CROSS-SECTION REGRESSION WITH COMMON SHOCKS

By

Donald W.K. Andrews

June 2003

COWLES FOUNDATION DISCUSSION PAPER NO. 1428

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Cross-section Regression with Common Shocks

Donald W. K. Andrews[1]
Cowles Foundation for Research in Economics
Yale University

April 2003

Abstract

This paper considers regression models for cross-section data that exhibit cross-section dependence due to common shocks, such as macroeconomic shocks. The paper analyzes the properties of least squares (LS) and instrumental variables (IV) estimators in this context. The results of the paper allow for any form of cross-section dependence and heterogeneity across population units. The probability limits of the LS and IV estimators are determined and necessary and sufficient conditions are given for consistency. The asymptotic distributions of the estimators are found to be mixed normal after re-centering and scaling. t, Wald, and F statistics are found to have asymptotic standard normal, χ^2, and scaled χ^2 distributions, respectively, under the null hypothesis when the conditions required for consistency of the parameter under test hold. But, the absolute values of t statistics and Wald and F statistics are found to diverge to infinity under the null hypothesis when these conditions fail. Confidence intervals exhibit similarly dichotomous behavior. Hence, common shocks are found to be innocuous in some circumstances, but quite problematic in others.

Models with factor structures for errors, regressors, and IV's are considered. Using the general results, conditions are determined under which consistency of the LS and IV estimators holds and fails in models with factor structures. The results are extended to cover heterogeneous and functional factor structures in which common factors have different impacts on different population units.

Extensions to generalized method of moments estimators are discussed.

Keywords: Asymptotics, common shocks, dependence, exchangeability, factor model, inconsistency, regression.

JEL Classification Numbers: C10, C12, C13.

1. Introduction

The regression model estimated by least squares (LS) or instrumental variables (IV) is the work-horse of econometrics. The properties of LS and IV estimators and related testing methods have been studied extensively. In particular, there has been extensive research on the effects on these estimators of key features of economic data such as simultaneity, measurement errors, left-out variables, heteroskedasticity, and autocorrelation.

Surprisingly, however, there has been little research on the effects of common shocks, such as macroeconomic, political, and environmental shocks, on the properties of LS and IV estimators in cross-section regressions. There has been some research on models with group effects, e.g., Moulton (1990) and other references listed below, and on models with spatial autocorrelation, e.g., see Case (1991) and Conley (1999) and other references listed below. But, this research focuses on shocks that are predominantly local in nature. Common shocks need not be of this form.

Below we argue that common shocks are, in fact, a likely feature of cross-section economic data. Then, in the body of the paper, we analyze the effects of common shocks on the properties of LS and IV estimators and related test statistics.

The population units in cross-section regression are often individuals, households, or firms, though sometimes they are industries, plants, cities, states, countries, or products. Common shocks influence each of these different types of population units. We now discuss a variety of types of common shocks and how they may affect population units. Our emphasis is on individuals, households, and firms.

Macroeconomic shocks, both financial and real, impact micro-level population units. For example, inflation affects nominal wages of individuals; interest rates affect individual consumption and firm investment; oil price shocks affect firm factor costs and production; stock market shocks affect individual wealth and firm assets; and financial crises (including stock market, banking, and foreign exchange crises) affect individual and firm consumption, investment, and production. Real shocks, such as aggregate demand and employment shocks, similarly affect the behavior of individuals and firms.

Federal and state governmental shocks in fiscal and monetary policy cause financial and real shocks that affect population units. For example, tax changes affect individual consumption; central bank shocks affect interest rates and inflation and, hence, individual consumption and nominal wages and firm investment.

Other macroeconomic shocks include changes in international economic agreements (e.g., GATT, WTO, free trade areas, customs unions) that affect individual consumption and firm level production via their competitive effects. In addition, large mergers, bankruptcies, and antitrust decisions can cause shocks that are common across firms and employees at the industry level.

Some results from the estimation of panel models with factor structures provide direct evidence that macroeconomic shocks affect individual population units. For example, stock market returns for different companies are found to have a strong common component.

Note that the *impact* of common shocks, such as macroeconomic shocks, typically is not the same across different population units. For example, a stock market shock affects wealthy individuals much more than poor individuals. An oil price shock affects airlines and auto companies much more than computer companies. In the extreme, some common shocks may have no affect on some population units. In the framework considered below, common shocks are allowed to have different impacts on different population units depending on the characteristics (possibly observed, possibly unobserved) of the population units.

In addition to macroeconomic common shocks, there are a myriad of other types of common shocks that affect micro-level population units. Technology shocks, such as advances in computer hardware and software, the internet, and other information technology, affect individual consumer demand and firm productivity.

Legal/institutional shocks include regulatory changes (as have occurred with airlines, railroads, trucks, electricity, communications, and banking in the U.S.); fundamental statutory changes (such as the Civil Rights Act, Tax Reform Act, and Americans with Disabilities Act (ADA)); and supreme court decisions. Each of these shocks affects micro-level units. For example, regulatory changes affect the factor prices that firms face, and the ADA affects opportunities of individuals with disabilities as well as costs that firms face.

Political shocks include changes in political regimes (such as the fall of communism in Eastern Europe, the rise of the European Union, changes in government in democratic societies—exemplified by the election of George W. Bush over Al Gore, and the rise and fall of dictators—exemplified by Saddam Hussein); changes in policy of existing regimes; war (including world wars, regional wars, local wars, and civil wars); terrorism—exemplified by the destruction of the World Trade Center; and changes in international relations. Again, it is easy to see that these shocks affect micro-level population units both in the countries in which the political shocks occur and in other countries.

Environmental shocks include meteorological, i.e., weather, shocks (including el nino and la nina effects); natural disasters (including floods, earthquakes, droughts, and forest fires); and man-made disasters (including oil spills, air pollution, overfishing, and forest fires). For example, droughts in predominantly agricultural countries affect food prices, individual consumption, and individual farm production. Some other environmental shocks, such as global warming and deforestation, may be sufficiently incremental that they do not have a discernible affect on a cross-section regression at a given point in time.

Health shocks include diseases (such as AIDS, SARS, and other epidemics) and medical advances (including vaccines, drugs, and other treatments). Especially in less developed countries, common health shocks can have pervasive effects on individuals.

Sociological shocks include styles and trends in consumer goods (e.g., mini-vans, SUVs, and bell bottoms) as well as changes in opportunities for women and minorities. As with some environmental shocks, some sociological shocks may be too incremental to have discernible effects on a cross-section regression at a given point in time.

In sum, it seems clear that numerous common shocks occur and affect individual population units in a given cross-section, whether the units are individuals, households, firms, or some other units. Furthermore, given the globalization of the economy, one would expect that the impact of common shocks on micro-level units to become more pervasive as time passes. The above discussion also indicates that one expects the effects of common shocks to vary considerably depending on the characteristics of the population unit. Hence, one needs to allow for a heterogeneous impact of common shocks across population units.

In this paper, we analyze the effects of common shocks on LS and IV estimators and related tests. We utilize a different probabilistic framework than is usual in statistics and econometrics. We start by defining random vectors for all units in the population, not just the observed units, on a given probability space. Then, we consider iid sampling from the population with the randomness in the sampling defined on the same probability space.[2] This framework allows for general patterns of cross-section dependence and heterogeneity, while at the same time yielding asymptotic results that are remarkably simple. The framework is similar to that used by Conley (1999) but does not impose a strong mixing assumption.

Using this framework, we address the question of when do common shocks cause problems for standard methods and when do they not. First, we determine the probability limit of LS and IV estimators in the general setting. We obtain necessary and sufficient conditions for consistency of the estimators. Next, we specify standard factor structures for the errors, regressors, and IV's. We show that consistency holds or fails to hold depending upon the properties of the common factors and the idiosyncratic components in the models. We extend these results to what we call heterogeneous factor structures and functional factor structures. In these factor structures, common shocks are infinite dimensional and the impact of a common shock on a population unit depends on the characteristics of that unit. Special cases of the factor structures considered include models with variance components and models with group structures. But, the factor models covered by the results are much more general than these models.

Returning to the general setting, we establish that the estimators (suitably normalized) have mixed normal asymptotic distributions. The asymptotic properties of t, Wald, and F statistics are determined. They are found to have asymptotic standard normal, χ^2, and scaled χ^2 distributions, respectively, under the null hypothesis when the necessary conditions for consistency hold. Similarly, the usual confidence intervals for regression parameters are shown to have asymptotically correct coverage probabilities when the necessary conditions for consistency hold.

On the other hand, when the conditions for consistency of the parameter under test do not hold, absolute values of t statistics and Wald and F statistics diverge to infinity in probability under the null. Correspondingly, the usual confidence intervals have coverage probabilities that converge to zero as the sample size goes to infinity. Such behavior, obviously, is problematic. We conclude that there is a sharp dichotomy in the behavior of test statistics when common shocks are present depending upon the assumptions. These results are applied easily to the models discussed above with standard, heterogeneous, and functional factor structures.

The asymptotic results are obtained by exploiting the exchangeability of the observations, which results from iid sampling from the population. A law of large numbers (LLN) for exchangeable random variables leads to the probability limit results for the estimators. A martingale difference sequence (MDS) central limit theorem (CLT) provides the mixed normal asymptotic distributional results. The necessary and sufficient condition for consistency of the LS slope coefficient estimator is that the errors are conditionally uncorrelated with the regressors given the σ-field $\mathcal{C}$ that is generated by common shocks. The form of $\mathcal{C}$ is simple in the case of models with standard, heterogeneous, or functional factor structures. As noted above, the necessary and sufficient condition holds or fails in the factor structure models depending on the properties of the factors and idiosyncratic components in the models.

The paper discusses extensions of the results to panel regression models with a fixed number of time periods T, clustered sampling, and generalized method of moments (GMM) estimators of moment condition models.

The existing literature on cross-section dependence in cross-section regression models includes a number of papers on models with group effects (and the closely related models with variance components and clustered sampling), see Kloek (1981), Scott and Holt (1982), Greenwald (1983), Pfeffermann and Smith (1985), Moulton (1986, 1987, 1990), Moulton and Randolf (1989), and Pepper (2002). Donald and Lang (2001) consider panel regression models with group effects. In these models, the errors for observations within any given group are correlated (typically equicorrelated), but the errors (and observations) across different groups are independent. Thus, these models allow for simple forms of common shocks, but not common shocks that affect all units in the population, such as many macroeconomic and political shocks among others.

Conley (1999) considers GMM estimation for cross-section observations that are assumed to form a stationary strong mixing random field. Conley's approach is a more sophisticated and flexible way of handling cross-section dependence than via models with group effects. The basic idea, however, is similar in that common shocks are presumed to have predominantly local effects (due to the strong mixing assumption). Numerous other papers in the spatial econometrics literature consider parametric models for cross-section dependence that is predominantly local in nature, e.g., see Anselin (1988), Case (1991), Kelejian and Prucha (1999), Chen and Conley (2001), and references cited therein. This literature is complementary to the present paper, which focuses on common shocks that may or may not be local in nature.

There is a growing literature on factor models for panel data in which the number of time series observations is large and the number of cross-section units may or may not be large, e.g., see Geweke (1977), Sargent and Sims (1977), Chamberlain and Rothschild (1983), Forni, Hallin, Lippi, and Reichlin (2000), Forni and Lippi (2001), Bai and Ng (2001, 2002), Moon and Perron (2002), Pesaran (2002), Phillips and Sul (2002), Bai (2003), and Stock and Watson (2003). These papers allow for common shocks in the errors (though not necessarily in the regressors). These papers differ from the present paper in that we consider common shocks in cross-section models, rather than in panel models with large T, and we allow for more general forms of common shocks. In future work, we plan to use the probabilistic framework adopted here to explore the properties of estimators and tests in panel data models with large T and large n.

The remainder of this paper is organized as follows. For simplicity and clarity, we consider results for the LS estimator first. Later we show how these results can be extended straightforwardly to regression models estimated by IV's. Section 2 specifies the regression model and the probabilistic framework employed in the paper. Section 3 establishes the probability limit of the LS estimator and provides conditions under which the LS estimator is consistent and inconsistent in standard, heterogeneous, and functional factor structure models. Section 4 establishes the asymptotic mixed normality of the LS estimator. Section 5 introduces covariance matrix estimators and determines their probability limits. Section 6 analyzes the asymptotic properties of t, Wald, and F tests under the null hypothesis. Section 7 extends the results for LS estimators to IV estimators. Section 8 discusses extensions to panel models with a fixed time dimension T, clustered sampling, and GMM estimators. Section 9 provides a brief conclusion. An Appendix provides proofs of results stated in the paper.

All limits are taken as $n \to \infty$, where n is the sample size.

2. Regression Model

The probabilistic framework that we adopt is somewhat unconventional because we want to be explicit about the cross-section dependence that may exist between all units in the population. We start by defining, for each cross-sectional unit in the population, the dependent and independent regression variables, as well as other characteristics of the unit that may or may not be observed. Then, we specify the sampling scheme used to draw observations from the population.

Let γ denote some unit in the population. Let Γ denote the set of all units in the population, where Γ is an arbitrary topological space. For population unit $\gamma \in \Gamma$, $Y(\gamma) \in R$ denotes the regression dependent variable, $X(\gamma) \in R^k$ denotes the regression independent variable vector, and $S(\gamma) \in \mathcal{S}$ denotes some supplementary variables that include other characteristics of population unit γ and/or some stochastic terms that are common to some or all of the units in the population, where $\mathcal{S}$ is an arbitrary topological space. Let

$$W(\gamma) = (Y(\gamma), X(\gamma), S(\gamma)). \tag{2.1}$$

For each $\gamma \in \Gamma$, $W(\gamma)$ is a random element defined on a (common) probability space $(\Omega, \mathcal{B}, P)$ (using the product Borel σ-field on $(R, R^k, \mathcal{S})$).

For each $\gamma \in \Gamma$, the vector $(Y(\gamma), X(\gamma))$ satisfies the regression model

$$Y(\gamma) = \alpha_0 + X(\gamma)'\beta_0 + U(\gamma), \tag{2.2}$$

where $U(\gamma)$ is a scalar error, β_0 is an unknown k-vector parameter, and α_0 is an unknown scalar parameter. Our interest centers on the properties of the least squares estimators of β_0 and α_0.

A standard assumption for a linear regression model to be well defined is for the error to have mean zero and to be uncorrelated with the regressors. For cross-section applications, another standard assumption is that the random elements $W(\gamma)$ are independent across different units γ. Thus, the following assumptions are standard (STD) for cross-section applications:

Assumption STD1. $E(1, X(\gamma)')U(\gamma) = 0$ for all $\gamma \in \Gamma$.

Assumption STD2. $\{W(\gamma) : \gamma \in \Gamma\}$ are independent across $\gamma \in \Gamma$.

We do not impose Assumptions STD1 and STD2. We state these assumptions for reference only.

Our results allow for arbitrary dependence between $W(\gamma_1)$ and $W(\gamma_2)$ for all $\gamma_1, \gamma_2 \in \Gamma$. In particular, $(W(\gamma_1), W(\gamma_2))$ may be subject to common shocks and, hence, be dependent. In addition, the effect of a common shock on the distribution of $(Y(\gamma_1), X(\gamma_1), U(\gamma_1))$ may depend on $S(\gamma_1)$ and, hence, may be different from its effect on $(Y(\gamma_2), X(\gamma_2), U(\gamma_2))$ when $S(\gamma_1) \neq S(\gamma_2)$. Arbitrary forms of heterogeneity (i.e., non-identical distributions) of $W(\gamma)$ across $\gamma \in \Gamma$ also are allowed.

Samples of size n for $n \geq 1$ are obtained by drawing indices $\{\gamma_i : i \geq 1\}$ randomly from Γ according to a probability distribution G on Γ (coupled with its Borel σ-field). (The random indices $\{\gamma_i : i \geq 1\}$ are defined on the same probability space $(\Omega, \mathcal{B}, P)$ as $\{W(\gamma) : \gamma \in \Gamma\}$.) That is, we assume:

Assumption 1. $\{\gamma_i : i \geq 1\}$ are iid indices, independent of $\{W(\gamma) : \gamma \in \Gamma\}$, each with some distribution G.

Assumption 1 allows for probabilistic over-sampling of some units or proportional sampling depending on the specification of the distribution G. Proportional sampling is obtained when G is a uniform distribution on Γ. For example, if Γ is a bounded subset of Euclidean space, then proportional sampling is obtained by taking G to have a density proportional to Lebesgue measure. Over-sampling of some units is obtained by taking G to be some non-uniform distribution on Γ. A special case of this is *multinomial* sampling, e.g., see Imbens and Lancaster (1996), which is a type of stratified sampling.

We denote

$$W_i = W(\gamma_i),\ Y_i = Y(\gamma_i),\ X_i = X(\gamma_i),\ S_i = S(\gamma_i), \text{ and } U_i = U(\gamma_i) \tag{2.3}$$

for $i = 1, 2, \ldots$. ($W(\gamma_i)$ is assumed to be a measurable function on $(\Omega, \mathcal{B}, P)$ with respect to the product Borel σ-field on $(R \times R^k \times \mathcal{S})$.) In the probability literature, $\{W_i : i \geq 1\}$ is called a *subordinated* stochastic process, subordinated to the process $\{W(\gamma) : \gamma \in \Gamma\}$ via the *directing* process $\{\gamma_i : i \geq 1\}$, see Feller (1966, Ch. X.7, p. 345). Subordinated processes have been used in economics by Mandelbrot and Taylor (1967) and Clark (1973), among others, for quite different purposes than those considered here and in econometrics by Conley (1999) for a similar purpose to that considered here.

