Econometrica, Vol. 72, No. 4 (July, 2004), 1127–1177

A PANIC ATTACK ON UNIT ROOTS AND COINTEGRATION

BY JUSHAN BAI AND SERENA NG[1]

This paper develops a new methodology that makes use of the factor structure of large dimensional panels to understand the nature of nonstationarity in the data. We refer to it as PANIC—Panel Analysis of Nonstationarity in Idiosyncratic and Common components. PANIC can detect whether the nonstationarity in a series is pervasive, or variable-specific, or both. It can determine the number of independent stochastic trends driving the common factors. PANIC also permits valid pooling of individual statistics and thus panel tests can be constructed. A distinctive feature of PANIC is that it tests the unobserved components of the data instead of the observed series. The key to PANIC is consistent estimation of the space spanned by the unobserved common factors and the idiosyncratic errors without knowing a priori whether these are stationary or integrated processes. We provide a rigorous theory for estimation and inference and show that the tests have good finite sample properties.

KEYWORDS: Panel data, common factors, common trends, principal components.

1. INTRODUCTION

KNOWLEDGE OF WHETHER a series is stationary or nonstationary is important for a wide range of economic analysis. As such, unit root testing is extensively conducted in empirical work. But in spite of the development of many elegant theories, the power of univariate unit root tests is severely constrained in practice by the short span of macroeconomic time series. Panel unit root tests have since been developed with the goal of increasing power through pooling information across units. But pooling is valid only if the units are independent, an assumption that is perhaps unreasonable given that many economic models imply, and the data support, the comovement of economic variables.

In this paper, we propose a new approach to understanding nonstationarity in the data, both on a series by series basis, and from the viewpoint of a panel. Rather than treating the cross-section correlation as a nuisance, we exploit these comovements to develop new univariate statistics and valid pooled tests for the null hypothesis of nonstationarity. Our tests are applied to two unobserved components of the data, one with the characteristic that it is strongly correlated with many series, and one with the characteristic that it is largely unit specific. More precisely, we consider a factor analytic model:

$$X_{it} = D_{it} + \lambda_i' F_t + e_{it},$$

where D_{it} is a polynomial trend function, F_t is an $r \times 1$ vector of common factors, and λ_i is a vector of factor loadings. The series X_{it} is the sum of a

[1]This paper was presented at the NSF 2001 Summer Symposium on Econometrics and Statistics in Berkeley, California, the CEPR/Banca d'Italia Conference in Rome, and at NYU, Princeton, Toronto, Maryland, Virginia, Michigan, and LSE. We thank three anonymous referees, the editor, and the seminar participants for many helpful comments. The authors acknowledge financial support from the NSF (Grants SBR 9896329, SES-0136923, and SES-0137084).

deterministic component D_{it}, a common component $\lambda_i' F_t$, and an error e_{it} that is largely idiosyncratic. A factor model with N variables has N idiosyncratic components but a small number of common factors.[2]

A series with a factor structure is nonstationary if one or more of the common factors are nonstationary, or the idiosyncratic error is nonstationary, or both. Except by assumption, there is nothing that restricts F_t to be all I(1) or all I(0). There is also nothing that rules out the possibility that F_t and e_{it} are integrated of different orders. These are not merely cases of theoretical interest, but also of empirical relevance. As an example, let X_{it} be real output of country i. It may consist of a global trend component F_{1t}, a global cyclical component F_{2t}, and an idiosyncratic component e_{it} that may or may not be stationary. As another example, the inflation rate of durable goods may consist of a component that is common to all prices, and a component that is specific to durable goods. Whether these components are stationary or nonstationary is an empirical matter.

It is well known that the sum of two time series can have dynamic properties very different from the individual series themselves. If one component is I(1) and one is I(0), it could be difficult to establish that a unit root exists from observations on X_{it} alone, especially if the stationary component is large. Unit root tests on X_{it} can be expected to be oversized while stationarity tests will have no power. The issue is documented in Schwert (1989), and formally analyzed in Pantula (1991), Ng and Perron (2001), and among others, in the context of a negative moving-average component in the first-differenced data.

Instead of testing for the presence of a unit root in X_{it}, the approach proposed in this paper is to test the common factors and the idiosyncratic components separately. We refer to such a Panel Analysis of Nonstationarity in the Idiosyncratic and Common components as PANIC. PANIC has two objectives. The first is to determine if nonstationarity comes from a pervasive or an idiosyncratic source. The second is to construct valid pooled tests for panel data when the units are correlated. PANIC can also potentially resolve three econometric problems. The first is the problem of size distortion just mentioned, namely, existing tests in the literature tend to over-reject the nonstationarity hypothesis when the series being tested is the sum of a weak I(1) component and a strong stationary component. The second is a consequence of the fact that the idiosyncratic components in a factor model can only be weakly correlated across i by design. In contrast, X_{it} will be strongly correlated across units if the data obey a factor structure. Thus, pooled tests based upon e_{it} are more likely to satisfy the cross-section independence assumption required for pooling. The third relates to power, and follows from the fact that pooled tests exploit cross-section information and are more powerful than univariate unit root tests.

[2]This is a static factor model, and is to be distinguished from the dynamic factor model being analyzed in Forni, Hallin, Lippi, and Reichlin (2000).

Since the factors and the idiosyncratic components are both unobserved, and our objective is to test if they have unit roots, the key to our analysis is consistent estimation of these components when it is not known a priori whether they are I(1) or I(0). To this end, we propose a robust common-idiosyncratic (I-C) decomposition of the data using large dimensional panels, that is, datasets in which the number of observations in the time (T) and the cross-section (N) dimensions are both large. Loosely speaking, the large N permits consistent estimation of the common variation whether or not it is stationary, while a large T enables application of the relevant central limit theorems so that limiting distributions of the tests can be obtained. Robustness is achieved by a "differencing and recumulating" estimation procedure so that I(1) and I(0) errors can be accommodated. We provide a rigorous development of the theory for this estimation procedure. Our results add to the growing literature on large dimensional factor analysis by showing how consistent estimates of the factors can be obtained using the method of principal components even without imposing stationarity on the errors. These results can be used to study other dynamic properties of the common factors, such as long memory, ARCH effects, and structural change, under very general conditions.

Several authors have also developed panel unit roots to resolve the problem of correlated errors. In Chang (2002), Moon and Perron (2003), Chang and Song (2002), and Phillips and Sul (2003), for example, cross-section dependence is treated as a nuisance. In contrast, the nature of the cross-section dependence is itself an object of interest in our analysis. We allow for the possibility of multiple factors, and the framework is thus more general than the one-way error component model of Choi (2002). Furthermore, these papers are ultimately interested in testing for unit roots in the observed data. We go beyond this to analyze the source of nonstationarity. In doing so, we provide a coherent framework for studying unit roots, common trends, and common cycles in large dimensional panels.

Our framework differs from conventional multivariate time-series models in which N is small. In small N analysis of cointegration, common trends, and cycles, the estimation methodology being employed typically depends on whether the variables considered are all I(1) or all I(0).[3] Pretesting for unit roots is thus necessary. Because N is small, what is extracted is the trend or the cycle common to just a small number of variables. Not only is the information in many potentially relevant series left unexploited, consistent estimation of the common factors is in fact not possible when the number of variables is small. In our analysis with N and T large, the common variation can be extracted without appealing to stationarity assumptions and/or cointegration restrictions. This makes it possible to decouple the extraction of common trends and cycles from the issue of testing stationarity.

[3] For example, King, Plosser, Stock, and Watson (1991), Engle and Kozicki (1993), and Gonzalo and Granger (1995).

The rest of the paper is organized as follows. In Section 2, we describe the PANIC procedures and present asymptotic results for testing single and multiple unit roots. We devote Section 3 to the large sample properties of the factor estimates. As an intermediate result, we establish uniform consistency of the factor estimates without assuming the errors are stationary. This result is of interest in much broader contexts than unit root testing. Section 4 uses simulations to illustrate the properties of the factor estimates and the tests in finite samples.

2. PANIC

The data X_{it} $(i=1,\ldots,N;\ t=1,\ldots,T)$ are assumed to be generated by

$$X_{it}=c_i+\beta_i t+\lambda_i' F_t+e_{it}, \tag{1}$$

$$(I-L)F_t=C(L)u_t, \tag{2}$$

$$(1-\rho_i L)e_{it}=D_i(L)\epsilon_{it}, \tag{3}$$

where $C(L)=\sum_{j=0}^{\infty}C_jL^j$ and $D_i(L)=\sum_{j=0}^{\infty}D_{ij}L^j$. The idiosyncratic error e_{it} is I(1) if $\rho_i=1$, and is stationary if $|\rho_i|<1$. We allow r_0 stationary factors and r_1 common trends, with $r=r_0+r_1$. Stated differently, the rank of $C(1)$ is r_1. The objective is to determine r_1 and test if $\rho_i=1$ when neither F_t nor e_{it} is observed and will be estimated by the method of principal components.

2.1. *Assumptions and Overview*

Let $M<\infty$ be a generic positive number, not depending on T or N. Let $\|A\|=\text{trace}(A'A)^{1/2}$. Our analysis is based on the following assumptions:

ASSUMPTION A: (i) *For nonrandom* λ_i, $\|\lambda_i\|\le M$; *for random* λ_i, $E\|\lambda_i\|^4\le M$; (ii) $N^{-1}\sum_{i=1}^N\lambda_i\lambda_i'\xrightarrow{p}\Sigma_\Lambda$, *an* $r\times r$ *positive definite matrix.*

ASSUMPTION B: (i) $u_t\sim\text{iid}(0,\Sigma_u)$, $E\|u_t\|^4\le M$; (ii) $\text{var}(\Delta F_t)=\sum_{j=0}^{\infty}C_j\Sigma_u C_j'>0$; (iii) $\sum_{j=0}^{\infty}j\|C_j\|<M$; *and* (iv) $C(1)$ *has rank* r_1, $0\le r_1\le r$.

ASSUMPTION C: (i) *For each* i, $\epsilon_{it}\sim\text{iid}(0,\sigma_{\epsilon i}^2)$, $E|\epsilon_{it}|^8\le M$, $\sum_{j=0}^{\infty}j|D_{ij}|<M$, $\omega_{\epsilon i}^2=D_i(1)^2\sigma_{\epsilon i}^2>0$; (ii) $E(\epsilon_{it}\epsilon_{jt})=\tau_{ij}$ *with* $\sum_{i=1}^N|\tau_{ij}|\le M$ *for all* j; (iii) $E|N^{-1/2}\sum_{i=1}^N[\epsilon_{is}\epsilon_{it}-E(\epsilon_{is}\epsilon_{it})]|^4\le M$, *for every* (t,s).

ASSUMPTION D: *The errors* ϵ_{it}, u_t, *and the loadings* λ_i *are three mutually independent groups*.

ASSUMPTION E: $E\|F_0\|\le M$, *and for every* $i=1,\ldots,N$, $E|e_{i0}|\le M$.

Assumption A is made on the factor loadings to ensure that the factor structure is identifiable. It is a common assumption in factor analysis. A set of factors F_t is deemed to be pervasive if and only if the corresponding loading coefficients are such that $N^{-1}\sum_{i=1}^{N}\lambda_i\lambda_i'$ converges to a positive definite matrix as $N\to\infty$. A variable that has only a finite number of nonzero loadings does not satisfy this condition and is not a factor in our large N framework. Instead, its variation will be considered idiosyncratic, and thus included in ϵ_{it}.

Under Assumption B, the short run variance of ΔF_t is required to be positive definite, but the long-run covariance of ΔF_t can be reduced rank to permit linear combinations of I(1) factors to be stationary. When $r_1 = 0$ and there are no stochastic trends, $C(1)$ is null because ΔF_t is overdifferenced. On the other hand, when $r_1 \neq 0$, one can consider a rotation of F_t by a matrix G such that the first r_1 elements of GF_t are integrated, while the final r_0 elements are stationary. One such rotation is given by $G = [\beta_\perp' \beta']'$, where $\beta_\perp$ is $r \times r_1$ satisfying $\beta_\perp'\beta_\perp = I_{r_1}$, and $\beta_\perp'\beta = 0$. We define $Y_t = \beta_\perp' F_t$ to be the r_1 common stochastic trends resulting from such a rotation.

Assumption C(i) allows some weak serial correlation in $(1-\rho_i L)e_{it}$ with ρ_i possibly different across i, while C(ii) and C(iii) allow weak cross-section correlation. Clearly, C(ii) holds if ε_{it} are cross-sectionally uncorrelated. The assumption obviously holds if there exists an ordering of the cross sections such that the ordered ε_{it} $(i = 1, 2, \ldots, N)$ is a mixing process.[4] But the assumption is more general. It allows weak cross-correlation in the errors, weak in the sense that even as N increases, the column sum of the error covariance matrix remains bounded. Chamberlain and Rothschild (1983) defined an approximate factor model as one in which the largest eigenvalue of Ω is bounded. But if e_t is stationary with $E(e_{it}e_{jt}) = \tau_{ij}$, then from matrix theory, the largest eigenvalue of Ω is bounded by $\max_j \sum_{i=1}^{N}|\tau_{ij}|$. Since C(ii) requires that $\sum_{i=1}^{N}|\tau_{ij}| \le M$ for all j and all N, we have an "approximate factor model" in the sense of Chamberlain and Rothschild (1983). Under Assumption D, ϵ_{it}, u_t, and λ_i are mutually independent across i and t. The assumption is stronger than the one used in Bai and Ng (2002), which permits u_t and ϵ_{it} to be weakly correlated. Assumption E is an initial condition assumption made commonly in unit root analysis.

Our factor estimates are based on the method of principal components. When e_{it} is I(0), the principal components estimators for F_t and λ_i have been shown to be consistent when all the factors are I(0) and when some or all of them are I(1). But consistent estimation of the factors when e_{it} is I(1) has not been considered in the literature. Indeed, when e_{it} has a unit root, a regression of X_{it} on F_t is spurious even if F_t was observed, and the estimates of λ_i and thus of e_{it} will not be consistent. The validity of PANIC thus hinges on the ability to obtain estimates of F_t and e_{it} that preserve their orders of integration,

[4]Such an assumption was made in Connor and Korajzcyk (1986).

both when e_{it} is I(1) and when it is I(0). A contribution of this paper is to show how this can be accomplished. Essentially, the trick is to apply the method of principal components to the first-differenced data.

To be precise, suppose we observe X, a data matrix with T time-series observations and N cross-section units. Suitably transform X to yield x, a set of $(T-1)\times N$ stationary variables. Let $f=(f_2, f_3, \dots, f_T)'$ and $\Lambda = (\lambda_1, \dots, \lambda_N)'$. The principal component estimator of f, denoted $\hat{f}$, is $\sqrt{T-1}$ times the r eigenvectors corresponding to the first r largest eigenvalues of the $(T-1)\times(T-1)$ matrix xx'. Under the normalization $\hat{f}'\hat{f}/(T-1)=I_r$, the estimated loading matrix is $\hat{\Lambda}=x'\hat{f}/(T-1)$.

Before turning to the details, an overview of the inference procedures gives an idea of what is to follow. If there is one factor, PANIC will test if it is a unit root process. If there are multiple factors, PANIC will determine r_1, the number of independent stochastic trends underlying the r common factors. In addition, PANIC will test if there is a unit root in each of the idiosyncratic errors. An important aspect of PANIC is that the idiosyncratic errors can be tested for the presence of a unit root without knowing if the factors are stationary, and vice versa. In fact, the tests on the factors are asymptotically independent of the tests on the idiosyncratic errors. In each case, we allow for the possibility that the differenced stationary series are serially correlated with (possibly infinite) autoregressive representations. The two univariate tests will be denoted by $ADF_{\hat{e}}(i)$ and $ADF_{\hat{F}}$ respectively, as they are based on the t test of Said and Dickey (1984) using an augmented autoregression with suitably chosen lag lengths. In the case when $r>1$, we consider two tests. The first filters the factors under the assumption that they have finite order VAR representations. The second corrects for serial correlation of arbitrary form by nonparametrically estimating the relevant nuisance parameters. Accordingly, the "filtered" test is denoted by MQ_f, and the "corrected" test is denoted by MQ_c. These are modified versions of the Q_f and Q_c tests developed in Stock and Watson (1988).

The definition of x depends on the deterministic trend function. We consider two specifications, leading to what we will call the intercept only model and the linear trend model. The superscripts c and τ will be used to distinguish these two cases. The focus of this section is unit root inference. The properties of the factor estimates will be deferred to Section 3. The theory proceeds assuming r is known. We will return to the determination of r in practice in Section 4.

