CROSS-SECTION REGRESSION WITH COMMON STOCKS

BY

DONALD W. K. ANDREWS

COWLES FOUNDATION PAPER NO. 1153

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2006

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

Econometrica, Vol. 73, No. 5 (September, 2005), 1551–1585

CROSS-SECTION REGRESSION WITH COMMON SHOCKS

BY DONALD W. K. ANDREWS[1]

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) 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 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 recentering and scaling. The 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. However, the absolute values of t, 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 and regressors are considered. Using the general results, conditions are determined under which consistency of the LS 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.

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

1. INTRODUCTION

THE REGRESSION MODEL ESTIMATED BY LEAST SQUARES (LS) is the workhorse of econometrics. The properties of LS 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 on the properties of LS 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

[1]This paper was prepared for the Ted Hannan Lecture at the Australasian Meetings of the Econometric Society 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 the co-editor and two anonymous referees for helpful comments, and Joe Altonji and Guido Imbens for references. The author gratefully acknowledges the research support of the National Science Foundation via Grant SES-0001706.

below, but this research focuses on shocks that are predominantly local in nature. Common shocks need not be of this form.

By common shocks, we mean macroeconomic, technological, legal/institutional, political, environmental, health, and sociological shocks. It seems apparent that common shocks are a likely feature of cross-section economic data (see Andrews (2003) for further discussion). This is true whether the population units in the cross-section regression are individuals, households, firms, industries, plants, cities, states, countries, or products.

The *impact* of common 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 analysis 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 this paper, we analyze the effects of common shocks on the properties of LS estimators and related test statistics. We start by assuming that there is a common shock σ-field, $\mathcal{C}$, such that the observations are i.i.d. conditional on $\mathcal{C}$. This is referred to as Assumption 1. For example, if there is a vector C of common shocks, then $\mathcal{C}$ is the σ-field generated by C. Typical factor models are of this type. As discussed below, Assumption 1 is shown to be surprisingly general. At the same time, it yields asymptotic results that are remarkably simple.

Using Assumption 1, 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 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 and regressors. 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, 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.

To justify the basic assumption, viz. Assumption 1, 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. We allow for general patterns of cross-section dependence and heterogeneity across the population units. Then we consider i.i.d. sampling from the population with the randomness in the sampling defined on the same probability space.[2] This sampling scheme leads to exchangeable observations, which have the property of being i.i.d. conditional on some σ-field $\mathcal{C}$ by de Finetti's theorem. That is, it leads to Assumption 1 holding. The framework is similar to that used by Conley (1999), but does not impose a strong mixing assumption.

The asymptotic results are obtained by exploiting the exchangeability of the observations, which results from i.i.d. 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 and clustered sampling. Extensions to instrumental variables estimators and generalized method of moments (GMM) estimators of moment condition models are discussed in Andrews (2003).

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):

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

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 (2002, 2004), Pesaran (2002), Stock and Watson (2002), Phillips and Sul (2003, 2004), Bai (2003), and Moon and Perron (2004). 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. Section 2 specifies the regression model 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 shows that the basic assumption that the observations are iid conditional on a common σ-field allows for very general dependence and heterogeneity across population units. Section 8 discusses extensions to panel models with a fixed time dimension T

and clustered sampling. 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 observations for sample size n are $\{(Y_i, X_i) : i = 1, \dots, n\}$. The model is

$$Y_i = \alpha_0 + X_i'\beta_0 + U_i \quad \text{for} \quad i = 1, \dots, n, \tag{2.1}$$

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. In some cases, an additional variable, S_i, or some component of S_i may be observed for $i = 1, \dots, n$. We suppose the random variables $\{W_i : i = 1, 2, \dots\}$, where $W_i = (Y_i, X_i, S_i)$, are defined on a probability space $(\Omega, \mathcal{B}, P)$.

We suppose that common shocks across observations are captured by a σ-field $\mathcal{C} \subset \mathcal{B}$. For example, if common shocks arise in the form of a vector of random variables C. Then, $\mathcal{C}$ is the σ-field generated by C. More generally, $\mathcal{C}$ could be the σ-field generated by an infinite-dimensional vector C or the σ-field generated by a random function $C(\cdot)$. Examples are provided below.

The main assumption we employ is the following.

ASSUMPTION 1: *There exists a σ-field $\mathcal{C} \subset \mathcal{B}$ such that, conditional on $\mathcal{C}$, $\{W_i : i = 1, 2, \dots\}$ are i.i.d.*

As shown in Section 7, Assumption 1 is surprisingly general. When observational units are drawn randomly from the population, Assumption 1 is compatible with arbitrary dependence between underlying units in the population. Assumption 1 is compatible with common shocks that have different effects on different population units. It is compatible with arbitrary forms of heterogeneity (i.e., nonidentical distributions) across population units.

A simple example of a regression model with a common error shock is the model in (2.1) with $U_i = C + \varepsilon_i$, where (X_i, ε_i) are i.i.d. across i and C is a random variable that is common for all observations i.

A second example is a model with a factor structure in the errors and regressors. In this case, the following assumption holds.

ASSUMPTION SF1: *The errors and regressors satisfy*

$$U_i = C_1'U_i^* \quad \text{and} \quad X_i = C_2 X_i^*, \tag{2.2}$$

where (a) *C_1 and U_i^* are random d_1 vectors, X_i^* is a random d_2 vector, and C_2 is a random $k \times d_2$ matrix for $d_2 \geq k$*; (b) *$\{(U_i^*, X_i^*) : i \leq n\}$ and (C_1, C_2) are mutually independent*; *and* (c) *(U_i^*, X_i^*) are independent across $i \leq n$.*

In this *standard factor* scenario, (C_1, C_2) are random common factors, (U_i^*, X_i^*) are random factor loadings, and $\mathcal{C} = \sigma(C_1, C_2)$.

The σ-field $\mathcal{C}$ consists of the *common* shocks to the random elements $\{W_i : i = 1, 2, \ldots\}$. 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 1 if the ith observation is in group g and equals 0 otherwise for $g = 1, \ldots, g_{\max}$. Let $C_1, \ldots, C_{g_{\max}}$ 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 $(C_1, \ldots, C_{g_{\max}})$ through the vector $S_i = (C_1 S_{1i}, C_2 S_{2i}, \ldots, C_{g_{\max}} S_{gi})'$. Thus, only observations in group g are affected by the gth 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 $(C_1, \ldots, C_{g_{\max}})$ are assumed to be independent, but in the present paper, there is no need to make this assumption. In fact, $(C_1, \ldots, C_{g_{\max}})$ could just denote the differential impacts of a single common shock on g different groups and, in this case, correlation between the elements of $(C_1, \ldots, C_{g_{\max}})$ 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_{1i}, 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 ith observation being through $C(S_{1i})$. Thus, the impact of the common shock varies continuously across i depending on the value of S_{1i}.

