}
E[\mathbf{\Lambda}_1] &= \sigma_{uu}^{1/2} \mathbf{\Sigma}_{vv}^{-1/2} \boldsymbol{\rho} \left[1 - 2\delta E((\chi_{q+2}^2(\delta))^{-1}) \right] \\
&= \sigma_{uu}^{1/2} \mathbf{\Sigma}_{vv}^{-1/2} \boldsymbol{\rho} \left[1 - \delta e^{-\delta} \frac{\Gamma(q/2)}{\Gamma(q/2+1)} H(q/2; q/2+1; \delta) \right] \\
&= \sigma_{uu}^{1/2} \mathbf{\Sigma}_{vv}^{-1/2} \boldsymbol{\rho} \, e^{-\delta} H(q/2 - 1; q/2; \delta).
\end{aligned}
$$

Thus, it follows easily from Lemma 5 that

$$E[\mathbf{\Lambda}_1] = \sigma_{uu}^{1/2} \mathbf{\Sigma}_{vv}^{-1/2} \boldsymbol{\rho} \frac{q}{2} \delta^{-1} \left[1 + O(\delta^{-1}) \right] = q \, \sigma_{uu}^{-1/2} \mathbf{\Sigma}_{vv}^{1/2} \, \boldsymbol{\rho} \, \mathbf{\Sigma}_{\beta,1} + O(T^{-2}),$$

since $\delta = 2(T-1) \mathbf{\Sigma}_{vv}^{-1/2} \mathbf{C}^\top \mathbf{\Omega} \mathbf{C} \mathbf{\Sigma}_{vv}^{-1/2} = O(T)$, which concludes that the first conclusion in the theorem holds. Finally, we establish the second assertion. It is easy to see that

$$
\begin{aligned}
E(\mathbf{\Lambda}_3) &= \sigma_{uu}^{1/2} \mathbf{\Sigma}_{vv}^{-1/2} E \left[E((\mathbf{Z}_2^\top \mathbf{Z}_2)^{-1} \mathbf{Z}_2^\top (\mathbf{Z}_2 \boldsymbol{\rho} + \mathbf{Z}^*) \mid \mathbf{Z}_2) \right] = \sigma_{uu}^{1/2} \mathbf{\Sigma}_{vv}^{-1/2} E[(\mathbf{Z}_2^\top \mathbf{Z}_2)^{-1} \mathbf{Z}_2^\top \mathbf{Z}_2 \boldsymbol{\rho}] \\
&= \sigma_{uu}^{1/2} \mathbf{\Sigma}_{vv}^{-1/2} \boldsymbol{\rho}.
\end{aligned}
$$

The proof of Theorem 2 is complete. Q.E.D

When T goes to infinite, to analyze the asymptotic distribution of $\boldsymbol{\theta}$, we need to consider the following three components: $N^{-1} T^{-1} \widetilde{\mathbf{X}}^{*\top} \widetilde{\mathbf{Z}}^*$, $N^{-1} T^{-1} \widetilde{\mathbf{Z}}^{*\top} \widetilde{\mathbf{Z}}^*$, and $N^{-1/2} T^{-1/2} \widetilde{\mathbf{Z}}^{*\top} \widetilde{\mathbf{U}}$. Before embarking on the proof of Theorem 4, we need the following two lemmas.

Lemma 6: *Under Assumption 2, we have*

$$N^{-1} T^{-1} \widetilde{\mathbf{Z}}^\top \widetilde{\mathbf{Z}} \to^p \mathbf{\Sigma}_{zz}, \quad N^{-1} T^{-1} \widetilde{\mathbf{W}}^\top \widetilde{\mathbf{Z}} \to^p \mathbf{\Sigma}_{wz}, \quad and \quad N^{-1} T^{-1} \widetilde{\mathbf{W}}^\top \widetilde{\mathbf{W}} \to^p \mathbf{\Sigma}_{ww}.$$