The observations for sample size n are $\{(Y_i, X_i) : i = 1, ..., n\}$. In addition, depending upon the context, S_i or some component of S_i may be observed for $i = 1, ..., n$.

In terms of the sample of the first n observations, the model is

$$Y_i = \alpha_0 + X_i'\beta_0 + U_i \text{ for } i = 1, ..., n, \tag{2.4}$$

where Y_i is an observed scalar dependent variable, X_i is an observed regressor k-vector, and U_i is an unobserved scalar error.

By iterated expectations, the definition that $W_i = W(\gamma_i)$, and the independence of $\{W(\gamma) : \gamma \in \Gamma\}$ and $\{\gamma_i : i \geq 1\}$, we have: for any vector-valued function $h(\cdot)$ with $E||h(W_i)|| < \infty$,

$$Eh(W_i) = E_{\gamma_i} E(h(W(\gamma_i))|\gamma_i) = \int Eh(W(\gamma))dG(\gamma), \tag{2.5}$$

where E_{γ_i} denotes expectation with respect to the randomness in γ_i.

To determine the large sample properties of LS estimators of β_0 and α_0, we make use of the fact that the random elements $\{W_i : i = 1, 2, ...\}$ are *exchangeable* given Assumption 1. (That is, $(W_{\pi(1)}, ..., W_{\pi(n)})$ has the same distribution as $(W_1, ..., W_n)$ for every permutation π of $(1, ..., n)$ for all $n \geq 2$.) In consequence, de Finetti's Theorem (e.g., see Hall and Heyde (1980, Thm. 7.1, p. 203)) applies, and we have the following result.

Lemma 1. *Suppose Assumption* 1 *holds. Then,* $\{W_i : i = 1, 2, ...\}$ *are exchangeable random elements and there exists a* σ*-field* $\mathcal{C} \subset \mathcal{B}$ *such that, conditional on* $\mathcal{C}$, $\{W_i : i = 1, 2, ...\}$ *are iid.*

Comments. 1. The σ-field $\mathcal{C}$ equals $\bigcap_{n=1}^{\infty} \mathcal{C}_n$, where $\mathcal{C}_n$ is the σ-field of n-symmetric random variables (that is, the σ-field generated by random variables that depend on $\{W_i : i = 1, 2, ...\}$ and are invariant to permutations of the first n random variables $\{W_i : i = 1, 2, ..., n\}$), see Hall and Heyde (1980, p. 202).

2. The σ-field $\mathcal{C}$ consists of the *common* shocks to the random elements $\{W_i : i = 1, 2, ...\}$. The effect of a common shock could be the same for all population units or it could depend on the characteristics of a given unit through the supplementary variable S_i. For example, a common shock could affect observations that are in a certain group or region, but not other observations. Suppose $S_{g,i}$ is a dummy variable that equals one if the i-th observation is in group g and zero otherwise for $g = 1, ..., g_{\max}$. Let $C_1, ..., C_{g_{\max}}$ denote common shocks, i.e., random variables that are $\mathcal{C}$-measurable. Then, the regression dependent and independent variables (Y_i, X_i) could depend on the common shocks $(C_1, ..., C_{g_{\max}})$ through the vector $S_i = (C_1 S_{1,i}, C_2 S_{2,i}, ..., C_{g_{\max}} S_{g,i})'$. Thus, only observations in group g are affected by the g-th common shock. In this case, the model is an example of a model with group effects, see the Introduction for references.

In the group effect literature, the shocks $(C_1, ..., C_{g_{\max}})$ are assumed to be independent. But, in the present paper, there is no need to make this assumption. In fact, $(C_1, ..., C_{g_{\max}})$ could just denote the differential impacts of a single common shock on g different groups and, in this case, correlation between the elements of $(C_1, ..., C_{g_{\max}})$ would be expected.

Furthermore, the effect of common shocks may differ across observations in a continuous manner. For example, suppose the effect of some macroeconomic shock, such as an interest rate change, depends on the characteristics of the population unit, such as its wealth holdings, as measured by some absolutely continuous component,

$S_{1,i}$, of S_i. The macro shock could take the form of a random function $C(\cdot)$ that is $\mathcal{C}$-measurable with the effect of the macro shock on the i-th observation being through $C(S_{1,i})$. Thus, the impact of the common shock varies continuously across i depending on the value of $S_{1,i}$.

In this case, the model could be akin to models in the spatial econometrics literature in which shocks are predominantly local in nature, e.g., due to the spatial autoregressive assumption in Case (1991) and the strong mixing assumption in Conley (1999). On the other hand, the model could be one in which some common shocks affect a sufficient number of population units that the effect is not local in nature. For example, the model could be such that all population units are effected in a manner that varies continuously, but the effect for all units is significant.

3. Probability Limit of the LS Estimator

3.1. Main Results

The LS estimator, $\widehat{\beta}_n$, of β_0 can be written as

$$\widehat{\beta}_n = \beta_0 + \left(n^{-1}\sum_{i=1}^n X_i X_i' - \overline{X}_n \overline{X}_n'\right)^{-1}\left(n^{-1}\sum_{i=1}^n X_i U_i - \overline{X}_n \overline{U}_n\right), \text{ where}$$
$$\overline{X}_n = n^{-1}\sum_{i=1}^n X_i \text{ and } \overline{U}_n = n^{-1}\sum_{i=1}^n U_i. \tag{3.1}$$

The LS estimator, $\widehat{\alpha}_n$, of α_0 can be written as

$$\widehat{\alpha}_n = \overline{Y}_n - \overline{X}_n'\widehat{\beta}_n = \alpha_0 + \overline{U}_n - \overline{X}_n'(\widehat{\beta}_n - \beta_0), \text{ where } \overline{Y}_n = n^{-1}\sum_{i=1}^n Y_i. \tag{3.2}$$

The probability limits of the terms in the expressions for $\widehat{\beta}_n$ and $\widehat{\alpha}_n$ are determined using the following LLN for exchangeable random variables, e.g., see Hall and Heyde (1980, (7.1), p. 202):

Lemma 2. *Suppose Assumption* 1 *holds. Let $h(\cdot)$ be a vector-valued function that satisfies $E||h(W_i)|| < \infty$. Then,*

$$n^{-1}\sum_{i=1}^n h(W_i) \to_p E(h(W_i)|\mathcal{C}) \text{ as } n \to \infty,$$

where $\mathcal{C}$ is the σ-field given in Lemma 1*.*

Comments. 1. The convergence in the lemma also holds almost surely (a.s.).
2. The random variable W_i that appears in the limit is $W(\gamma_i)$, which is a draw from the population $\{W(\gamma) : \gamma \in \Gamma\}$ according to the distribution G. In consequence, by

iterated expectations and the independence of $\{W(\gamma) : \gamma \in \Gamma\}$ and $\{\gamma_i : i \geq 1\}$ conditional on $\mathcal{C}$, the limit random variable in the lemma can be written as

$$E(h(W_i)|\mathcal{C}) = E_{\gamma_i} E(h(W(\gamma_i))|\mathcal{C}, \gamma_i) = \int Eh(W(\gamma)|\mathcal{C}) dG(\gamma), \tag{3.3}$$

where E_{γ_i} denotes expectation with respect to the randomness in γ_i.

To establish the probability limits of $\widehat{\beta}_n$ and $\widehat{\alpha}_n$, we require (i) some moment conditions and (ii) that the regressor variables contain sufficient idiosyncratic variability that their conditional covariance matrix given the common shocks $\mathcal{C}$ is nonsingular:

Assumption 2. (a) $E||X_i||^2 = \int E||X(\gamma)||^2 dG(\gamma) < \infty$.
(b) $E|U_i| = \int E|U(\gamma)| dG(\gamma) < \infty$.
(c) $E||X_i U_i|| = \int E||X(\gamma)U(\gamma)|| dG(\gamma) < \infty$.
(d) $E(X_i X_i'|\mathcal{C}) - E(X_i|\mathcal{C}) E(X_i|\mathcal{C})' > 0$ a.s.

In terms of the population random elements $\{W(\gamma) : \gamma \in \Gamma\}$, Assumption 2(d) is

$$\int E\left(X(\gamma)X(\gamma)'|\mathcal{C}\right) dG(\gamma) - \int E(X(\gamma)|\mathcal{C}) dG(\gamma) \int E(X(\gamma)'|\mathcal{C}) dG(\gamma) > 0 \text{ a.s.} \tag{3.4}$$

The deviation of the probability limit of $\widehat{\beta}_n$ from β_0 is given by

$$r(\mathcal{C}) = \left(E(X_i X_i'|\mathcal{C}) - E(X_i|\mathcal{C}) E(X_i|\mathcal{C})'\right)^{-1} \left(E(X_i U_i|\mathcal{C}) - E(X_i|\mathcal{C}) E(U_i|\mathcal{C})\right). \tag{3.5}$$

Note that the term $E(X_i U_i|\mathcal{C}) - E(X_i|\mathcal{C}) E(U_i|\mathcal{C})$ in (3.5) is the conditional covariance given $\mathcal{C}$ between X_i and U_i. Also note that $r(\mathcal{C})$ is the solution to the conditional population least squares minimization problem

$$\min_{\beta \in R^k} E(U_i - X_i'\beta|\mathcal{C})' E(U_i - X_i'\beta|\mathcal{C}). \tag{3.6}$$

In terms of the population random elements $\{W(\gamma) : \gamma \in \Gamma\}$, $r(\mathcal{C})$ is

$$\begin{aligned} r(\mathcal{C}) = & \left(\int E(X(\gamma)X(\gamma)'|\mathcal{C}) dG(\gamma) - \int E(X(\gamma)|\mathcal{C}) dG(\gamma) \int E(X(\gamma)'|\mathcal{C}) dG(\gamma)\right)^{-1} \\ & \times \left(\int E(X(\gamma)U(\gamma)|\mathcal{C}) dG(\gamma) - \int E(X(\gamma)|\mathcal{C}) dG(\gamma) \int E(U(\gamma)|\mathcal{C}) dG(\gamma)\right). \end{aligned} \tag{3.7}$$

The deviation of the probability limit of $\widehat{\alpha}_n$ from α_0 is given by

$$s(\mathcal{C}) = E(U_i|\mathcal{C}) - E(X_i|\mathcal{C})' r(\mathcal{C}). \tag{3.8}$$

Using (3.1), (3.2), and Lemma 2, the probability limits of $\widehat{\beta}_n$ and $\widehat{\alpha}_n$ are easily obtained:

Theorem 3. *Suppose Assumptions* 1 *and* 2 *hold. Then,*

$$\widehat{\beta}_n \to_p \beta_0 + r(\mathcal{C}) \text{ and}$$
$$\widehat{\alpha}_n \to_p \alpha_0 + s(\mathcal{C}).$$

Comments. 1. The convergence in the Theorem holds jointly and almost surely.
2. Theorem 3 states that the probability limit of $\widehat{\beta}_n$ is β_0 plus a term, $r(\mathcal{C})$, that may be zero, random, or in some cases a non-zero constant. Similarly, $\widehat{\alpha}_n$ equals α_0 plus a term, $s(\mathcal{C})$, that may be zero, random, or a non-zero constant.
3. The term $r(\mathcal{C})$ is zero if and only if the conditional correlation given $\mathcal{C}$ between X_i and U_i is zero. Note that the standard assumption employed in the literature, Assumption STD1, coupled with Assumption 1, implies that the unconditional correlation between X_i and U_i is zero. This does *not,* however, imply that their conditional correlation given $\mathcal{C}$ is zero. Hence, under Assumption STD1, $r(\mathcal{C})$ is not necessarily zero.
4. Donald and Lang (2001) consider a regression model with a group structure and find inconsistency of the LS estimator due to correlations within each of a finite number of groups. This is an example of the result of Theorem 3.

For any random vectors A and B and any random vector or σ-field D, let $Cov(A, B|D)$ denote the conditional covariance between A and B given D, i.e., $E(AB'|D) - E(A|D)E(B|D)'$.

It is easy to see that a necessary and sufficient condition for $r(\mathcal{C}) = 0$ is the following:

Assumption CU. $Cov(X_i, U_i|\mathcal{C}) = 0$ a.s.

(CU abbreviates *conditionally uncorrelated.*)

Necessary and sufficient conditions for $r(\mathcal{C}) = 0$ and $s(\mathcal{C}) = 0$ are Assumption CU plus the following:

Assumption CMZ. $E(U_i|\mathcal{C}) = 0$ a.s.

(CMZ abbreviates *conditionally mean zero.*)

Given Theorem 3, we have the following necessary and sufficient condition for consistency of $\widehat{\beta}_n$ and $\widehat{\alpha}_n$.

Corollary 4. *Suppose Assumptions* 1 *and* 2 *hold. Then,* $\widehat{\beta}_n \to_p \beta_0$ *if and only if Assumption* CU *holds; and* $(\widehat{\beta}_n, \widehat{\alpha}_n) \to_p (\beta_0, \alpha_0)$ *if and only if Assumptions* CU *and* CMZ *hold.*

Comments. 1. Assumptions CU and CMZ are necessary for consistency of the LS estimators, but they are not necessary for *unbiasedness.* Unbiasedness holds (by trivial calculations) under the following standard condition:

Assumption STD3. (a) $E(U_i|X_i) = 0$ a.s.
(b) $E||\widehat{\beta}_n|| < \infty$ and $E|\widehat{\alpha}_n| < \infty$.

In consequence, if Assumption STD3 holds and $\{\widehat{\beta}_n : n \geq 1\}$ is uniformly integrable, then $r(\mathcal{C})$ has mean zero.[3] Hence, $r(\mathcal{C})$ is either zero or random. It cannot be a non-zero constant. In this case, inconsistency of $\widehat{\beta}_n$ is due to randomness that does not die out as $n \to \infty$. Inconsistency is *not* due to improper centering of $\widehat{\beta}_n$ that persists as $n \to \infty$. Analogous comments apply to $\widehat{\alpha}_n$.

2. If Assumption CU fails, it is still possible to construct a consistent estimator of β_0 if IV's are available that are uncorrelated with U_i conditional on $\mathcal{C}$, see Section 7 below.

Sufficient conditions for Assumptions CU and CMZ in terms of the population quantities, $(X(\gamma), U(\gamma))$, rather than the observed quantities, (X_i, U_i), are:

Assumption CUγ. (a) $Cov(X(\gamma), U(\gamma)|\mathcal{C}) = 0$ a.s. for all $\gamma \in \Gamma$.
(b) Either $E(U(\gamma)|\mathcal{C})$ or $E(X(\gamma)|\mathcal{C})$ does not depend on γ a.s. for all $\gamma \in \Gamma$.

Assumption CMZγ. $E(U(\gamma)|\mathcal{C}) = 0$ a.s. for all $\gamma \in \Gamma$.

Lemma 5 (a) *Assumptions* 1 *and CUγ imply Assumption CU.*
(b) *Assumptions* 1 *and CMZγ imply Assumption CMZ.*

Comment. It is interesting to note that zero conditional covariance given $\mathcal{C}$ between the population quantities $X(\gamma)$ and $U(\gamma)$ does not imply zero conditional covariance given $\mathcal{C}$ between the observed regressor X_i and the corresponding error U_i. The same is true in terms of unconditional covariances or correlations. Thus, zero covariance between $X(\gamma)$ and $U(\gamma)$ does not imply that X_i and U_i have zero covariance. The former plus the condition that either $EU(\gamma)$ or $EX(\gamma)$ does not depend on γ for all $\gamma \in \Gamma$ suffices for X_i and U_i have zero covariance. Of course, if $EU(\gamma) = 0$ for all $\gamma \in \Gamma$, then the additional condition holds. In the present context, this additional condition may seem innocuous, but in the factor structures considered below the additional condition is not necessarily innocuous.

3.2. Standard Factor Structure

Corollary 4 shows that a necessary and sufficient condition for consistency of the LS estimator of β_0 (or (β_0, α_0)) is Assumption CU (or Assumptions CU and CMZ). We now provide sufficient conditions for Assumption CU (or CU and CMZ) in terms of a *standard factor* structure for the regressors and errors.[4] (We use the term "standard" here because (i) the factor structure considered here is akin to factor structures considered in the literature and (ii) we want to differentiate the factor structure considered here from the *heterogeneous* and *functional* factor structures considered below.)

Assumption SF1. For all $\gamma \in \Gamma$,

$$\begin{aligned} U(\gamma) &= C_1' U^*(\gamma), \\ X(\gamma) &= C_2 X^*(\gamma), \end{aligned}$$

and $S(\gamma) = (C_1, C_2)$, where (a) C_1 and $U^*(\gamma)$ are random d_1 vectors; $X^*(\gamma)$ is a random d_2 vector; and C_2 is a random $k \times d_2$ matrix for $d_2 \geq k$; (b) $\{(U^*(\gamma), X^*(\gamma)) : \gamma \in \Gamma\}$, (C_1, C_2), and $\{\gamma_i : i \geq 1\}$ are mutually independent; and (c) $(U^*(\gamma), X^*(\gamma))$ are independent across $\gamma \in \Gamma$.[5]

Assumption SF1 defines a factor structure with random common factor vectors and matrices (C_1, C_2) and random factor loadings $(U^*(\gamma), X^*(\gamma))$. It is easy to see that

(C_1, C_2) is measurable with respect to $\mathcal{C}$. In fact, $\mathcal{C} = \sigma(C_1, C_2)$ using Assumptions SF1(b) and (c) and the definition of $S(\gamma)$.

Given Assumption SF1, Assumption 2(d) holds provided

$$EX_i^* X_i^{*\prime} - EX_i^* EX_i^{*\prime} > 0 \tag{3.9}$$

and C_2 has full row rank d_2 a.s.