2.2. *The Intercept Only Case*

The factor model in the intercept only case is

$$X_{it} = c_i + \lambda_i' F_t + e_{it}. \tag{4}$$

Denote

$$x_{it} = \Delta X_{it}, \quad f_t = \Delta F_t, \quad \text{and} \quad z_{it} = \Delta e_{it}. \tag{5}$$

Then the model in first-differenced form is

$$x_{it} = \lambda_i' f_t + z_{it}. \tag{6}$$

Applying the method of principal components to x yields r estimated factors $\hat{f}_t$, the associated loadings $\hat{\lambda}_i$, and the estimated residuals, $\hat{z}_{it} = x_{it} - \hat{\lambda}_i' \hat{f}_t$. Define for $t = 2, \ldots, T$:

$$\hat{e}_{it} = \sum_{s=2}^{t} \hat{z}_{it} \qquad (i = 1, \ldots, N),$$

$$\hat{F}_t = \sum_{s=2}^{t} \hat{f}_s, \quad \text{an } r \times 1 \text{ vector}.$$

1. Let $ADF^c_{\hat{e}}(i)$ be the t statistic for testing $d_{i0} = 0$ in the univariate augmented autoregression (with no deterministic terms)

$$\Delta \hat{e}_{it} = d_{i0} \hat{e}_{it-1} + d_{i1} \Delta \hat{e}_{it-1} + \cdots + d_{ip} \Delta \hat{e}_{it-p} + \text{error}.$$

2. If $r = 1$, let $ADF^c_{\hat{F}}$ be the t statistic for testing $\delta_0 = 0$ in the univariate augmented autoregression (with an intercept):

$$\Delta \hat{F}_t = c + \delta_0 \hat{F}_{t-1} + \delta_1 \Delta \hat{F}_{t-1} + \cdots + \delta_p \Delta \hat{F}_{t-p} + \text{error}.$$

3. If $r > 1$, demean $\hat{F}_t$ and define $\hat{F}^c_t = \hat{F}_t - \overline{\hat{F}}$, where $\overline{\hat{F}} = (T-1)^{-1} \sum_{t=2}^{T} \hat{F}_t$. Start with $m = r$:

A: Let $\hat{\beta}_\perp$ be the m eigenvectors associated with the m largest eigenvalues of $T^{-2} \sum_{t=2}^{T} \hat{F}^c_t \hat{F}^{c\prime}_t$. Let $\hat{Y}^c_t = \hat{\beta}'_\perp \hat{F}^c_t$. Two statistics can be considered:

B.I: Let $K(j) = 1 - j/(J+1)$, $j = 0, 1, \ldots, J$:

(i) Let $\hat{\xi}^c_t$ be the residuals from estimating a first-order VAR in $\hat{Y}^c_t$, and let

$$\hat{\Sigma}^c_1 = \sum_{j=1}^{J} K(j) \left(T^{-1} \sum_{t=2}^{T} \hat{\xi}^c_{t-j} \hat{\xi}^{c\prime}_t \right).$$

(ii) Let $\nu^c_c(m)$ be the smallest eigenvalue of

$$\hat{\Phi}^c_c(m) = .5 \left[\sum_{t=2}^{T} (\hat{Y}^c_t \hat{Y}^{c\prime}_{t-1} + \hat{Y}^c_{t-1} \hat{Y}^{c\prime}_t) - T(\hat{\Sigma}^c_1 + \hat{\Sigma}^{c\prime}_1) \right] \left(\sum_{t=2}^{T} \hat{Y}^c_{t-1} \hat{Y}^{c\prime}_{t-1} \right)^{-1}.$$

(iii) Define $MQ_c^c(m) = T[\hat{\nu}_c^c(m) - 1]$.

B.II: For p fixed that does not depend on N or T:

(i) Estimate a VAR of order p in $\Delta\hat{Y}_t^c$ to obtain $\hat{\Pi}(L) = I_m - \hat{\Pi}_1 L - \cdots - \hat{\Pi}_p L^p$. Filter $\hat{Y}_t^c$ by $\hat{\Pi}(L)$ to get $\hat{y}_t^c = \hat{\Pi}(L)\hat{Y}_t^c$.

(ii) Let $\hat{\nu}_f^c(m)$ be the smallest eigenvalue of

$$\hat{\Phi}_f^c(m) = .5\left[\sum_{t=2}^{T}(\hat{y}_t^c\hat{y}_{t-1}^{c\prime} + \hat{y}_{t-1}^c\hat{y}_t^{c\prime})\right]\left(\sum_{t=2}^{T}\hat{y}_{t-1}^c\hat{y}_{t-1}^{c\prime}\right)^{-1}.$$

(iii) Define the statistic $MQ_f^c(m) = T[\hat{\nu}_f^c(m) - 1]$.

C: If $H_0 : r_1 = m$ is rejected, set $m = m - 1$ and return to step A. Otherwise, $\hat{r}_1 = m$ and stop.

THEOREM 1 (The Intercept Only Case): *Suppose the data are generated by* (2), (3), *and* (4) *and Assumptions* A–E *hold. Let* W_u *and* $W_{\epsilon i}$ $(i = 1, \ldots, N)$ *be standard Brownian motions. The following results hold as* $N, T \to \infty$.

1. *Let* p *be the order of autoregression chosen such that* $p \to \infty$ *and* $p^3/\min[N, T] \to 0$. *Under the null hypothesis that* $\rho_i = 1$,

$$ADF_{\hat{e}}^c(i) \Rightarrow \frac{\int_0^1 W_{\epsilon i}(s)\,dW_{\epsilon i}(s)}{(\int_0^1 W_{\epsilon i}(s)^2\,ds)^{1/2}} \qquad (i = 1, \ldots, N).$$

2. $(r = 1)$. *Let* p *be the order of autoregression chosen such that* $p \to \infty$ *and* $p^3/\min[N, T] \to 0$. *Let* $W_u^c(s) = W_u(s) - \int_0^1 W_u(s)\,ds$. *Under the null hypothesis that* F_t *has a unit root*,

$$ADF_{\hat{F}}^c \Rightarrow \frac{\int_0^1 W_u^c(s)\,dW_u(s)}{(\int_0^r W_u^c(s)^2\,ds)^{1/2}}.$$

3. $(r > 1)$. *Let* W_m *be an* m*-vector standard Brownian motion,* $W_m^c = W_m - \int_0^1 W_m$. *Let* $\nu_*^c(m)$ *be the smallest eigenvalue of*

$$\Phi_*^c = \frac{1}{2}[W_m^c(1)W_m^c(1)' - I_m]\left[\int_0^1 W_m^c(s)W_m^c(s)'\,ds\right]^{-1}.$$

(i) *Let* J *be the truncation lag of the Bartlett kernel, chosen such that* $J \to \infty$ *and* $J/\min[\sqrt{N}, \sqrt{T}\,] \to 0$. *Then under the null hypothesis that* F_t *has* m *stochastic trends,* $T[\hat{\nu}_c^c(m) - 1] \xrightarrow{d} \nu_*^c(m)$.

(ii) *Under the null hypothesis that* F_t *has* m *stochastic trends with a finite* VAR$(\bar{p})$ *representation and a* VAR(p) *is estimated with* $p \geq \bar{p}$, $T[\hat{\nu}_f^c(m) - 1] \xrightarrow{d} \nu_*^c(m)$.

Separately testing F_t and e_{it} allows us to disentangle the source of nonstationarity. If F_t is nonstationary but e_{it} is stationary, we say the nonstationarity of X_{it} is due to a pervasive source. On the other hand, if F_t is stationary but e_{it} is nonstationary, then the nonstationarity of X_{it} is due to a series-specific source. Evidently, if both F_t and e_{it} are nonstationary, both common and idiosyncratic variations contribute to the integratedness of X_{it}. Although direct testing of F_t and e_{it} is not feasible, Theorem 1 shows that testing $\hat{e}_{it}$ and $\hat{F}_t$ are the same as if e_{it} and F_t were observable.

As shown in the Appendix, $T^{-1/2}\hat{e}_{it} = T^{-1/2}e_{it} + o_p(1)$, where the $o_p(1)$ term is uniform in t. The asymptotic distribution of $ADF^c_{\hat{e}}(i)$ coincides with the DF test developed by Dickey and Fuller (1979) for the case of no constant. The critical value of the test at the 5% significance level is -1.95.

When the first-differenced data ΔX_{it} contain no deterministic terms, $T^{-1/2}\hat{F}_t = T^{-1/2}HF_t + o_p(1)$, where H is a full rank matrix and the $o_p(1)$ term is uniform in t. This means that the difference between the space spanned by estimated factors and the true factors is small. Testing for a unit root in demeaned $\hat{F}_t$ is asymptotically the same as testing for a unit root in demeaned F_t. When $r = 1$, this is a simple univariate test. The $ADF^c_{\hat{F}}$ has the same limiting distribution as the DF test for the constant only case. The 5% asymptotic critical value is -2.86.

Assuming that the series to be tested is observed, Said and Dickey (1984) showed that the ADF based upon an augmented autoregression has the same limiting distribution as the DF if the number of lags is chosen such that $p^3/T \to 0$ as $p, T \to \infty$. In our analysis, the series to be tested are $\hat{e}_{it}$ and $\hat{F}_t$. Since these are estimates of e_{it} and F_t, the allowed rate of increase in p depends on the rate at which the estimation errors vanish, giving the result $p^3/\min[N, T] \to 0$ as stated.

If all factors are I(1), linear combination of the factors will be I(1). If all factors are I(0), their linear combinations will still be I(0). However, linear combinations of I(1) and I(0) factors can remain I(1). Since we can only estimate the space spanned by the factors, individually testing each of the factors for the presence of a unit root will, in general, overstate the number of common trends. Accordingly, we need to determine the number of basis functions spanning the nonstationary space of F_t. Stock and Watson (1988) proposed two statistics, denoted Q_f and Q_c, designed to test if the real part of the smallest eigenvalue of an autoregressive coefficient matrix is unity. While the Q^c_f assumes the nonstationary components of F_t to be finite order vector-autoregressive processes, the Q^c_c allows the unit root process to have more general dynamics, including moving-average errors.

Our proposed MQ^c_c and MQ^c_f tests are modified variants of Stock and Watson's Q^c_c and Q^c_f. The basic difference is in the numerator of the tests. Instead of $\sum_{t=1}^{T} \hat{y}^c_t \hat{y}^{c\prime}_{t-1}$ in Q^c_f, we use as numerator $.5[\sum_{t=1}^{T}(\hat{y}^c_t \hat{y}^{c\prime}_{t-1} + \hat{y}^c_{t-1}\hat{y}^{c\prime}_t)]$, which can be thought of as an average of the autocovariance of $\hat{y}^c_t$ at lead and lag one. Such a numerator was also considered by Phillips and Durlauf (1986) for testing unit roots in vector time series. In the present context, the modification serves two purposes. First, the modified numerator is symmetric, and thus the eigenvalues are always real. Second, when $r = 1$, the identity $\sum_{t=2}^{T} \hat{y}_{t-1}^c \Delta \hat{y}_t^c = .5[(\hat{y}_T^c)^2 - \sum_{t=2}^{T} (\Delta \hat{y}_t^c)^2]$ holds under the null hypothesis of a unit root, leading in the limit to Ito's Lemma, $\int_0^1 W_c(s)\,dW(s) = .5[W_c(1)^2 - 1]$. As is well known, the numerator of the *ADF* test can be represented either way. In the multivariate case, the analogous identity is $\sum_{t=2}^{T} (\hat{y}_{t-1}^c \Delta \hat{y}_t^{c\prime} + \Delta \hat{y}_t^c \hat{y}_{t-1}^{c\prime}) = \hat{y}_T^c \hat{y}_T^{c\prime} - \sum_{t=2}^{T} \Delta \hat{y}_t^c \Delta \hat{y}_t^{c\prime}$, and so, in the limit, the numerators of our MQ^c tests still have two equivalent representations. At a more technical level, the modifications allow us to exploit use of the identity, which substantially simplifies the proofs.

The limiting distributions of $T[\hat{\Phi}_f^c(m) - I_m]$ and $T[\hat{\Phi}_c^c(m) - I_m]$ are of the form $A\Phi_*^c A^{-1}$, which has the same eigenvalues as Φ_*^c.[5] The critical values of both tests can thus be obtained by simulating Φ_*^c, which is based upon a vector of standard Brownian motions. These are reported in Table I.

Strictly speaking, the $MQ_f^c(m)$ test is valid only when the common trends can be represented as finite order AR(p) processes. From a theoretical point of view, the $MQ_c^c(m)$ is more general, as it only requires the weakly dependent errors to satisfy the moment conditions of Assumption B. We can then perform kernel estimation of the long-run minus the short-run residual variance of a VAR(1). Theorem 1, based upon the Bartlett kernel as in Newey and West (1987), shows that so long as the number of autocovariances, J, does not increase too fast, serial correlation of arbitrary form can be effectively removed nonparametrically. One can expect the results to generalize to other kernels, with appropriate restrictions to the truncation point.

TABLE I

CRITICAL VALUES FOR MQ_c AND MQ_f FOR TESTING $H_0: r_1 = m$ AT SIGNIFICANCE LEVEL φ

	$MQ_{c,f}^c$			$MQ_{c,f}^\tau$		
$m \setminus \varphi$	.01	.05	.10	.01	.05	.10
1	−20.151	−13.730	−11.022	−29.246	−21.313	−17.829
2	−31.621	−23.535	−19.923	−38.619	−31.356	−27.435
3	−41.064	−32.296	−28.399	−50.019	−40.180	−35.685
4	−48.501	−40.442	−36.592	−58.140	−48.421	−44.079
5	−58.383	−48.617	−44.111	−64.729	−55.818	−55.286
6	−66.978	−57.040	−52.312	−74.251	−64.393	−59.555

[5] Stock and Watson (1988) suggest normalizing the $\Pi(L)$ estimates so that $\Delta\hat{y}_t$ has unit variance. This is in fact not necessary.

In our analysis, r_1 is estimated by successive application of the MQ tests. If the chosen significance level at each stage is φ, then $P(\hat{r}_1 = r_1) \to 1 - \varphi$, and the overall asymptotic type I error is also φ. This property is generic of successive testing procedures, including Johansen's trace and eigenvalue tests for the number of cointegrating vectors.[6] The result is a consequence of the fact that the tests are applied to the same $\hat{F}_t$ and thus not independent across stages.

When an observed series is tested for a unit root using the ADF, it is known that the statistic diverges at rate $\sqrt{T}$ under the alternative of stationarity when p is chosen to increase with T such that $p^3/T \to 0$ as $T \to \infty$.

THEOREM 2: *Suppose $\hat{e}_{it}$ is tested for a unit root using the ADF, the assumptions of Theorem 1 hold, and p is chosen such that $p \to \infty$ and $p^4/\min[N, T] \to 0$ as N and $T \to \infty$. Under the alternative of stationarity, the statistic diverges at rate* $\min[\sqrt{N}, \sqrt{T}]$.

Although $\hat{e}_{it}$ yields asymptotically valid inference about nonstationarity of e_{it}, the fact that e_{it} is unobserved is not innocuous. As indicated in Theorem 1, the usual Dickey–Fuller limiting distribution obtains only when $p^3/\min[N, T] \to 0$ as $p, N, T \to \infty$. Theorem 2 shows that for the test to be consistent, we need $p^4/\min[N, T] \to 0$. As shown in the Appendix, the divergence rate is $\sqrt{N}$ if $T/N \to \infty$, and is $\sqrt{T}$ if T/N is bounded. The overall rate is thus $\min[\sqrt{N}, \sqrt{T}]$. Essentially, the power of the test is determined by how fast the error in estimating the factors vanishes, and thus depends on both N and T. The above rate of divergence also applies to testing $\hat{F}_t$ when $r = 1$.[7]

2.3. *The Linear Trend Case*

Consider now the factor model in the case of a linear trend:

$$X_{it} = c_i + \beta_i t + \lambda_i' F_t + e_{it} \tag{7}$$

and thus $\Delta X_{it} = \beta_i + \lambda_i' \Delta F_t + \Delta e_{it}$. Let $\overline{\Delta F} = (T-1)^{-1} \sum_{t=2}^{T} \Delta F_t$, $\overline{\Delta e_i} = (T-1)^{-1} \sum_{t=2}^{T} \Delta e_{it}$, and $\overline{\Delta X}_i = (T-1)^{-1} \sum_{t=2}^{T} \Delta X_{it}$. PANIC proceeds as follows for the case of a linear trend. Define

$$\Delta X_{it} - \overline{\Delta X}_i = \lambda_i' (\Delta F_t - \overline{\Delta F}) + (\Delta e_{it} - \overline{\Delta e_i}),$$

which can be rewritten as

$$x_{it} = \lambda_i' f_t + z_{it}, \tag{8}$$

[6]For a detailed discussion, see Johansen (1995, pp. 169 and 170).

[7]Stock and Watson showed (in an unpublished appendix) consistency of the Q tests. Even for observed data, the proof is quite involved. We conjecture a similar result to hold for the MQ tests.

where

$$x_{it} = \Delta X_{it} - \overline{\Delta X}_i, \quad f_t = \Delta F_t - \overline{\Delta F}, \quad z_{it} = \Delta e_{it} - \overline{\Delta e_i}. \tag{9}$$

Note that the differenced and demeaned data x_{it} are invariant to c_i and β_i. As a consequence, there is no loss of generality to assume $E(\Delta F_t) = 0$. For example, if $F_t = a + bt + \eta_t$ such that $E(\Delta\eta_t) = 0$, then we can rewrite model (7) with F_t replaced by η_t and $c_i + \beta_i t$ replaced by $c_i + \lambda_i' a + (\beta_i + \lambda_i' b)t$.

Let $\hat{f}_t$ and $\hat{\lambda}_i$ be the estimates obtained by applying the method of principal components to the differenced and demeaned data, x, with $\hat{z}_{it} = x_{it} - \hat{\lambda}_i' \hat{f}_t$, and $\hat{e}_{it} = \sum_{s=2}^{t} \hat{z}_{it}$. Also let $\hat{F}_t = \sum_{s=2}^{t} \hat{f}_s$, an $r \times 1$ vector.

1. Let $ADF_{\hat{e}}^{\tau}(i)$ be the t statistic for testing $d_{i0} = 0$ in the univariate augmented autoregression (with no deterministic terms)

$$\Delta\hat{e}_{it} = d_{i0}\hat{e}_{it-1} + d_{i1}\Delta\hat{e}_{it-1} + \cdots + d_{ip}\Delta\hat{e}_{it-p} + \text{error}.$$

2. If $r = 1$, let $ADF_{\hat{F}}^{\tau}$ be the t statistic for testing $\delta_0 = 0$ in the univariate augmented autoregression (with an intercept and a time trend)

$$\Delta\hat{F}_t = c_0 + c_1 t + \delta_0 \hat{F}_{t-1} + \delta_1 \Delta\hat{F}_{t-1} + \cdots + \delta_p \Delta\hat{F}_{t-p} + \text{error}.$$

3. If $r > 1$, let $\hat{F}_t^{\tau}$ be the residuals from a regression of $\hat{F}_t$ on a constant and a time trend. Repeat step (3) for the intercept only case with $\hat{F}_t^{\tau}$ replacing $\hat{F}_t^{c}$ to yield $\hat{Y}_t^{\tau}$ and $\hat{y}_t^{\tau}$. Denote the tests by $MQ_f^{\tau}(m)$, and $MQ_c^{\tau}(m)$.