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.

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 observations are independent across i. Thus, the following assumptions are standard (STD) for cross-section applications.

ASSUMPTION STD1: $E(1, X_i')U_i = 0$ *for all* $i \leq n$, ASSUMPTION STD2: *$\{W_i : i \leq n\}$ are independent across $i \leq n$.*

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

3. PROBABILITY LIMIT OF THE LS ESTIMATOR

3.1. *Main Results*

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

$$\begin{aligned}(3.1)\qquad \widehat{\beta}_n &= \beta_0 + \left(n^{-1}\sum_{i=1}^{n} X_i X_i' - \overline{X}_n \overline{X}_n'\right)^{-1} \\ &\quad \times \left(n^{-1}\sum_{i=1}^{n} X_i U_i - \overline{X}_n \overline{U}_n\right), \quad \text{where}\end{aligned}$$

$$\overline{X}_n = n^{-1}\sum_{i=1}^{n} X_i \quad \text{and} \quad \overline{U}_n = n^{-1}\sum_{i=1}^{n} U_i.$$

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

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

Under Assumption 1, the random variables $\{W_i : i \geq 1\}$ are exchangeable. (That is, $(W_{\pi(1)}, \ldots, W_{\pi(n)})$ has the same distribution as $(W_1, \ldots, W_n)$ for every permutation π of $(1, \ldots, n)$ for all $n \geq 2$.) 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 1: *Suppose Assumption* 1 *holds. Let $h(\cdot)$ be a vector-valued function that satisfies $E\|h(W_i)\| < \infty$. Then*

$$n^{-1}\sum_{i=1}^{n} h(W_i) \xrightarrow{p} E\big(h(W_i)|\mathcal{C}\big) \quad as \quad n \to \infty,$$

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

COMMENT: The convergence in Lemma 1 also holds almost surely (a.s.).

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) $E\|X_i\|^2 < \infty$.
(b) $E|U_i| < \infty$.
(c) $E\|X_i U_i\| < \infty$.
(d) $E(X_i X_i'|\mathcal{C}) - E(X_i|\mathcal{C})E(X_i|\mathcal{C})' > 0$ *a.s.*

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

$$r(\mathcal{C}) = \big(E(X_i X_i'|\mathcal{C}) - E(X_i|\mathcal{C})E(X_i|\mathcal{C})'\big)^{-1} \times \big(E(X_i U_i|\mathcal{C}) - E(X_i|\mathcal{C})E(U_i|\mathcal{C})\big). \tag{3.3}$$

Note that the term $E(X_i U_i|\mathcal{C}) - E(X_i|\mathcal{C})E(U_i|\mathcal{C})$ in (3.3) 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

$$\min_{\beta \in \mathbb{R}^k} E(U_i - X_i'\beta|\mathcal{C})' E(U_i - X_i'\beta|\mathcal{C}). \tag{3.4}$$

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

$$s(\mathcal{C}) = E(U_i|\mathcal{C}) - E(X_i|\mathcal{C})' r(\mathcal{C}). \tag{3.5}$$

Using (3.1), (3.2), and Lemma 1, the probability limits of $\widehat{\beta}_n$ and $\widehat{\alpha}_n$ are easily obtained:

THEOREM 1: *Suppose Assumptions* 1 *and* 2 *hold. Then*

$$\widehat{\beta}_n \xrightarrow{p} \beta_0 + r(\mathcal{C}) \quad \textit{and} \quad \widehat{\alpha}_n \xrightarrow{p} \alpha_0 + s(\mathcal{C}).$$

COMMENTS: 1. The convergence in Theorem 1 holds jointly and almost surely.

2. Theorem 1 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 nonzero constant. Similarly, $\widehat{\alpha}_n$ equals α_0 plus a term, $s(\mathcal{C})$, that may be zero, random, or a nonzero 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, 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.

For any random vectors A and B and any random vector or σ-field D, let $\text{Cov}(A, B|D)$ 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: $\text{Cov}(X_i, U_i|\mathcal{C}) = 0$ *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 1, we have the following necessary and sufficient condition for consistency of $\widehat{\beta}_n$ and $\widehat{\alpha}_n$.

COROLLARY 1: *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 nonzero 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 instrumental variables are available that are uncorrelated with U_i conditional on $\mathcal{C}$; see Andrews (2003, Sec. 7).

[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) $E|\xi_n - \xi| \to 0$ if and only if $\xi_n - \xi \to_p 0$ and $\{\xi_n : n \geq 1\}$ is uniformly integrable, e.g., see Dudley (1989, Theorem 10.3.6, p. 279).

3.2. *Standard Factor Structure*

Corollary 1 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 Assumptions CU and CMZ) in terms of the *standard factor* structure for the regressors and errors (see Assumption SF1).[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.)

Given Assumption SF1, Assumption 2(d) holds provided

$$EX_i^* X_i^{*\prime} - EX_i^* EX_i^{*\prime} > 0 \tag{3.6}$$

and C_2 has full row rank d_2 a.s.

For any random vectors A and B, let $\mathrm{Cov}(A, B)$ denote the covariance between A and B.

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

ASSUMPTION SF2: $\mathrm{Cov}(X_i^*, U_i^*) = 0$.

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 1, 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

$$\begin{aligned} E(U_i|\mathcal{C}) &= E(C_1' U_i^*|\mathcal{C}) = C_1' E(U_i^*|\mathcal{C}) = C_1' EU_i^*, \\ E(X_i|\mathcal{C}) &= E(C_2 X_i^*|\mathcal{C}) = C_2 E(X_i^*|\mathcal{C}) = C_2 EX_i^*, \\ E(X_i U_i|\mathcal{C}) &= E(C_2 X_i^* U_i^{*\prime} C_1|\mathcal{C}) \\ &= C_2 E(X_i^* U_i^{*\prime}|\mathcal{C}) C_1 = C_2 E(X_i^* U_i^{*\prime}) C_1, \end{aligned} \tag{3.7}$$

[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.

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 independent of $\mathcal{C} = \sigma(C_1, C_2)$.