In keeping with the notation used above, we let $U_i^* = U^*(\gamma_i)$ and $X_i^* = X^*(\gamma_i)$. For any random vectors A and B, let $Cov(A, B)$ denote the covariance between A and B.

To obtain consistency of the LS slope coefficient estimator $\widehat{\beta}_n$ we require:

Assumption SF2. $Cov(X_i^*, U_i^*) = 0$.

Note that Assumption SF2 does not require that the error factor loading vector, U_i^*, has mean zero. This allows one element of both U_i^* and X_i^* to equal one, which means that the errors and regressors may contain a purely common component.

However, to obtain consistency of the LS intercept estimator, $\widehat{\alpha}_n$, U_i^* must have mean zero:

Assumption SF3. $EU_i^* = 0$.

Assumption SF3 rules out a purely common component in U_i.

We now show that Assumptions 1, SF1, and SF2 imply Assumption CU. Using Assumptions 1 and SF1, we have

$$\begin{aligned}
&E(U_i|\mathcal{C}) = E(C_1' U_i^*|\mathcal{C}) = C_1' E(U_i^*|\mathcal{C}) = C_1' EU_i^*,\\
&E(X_i|\mathcal{C}) = E(C_2 X_i^*|\mathcal{C}) = C_2 E(X_i^*|\mathcal{C}) = C_2 EX_i^*, \text{ and}\\
&E(X_i U_i|\mathcal{C}) = E(C_2 X_i^* U_i^{*\prime} C_1|\mathcal{C}) = C_2 E(X_i^* U_i^{*\prime}|\mathcal{C}) C_1 = C_2 E(X_i^* U_i^{*\prime}) C_1,
\end{aligned} \tag{3.10}$$

where the second equality in each line holds because $\mathcal{C} = \sigma(C_1, C_2)$ and the third equality in each line holds because (U_i^*, X_i^*) is a function of $\{(U^*(\gamma), X^*(\gamma)) : \gamma \in \Gamma\}$ and γ_i and the latter are independent of $\mathcal{C} = \sigma(C_1, C_2)$.

Combining the results in (3.10) gives

$$\begin{aligned}
E(X_i U_i|\mathcal{C}) - E(X_i|\mathcal{C}) E(U_i|\mathcal{C}) &= C_2 E(X_i^* U_i^{*\prime}) C_1 - C_2 EX_i^* EU_i^{*\prime} C_1\\
&= C_2 (EX_i^* U_i^{*\prime} - EX_i^* EU_i^{*\prime}) C_1\\
&= 0,
\end{aligned} \tag{3.11}$$

where the last equality holds by Assumption SF2.

Assumptions 1, SF1, and SF3 imply Assumption CMZ by the first line of (3.10).

Sufficient conditions for Assumptions SF2 and SF3 in terms of population quantities are:

Assumption SF2γ. (a) $Cov(X^*(\gamma), U^*(\gamma)) = 0$ for all $\gamma \in \Gamma$.
(b) Either $EU^*(\gamma)$ or $EX^*(\gamma)$ does not depend on γ for all $\gamma \in \Gamma$.

Assumption SF3γ. $EU^*(\gamma) = 0$ for all $\gamma \in \Gamma$.

Assumption SF2γ(b) requires a certain degree of homogeneity across population units. See the Comment following Lemma 5.

The following Corollary is a special case of a more general result (viz., Theorem 9) given below. The first two parts of the Corollary are the results proved above in (3.10) and (3.11). (The proof is given above because it is instructive.)

Corollary 6. (a) *Suppose Assumptions* 1, SF1, *and* SF2 *hold. Then, Assumption* CU *holds and* $r(\mathcal{C}) = 0$.
(b) *Suppose Assumptions* 1 *and* SF1-SF3 *hold. Then, Assumptions* CU *and* CMZ *hold,* $r(\mathcal{C}) = 0$, *and* $s(\mathcal{C}) = 0$.
(c) *Assumptions* 1 *and* SF2γ *imply Assumption* SF2.
(d) *Assumptions* 1 *and* SF3γ *imply Assumption* SF3.

Comments. 1. Theorem 3 and Corollary 6 combine to show that $\widehat{\beta}_n$ is consistent under Assumptions 1, 2, SF1, and SF2 and $(\widehat{\beta}_n, \widehat{\alpha}_n)$ is consistent under Assumptions 1, 2, and SF1-SF3.
2. If Assumptions SF2 and SF3 are strengthened to $E(U_i^*|X_i^*) = 0$ a.s., then Assumption STD3(a) holds. In this case, $\widehat{\beta}_n$ and $\widehat{\alpha}_n$ are unbiased (provided their expectation exists). This holds because

$$\begin{aligned} &E(U_i|X_i) = E_{X_i^*,\mathcal{C}}E(U_i|X_i, X_i^*, \mathcal{C}) = E_{X_i^*,\mathcal{C}}C_1'E(U_i^*|X_i, X_i^*, \mathcal{C}) \\ &= E_{X_i^*,\mathcal{C}}C_1'E(U_i^*|X_i^*) = 0 \text{ a.s.,} \end{aligned} \tag{3.12}$$

where $E_{X_i^*,\mathcal{C}}$ denotes expectation with respect to $(X_i^*, \mathcal{C})$.

We now show that the regressors and errors may satisfy the standard factor structure of Assumption SF1 and the standard assumptions of mean zero errors and lack of covariance between the errors and regressors, viz., Assumption STD1, yet fail Assumption CU. In this case, consistency of the LS estimator of β_0 does not hold due to the effect of common shocks.

Instead of Assumptions SF2γ and SF3γ, consider the following assumption:

Assumption SF4. (a) $Cov(X^*(\gamma), U_1^*(\gamma)) = 0$ and $EU_1^*(\gamma) = 0$ for all $\gamma \in \Gamma$, where $U^*(\gamma) = (U_1^*(\gamma), U_2^*(\gamma))' \in R^2$.
(b) $C_1 = (1, C_{11})' \in R^2$.
(c) $EC_{11} = 0$ and $EC_2C_{11} = 0$.
(d) $\int EX^*(\gamma)U_2^*(\gamma)dG(\gamma) \neq 0$; $EU_2^*(\gamma) = 0$ for all $\gamma \in \Gamma$; $C_{11} \neq 0$ with probability one; and C_2 is full row rank with probability one.

Under Assumption SF4,

$$U(\gamma) = U_1^*(\gamma) + C_{11}U_2^*(\gamma), \tag{3.13}$$

where $U_1^*(\gamma)$ has mean zero and is uncorrelated with the idiosyncratic component of the regressor $X_2^*(\gamma)$; the error factor C_{11} has mean zero and is uncorrelated with the regressor factors C_2; and $U_2^*(\gamma)$ has mean zero but is correlated with the idiosyncratic component of the regressor $X_2^*(\gamma)$.

Theorem 7. *Suppose Assumptions* 1, SF1, *and* SF4 *hold. Then, Assumption* STD1 *holds, but Assumption* CU *does not hold.*

Comments. 1. Under Assumptions 1, 2, SF1, and SF4, we have

$$\begin{aligned} r(\mathcal{C}) &= \left(C_2 E^*[X_i^* - E^* X_i^*][X_i^* - E^* X_i^*]' C_2'\right)^{-1} \left(C_2 E^*[X_i^* - E^* X_i^*] U_{2,i}^* C_{11}\right) \text{ and} \\ s(\mathcal{C}) &= E^* U_{2,i}^* C_{11} + (E^* X_i^*)' C_2' r(\mathcal{C}), \text{ where} \\ U_{2,i}^* &= U_2^*(\gamma_i) \end{aligned} \tag{3.14}$$

and E^* denotes expectation with respect to (X_i^*, U_i^*) alone.
2. Theorems 3 and 7 combine to show that $\widehat{\beta}_n$ is *not* consistent for β_0 under Assumptions 1, 2, SF1, and SF4.
3. In the proof of the Theorem, Assumption SF4(c) is used only to show that Assumption STD1 holds and Assumption SF4(d) is used only to show that Assumption CU does not hold.

3.3. Heterogeneous Factor Structure

In this subsection, we generalize the standard factor structure to a *heterogeneous factor structure*. The heterogeneous factor structure allows the effects of the common shocks to differ across population units depending on the characteristics of the unit. In particular, the common shocks for the γ-th unit are of the form $(C_1(S_0(\gamma)), C_2(S_0(\gamma)))$, where $S_0(\gamma)$ is a vector of characteristics of the γ-th unit. Hence, the common shocks take the form of stochastic functions $(C_1(\cdot), C_2(\cdot))$. The random element $S_0(\gamma)$ may or may not be observed.

For a random element ξ, let $supp(\xi)$ denote the support of ξ. Let $supp(S_0) = \cup_{\gamma \in \Gamma} supp(S_0(\gamma))$.

The heterogeneous factor structure is specified in the following assumptions:

Assumption HF1. For all $\gamma \in \Gamma$,

$$\begin{aligned} U(\gamma) &= C_1(S_0(\gamma))' U^*(\gamma), \\ X(\gamma) &= C_2(S_0(\gamma)) X^*(\gamma), \text{ and} \\ S(\gamma) &= (S_0(\gamma), C_1(\cdot), C_2(\cdot)), \end{aligned}$$

where (a) $U^*(\gamma)$ is a random d_1 vector; $C_1(\cdot)$ is a random d_1 vector-valued function with domain $supp(S_0)$; $X^*(\gamma)$ is a random d_2 vector; and $C_2(\cdot)$ is a random $k \times d_2$ matrix-valued function with domain $supp(S_0)$ for $d_2 \geq k$; (b) $\{(U^*(\gamma), X^*(\gamma), S_0(\gamma)) : \gamma \in \Gamma\}$, $(C_1(\cdot), C_2(\cdot))$, and $\{\gamma_i : i \geq 1\}$ are mutually independent; and (c) $(U^*(\gamma), X^*(\gamma), S_0(\gamma))$ are independent across $\gamma \in \Gamma$.[6]

For notational convenience, let $X_i^* = X^*(\gamma_i)$, $U_i^* = U^*(\gamma_i)$, and $S_{0,i} = S_0(\gamma_i)$.

With the heterogeneous factor structure, to obtain $r(\mathcal{C}) = 0$ and consistency of $\widehat{\beta}_n$, we need a strengthened version of Assumption SF2 to hold.

Assumption HF2. (a) $Cov(X_i^*, U_i^* | S_{0,i}) = 0$ a.s.
(b) Either $E(U_i^* | S_{0,i})$ or $E(X_i^* | S_{0,i})$ does not depend on $S_{0,i}$ a.s.

Similarly, for $s(\mathcal{C}) = 0$ and consistency of $\widehat{\alpha}_n$, we need a strengthened version of Assumption SF3 to hold:

Assumption HF3. $E(U_i^*|S_{0,i}) = 0$ a.s.

A sufficient condition for Assumptions HF2 and HF3 is $E(U_i^*|X_i^*, S_{0,i}) = 0$ a.s.

Sufficient conditions for Assumptions HF2 and HF3 in terms of population quantities are:

Assumption HF2γ. (a) $Cov(X^*(\gamma), U^*(\gamma)|S_0(\gamma)) = 0$ a.s. for all $\gamma \in \Gamma$.
(b) Either $E(U^*(\gamma)|S_0(\gamma))$ or $E(X^*(\gamma)|S_0(\gamma))$ does not depend on $S_0(\gamma)$ a.s. or on γ for all $\gamma \in \Gamma$.

Assumption HF3γ. $E(U^*(\gamma)|S_0(\gamma)) = 0$ a.s. for all $\gamma \in \Gamma$.

As in the previous subsection, the common σ-field $\mathcal{C}$ is the σ-field generated by the common shocks:

$$\mathcal{C} = \sigma(C_1(\cdot), C_2(\cdot)). \tag{3.15}$$

The following result, like Corollary 6, is a special case of Theorem 9 given below.

Corollary 8. (a) *Suppose Assumptions* 1, HF1, *and* HF2 *hold. Then, Assumption* CU *holds and* $r(\mathcal{C}) = 0$.
(b) *Suppose Assumptions* 1 *and* HF1-HF3 *hold. Then, Assumptions* CU *and* CMZ *hold,* $r(\mathcal{C}) = 0$, *and* $s(\mathcal{C}) = 0$.
(c) *Assumptions* 1 *and* HF2γ *imply Assumption* HF2.
(d) *Assumptions* 1 *and* HF3γ *imply Assumption* HF3.

Comment. Corollary 8 and Theorem 3 show that Assumptions 1, HF1 and HF2 are sufficient for consistency of $\widehat{\beta}_n$ and, with the addition of Assumption HF3, for $\widehat{\alpha}_n$.

3.4. Functional Factor Structure

We now provide sets of sufficient conditions for Assumption CU and Assumptions CU and CMZ that are as general as we can find. We call the structures considered *functional factor structures.* These structures are sufficiently general that they contain both standard and heterogeneous factor structures. The conditions allow the effect of common shocks on a population unit to depend on the characteristics of the population unit via a component $S_0(\gamma)$ of $S(\gamma)$. The common shocks are characterized by a function $C(\cdot)$. In particular, the effects of the common shocks on unit γ is through $C(S_0(\gamma))$. The errors and regressors are determined by stochastic processes $U(\cdot, \gamma)$ and $X(\cdot, \gamma)$ that are uncorrelated conditional on $S_0(\gamma)$ for each $\gamma \in \Gamma$. Specifically, we have:

Assumption FF1. (a) $S_0(\gamma)$ is a component of $S(\gamma)$.
(b) $C(\cdot)$ is a random function that does not depend on γ, has domain *supp*(S_0), and is a component of $S(\gamma)$ for all $\gamma \in \Gamma$.
(c) For each $\gamma \in \Gamma$, $U(\cdot, \gamma)$ and $X(\cdot, \gamma)$ are random functions with ranges R and R^k, respectively, and domain $\cup_{\gamma \in \Gamma} \text{supp}(C(S_0(\gamma)))$.

(d) For each $\gamma \in \Gamma$, $U(\gamma) = U(C(S_0(\gamma)), \gamma)$ and $X(\gamma) = X(C(S_0(\gamma)), \gamma)$.
(e) $\{(U(\cdot,\gamma), X(\cdot,\gamma), S_0(\gamma)) : \gamma \in \Gamma\}$, $C(\cdot)$, and $\{\gamma_i : i \geq 1\}$ are mutually independent.
(f) $(U(\cdot,\gamma), X(\cdot,\gamma), S_0(\gamma))$ are independent across $\gamma \in \Gamma$.

Assumption FF1 allows the whole distributions of $U(\cdot,\gamma)$ and $X(\cdot,\gamma)$ to vary with $C(S_0(\gamma))$. In contrast, with standard or heterogeneous factor structures, $C(S_0(\gamma))$ only affects the multivariate location and scale of the regressors and errors.

Let $X_i(c) = X(c,\gamma_i)$, $U_i(c) = U(c,\gamma_i)$, and $S_{0,i} = S_0(\gamma_i)$.

Let $supp(C) = \cup_{\gamma\in\Gamma} supp(C(S_0(\gamma)))$.

Assumption FF2. (a) $Cov(X_i(c), U_i(c)|S_{0,i}) = 0$ a.s. for all $c \in supp(C)$.
(b) Either $E(U_i(c)|S_{0,i})$ or $E(X_i(c)|S_{0,i})$ does not depend on $S_{0,i}$ for all $c \in supp(C)$ a.s.

Assumptions FF1 and FF2 are sufficient for consistency of $\widehat{\beta}_n$. To obtain consistency of $\widehat{\alpha}_n$, we also need:

Assumption FF3. $E(U_i(c)|S_{0,i}) = 0$ a.s. for all $c \in supp(C)$.

Sufficient conditions for Assumptions FF2 and FF3 in terms of the population random quantities are:

Assumption FF2γ. (a) $Cov(X(c,\gamma), U(c,\gamma)|S_0(\gamma)) = 0$ a.s. for all $c \in supp(C)$ and all $\gamma \in \Gamma$.
(b) Either $E(U(c,\gamma)|S_0(\gamma))$ or $E(X(c,\gamma)|S_0(\gamma))$ does not depend on $S_0(\gamma)$ a.s. or on γ for all $c \in supp(C)$ and all $\gamma \in \Gamma$.

Assumption FF3γ. $E(U(c,\gamma)|S_0(\gamma)) = 0$ a.s. for all $c \in supp(C)$ and all $\gamma \in \Gamma$.

Sufficiency of Assumptions FF1 and FF2 for Assumption CU etc. are established in the following theorem:

Theorem 9. (a) *Suppose Assumptions* 1, FF1, *and* FF2 *hold. Then, Assumption* CU *holds and* $r(\mathcal{C}) = 0$.
(b) *Suppose Assumptions* 1, FF1, FF2, *and* FF3 *hold. Then, Assumptions* CU *and* CMZ *hold,* $r(\mathcal{C}) = 0$*, and* $s(\mathcal{C}) = 0$.
(c) *Assumptions* 1 *and* FF2γ *imply Assumption* FF2.
(d) *Assumptions* 1 *and* FF3γ *imply Assumption* FF3.

Comments. 1. Assumptions SF1 and SF2 imply Assumptions FF1 and FF2 with $S_0(\gamma) = 0$, $C(\cdot) = (C_1, C_2)$, $U(c,\gamma) = c_1' U^*(\gamma)$, and $X(c,\gamma) = c_2 X^*(\gamma)$, where $c = (c_1, c_2)$. Analogously, Assumptions SF1-SF3 imply Assumptions FF1-FF3.
2. Assumptions HF1 and HF2 imply Assumptions FF1 and FF2 with $C(\cdot) = (C_1(\cdot), C_2(\cdot))$, $U(c,\gamma) = c_1' U^*(\gamma)$, and $X(c,\gamma) = c_2 X^*(\gamma)$, where $c = (c_1, c_2)$. Analogously, Assumptions HF1-HF3 imply Assumptions FF1-FF3.