THEOREM 3 (The Linear Trend Case): *Suppose the data are generated by* (2), (3), *and* (7), *and the assumptions of Theorem* 1 *hold. Let* W_u *and* $W_{\epsilon i}$, $i = 1, \ldots, N$ *be standard Brownian motions. The following hold as* $N, T \to \infty$:

1. *Let* p *be the order of autoregression chosen such that* $p \to \infty$ *and* $p^3/\min[N, T] \to 0$. *Let* $V_{\epsilon i}(s) = W_{\epsilon i}(s) - sW_{\epsilon i}(1)$ *be a Brownian bridge. Under the null hypothesis that* $\rho_i = 1$,

$$ADF_{\hat{e}}^{\tau}(i) \Rightarrow -\frac{1}{2}\left(\int_0^1 V_{\epsilon i}(s)^2\, ds\right)^{-1/2} \qquad (i = 1, \ldots, N).$$

2. ($r = 1$). *Let* p *be the order of autoregression chosen such that* $p \to \infty$ *and* $p^3/\min[N, T] \to 0$. *Let* $W_u^{\tau}(t) = W_u(t) - \int_0^1 (4 - 7s) W_u(s)\, ds - t\int_0^1 (12 - 6s) W_u(s)\, ds$. *When* $r = 1$ *and under the null hypothesis that* F_t *has a unit root*,

$$ADF_{\hat{F}}^{\tau} \Rightarrow \frac{\int_0^1 W_u^{\tau}(s)\, dW_u(s)}{(\int_0^1 W_u^{\tau}(s)^2\, ds)^{1/2}}.$$

3. $(r > 1)$. *Let W_m^τ be a vector of m-dimensional detrended Brownian motions. Let $\nu_*^\tau(m)$ be the smallest eigenvalue of*

$$\Phi_*^\tau = \frac{1}{2}[W_m^\tau(1)W_m^\tau(1)' - I_m]\left[\int_0^1 W_m^\tau(s)W_m^\tau(s)'\,ds\right]^{-1}.$$

(i) *Let J be the truncation point of the Bartlett kernel, chosen such that $J/\min[\sqrt{N}, \sqrt{T}] \to 0$ as $J, N, T \to \infty$. Then under the null hypothesis that F_t has m stochastic trends, $MQ_c^\tau(m) \xrightarrow{d} \nu_*^\tau(m)$.*

(ii) *Under the null hypothesis that F_t has m stochastic trends with a finite $\text{VAR}(\bar{p})$ representation, and a $\text{VAR}(p)$ is estimated with $p \geq \bar{p}$, then $MQ_f^\tau(m) \xrightarrow{d} \nu_*^\tau(m)$.*

The limiting distribution of $ADF_{\hat{F}}^\tau$ coincides with the DF for the case with a constant and a linear trend. However, as shown in the Appendix, the consequence of having to demean ΔX_{it} is that $T^{-1/2}\hat{e}_{it}$ converges to a Brownian bridge instead of a Brownian motion. The limiting distribution of the $ADF_{\hat{e}}^\tau$ is not a DF type distribution, but is proportional to the reciprocal of a Brownian bridge.

There are three important features of PANIC that are worthy of highlighting. First, the tests on the factors can be performed without knowing if the idiosyncratic errors are stationary or nonstationary. Second, the unit root test for e_{it} is valid whether e_{jt}, $j \neq i$, is I(1) or I(0), and in any event, such knowledge is not necessary. Third, the test on the idiosyncratic errors do not depend on whether F_t is I(1) or I(0). In fact, the limiting distributions of $ADF_{\hat{e}}^c(i)$ and $ADF_{\hat{e}}^\tau(i)$ do not depend on the common factors. This property is useful for constructing pooled tests.

2.4. *Pooled Tests*

A common criticism of univariate unit root tests is low power, especially when T is small. This has generated substantial interest in improving power. A popular method is to pool information across units, leading to panel unit root tests. Recent surveys of panel unit root tests can be found in Maddala and Wu (1999) and Baltagi and Kao (2001). The early test developed in Quah (1994) imposed substantial homogeneity in the cross-section dimension. Subsequent tests such as that of Levin, Lin, and Chu (2002) and Im, Pesaran, and Shin (2003) allow for heterogeneous intercepts and slopes, while maintaining the assumption of independence across units. This assumption is restrictive, and if violated, can lead to over-rejections of the null hypothesis. Banerjee, Marcellino, and Osbat (2001) argued against use of panel unit root tests because of this potential problem. O'Connell (1998) provides a GLS solution to this problem, but the approach is theoretically valid only when N is fixed.

When N also tends to infinity, as is the case under consideration, consistent estimation of the GLS transformation matrix is not a well defined concept since the sample cross-section covariance matrix will have rank T when $N > T$ even when the population covariance matrix is rank N.

If cross-section correlation can be represented by common factors, then Theorems 1 and 2 show that univariate tests for $\hat{e}_{it}$ do not depend on Brownian motions driven by the common innovations u_t asymptotically. Thus, if e_{it} is independent across i, tests based upon $\hat{e}_{it}$ are asymptotically independent across i. Consider the following:

THEOREM 4: *Suppose e_{it} is independent across i and consider testing $H_0: \rho_i = 1\ \forall i$ against $H_1: \rho_i < 1$ for some i. Let $p^c_{\hat{e}}(i)$ and $p^\tau_{\hat{e}}(i)$ be the p-values associated with $ADF^c_{\hat{e}}(i)$ and $ADF^\tau_{\hat{e}}(i)$, respectively. Then*

$$P^c_{\hat{e}} = \frac{-2\sum_{i=1}^N \log p^c_{\hat{e}}(i) - 2N}{\sqrt{4N}} \xrightarrow{d} N(0,1),$$

$$P^\tau_{\hat{e}} = \frac{-2\sum_{i=1}^N \log p^\tau_{\hat{e}}(i) - 2N}{\sqrt{4N}} \xrightarrow{d} N(0,1).$$

Under the assumption that e_{it} is independent across i, tests for $\hat{e}_{it}$ are independent across i asymptotically. The p-values are thus independent U[0,1] random variables. This implies that minus two times the logarithm of the p-value is a χ^2 random variable with two degrees of freedom. The test $-2\sum_{i=1}^N \ln p_X(i)$ was first proposed in Maddala and Wu (1999) for testing a fixed number of observed series. Choi (2001) extended the analysis to allow $N \to \infty$ by standardization. Pooling on the basis of p-values is widely used in meta analysis. It has the advantage of allowing for as much heterogeneity across units as possible. For example, it can be used even when the panel is nonbalanced. Alternatively, one can also test if the pooled coefficient estimated by regressing $\hat{e}_{it}$ on $\hat{e}_{it-1}$ is statistically different from unity. Such a pooled test would be in the spirit of Levin, Lin, and Chu (2002).

A pooled test of the idiosyncratic errors can be seen as a panel test of no cointegration, as the null hypothesis that $\rho_i = 1$ for every i holds only if no stationary combination of X_{it} can be formed. It differs from other panel cointegration tests in the literature, such as developed in Pedroni (1995), in that our framework is based on a large N, and the test is applied to $\hat{e}_{it}$ instead of X_{it}. While panel unit root tests for X_{it} are inappropriate if the data admit a factor structure, pooling of tests for $\hat{e}_{it}$ is asymptotically valid under the more plausible assumption that e_{it} is independent across i. It should be made clear that the univariate tests proposed in Theorems 1 and 3 permit weak cross-section correlation of the idiosyncratic errors. It is only in developing pooled tests that independence of the idiosyncratic errors is assumed. The independence assumption can, in principle, be relaxed by allowing the number of cross-correlated errors to be finite so that as N increases, the p-values are averaged over infinitely many units that are not cross-correlated.

3. CONSISTENCY OF $\hat{F}_t$

The asymptotic results stated in the previous section require consistent estimation of F_t and e_{it} when some, none, or all of these components are I(1). Bai and Ng (2002) considered estimation of r and showed that the squared deviations between the estimated factors and the true factors vanish, while Bai (2003) derived the asymptotic distributions for the estimated F_t and λ_i. Both studies assume the errors are all I(0). However, we need consistent estimates not just when e_{it} is I(0), but also when it is I(1).

The insight of the present analysis is that, by applying the method of principal components to the first-differenced data, it is possible to obtain consistent estimates of F_t and e_{it}, regardless of the dynamic properties of F_t and e_{it}. To sketch the idea why this is the case, assume $\beta_i = 0$. The factor model in differenced form is $\Delta X_{it} = \lambda_i' \Delta F_t + \Delta e_{it}$. Clearly, differencing removes the fixed effect c_i. This is desirable because a consistent estimate of it cannot be obtained when e_{it} is I(1). Now if e_{it} is I(1), $\Delta e_{it} = z_{it}$ will be I(0). Under Assumption C, z_{it} has weak cross-section and serial correlation. Consistent estimates of ΔF_t can thus be obtained. If e_{it} is I(0), Δe_{it}, although over-differenced, is still stationary and weakly correlated. Thus, consistent estimation of ΔF_t can once again be shown. We summarize these arguments in the following lemma.

LEMMA 1: *Let f_t be defined by* (5). *Consider estimation of* (6) *by the method of principal components and suppose Assumptions* A–E *hold. Then there exists an H with rank r such that as $N, T \to \infty$,*

(a) $\min\{N, T\} T^{-1} \sum_{t=2}^{T} \|\hat{f}_t - H f_t\|^2 = O_p(1)$,

(b) $\min\{\sqrt{N}, T\}(\hat{f}_t - H f_t) = O_p(1)$, *for each given t*,

(c) $\min\{\sqrt{T}, N\}(\hat{\lambda}_i - H'^{-1}\lambda_i) = O_p(1)$, *for each given i*.

The results also hold when f_t is defined by (9) *and* (8) *is estimated.*

As is well known in factor analysis, λ_i and f_t are not directly identifiable. Therefore, when assessing the properties of the estimates, we can only consider the difference in the space spanned by $\hat{f}_t$ and f_t, and likewise between $\hat{\lambda}_i$ and λ_i. The matrix H is defined (in the Appendix) such that $H f_t$ is the projection of $\hat{f}_t$ on the space spanned by the factors, f_t. Result (a) is proved in Bai and Ng (2002), while (b) and (c) are proved in Bai (2003). It should be remarked that when e_{it} is I(0), estimation using the data in level form will give a direct and consistent estimate on F_t. Although these estimates could be more efficient than the ones based upon first differencing, they are not consistent when e_{it} is I(1).

In Pesaran and Smith (1995), it was shown that spurious correlations between two I(1) variables do not arise in cross-section regressions estimated with time averaged data under the assumption of strictly exogenous regressors, i.i.d. errors, and T fixed. Phillips and Moon (1999) showed that an average long-run relation, defined from long-run covariance matrices of a panel of I(1) variables, can be identified when N and T are both large. Lemma 1 shows that the individual relations (not just the average) can be consistently estimated under a much wider range of conditions: the regressors are unobserved, they can be I(1) or I(0), and the individual regressions may or may not be spurious.

Although λ_i and f_t can be consistently estimated, the series we are interested in testing are $\hat{F}_t = \sum_{s=2}^t \hat{f}_s$ and $\hat{e}_{it} = \sum_{s=2}^t \hat{z}_{it}$. Thus, we need to show that given estimates of f_t and z_{it}, $\hat{F}_t$ and $\hat{e}_{it}$ are consistent for F_t and e_{it}, respectively.

LEMMA 2: *Under the assumptions of Lemma* 1,

$$\max_{1\le t\le T} \frac{1}{\sqrt{T}} \left\| \sum_{s=2}^{t} (\hat{f}_s - Hf_s) \right\| = O_p(N^{-1/2}) + O_p(T^{-3/4}).$$

The lemma says that the cumulative sum of $\hat{f}_t$ is uniformly close to the cumulative sum of f_t provided $N, T \to \infty$.[8] Because $\hat{F}_t = \sum_{s=2}^t \hat{f}_s$ and $H\sum_{s=2}^t f_t = H\sum_{s=2}^t \Delta F_s = HF_t - HF_1$, Lemma 2 can be stated as

$$\max_{1\le t\le T} \frac{1}{\sqrt{T}} \|\hat{F}_t - HF_t + HF_1\| = O_p(N^{-1/2}) + O_p(T^{-3/4}). \tag{10}$$

Since a location shift does not change the nonstationarity property of a series, testing the demeaned process $\hat{F}_t - \overline{\hat{F}}$ is asymptotically the same as testing $H(F_t - \overline{F})$. This result is instrumental in obtaining the limiting distributions of unit root tests for F_t. It would seem that for testing $\hat{e}_{it}$, this result may not be sufficient since $\hat{e}_{it}$ also depends on $\hat{\lambda}_i$. But as shown in the Appendix, we only require $(\hat{\lambda}_i - H'^{-1}\lambda_i)$ to be $o_p(1)$ for unit root tests on $\hat{e}_{it}$ to yield the same inference as testing e_{it}, and by Lemma 1(c), this holds provided N and T tend to infinity. Thus, the conditions for valid testing of F_t and e_{it} using $\hat{F}_t$ and $\hat{e}_{it}$ are the same.

An implication of Lemma 2 is that $T^{-2}\sum_{t=2}^T \hat{F}_t\hat{F}_t' - T^{-2}H(\sum_{t=2}^T F_tF_t')H' \xrightarrow{p} 0$. That is, the sample variation generated by $\hat{F}_t$ is the same order as F_t. If $T^{-2}\sum_{t=2}^T F_tF_t'$ has r_1 nondegenerate eigenvalues, $T^{-2}\sum_{t=2}^T \hat{F}_t\hat{F}_t'$ also has r_1 nondegenerate eigenvalues. Thus if F_t has r_1 common trends, $\hat{F}_t$ will also have r_1 common trends. This result is instrumental in the development of the MQ_c and MQ_f tests.

[8]The $O_p(T^{-3/4})$ can be replaced by $O_p(\log T/T)$ if the moment generating function of f_t exists (i.e., if $Ee^{\tau\|f_t\|} \le M$ for all t and for some $\tau > 0$).

Uniform convergence of the factor estimates in large panels was proved in Stock and Watson (2002) under the assumption that $N \gg T^2$ and that F_t and e_{it} are stationary. However, our analysis provides a more general uniform consistency result as a by-product. Upon multiplying (A.2) by $\sqrt{T}$, we have

$$(11) \qquad \max_{1\le t\le T} \|\hat{F}_t - HF_t + HF_1\| = O_p(T^{1/2}N^{-1/2}) + O_p(T^{-1/4}).$$

As stated in (11), $\hat{F}_t$ is uniformly consistent for HF_t (up to a shift factor HF_1) provided $T/N \to 0$ as $N, T \to \infty$. This result is quite remarkable in that the common stochastic trends can be consistently estimated by the method of principal components, up to a rotation and a shift in level, without knowing whether F_t or e_{it} is I(0) or I(1). This means that even if each cross-section equation is a spurious regression, the common stochastic trends are well defined and can be consistently estimated, if they exist. This is not possible within the framework of traditional time-series analysis, in which N is fixed.

The result that when N and T are large, the space spanned by the common factors can be consistently estimated under very general conditions is not merely a strong result of theoretical interest. It is also of practical interest because it opens the possibility of testing other properties of F_t using $\hat{F}_t$. For example, ARCH and long memory effects can be assessed, parameter instability tests can be devised, and the relative importance of the common and the idiosyncratic components can be evaluated even when neither is observed. Because Lemma 2 is potentially useful in contexts other than unit root testing, we stated it as a primary result. It should be made clear that uniform consistency is not necessary for PANIC, and thus we do not require $T/N \to 0$, though our results will hold under these stronger conditions. For PANIC to be valid, only Lemmas 1 and 2 are necessary.

4. MONTE CARLO SIMULATIONS

We begin by using a model with one factor to show that $\hat{F}_t$ constructed as $\sum_{s=2}^{t} \hat{f}_t$ is robust to different stationarity assumptions about e_{it}, where $\hat{f}_t$ is estimated from first-differenced data. We generate F_t as an independent random walk of $N(0,1)$ errors with $T = 100$, and λ_i is i.i.d. $N(1,1)$. Data are generated according to $X_{it} = \lambda_i F_t + e_{it}$. We then construct $\hat{F}_t$ as discussed in Section 2 for the intercept only model. In practice, a comparison of F_t with $\hat{F}_t$ cannot be made because the former is unobservable. But F_t is known in simulations. Thus, for the sake of illustration, we compare the fitted values from the regression $F_t = a + b\hat{F}_t + \text{error}$ with F_t. An implication of Lemma 2 is that this fitted value (which we will continue to call $\hat{F}_t$) should be increasingly close to F_t as N increases. On the other hand, estimation using the data in levels will not have this consistency property.

For the case when e_{it} is I(1), we simulate a random walk driven by i.i.d. $N(0, 1)$ errors for $N = 20$, 50, and 100, respectively. We then estimate the factors using (i) differenced data, and (ii) the data in level form. Figures 1(b), (c), and (d) display the true factor process F_t along with $\hat{F}_t$. Evidently, $\hat{F}_t$ gets closer to F_t as N increases if the data are differenced. In fact $\hat{F}_t$ is close to the true process even when $N = 20$. On the other hand, when the method of principal components is applied to levels of the same data, all the estimated series are far from the true series, showing that estimation using the data in levels is not consistent when e_{it} is I(1). We next assume the idiosyncratic errors are all I(0) by drawing e_{it} from an i.i.d. $N(0, 1)$ distribution. Figure 2 illustrates that even though the data are over-differenced, the estimates are very precise. In this case, both the level and differenced methods give almost identical estimates.