Combining the results in (3.7) gives

$$
\begin{aligned}
(3.8)\qquad E(X_iU_i|\mathcal{C}) - E(X_i|\mathcal{C})E(U_i|\mathcal{C}) &= C_2E(X_i^*U_i^{*\prime})C_1 - C_2EX_i^*EU_i^{*\prime}C_1 \\
&= C_2(EX_i^*U_i^{*\prime} - EX_i^*EU_i^{*\prime})C_1 \\
&= 0,
\end{aligned}
$$

where the last equality holds by Assumption SF2.

Assumptions 1, SF1, and SF3 imply Assumption CMZ by the first line of (3.7).

The following corollary is a special case of a more general result (viz. Theorem 3) given below. The first two parts of the corollary are the results proved above in (3.7) and (3.8). (The proof is given above because it is instructive.)

COROLLARY 2: (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* $s(\mathcal{C}) = 0$.

COMMENTS: 1. Theorem 1 and Corollary 2 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

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

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.

Instead of Assumptions SF2 and SF3, consider the following assumption:

ASSUMPTION SF4: (a) $\mathrm{Cov}(X_i^*, U_{1i}^*) = 0$ *and* $EU_{1i}^* = 0$, *where* $U_i^* = (U_{1i}^*, U_{2i}^*)' \in R^2$.

(b) $C_1 = (1, C_{11})' \in R^2$.

(c) $EC_{11} = 0$ *and* $EC_2C_{11} = 0$.
(d) $EX_i^* U_{2i}^* \neq 0$, $EU_{2i}^* = 0$, $C_{11} \neq 0$ *with probability* 1, *and* C_2 *is full row rank with probability* 1.

Under Assumption SF4,

$$U_i = U_{1i}^* + C_{11} U_{2i}^*, \tag{3.10}$$

where U_{1i}^* has mean zero and is uncorrelated with the idiosyncratic component of the regressor X_{2i}^*; the error factor C_{11} has mean zero and is uncorrelated with the regressor factors C_2; and U_{2i}^* has mean zero but is correlated with the idiosyncratic component of the regressor X_{2i}^*.

THEOREM 2: *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

$$\begin{aligned} r(\mathcal{C}) &= (C_2 E^*[X_i^* - E^* X_i^*][X_i^* - E^* X_i^*]' C_2')^{-1} \\ &\quad \times (C_2 E^*[X_i^* - E^* X_i^*] U_{2i}^* C_{11}), \\ s(\mathcal{C}) &= E^* U_{2i}^* C_{11} + (E^* X_i^*)' C_2' r(\mathcal{C}), \end{aligned} \tag{3.11}$$

where E^* denotes expectation with respect to (X_i^*, U_i^*) alone.
2. Theorems 1 and 2 combine to show that $\widehat{\beta}_n$ is *not* consistent for β_0 under Assumptions 1, 2, SF1, and SF4 in spite of the fact that Assumption STD1 holds.
3. In the proof of Theorem 2, 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 ith observation are of the form $(C_1(S_{0,i}), C_2(S_{0,i}))$, where $S_{0,i}$ is a vector of characteristics of the ith observation. Hence, the common shocks take the form of stochastic functions $(C_1(\cdot), C_2(\cdot))$. The random element $S_{0,i}$ may or may not be observed.
For a random element ξ, let $\mathrm{supp}(\xi)$ denote the support of ξ.
The heterogeneous factor structure is specified in the following assumptions:

ASSUMPTION HF1: *For all* $i \geq 1$,

$$U_i = C_1(S_{0,i})' U_i^*,$$
$$X_i = C_2(S_{0,i}) X_i^*,$$
$$S_i = \big(S_{0,i}, C_1(\cdot), C_2(\cdot)\big),$$

where (a) U_i^* *is a random* d_1 *vector*, $C_1(\cdot)$ *is a random* d_1 *vector-valued function with domain* $\text{supp}(S_{0,i})$, X_i^* *is a random* d_2 *vector, and* $C_2(\cdot)$ *is a random* $k \times d_2$ *matrix-valued function with domain* $\text{supp}(S_{0,i})$ *for* $d_2 \geq k$; (b) $\{(U_i^*, X_i^*, S_{0,i}) : i \geq 1\}$ *and* $(C_1(\cdot), C_2(\cdot))$ *are mutually independent*; *and* (c) $(U_i^*, X_i^*, S_{0,i})$ *are independent across* $i \geq 1$.

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) $\text{Cov}(X_i^*, U_i^* | S_{0,i}) = 0$ *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.

The common σ-field $\mathcal{C}$ is the σ-field generated by the common shocks:

$$\mathcal{C} = \sigma\big(C_1(\cdot), C_2(\cdot)\big). \tag{3.12}$$

The following result, like Corollary 2, is a special case of Theorem 3.

COROLLARY 3: (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* $s(\mathcal{C}) = 0$.

COMMENT: Corollary 3 and Theorem 1 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 an observation to depend on the characteristics of the observation via a component $S_{0,i}$ of S_i. The common shocks are characterized by a function $C(\cdot)$. In particular, the effects of the common shocks on observation i are through $C(S_{0,i})$. The errors and regressors are determined by stochastic processes $U_i(\cdot)$ and $X_i(\cdot)$ that are uncorrelated conditional on $S_{0,i}$ for each i. Specifically, we have the following assumption:

ASSUMPTION FF1: (a) *$S_{0,i}$ is a component of S_i.*

(b) *$C(\cdot)$ is a random function that does not depend on i, has domain* $\text{supp}(S_{0,i})$, *and is a component of S_i for all i.*

(c) *For all i, $U_i(\cdot)$ and $X_i(\cdot)$ are random functions with ranges $\mathbb{R}$ and $\mathbb{R}^k$, respectively, and domain* $\text{supp}(C(S_{0,i}))$.

(d) *For all i, $U_i = U_i(C(S_{0,i}))$ and $X_i = X_i(C(S_{0,i}))$.*

(e) *$\{(U_i(\cdot), X_i(\cdot), S_{0,i}) : i \geq 1\}$ and $C(\cdot)$ are mutually independent.*

(f) *$(U_i(\cdot), X_i(\cdot), S_{0,i})$ are independent across $i \geq 1$.*

Assumption FF1 allows the whole distributions of U_i and X_i to vary with $C(S_{0,i})$. In contrast, with standard or heterogeneous factor structures, $C(S_{0,i})$ only affects the multivariate location and scale of the errors and regressors.