4. Asymptotic Mixed Normality of the LS Estimator

In this section, we establish the asymptotic distribution of $\widehat{\beta}_n$ suitably centered and scaled. These results allow one to determine the effect of cross-section dependence on the null rejection rates of hypothesis tests and on the coverage probabilities of confidence intervals constructed using LS estimators.

To establish asymptotic normality of the estimator, we use the following additional moment conditions:

Assumption 3. (a) $EU_i^2 = \int EU^2(\gamma)dG(\gamma) < \infty$.
(b) $E||X_i U_i||^2 = \int E||X(\gamma)U(\gamma)||^2 dG(\gamma) < \infty$.

The following quantity is used to center the LS estimator in order to establish its asymptotic distribution:

$$r_n(\mathcal{C}) = \left(n^{-1} \sum_{i=1}^{n} X_i X_i' - \overline{X}_n \overline{X}_n' \right)^{-1} \left(E(X_i U_i|\mathcal{C}) - E(X_i|\mathcal{C})E(U_i|\mathcal{C}) \right). \qquad (4.1)$$

Note that $r_n(\mathcal{C})$ converges in probability to $r(\mathcal{C})$ as $n \to \infty$ under Assumptions 1 and 2 by Lemma 2. Also note that $r_n(\mathcal{C}) = 0$ if and only if Assumption CU holds.

The conditional asymptotic variance, $V_{\mathcal{C}}$, of the normalized LS estimator of β_0 given $\mathcal{C}$ is defined as follows:

$$\begin{aligned} V_{\mathcal{C}} &= B_{\mathcal{C}}^{-1} \Omega_{\mathcal{C}} B_{\mathcal{C}}^{-1}, \text{ where} \\ B_{\mathcal{C}} &= E([X_i - E(X_i|\mathcal{C})][X_i - E(X_i|\mathcal{C})]'|\mathcal{C}), \qquad (4.2) \\ \Omega_{\mathcal{C}} &= E(\xi_i \xi_i'|\mathcal{C}), \text{ and} \\ \xi_i &= [X_i - E(X_i|\mathcal{C})]U_i - E([X_i - E(X_i|\mathcal{C})]U_i|\mathcal{C}) - [X_i - E(X_i|\mathcal{C})]E(U_i|\mathcal{C}). \end{aligned}$$

Note that $B_{\mathcal{C}}$ is positive definite a.s. by Assumption 2(d).

In contrast, under standard assumptions for cross-section data, viz., Assumptions STD1, STD2, and 1-3, the asymptotic variance of the normalized LS estimator of β_0 is given by

$$\begin{aligned} V &= B^{-1} \Omega B^{-1}, \text{ where} \\ B &= E[X_i - EX_i][X_i - EX_i]', \\ \Omega &= E\xi_i^S \xi_i^{S\prime}, \text{ and} \\ \xi_i^S &= [X_i - EX_i]U_i. \qquad (4.3) \end{aligned}$$

Note that the last two of the three terms in the definition of ξ_i in (4.2) do not appear in the definition of ξ_i^S in (4.3). The second term of ξ_i does not appear in ξ_i^S because it is the mean of the first term of ξ_i conditional on $\mathcal{C}$ and the mean of ξ_i^S is zero. Also, the third term of ξ_i does not appear in ξ_i^S because the third term of ξ_i arises due to the lack of asymptotic equivalence between $n^{1/2}$ times the $\overline{X}_n \overline{U}_n$ term in the definition of $\widehat{\beta}_n$, see (3.1), and $n^{1/2}(plim_{n\to\infty} \overline{X}_n)\overline{U}_n$, which occurs because $E\overline{U}_n$ is not necessarily zero in (4.2), whereas these quantities are asymptotically equivalent under Assumption STD1 because $E\overline{U}_n$ is zero.

If Assumption CU holds, then ξ_i and $\Omega_{\mathcal{C}}$ simplify because the second term in the definition of ξ_i in (4.2) is zero. If Assumption CMZ holds, the third term of ξ_i is zero.

If Assumption CU holds, we have

$$\begin{aligned}\xi_i &= [X_i - E(X_i|\mathcal{C})][U_i - E(U_i|\mathcal{C})] \text{ and} \\ \Omega_{\mathcal{C}} &= \Omega_{\mathcal{C}}^0, \text{ where} \\ \Omega_{\mathcal{C}}^0 &= E([U_i - E(U_i|\mathcal{C})]^2[X_i - E(X_i|\mathcal{C})][X_i - E(X_i|\mathcal{C})]'|\mathcal{C}). \end{aligned} \tag{4.4}$$

In particular, under Assumptions SF1 and SF2, we have

$$\begin{aligned}\Omega_{\mathcal{C}} &= C_2 E^*[C_1'(U_i^* - EU_i^*)]^2[X_i^* - E^*X_i^*][X_i^* - E^*X_i^*]'C_2' \text{ and} \\ B_{\mathcal{C}} &= C_2 E^*[X_i^* - E^*X_i^*][X_i^* - E^*X_i^*]'C_2', \end{aligned} \tag{4.5}$$

where E^* denotes expectation with respect to (U_i^*, X_i^*) alone.

Under Assumptions HF1 and HF2, we have

$$\begin{aligned}\Omega_{\mathcal{C}} &= E^*[C_1(S_{0,i})'(U_i^* - EU_i^*)]^2 C_2(S_{0,i})[X_i^* - E^*X_i^*][X_i^* - E^*X_i^*]'C_2(S_{0,i})' \text{ and} \\ B_{\mathcal{C}} &= E^*C_2(S_{0,i})[X_i^* - E^*X_i^*][X_i^* - E^*X_i^*]'C_2(S_{0,i})', \end{aligned} \tag{4.6}$$

where E^* denotes expectation with respect to $(U_i^*, X_i^*, S_{0,i})$ alone.

Next, define

$$\sigma_{\mathcal{C}}^2 = Var(U_i|\mathcal{C}) = E([U_i - E(U_i|\mathcal{C})]^2|\mathcal{C}). \tag{4.7}$$

Suppose the errors are homoskedastic conditional on $\mathcal{C}$, i.e.,

$$E([U_i - E(U_i|\mathcal{C})]^2|\mathcal{C}, X_i) = \sigma_{\mathcal{C}}^2 \text{ a.s.} \tag{4.8}$$

Then, if Assumption CU holds, $\Omega_{\mathcal{C}}$ and $V_{\mathcal{C}}$ simplify to

$$\Omega_{\mathcal{C}} = \sigma_{\mathcal{C}}^2 B_{\mathcal{C}} \text{ and } V_{\mathcal{C}} = \sigma_{\mathcal{C}}^2 B_{\mathcal{C}}^{-1}, \tag{4.9}$$

respectively. Note that Assumption CMZ is not needed for these simplifications to hold.

The asymptotic distribution of $\widehat{\beta}_n$ after centering and scaling is given in the following theorem.

Theorem 10. *Suppose Assumptions* 1-3 *hold. Let* $Z \sim N(0, I_k)$ *be a standard normal k-vector that is independent of* $\mathcal{C}$. *Then,*
(a) $n^{1/2}(\widehat{\beta}_n - \beta_0 - r_n(\mathcal{C})) \to_d V_{\mathcal{C}}^{1/2} Z$,
(b) $V_{\mathcal{C}}^{-1/2} n^{1/2}(\widehat{\beta}_n - \beta_0 - r_n(\mathcal{C})) \to_d Z$ *provided* $V_{\mathcal{C}} > 0$ *a.s., and*
(c) $r_n(\mathcal{C}) \to_p r(\mathcal{C})$.

Comments. 1. Part (a) of the Theorem implies that $n^{1/2}(\widehat{\beta}_n - \beta_0 - r_n(\mathcal{C}))$ has a mixed normal asymptotic distribution.
2. Under Assumption CU, part (a) of the Theorem gives the asymptotic distribution of $n^{1/2}(\widehat{\beta}_n - \beta_0)$ because $r_n(\mathcal{C}) = 0$. Hence, if the errors and regressors have factor

structures that satisfy Assumptions SF1 and SF2, HF1 and HF2, or FF1 and FF2, then $n^{1/2}(\widehat{\beta}_n - \beta_0)$ has the asymptotic mixed normal distribution given by $V_{\mathcal{C}}^{1/2}Z$.
3. Parts (a) and (b) are established using an MDS CLT, e.g., see Hall and Heyde (1980, Thm. 3.2, p. 58). Part (c) is established using Lemma 2.
4. The asymptotic distribution of $\widehat{\alpha}_n$, after suitable centering and scaling, can be obtained by the same argument as for $\widehat{\beta}_n$. For brevity, we do not do so here.

5. Covariance Matrix Estimation

The usual heteroskedasticity-robust estimator of the asymptotic variance of $\widehat{\beta}_n$ is denoted $\widehat{V}_n$. It is defined by

$$\begin{aligned} \widehat{V}_n &= \widehat{B}_n^{-1}\widehat{\Omega}_n\widehat{B}_n^{-1}, \text{ where} \\ \widehat{B}_n &= n^{-1}\sum_{i=1}^{n}[X_i - \overline{X}_n][X_i - \overline{X}_n]', \\ \widehat{\Omega}_n &= n^{-1}\sum_{i=1}^{n}\widehat{U}_i^2[X_i - \overline{X}_n][X_i - \overline{X}_n]', \text{ and} \\ \widehat{U}_i &= Y_i - \widehat{\alpha}_n - X_i'\widehat{\beta}_n. \end{aligned} \tag{5.1}$$

The usual estimator of the asymptotic variance of $\widehat{\beta}_n$ that relies on homoskedasticity of the errors is

$$\widehat{V}_{\sigma,n} = \widehat{\sigma}_n^2\widehat{B}_n^{-1}, \text{ where } \widehat{\sigma}_n^2 = (n-k-1)^{-1}\sum_{i=1}^{n}\widehat{U}_i^2. \tag{5.2}$$

To obtain the probability limits of the covariance matrix estimators, we strengthen the moment conditions used:

Assumption 4. (a) $E||X_i||^4 = \int E||X(\gamma)||^4 dG(\gamma) < \infty$.
(b) $E||X_i||^3|U_i| = \int E||X(\gamma)||^3|U(\gamma)|dG(\gamma) < \infty$.

The probability limit of $\widehat{\Omega}_n$ depends on $\Omega_{\mathcal{C}}^0$, defined in (4.4), and the following random matrix:

$$\begin{aligned} \eta_{\mathcal{C}} &= E\big([r(\mathcal{C})'(X_i - E(X_i|\mathcal{C}))]^2[X_i - E(X_i|\mathcal{C})][X_i - E(X_i|\mathcal{C})]'|\mathcal{C}\big) \\ &\quad -2E\big([r(\mathcal{C})'(X_i - E(X_i|\mathcal{C}))][U_i - E(U_i|\mathcal{C})][X_i - E(X_i|\mathcal{C})][X_i - E(X_i|\mathcal{C})]'|\mathcal{C}\big). \end{aligned} \tag{5.3}$$

If Assumption CU holds, then $r(\mathcal{C}) = 0$ and $\eta_{\mathcal{C}} = 0$.

The probability limit of $\widehat{\sigma}_n^2$ depends on $\sigma_{\mathcal{C}}^2$ and the following random variable:

$$\begin{aligned} \tau_{\mathcal{C}} &= E\big([r(\mathcal{C})'(X_i - E(X_i|\mathcal{C}))]^2|\mathcal{C}\big) \\ &\quad -2E\big([r(\mathcal{C})'(X_i - E(X_i|\mathcal{C}))][U_i - E(U_i|\mathcal{C})]|\mathcal{C}\big). \end{aligned} \tag{5.4}$$

If Assumption CU holds, then $r(\mathcal{C}) = 0$ and $\tau_{\mathcal{C}} = 0$.

The asymptotic properties of the covariance matrix estimators $\widehat{V}_n$ and $\widehat{V}_{\sigma,n}$ are given in the following theorem:

Theorem 11. *Suppose Assumptions* 1-4 *hold. Then,*
(a) $\widehat{B}_n \to_p B_{\mathcal{C}}$,
(b) $\widehat{\Omega}_n \to_p \Omega^0_{\mathcal{C}} + \eta_{\mathcal{C}}$,
(c) $\widehat{V}_n \to_p B^{-1}_{\mathcal{C}}[\Omega^0_{\mathcal{C}} + \eta_{\mathcal{C}}]B^{-1}_{\mathcal{C}}$,
(d) $\widehat{\sigma}^2_n \to_p \sigma^2_{\mathcal{C}} + \tau_{\mathcal{C}}$, *and*
(e) $\widehat{V}_{\sigma,n} \to_p (\sigma^2_{\mathcal{C}} + \tau_{\mathcal{C}})B^{-1}_{\mathcal{C}}$.

Comments. 1. The quantities $\eta_{\mathcal{C}}$ and $\tau_{\mathcal{C}}$ arise in the Theorem because the residuals, $\{\widehat{U}_i : i = 1, ..., n\}$, are not consistent estimators of the errors, $\{U_i : i = 1, ..., n\}$, if Assumptions CU and CMZ do not hold. In fact, only Assumption CU is needed for $\eta_{\mathcal{C}} = 0$ and $\tau_{\mathcal{C}} = 0$. Hence, if Assumption CU holds but Assumption CMZ does not hold, then the residuals are not consistent estimators of the errors, but $\widehat{\Omega}_n$ and $\widehat{\sigma}^2_n$ are still consistent for $\Omega^0_{\mathcal{C}}$ and $\sigma^2_{\mathcal{C}}$, respectively. The reason is that $\widehat{U}_i$ is consistent for $U_i - E(U_i|\mathcal{C})$.
2. If Assumption CU holds (as well as Assumptions 1-4), then $\eta_{\mathcal{C}} = 0$, $\Omega_{\mathcal{C}} = \Omega^0_{\mathcal{C}}$, $\widehat{\Omega}_n \to_p \Omega^0_{\mathcal{C}}$, and $\widehat{V}_n \to_p B^{-1}_{\mathcal{C}}\Omega^0_{\mathcal{C}}B^{-1}_{\mathcal{C}} = V_{\mathcal{C}}$. If Assumption CU and (4.8) hold, then $\tau_{\mathcal{C}} = 0$, $\widehat{\sigma}^2_n \to_p \sigma^2_{\mathcal{C}}$, and $\widehat{V}_{\sigma,n} \to_p \sigma^2_{\mathcal{C}}B^{-1}_{\mathcal{C}} = V_{\mathcal{C}}$.
3. If Assumption CU does not hold, then $\Omega^0_{\mathcal{C}} + \eta_{\mathcal{C}}$ does not equal $\Omega_{\mathcal{C}}$ in general and $\widehat{\Omega}_n \to_p \Omega^0_{\mathcal{C}} + \eta_{\mathcal{C}} \neq \Omega_{\mathcal{C}}$. Hence, if Assumption CU does not hold, $\widehat{V}_n = \widehat{B}^{-1}_n\widehat{\Omega}_n\widehat{B}^{-1}_n$ is not a consistent estimator of $V_{\mathcal{C}} = B^{-1}_{\mathcal{C}}\Omega_{\mathcal{C}}B^{-1}_{\mathcal{C}}$ in general. Similarly, if (4.8) holds, but Assumption CU does not hold, then $(\sigma^2_{\mathcal{C}}+\tau_{\mathcal{C}})B_{\mathcal{C}}$ does not equal $\Omega_{\mathcal{C}}$ in general and $\widehat{\sigma}^2_n\widehat{B}_n \to_p (\sigma^2_{\mathcal{C}} + \tau_{\mathcal{C}})B_{\mathcal{C}} \neq \Omega_{\mathcal{C}}$. Hence, in this case, $\widehat{V}_{\sigma,n} = \widehat{\sigma}^2_n\widehat{B}^{-1}_n$ is not a consistent estimator of $V_{\mathcal{C}} = B^{-1}_{\mathcal{C}}\Omega_{\mathcal{C}}B^{-1}_{\mathcal{C}}$ in general.

The probability limits of $\widehat{V}_n$ and $\widehat{V}_{\sigma,n}$ are nonsingular a.s. under Assumption 2(d) and the following assumption:

Assumption 5. (a) $\Omega^0_{\mathcal{C}} + \eta_{\mathcal{C}} > 0$ a.s.
(b) $\sigma^2_{\mathcal{C}} + \tau_{\mathcal{C}} > 0$ a.s.