Proof of Lemma 6: Define $\mathbf{F}_1 = N^{-1}T^{-2}(T-1)\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbf{z}_{it}\mathbf{z}_{it}^{\top} - E(\mathbf{z}_{it})E(\mathbf{z}_{it}^{\top})\right]$ and $\mathbf{F}_2 = N^{-1}T^{-2}\sum_{i=1}^{N}\left[T(T-1)\,E(\mathbf{z}_{it})E(\mathbf{z}_{it}^{\top}) - \sum_{s \neq t}\mathbf{z}_{is}\mathbf{z}_{it}^{\top}\right]$. Then, $N^{-1}T^{-1}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{Z}} = \mathbf{F}_1 + \mathbf{F}_2$. Since $\{(\mathbf{w}_{it}^{\top}, \mathbf{z}_{it}^{\top})^{\top}\}$ are iid across both individuals and time, so are $\{\mathbf{z}_{it}\mathbf{z}_{it}^{\top} - E(\mathbf{z}_{it})E(\mathbf{z}_{it}^{\top})\}$. By the law of large numbers,

$$N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbf{z}_{it}\mathbf{z}_{it}^{\top} - E(\mathbf{z}_{it})E(\mathbf{z}_{it}^{\top})\right] \to^p E(\mathbf{z}_{it}\mathbf{z}_{it}^{\top}) - E(\mathbf{z}_{it})E(\mathbf{z}_{it}^{\top}) = \mathbf{\Sigma}_{zz},$$

so that $\mathbf{F}_1 \to^p \mathbf{\Sigma}_{zz}$. It is clear that to accomplish the proof of the lemma, it suffices to show that $\mathbf{F}_2 = o_p(1)$. To this end, let F_{2,j_1j_2} denote the (j_1, j_2)th $(1 \le j_1, j_2 \le q)$ element of $\mathbf{F}_2$. Then, F_{2,j_1j_2} can be re-expressed as

$$F_{2,j_1j_2} = 2\,N^{-1}T^{-2}\sum_{i=1}^{N}\sum_{1 \le s < t \le T}\left[-z_{j_1is}z_{j_2it} + E(z_{j_1it})E(z_{j_2it})\right] = 2\,N^{-1}T^{-2}\sum_{m=1}^{NT(T-1)/2} z_{m,j_1j_2},$$

where $z_{m,j_1j_2} = -z_{j_1is}z_{j_2it} + E(z_{j_1it})E(z_{j_2it})$ for $m = (i-1)N + (T-1)(s-1) + t - s = 1, 2, \cdots, NT(T-1)/2$. Clearly, $E[z_{m,j_1j_2}] = 0$, and $\{z_{m,j_1j_2}\}_{m=1}^{NT(T-1)/2}$ are iid with the finite second moment by Assumption 2. Then, the law of large numbers implies that $F_{2,j_1j_2} = o_p(1)$. Thus, $\mathbf{F}_2 = o_p(1)$, and $N^{-1}T^{-1}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{Z}} \to^p \mathbf{\Sigma}_{zz}$. Similarly, $N^{-1}T^{-1}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{Z}} \to^p \mathbf{\Sigma}_{wz}$, and $N^{-1}T^{-1}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{W}} \to^p \mathbf{\Sigma}_{ww}$. Therefore, Lemma 6 is proved. Q.E.D

Lemma 7: *Under Assumptions 2 and 3, we have*

$$T^{-1/2}N^{-1/2}(\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{U}},\ \widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{U}}, \widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{V}},\ \widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{V}}) \Rightarrow (\mathbf{Z}_{wu}, \mathbf{Z}_{zu}, \mathbf{Z}_{wv}, \mathbf{Z}_{zv}).$$

Proof of Lemma 7: Observe that

$$N^{-1/2}T^{-1/2}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{U}} \equiv N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}[\mathbf{z}_{it} - E(\mathbf{z}_{it})]u_{it} + \mathbf{F}_3,$$

where

$$\mathbf{F}_3 = N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[E(\mathbf{z}_{it}) - T^{-1}\sum_{s=1}^{T}\mathbf{z}_{is}\right]u_{it}.$$