We now use simulations to illustrate the finite sample properties of the proposed tests. Throughout, the number of replications is 5000. In theory, r is not known. We showed in Bai and Ng (2002) that the number of factors in stationary data can be consistently determined by information criteria (PC_p) if the penalty on an additional factor is specified as a function of both N and T. In the present context, we can consistently estimate r from the first-differenced

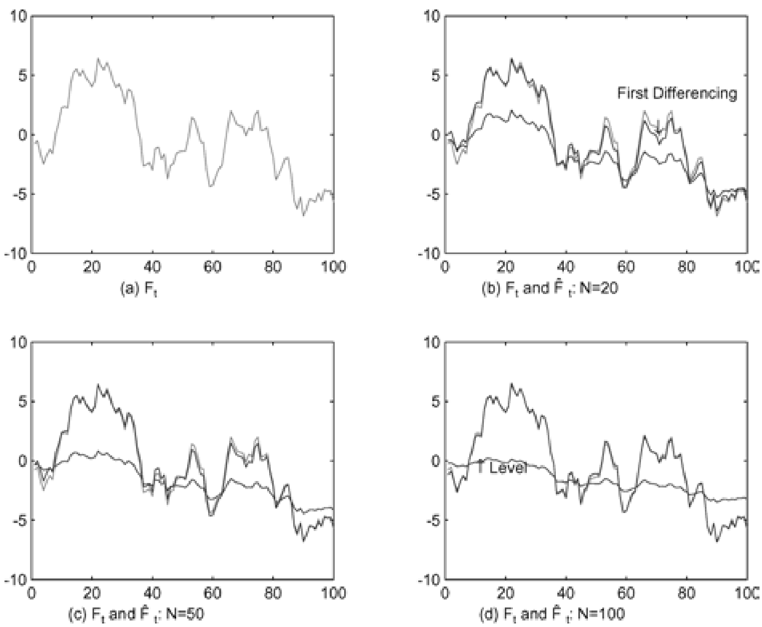

FIGURE 1.—True and estimated F_t when e_{it} is I(1).

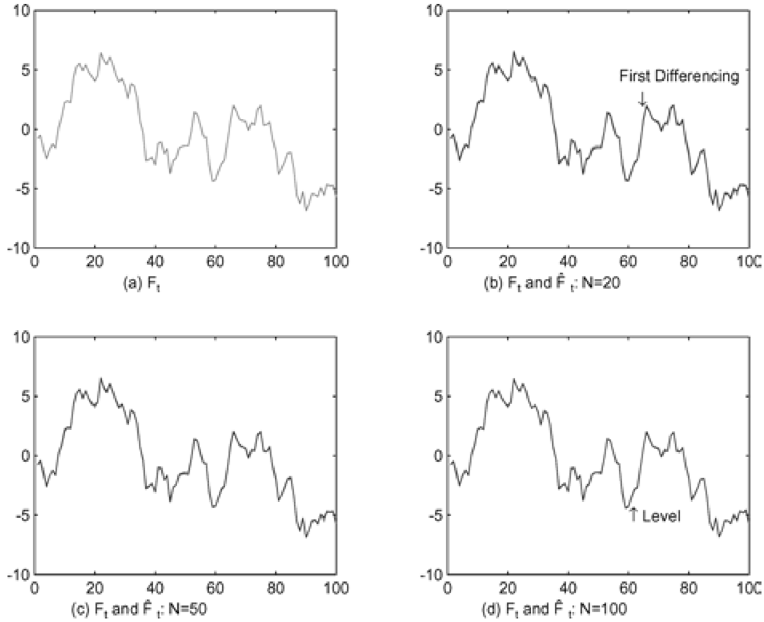

FIGURE 2.—True and estimated F_t when e_{it} is I(0).

data. In the simulations, the following is used:

$$\hat{r} = \underset{k=0,\dots,k_{\max}}{\arg\min}\, IC_1(k), \quad \text{where}$$

$$IC_1(k) = \log \hat{\sigma}^2(k) + k \log\left(\frac{NT}{N+T}\right)\frac{N+T}{NT},$$

where $\hat{\sigma}^2(k) = N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{z}_{it}^2$, $\hat{z}_{it}$ are the estimated residuals from principal components estimation of the first-differenced data, and $k_{\max} = 6$. In all the configurations considered (up to 3 true factors), the criterion always selects $\hat{r} = r$.[9]

4.1. *The Case* $r = 1$

We simulate data using $X_{it} = \lambda_i F_t + e_{it}$, with $e_{it} = \rho e_{it-1} + \epsilon_{it}$, and $F_t = \alpha F_{t-1} + u_t$, with $\lambda_i \sim N(0,1)$, $\epsilon_{it} \sim N(0,1)$, and $u_t \sim N(0, \sigma_F^2)$. We consider three values of σ_F^2 with the importance of the common component increasing in the value of σ_F^2. In the simulations, ρ_i is the same across i. We also consider

[9]Using IC_2 in Bai and Ng (2002), $P(\hat{r} = r)$ is sometimes .98. The choice of a penalty that satisfies the conditions of Bai and Ng (2002) is important.

fourteen pairs of (ρ_i, α). When $\rho_i = 1$ but $\alpha < 1$, the errors are nonstationary but the factors are stationary. When $\alpha = 1$ but $\rho_i < 1$, the factors are unit root processes but the errors are stationary.

We report results for $T = 100$, and $N = 40,\ 100$ in Table II. The column labeled $\hat{F}$ is the rejection rate of the *ADF* test applied to the estimated common factor. The remaining three columns are the average rejection rates, where the average is taken across N units over 5000 trials. Results for a particular i are

TABLE IIA

REJECTION RATES FOR THE NULL HYPOTHESIS OF A UNIT ROOT, INTERCEPT ONLY, $r = 1$

		$\sigma^F = \sqrt{10}$					$\sigma^F = 1$					$\sigma^F = \sqrt{.5}$				
ρ_i	α	X	$\hat{F}$	$\hat{e}$	P_X^c	$P_{\hat{e}}^c$	X	$\hat{F}$	$\hat{e}$	P_X^c	$P_{\hat{e}}^c$	X	$\hat{F}$	$\hat{e}$	P_X^c	$P_{\hat{e}}^c$
							$T = 100, N = 40$									
1.00	.00	.18	.96	.06	.90	.05	.07	.53	.06	.21	.06	.07	.33	.06	.16	.05
1.00	.50	.25	.92	.06	.97	.06	.09	.64	.06	.39	.06	.08	.47	.05	.25	.05
1.00	.80	.23	.57	.05	.91	.05	.10	.47	.06	.54	.06	.08	.40	.06	.36	.05
1.00	.90	.15	.27	.06	.72	.06	.09	.25	.05	.47	.06	.08	.23	.06	.34	.06
1.00	.95	.10	.13	.06	.51	.05	.08	.12	.05	.36	.05	.07	.12	.05	.27	.05
.00	1.00	.11	.07	.44	.42	1.00	.21	.07	.43	.68	1.00	.26	.06	.43	.81	1.00
.50	1.00	.13	.07	.58	.45	1.00	.25	.07	.58	.77	1.00	.32	.06	.58	.88	1.00
.80	1.00	.13	.07	.58	.46	1.00	.22	.07	.58	.75	1.00	.26	.07	.58	.87	1.00
.90	1.00	.09	.07	.43	.41	1.00	.14	.07	.43	.67	1.00	.16	.07	.43	.79	1.00
.95	1.00	.08	.06	.25	.37	1.00	.10	.07	.25	.55	1.00	.10	.06	.25	.63	1.00
1.00	1.00	.07	.07	.06	.32	.06	.07	.07	.06	.26	.06	.07	.07	.05	.23	.05
.50	.80	.68	.59	.67	1.00	1.00	.80	.60	.67	1.00	1.00	.84	.62	.67	1.00	1.00
.80	.50	.82	.96	.64	1.00	1.00	.69	.94	.64	1.00	1.00	.65	.93	.64	1.00	1.00
.00	.90	.35	.28	.57	.91	1.00	.49	.27	.57	.99	1.00	.57	.27	.57	1.00	1.00
.90	.00	.54	1.00	.46	1.00	1.00	.31	.94	.46	1.00	1.00	.28	.85	.46	1.00	1.00
							$T = 100, N = 100$									
1.00	.00	.18	.99	.06	.98	.06	.07	.78	.06	.35	.06	.07	.58	.05	.25	.06
1.00	.50	.25	.95	.06	.99	.06	.09	.82	.06	.62	.06	.08	.70	.06	.43	.06
1.00	.80	.23	.59	.06	.95	.06	.10	.53	.05	.75	.05	.08	.48	.06	.58	.05
1.00	.90	.15	.27	.06	.81	.06	.09	.25	.05	.66	.05	.08	.24	.05	.54	.05
1.00	.95	.10	.13	.05	.59	.05	.08	.13	.06	.51	.05	.07	.13	.05	.44	.05
.00	1.00	.11	.06	.44	.46	1.00	.21	.07	.44	.75	1.00	.27	.07	.43	.85	1.00
.50	1.00	.13	.07	.58	.50	1.00	.25	.07	.58	.81	1.00	.32	.07	.58	.92	1.00
.80	1.00	.12	.07	.58	.50	1.00	.22	.07	.58	.81	1.00	.27	.07	.58	.91	1.00
.90	1.00	.09	.06	.43	.47	1.00	.14	.06	.43	.74	1.00	.16	.07	.43	.86	1.00
.95	1.00	.08	.06	.25	.43	1.00	.10	.07	.25	.64	1.00	.10	.06	.25	.74	1.00
1.00	1.00	.07	.07	.06	.37	.05	.07	.07	.06	.35	.06	.07	.06	.06	.33	.06
.50	.80	.68	.59	.67	1.00	1.00	.80	.61	.68	1.00	1.00	.84	.60	.67	1.00	1.00
.80	.50	.82	.96	.64	1.00	1.00	.69	.96	.64	1.00	1.00	.65	.95	.64	1.00	1.00
.00	.90	.35	.27	.57	.93	1.00	.49	.28	.57	1.00	1.00	.57	.26	.57	1.00	1.00
.90	.00	.54	1.00	.46	1.00	1.00	.31	.98	.46	1.00	1.00	.29	.96	.46	1.00	1.00

Note: The data are generated as $e_{it} = \rho_i e_{it-1} + \epsilon_{it}$ and $F_t = \alpha F_{t-1} + u_t$. Columns under X and $\hat{e}$ are reject rates of the *ADF*. P^c and P^τ are rejection rates of the pooled tests.

TABLE IIB

REJECTION RATES FOR THE NULL HYPOTHESIS OF A UNIT ROOT, LINEAR TREND MODEL, $r = 1$

		$\sigma^F = \sqrt{10}$					$\sigma^F = 1$					$\sigma^F = \sqrt{.5}$				
ρ_i	α	X	$\hat{F}$	$\hat{e}$	P_X^τ	$P_{\hat{e}}^\tau$	X	$\hat{F}$	$\hat{e}$	P_X^τ	$P_{\hat{e}}^\tau$	X	$\hat{F}$	$\hat{e}$	P_X^τ	$P_{\hat{e}}^\tau$
$T = 100, N = 40$																
1.00	.00	.22	.95	.05	.94	.07	.08	.63	.05	.32	.07	.07	.45	.05	.24	.06
1.00	.50	.28	.81	.05	.95	.07	.10	.65	.05	.54	.06	.08	.51	.05	.37	.06
1.00	.80	.21	.38	.05	.82	.06	.11	.35	.05	.61	.06	.09	.32	.05	.47	.06
1.00	.90	.13	.17	.05	.59	.06	.09	.17	.05	.49	.06	.08	.16	.05	.42	.06
1.00	.95	.09	.10	.05	.43	.06	.08	.10	.05	.38	.06	.08	.10	.05	.33	.06
.00	1.00	.12	.07	.35	.45	1.00	.24	.07	.34	.76	1.00	.29	.07	.34	.87	1.00
.50	1.00	.14	.06	.48	.48	1.00	.27	.07	.48	.82	1.00	.33	.07	.48	.93	1.00
.80	1.00	.12	.07	.37	.48	1.00	.19	.07	.36	.79	1.00	.23	.08	.36	.89	1.00
.90	1.00	.09	.07	.20	.42	1.00	.12	.08	.20	.65	1.00	.13	.07	.20	.78	1.00
.95	1.00	.08	.07	.10	.37	.83	.09	.08	.10	.50	.83	.09	.08	.10	.55	.83
1.00	1.00	.07	.07	.05	.34	.06	.07	.08	.05	.31	.06	.07	.07	.05	.28	.07
.50	.80	.49	.39	.53	.97	1.00	.62	.40	.53	1.00	1.00	.67	.42	.53	1.00	1.00
.80	.50	.65	.85	.38	1.00	1.00	.48	.82	.38	1.00	1.00	.45	.79	.38	1.00	1.00
.00	.90	.25	.18	.41	.77	1.00	.39	.18	.41	.97	1.00	.46	.17	.41	.99	1.00
.90	.00	.42	.97	.20	1.00	1.00	.20	.85	.20	1.00	1.00	.19	.71	.20	1.00	1.00
$T = 100, N = 100$																
1.00	.00	.22	.96	.05	.98	.07	.08	.84	.05	.51	.06	.07	.69	.05	.40	.07
1.00	.50	.28	.83	.05	.98	.06	.10	.76	.05	.75	.07	.08	.69	.05	.61	.06
1.00	.80	.21	.39	.05	.88	.06	.11	.37	.05	.79	.06	.09	.36	.05	.70	.07
1.00	.90	.13	.17	.05	.67	.06	.09	.16	.05	.65	.06	.08	.17	.05	.60	.06
1.00	.95	.09	.10	.05	.50	.06	.08	.10	.05	.49	.06	.08	.11	.05	.49	.06
.00	1.00	.13	.07	.35	.49	1.00	.24	.07	.35	.81	1.00	.30	.07	.34	.91	1.00
.50	1.00	.14	.07	.48	.52	1.00	.27	.07	.48	.87	1.00	.34	.07	.48	.96	1.00
.80	1.00	.12	.07	.36	.52	1.00	.19	.07	.36	.83	1.00	.23	.07	.37	.94	1.00
.90	1.00	.09	.08	.20	.47	1.00	.12	.07	.20	.74	1.00	.13	.07	.20	.86	1.00
.95	1.00	.08	.07	.10	.42	.99	.09	.07	.10	.60	.99	.09	.07	.10	.69	.99
1.00	1.00	.07	.07	.05	.40	.06	.07	.07	.05	.41	.06	.07	.07	.05	.41	.06
.50	.80	.49	.40	.53	.98	1.00	.62	.41	.53	1.00	1.00	.67	.41	.53	1.00	1.00
.80	.50	.64	.84	.38	1.00	1.00	.48	.84	.38	1.00	1.00	.45	.82	.38	1.00	1.00
.00	.90	.24	.17	.41	.81	1.00	.38	.18	.41	.99	1.00	.46	.17	.41	1.00	1.00
.90	.00	.42	.97	.20	1.00	1.00	.20	.93	.20	1.00	1.00	.19	.88	.20	1.00	1.00

Note: The data are generated as $e_{it} = \rho_i e_{it-1} + \epsilon_{it}$ and $F_t = \alpha F_{t-1} + u_t$. Columns under X and $\hat{e}$ are reject rates of the *ADF*. P^c and P^τ are rejection rates of the pooled tests.

similar. The augmented autoregressions have $p = 4[\min[N, T]/100]^{1/4}$ lags. Critical values at the 5% level were used.

The ADF test applied to X_{it} should have a rejection rate of .05 when $\alpha = 1$ or $\rho = 1$. In finite samples, this is true only when $\rho = 1$ and σ^F is small. When $\sigma^F = 10$ and $\alpha = .5$, for example, the ADF test rejects a unit root in X_{it} with probability around .25 in the intercept model, and .28 in the linear trend model, even though $\rho = 1$. As noted earlier, testing for a unit root in X_{it} when it has components with different degrees of integration is difficult because of the negative moving average component in ΔX_{it}. Because our procedure separately tests these components, our tests are also less sensitive to the choice of truncation lag compared to conventional testing of X_{it}.

Turning now to $\hat{F}_t$, the rejection rate is close to the nominal size of .05 when α is 1. At other values of α, the rejection rates are comparable to the power of other unit root tests that are based on least squares detrending. The $ADF_{\hat{e}}(i)$ has similar properties, with rejection rates around 5% when $\rho_i = 1$. These results suggest that the error in estimating F_t is small even when $N=40$. Indeed, the results for $N = 100$ are similar except for small values of α.

Results for the pooled tests are also reported in Tables IIA, B.[10] Both $P^c_{\hat{e}}$ and $P^\tau_{\hat{e}}$ correctly reject the null hypothesis when e_{it} is in fact stationary. When each of the e_{it} is nonstationary, the rejection rates roughly equal the nominal size of .05. Consider the standard pooled tests for X_{it} (see column under P^c_X, P^τ_X). When $\rho = .5$ and $\alpha = 1$, all N series are nonstationary in view of the common stochastic trend. The standard pooled test should have a rejection rate close to .05. However, the rejection rate ranges from .45 to .88 depending on σ^F. Consider also $(\rho_i, \alpha) = (1, 0)$. The common factor is i.i.d.; the pooled test has a rejection rate of .16 when σ^F is small and deteriorates to .90 when σ^F is large. These results are consistent with the findings of O'Connell (1998) that cross section correlation leads the standard pooled test to over-reject the null hypothesis.

4.2. $r > 1$

In cases of multiple factors, we generate the I(1) factors as simple random walks and the stationary factors as AR(1) processes with coefficient α. We continue to assume that e_{it} is AR(1) with parameter ρ_i. The factor loadings are taken from an $N \times r$ matrix of $N(0, 1)$ variables. We consider three cases of σ_F as in the previous section. As the results in Table III illustrate quite well the consequence of increasing N from 40 to 100, we simply report results for $N = 40$ to conserve space. Although we only present results for $r = 3$, many additional configurations were considered and are available on request.