Using Assumption 5, Theorems 10 and 11 combine to give the following results for the LS estimator of β_0 normalized by an estimated covariance matrix:

Corollary 12. *Suppose Assumptions* 1-5 *hold. Let* $Z \sim N(0, I_k)$ *be a standard normal k-vector that is independent of* $\mathcal{C}$. *Then,*
(a) $(\widehat{V}_n)^{-1/2}n^{1/2}(\widehat{\beta}_n - \beta_0 - r_n(\mathcal{C})) \to_d (B^{-1}_{\mathcal{C}}[\Omega^0_{\mathcal{C}} + \eta_{\mathcal{C}}]B^{-1}_{\mathcal{C}})^{-1/2} \times V^{1/2}_{\mathcal{C}} \times Z$,
(b) $\widehat{V}^{-1/2}_n n^{1/2}(\widehat{\beta}_n - \beta_0) \to_d Z$ *provided Assumption* CU *also holds,*
(c) $\widehat{V}^{-1/2}_{\sigma,n} n^{1/2}(\widehat{\beta}_n - \beta_0 - r_n(\mathcal{C})) \to_d (\sigma^2_{\mathcal{C}} + \tau_{\mathcal{C}})^{-1/2}B^{1/2}_{\mathcal{C}} \times V^{1/2}_{\mathcal{C}} \times Z$, *and*
(d) $\widehat{V}^{-1/2}_{\sigma,n} n^{1/2}(\widehat{\beta}_n - \beta_0) \to_d Z$ *provided Assumption* CU *and* (4.8) *also hold.*

6. Test Statistics

Asymptotic results for t and Wald (or equivalently, F) tests can be obtained by using the results of Theorems 10 and 11. Consider the hypotheses $H_0 : \beta_j = \beta_{0,j}$ and

$H_1 : \beta_j \neq \beta_{0,j}$ for some $j \leq k$, where $\beta = (\beta_1, ..., \beta_k)'$ and $\beta_0 = (\beta_{0,1}, ..., \beta_{0,k})'$. The t statistic for testing H_0 against H_1 is

$$T_n = \frac{\sqrt{n}\left(\widehat{\beta}_{n,j} - \beta_{0,j}\right)}{\sqrt{[\widehat{V}_n]_{j,j}}}, \tag{6.1}$$

where $\widehat{\beta}_n = (\widehat{\beta}_{n,1}, ..., \widehat{\beta}_{n,k})'$ and $[D]_{j,j}$ denotes the j-th diagonal element of a square matrix D. The usual two-sided t test with nominal significance level α rejects the null hypothesis when $|T_n| > z_{1-\alpha/2}$, where z_α denotes the α quantile of the standard normal distribution. A one-sided t test with nominal significance level α rejects H_0 in favor of $H_1' : \beta_j > \beta_{0,j}$ when $T_n > z_{1-\alpha}$.

Next, consider the hypotheses $H_0 : R\beta_0 = a$ and $H_1 : R\beta_0 \neq a$, where R is a (non-stochastic) full row rank $q \times k$ matrix and a is a (non-stochastic) q vector. Define the Wald test statistic W_n as follows:

$$W_n = ||(R\widehat{V}_n R')^{-1/2} n^{1/2} (R\widehat{\beta}_n - a)||^2. \tag{6.2}$$

The Wald test with nominal significance level α rejects H_0 if $W_n > \chi^2_{q,1-\alpha}$, where $\chi^2_{q,\alpha}$ is the α quantile of a χ^2 random variable with q degrees of freedom. The Wald statistic can also be defined using the covariance matrix estimator $\widehat{V}_{\sigma,n}$. In this case, the Wald statistic divided by q equals the F statistic. Hence, the results given below are applicable to the F test (with the $/q$ modification).

Let $r(\mathcal{C})_j$ denote the j-th element of $r(\mathcal{C})$.

Properties of the t and Wald tests are given in the following theorem.

Theorem 13. *Suppose Assumptions* 1-5 *hold. Let R be a full row rank $q \times k$ matrix. Then, under H_0,*
(a) $P\left(|T_n| > z_{1-\alpha/2}\right) \to \alpha$ *and* $P\left(T_n > z_{1-\alpha}\right) \to \alpha$ *when Assumption* CU *holds,*
(b) $P\left(|T_n| > z_{1-\alpha/2}\right) \to 1$ *when* $r(\mathcal{C})_j \neq 0$ *a.s.,*
(c) $P\left(T_n > z_{1-\alpha}\right) \to 1$ *when* $r(\mathcal{C})_j > 0$ *a.s.,*
(d) $P\left(W_n > \chi^2_{q,1-\alpha}\right) \to \alpha$ *when Assumption* CU *holds, and*
(e) $P\left(W_n > \chi^2_{q,1-\alpha}\right) \to 1$ *when* $Rr(\mathcal{C}) \neq 0$ *a.s.*

Comments. 1. The results of the Theorem continue to hold if the t and Wald statistics are defined with $\widehat{V}_{\sigma,n}$ in place of $\widehat{V}_n$, provided (4.8) holds in parts (a) and (d). Hence, the results of the Theorem for the Wald test also apply to the F test.
2. Parts (a) and (d) of the Theorem show that t, Wald, and F tests are asymptotically valid in the presence of common shocks provided Assumption CU holds. On the other hand, parts (b), (c), and (e) of the Theorem show that t, Wald, and F tests typically reject the null hypothesis with probability that goes to one when Assumption CU fails to hold. This occurs because $|\sqrt{n} r_n(\mathcal{C})_j| \to_p \infty$, $\sqrt{n} r_n(\mathcal{C})_j \to_p \infty$, and $||\sqrt{n} R r_n(\mathcal{C})||^2 \to_p \infty$ in parts (b), (c), and (e), respectively. In this case, the probability of over-rejection increases as the sample size increases.
3. As stated, Theorem 13 does not cover the case where Assumption *CU* does not hold, but $r(\mathcal{C})_j = 0$ a.s. when a t test is considered, or $Rr(\mathcal{C}) = 0$ a.s. when a

Wald test is considered. Results for these cases, however, can be determined using Theorems 10 and 11. In these cases, $|\sqrt{n}r_n(\mathcal{C})_j| = 0 \nrightarrow_p \infty$, $\sqrt{n}r_n(\mathcal{C})_j = 0 \nrightarrow_p \infty$, and $||\sqrt{n}Rr_n(\mathcal{C})||^2 = 0 \nrightarrow_p \infty$, which means that the t and Wald test statistics have well-defined asymptotic distributions under the null hypothesis and, hence, do not reject the null with probability that goes to one under the null hypothesis. But, $\widehat{V}_n$ is not consistent for $V_{\mathcal{C}}$ in general when Assumption CU does not hold. Hence, T_n and W_n do not have standard normal and χ^2 distributions under the null hypothesis and do not reject the null with asymptotic probability equal to α in general. Thus, t, Wald, and F tests are not asymptotically valid in the case under consideration, but their behavior is likely to be much superior to that when $r(\mathcal{C})_j \neq 0$ a.s., $r(\mathcal{C})_j > 0$ a.s., or $Rr(\mathcal{C}) \neq 0$ a.s.

4. The standard $100(1-\alpha)\%$ confidence interval for $\beta_{0,j}$ based on $\widehat{\beta}_{n,j}$ is

$$CI_{\beta_{0,j}} = [\widehat{\beta}_{n,j} - \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{[\widehat{V}_n]_{j,j}},\ \widehat{\beta}_{n,j} + \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{[\widehat{V}_n]_{j,j}}]. \tag{6.3}$$

By a standard and simple argument, the behavior of $CI_{\beta_{0,j}}$ is determined by the behavior under the null hypothesis of the t statistic T_n. In consequence, the results of Theorem 13 imply that under Assumptions 1-5 and Assumption CU, the coverage probability of $CI_{\beta_{0,j}}$ converges to $1-\alpha$ as $n \to \infty$, as desired. On the other hand, under Assumptions 1-5, if $r(\mathcal{C})_j \neq 0$ a.s., then the coverage probability of $CI_{\beta_{0,j}}$ converges to zero as $n \to \infty$, which is not desired.

7. Instrumental Variables Estimator

In this section, we analyze the standard IV estimator of (β_0, α_0) in the regression model of (2.2). This estimator is employed when it is believed that the regressors and errors may be correlated. We consider the IV estimator based on a k_{IV}-vector of non-constant IV's denoted $Z(\gamma)$ for $\gamma \in \Gamma$ and the constant 1. Thus, the IV vector is $Z^+(\gamma) = (1, Z(\gamma)')'$. We use the same framework as above: the population random elements are $\{W(\gamma) = (Y(\gamma), X(\gamma), S(\gamma)) : \gamma \in \Gamma\}$ and the sample is determined by the random indices $\{\gamma_i : i \geq 1\}$ which satisfy Assumption 1. In the present situation, $Z(\gamma)$ is a component of $S(\gamma)$.

Let $Z_i = Z(\gamma_i)$ and $Z_i^+ = Z^+(\gamma_i)$. Let $\overline{Z}_n = n^{-1}\sum_{i=1}^n Z_i$. Let $Y = (Y_1, ..., Y_n)'$, $X = (X_1, ..., X_n)'$, $U = (U_1, ..., U_n)'$, $Z = (Z_1, ..., Z_n)'$, $Z^+ = (Z_1^+, ..., Z_n^+)'$, and $1_n = (1, ..., 1)'$. For any full column rank matrix A, let $P_A = A(A'A)^{-1}A'$.

7.1. Probability Limit of the IV Estimator

The IV slope coefficient estimator, $\widetilde{\beta}_n$, equals the LS slope coefficient estimator from the regression of $P_{Z^+}Y$ on $P_{Z^+}1_n$ and $P_{Z^+}X$. Note that $P_{Z^+}P_{1_n} = P_{1_n}$ and $P_{Z^+} = P_{Z-1_n\overline{Z}_n'} + P_{1_n}$. Using these properties and the LS partitioned regression formula, we obtain

$$
\begin{aligned}
\widetilde{\beta}_n &= \left((P_{Z^+}X)'(I_n - P_{1_n})P_{Z^+}X\right)^{-1}(P_{Z^+}X)'(I_n - P_{1_n})P_{Z^+}Y \\
&= \left(X'P_{Z-1_n\overline{Z}'_n}X\right)^{-1}X'P_{Z-1_n\overline{Z}'_n}Y \\
&= \beta_0 + \left(X'P_{Z-1_n\overline{Z}'_n}X\right)^{-1}X'P_{Z-1_n\overline{Z}'_n}U \qquad (7.1)\\
&= \beta_0 + \\
&\left(n^{-1}\sum_{i=1}^{n}X_i(Z_i-\overline{Z}_n)'\left(n^{-1}\sum_{i=1}^{n}(Z_i-\overline{Z}_n)(Z_i-\overline{Z}_n)'\right)^{-1}n^{-1}\sum_{i=1}^{n}(Z_i-\overline{Z}_n)X_i'\right)^{-1} \\
&\times\left(n^{-1}\sum_{i=1}^{n}X_i(Z_i-\overline{Z}_n)'\left(n^{-1}\sum_{i=1}^{n}(Z_i-\overline{Z}_n)(Z_i-\overline{Z}_n)'\right)^{-1}n^{-1}\sum_{i=1}^{n}(Z_i-\overline{Z}_n)U_i\right).
\end{aligned}
$$

The IV intercept estimator, $\widetilde{\alpha}_n$, equals the LS intercept estimator from the regression of $P_{Z^+}Y$ on $P_{Z^+}1_n$ and $P_{Z^+}X$. Thus, we have

$$
\begin{aligned}
\widetilde{\alpha}_n &= n^{-1}1_n'P_{Z^+}Y - n^{-1}1_n'P_{Z^+}X\widetilde{\beta}_n \\
&= n^{-1}1_n'Y - n^{-1}1_n'X\widetilde{\beta}_n \\
&= \alpha_0 + \overline{U}_n - \overline{X}_n(\widetilde{\beta}_n - \beta_0), \qquad (7.2)
\end{aligned}
$$

using $P_{Z^+}1_n = 1_n$.

Let

$$
\begin{aligned}
B_{\mathcal{C}} &= \Delta_{\mathcal{C}}\zeta_{\mathcal{C}}^{-1}\Delta_{\mathcal{C}}', \text{ where} \\
\Delta_{\mathcal{C}} &= E(X_iZ_i'|\mathcal{C}) - E(X_i|\mathcal{C})E(Z_i|\mathcal{C})' \text{ and} \\
\zeta_{\mathcal{C}} &= E(Z_iZ_i'|\mathcal{C}) - E(Z_i|\mathcal{C})E(Z_i|\mathcal{C})'. \qquad (7.3)
\end{aligned}
$$

To establish the probability limits of $\widetilde{\beta}_n$ and $\widetilde{\alpha}_n$, we replace Assumption 2 with the following:

Assumption IV-2. (a) $E||Z_i||^2 = \int E||Z(\gamma)||^2 dG(\gamma) < \infty$.
(b) $E||Z_iU_i|| = \int E||Z(\gamma)U(\gamma)||dG(\gamma) < \infty$.
(c) $E||Z_iX_i'|| = \int E||Z(\gamma)X(\gamma)'||dG(\gamma) < \infty$.
(d) $B_{\mathcal{C}} > 0$ a.s.

The probability limit of $\widetilde{\beta}_n$ deviates from β_0 by

$$r_{IV}(\mathcal{C}) = B_{\mathcal{C}}^{-1}\Delta_{\mathcal{C}}\zeta_{\mathcal{C}}^{-1}[E(Z_iU_i|\mathcal{C}) - E(Z_i|\mathcal{C})E(U_i|\mathcal{C})]. \qquad (7.4)$$

The probability limit of $\widetilde{\alpha}_n$ deviates from α_0 by

$$s_{IV}(\mathcal{C}) = E(U_i|\mathcal{C}) - E(X_i|\mathcal{C})r_{IV}(\mathcal{C}). \qquad (7.5)$$

Using (7.1), (7.2), and Lemma 2, the probability limits of $\widetilde{\beta}_n$ and $\widetilde{\alpha}_n$ are easily obtained:

Theorem 14. *Suppose Assumptions* 1 *and* IV-2 *hold. Then,*

$$\widetilde{\beta}_n \to_p \beta_0 + r_{IV}(\mathcal{C}) \text{ and}$$
$$\widetilde{\alpha}_n \to_p \alpha_0 + s_{IV}(\mathcal{C}).$$

Comments. 1. The convergence in the Theorem holds jointly and almost surely.
2. A necessary and sufficient condition for $r_{IV}(\mathcal{C}) = 0$ is

Assumption IV-CU. $Cov(Z_i, U_i|\mathcal{C}) = 0$ a.s.

Assumption IV-CU is not implied by the standard assumption for IV estimators that $Cov(Z_i, U_i) = 0$ for $i \geq 1$.
3. Necessary and sufficient conditions for $r_{IV}(\mathcal{C}) = 0$ and $s_{IV}(\mathcal{C}) = 0$ are Assumption IV-CU plus the following assumption:

Assumption IV-CMZ. $E(U_i|\mathcal{C}) = 0$ a.s.

4. The Theorem shows that under Assumptions 1 and IV-2, $\widetilde{\beta}_n \to_p \beta_0$ if and only if Assumption IV-CU holds, and $(\widetilde{\beta}_n, \widetilde{\alpha}_n) \to_p (\beta_0, \alpha_0)$ if and only if Assumptions IV-CU and IV-CMZ hold.
5. The Theorem indicates that consistent estimators of β_0 can be constructed even if Assumption CU fails to hold, which implies that the LS estimator is inconsistent, provided IV's Z_i are available that are uncorrelated with U_i conditional on $\mathcal{C}$.

One can provide sufficient conditions for Assumption IV-CU to hold, and to fail to hold, in terms of a *factor structure* for the errors and IV's that is analogous to that considered in Section 3.2 for the errors and regressors. Specifically, define Assumptions IV-SF1, ..., IV-SF4, and IV-STD1 as Assumptions SF1, ..., SF4, and STD1 are defined, respectively, but with X replaced by Z throughout, with C_2 being a random $k_{IV} \times d_2$ matrix for $d_2 \geq k_{IV} \geq k$, and with $Z(\gamma)$ added as a component of $S(\gamma)$. Then, $\mathcal{C} \supseteq \sigma(C_1, C_2)$ (where " $\supseteq$ " appears rather than " $=$ " because the regressors could add some randomness to $\mathcal{C}$).

In this case, results analogous to those of Corollary 6 and Theorem 7 hold: (i) Suppose Assumptions 1, IV-SF1, and IV-SF2 hold. Then, Assumption IV-CU holds and $r_{IV}(\mathcal{C}) = 0$. (ii) Suppose Assumptions 1, IV-SF1, IV-SF2, and IV-SF3 hold. Then, Assumptions IV-CU and IV-CMZ hold, $r_{IV}(\mathcal{C}) = 0$, and $s_{IV}(\mathcal{C}) = 0$. (iii) Suppose Assumptions 1, IV-SF1, and IV-SF4 hold. Then, Assumption IV-STD1 holds, but Assumption IV-CU does not hold. The proofs of these results are analogous to those of Corollary 6 and Theorem 7 and, hence, are not given.

Similarly, one can provide sufficient conditions for Assumption IV-CU to hold in terms of a *heterogeneous factor structure* for the errors and IV's that is analogous to that in Section 3.3 for the errors and regressors. Define Assumptions IV-HF1, IV-HF2, and IV-HF3 as Assumptions HF1, HF2, and HF3 are defined, respectively, but with X replaced by Z throughout, with $C_2(\cdot)$ being a random $k_{IV} \times d_2$ matrix-valued function for $d_2 \geq k_{IV} \geq k$, and with $Z(\gamma)$ added as a component of $S(\gamma)$. Then, $\mathcal{C} \supseteq \sigma(C_1(\cdot), C_2(\cdot))$. Results analogous to those of Corollary 8 hold: (i) Suppose Assumptions 1, IV-HF1, and IV-HF2 hold. Then, Assumption IV-CU holds and

$r_{IV}(\mathcal{C}) = 0$. (ii) Suppose Assumptions 1, IV-HF1, IV-HF2, and IV-HF3 hold. Then, Assumptions IV-CU and IV-CMZ hold, $r_{IV}(\mathcal{C}) = 0$, and $s_{IV}(\mathcal{C}) = 0$.