Let $\text{supp}(C)$ denote $\text{supp}(C(S_{0,i}))$.

ASSUMPTION FF2: (a) *For all $c \in \text{supp}(C)$, $\text{Cov}(X_i(c), U_i(c)|S_{0,i}) = 0$ a.s.*

(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 $c \in \text{supp}(C)$ a.s.*

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

ASSUMPTION FF3: *For all $c \in \text{supp}(C)$, $E(U_i(c)|S_{0,i}) = 0$ a.s.*

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

THEOREM 3: (a) *Suppose Assumptions* 1, FF1, *and* FF2 *hold. Then Assumption* CU *holds and $r(\mathcal{C}) = 0$.*

(b) *Suppose Assumptions* 1 *and* FF1–FF3 *hold. Then Assumptions* CU *and* CMZ *hold, $r(\mathcal{C}) = 0$, and $s(\mathcal{C}) = 0$.*

COMMENTS: 1. Assumptions SF1 and SF2 imply Assumptions FF1 and FF2 with $S_{0,i} = 0$, $C(\cdot) = (C_1, C_2)$, $U_i(c) = c_1'U_i^*$, and $X_i(c) = c_2X_i^*$, 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 $C(\cdot) = (C_1(\cdot), C_2(\cdot))$, $U_i(c) = c_1' U_i^*$, and $X_i(c) = c_2 X_i^*$, where $c = (c_1, c_2)$. Analogously, Assumptions HF1–HF3 imply Assumptions FF1–FF3.

3.5. *Probability Limit of Parameter Subvectors*

Theorem 1 provides necessary and sufficient conditions for the consistency of the LS estimator of the whole parameter β and of subvectors of β. For example, the first element of $\widehat{\beta}$ is consistent if and only if the first element of $r(\mathcal{C})$ equals zero. Alternatively, one can use the partitioned regression formula for the LS estimator to obtain equivalent conditions for consistency of subvectors of β. Suppose

$$X_i = (X_{1i}', X_{2i}')' \quad \text{and} \quad \beta = (\beta_1', \beta_2')', \tag{3.13}$$

where $X_{1i}, \beta_1 \in \mathbb{R}^{k_1}$. Let $A_i = (1, X_{2i}')'$. Then, using Lemma 1, some calculations show that under Assumptions 1 and 2, $\widehat{\beta}_1$ is consistent if and only if

$$E(X_{1i}U_i|\mathcal{C}) - E(X_{1i}A_i|\mathcal{C})(E(A_iA_i|\mathcal{C}))^{-1}E(A_iU_i|\mathcal{C}) = 0 \quad \text{a.s.} \tag{3.14}$$

For example, suppose the model is as in (2.1) with $U_i = A_i'C + \varepsilon_i$, where $\{(X_i, A_i) : i \geq 1\}$ and $\{\varepsilon_i : i \geq 1\}$ are independent and i.i.d. with $E\varepsilon_i = 0$ and C is a common random vector. Then (3.14) holds and $\widehat{\beta}_1$ is consistent even though the error contains a common shock whose impact depends on the regressors X_{2i}.

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) $EU_i^2 < \infty$.
(b) $E\|X_iU_i\|^2 < \infty$.

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

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

Note that $r_n(\mathcal{C})$ converges in probability to $r(\mathcal{C})$ as $n \to \infty$ under Assumptions 1 and 2 by Lemma 1. 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

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

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

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

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 $n^{1/2}(\text{plim}_{n\to\infty}\overline{X}_n)\overline{U}_n$, 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

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

In particular, under Assumptions SF1 and SF2, we have

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

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

Under Assumptions HF1 and HF2, we have

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

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

Next, define

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

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

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

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

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

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 4: *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.*;

(c) $r_n(\mathcal{C}) \to_p r(\mathcal{C})$.

COMMENTS: 1. Part (a) of Theorem 4 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, Theorem 3.2, p. 58). Part (c) is established using Lemma 1.

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

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

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

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

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

ASSUMPTION 4: (a) $E\|X_i\|^4 < \infty$.
(b) $E\|X_i\|^3|U_i| < \infty$.

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

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

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 random variable

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

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 5: *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) $\widehat{\sigma}^2_n \to_p \sigma^2_{\mathcal{C}} + \tau_{\mathcal{C}}$;
(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 Theorem 5 because the residuals, $\{\widehat{U}_i : i=1,\dots,n\}$, are not consistent estimators of the errors, $\{U_i : i=1,\dots,n\}$, 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 4 and 5 combine to give the following results for the LS estimator of β_0 normalized by an estimated covariance matrix:

COROLLARY 4: *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) $\widehat{V}^{-1/2}_{\sigma,n} n^{1/2}(\widehat{\beta}_n - \beta_0 - r_n(\mathcal{C})) \to_d (\sigma^2_{\mathcal{C}} + \tau_{\mathcal{C}})^{-1/2}B^{1/2}_{\mathcal{C}} \times V^{1/2}_{\mathcal{C}} \times Z$;
(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 4 and 5. 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 $\beta = (\beta_1, \dots, \beta_k)'$ and $\beta_0 = (\beta_{0,1}, \dots, \beta_{0,k})'$. The t statistic for testing H_0 against H_1 is

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

where $\widehat{\beta}_n = (\widehat{\beta}_{n,1}, \dots, \widehat{\beta}_{n,k})'$ and $[D]_{j,j}$ denotes the jth 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 (nonstochastic) full row rank $q \times k$ matrix and a is a (nonstochastic) q vector. Define the Wald test statistic W_n as

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

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 also can 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 jth element of $r(\mathcal{C})$.

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

THEOREM 6: *Suppose Assumptions* 1–5 *hold. Let R be a full row rank $q \times k$ matrix. Then, under H_0*:

(a) $P(|T_n| > z_{1-\alpha/2}) \to \alpha$ *and* $P(T_n > z_{1-\alpha}) \to \alpha$ *when Assumption* CU *holds*;
(b) $P(|T_n| > z_{1-\alpha/2}) \to 1$ *when* $r(\mathcal{C})_j \neq 0$ *a.s.*;
(c) $P(T_n > z_{1-\alpha}) \to 1$ *when* $r(\mathcal{C})_j > 0$ *a.s.*;
(d) $P(W_n > \chi^2_{q,1-\alpha}) \to \alpha$ *when Assumption* CU *holds*;
(e) $P(W_n > \chi^2_{q,1-\alpha}) \to 1$ *when* $Rr(\mathcal{C}) \neq 0$ *a.s.*

COMMENTS: 1. The results of Theorem 6 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 Theorem 6 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 1 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 $\|\sqrt{n}Rr_n(\mathcal{C})\|^2 \to_p \infty$ in parts (b), (c), and (e), respectively. In this case, the probability of overrejection increases as the sample size increases.