First, we show that $\mathbf{F}_3 = o_p(1)$. Clearly, $E(\mathbf{F}_3) = 0$, and

$$\begin{aligned}
\mathrm{Var}(\mathbf{F}_3) &= \mathrm{Var}\left[\left\{E(\mathbf{z}_{it}) - T^{-1}\sum_{s=1}^{T}\mathbf{z}_{is}\right\}u_{it}\right] = \sigma_{uu}\mathrm{Var}\left(E(\mathbf{z}_{it}) - T^{-1}\sum_{s=1}^{T}\mathbf{z}_{is}\right) \\
&= \sigma_{uu}\mathrm{Var}\left(T^{-1}\sum_{s=1}^{T}\mathbf{z}_{is}\right) = T^{-1}\sigma_{uu}\mathbf{\Sigma}_{zz}.
\end{aligned}$$

Clearly, $\text{Var}(\mathbf{F}_3) \to 0$ as $T \to \infty$ so that $\mathbf{F}_3 = o_p(1)$, and

$$N^{-1/2}T^{-1/2}\widetilde{\mathbf{Z}}^\top\widetilde{\mathbf{U}} = N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}[\mathbf{z}_{it} - E(\mathbf{z}_{it})]u_{it} + o_p(1).$$

Similarly,

$$N^{-1/2}T^{-1/2}\widetilde{\mathbf{W}}^\top\widetilde{\mathbf{U}} = N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}[\mathbf{w}_{it} - E(\mathbf{w}_{it})]u_{it} + o_p(1),$$

$$N^{-1/2}T^{-1/2}\widetilde{\mathbf{W}}^\top\widetilde{\mathbf{V}} = N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}[\mathbf{w}_{it} - E(\mathbf{w}_{it})]\mathbf{v}_{it}^\top + o_p(1),$$

and

$$N^{-1/2}T^{-1/2}\widetilde{\mathbf{Z}}^\top\widetilde{\mathbf{V}} = N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}[\mathbf{z}_{it} - E(\mathbf{z}_{it})]\mathbf{v}_{it}^\top + o_p(1).$$

Define,

$$\boldsymbol{\eta}_{i,t} = \begin{pmatrix} (\mathbf{w}_{it} - E(\mathbf{w}_{it}))u_{it} \\ (\mathbf{z}_{it} - E(\mathbf{z}_{it}))u_{it} \\ \text{Vec}\begin{pmatrix} (\mathbf{w}_{it} - E(\mathbf{w}_{it}))\mathbf{v}_{it}^\top \\ (\mathbf{z}_{it} - E(\mathbf{z}_{it}))\mathbf{v}_{it}^\top \end{pmatrix} \end{pmatrix}.$$

Similar to the proof of Lemma 2, it is easy to show that $E(\boldsymbol{\eta}_{i,t}) = 0$, and the covariance matrix $\text{Cov}(\boldsymbol{\eta}_{i,t}) = \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}_0$. Since $\{(\mathbf{w}_{it}^\top, \mathbf{z}_{it}^\top)^\top\}$ and $\{(u_{it}, \mathbf{v}_{it})\}$ are iid across both individuals and time, then $\{\boldsymbol{\eta}_{i,t}\}$ are iid across both individuals and time. It follows from the central limit theorem that

$$N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\eta}_{i,t} \Rightarrow \begin{pmatrix} \mathbf{Z}_{wu} \\ \mathbf{Z}_{zu} \\ \text{Vec}\begin{pmatrix} \mathbf{Z}_{wv} \\ \mathbf{Z}_{zv} \end{pmatrix} \end{pmatrix} \sim N(\mathbf{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}_0).$$

By applying Slutsky theorem, we have

$$T^{-1/2}N^{-1/2}(\widetilde{\mathbf{W}}^\top\widetilde{\mathbf{U}}, \widetilde{\mathbf{Z}}^\top\widetilde{\mathbf{U}}, \widetilde{\mathbf{W}}^\top\widetilde{\mathbf{V}}, \widetilde{\mathbf{Z}}^\top\widetilde{\mathbf{V}}) \Rightarrow (\mathbf{Z}_{wu}, \mathbf{Z}_{zu}, \mathbf{Z}_{wv}, \mathbf{Z}_{zv}).$$

Therefore, we conclude the proof of Lemma 7. Q.E.D

Proof of Theorem 3: Recall (4),

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \left[\widetilde{\mathbf{X}}^{*\top}\widetilde{\mathbf{Z}}^*\left(\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{Z}}^*\right)^{-1}\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^*\right]^{-1}\widetilde{\mathbf{X}}^{*\top}\widetilde{\mathbf{Z}}^*\left(\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{Z}}^*\right)^{-1}\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{U}}.$$