Finally, general sufficient conditions for Assumption IV-CU can be given by defining Assumptions IV-FF1, IV-FF2, and IV-FF3 to be the same as Assumptions FF1, FF2, and FF3, respectively, but with X replaced by Z throughout and with the range of $Z(\cdot, \gamma)$ being $R^{k_{IV}}$ rather than R^k. By the same proof as for Theorem 9, Assumptions 1, IV-FF1, and IV-FF2 imply Assumption IV-CU; and Assumptions 1, IV-FF1, IV-FF2, and IV-FF3 imply Assumption IV-CMZ.

7.2. Asymptotic Distributions of the IV Estimator and Test Statistics

Results analogous to those of Theorems 10, 11, and 13 hold for IV estimators and test statistics. Specifically, when Assumption IV-CU holds, $n^{1/2}(\widetilde{\beta}_n - \beta_0)$ is asymptotically mixed normal; IV-based covariance matrix estimators are consistent for the random asymptotic mixing matrix; t and Wald statistics based on IV estimators have standard normal and chi-squared asymptotic distributions, respectively, under the null hypothesis; and confidence intervals based on IV estimators have asymptotically correct coverage probabilities. When Assumption IV-CU does not hold, none of these results hold. For brevity, we only provide explicit results here for the case where Assumption IV-CU holds.

Under Assumption IV-CU, the conditional asymptotic variance, $V_{\mathcal{C}}$, of the normalized IV estimator of β_0 given $\mathcal{C}$ is

$$\begin{aligned} V_{\mathcal{C}} &= B_{\mathcal{C}}^{-1} \Omega_{\mathcal{C}} B_{\mathcal{C}}^{-1}, \text{ where} \\ \Omega_{\mathcal{C}} &= \Delta_{\mathcal{C}} \zeta_{\mathcal{C}}^{-1} M_{\mathcal{C}} \zeta_{\mathcal{C}}^{-1} \Delta_{\mathcal{C}}' \text{ and} \\ M_{\mathcal{C}} &= E([U_i - E(U_i|\mathcal{C})]^2 Z_i Z_i' | \mathcal{C}). \end{aligned} \tag{7.6}$$

If the errors are homoskedastic conditional on $\mathcal{C}$ and Z_i, i.e., $E([U_i - E(U_i|\mathcal{C})]^2 | \mathcal{C}, Z_i) = \sigma_{\mathcal{C}}^2$ a.s. for $\sigma_{\mathcal{C}}^2 = Var(U_i|\mathcal{C})$, then $\Omega_{\mathcal{C}}$ and $V_{\mathcal{C}}$ simplify to $\Omega_{\mathcal{C}} = \sigma_{\mathcal{C}}^2 B_{\mathcal{C}}$ and $V_{\mathcal{C}} = \sigma_{\mathcal{C}}^2 B_{\mathcal{C}}^{-1}$, respectively.

The usual heteroskedasticity-robust estimator of the asymptotic variance of $\widetilde{\beta}_n$ is

$$\begin{aligned} \widetilde{V}_n &= \widetilde{B}_n^{-1} \widetilde{\Omega}_n \widetilde{B}_n^{-1}, \text{ where} \\ \widetilde{B}_n &= \widetilde{\Delta}_n \widetilde{\zeta}_n^{-1} \widetilde{\Delta}_n', \ \widetilde{\Omega}_n = \widetilde{\Delta}_n \widetilde{\zeta}_n^{-1} \widetilde{M}_n \widetilde{\zeta}_n^{-1} \widetilde{\Delta}_n', \\ \widetilde{\Delta}_n &= n^{-1} \sum_{i=1}^n (X_i - \overline{X}_n) Z_i', \ \widetilde{\zeta}_n = n^{-1} \sum_{i=1}^n Z_i Z_i', \\ \widetilde{M}_n &= n^{-1} \sum_{i=1}^n \widetilde{U}_i^2 Z_i Z_i', \text{ and } \widetilde{U}_i = Y_i - \widetilde{\alpha}_n - X_i' \widetilde{\beta}_n. \end{aligned} \tag{7.7}$$

The usual estimator of the asymptotic variance of $\widetilde{\beta}_n$ that relies on homoskedasticity of the errors is $\widetilde{V}_{\sigma,n} = \widetilde{\sigma}_n^2 \widetilde{B}_n^{-1}$, where $\widetilde{\sigma}_n^2 = (n-k-1)^{-1} \sum_{i=1}^n \widetilde{U}_i^2$.

We consider the same hypotheses and the same t and Wald statistics, T_n and W_n, as in Section 6, but with the LS estimator replaced by the IV estimator and with $\widehat{V}_n$ replaced by $\widetilde{V}_n$.

Let Assumption IV-3 denote Assumption 3 with X_i replaced by Z_i. Let Assumption IV-4 denote Assumption 4 with $||X_i||^4$ replaced by $||X_i||^2||Z_i||^2$ and with $||X_i||^3|U_i|$ replaced by $||Z_i||^2||X_i|| \cdot |U_i|$. Let Assumption IV-5 state that $\Omega_{\mathcal{C}} > 0$ a.s. and $\sigma^2_{\mathcal{C}} > 0$ a.s.

The asymptotic distribution of $n^{1/2}(\widetilde{\beta}_n - \beta_0)$, the probability limits of $\widetilde{V}_n$ and $\widetilde{V}_{\sigma,n}$, and the asymptotic null distributions of the IV-based t and Wald statistics are given in the following theorem for the case where Assumption IV-CU holds.

Theorem 15. *Suppose Assumptions* IV-CU, 1, *and* IV-2,..., IV-5 *hold. Let* $Z \sim N(0, I_k)$ *be a standard normal* k*-vector that is independent of* $\mathcal{C}$. *Let* R *be a full row rank* $q \times k$ *matrix. Then,*
(a) $n^{1/2}(\widetilde{\beta}_n - \beta_0) \to_d V_{\mathcal{C}}^{1/2} Z$,
(b) $V_{\mathcal{C}}^{-1/2} n^{1/2}(\widetilde{\beta}_n - \beta_0) \to_d Z$,
(c) $\widetilde{B}_n \to_p B_{\mathcal{C}}$, $\widetilde{\Omega}_n \to_p \Omega_{\mathcal{C}}$, $\widetilde{V}_n \to_p V_{\mathcal{C}}$, $\widetilde{\sigma}^2_n \to_p \sigma^2_{\mathcal{C}}$, *and* $\widetilde{V}_{\sigma,n} \to_p \sigma^2_{\mathcal{C}} B_{\mathcal{C}}^{-1}$,
(d) $P\left(|T_n| > z_{1-\alpha/2}\right) \to \alpha$ *and* $P\left(T_n > z_{1-\alpha}\right) \to \alpha$ *under* H_0, *and*
(e) $P\left(W_n > z_{1-\alpha}\right) \to \alpha$ *under* H_0.

Comments. 1. Parts (d) and (e) of the Theorem continue to hold if the t and Wald statistics are defined with $\widetilde{V}_{\sigma,n}$ in place of $\widetilde{V}_n$, provided $E([U_i - E(U_i|\mathcal{C})]^2|\mathcal{C}, Z_i) = \sigma^2_{\mathcal{C}}$ a.s.
2. Parts (d) and (e) of the Theorem show that if Assumption IV-CU holds then t and Wald tests are asymptotically valid in the presence of common shocks. On the other hand, in analogy to parts (b), (c), and (e) of Theorem 13, IV-based t and Wald tests typically reject the null hypothesis with probability that goes to one when Assumption CU fails to hold. For brevity, such results are not given here.
3. Results for IV-based confidence intervals, analogous to those in Comment 4 following Theorem 13, hold.
4. The proof of Theorem 15 is quite similar to that of Theorems 10, 11, and 13 and, hence, is not given.

8. Extensions

8.1. Panel Models with Fixed T

The results of this paper can be extended to cover panel regression models with a fixed number of time periods T. In a panel model, $W(\gamma)$ is defined to include random variables for all time periods $t = 1, ..., T$ for population unit γ, and all random variables have a t subscript added, e.g., $Y(\gamma)$ is replaced by $Y_t(\gamma)$. The model is given by

$$Y_t(\gamma) = \alpha_0 + X_t(\gamma)'\beta_0 + U_t(\gamma) \text{ for } t = 1, ..., T \tag{8.1}$$

and $\gamma \in \Gamma$. Samples of n population units for $n \geq 1$ are obtained by drawing indices $\{\gamma_i : i \geq 1\}$ according to Assumption 1. The LS and IV estimators of β_0 and α_0 are defined as above but with all sums taken over $t = 1, ..., T$ as well as $i = 1, ..., n$ and with normalization by $(nT)^{-1}$ rather than n^{-1}. In the present case, for the LS estimator, $r(\mathcal{C})$ and $s(\mathcal{C})$ are defined with $E(X_i X_i'|\mathcal{C})$ replaced by $T^{-1}\sum_{t=1}^T E(X_{it}X_{it}'|\mathcal{C})$, where $X_{it} = X_t(\gamma_i)$, and likewise for $E(X_i|\mathcal{C})$, $E(X_i U_i|\mathcal{C})$, and $E(U_i|\mathcal{C})$. Consistency of the LS estimator of β_0 depends on whether $r(\mathcal{C}) = 0$ a.s. just as above.

With a panel regression model, one might want to analyze the properties of the *within* and *between* estimators. This can be done in an analogous fashion to the analysis of the LS and IV estimators. For the within estimator, the model we consider is

$$Y_t(\gamma) = \alpha(\gamma) + X_t(\gamma)'\beta_0 + U_t(\gamma) \text{ for } t = 1, ..., T \tag{8.2}$$

and $\gamma \in \Gamma$, where $\alpha(\gamma)$ is a population unit γ fixed effect that may be random or non-random. The within estimator, $\widehat{\beta}_{W,n}$, is

$$\begin{aligned}\widehat{\beta}_{W,n} &= \left((nT)^{-1}\sum_{i=1}^n\sum_{t=1}^T X_{it}X_{it}' - \overline{X}_{T,i}\overline{X}_{T,i}'\right)^{-1}\left((nT)^{-1}\sum_{i=1}^n\sum_{t=1}^T X_{it}Y_{it} - \overline{X}_{T,i}\overline{Y}_{T,i}\right)\\ &= \beta_0 + \left((nT)^{-1}\sum_{i=1}^n\sum_{t=1}^T X_{it}X_{it}' - \overline{X}_{T,i}\overline{X}_{T,i}'\right)^{-1}\left((nT)^{-1}\sum_{i=1}^n\sum_{t=1}^T X_{it}U_{it} - \overline{X}_{T,i}\overline{U}_{T,i}\right),\end{aligned}$$

where

$$\begin{aligned}&X_{it} = X_t(\gamma_i),\ Y_{it} = Y_t(\gamma_i),\ U_{it} = U_t(\gamma_i),\ \overline{X}_{T,i} = (nT)^{-1}\sum_{i=1}^n\sum_{t=1}^T X_{it},\\ &\overline{Y}_{T,i} = (nT)^{-1}\sum_{i=1}^n\sum_{t=1}^T Y_{it},\ \text{and } \overline{U}_{T,i} = (nT)^{-1}\sum_{i=1}^n\sum_{t=1}^T U_{it}.\end{aligned} \tag{8.3}$$

The probability limit of $\widehat{\beta}_{W,n}$ is $\beta_0 + r_W(\mathcal{C})$, where

$$r_W(\mathcal{C}) = \left(T^{-1}\sum_{t=1}^T E(X_{it}X_{it}' - \overline{X}_{T,i}\overline{X}_{T,i}'|\mathcal{C})\right)^{-1}\left(T^{-1}\sum_{t=1}^T E(X_{it}U_{it} - \overline{X}_{T,i}\overline{U}_{T,i}|\mathcal{C})\right). \tag{8.4}$$

The analogue of Assumption CU for the within estimator is

$$T^{-1}\sum_{t=1}^T E(X_{it}U_{it} - \overline{X}_{T,i}\overline{U}_{T,i}|\mathcal{C}) = 0 \text{ a.s.} \tag{8.5}$$

Consistency of the within estimator depends on whether (8.5) holds.

The asymptotic distributions of the within estimator and test statistics based on it can be determined in a manner analogous to that used above for the LS estimator. The asymptotic properties of the between estimator can be determined in a similar fashion.

8.2. Clustered Sampling

The results of the paper can be extended to cover *clustered* sampling. In this case, γ is taken to be a cluster, and Γ is the population of clusters. Then, $W(\gamma)$ is defined to include random variables for all population units in the γ-th cluster. Population units in the γ-th cluster are indexed by $b = 1, ..., B$, where $B \leq \infty$ denotes the cluster size. A sample of n clusters is selected via iid indices $\{\gamma_i : i = 1, ..., n\}$ that satisfy Assumption 1. For each cluster γ_i selected, a random sample of T population units from the cluster is drawn.

The population units selected from the γ_i-th cluster are denoted with t subscripts for $t = 1, ..., T$. For example, the regressor variables are $X_t(\gamma_i)$ for $t = 1, ..., T$. Then, as in the panel model of (8.1), the LS, IV, and covariance matrix estimators are defined with sums taken over $t = 1, ..., T$ as well as $i = 1, ..., n$ and with normalization by $(nT)^{-1}$ rather than n^{-1}. The definitions of $r(\mathcal{C})$ and $s(\mathcal{C})$ are altered as in the panel model of (8.1). The total sample size in this case is nT.

8.3. Generalized Methods of Moments Estimators

The results of this paper can be extended to nonlinear estimators, such as GMM and maximum likelihood estimators. Here we show how probability limit results can be obtained for GMM estimators. For brevity, we do not provide asymptotic distributional results. Such results can be obtained using the same sort of method as for LS estimators.

Consider the moment functions

$$g(Z(\gamma), \theta), \tag{8.6}$$

where $g(\cdot, \cdot)$ is a known k_g vector-valued function, $Z(\gamma)$ is a vector of random variables, θ is a k_θ vector of unknown parameters, and $\gamma \in \Gamma$ is a population unit index. The standard assumption used to obtain consistency of the GMM estimator for the true parameter $\theta_0 \in \Theta$ is

Assumption GMM-STD. (a) $Eg(Z(\gamma), \theta_0) = 0$ for all $\gamma \in \Gamma$.
(b) $Eg(Z(\gamma), \theta) \neq 0$ if $\theta \neq \theta_0$ and $\theta \in \Theta$ for all $\gamma \in \Gamma$.

As above, we let $S(\gamma) \in \mathcal{S}$ denote some supplementary variables that may include characteristics of population unit γ and/or stochastic terms that are common to some or all of the units in the population. We let $W(\gamma) = (Z(\gamma), S(\gamma))$.

As above, samples of size n for $n \geq 1$ are obtained by drawing indices $\{\gamma_i : i = 1, ..., n\}$ randomly from Γ such that Assumption 1 holds. Let $W_i = W(\gamma_i)$, $Z_i = Z(\gamma_i)$, and $S_i = S(\gamma_i)$. The random variables $\{Z_i : i = 1, ..., n\}$ constitute the observed sample.

The one-step GMM estimator, $\widehat{\theta}_{1,n}$, of θ_0 minimizes

$$Q_{1,n}(\theta) = \left(n^{-1} \sum_{i=1}^{n} g(Z_i, \theta) \right)' \Sigma^{-1} \left(n^{-1} \sum_{i=1}^{n} g(Z_i, \theta) \right) \tag{8.7}$$

over the parameter space $\Theta \subset R^{k_\theta}$ for some non-random symmetric positive definite $k_g \times k_g$ matrix Σ, such as $\Sigma = I_{k_g}$.

The two-step GMM estimator, $\widehat{\theta}_{2,n}$, of θ_0 minimizes

$$Q_{2,n}(\theta) = \left(n^{-1}\sum_{i=1}^{n} g(Z_i,\theta)\right)' \widehat{\Sigma}_n^{-1} \left(n^{-1}\sum_{i=1}^{n} g(Z_i,\theta)\right) \tag{8.8}$$

over the parameter space Θ, where

$$\widehat{\Sigma}_n = n^{-1}\sum_{i=1}^{n} g(Z_i,\widehat{\theta}_{1,n})g(Z_i,\widehat{\theta}_{1,n})'. \tag{8.9}$$

Note that $\widehat{\Sigma}_n$ could be defined with $g(Z_i,\widehat{\theta}_{1,n})$ replaced by the deviation of $g(Z_i,\widehat{\theta}_{1,n})$ from its the sample average.

One could also consider the continuous updating GMM estimator. For brevity, we do not do so.

We employ the following uniqueness assumptions:

Assumption GMM-1. (a) $Q_{1,\mathcal{C}}(\theta) = E(g(Z_i,\theta)|\mathcal{C})'\Sigma^{-1}E(g(Z_i,\theta)|\mathcal{C})$ has a unique minimum over Θ a.s., denoted $\theta_1(\mathcal{C})$.
(b) $Q_{2,\mathcal{C}}(\theta) = E(g(Z_i,\theta)|\mathcal{C})'\Sigma_{\mathcal{C}}^{-1}E(g(Z_i,\theta)|\mathcal{C})$ has a unique minimum over Θ a.s., denoted $\theta_2(\mathcal{C})$, where $\Sigma_{\mathcal{C}} = E(g(Z_i,\theta_1(\mathcal{C}))g(Z_i,\theta_1(\mathcal{C}))'|\mathcal{C})$.

Next, we introduce some basic assumptions on the parameter space, moment conditions, and weight matrices Σ and $\Sigma_{\mathcal{C}}$:

Assumption GMM-2. (a) Θ is a compact subset of R^{k_θ}.
(b) $g(Z_i,\theta)$ is continuously differentiable in θ on Θ with probability one.
(c) $E\sup_{\theta\in\Theta}||g(Z_i,\theta)||^2 < \infty$ and $E\sup_{\theta\in\Theta}||(\partial/\partial\theta')g(Z_i,\theta)||\cdot(1+\sup_{\theta\in\Theta}||g(Z_i,\theta)||) < \infty$.
(d) Σ is positive definite, and $\Sigma_{\mathcal{C}}$ is positive definite a.s.