We begin with results for testing X_{it} and $\hat{e}_{it}$. With $r = 3$, we can vary r_1 from 0 to 3 to assess the case of none, one, two, and three common trends. Regardless

[10]The p-values required to construct the pooled tests are obtained as follows. We first simulate the asymptotic distributions reported in Theorems 1 and 2 by using partial sums of 500 $N(0, 1)$ errors to approximate the standard Brownian motion in each of the 10,000 replications. A look-up table is then constructed to map 300 points on the asymptotic distributions to the corresponding p-values. In particular, 100 points are used to approximate the upper tail, 100 to approximate the lower tail, and 100 points for the middle part of the asymptotic distributions. The p-values match Table IV of MacKinnon (1994) very well , whenever they are available. These look-up tables are available from the authors.

TABLE IIIA

REJECTION RATES, UNIVARIATE AND POOLED UNIT ROOT TESTS, $r=3$, $T=100$, $N=40$, INTERCEPT MODEL

				$\sigma^F=\sqrt{10}$				$\sigma^F=1$				$\sigma^F=\sqrt{.5}$			
r	r_1	ρ_i	α	X	$\hat{e}$	P^c_X	$P^c_{\hat{e}}$	X	$\hat{e}$	P^c_X	$P^c_{\hat{e}}$	X	$\hat{e}$	P^c_X	$P^c_{\hat{e}}$
3	3	.00	–	.07	.27	.30	.97	.07	.26	.32	.96	.08	.25	.35	.95
3	3	.50	–	.07	.47	.30	1.00	.08	.46	.37	1.00	.10	.46	.47	1.00
3	3	.80	–	.07	.54	.32	1.00	.09	.54	.44	1.00	.12	.53	.56	1.00
3	3	.90	–	.07	.42	.31	1.00	.08	.42	.42	1.00	.10	.42	.53	1.00
3	3	1.00	–	.06	.06	.29	.06	.07	.06	.26	.06	.07	.05	.24	.06
3	0	.00	.00	1.00	.59	1.00	1.00	1.00	.59	1.00	1.00	1.00	.59	1.00	1.00
3	0	.50	.50	.96	.73	1.00	1.00	.96	.72	1.00	1.00	.96	.73	1.00	1.00
3	0	.80	.50	.93	.72	1.00	1.00	.79	.72	1.00	1.00	.74	.73	1.00	1.00
3	0	.00	.50	.96	.59	1.00	1.00	.97	.60	1.00	1.00	.98	.60	1.00	1.00
3	0	.90	.00	.82	.52	1.00	1.00	.40	.52	1.00	1.00	.33	.52	1.00	1.00
3	0	1.00	.00	.39	.06	1.00	.06	.09	.05	.43	.05	.07	.06	.24	.06
3	1	.00	.00	.25	.44	.77	1.00	.29	.43	.85	1.00	.32	.43	.91	1.00
3	1	.50	.50	.30	.62	.83	1.00	.35	.61	.91	1.00	.40	.61	.95	1.00
3	1	.80	.50	.29	.65	.82	1.00	.31	.65	.90	1.00	.33	.65	.94	1.00
3	1	.00	.50	.29	.44	.83	1.00	.33	.44	.89	1.00	.36	.43	.92	1.00
3	1	.90	.00	.21	.48	.70	1.00	.18	.48	.75	1.00	.18	.48	.82	1.00
3	1	1.00	.00	.14	.06	.58	.06	.07	.06	.32	.06	.07	.06	.24	.05
3	2	.00	.00	.09	.34	.38	.99	.11	.33	.46	.99	.13	.32	.56	.98
3	2	.50	.50	.11	.53	.47	1.00	.15	.53	.59	1.00	.18	.53	.70	1.00
3	2	.80	.50	.11	.59	.47	1.00	.15	.59	.63	1.00	.18	.59	.74	1.00
3	2	.00	.50	.11	.34	.46	.99	.13	.34	.54	.99	.15	.33	.60	.98
3	2	.90	.00	.09	.45	.40	1.00	.11	.45	.52	1.00	.12	.44	.63	1.00
3	2	1.00	.00	.08	.05	.36	.06	.07	.05	.28	.05	.07	.06	.24	.06

of the number of common trends, testing X_{it} remains imprecise frequently. For example, if $r_1=0$ and $\rho_i=1$, there is a unit root in X_{it} and the ADF test should reject roughly with probability .05. Instead, the rejection rates are .39 for the intercept model, and .48 for the linear trend model. The rejection rates for $\hat{e}$ mirror the results for $r=1$, showing that the behavior of $ADF_{\hat{e}}(i)$ is not sensitive to the true number of factors in the data. With one random walk factor and two stationary AR(1) factors, the $ADF_{\hat{e}}(i)$ has a rejection rate of .65 when $(\rho_i,\alpha)=(.8,.5)$. This is almost the same rejection rate as when there was only one stationary factor.

The simulated critical values of the $MQ^{c,\tau}$ tests are extremely close to those reported in Stock and Watson for Q_c and Q_f. We conjecture that their tests are also valid in the present context. Indeed, because $\hat{F}_t$ consistently estimates the space spanned by F_t, we conjecture that other tests that assume F_t is observed remain valid when F_t is estimated using our proposed methodology. To investigate this and for the sake of comparison, we also consider the trace test of

TABLE IIIB

REJECTION RATES, UNIVARIATE AND POOLED UNIT ROOT TESTS, $r = 3$, $T = 100$, $N = 40$, LINEAR TREND MODEL

				$\sigma^F = \sqrt{10}$				$\sigma^F = 1$				$\sigma^F = \sqrt{.5}$			
r	r_1	ρ_i	α	X	$\hat{e}$	P_X^τ	$P_{\hat{e}}^\tau$	X	$\hat{e}$	P_X^τ	$P_{\hat{e}}^\tau$	X	$\hat{e}$	P_X^τ	$P_{\hat{e}}^\tau$
3	3	.00	–	.07	.23	.32	.97	.08	.22	.36	.96	.09	.21	.41	.95
3	3	.50	–	.07	.40	.33	1.00	.09	.40	.44	1.00	.12	.40	.56	1.00
3	3	.80	–	.08	.35	.36	1.00	.10	.35	.50	1.00	.12	.35	.62	1.00
3	3	.90	–	.07	.20	.35	1.00	.09	.20	.46	1.00	.10	.20	.56	1.00
3	3	1.00	–	.07	.05	.32	.07	.07	.05	.31	.06	.07	.05	.29	.06
3	0	.00	.00	.98	.45	1.00	1.00	.98	.45	1.00	1.00	.98	.45	1.00	1.00
3	0	.50	.50	.86	.56	1.00	1.00	.85	.56	1.00	1.00	.85	.56	1.00	1.00
3	0	.80	.50	.79	.41	1.00	1.00	.60	.41	1.00	1.00	.53	.41	1.00	1.00
3	0	.00	.50	.86	.45	1.00	1.00	.89	.45	1.00	1.00	.90	.45	1.00	1.00
3	0	.90	.00	.71	.21	1.00	1.00	. 28	.21	1.00	1.00	.22	.21	1.00	1.00
3	0	1.00	.00	.48	.05	1.00	.07	.11	.05	.63	.06	.08	.05	.37	.07
3	1	.00	.00	.28	.35	.82	1.00	.32	.35	.91	1.00	.37	.34	.95	1.00
3	1	.50	.50	.31	.50	.86	1.00	.37	.50	.94	1.00	.41	.50	.97	1.00
3	1	.80	.50	.30	.39	.86	1.00	.29	.39	.92	1.00	.30	.39	.96	1.00
3	1	.00	.50	.30	.35	.86	1.00	.35	.35	.93	1.00	.38	.35	.95	1.00
3	1	.90	.00	.21	.21	.74	1.00	.15	.21	.76	1.00	.15	.21	.82	1.00
3	1	1.00	.00	.16	.05	.65	.06	.08	.05	.39	.06	.07	.05	.31	.07
3	2	.00	.00	.10	.28	.43	.99	.13	.27	.52	.99	.16	.27	.64	.99
3	2	.50	.50	.12	.45	.53	1.00	.17	.45	.67	1.00	.21	.44	.79	1.00
3	2	.80	.50	.13	.37	.53	1.00	.16	.37	.71	1.00	.18	.37	.81	1.00
3	2	.00	.50	.12	.28	.51	.99	.15	.28	.59	.99	.18	.27	.69	.98
3	2	.90	.00	.09	.20	.44	1.00	.11	.20	.56	1.00	.12	.20	.65	1.00
3	2	1.00	.00	.09	.05	.40	.06	.07	.05	.33	.07	.07	.05	.30	0.07

Note: The idiosyncratic errors are generated as $e_{it} = \rho_i e_{i,t-1} + \epsilon_{it}$, $\epsilon_{it} \sim N(0, 1)$, ρ_i identical across i. The r_1 nonstationary factors are generated as $\Delta F_t = u_t$. The r_0 stationary factors are generated as $F_t = \alpha F_{t-1} + u_t$, $u_t \sim N(0, 1)$. Columns under X and $\hat{e}$ are reject rates of the *ADF*. P^c and P^τ are rejection rates of the pooled tests.

Johansen (1988),[11] and the information criteria developed by Aznar and Salvador (2002). The trace test uses the residuals from projections of $\hat{Y}_t$ and $\hat{Y}_{t-1}$ on p lags of $\Delta\hat{Y}_t$. It thus uses a slightly different way of controlling for serial correlation than the MQ_f. The information criterion, which we denote by ASIC, determines r and p simultaneously. Both statistics are designed to test the number of cointegrating vectors assuming p is finite. In contrast, $MQ_f(m)$ is a test for the number of common trends.

In the simulations, we use the Bartlett kernel with $J = 4\,\text{ceil}[\min[N, T]/100]^{1/4}$ for the MQ_c. For the data generating processes considered, the test is not very sensitive to the choice of J. However, both Johansen's trace and our MQ_f tests are sensitive to the choice of the order of the VAR. To be precise,

[11]Results for his max λ test are similar.

they lose power when too many lags are selected. A data dependent method for selecting the VAR order is thus important. Since a by-product of the ASIC is an estimate of p, we use the same $\hat{p}$ in both the MQ_f and the trace test. In simulations, the results are somewhat better than when the BIC was used to determine the lag length of the appropriate VARs, and both dominate fixing p at, say, $4[T/100]^{1/4}$.

Table IV reports the probability of selecting the true number of common stochastic trends. We used the 5% critical values in successive testing. The two MQ tests have rather similar properties. When all factors are I(1), both tests select $r_1 = 3$ with probability around .95. When all factors are I(0), the tests correctly select $r_1 = 0$ with probability one. When some factors are I(1) and some are I(0), the tests still maintain a very high accuracy rate (over .9).

As discussed earlier, the MQ statistics involve successive testing of a sequence of hypotheses, much like the trace and maximal eigenvalue tests for

TABLE IVA

PROBABILITY OF SELECTING THE CORRECT NUMBER OF COMMON STOCHASTIC TRENDS, $T = 100$, $N = 40$, INTERCEPT MODEL

				$\sigma^F = \sqrt{10}$				$\sigma^F = 1$				$\sigma^F = \sqrt{.5}$			
r	r_1	ρ_i	α	MQ_c^c	MQ_f^c	asic	trace	MQ_c^c	MQ_f^c	asic	trace	MQ_c^c	MQ_f^c	asic	trace
3	3	.00	–	.95	.95	.81	.82	.94	.93	.83	.83	.90	.90	.84	.81
3	3	.50	–	.96	.96	.82	.83	.95	.95	.81	.83	.93	.94	.83	.82
3	3	.80	–	.96	.97	.80	.83	.95	.95	.82	.83	.96	.95	.82	.83
3	3	.90	–	.96	.96	.80	.82	.96	.95	.81	.84	.96	.95	.85	.85
3	3	1.00	–	.96	.96	.80	.83	.96	.96	.81	.82	.96	.96	.81	.83
3	0	.00	.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	0	.50	.50	1.00	1.00	.89	1.00	1.00	1.00	.90	1.00	1.00	1.00	.90	1.00
3	0	.80	.50	1.00	1.00	.88	1.00	1.00	1.00	.88	1.00	1.00	1.00	.89	1.00
3	0	.00	.50	1.00	1.00	.89	1.00	1.00	1.00	.91	1.00	1.00	1.00	.93	1.00
3	0	.90	.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	0	1.00	.00	1.00	1.00	1.00	1.00	.99	.93	.93	.81	.92	.69	.70	.48
3	1	.00	.00	.92	.92	.91	.95	.91	.92	.91	.94	.90	.91	.90	.94
3	1	.50	.50	.92	.92	.79	.95	.91	.90	.79	.94	.91	.90	.79	.94
3	1	.80	.50	.92	.92	.78	.95	.92	.92	.78	.95	.91	.90	.77	.94
3	1	.00	.50	.92	.92	.79	.95	.92	.90	.81	.94	.90	.88	.83	.92
3	1	.90	.00	.92	.92	.91	.95	.92	.92	.92	.95	.91	.91	.91	.94
3	1	1.00	.00	.92	.92	.91	.95	.92	.87	.82	.87	.89	.75	.67	.74
3	2	.00	.00	.91	.90	.92	.91	.89	.89	.92	.90	.87	.87	.91	.89
3	2	.50	.50	.92	.92	.81	.91	.91	.91	.81	.91	.90	.89	.81	.90
3	2	.80	.50	.92	.92	.81	.91	.93	.92	.80	.91	.91	.91	.79	.90
3	2	.00	.50	.93	.93	.81	.91	.90	.89	.82	.90	.87	.85	.83	.87
3	2	.90	.00	.91	.91	.92	.91	.91	.90	.91	.91	.91	.90	.92	.91
3	2	1.00	.00	.92	.91	.92	.90	.92	.89	.88	.90	.91	.87	.83	.88

TABLE IVB

PROBABILITY OF SELECTING THE CORRECT NUMBER OF COMMON STOCHASTIC TRENDS, $T = 100$, $N = 40$, LINEAR TREND MODEL

				$\sigma^F = \sqrt{10}$				$\sigma^F = 1$				$\sigma^F = \sqrt{.5}$			
r	r_1	ρ_i	α	MQ_τ^c	MQ_f^τ	asic	trace	MQ_c^τ	MQ_f^τ	asic	trace	MQ_c^τ	MQ_f^τ	asic	trace
3	3	.00	–	.96	.96	.96	.82	.94	.94	.96	.83	.91	.90	.95	.81
3	3	.50	–	.97	.97	.96	.83	.96	.96	.96	.83	.95	.95	.96	.82
3	3	.80	–	.97	.97	.97	.83	.97	.97	.97	.83	.97	.96	.96	.83
3	3	.90	–	.97	.97	.97	.82	.97	.97	.97	.84	.96	.96	.97	.85
3	3	1.00	–	.97	.97	.96	.83	.97	.97	.97	.82	.97	.97	.97	.83
3	0	.00	.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	0	.50	.50	1.00	1.00	.98	1.00	1.00	1.00	.98	1.00	1.00	1.00	.98	1.00
3	0	.80	.50	1.00	1.00	.97	1.00	1.00	1.00	.97	1.00	1.00	1.00	.97	1.00
3	0	.00	.50	1.00	1.00	.98	1.00	1.00	1.00	.98	1.00	1.00	1.00	.99	1.00
3	0	.90	.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	0	1.00	.00	1.00	1.00	1.00	1.00	1.00	.99	.99	.81	1.00	.92	.95	.48
3	1	.00	.00	.88	.89	.82	.95	.87	.89	.80	.94	.85	.86	.78	.94
3	1	.50	.50	.90	.90	.70	.95	.88	.88	.70	.94	.88	.88	.69	.94
3	1	.80	.50	.89	.90	.69	.95	.89	.89	.68	.95	.88	.88	.66	.94
3	1	.00	.50	.90	.90	.70	.95	.89	.89	.70	.94	.87	.85	.70	.92
3	1	.90	.00	.89	.90	.82	.95	.89	.89	.81	.95	.89	.89	.79	.94
3	1	1.00	.00	.89	.90	.82	.95	.90	.90	.78	.87	.89	.85	.73	.74
3	2	.00	.00	.88	.88	.94	.91	.85	.85	.93	.90	.82	.82	.91	.89
3	2	.50	.50	.88	.88	.75	.91	.88	.88	.75	.91	.86	.86	.76	.90
3	2	.80	.50	.88	.88	.77	.91	.87	.87	.74	.91	.85	.85	.71	.90
3	2	.00	.50	.89	.89	.77	.91	.86	.86	.78	.90	.83	.82	.78	.87
3	2	.90	.00	.89	.89	.94	.91	.88	.88	.94	.91	.88	.88	.93	.91
3	2	1.00	.00	.89	.89	.94	.90	.89	.88	.92	.90	.88	.86	.88	.88

The data are generated as in note to Table II. MQ_c and MQ_f are tests for the number of common trends, r_1. "Trace" is Johansen's statistic for determining r_0, and "asic" is the information criteria that jointly determine p and r_0.

the number of cointegrating vectors developed by Johansen (1988). Whether or not a hypothesis is entertained depends on the outcome of the preceeding hypothesis being tested. As such, if the chosen level of significance is φ, the probability of selecting the true number of common trends converges to $(1 - \varphi) < 1$.[12] In theory, the ASIC is immune to this problem. Although its accuracy rate is also very high, in finite examples and at least for the configurations considered, it does not appear to have an obvious advantage over the MQ tests.

[12]Consistent rank selection using information criteria was also discussed in Chao and Phillips (1999).

5. CONCLUSION

This paper makes use of the observation that if a panel of data has a factor structure, then testing for the presence of a unit root in the common and the idiosyncratic terms separately should be more effective than unit root testing of the observed data. Accordingly, we first consider how the common factors can be consistently estimated irrespective of the stationarity property of the idiosyncratic errors. We then show that inference about unit roots is not affected by the fact that the true factors and errors are not observed. Our tests for the number of common stochastic trends do not depend on whether the idiosyncratic errors are stationary. Similarly, the test of whether the errors are stationary does not depend on the presence or absence of common stochastic trends. An appeal of PANIC is that r_1 can be determined without pretesting for the presence of a unit root in the data. While pooling is inappropriate when the observed data are cross-correlated, pooling over tests based on the idiosyncratic components are more likely to be valid. Simulations show that the proposed tests have good finite sample properties even for panels with only 40 units. In view of the documented problems concerning unit root tests applied to observed data, the results using PANIC are striking.