By Lemma 6,

$$N^{-1}T^{-1}\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{Z}}^* = N^{-1}T^{-1}\begin{pmatrix} \widetilde{\mathbf{W}}^\top\widetilde{\mathbf{W}} & \widetilde{\mathbf{W}}^\top\widetilde{\mathbf{Z}} \\ \widetilde{\mathbf{Z}}^\top\widetilde{\mathbf{W}} & \widetilde{\mathbf{Z}}^\top\widetilde{\mathbf{Z}} \end{pmatrix} \to^p \boldsymbol{\Sigma}_0.$$

20

By Assumption 1, one has

$$\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_1^{-1} = \begin{pmatrix} K_N^{-1}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{Z}}\mathbf{C} + K_N^{-1}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{W}}\mathbf{C}_1 + \widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{V}} & N^{-1/2}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{W}} \\ K_N^{-1}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{Z}}\mathbf{C} + K_N^{-1}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{W}}\mathbf{C}_1 + \widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{V}} & N^{-1/2}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{W}} \end{pmatrix},$$

and

$$\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_2^{-1} = N^{-1/2}\begin{pmatrix} \widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{Z}}\mathbf{C} + \widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{W}}\mathbf{C}_1 + K_N\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{V}} & \widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{W}} \\ \widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{Z}}\mathbf{C} + \widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{W}}\mathbf{C}_1 + K_N\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{V}} & \widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{W}} \end{pmatrix}$$

First, we consider the weak case. Since $K_N^{-1}N^{1/2} \to 1$, as $N \to \infty$, it follows from Lemmas 6 and 7 that

$$\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_1^{-1}N^{-1/2}T^{-1} \Rightarrow \boldsymbol{\Delta},$$

and

$$
\begin{aligned}
&T^{1/2}\mathbf{H}_1(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\
&= \left[\left(\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_1^{-1}N^{-1/2}T^{-1}\right)^{\top}\left(N^{-1}T^{-1}\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{Z}}^{*}\right)^{-1}\left(\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_1^{-1}N^{-1/2}T^{-1}\right)\right]^{-1} \\
&\quad \times \left(\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_1^{-1}N^{-1/2}T^{-1}\right)^{\top}\left(N^{-1}T^{-1}\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{Z}}^{*}\right)^{-1}\left(N^{-1/2}T^{-1}\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{U}}\right) \\
&\Rightarrow (\boldsymbol{\Delta}^{\top}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}^{\top}\boldsymbol{\Sigma}_0^{-1}\mathbf{Z}_u.
\end{aligned}
$$

Second, we focus on the near weak case. Since $K_N^{-1}N^{1/2} \to \infty$, as $N \to \infty$, by Lemmas 6 and 7, we have

$$\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_2^{-1}N^{-1/2}T^{-1} \Rightarrow \boldsymbol{\Delta},$$

and

$$
\begin{aligned}
&T^{1/2}\mathbf{H}_2(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\
&= \left[\left(\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_2^{-1}N^{-1/2}T^{-1}\right)^{\top}\left(N^{-1}T^{-1}\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{Z}}^{*}\right)^{-1}\left(\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_2^{-1}N^{-1/2}T^{-1}\right)\right]^{-1} \\
&\quad \times \left(\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{X}}^{*}\mathbf{H}_1^{-1}N^{-1/2}T^{-1}\right)^{\top}\left(N^{-1}T^{-1}\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{Z}}^{*}\right)^{-1}\left(N^{-1/2}T^{-1}\widetilde{\mathbf{Z}}^{*\top}\widetilde{\mathbf{U}}\right) \\
&\Rightarrow (\boldsymbol{\Delta}^{\top}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}^{\top}\boldsymbol{\Sigma}_0^{-1}\mathbf{Z}_u.
\end{aligned}
$$

The final consideration goes to the nearly non-identified case. Since $K_N/\sqrt{NT} \to 0$ as N and $T \to \infty$, an application of Lemma 7 gives

$$K_N N^{-1}T^{-1}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{V}} = K_N N^{-1}\left(N^{-1/2}T^{-1/2}\widetilde{\mathbf{W}}^{\top}\widetilde{\mathbf{V}}\right)N^{1/2}T^{-1/2} \to^{p} \mathbf{0},$$

and

$$K_N N^{-1}T^{-1}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{V}} = K_N N^{-1}\left(N^{-1/2}T^{-1/2}\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{V}}\right)N^{1/2}T^{-1/2} \to^{p} \mathbf{0}.$$