The probability limits of the one-step and two-step GMM estimators are given in the following theorem:

Theorem 16. *Suppose Assumptions* GMM-1 *and* GMM-2 *hold. Then,* $\widehat{\theta}_{1,n} \to_p \theta_1(\mathcal{C})$ *and* $\widehat{\theta}_{2,n} \to_p \theta_2(\mathcal{C})$.

Next, we give a sufficient condition for consistency of GMM estimators.

Assumption GMM-CON. (a) $E(g(Z_i,\theta_0)|\mathcal{C}) = 0$ a.s.
(b) $E(g(Z_i,\theta)|\mathcal{C}) \neq 0$ if $\theta \neq \theta_0$ and $\theta \in \Theta$ a.s.

Theorem 17. *Suppose Assumptions* GMM-2 *and* GMM-CON *hold. Then, the one-step and two-step GMM estimators,* $\widehat{\theta}_{1,n}$ *and* $\widehat{\theta}_{2,n}$*, are consistent.*

Comments. **1.** Assumption GMM-CON requires that the standard condition for consistency of GMM estimators holds conditionally on the common shocks a.s., not unconditionally.

2. Assumption GMM-CON is close to being a necessary condition for consistency of $\widehat{\theta}_{1,n}$ and $\widehat{\theta}_{2,n}$. The reason is as follows. Under Assumption GMM-2, one can show that the distance between $\widehat{\theta}_{j,n}$ and the set of minimizers of $Q_{j,\mathcal{C}}(\theta)$ over Θ converges in probability to zero for $j = 1, 2$. If Assumption GMM-CON fails to hold, then with positive probability $E(g(Z_i, \theta_{\mathcal{C}}^*)|\mathcal{C}) = 0$ for some $\theta_{\mathcal{C}}^* \in \Theta$ with $\theta_{\mathcal{C}}^* \neq \theta_0$. Hence, the set of minimizers of $Q_{j,\mathcal{C}}(\theta)$ over Θ includes both θ_0 and $\theta_{\mathcal{C}}^*$ with positive probability for $j = 1, 2$. In this case, one typically has $\limsup_{n\to\infty} P(|\widehat{\theta}_{j,n} - \theta_{\mathcal{C}}^*| < \varepsilon) > 0$ for all $\varepsilon > 0$ for $j = 1, 2$, which implies that $\widehat{\theta}_{j,n}$ does not converge in probability to θ_0. (However, providing simple conditions under which the latter *must* occur does not appear to be easy.)

9. Conclusion

This paper calls into question the standard assumption that observations in cross-section econometric models are independent. The paper takes a further step away from independence than does the literature on models with group effects or spatial correlation. The paper allows for common shocks of a very general nature. They may affect all population units or just some population units. Their effect may depend on characteristics of the population unit in a discrete or continuous fashion. Their effect may be local or global in nature.

The paper shows that necessary and sufficient conditions for consistency of LS (or IV) slope coefficient estimators in regression models with common shocks are that the errors are uncorrelated with the regressors (or IVs) conditional on the σ-field generated by the common shocks. The LS and IV estimators are shown to have a mixed normal asymptotic distribution after suitable centering and scaling. The paper shows that, when the LS (or IV) estimators are consistent, the t, Wald, and F tests and confidence intervals based on them are asymptotically valid.

On the other hand, when the errors are correlated with the regressors (or IVs) conditional on the common shocks a.s., then the null rejection probabilities of t, Wald, and F tests based on the LS (or IV) estimators converge to one as $n \to \infty$ and confidence interval coverage probabilities converge to zero as $n \to \infty$. Hence, common shocks can have an innocuous or detrimental effect on estimators and tests depending on the properties of the errors, regressors, and IVs conditional on the common shocks.

10. Appendix of Proofs

Proof of Theorem 3. With convergence in probability replaced by convergence almost surely, the Theorem follows straightforwardly from (3.1), (3.2), and Lemma 2 using Assumptions 1 and 2. The convergence in probability result then follows from the almost sure convergence result. □

Proof of Lemma 5. Using (3.3), which relies on Assumption 1, we have

$$\begin{aligned} E(U_i|\mathcal{C}) &= \int E(U(\gamma)|\mathcal{C})dG(\gamma), \\ E(X_i|\mathcal{C}) &= \int E(X(\gamma)|\mathcal{C})dG(\gamma), \text{ and} \\ E(X_iU_i|\mathcal{C}) &= \int E(X(\gamma)U(\gamma)|\mathcal{C})dG(\gamma). \end{aligned} \tag{10.1}$$

Equation (10.1) and Assumption CMZ give

$$E(U_i|\mathcal{C}) = \int E(U(\gamma)|\mathcal{C})dG(\gamma) = 0. \tag{10.2}$$

Combining the results of (10.1) gives

$$\begin{aligned} & E(X_iU_i|\mathcal{C}) - E(X_i|\mathcal{C})E(U_i|\mathcal{C}) \\ = & \int E(X(\gamma)U(\gamma)|\mathcal{C})dG(\gamma) - \int E(X(\gamma)|\mathcal{C})dG(\gamma) \int E(U(\gamma)|\mathcal{C})dG(\gamma), \\ = & \int [E(X(\gamma)U(\gamma)|\mathcal{C}) - E(X(\gamma)|\mathcal{C})E(U(\gamma)|\mathcal{C})]dG(\gamma) \\ = & \ 0, \end{aligned} \tag{10.3}$$

where the second equality holds by Assumption CUγ(b) and the third equality holds by Assumption CUγ(a). □

Proof of Theorem 7. Assumption STD1 holds by the following calculations:

$$\begin{aligned} EX(\gamma)U(\gamma) &= E_{\mathcal{C}}C_2E(X^*(\gamma)U^*(\gamma)'|\mathcal{C})C_1 \\ &= E_{\mathcal{C}}C_2E(X^*(\gamma)U^*(\gamma)')C_1 \\ &= E_{\mathcal{C}}C_2E(X^*(\gamma)U_1^*(\gamma)) + E_{\mathcal{C}}C_2E(X^*(\gamma)U_2^*(\gamma)')C_{11} \\ &= 0, \end{aligned} \tag{10.4}$$

for all $\gamma \in \Gamma$, where $E_{\mathcal{C}}$ denotes expectation with respect to the randomness in $\mathcal{C}$, the first equality holds by Assumption SF1 and iterated expectations, the second equality holds by Assumption SF1(b), the third equality holds by Assumption SF4(a) and (b), and the fourth equality holds by Assumptions SF1(b) and SF4(a) and (c). Analogous calculations give $EU(\gamma) = 0$.

Next we show that Assumption CU does not hold. We have

$$
\begin{aligned}
E(U_i|\mathcal{C}) &= \int E(U(\gamma)|\mathcal{C})dG(\gamma) \\
&= \int E(U^*(\gamma))'dG(\gamma)C_1 \\
&= EU_1^*(\gamma) + C_{11}\int EU_2^*(\gamma)dG(\gamma) \\
&= 0, \qquad (10.5)
\end{aligned}
$$

where the first equality holds by (3.3), the second equality holds by Assumption SF1, the third equality holds by Assumption SF4(a) and (b), and the fourth equality holds by Assumption SF4(a) and (d).

Given (10.5), we have

$$
\begin{aligned}
Cov(X_i, U_i|\mathcal{C}) &= E(X_iU_i|\mathcal{C}) \\
&= \int E(X(\gamma)U(\gamma)|\mathcal{C})dG(\gamma) \\
&= C_2\int E(X^*(\gamma)U^*(\gamma)')dG(\gamma)C_1 \\
&= C_2\int EX^*(\gamma)U_1^*(\gamma)dG(\gamma) + C_2\int E(X^*(\gamma)U_2^*(\gamma))dG(\gamma)C_{11} \\
&= C_2\int E(X^*(\gamma)U_2^*(\gamma))dG(\gamma)C_{11} \\
&\neq 0 \text{ with positive probability}, \qquad (10.6)
\end{aligned}
$$

where the second through fourth equalities hold by the same arguments as in (10.5) and the fifth equality and the inequality hold by Assumption SF4(a) and (d). □

Proof of Theorem 9. We have

$$
\begin{aligned}
E(U_i|\mathcal{C}) &= E(U_i(C(S_{0,i}))|\mathcal{C}) \\
&= E_{S_{0,i}}E(U_i(C(S_{0,i}))|\mathcal{C}, S_{0,i}) \\
&= E_{S_{0,i}}E_{X_i(\cdot),S_{0,i}}(U_i(C(S_{0,i}))|S_{0,i}), \qquad (10.7)
\end{aligned}
$$

where $E_{S_{0,i}}$ denotes expectation with respect to $S_{0,i}$ alone, $E_{X_i(\cdot),S_{0,i}}(\cdot|S_{0,i})$ denotes conditional expectation with respect to $(X_i(\cdot), S_{0,i})$ alone given $S_{0,i}$, the first equality holds by Assumption FF1(d), the second equality holds by iterated expectations, and the third equality holds by Assumption FF1(e) and the fact that $\mathcal{C} = \sigma(C(\cdot))$, which holds by Assumptions FF1(b) and (f).

By similar arguments, we obtain

$$
\begin{aligned}
E(X_iU_i|\mathcal{C}) &= E_{S_{0,i}}E_{X_i(\cdot),S_{0,i}}(X_i(C(S_{0,i}))U_i(C(S_{0,i}))|S_{0,i}) \text{ and} \\
E(X_i|\mathcal{C}) &= E_{S_{0,i}}E_{X_i(\cdot),S_{0,i}}(X_i(C(S_{0,i}))|S_{0,i}). \qquad (10.8)
\end{aligned}
$$

Combining (10.7) and (10.8) gives

$$\begin{aligned}
&E(X_iU_i|\mathcal{C}) - E(X_i|\mathcal{C})E(U_i|\mathcal{C}) \\
&= E_{S_{0,i}}E_{X_i(\cdot),S_{0,i}}(X_i(C(S_{0,i}))U_i(C(S_{0,i}))|S_{0,i}) \\
&\quad -E_{S_{0,i}}E_{X_i(\cdot),S_{0,i}}(X_i(C(S_{0,i}))|S_{0,i}) \cdot E_{S_{0,i}}E_{X_i(\cdot),S_{0,i}}(U_i(C(S_{0,i}))|S_{0,i}) \\
&= E_{S_{0,i}}[E_{X_i(\cdot),S_{0,i}}(X_i(C(S_{0,i}))U_i(C(S_{0,i}))|S_{0,i}) \\
&\quad -E_{X_i(\cdot),S_{0,i}}(X_i(C(S_{0,i}))|S_{0,i}) \cdot E_{X_i(\cdot),S_{0,i}}(U_i(C(S_{0,i}))|S_{0,i})] \\
&= 0, \qquad (10.9)
\end{aligned}$$

where the second equality holds by Assumption FF2(b) because (i) $C(\cdot)$ is independent of $(X_i(\cdot), U_i(\cdot), S_{0,i})$ and, hence, can be conditioned on and (ii) $C(S_{0,i})$ is a constant conditional on $C(\cdot)$ and $S_{0,i}$; and the third equality holds by Assumption FF2(a). This result implies Assumption CU.

By (i) and (ii) of the last paragraph applied to the right-hand side of (10.7) and Assumption FF3, the right-hand side of (10.7) equals zero a.s. Hence, Assumption CMZ holds.

The proof of parts (c) and (d) is the same as the proof of Lemma 5 with X_i, U_i, $X(\gamma)$, $U(\gamma)$, and $E(\cdot|\mathcal{C})$ replaced by $X_i(c)$, $U_i(c)$, $X(c,\gamma)$, $U(c,\gamma)$, and $E(\cdot)$, respectively, using (2.5) in place of (3.3). $\square$

Proof of Theorem 10. To prove part (a), we write

$$\begin{aligned}
&n^{1/2}(\widehat{\beta}_n - \beta_0 - r_n(\mathcal{C})) \\
&= \left(n^{-1}\sum_{i=1}^n X_iX_i' - \overline{X}_n\overline{X}_n'\right)^{-1} n^{-1/2}\sum_{i=1}^n \left([X_i - \overline{X}_n]U_i - E([X_i - E(X_i|\mathcal{C})]U_i|\mathcal{C})\right) \\
&= \left(n^{-1}\sum_{i=1}^n X_iX_i' - \overline{X}_n\overline{X}_n'\right)^{-1} \\
&\times \left(n^{-1/2}\sum_{i=1}^n \{[X_i - E(X_i|\mathcal{C})]U_i - E([X_i - E(X_i|\mathcal{C})]U_i|\mathcal{C}) - [X_i - E(X_i|\mathcal{C})]E(U_i|\mathcal{C})\}\right. \\
&\qquad \left. -[\overline{X}_n - E(X_i|\mathcal{C})]n^{-1/2}\sum_{i=1}^n [U_i - E(U_i|\mathcal{C})]\right) \\
&= \left(B_{\mathcal{C}}^{-1} + o_p(1)\right)\left(n^{-1/2}\sum_{i=1}^n \xi_i\right) + o_p(1), \qquad (10.10)
\end{aligned}$$

where ξ_i is defined in (4.2) and the third equality of (10.10) holds using Lemma 2 to obtain the $B_{\mathcal{C}}^{-1} + o_p(1)$ result, using Lemma 2 to obtain $\overline{X}_n - E(X_i|\mathcal{C}) = o_p(1)$, and using a MDS CLT to obtain

$$n^{-1/2}\sum_{i=1}^n [U_i - E(U_i|\mathcal{C})] = O_p(1). \qquad (10.11)$$

In particular, we apply Corollary 3.1 of Hall and Heyde (1980, p. 59) to obtain (10.11). For $i \geq 1$, let $\mathcal{F}_i$ denote the σ-field generated by $\mathcal{C}$ and $(W_1, ..., W_i)$. Then,

$\{U_i - E(U_i|\mathcal{C}), \mathcal{F}_i : i \geq 1\}$ is a MDS because $\{U_i : i \geq 1\}$ are iid conditional on $\mathcal{C}$, and hence, $E(U_i|\mathcal{F}_{i-1}) = E(U_i|\mathcal{C})$ a.s. A conditional Lindeberg condition holds because, for all $\varepsilon > 0$,

$$\begin{aligned}
&\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} E([U_i - E(U_i|\mathcal{C})]^2 1(|U_i - E(U_i|\mathcal{C})| > n^{1/2}\varepsilon)|\mathcal{F}_{i-1}) \\
&= \lim_{n\to\infty} E([U_i - E(U_i|\mathcal{C})]^2 1(|U_i - E(U_i|\mathcal{C})| > n^{1/2}\varepsilon)|\mathcal{C}) = 0 \text{ a.s.,} \qquad (10.12)
\end{aligned}$$

where the first equality holds because $\{U_i : i \geq 1\}$ are iid conditional on $\mathcal{C}$ and the second equality holds by the dominated convergence theorem using $E([U_i - E(U_i|\mathcal{C})]^2|\mathcal{C}) \leq E(U_i^2|\mathcal{C}) < \infty$ a.s. by Assumption 3(a). In addition, the normalized sums of conditional variances converge as $n \to \infty$ because they do not depend on n:

$$n^{-1} \sum_{i=1}^{n} E([U_i - E(U_i|\mathcal{C})]^2|\mathcal{F}_{i-1}) = E([U_i - E(U_i|\mathcal{C})]^2|\mathcal{C}). \qquad (10.13)$$

Equation (10.13) holds because the conditional variances given $\mathcal{F}_{i-1}$ equal the conditional variances given $\mathcal{C}$ and the latter are identically distributed by exchangeability. Hence, the MDS CLT implies that

$$n^{-1/2} \sum_{i=1}^{n} [U_i - E(U_i|\mathcal{C})] \to_d E([U_i - E(U_i|\mathcal{C})]^2|\mathcal{C}) \times Z^*, \qquad (10.14)$$

where Z^* and $E([U_i - E(U_i|\mathcal{C})]^2|\mathcal{C})$ are independent and $Z^* \sim N(0, 1)$. This, in turn, gives (10.11).

Next, $\{\xi_i, \mathcal{F}_i : i \geq 1\}$ is a MDS by the same argument as above for $\{U_i - E(U_i|\mathcal{C}), \mathcal{F}_i : i \geq 1\}$. By application of the same MDS CLT as above, we obtain

$$n^{-1/2} \sum_{i=1}^{n} \xi_i \to_d \Omega_{\mathcal{C}} \times Z, \qquad (10.15)$$

where $(\Omega_{\mathcal{C}}, B_{\mathcal{C}})$ and Z are independent and $Z \sim N(0, I_k)$. To establish the CLT, we note that a conditional Lindeberg condition holds using the moment conditions of Assumptions 2(a) and 3 and the dominated convergence theorem as above and the conditional variances converge by the same argument as in (10.13). Combining (10.10) and (10.15) gives the result of part (a).

Part (b) of the Theorem holds by the same argument as for part (a), but with all of the terms pre-multiplied by $V_{\mathcal{C}}^{-1/2}$.

Part (c) of the Theorem holds using Lemma 2. □

Proof of Theorem 11. Part (a) holds by Lemma 2.