3. As stated, Theorem 6 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 4 and 5. 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 1 under the null hypothesis. However, $\widehat{V}_n$ 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

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

By a standard and simple argument, the behavior of $\mathrm{CI}_{\beta_{0,j}}$ is determined by the behavior under the null hypothesis of the t statistic T_n. In consequence, the results of Theorem 6 imply that under Assumptions 1–5 and Assumption CU, the coverage probability of $\mathrm{CI}_{\beta_{0,j}}$ 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 $\mathrm{CI}_{\beta_{0,j}}$ converges to zero as $n \to \infty$, which is not desired.

7. SAMPLING SCHEME

In this section, we show that Assumption 1 holds in the presence of arbitrary forms of dependence and heterogeneity between population units provided observations in the sample are obtained from the population via a random sampling scheme.

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

$\gamma \in \Gamma$, $Y(\gamma) \in \mathbb{R}$ denotes the regression dependent variable, $X(\gamma) \in \mathbb{R}^k$ 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

$$W(\gamma) = \big(Y(\gamma), X(\gamma), S(\gamma)\big). \tag{7.1}$$

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 $(\mathbb{R}, \mathbb{R}^k, \mathcal{S})$).

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

$$Y(\gamma) = \alpha_0 + X(\gamma)'\beta_0 + U(\gamma), \tag{7.2}$$

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.

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., nonidentical distributions) of $W(\gamma)$ across $\gamma \in \Gamma$ also are allowed.

Samples of size n for $n \geq 1$ are obtained by drawing indices $\{\gamma_i : \geq 1\}$ 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 make the following assumption:

ASSUMPTION S: *The indices* $\{\gamma_i : i \geq 1\}$ *are i.i.d. indices, independent of* $\{W(\gamma) : \gamma \in \Gamma\}$, *each with some distribution* G.

Assumption S allows for probabilistic oversampling 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. Oversampling of some units is obtained by taking G to be some nonuniform 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

$$\begin{aligned} &W_i = W(\gamma_i), \quad Y_i = Y(\gamma_i), \quad X_i = X(\gamma_i), \\ &S_i = S(\gamma_i), \quad \text{and} \quad U_i = U(\gamma_i) \end{aligned} \tag{7.3}$$

for $i = 1, 2, \ldots$. (The random vector $W(\gamma_i)$ is assumed to be a measurable function on $(\Omega, \mathcal{B}, P)$ with respect to the product Borel σ-field on $(\mathbb{R} \times \mathbb{R}^k \times \mathcal{S})$.) 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, Chap. 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 $\{(Y_i, X_i) : i = 1, \ldots, n\}$. In addition, depending upon the context, S_i or some component of S_i may be observed for $i = 1, \ldots, n$. In terms of the sample of the first n observations, the model is as defined in (2.1).

The sampling scheme given in Assumption S leads to exchangeable observations $\{W_i : i = 1, 2, \ldots\}$. Hence, de Finetti's theorem applies (e.g., see Hall and Heyde (1980, (7.1), p. 202)), which means that Assumption S implies Assumption 1.

LEMMA 2: *Suppose Assumption* S *holds. Then* $\{W_i : i = 1, 2, \ldots\}$ *are exchangeable random elements and Assumption* 1 *holds*.

COMMENTS: 1. Under Assumption S, the σ-field $\mathcal{C}$ in Assumption 1 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 $\{W_i : i = 1, 2, \ldots\}$ and are invariant to permutations of the first n random variables $\{W_i : i = 1, 2, \ldots, n\}$); see Hall and Heyde (1980, p. 202).

2. The σ-field $\mathcal{C}$ consists of the *common* shocks to the random elements $\{W_i : i = 1, 2, \ldots\}$.

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 $E\|h(W_i)\| < \infty$,

$$Eh(W_i) = E_{\gamma_i} E\big(h(W(\gamma_i))|\gamma_i\big) = \int Eh(W(\gamma))\,dG(\gamma), \tag{7.4}$$

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

The random variable W_i that appears in the limit in Lemma 1 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 Lemma 1 can be written as

$$E\big(h(W_i)|\mathcal{C}\big) = E_{\gamma_i} E\big(h(W(\gamma_i))|\mathcal{C}, \gamma_i\big) = \int Eh\big(W(\gamma)|\mathcal{C}\big)\,dG(\gamma), \tag{7.5}$$

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

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

$$\begin{aligned}(7.6)\qquad &\int E\big(X(\gamma)X(\gamma)'|\mathcal{C}\big)\,dG(\gamma)\\ &\quad -\int E\big(X(\gamma)|\mathcal{C}\big)\,dG(\gamma)\int E\big(X(\gamma)'|\mathcal{C}\big)\,dG(\gamma) > 0 \quad \text{a.s.}\end{aligned}$$

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

$$\begin{aligned}(7.7)\qquad r(\mathcal{C}) = &\bigg(\int E\big(X(\gamma)X(\gamma)'|\mathcal{C}\big)\,dG(\gamma)\\ &\quad -\int E\big(X(\gamma)|\mathcal{C}\big)\,dG(\gamma)\int E\big(X(\gamma)'|\mathcal{C}\big)\,dG(\gamma)\bigg)^{-1}\\ &\times\bigg(\int E\big(X(\gamma)U(\gamma)|\mathcal{C}\big)\,dG(\gamma)\\ &\quad -\int E\big(X(\gamma)|\mathcal{C}\big)\,dG(\gamma)\int E\big(U(\gamma)|\mathcal{C}\big)\,dG(\gamma)\bigg).\end{aligned}$$

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 the following:

ASSUMPTION CUγ: (a) *For all* $\gamma\in\Gamma$, $\mathrm{Cov}(X(\gamma), U(\gamma)|\mathcal{C}) = 0$ *a.s.*
(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: *For all* $\gamma\in\Gamma$, $E(U(\gamma)|\mathcal{C}) = 0$ *a.s.*

LEMMA 3: (a) *Assumptions* S *and* CUγ *imply Assumption* CU.
(b) *Assumptions* S *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 structure discussed below the additional condition is not necessarily innocuous.