21

Thus, it follows from Lemmas 6 and that

$$\widetilde{\mathbf{Z}}^{*\top} \widetilde{\mathbf{X}}^{*} \mathbf{H}_2^{-1} N^{-1/2} T^{-1} \;\Rightarrow\; \boldsymbol{\Delta},$$

and

$$T^{1/2} \mathbf{H}_2 (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \;\Rightarrow\; (\boldsymbol{\Delta}^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Delta})^{-1} \boldsymbol{\Delta}^\top \boldsymbol{\Sigma}_0^{-1} \mathbf{Z}_u.$$

Hence, the proof of Theorem 3 is complete. Q.E.D

Proof of Corollary 4: First, we calculate $(\boldsymbol{\Delta}^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Delta})^{-1}$. Since

$$\boldsymbol{\Sigma}_0^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{ww}^{-1} + \boldsymbol{\Sigma}_{ww}^{-1} \boldsymbol{\Sigma}_{wz} \boldsymbol{\Omega}^{-1} \boldsymbol{\Sigma}_{zw} \boldsymbol{\Sigma}_{ww}^{-1} & -\boldsymbol{\Sigma}_{ww}^{-1} \boldsymbol{\Sigma}_{wz} \boldsymbol{\Omega}^{-1} \\ -\boldsymbol{\Omega}^{-1} \boldsymbol{\Sigma}_{zw} \boldsymbol{\Sigma}_{ww}^{-1} & \boldsymbol{\Omega}^{-1} \end{pmatrix},$$

then,

$$\boldsymbol{\Delta}^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Delta} = \begin{pmatrix} \mathbf{C}_1^\top \boldsymbol{\Sigma}_{wz} \mathbf{C} + \mathbf{C}^\top \boldsymbol{\Sigma}_{zw} \mathbf{C}_1 + \mathbf{C}_1^\top \boldsymbol{\Sigma}_{ww} \mathbf{C}_1 + \mathbf{C}^\top \boldsymbol{\Sigma}_{zz} \mathbf{C} & \mathbf{C}_1^\top \boldsymbol{\Sigma}_{ww} + \mathbf{C}^\top \boldsymbol{\Sigma}_{zw} \\ \boldsymbol{\Sigma}_{ww} \mathbf{C}_1 + \boldsymbol{\Sigma}_{wz} \mathbf{C} & \boldsymbol{\Sigma}_{ww} \end{pmatrix}.$$

Therefore, by the inverse of a partitioned matrix,

$$\sigma_{uu} (\boldsymbol{\Delta}^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Delta})^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{\beta,2} & -\boldsymbol{\Sigma}_{\beta,2} \mathbf{C}_2^\top \\ -\mathbf{C}_2 \boldsymbol{\Sigma}_{\beta,2} & \boldsymbol{\Sigma}_{\gamma,2} \end{pmatrix}.$$

Then, Corollary 4 holds from Theorem 3. Q.E.D

References

Angrist, J.D. and A.B. Krueger (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics*, **106**, 979-1014.

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

Baltagi, H. and Q. Li (2002). On instrumental Variable Estimation of Semiparametric Dynamic Panel Data Models. *Economics Letters*, **76**, 1-9.

Baltagi, B. (2005). *Econometric Analysis of Panel Data*, 2nd. New York: John Wiley and Sons.

Bound, J., D.A. Jaeger and R.M. Baker (1995). Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak. *Journal of the American Statistical Association*, **90**, 443-450.

Cai, Z. and H. Li (2008). Convergency and Divergency of Functional Coefficient Weak Instrumental Variables Models. *Statistics and Its Interface*, **2**,

Cai, Z. and Q. Li (2008). Nonparametric Estimation of Varying Coefficient Dynamic Panel Models. *Econometric Theory*, **24**, 1321-1342.

Campbell, J.Y. (2003). Consumption-Based Asset Pricing. *Handbook of the Economics of Finance*, Amstedan: Elsevier, 801-855.