To prove part (b), for notational simplicity, suppose X_i is a scalar (otherwise one can establish the results element by element). Using Theorem 3, we have

$$\begin{aligned}
\widehat{U}_i &= [U_i - E(U_i|\mathcal{C})] - [\widehat{\alpha}_n - \alpha_0 - E(U_i|\mathcal{C})] - X_i(\widehat{\beta}_n - \beta_0) \\
&= [U_i - E(U_i|\mathcal{C})] + E(X_i|\mathcal{C})r(\mathcal{C}) + o_p(1) - X_i(r(\mathcal{C}) + o_p(1)) \\
&= [U_i - E(U_i|\mathcal{C})] - [X_i - E(X_i|\mathcal{C})][r(\mathcal{C}) + o_p(1)] + o_p(1) \qquad (10.16)
\end{aligned}$$

(where $o_p(1)$ does not depend on i). Using (10.16), we can write $\widehat{\Omega}_n$ as

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^{n}[U_i - E(U_i|\mathcal{C})]^2(X_i - \overline{X}_n)^2 \\
&+[r(\mathcal{C}) + o_p(1)]^2\left(n^{-1}\sum_{i=1}^{n}[X_i - E(X_i|\mathcal{C})]^2(X_i - \overline{X}_n)^2\right) \\
&-2[r(\mathcal{C}) + o_p(1)]n^{-1}\sum_{i=1}^{n}[X_i - E(X_i|\mathcal{C})][U_i - E(U_i|\mathcal{C})](X_i - \overline{X}_n)^2 \\
&+o_p(1)2n^{-1}\sum_{i=1}^{n}\left([U_i - E(U_i|\mathcal{C})] - [X_i - E(X_i|\mathcal{C})][r(\mathcal{C}) + o_p(1)]\right)(X_i - \overline{X}_n)^2 \\
&+o_p(1)n^{-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2. \qquad (10.17)
\end{aligned}
$$

The probability limit of (10.17) is $\Omega^0_{\mathcal{C}} + \eta_{\mathcal{C}}$ using Lemma 2. Hence, part (b) holds.

Part (c) follows from parts (a) and (b).

To prove part (d), note that $\widehat{\sigma}^2_n$ equals the expression in (10.17) for $\widehat{\Omega}_n$ with $X_i - \overline{X}_n$ replaced by 1. This, combined with Lemma 2, establishes part (d).

Part (e) follows from parts (a) and (d). $\square$

Proof of Theorem 13. We have

$$T_n = [\widehat{V}_n]^{-1/2}_{j,j}\sqrt{n}\left(\widehat{\beta}_{n,j} - \beta_{0,j} - r_n(\mathcal{C})_j\right) + [\widehat{V}_n]^{-1/2}_{j,j}\sqrt{n}r_n(\mathcal{C})_j, \qquad (10.18)$$

where $r_n(\mathcal{C})_j$ denotes the j-th element of $r_n(\mathcal{C})$. When Assumption CU holds, we have $r_n(\mathcal{C})_j = 0$ and $[\widehat{V}_n]^{-1/2}_{j,j}\sqrt{n}(\widehat{\beta}_{n,j} - \beta_{0,j}) \to_d N(0,1)$ by the combination of Theorems 10(a) and 11(c) and Comment 2 to Theorem 11. Hence, part (a) of the Theorem holds.

When $r(\mathcal{C})_j \neq 0$ a.s., we have

$$|T_n| = |O_p(1) + ([B^{-1}_{\mathcal{C}}(\Omega^0_{\mathcal{C}} + \eta_{\mathcal{C}})B^{-1}_{\mathcal{C}}]_{j,j} + o_p(1))^{-1/2}\sqrt{n}r_n(\mathcal{C})_j| \to_p \infty, \qquad (10.19)$$

where the equality holds using Theorems 10(a) and 11(c) and the divergence to infinity holds because $[B^{-1}_{\mathcal{C}}(\Omega^0_{\mathcal{C}} + \eta_{\mathcal{C}})B^{-1}_{\mathcal{C}}]_{j,j}$ is positive a.s. (by Assumptions 2(d) and 5(a)) and $r_n(\mathcal{C})_j \to_p r(\mathcal{C})_j$ by Theorem 10(c). In consequence, part (b) of the Theorem holds. Part (c) holds by a similar argument.

To establish part (d), under H_0, we have

$$W_n = ||(R\widehat{V}_nR')^{-1/2}n^{1/2}R(\widehat{\beta}_n - \beta_0 - r_n(\mathcal{C})) + (R\widehat{V}_nR')^{-1/2}n^{1/2}Rr_n(\mathcal{C})||^2. \qquad (10.20)$$

When Assumption CU holds, we have $Rr_n(\mathcal{C}) = 0$ a.s. and $(R\widehat{V}_nR')^{-1/2}n^{1/2}R(\widehat{\beta}_n - \beta_0) \to_d N(0, I_q)$ by Theorems 10(a) and 11(c). Hence, part (d) of the Theorem holds.

When Assumption $Rr(\mathcal{C}) \neq 0$ a.s., we have

$$W_n = ||O_p(1) + n^{1/2}(RB^{-1}_{\mathcal{C}}[\Omega^0_{\mathcal{C}} + \eta_{\mathcal{C}}]B^{-1}_{\mathcal{C}}R' + o_p(1))^{-1/2}Rr_n(\mathcal{C})||^2 \to_p \infty, \qquad (10.21)$$

where the first equality uses (10.20) and Theorems 10(a) and 11(c) and the divergence to infinity uses the fact that $B_{\mathcal{C}}^{-1}[\Omega_{\mathcal{C}}^{0}+\eta_{\mathcal{C}}]B_{\mathcal{C}}^{-1}$ is nonsingular a.s. by Assumptions 2(d) and 5(a). Hence, part (e) of the Theorem holds. □

Proof of Theorem 16. The proof is a variation of a standard argument for determining the probability limit of an extremum estimator. First, we show that

$$\sup_{\theta\in\Theta}|Q_{1,n}(\theta)-Q_{1,\mathcal{C}}(\theta)|\rightarrow_p 0. \tag{10.22}$$

This holds because (i) $Q_{1,n}(\theta)-Q_{1,\mathcal{C}}(\theta)\rightarrow_p 0$ for all $\theta\in\Theta$ by Lemma 2 and (ii) stochastic equicontinuity of $\{Q_{1,n}(\theta)-Q_{1,\mathcal{C}}(\theta):\theta\in\Theta\}$ for $n\geq 1$ holds by Lemma 1(a) of Andrews (1992) with $\widehat{Q}_n(\theta)=Q_{1,n}(\theta)-Q_{1,\mathcal{C}}(\theta)$ and $Q_n(\theta)=0$ using a mean value expansion and Assumption GMM-2(b) and (c) to verify Assumption SE-1 of Andrews (1992). Results (i) and (ii) combine to establish (10.22) by a generic uniform convergence result, e.g., see Andrews (1992, Thm. 1).

Next, we have

$$\inf_{\theta\in\Theta,\theta\notin B(\theta_1(\mathcal{C}),\varepsilon)}Q_{1,\mathcal{C}}(\theta)>Q_{1,\mathcal{C}}(\theta_0)\text{ a.s. for all }\varepsilon>0, \tag{10.23}$$

where $B(\theta_1(\mathcal{C}),\varepsilon)$ denotes an open ball of radius ε centered at $\theta_1(\mathcal{C})$. This holds because $Q_{1,\mathcal{C}}(\theta)$ is a continuous function a.s. defined on a compact set by Assumption GMM-2(a) and (b) and it has a unique minimum at $\theta_1(\mathcal{C})$ a.s. by Assumption GMM-1(a). Equation (10.23) implies that

$$\lim_{\delta\rightarrow 0}P(\inf_{\theta\in\Theta,\theta\notin B(\theta_1(\mathcal{C}),\varepsilon)}Q_{1,\mathcal{C}}(\theta)>Q_{1,\mathcal{C}}(\theta_0)+\delta)=1 \tag{10.24}$$

by the bounded convergence theorem. Hence, given any $\varepsilon,\varepsilon_1>0$, there exists a constant $\delta>0$ such that

$$P(\inf_{\theta\in\Theta,\theta\notin B(\theta_1(\mathcal{C}),\varepsilon)}Q_{1,\mathcal{C}}(\theta)\geq Q_{1,\mathcal{C}}(\theta_0)+\delta)\geq 1-\varepsilon_1. \tag{10.25}$$

Using (10.25), we have

$$\begin{aligned}
&P(\widehat{\theta}_{1,n}\notin B(\theta_1(\mathcal{C}),\varepsilon))\\
\leq\ &P(Q_{1,\mathcal{C}}(\widehat{\theta}_{1,n})-Q_{1,\mathcal{C}}(\theta_0)\geq\delta)+\varepsilon_1\\
=\ &P(Q_{1,\mathcal{C}}(\widehat{\theta}_{1,n})-Q_{1,n}(\widehat{\theta}_{1,n})+Q_{1,n}(\widehat{\theta}_{1,n})-Q_{1,\mathcal{C}}(\theta_0)\geq\delta)+\varepsilon_1\\
\leq\ &P(Q_{1,\mathcal{C}}(\widehat{\theta}_{1,n})-Q_{1,n}(\widehat{\theta}_{1,n})+Q_{1,n}(\theta_0)-Q_{1,\mathcal{C}}(\theta_0)\geq\delta)+\varepsilon_1\\
\leq\ &P(2\sup_{\theta\in\Theta}|Q_{1,n}(\theta)-Q_{1,\mathcal{C}}(\theta)|\geq\delta)+\varepsilon_1\\
\rightarrow\ &\varepsilon_1,
\end{aligned} \tag{10.26}$$

where the second inequality holds because $Q_{1,n}(\widehat{\theta}_{1,n})\leq Q_{1,n}(\theta_0)$ by the definition of $\widehat{\theta}_{1,n}$, and the convergence to ε_1 holds by (10.22). Because $\varepsilon_1>0$ is arbitrary in (10.26), the limit of $P(\widehat{\theta}_{1,n}\notin B(\theta_1(\mathcal{C}),\varepsilon))$ is zero for all $\varepsilon>0$ and the result of the theorem for $\widehat{\theta}_{1,n}$ is proved.

The corresponding result for $\widehat{\theta}_{2,n}$ holds by an analogous argument provided $\widehat{\Sigma}_n \rightarrow_p \Sigma_{\mathcal{C}}$. The latter holds using (i) Lemma 2 with $h(W_i) = g(Z_i, \theta_0)g(Z_i, \theta_0)'$, (ii) mean value expansions of $g(Z_i, \widehat{\theta}_{1,n})$ around θ_0 using Assumptions GMM-2(b) and (c), and (iii) $\widehat{\theta}_{1,n} \rightarrow_p \theta_1(\mathcal{C})$. $\square$

Proof of Theorem 17. By Assumption GMM-CON(a), $Q_{j,\mathcal{C}}(\theta_0) = 0$ for $j = 1, 2$. By Assumption GMM-CON(b) and Assumption GMM-2(d), $Q_{j,\mathcal{C}}(\theta) > 0$ for all $\theta \in \Theta$ with $\theta \neq \theta_0$ a.s. for $j = 1, 2$. Thus, Assumption GMM-1 holds with $\theta_1(\mathcal{C}) = \theta_0$ and $\theta_2(\mathcal{C}) = \theta_0$ a.s. Theorem 16 now gives $\widehat{\theta}_{1,n} \rightarrow_p \theta_0$ and $\widehat{\theta}_{2,n} \rightarrow_p \theta_0$. $\square$

Footnotes

[1] This paper has been prepared for the Ted Hannan Lecture at the Australasian Meetings of the Econometric Society to be held in Sydney, Australia in July 2003. The author thanks the organizers of these meetings for their work. The author thanks Peter Phillips for remarks and comments made over the years on the general topic considered in this paper. He also thanks Joe Altonji and Guido Imbens for references and gratefully acknowledges the research support of the National Science Foundation via grant number SES-0001706.

[2] This approach allows for multinomial sampling, which is a type of stratified sampling. Extensions to clustered sampling are also possible.

[3] This holds because, for any integrable random variables $\{\xi_n : n \geq 1\}$ and ξ, we have (i) $|E\xi_n - E\xi| \leq E|\xi_n - \xi|$ and (ii) $E|\xi_n - \xi| \to 0$ iff $\xi_n - \xi \to_p 0$ and $\{\xi_n : n \geq 1\}$ is uniformly integrable, e.g., see Dudley (1989, Thm. 10.3.6, p. 279).

[4] Note that some authors refer to the following type of structure as an *approximate* factor structure because it allows for a purely idiosyncratic component as well as common factors.

[5] We take $S(\gamma) = (C_1, C_2)$ for convenience because it guarantees that $\mathcal{C} = \sigma(C_1, C_2)$. Instead, if we took $S(\gamma) = 0$, then $\mathcal{C}$ would not necessarily equal $\sigma(C_1, C_2)$ without some additional assumptions, which would just clutter the presentation. For example, if some element of $U^*(\gamma)$ equaled zero for all $\gamma \in \Gamma$, then the corresponding element of C_1 would not appear in $U(\gamma)$ and, hence, would not be part of the random variables that generate $\mathcal{C}$.

[6] The common shocks $(C_1(\cdot), C_2(\cdot))$ are included as part of $S(\gamma)$ for the same reason as in footnote 5.

References

Andrews, D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory,* 8, 241-257.

Anselin, L. (1988): *Spatial Econometric Methods and Models.* Boston: Kluwer.

Bai, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica,* 71, 135-171.

Bai, J. and S. Ng (2001): "A PANIC Attack on Unit Roots and Cointegration," unpublished manuscript, Department of Economics, Boston College.

——— (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica,* 70, 191-221.

Case, A. C. (1991): "Spatial Patterns in Household Demand," *Econometrica,* 59, 953-965.

Chamberlain, G. and M. Rothschild (1983): "Arbitrage, Factor Structure, and Mean-variance Analysis on Large Asset Markets," *Econometrica,* 51, 1281-1304.

Chen, X. and T. Conley (2001): "A New Semiparametric Spatial Model for Panel Time Series," *Journal of Econometrics,* 105, 59-84.

Clark, P. K. (1973): "Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica,* 41, 135-155.

Conley, T. G. (1999): "GMM Estimation with Cross-sectional Dependence," *Journal of Econometrics,* 92, 1-45.

Donald, S. G. and K. Lang (2001): "Inference with Difference in Differences and Other Panel Data," unpublished manuscript, Department of Economics, University of Texas at Austin.

Dudley, R. M. (1989): *Real Analysis and Probability.* Pacific Grove, CA: Wadsworth and Brooks/Cole.

Feller, W. (1966): *An Introduction to Probability Theory and Its Applications,* Vol. II, 2nd Ed. New York: Wiley.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000): "The Generalized Dynamic Factor Model: Identification and Estimation," *Review of Economics and Statistics,* 82, 540-554.

Forni, M. and M. Lippi (2001): "The Generalized Dynamic Factor Model: Representation Theory," *Econometric Theory,* 17, 1113-1141.

Geweke, J. (1977): "The Dynamic Factor Analysis of Economic Time-series Models," in D. J. Aigner and A. S. Goldberger eds., *Latent Variables in Socio-economic Models.* Amsterdam: North-Holland.

Greenwald, B. C. (1983): "A General Analysis of Bias in the Estimated Standard Errors of Least Squares Coefficients," *Journal of Econometrics,* 22, 323-338.

Hall, P. and C. C. Heyde (1980): *Martingale Limit Theory and Its Applications.* New York: Academic Press.

Imbens, G. and T. Lancaster (1996): "Efficient Estimation and Stratified Sampling," *Journal of Econometrics,* 74, 289-318.

Kelejian, H. H. and I. R. Prucha (1999): "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," *International Economic Review,* 40, 509-533.

Kloek, T. (1981): "OLS Estimation in a Model Where a Microvariable Is Explained by Aggregates and Contemporaneous Disturbances Are Equicorrelated," *Econometrica,* 49, 205-207.

Mandelbrot, B. and H. Taylor (1967): "On the Distribution of Stock Price Differences," *Operations Research,* 15, 1057-1062.

Moon, H. R. and B. Perron (2002): "Testing for a Unit Root in Panels with Dynamic Factors," unpublished working paper, Department of Economics, University of Southern California.

Moulton, B. R. (1986): "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics,* 32, 385-397.

——— (1987): "Diagnostics for Group Effects in Regression Analysis," *Journal of Business and Economic Statistics,* 5, 275-282.

——— (1990): "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *Review of Economics and Statistics,* 72, 334-338.

Moulton, B. R. and W. C. Randolf (1989): "Alternative Tests of the Error Components Model," *Econometrica,* 57, 685-693.

Pepper, J. V. (2002): "Robust Inferences from Random Clustered Samples: An Application Using Data from the Panel Study of Income Dynamics," *Economics Letters,* 75, 341-345.

Pesaran, H. M. (2002): "Estimation and Inference in Large Heterogeneous Panels with Cross Section Dependence," unpublished manuscript, Trinity College, Cambridge.

Pfeffermann, D. and T. M. F. Smith (1985): "Regression Models for Grouped Populations in Cross-Section Surveys," *International Statistical Review,* 53, 37-59.

Phillips, P. C. B. and D. Sul (2002): “Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence,” Cowles Foundation Discussion Paper No. 1362, Yale University.

Sargent, T. J. and C. A. Sims (1977): “Business Cycle Modeling Without Pretending to Have Too Much *A Priori* Economic Theory,” in C. A. Sims ed., *New Methods in Business Cycle Research*: *Proceedings from a Conference*. Minneapolis: Federal Reserve Bank of Minneapolis.

Scott, A. J. and D. Holt (1982): “The Effect of Two-Stage Sampling on Ordinary Least Squares Methods,” *Journal of the American Statistical Society,* 77, 848-854.

Stock, J. H. and M. W. Watson (2003): “Macroeconomic Forecasting Using Diffusion Indexes,” *Journal of Business and Economic Statistics,* 21, forthcoming.