The present analysis can be extended in several ways. The common-idiosyncratic decomposition enables inferential analysis in general. The deterministic terms in the factor model are estimated in the present paper by the method of least squares. As such, the unit root tests are implicitly based on least squares detrending. But as Elliott, Rothenberg, and Stock (1996) showed, unit root tests based on GLS detrending are more powerful. The tests developed in this paper can potentially be improved along this direction. Using the results in the Appendix, other unit root and cointegration tests of choice can be developed. Asymptotic analysis can also be developed to analyze time-series processes with roots local to unity. In theory, the machinery developed in this paper can also be used to test long memory, ARCH effects, and other time-series features in the data.

Dept. of Economics, New York University, New York, NY 10003, U.S.A.; jushan.bai@nyu.edu

and

Dept. of Economics, University of Michigan, Ann Arbor, MI 48109, U.S.A.; serena.ng@umich.edu.

Manuscript received January, 2002; final revision received August, 2003.

APPENDIX

By definition, $\hat{e}_{it} = \sum_{s=2}^{t} \hat{z}_{it}$ with $\hat{e}_{i1} = 0$. It follows that $\Delta\hat{e}_{it} = \hat{e}_{it} - \hat{e}_{it-1} = \hat{z}_{it}$. Now

$$x_{it} = \lambda_i' f_t + z_{it} = \lambda_i' H^{-1} H f_t + z_{it}$$

and

$$x_{it} = \hat{\lambda}_i' \hat{f}_t + \hat{z}_{it} = \hat{\lambda}_i' \hat{f}_t + \Delta\hat{e}_{it}.$$

Subtracting the first equation from the second, we obtain

$$\begin{aligned}\Delta\hat{e}_{it} &= z_{it} + \lambda_i' H^{-1} H f_t - \hat{\lambda}_i' \hat{f}_t \\ &= z_{it} - \lambda_i' H^{-1}(\hat{f}_t - H f_t) - (\hat{\lambda}_i - H^{-1\prime}\lambda_i)' \hat{f}_t \\ &= z_{it} - \lambda_i' H^{-1} v_t - d_i' \hat{f}_t,\end{aligned} \tag{12}$$

where $v_t = \hat{f}_t - H f_t$ and $d_i = \hat{\lambda}_i - H^{-1\prime}\lambda_i$. These representations hold for both the intercept and the linear trend case and will be used throughout.

For the intercept model, $z_{it} = \Delta e_{it}$. We can rewrite the above as

$$\Delta\hat{e}_{it} = \Delta e_{it} - \lambda_i' H^{-1} v_t - d_i' \hat{f}_t, \tag{13}$$

$$\hat{e}_{it} = e_{it} - e_{i1} - \lambda_i' H^{-1} \sum_{s=2}^{t} v_s - d_i' \sum_{s=2}^{t} \hat{f}_s \tag{14}$$

$$= e_{it} - e_{i1} - \lambda_i' H^{-1} V_t - d_i' \hat{F}_t, \tag{15}$$

with $V_t = \sum_{s=2}^{t} v_s$ and $\hat{F}_t = \sum_{s=2}^{t} \hat{f}_s$.

For the linear trend model,

$$z_{it} = \Delta e_{it} - \overline{\Delta e_i} = \Delta e_{it} - \frac{e_{iT} - e_{i1}}{T-1}.$$

We have by (12)

$$\Delta\hat{e}_{it} = \Delta e_{it} - \overline{\Delta e_i} - \lambda_i' H^{-1} v_t - d_i' \hat{f}_t, \tag{16}$$

$$\hat{e}_{it} = e_{it} - e_{i1} - \frac{e_{iT} - e_{i1}}{T-1}(t-1) - \lambda_i' H^{-1} \sum_{s=2}^{t} v_s - d_i \sum_{s=2}^{t} \hat{f}_s \tag{17}$$

$$= e_{it} - e_{i1} - \frac{e_{iT} - e_{i1}}{T-1}(t-1) - \lambda_i' H^{-1} V_t - d_i \hat{F}_t. \tag{18}$$

Throughout, we denote $C_{NT} = \min[\sqrt{N}, \sqrt{T}\,]$. In this notation, Lemma 1(a) gives

$$\frac{1}{T}\sum_{t=1}^{T} \|v_t\|^2 = O_p(C_{NT}^{-2})$$

and Lemma 1(c) gives

$$\|d_i\|^2 = O_p\left(\frac{1}{\min[T, N^2]}\right) \le O_p(C_{NT}^{-2}).$$

A. *Proof of Lemma 2*

For notational simplicity, we assume there are $T+1$ observations $(t = 0, 1, \dots, T)$ for this lemma. The differenced data have T observations so that x is $T \times N$. Let V_{NT} be the $r \times r$ diagonal matrix of the first r largest eigenvalues of $(NT)^{-1}xx'$ in decreasing order. By the definition of eigenvectors and eigenvalues, we have $(NT)^{-1}xx'\hat{f} = \hat{f}V_{NT}$ or $(NT)^{-1}xx'\hat{f}V_{NT}^{-1} = \hat{f}$. We make use of an $r \times r$ matrix H defined as follows: $H = V_{NT}^{-1}(\hat{f}'f/T)(\Lambda'\Lambda/N)$. Then the following is a mathematical identity:

$$\hat{f}_t - H f_t = V_{NT}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T} \hat{f}_s \gamma_N(s,t) + \frac{1}{T}\sum_{s=1}^{T} \hat{f}_s \zeta_{st} + \frac{1}{T}\sum_{s=1}^{T} \hat{f}_s \eta_{st} + \frac{1}{T}\sum_{s=1}^{T} \hat{f}_s y_{st}\right), \tag{A.1}$$

where for $z_t = (z_{1t}, z_{2t}, \dots, z_{Nt})'$,

(A.2) $$\zeta_{st} = \frac{z_s' z_t}{N} - \gamma_N(s,t), \quad \eta_{st} = f_s' \Lambda' z_t / N, \quad y_{st} = f_t' \Lambda' z_s / N.$$

Bai (2003) showed that $\|V_{NT}^{-1}\| = O_p(1)$. Using $\hat{f}'\hat{f}/T = I_r$, together with Assumptions A and B, $\|H\| = O_p(1)$. To prove Lemma 2, we need additional results:

LEMMA A.1: *Under Assumptions* A–D, *we have*:
1. $E(z_{it}) = 0$, $E|z_{it}|^8 \le M$;
2. $E(N^{-1}\sum_{i=1}^N z_{is} z_{it}) = \gamma_N(s,t)$, $\sum_{s=1}^T |\gamma_N(s,t)| \le M$ *for all* t;
3. $E(z_{it} z_{jt}) = \phi_{ij}$ *with* $\sum_{i=1}^N |\phi_{ij}| \le M$ *for all* j;
4. $E(\max_{1\le k\le T} \frac{1}{\sqrt{NT}} \| \sum_{t=1}^k \sum_{i=1}^N \lambda_i z_{it} \|) \le M$;
5. $E|N^{-1/2}\sum_{i=1}^N [z_{is} z_{it} - E(z_{is} z_{it})]|^4 \le M$, *for every* (t,s);
6. $E(\max_{1\le k\le T} \frac{1}{\sqrt{NT}} | \sum_{t=1}^k \sum_{i=1}^N (z_{is} z_{it} - E(z_{is} z_{it}))|)^2 \le M$, *for every* s.

The proof of this lemma is elementary and thus is omitted. Note that this lemma does not involve estimated variables.

LEMMA A.2: *Under Assumptions* A–D, *we have for* $C_{NT} = \min[\sqrt{N}, \sqrt{T}]$:
(a) $T^{-3/2} \sup_{1\le k\le T} \| \sum_{t=1}^k \sum_{s=1}^T \hat{f}_s \gamma_N(s,t) \| = O_p(1/(\sqrt{T} C_{NT})) + O_p(T^{-3/4})$;
(b) $T^{-3/2} \sup_{1\le k\le T} \| \sum_{t=1}^k \sum_{s=1}^T \hat{f}_s \zeta_{st} \| = O_p(1/\sqrt{N})$;
(c) $T^{-3/2} \sup_{1\le k\le T} \| \sum_{t=1}^k \sum_{s=1}^T \hat{f}_s \eta_{st} \| = O_p(1/\sqrt{N})$;
(d) $T^{-3/2} \sup_{1\le k\le T} \| \sum_{t=1}^k \sum_{s=1}^T \hat{f}_s y_{st} \| = O_p(1/(\sqrt{N} C_{NT}))$.

PROOF: Consider part (a). By adding and subtracting terms,

$$\sum_{t=1}^k \sum_{s=1}^T \hat{f}_s \gamma_N(s,t) = \sum_{s=1}^T (\hat{f}_s - H f_s) \sum_{t=1}^k \gamma_N(s,t) + H \sum_{s=1}^T f_s \sum_{t=1}^k \gamma_N(s,t).$$

Consider the first term:

$$\left\| \sum_{s=1}^T (\hat{f}_s - H f_s) \sum_{t=1}^k \gamma_N(s,t) \right\| \le \left(\sum_{s=1}^T \|\hat{f}_s - H f_s\|^2 \right)^{1/2} \left(\sum_{s=1}^T \left| \sum_{t=1}^k \gamma_N(s,t) \right|^2 \right)^{1/2}.$$

By Lemma 1(i), $(\sum_{s=1}^T \|\hat{f}_s - H f_s\|^2)^{1/2} = T^{1/2} O_p(C_{NT}^{-1})$. Because $|\sum_{t=1}^k \gamma_N(s,t)| \le M$ for all k and s, $(\sum_{s=1}^T |\sum_{t=1}^k \gamma_N(s,t)|^2)^{1/2} \le M\sqrt{T}$. Thus $T^{-3/2} \| \sum_{s=1}^T (\hat{f}_s - H f_s) \sum_{t=1}^k \gamma_N(s,t) \| = O_p((\sqrt{T} C_{NT})^{-1})$. Consider the second term. We use the following fact: Let $X_1, X_2, \dots, X_T$ be an arbitrary sequence of random variables. If $\max_{1\le k\le T} E|X_k|^\alpha \le M$ $(\alpha > 0)$; then $\max_{1\le k\le T} |X_k| = O_p(T^{1/\alpha})$. Let $a_{sk} = \sum_{t=1}^k \gamma_N(s,t)$; then $E\|T^{-1/2}\sum_{s=1}^T f_s a_{sk}\|^4 \le M$ by Assumption B and Lemma A.1(2). This implies that with $\alpha = 4$ and $X_k = T^{-1/2}\sum_{s=1}^T f_s a_{sk}$

$$T^{-3/2} \sup_{1\le k\le T} \left\| \sum_{s=1}^T f_s a_{sk} \right\| = O_p(T^{-3/4}),$$

proving (a). Consider part (b).

$$T^{-3/2} \sum_{t=1}^k \sum_{s=1}^T \hat{f}_s \zeta_{st} = T^{-1} \sum_{s=1}^T (\hat{f}_s - H f_s) \frac{1}{\sqrt{T}} \sum_{t=1}^k \zeta_{st} + H T^{-1} \sum_{s=1}^T f_s \frac{1}{\sqrt{T}} \sum_{t=1}^k \zeta_{st}.$$

For the first term,

$$\left\| T^{-1}\sum_{s=1}^{T}(\hat{f}_s - Hf_s)\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\zeta_{st}\right\| \leq \left(\frac{1}{T}\sum_{s=1}^{T}\|\hat{f}_s - Hf_s\|^2\right)^{1/2}\left[\frac{1}{T}\sum_{s=1}^{T}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\zeta_{st}\right)^2\right]^{1/2}.$$

Furthermore,

$$\begin{aligned}\frac{1}{T}\sum_{s=1}^{T}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\zeta_{st}\right)^2 &= \frac{1}{T}\sum_{s=1}^{T}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\left(\frac{z_s'z_t}{N} - \gamma_N(s,t)\right)\right]^2 \\ &= \frac{1}{T}\sum_{s=1}^{T}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\left(\frac{z_s'z_t}{N} - \frac{E(z_s'z_t)}{N}\right)\right]^2 \\ &= \frac{1}{N}\frac{1}{T}\sum_{s=1}^{T}\left[\frac{1}{\sqrt{NT}}\sum_{t=1}^{k}\sum_{i=1}^{N}(z_{is}z_{it} - E(z_{is}z_{it}))\right]^2 = O_p\left(\frac{1}{N}\right),\end{aligned}$$

uniformly in k by Lemma A.1(6). Thus the first term is $O_p((C_{NT})^{-1})O_p(N^{-1/2})$. Next,

$$T^{-3/2}\sum_{s=1}^{T}f_s\sum_{t=1}^{k}\zeta_{st} = \frac{1}{T\sqrt{N}}\sum_{s=1}^{T}f_s\frac{1}{\sqrt{TN}}\sum_{t=1}^{k}\sum_{i=1}^{N}(z_{is}z_{it} - E(z_{is}z_{it})) = \frac{1}{\sqrt{N}T}\sum_{s=1}^{T}f_s\phi_{k,s},$$

where $\phi_{k,s}$ is implicitly defined in the above expression. Lemma A.1(6) implies that $E(\max_{1\le k\le T}|\phi_{k,s}|) \le M$. Thus

$$E\left(\max_{1\le k\le T}(\sqrt{N}T)^{-1}\left\|\sum_{s=1}^{T}f_s\phi_{k,s}\right\|\right) \le (\sqrt{N}T)^{-1}\sum_{s=1}^{T}E\left(\|f_s\|\max_{1\le k\le T}|\phi_{k,s}|\right) = O(N^{-1/2}),$$

because $E(\|f_s\|\max_{1\le k\le T}|\phi_{k,s}|) = E\|f_s\|\cdot E(\max_{1\le k\le T}|\phi_{k,s}|) \le M_1$ $(M_1 < \infty)$ by the independence of f_s and the z_{it}'s. Thus, uniformly in k,

$$T^{-3/2}\sum_{s=1}^{T}\hat{f}_s\sum_{t=1}^{k}\zeta_{st} = O_p\left(\frac{1}{C_{NT}}\right)\cdot O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) = O_p\left(\frac{1}{\sqrt{N}}\right).$$

Consider part (c):

$$T^{-3/2}\sum_{s=1}^{T}\sum_{t=1}^{k}\hat{f}_s\eta_{st} = T^{-1}\sum_{s=1}^{T}(\hat{f}_s - Hf_s)\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\eta_{st} + HT^{-1}\sum_{s=1}^{T}f_s\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\eta_{st}.$$

But $T^{-1}\sum_{s=1}^{T}f_sT^{-1/2}\sum_{t=1}^{k}\eta_{st} = (T^{-1}\sum_{s=1}^{T}f_sf_s')(N\sqrt{T})^{-1}\sum_{t=1}^{k}\sum_{i=1}^{N}\lambda_i z_{it} = O_p(N^{-1/2})$, uniformly in k by Lemma A.1(4). Next,

$$\left\| T^{-1}\sum_{s=1}^{T}(\hat{f}_s - Hf_s)\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\eta_{st}\right\| \leq \left(\frac{1}{T}\sum_{s=1}^{T}\|\hat{f}_s - Hf_s\|^2\right)^{1/2}\cdot\left[\frac{1}{T}\sum_{s=1}^{T}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\eta_{st}\right)^2\right]^{1/2}.$$

The first expression is $O_p(1/C_{NT})$ by Lemma 1. For the second expression,

$$T^{-1}\sum_{s=1}^{T}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\eta_{st}\right)^2=\frac{1}{N}\frac{1}{T}\sum_{s=1}^{T}\left(f_s'\frac{1}{\sqrt{TN}}\sum_{t=1}^{k}\sum_{i=1}^{N}\lambda_i z_{it}\right)^2=O_p\left(\frac{1}{N}\right),$$

uniformly in k. Thus, (c) is $O_p(N^{-1/2})+O_p((\sqrt{N}C_{NT})^{-1})=O_p(N^{-1/2})$.

Finally for part (d),

$$\begin{aligned}T^{-3/2}\sum_{t=1}^{k}\sum_{s=1}^{T}\hat{f}_s y_{st}&=T^{-3/2}\sum_{t=1}^{k}\sum_{s=1}^{T}\hat{f}_s f_t'\Lambda z_s/N\\&=T^{-1}\sum_{s=1}^{T}(\hat{f}_s z_s'\Lambda/N)\frac{1}{\sqrt{T}}\sum_{t=1}^{k}f_t.\end{aligned}$$

It is proved in Bai (2003) that $T^{-1}\sum_{s=1}^{T}(\hat{f}_s z_s'\Lambda/N)=O_p((\sqrt{N}C_{NT})^{-1})$ (this can also be proved directly). Assumption B implies that $T^{-1/2}\sum_{t=1}^{k}f_t=O_p(1)$ uniformly in k. Thus (d) is equal to $O_p((\sqrt{N}C_{NT})^{-1})$ uniformly in k. The proof of Lemma A.2 is complete. *Q.E.D.*

From (A.1) and Lemma A.2

$$\begin{aligned}&\max_{1\le k\le T}\frac{1}{\sqrt{T}}\left\|\sum_{t=1}^{k}(\hat{f}_t-Hf_t)\right\|\\&=O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right)+O_p\left(\frac{1}{T^{3/4}}\right)+O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\sqrt{N}C_{NT}}\right)\\&=O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{T^{3/4}}\right).\end{aligned}$$

By definition, $V_t=\sum_{s=2}^{t}v_s=\sum_{s=1}^{t}(\hat{f}_s-HF_s)$. Lemma 2 can be stated as

$$\max_{2\le t\le T}\frac{1}{\sqrt{T}}\|V_t\|=\max_{2\le t\le T}\frac{1}{\sqrt{T}}\left\|\sum_{s=2}^{t}v_s\right\|=O_p(C_{NT}^{-1}).\tag{A.3}$$

From $\|V_t\|=O_p(T/N)$ uniformly in t, we also have

$$\frac{1}{T}\sum_{t=2}^{T}\|V_t\|^2=O_p\left(\frac{T}{N}\right).\tag{A.4}$$

B. *Preliminaries for Theorem 1*

LEMMA B.1: *For* $\rho_i=1$ *or* $|\rho_i|<1$:

(i) $(1/\sqrt{T})\hat{e}_{it}=(1/\sqrt{T})e_{it}+O_p(C_{NT}^{-1})$, *uniformly in* $t\in[1,T]$;

(ii) $(1/T^2)\sum_{t=2}^{T}\hat{e}_{it}^2=(1/T^2)\sum_{t=2}^{T}e_{it}^2+O_p(C_{NT}^{-1})$;

(iii) $(1/T)\sum_{t=2}^{T}(\Delta\hat{e}_{it})^2=(1/T)\sum_{t=2}^{T}(\Delta e_{it})^2+O_p(C_{NT}^{-1})$;

(iv) $(1/T)\sum_{t=2}^{T}\hat{e}_{it-1}\Delta\hat{e}_{it}=(1/T)\sum_{t=2}^{T}e_{it-1}\Delta e_{it}+O_p(C_{NT}^{-1})$.