Chao, J. and N. Swanson (2007). Alternative Approximations of the Bias and MSE of the IV Estimator Under Weak Identification with Application to Bias Correction. *Journal of Econometrics*, **137**, 515-555.

Hahn, J., J. Hausman and G. Kuersteiner (2004). Estimation with Weak Instruments: Accuracy of Higher Order Bias and MSE Approximations. *Econometrics Journal*, **7**, 272-306.

Hahn, J. and G. Kuersteiner (2002). Discontinuities of Weak Instrument Limiting Distributions. *Economics Letters*, **75**, 325-331.

Hall, R.E. (1978). Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence. *Journal of Political Economy*, **86**, 971-987.

Hausman, J., J. Stock and M. Yogo (2005). Asymptotic Properties of the Hahn-Hausman Test for Weak Instruments. *Economics Letters*, **89**, 333-342.

Horowitz, J.L. and M. Markatou (1996). Semiparametric Estimation of Regression Models for Panel data. *The Review of Economic Studies*, **63**, 145-168.

Hsiao, C. (2003). *Analysis of panel data*, 2nd. Cambridge: Cambridge University Press.

Judge, G.G. and M.E. Bock (1978). *The statistical Implications of Pre-test and Stein-Rule Estimators in Econometrics*. Amsterdam: North-Holland Publishing Company.

Kao, C. (1999). Spurious Regression and Residual-Based Tests for Cointegration in Panel Data. *Journal of Econometrics*, **90**, 1-44.

Lebedev, N.N. (1972). *Special Functions and Their Applications*. New York: Dover Publications.

Li, H. (2006). Semiparametric Weak Instrumental Variables Models for Panel and Cross-Sectional Data. *Ph.D. Dissertation*, Department of Mathematics and Statistics, University of North Carolina at Charlotte.

Li, Q. and C. Hsiao (1998). Testing Serial Correlation in Semiparametric Panel Data models. *Journal of Econometrics*, **87**, 207-237.

Li, Q. and T. Stengos, T. (1996). Semiparametric Estimation of Partially Linear Panel Data Models. *Journal of Econometrics*, **71**, 389-397.

Li, Q. and A. Ullah, A.(1998). Estimating Partially Linear Models with One-way Error Components. *Econometric Reviews*, **17**, 145-166.

Maddala, G.S. and J. Jeong (1992). On the Exact Small Sample Distribution of Instrumental Variable Estimator. *Econometrica*, **60**, 181-183.

Mavroeidis, S. (2004). Weak Identification of Forward-Looking Models in Monetary Models. *Oxford Bulletin of Economics and Statistics*, **66**, Supplement, 609-635.

Nason, J.M. and G.W. Smith (2005). Identifying the New Keynesian Phillips Curve. *Working Paper 2005-1*, Federal Reserve Bank of Atlanta.

Neeley, C.J., A. Roy and C.H. Whiteman (2001). Risk Aversion Versus Inter-temporal Substitution: A Case Study of Identification Failure in the Intertemporal Consumption Capital Asset Pricing Model. *Journal of Business and Economic Statistics*, **19**, 395-403.

Nelson, C. and R. Startz (1990). Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator. *Econometrica*, **58**, 967-976.

Phillips, P.C.B. and H. Moon (1999). Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica*, **50**, 1057-1112.

Staiger, D. and J. Stock (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, **65**, 556-586.

Stock, J. (2002). Instrumental Variables in Economics and Statistics. *International Encyclopedia of the Social Sciences*, Amstedan: Elsevier, 7577-7582.

Stock, J.H. and J.H. Wright (2000). GMM with Weak Instruments. *Econometrica*, **68**, 1055-1096.

Stock, J.H., J.H. Wright and M. Yogo (2002). A survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business and Economic Statistics*, **4**, 518-529.

Woodford, M.D. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton: Princeton University Press.

Yogo, M. (2004). Estimating the Elasticity of Intertemporal Substitution When Instruments are Weak. *Review of Economic Studies*, **86**, 797-810.

Ullah, A. (1974). On the Sampling Distribution of Improved Estimators for Coefficients in Linear Regression. *Journal of Econometrics*, **2**, 143-150.