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

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

ASSUMPTION SF3γ: *For all* $\gamma \in \Gamma$, $EU^*(\gamma) = 0$.

Assumption SF2γ(b) requires a certain degree of homogeneity across population units. See the comment following Lemma 3.

LEMMA 4: (a) *Assumptions* S *and* SF2γ *imply Assumption* SF2.
(b) *Assumptions* S *and* SF3γ *imply Assumption* SF3.

Analogous sufficient conditions for Assumptions HF2, HF3, FF2, and FF3 in terms of population quantities are given in Andrews (2003).

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 $t = 1, \ldots, T$ 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

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

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 estimators of β_0 and α_0 are defined as above but with all sums taken over $t = 1, \ldots, T$ as well as $i = 1, \ldots, n$ 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 $E(X_iX_i'|\mathcal{C})$ 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 $E(X_i|\mathcal{C})$, $E(X_iU_i|\mathcal{C})$, 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 estimators. For the within estimator, the model we consider is

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

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

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

where

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

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

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

The analogue of Assumption CU for the within estimator is

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

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 γth cluster. Population units in the γth cluster are indexed by $b = 1, \ldots, B$, where $B \leq \infty$ denotes the cluster size. A sample of n clusters is selected via i.i.d. indices $\{\gamma_i : i = 1, \ldots, n\}$ 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_ith cluster are denoted with t subscripts for $t = 1, \ldots, T$. For example, the regressor variables are $X_t(\gamma_i)$ for $t = 1, \ldots, T$. Then, as in the panel model of (8.1), the LS and covariance matrix estimators are defined with sums taken over $t = 1, \ldots, T$ as well as $i = 1, \ldots, n$ 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.

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 slope coefficient estimators in regression models with common shocks are that the errors are uncorrelated with the regressors conditional on the σ-field generated by the common shocks. The LS estimators are shown to have a mixed normal asymptotic distribution after suitable centering and scaling. The paper shows that when the LS 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 conditional on the common shocks a.s., then the null rejection probabilities of t, Wald, and F tests based on the LS estimators converge to 1 as $n \to \infty$ and confidence interval coverage probabilities converge to 0 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 and regressors conditional on the common shocks.

Cowles Foundation for Research in Economics, Dept. of Economics, Yale University, P.O. Box 208281, Yale Station, New Haven, CT 06520-8281, U.S.A.; donald.andrews@yale.edu.

Manuscript received June, 2003; final revision received August, 2004.

APPENDIX: PROOFS

PROOF OF THEOREM 1: With convergence in probability replaced by convergence almost surely, the theorem follows straightforwardly from (3.1), (3.2), and Lemma 1 using Assumptions 1 and 2. The convergence in probability result then follows from the almost sure convergence result. *Q.E.D.*

PROOF OF THEOREM 2: Assumption STD1 holds by the calculations:

$$\begin{aligned} EX_iU_i &= E_{\mathcal{C}}C_2E(X_i^*U_i^{*\prime}|\mathcal{C})C_1 \\ &= E_{\mathcal{C}}C_2E(X_i^*U_i^{*\prime})C_1 \\ &= E_{\mathcal{C}}C_2E(X_i^*U_{1i}^*) + E_{\mathcal{C}}C_2E(X_i^*U_{2i}^{*\prime})C_{11} \\ &= 0 \end{aligned} \tag{A.1}$$

for all $i \geq 1$, 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 $EU_i = 0$.

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

$$E(U_i|\mathcal{C}) = E(U_i^*)'C_1 = EU_{1i}^* + C_{11}EU_{2i}^* = 0, \tag{A.2}$$

where the first equality holds by (7.5), 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 (A.2), we have

$$\begin{aligned} \mathrm{Cov}(X_i, U_i|\mathcal{C}) &= E(X_iU_i|\mathcal{C}) \\ &= C_2E(X_i^*U_i^{*\prime})C_1 \\ &= C_2EX_i^*U_{1i}^* + C_2E(X_i^*U_{2i}^*)C_{11} \\ &= C_2E(X_i^*U_{2i}^*)C_{11} \\ &\neq 0 \quad \text{with positive probability,} \end{aligned} \tag{A.3}$$

where the second and third equalities hold by the same arguments as in (A.2) and the fourth equality and the inequality hold by Assumption SF4(a) and (d). *Q.E.D.*

PROOF OF THEOREM 3: We have

$$\begin{aligned} E(U_i|\mathcal{C}) &= E\big(U_i(C(S_{0,i}))|\mathcal{C}\big) \\ &= E_{S_0,i}E\big(U_i(C(S_{0,i}))|\mathcal{C}, S_{0,i}\big) \\ &= E_{S_0,i}E_{X_i(\cdot),S_{0,i}}\big(U_i(C(S_{0,i}))|S_{0,i}\big), \end{aligned} \tag{A.4}$$

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 Assumption FF1(b) and (f).

By similar arguments, we obtain

$$\text{(A.5)} \qquad \begin{aligned} E(X_iU_i|\mathcal{C}) &= E_{S_{0,i}}E_{X_i(\cdot),S_{0,i}}\big(X_i(C(S_{0,i}))U_i(C(S_{0,i}))|S_{0,i}\big), \\ E(X_i|\mathcal{C}) &= E_{S_{0,i}}E_{X_i(\cdot),S_{0,i}}\big(X_i(C(S_{0,i}))|S_{0,i}\big). \end{aligned}$$

Combining (A.4) and (A.5) gives

$$\text{(A.6)} \qquad \begin{aligned} &E(X_iU_i|\mathcal{C}) - E(X_i|\mathcal{C})E(U_i|\mathcal{C}) \\ &\quad = E_{S_{0,i}}E_{X_i(\cdot),S_{0,i}}\big(X_i(C(S_{0,i}))U_i(C(S_{0,i}))|S_{0,i}\big) \\ &\qquad - E_{S_{0,i}}E_{X_i(\cdot),S_{0,i}}\big(X_i(C(S_{0,i}))|S_{0,i}\big) \cdot E_{S_{0,i}}E_{X_i(\cdot),S_{0,i}}\big(U_i(C(S_{0,i}))|S_{0,i}\big) \\ &\quad = E_{S_{0,i}}\big[E_{X_i(\cdot),S_{0,i}}\big(X_i(C(S_{0,i}))U_i(C(S_{0,i}))|S_{0,i}\big) \\ &\qquad\quad - E_{X_i(\cdot),S_{0,i}}\big(X_i(C(S_{0,i}))|S_{0,i}\big) \cdot E_{X_i(\cdot),S_{0,i}}\big(U_i(C(S_{0,i}))|S_{0,i}\big)\big] \\ &\quad = 0, \end{aligned}$$