PROOF: (i) From (14),

$$\frac{\hat{e}_{it}}{\sqrt{T}} = \frac{e_{it}}{\sqrt{T}} - \frac{e_{i1}}{\sqrt{T}} - \lambda_i' H^{-1}\left(\frac{1}{\sqrt{T}}\sum_{s=2}^{t} v_s\right) - d_i' \frac{1}{\sqrt{T}}\sum_{s=2}^{t}\hat{f}_s.$$

Now $e_{i1}/\sqrt{T} = O_p(T^{-1/2}) = O_p(C_{NT}^{-1})$. The third term is $O_p(C_{NT}^{-1})$ by (A.3). By Lemma 1(c), $d_i = O_p(\max[T^{-1/2}, N^{-1}]) = O_p(C_{NT}^{-1})$, and

$$\left\|\frac{1}{\sqrt{T}}\sum_{s=2}^{t}\hat{f}_s\right\| \le \left\|\frac{1}{\sqrt{T}}\sum_{s=2}^{t}(\hat{f}_s - Hf_s)\right\| + \left\|\frac{1}{\sqrt{T}}\sum_{s=2}^{t} f_s\right\| \cdot \|H\|$$
$$= O_p(C_{NT}^{-1}) + O_p(1) = O_p(1).$$

Thus the last term is also $O_p(C_{NT}^{-1})$, proving (i). Part (ii) is a direct consequence of (i).

Consider (iii). From (13), $\Delta\hat{e}_{it} = \Delta e_{it} - a_{it}$, where $a_{it} = \lambda_i' H^{-1} v_t + d_i'\Delta\hat{F}_t$. Thus,

$$\frac{1}{T}\sum_{t=2}^{T}(\Delta\hat{e}_{it})^2 = \frac{1}{T}\sum_{t=2}^{T}(\Delta e_{it})^2 - \frac{2}{T}\sum_{t=2}^{T}(\Delta e_{it})a_{it} + \frac{1}{T}\sum_{t=2}^{T} a_{it}^2.$$

The middle term is $O_p(C_{NT}^{-1})$ by the Cauchy–Schwartz inequality and $\sum_{t=2}^{T} a_{it}^2/T = O_p(C_{NT}^{-2})$. The latter follows from $a_{it}^2 \le 2\|\lambda_i' H^{-1}\|^2\|v_t\|^2 + 2\|d_i\|^2\|\hat{f}_t\|^2$ and

$$\frac{1}{T}\sum_{t=2}^{T} a_{it}^2 \le 2\|\lambda_i' H^{-1}\|^2 \cdot \frac{1}{T}\sum_{t=2}^{T}\|v_t\|^2 + 2\|d_i\|^2 \frac{1}{T}\sum_{t=2}^{T}\|\hat{f}_t\|^2$$
$$= O_p(1)O_p(C_{NT}^{-2}) + O_p(C_{NT}^{-2})O_p(1) = O_p(C_{NT}^{-2})$$

by Lemma 1(a) and $\sum_{t=2}^{T}\|\hat{f}_t\|^2/T = O_p(1)$. This proves (iii).

Consider (iv). From $\hat{e}_{it}^2 = (\hat{e}_{it-1} + \Delta\hat{e}_{it})^2 = \hat{e}_{it-1}^2 + (\Delta\hat{e}_{it})^2 + 2\hat{e}_{it-1}\Delta\hat{e}_{it}$, we have the identity

$$\frac{1}{T}\sum_{t=2}^{T}\hat{e}_{it-1}\Delta\hat{e}_{it} = \frac{\hat{e}_{iT}^2}{2T} - \frac{\hat{e}_{i1}^2}{2T} - \frac{1}{2T}\sum_{t=2}^{T}(\Delta\hat{e}_{it})^2.$$

A similar identity holds for $T^{-1}\sum_{t=2}^{T} e_{it-1}\Delta e_{it}$. Comparing the right-hand side of the two identities, we have $\hat{e}_{iT}^2/T - e_{iT}^2/T = O_p(C_{NT}^{-1})$ by part (i) with $t = T$, and $T^{-1}\sum_{t=2}^{T}(\Delta\hat{e}_{it})^2 - T^{-1}\sum_{t=2}^{T}(\Delta e_{it})^2 = O_p(C_{NT}^{-1})$ by part (iii), proving (iv). *Q.E.D.*

Let $\overline{\hat{F}} = \sum_{t=2}^{T}\hat{F}_t/(T-1)$ and $\overline{F} = \sum_{t=2}^{T} F_t/(T-1)$ be the sample means. Let $\hat{F}_t^c = \hat{F}_t - \overline{\hat{F}}$ be the demeaned series and we define F_t^c similarly.

LEMMA B.2: *Under Assumptions* A–E:

(i) $(1/\sqrt{T})\hat{F}_t = H(1/\sqrt{T})F_t + O_p(C_{NT}^{-1})$ *uniformly in* $t \in [2, T]$;

(ii) $(1/T^2)\sum_{t=2}^{T}\hat{F}_t\hat{F}_t' = H((1/T^2)\sum_{t=2}^{T} F_t F_t')H' + O_p(C_{NT}^{-1})$;

(iii) $(1/T)\sum_{t=2}^{T}\Delta\hat{F}_t\Delta\hat{F}_t' = H((1/T)\sum_{t=2}^{T}\Delta F_t\Delta F_t')H' + O_p(C_{NT}^{-1})$;

(iv) $(1/T)\sum_{t=2}^{T}(\hat{F}_{t-1}\Delta\hat{F}_t' + \Delta\hat{F}_t\hat{F}_{t-1}') = (1/T)H\sum_{t=2}^{T}(F_{t-1}\Delta F_t' + \Delta F_t F_{t-1}')H' + O_p(C_{NT}^{-1})$;

(v) $(1/\sqrt{T})\overline{\hat{F}} = (1/\sqrt{T})H\overline{F} + O_p(C_{NT}^{-1})$;

(vi) $(1/\sqrt{T})\hat{F}^c = (1/\sqrt{T})HF^c + O_p(C_{NT}^{-1})$;

(vii) $(1/T^2)\sum_{t=2}^{T}\hat{F}_t^c\hat{F}_t^{c\prime} = H((1/T^2)\sum_{t=2}^{T} F_t^c F_t^{c\prime})H' + O_p(C_{NT}^{-1})$;

(viii) $(1/T)\sum_{t=2}^{T}(\hat{F}_{t-1}^c\Delta\hat{F}_t' + \Delta\hat{F}_t\hat{F}_{t-1}^{c\prime}) = H[(1/T)\sum_{t=2}^{T}(F_{t-1}^c\Delta F_t' + \Delta F_t F_{t-1}^{c\prime})]H' + O_p(C_{NT}^{-1})$.

PROOF: Because $\hat{F}_t = \sum_{s=2}^t \hat{f}_s$, we have $\Delta\hat{F}_t = \hat{F}_t - \hat{F}_{t-1} = \hat{f}_t$. Thus $v_t = \hat{f}_t - Hf_t = \Delta\hat{F}_t - H\Delta F_t$, or $\Delta\hat{F}_t = H\Delta F_t + v_t$. The cumulative sum implies

$$\hat{F}_t = HF_t - HF_1 + \sum_{s=2}^t v_s \tag{B.1}$$

for $t = 2, \dots, T$. Define $\hat{F}_1 = 0$. Thus $T^{-1/2}\hat{F}_t = HT^{-1/2}F_t - HT^{-1/2}F_1 + T^{-1/2}\sum_{s=2}^t v_s$. The second term on the right-hand side is $O_p(T^{-1/2})$ and the third term is $O_p(C_{NT}^{-1})$ uniformly in t by (A.3), proving (i). (ii) is an immediate consequence of (i).

Consider (iii). From $\Delta\hat{F}_t = H\Delta F_t + v_t$, we have

$$\frac{1}{T}\sum_{t=2}^T \Delta\hat{F}_t\Delta\hat{F}_t' = H\frac{1}{T}\sum_{t=2}^T \Delta F_t\Delta F_t'H' + \frac{1}{T}H\sum_{t=2}^T \Delta F_t v_t' + \frac{1}{T}\sum_{t=2}^T v_t\Delta F_t'H' + \frac{1}{T}\sum_{t=2}^T v_t v_t'. \tag{B.2}$$

The conclusion follows from $\|T^{-1}\sum_{t=2}^T v_t v_t'\| \le T^{-1}\sum_{t=2}^T \|v_t\|^2 = O_p(C_{NT}^{-2})$ and

$$\begin{aligned} T^{-1}\left\|H\sum_{t=2}^T \Delta F_t v_t'\right\| &\le \left(T^{-1}\sum_{t=2}^T \|\Delta F_t\|^2\right)^{1/2}\left(T^{-1}\sum_{t=2}^T \|v_t\|^2\right)^{1/2}\|H\| \\ &= O_p(C_{N,T}^{-1})\cdot O_p(1). \end{aligned}$$

For (iv), we use the identity

$$\frac{1}{T}\left[\sum_{t=2}^T(\Delta\hat{F}_t\hat{F}_{t-1}' + \hat{F}_{t-1}\Delta\hat{F}_t')\right] = \frac{\hat{F}_T\hat{F}_T'}{T} - \frac{\hat{F}_1\hat{F}_1'}{T} - \frac{1}{T}\sum_{t=2}^T \Delta\hat{F}_t\Delta\hat{F}_t'$$

and

$$\frac{1}{T}\left[\sum_{t=2}^T \Delta F_t F_{t-1}' + F_{t-1}\Delta F_t'\right] = \frac{F_T F_T'}{T} - \frac{F_1F_1'}{T} - \frac{1}{T}\sum_{t=2}^T \Delta F_t\Delta F_t'.$$

Note that $\hat{F}_1 = 0$ and $F_1F_1'/T = O_p(T^{-1})$. Part (i) of this lemma when $t = T$ implies that $\hat{F}_T\hat{F}_T'/T = H(F_TF_T'/T)H' + O_p(C_{NT}^{-1})$. These together with part (iii) imply (iv).

Consider (v). Averaging over (B.1), we obtain $\overline{\hat{F}} = H\overline{F} - HF_1 + \sum_{t=2}^T\sum_{s=2}^t v_s/(T-1)$. Hence,

$$\frac{1}{\sqrt{T}}\overline{\hat{F}} = H\frac{1}{\sqrt{T}}\overline{F} - H\frac{F_1}{\sqrt{T}} + \frac{1}{(T-1)}\sum_{t=2}^T\left(\frac{1}{\sqrt{T}}\sum_{s=2}^t v_s\right).$$

The second term on the right is $O_p(T^{-1/2})$ and the last term is $O_p(C_{NT}^{-1})$ because it is the average of $(T^{-1/2}\sum_{s=2}^t v_s)$ and thus must be no larger than its maximum, which is $O_p(C_{NT}^{-1})$ by (A.3), proving (v). The difference of (i) and (v) yields (vi). Result (vii) is an immediate consequence of (vi).

For (viii), the sum of demeaned series can be expressed as the sum of nondemeaned series plus extra terms

$$\begin{aligned} &\frac{1}{T}\sum_{t=2}^T(\hat{F}_t^c\Delta\hat{F}_t' + \Delta\hat{F}_t\hat{F}_t^{c\prime}) \\ &\quad = \frac{1}{T}\sum_{t=2}^T(\hat{F}_{t-1}\Delta\hat{F}_t' + \Delta\hat{F}_t\hat{F}_t') - \left(\frac{1}{\sqrt{T}}\overline{\hat{F}}\right)\left(\frac{1}{\sqrt{T}}\hat{F}_T'\right) - \left(\frac{1}{\sqrt{T}}\hat{F}_T\right)\left(\frac{1}{\sqrt{T}}\overline{\hat{F}}'\right). \end{aligned} \tag{B.3}$$

A similar identity holds for the true series F_t. Part (viii) is obtained by comparing the right-hand sides of the two identities and by invoking parts (iv), (v), and (i). *Q.E.D.*

Before proving the general case of serially correlated disturbances, we first consider the case of uncorrelated disturbances. That is, $e_{it} = \rho_i e_{it-1} + \epsilon_{it}$, with ϵ_{it} being i.i.d. This simple setup still provides substantial insight.

PROPOSITION 1: *If $D_i(L) = 1$, i.e., $e_{it} = \rho_i e_{it-1} + \epsilon_{it}$, then under the null hypothesis that $\rho_i = 1$,*

$$DF_{\hat{e}}^c(i) = \frac{\sum_{t=2}^T \hat{e}_{it-1}\Delta\hat{e}_{it}}{(\hat{\sigma}_{\epsilon i}^2 \sum_{t=2}^T \hat{e}_{it-1}^2)^{1/2}} \Rightarrow \frac{\frac{1}{2}(W_{\epsilon i}(1)^2 - 1)}{(\int_0^1 W_{\epsilon i}^2\, dr)^{1/2}}, \tag{B.4}$$

where $\hat{\sigma}_{\epsilon i}^2 = \sum_{t=2}^T (\Delta\hat{e}_{it} - \hat{b}_i \hat{e}_{it-1})^2/(T-1)$ and $\hat{b}_i$ is the OLS estimator when regressing $\Delta\hat{e}_{it}$ on $\hat{e}_{it-1}$.

Proposition 1 is an immediate consequence of the following:

LEMMA B.3: *Under the assumptions of Proposition 1 with $\rho_i = 1$, then as $N, T \to \infty$:*
(i) $(1/T^2)\sum_{t=2}^T \hat{e}_{it}^2 \Rightarrow \sigma_{\epsilon i}^2 \int_0^1 W_{\epsilon i}(r)^2\, dr$;
(ii) $(1/T)\sum_{t=2}^T \hat{e}_{it-1}\Delta\hat{e}_{it} \Rightarrow (\sigma_{\epsilon i}^2/2)(W_{\epsilon i}(1)^2 - 1)$.

These results are implied by Lemma B.1 parts (ii) and (iv) and the corresponding weak convergence of $\sum_{t=2}^T e_{it}^2/T^2$ and $\sum_{t=2}^T e_{it-1}\Delta e_{it}/T$. Note that $\hat{\sigma}_{\epsilon i}^2 \to \sigma_{\epsilon i}^2$. To see this, $\hat{\sigma}_{\epsilon i}^2 = \sum_{t=2}^T \Delta\hat{e}_{it}^2/(T-1) - 2\hat{b}_i \sum_{t=2}^T \Delta\hat{e}_{it}\hat{e}_{it-1}/(T-1) + \hat{b}_i^2 \sum_{t=2}^T \hat{e}_{it}^2/(T-1)$. From $T\hat{b}_i = O_p(1)$, the last two terms are each $o_p(1)$, and the first term converges to $\sigma_{\epsilon i}^2$ by Lemma B.1(iii).

Next consider the Dickey–Fuller test based on $\hat{F}_t$ with demeaning, when $r = 1$ (and hence H is scalar).

PROPOSITION 2: *If $C(L) = 1$, i.e., $F_t = F_{t-1} + u_t$, then*

$$DF_{\hat{F}}^c = \frac{\sum_{t=2}^T (\hat{F}_{t-1} - \overline{\hat{F}})\Delta\hat{F}_t}{(\hat{\sigma}_u^2 \sum_{t=2}^T (\hat{F}_{t-1} - \overline{\hat{F}})^2)^{1/2}} \Rightarrow \frac{\int_0^1 W_u^c(r)\, dW_u(r)}{(\int_0^1 W_u^c(r)^2\, dr)^{1/2}}, \tag{B.5}$$

where $\overline{\hat{F}} = \sum_{t=2}^T \hat{F}_t/(T-1)$, $\hat{\sigma}_u^2 = \sum_{t=2}^T (\Delta\hat{F}_t - \hat{a}\hat{F}_{t-1})^2/(T-1)$ with $\hat{a}$ being the OLS estimator when regressing $\Delta\hat{F}_t$ on $\hat{F}_{t-1}$, and $W_u^c(r) = W_u(r) - \int_0^1 W_u(r)\, dr$ is a demeaned Brownian motion.

The proposition is implied by the following:

LEMMA B.4: *If $C(L) = 1$, i.e., $F_t = F_{t-1} + u_t$, then as $N, T \to \infty$:*
(i) $(1/T)\sum_{t=2}^T (\hat{F}_{t-1} - \overline{\hat{F}})\Delta\hat{F}_t \Rightarrow H^2\sigma_u^2 \int_0^1 W_u^c(r)\, dW_u(r)$;
(ii) $(1/T^2)\sum_{t=2}^T (\hat{F}_{t-1} - \overline{\hat{F}})^2 \Rightarrow H^2\sigma_u^2 \int_0^1 W_u^c(r)^2\, dr$.

These results follow from Lemma B.2 parts (viii) and (vii), respectively, and from the known weak convergence for the series F_t.