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 (A.4) and Assumption FF3, the right-hand side of (A.4) equals zero a.s. Hence, Assumption CMZ holds. *Q.E.D.*

PROOF OF THEOREM 4: To prove part (a), we write

$$\text{(A.7)} \qquad \begin{aligned} &n^{1/2}\big(\widehat{\beta}_n - \beta_0 - r_n(\mathcal{C})\big) \\ &\quad = \left(n^{-1}\sum_{i=1}^{n} X_iX_i' - \overline{X}_n\overline{X}_n'\right)^{-1} \\ &\qquad \times n^{-1/2}\sum_{i=1}^{n}\big([X_i - \overline{X}_n]U_i - E\big([X_i - E(X_i|\mathcal{C})]U_i|\mathcal{C}\big)\big) \\ &\quad = \left(n^{-1}\sum_{i=1}^{n} X_iX_i' - \overline{X}_n\overline{X}_n'\right) \end{aligned}$$

$$
\begin{aligned}
&\times\Bigg(n^{-1/2}\sum_{i=1}^{n}\big\{[X_i - E(X_i|\mathcal{C})]U_i - E\big([X_i - E(X_i|\mathcal{C})]U_i|\mathcal{C}\big)\\
&\qquad - [X_i - E(X_i|\mathcal{C})]E(U_i|\mathcal{C})\big\}\\
&\qquad - [\overline{X}_n - E(X_i|\mathcal{C})]n^{-1/2}\sum_{i=1}^{n}[U_i - E(U_i|\mathcal{C})]\Bigg)\\
&= \big(B_{\mathcal{C}}^{-1} + o_p(1)\big)\left(n^{-1/2}\sum_{i=1}^{n}\xi_i\right) + o_p(1),
\end{aligned}
$$

where ξ_i is defined in (4.2) and the third equality of (A.7) holds using Lemma 1 to obtain the $B_{\mathcal{C}}^{-1} + o_p(1)$ result, using Lemma 1 to obtain $\overline{X}_n - E(X_i|\mathcal{C}) = o_p(1)$, and using an MDS CLT to obtain

$$
n^{-1/2}\sum_{i=1}^{n}[U_i - E(U_i|\mathcal{C})] = O_p(1). \tag{A.8}
$$

In particular, we apply Corollary 3.1 of Hall and Heyde (1980, p. 59) to obtain (A.8). For $i \geq 1$, let $\mathcal{F}_i$ denote the σ-field generated by $\mathcal{C}$ and $(W_1, \ldots, W_i)$. Then $\{U_i - E(U_i|\mathcal{C}), \mathcal{F}_i : i \geq 1\}$ is an MDS because $\{U_i : i \geq 1\}$ are i.i.d. 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 $\varepsilon > 0$,

$$
\begin{aligned}
&\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\big([U_i - E(U_i|\mathcal{C})]^2 1\big(|U_i - E(U_i|\mathcal{C})| > n^{1/2}\varepsilon\big)|\mathcal{F}_{i-1}\big)\\
&\quad = \lim_{n\to\infty} E\big([U_i - E(U_i|\mathcal{C})]^2 1\big(|U_i - E(U_i|\mathcal{C})| > n^{1/2}\varepsilon\big)|\mathcal{C}\big) = 0 \quad \text{a.s.},
\end{aligned} \tag{A.9}
$$

where the first equality holds because $\{U_i : i \geq 1\}$ are i.i.d. 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:

$$
n^{-1}\sum_{i=1}^{n} E\big([U_i - E(U_i|\mathcal{C})]^2|\mathcal{F}_{i-1}\big) = E\big([U_i - E(U_i|\mathcal{C})]^2|\mathcal{C}\big). \tag{A.10}
$$

Equation (A.10) 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

$$(A.11)\quad n^{-1/2}\sum_{i=1}^{n}[U_i - E(U_i|\mathcal{C})] \xrightarrow{d} E\big([U_i - E(U_i|\mathcal{C})]^2|\mathcal{C}\big) \times Z^*,$$

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 (A.8).

Next, $\{\xi_i, \mathcal{F}_i : i \geq 1\}$ is an 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

$$(A.12)\quad n^{-1/2}\sum_{i=1}^{n}\xi_i \xrightarrow{d} \Omega_{\mathcal{C}} \times Z,$$

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 (A.10). Combining (A.7) and (A.12) 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 premultiplied by $V_{\mathcal{C}}^{-1/2}$.

Part (c) of the theorem holds using Lemma 1. *Q.E.D.*

PROOF OF THEOREM 5: Part (a) holds by Lemma 1.

To prove part (b), for notational simplicity, suppose X_i is a scalar (otherwise one can establish the results element by element). Using Theorem 1, we have

$$\begin{aligned}(A.13)\quad \widehat{U}_i &= [U_i - E(U_i|\mathcal{C})] - [\widehat{\alpha}_n - \alpha_0 - E(U_i|\mathcal{C})] - X_i(\widehat{\beta}_n - \beta_0)\\ &= [U_i - E(U_i|\mathcal{C})] + E(X_i|\mathcal{C})r(\mathcal{C}) + o_p(1) - X_i\big(r(\mathcal{C}) + o_p(1)\big)\\ &= [U_i - E(U_i|\mathcal{C})] - [X_i - E(X_i|\mathcal{C})][r(\mathcal{C}) + o_p(1)] + o_p(1)\end{aligned}$$

(where $o_p(1)$ does not depend on i). Using (A.13), we can write $\widehat{\Omega}_n$ as

$$\begin{aligned}(A.14)\quad & n^{-1}\sum_{i=1}^{n}[U_i - E(U_i|\mathcal{C})]^2(X_i - \overline{X}_n)^2\\ &+ [r(\mathcal{C}) + o_p(1)]^2\left(n^{-1}\sum_{i=1}^{n}[X_i - E(X_i|\mathcal{C})]^2(X_i - \overline{X}_n)^2\right)\\ &- 2[r(\mathcal{C}) + o_p(1)]n^{-1}\sum_{i=1}^{n}[X_i - E(X_i|\mathcal{C})][U_i - E(U_i|\mathcal{C})](X_i - \overline{X}_n)^2\end{aligned}$$

$$+\, o_p(1)2n^{-1}\sum_{i=1}^{n}\bigl([U_i - E(U_i|\mathcal{C})] - [X_i - E(X_i|\mathcal{C})][r(\mathcal{C}) + o_p(1)]\bigr)$$
$$\times\, (X_i - \overline{X}_n)^2$$
$$+\, o_p(1)n^{-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

The probability limit of (A.14) is $\Omega^0_{\mathcal{C}} + \eta_{\mathcal{C}}$ using Lemma 1. 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 (A.14) for $\widehat{\Omega}_n$ with $X_i - \overline{X}_n$ replaced by 1. This, combined with Lemma 1, establishes part (d).

Part (e) follows from parts (a) and (d). *Q.E.D.*

PROOF OF THEOREM 6: We have

$$T_n = [\widehat{V}_n]^{-1/2}_{j,j}\sqrt{n}\bigl(\widehat{\beta}_{n,j} - \beta_{0,j} - r_n(\mathcal{C})_j\bigr) + [\widehat{V}_n]^{-1/2}_{j,j}\sqrt{n}r_n(\mathcal{C})_j, \tag{A.15}$$

where $r_n(\mathcal{C})_j$ denotes the jth 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 4(a) and 5(c) and comment 2 to Theorem 5. Hence, part (a) of the Theorem 6 holds.

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

$$|T_n| = \bigl|O_p(1) + \bigl([B^{-1}_{\mathcal{C}}(\Omega^0_{\mathcal{C}} + \eta_{\mathcal{C}})B^{-1}_{\mathcal{C}}]_{j,j} + o_p(1)\bigr)^{-1/2}\sqrt{n}r_n(\mathcal{C})_j\bigr| \xrightarrow{p} \infty, \tag{A.16}$$

where the equality holds using Theorems 4(a) and 5(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 4(c). In consequence, part (b) of the Theorem 6 holds. Part (c) holds by a similar argument.

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

$$W_n = \bigl\|(R\widehat{V}_nR')^{-1/2}n^{1/2}R\bigl(\widehat{\beta}_n - \beta_0 - r_n(\mathcal{C})\bigr) + (R\widehat{V}_nR')^{-1/2}n^{1/2}Rr_n(\mathcal{C})\bigr\|^2. \tag{A.17}$$

When Assumption CU holds, we have $Rr_n(\mathcal{C}) = 0$ a.s. and $(R\widehat{V}_nR')^{-1/2}n^{1/2} \times R(\widehat{\beta}_n - \beta_0) \to_d N(0, I_q)$ by Theorems 4(a) and 5(c). Hence, part (d) of the theorem holds.

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

$$W_n = \bigl\|O_p(1) + n^{1/2}\bigl(RB^{-1}_{\mathcal{C}}[\Omega^0_{\mathcal{C}} + \eta_{\mathcal{C}}]B^{-1}_{\mathcal{C}}R' + o_p(1)\bigr)^{-1/2}Rr_n(\mathcal{C})\bigr\|^2 \xrightarrow{p} \infty, \tag{A.18}$$

where the first equality uses (A.17) and Theorems 4(a) and 5(c), and the divergence to infinity uses the fact that $B^{-1}_{\mathcal{C}}[\Omega^0_{\mathcal{C}} + \eta_{\mathcal{C}}]B^{-1}_{\mathcal{C}}$ is nonsingular a.s. by Assumptions 2(d) and 5(a). Hence, part (e) of the theorem holds. *Q.E.D.*

PROOF OF LEMMA 3: Using (7.5), which relies on Assumption S, we have

$$
\begin{aligned}
\text{(A.19)}\quad & E(U_i|\mathcal{C}) = \int E\big(U(\gamma)|\mathcal{C}\big)\, dG(\gamma),\\
& E(X_i|\mathcal{C}) = \int E\big(X(\gamma)|\mathcal{C}\big)\, dG(\gamma),\\
& E(X_iU_i|\mathcal{C}) = \int E\big(X(\gamma)U(\gamma)|\mathcal{C}\big)\, dG(\gamma).
\end{aligned}
$$

Equation (A.19) and Assumption CMZ give

$$
\text{(A.20)}\quad E(U_i|\mathcal{C}) = \int E\big(U(\gamma)|\mathcal{C}\big)\, dG(\gamma) = 0.
$$

Combining the results of (A.19) gives

$$
\begin{aligned}
\text{(A.21)}\quad & E(X_iU_i|\mathcal{C}) - E(X_i|\mathcal{C})E(U_i|\mathcal{C})\\
&= \int E\big(X(\gamma)U(\gamma)|\mathcal{C}\big)\, dG(\gamma)\\
&\quad - \int E\big(X(\gamma)|\mathcal{C}\big)\, dG(\gamma) \int E\big(U(\gamma)|\mathcal{C}\big)\, dG(\gamma),\\
&= \int \big[E\big(X(\gamma)U(\gamma)|\mathcal{C}\big) - E\big(X(\gamma)|\mathcal{C}\big)E\big(U(\gamma)|\mathcal{C}\big)\big]\, dG(\gamma)\\
&= 0,
\end{aligned}
$$

where the second equality holds by Assumption CUγ(b) and the third equality holds by Assumption CUγ(a). *Q.E.D.*

PROOF OF LEMMA 4: The proof is analogous to that of Lemma 3 using (7.4) in place of (7.5). *Q.E.D.*

REFERENCES

ANDREWS, D. W. K. (2003): "Cross-Section Regression with Common Shocks," Discussion Paper 1428, Cowles Foundation, Yale University. Available at http://cowles.econ.yale.edu.

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

——— (2004): "A PANIC Attack on Unit Roots and Cointegration," *Econometrica*, 72, 1127–1177.

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 *Latent Variables in Socio-Economic Models*, ed. by D. J. Aigner and A. S. Goldberger. Amsterdam: North-Holland, 365–383.

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 (2004): "Testing for a Unit Root in Panels with Dynamic Factors," *Journal of Econometrics*, 122, 81–126.

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 & 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 (2003): "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence," *The Econometrics Journal*, 6, 217–259.

——— (2004): "Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence," Discussion Paper 1438, Cowles Foundation, Yale University.

SARGENT, T. J., AND C. A. SIMS (1977): "Business Cycle Modeling Without Pretending to Have Too Much *a priori* Economic Theory," in *New Methods in Business Cycle Research*: *Proceedings from a Conference*, ed. by C. A. Sims. Minneapolis: Federal Reserve Bank of Minneapolis, 45–110.

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 (2002): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics*, 20, 147–162.