C. *Testing $\hat{e}_{it}$ Using the ADF Test, Intercept Only*

Under $H_0: \rho_i = 1$, $\Delta e_{it} = D_i(L)\epsilon_{it}$ is a stationary process and $D_i(L)$ is invertible. We can write

$$\Delta e_{it} = \sum_{j=1}^{\infty} \delta_{ij}\Delta e_{i,t-j} + \epsilon_{it}.$$

Let $\omega_{\epsilon i}^2 = D_i(1)^2\sigma_{\epsilon i}^2$ be the long run variance of Δe_{it}. The functional central limit theorem gives $T^{-1/2}\sum_{j=1}^{[Ts]}\Delta e_{ij} \Rightarrow \omega_{\epsilon i}W_{\epsilon i}(s)$. The regression when e_{it} is observed is

$$\Delta e_{it} = \delta_{i0}e_{it-1} + \sum_{j=1}^{p}\delta_{ij}\Delta e_{i,t-j} + \epsilon_{i,tp}, \tag{C.1}$$

$$\epsilon_{i,tp} = \epsilon_{it} + \sum_{j=p+1}^{\infty}\delta_{ij}\Delta e_{i,t-j}. \tag{C.2}$$

Let $Z_{it} = (\Delta e_{i,t-1}, \dots, \Delta e_{i,t-p})'$, $x_{it} = (e_{i,t-1}Z_{it}')'$, and $D_T = \operatorname{diag}(T^{-1}, T^{-1/2}, \dots, T^{-1/2})$. Let

$$\tilde{M}_{ip} = D_T\sum_{t=p}^{T}x_{it}x_{it}'D_T = D_T\begin{bmatrix}\sum_{t=p}^{T}e_{i,t-1}^2 & \sum_{t=p}^{T}e_{i,t-1}Z_{it}' \\ \sum_{t=p}^{T}e_{i,t-1}Z_{it} & \sum_{t=p}^{T}Z_{it}Z_{it}'\end{bmatrix}D_T,$$

$$M_{ip} = \begin{bmatrix}\int_0^1 W_{\epsilon i}(r)^2\,dr & 0 \\ 0 & \Gamma_{iz}(p)\end{bmatrix},$$

where $\Gamma_{iz}(p) = E(Z_{it}Z_{it}')$. Consider $ADF_e(i)$, the t test on $\delta_{i0} = 0$. Let $\tilde{\delta}_i(p) = (\tilde{\delta}_{i0}, \tilde{\delta}_{i1}, \dots, \tilde{\delta}_{ip})$ be the least squares estimates from regressing Δe_{it} on e_{it-1} and lags of Δe_{it}. Let $e = (1, 0, \dots, 0)$ be a selection vector. Note

$$\begin{aligned}T(\tilde{\delta}_{i0} - \delta_{i0}) &= e'D_T^{-1}[\tilde{\delta}_i(p) - \delta_i(p)] \\ &= e'M_{ip}^{-1}D_T\sum_{t=p}^{T}x_{it}\epsilon_{it} + e'\big(\tilde{M}_{ip}^{-1} - M_{ip}^{-1}\big)D_T\sum_{t=p}^{T}x_t\epsilon_{i,tp} \\ &\quad + e'M_{ip}^{-1}D_T\sum_{t=p}^{T}x_{it}(\epsilon_{i,tp} - \epsilon_{it}).\end{aligned}$$

Said and Dickey (1984) showed that $\|M_{ip}\| = O_p(1)$, $\|M_{ip}^{-1}\| = O_p(1)$, $D_T\sum_{t=p}^{T}x_{it}\epsilon_{i,tp} = O_p(\sqrt{p})$, $p^{1/2}\|\tilde{M}_{ip}^{-1} - M_{ip}^{-1}\| \to 0$ if $p^3/T \to 0$ as $p, T \to \infty$. Since $\tilde{\sigma}_{\epsilon i}^2 = T^{-1}\sum_{t=p}^{T}\tilde{\epsilon}_{it}^2 \xrightarrow{p} \sigma_{\epsilon i}^2$, where $\tilde{\epsilon}_{it} = \Delta e_{it} - \tilde{\delta}_i(p)'x_{it}$, under the null that $\delta_{i0} = 0$,

$$ADF_e(i) = \frac{T\tilde{\delta}_{i0}}{\sqrt{\tilde{\sigma}_{\epsilon i}^2[\tilde{M}_{ip}^{-1}]_{11}}} \Rightarrow \frac{\int_0^1 W_{\epsilon i}(r)\,dW_{\epsilon i}(r)}{(\int_0^1 W_{\epsilon i}(r)^2\,dr)^{1/2}}.$$

We use $\hat{e}_{it}$ instead of e_{it} for testing, where $\Delta\hat{e}_{it}$ and $\hat{e}_{it}$ are defined in (13) and (15). Define $\hat{M}_{ip}$, $\hat{\delta}_i(p)$ with $\hat{e}_{it}$ in place of e_{it}. Then $\hat{\delta}_i(p)$ are the least squares estimates from regressing $\Delta\hat{e}_{it}$ on $\hat{e}_{i,t-1}$ and lags of $\Delta\hat{e}_{it}$. Furthermore, $\hat{\epsilon}_{it} = \Delta\hat{e}_{it} - \hat{\delta}_i(p)'\hat{x}_{it}$ are the estimated residuals, and $\hat{\sigma}_{\epsilon i}^2 = T^{-1}\sum_{t=p}^{T}\hat{\epsilon}_{it}^2$. The test statistic is

$$ADF_{\hat{e}}(i) = \frac{T\hat{\delta}_{i0}}{\sqrt{\hat{\sigma}_{\epsilon i}^2[\hat{M}_{ip}^{-1}]_{11}}}.$$

We will prove $ADF_{\hat{e}}(i) - ADF_e(i) = o_p(1)$ by showing $T(\tilde{\delta}_{i0} - \hat{\delta}_{i0}) = o_p(1)$ and $\hat{\sigma}_{\epsilon i}^2[\hat{M}_{ip}^{-1}]_{11} - \tilde{\sigma}_{\epsilon i}^2[\tilde{M}_{ip}^{-1}]_{11} = o_p(1)$ under the condition $p^3/\min[N, T] \to 0$.

From $T\hat{\delta}_{i0} = e'D_T^{-1}\hat{\delta}_i(p) = e'\hat{M}_{ip}^{-1}D_T\sum_{t=p}^T \hat{x}_{it}\Delta\hat{e}_{it}$ and $T\tilde{\delta}_{i0} = e'\tilde{M}_{ip}^{-1}D_T\sum_{t=p}^T x_{it}\Delta e_{it}$,

$$T(\hat{\delta}_{i0}-\tilde{\delta}_{i0}) = e'\big(\hat{M}_{ip}^{-1}-\tilde{M}_{ip}^{-1}\big)D_T\sum_{t=p}^T \hat{x}_{it}\Delta\hat{e}_{it} + e'\tilde{M}_{ip}^{-1}D_T\sum_{t=p}^T(\hat{x}_{it}\Delta\hat{e}_{it}-x_{it}\Delta e_{it}).$$

For the first term, Lemma C.1 shows that $\|\hat{M}_{ip}^{-1}-\tilde{M}_{ip}^{-1}\| = O_p(p/\min[\sqrt{N},\sqrt{T}])$. Thus,

$$\begin{aligned}\left|e'\big(\hat{M}_{ip}^{-1}-\tilde{M}_{ip}^{-1}\big)D_T\sum_{t=p}^T \hat{x}_{it}\Delta\hat{e}_{i,t}\right| &\le \|\hat{M}_{ip}^{-1}-\tilde{M}_{ip}^{-1}\|\cdot\left\|D_T\sum_{t=p}^T\hat{x}_{it}\Delta\hat{e}_{it}\right\| \\ &= O_p\left(\frac{p}{\min[\sqrt{N},\sqrt{T}]}\right)\cdot O_p(1)\cdot\sqrt{p} \\ &= O_p\left(\frac{p^{3/2}}{\min[\sqrt{N},\sqrt{T}]}\right),\end{aligned}$$

which vanishes if $p^3/\min[N,T]\to 0$ as $p,N,T\to\infty$. Next we show the second term is $o_p(1)$.

$$\begin{aligned}e'\tilde{M}_{ip}^{-1}D_T\sum_{t=p}^T(\hat{x}_{it}\Delta\hat{e}_{it}-x_{it}\Delta e_{it}) &= e'\big(\tilde{M}_{ip}^{-1}-M_{ip}^{-1}\big)D_T\sum_{t=p}^T(\hat{x}_{it}\Delta\hat{e}_{it}-x_{it}\Delta e_{it}) \\ &\quad + e'M_{ip}^{-1}D_T\sum_{t=p}^T(\hat{x}_{it}\Delta\hat{e}_{it}-x_{it}\Delta e_{it}).\end{aligned}$$

The first term is $o_p(1)$ because $p^{1/2}\|\tilde{M}_{ip}^{-1}-M_{ip}^{-1}\| = o_p(1)$ and $\|D_T\sum_{t=p}^T(\hat{x}_{it}\Delta\hat{e}_{it}-x_{it}\Delta e_{it})\| = p^{1/2}O_p(1)$. Since

$$e'M_{ip} = \left(\omega_{\epsilon i}^2\int_0^1 W_{\epsilon i}^2(s)\,ds, 0_{1\times p}\right)$$

it follows that

$$e'M_{ip}^{-1}D_T\sum_{t=p}^T(\hat{x}_{it}\Delta\hat{e}_{it}-x_{it}\Delta e_{it}) = \frac{1}{\omega_{\epsilon i}^2\int_0^1 W_{\epsilon i}^2(s)\,ds}\frac{1}{T}\sum_{t=p}^T(\hat{e}_{it-1}\Delta\hat{e}_{it}-e_{it-1}\Delta e_{it}).$$

But $T^{-1}\sum_{t=p}^T(\hat{e}_{it-1}\Delta\hat{e}_{it}-e_{it-1}\Delta e_{it}) = o_p(1)$ by Lemma B.1(iv). Thus, $T(\hat{\delta}_{i0}-\tilde{\delta}_{i0}) = o_p(1)$.

Next we show $\hat{\sigma}_{\epsilon i}^2[\hat{M}_{ip}^{-1}]_{11} - \tilde{\sigma}_{\epsilon i}^2[\tilde{M}_{ip}^{-1}]_{11} = o_p(1)$. But this follows from $\hat{\sigma}_{\epsilon i}^2 - \tilde{\sigma}_{\epsilon i}^2 \xrightarrow{p} 0$, which is easy to verify, and $[\hat{M}_{ip}^{-1}]_{11} - [\tilde{M}_{ip}^{-1}]_{11} = o_p(1)$ by Lemma C.1(ii). In summary, $ADF_{\hat{e}}(i) - ADF_e(i) = o_p(1)$ if $p^{3/2}/\min[\sqrt{N},\sqrt{T}]\to 0$.

LEMMA C.1:

(i) $$\|\hat{M}_{ip}-\tilde{M}_{ip}\| = O_p\left(\frac{p}{\min[\sqrt{N},\sqrt{T}]}\right);$$

(ii) $$\|\hat{M}_{ip}^{-1}-\tilde{M}_{ip}^{-1}\| = O_p\left(\frac{p}{\min[\sqrt{N},\sqrt{T}]}\right).$$

PROOF OF LEMMA C.1(i): From the definition of $\hat{M}_{ip}$ and $\tilde{M}_{ip}$, we have

$$\hat{M}_{ip}-\tilde{M}_{ip} = \begin{bmatrix} T^{-2}\sum_{t=p}^T(\hat{e}_{i,t-1}^2-e_{i,t-1}^2) & T^{-3/2}\sum_{t=p}^T(\hat{e}_{i,t-1}\hat{Z}_{it}'-e_{i,t-1}Z_{it}') \\ T^{-3/2}\sum_{t=p}^T(\hat{e}_{i,t-1}\hat{Z}_{it}-e_{i,t-1}Z_{it}) & T^{-1}\sum_{t=p}^T(\hat{Z}_{it}\hat{Z}_{it}'-Z_{it}Z_{it}')\end{bmatrix}.$$

(a) By Lemma B.1(ii), $(1/T^2)\sum_{t=p}^{T}(\hat{e}_{it-1}^2 - e_{it-1}^2) = O_p(C_{NT}^{-1})$, thus $|(1/T^2)\sum_{t=p}^{T}(\hat{e}_{it-1}^2 - e_{it-1}^2)|^2 = O_p(C_{NT}^{-2})$.

(b) Consider now the upper off-diagonal block of $\hat{M}_{ip} - \tilde{M}_{ip}$:

$$\left\| T^{-3/2}\sum_{t=p}^{T}\hat{e}_{it-1}\hat{Z}_{it}' - e_{it-1}Z_{it}' \right\|^2$$
$$\leq \sum_{j=1}^{p}\left\| T^{-3/2}\sum_{t=p}^{T}\hat{e}_{i,t-1}\Delta\hat{e}_{i,t-j} - e_{i,t-j}\Delta e_{i,t-j}\right\|^2$$
$$\leq 2\sum_{j=1}^{p}\left\| T^{-3/2}\sum_{t=p}^{T}\hat{e}_{i,t-1}(\Delta\hat{e}_{i,t-j} - \Delta e_{i,t-j})\right\|^2 + 2\sum_{j=1}^{p}\left\| T^{-3/2}\sum_{t=p}^{T}(\hat{e}_{i,t-1} - e_{i,t-1})\Delta e_{i,t-j}\right\|^2$$
$$\leq 2\sum_{j=1}^{p}\frac{1}{T^2}\sum_{t=p}^{T}\hat{e}_{i,t-1}^2 \cdot \frac{1}{T}\sum_{t=p}^{T}(\Delta\hat{e}_{i,t-j} - \Delta e_{i,t-j})^2$$
$$+ 2\sum_{j=1}^{p}\frac{1}{T^2}\sum_{t=p}^{T}(\hat{e}_{i,t-1} - e_{i,t-1})^2\frac{1}{T}\sum_{t=p}^{T}(\Delta e_{t-j})^2.$$

Consider first $\hat{\Delta}e_{i,t-j} - \Delta e_{i,t-j} = \lambda_i' H^{-1} v_{t-j} - d_i'\hat{f}_{t-j}$:

$$\frac{1}{T}\sum_{t=p}^{T}(\hat{\Delta}e_{i,t-j} - \Delta e_{i,t-j})^2 \leq \|\lambda_i H^{-1}\|\frac{1}{T}\sum_{t=p}^{T}\|v_{t-j}\|^2 + \|d_i\|^2\frac{1}{T}\sum_{t=p}^{T}\|\hat{f}_{t-j}\|^2$$
$$= O_p(C_{NT}^{-2}) + O_p\left(\frac{1}{\min[T, N^2]}\right)\cdot O_p(1) = O_p(C_{NT}^{-2}).$$

Furthermore, from (15), $\hat{e}_{it} = e_{it} + A_{it}$, where $A_{it} = -e_{i1} - \lambda_i H^{-1}V_t - d_i'\hat{F}_t$. Note that

$$\frac{1}{T^2}\sum_{t=1}^{T}A_{it}^2 \leq \frac{3e_{i1}^2}{T} + \frac{3}{T^2}\sum_{t=1}^{T}\|V_t\|^2\cdot\|\lambda_i H^{-1}\|^2 + \|d_i\|^2\frac{1}{T^2}\sum_{t=1}^{T}\|\hat{F}_t^2\|^2$$
$$= O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{T}\right)O_p\left(\frac{T}{N}\right) + O_p\left(\frac{1}{\min[N^2, T]}\right)\cdot O_p(1) = O_p(C_{NT}^{-2}).$$

Thus, $(1/T^2)\sum_{t=p}^{T}(\hat{e}_{i,t-1} - e_{i,t-1})^2 = (1/T^2)\sum_{t=p}^{T}A_{it-j}^2 \leq (1/T^2)\sum_{t=1}^{T}A_{it}^2 = O_p(C_{NT}^{-2})$. Putting everything together,

$$\left\| T^{-3/2}\sum_{t=p}^{T}\hat{e}_{it-1}\hat{Z}_{it}' - e_{it-1}Z_{it}\right\|^2 = p\cdot O_p(1)O_p(C_{NT}^{-2}) + pO_p(C_{NT}^{-2})\cdot O_p(1)$$
$$= O_p(C_{NT}^{-2})\cdot p.$$

(c) Consider the lower diagonal block $\hat{M}_{ip} - \tilde{M}_{ip}$:

$$\frac{1}{T}\sum_{t=p}^{T}(\hat{Z}_{it}\hat{Z}_{it}' - Z_{it}Z_{it}') = \frac{1}{T}\sum_{t=p}^{T}(\hat{Z}_{it} - Z_{it})(\hat{Z}_{it} - Z_{it})'$$
$$+ \frac{1}{T}\sum_{t=p}^{T}(\hat{Z}_{it} - Z_{it})Z_{it}' + \frac{1}{T}\sum_{t=p}^{T}Z_{it}(\hat{Z}_{it} - Z_{it})',$$

which is dominated by the two cross-product terms. Now

$$\text{(C.3)}\qquad \left\|\frac{1}{T}\sum_{t=p}^{T}(\hat{Z}_{it}-Z_{it})Z_{it}'\right\|^2 \le \frac{1}{T}\sum_{t=p}^{T}\|Z_{it}\|^2\cdot\frac{1}{T}\sum_{t=p}^{T}\|\hat{Z}_{it}-Z_{it}\|^2$$

$$\le \left(\sum_{j=1}^{p}\frac{1}{T}\sum_{t=p}^{T}(\Delta e_{i,t-j})^2\right)\left(\sum_{j=1}^{p}\frac{1}{T}\sum_{t=p}^{T}(\Delta\hat{e}_{i,t-j}-\Delta e_{i,t-j})^2\right)$$

$$= p^2 O_p(C_{NT}^{-2}).$$

Combining parts (a)–(c), $\|\hat{M}_{ip}-\tilde{M}_{ip}\|^2 = p^2O_p(C_{NT}^{-2})$ or $\|\hat{M}_{ip}-\tilde{M}_{ip}\| = pO_p(C_{NT}^{-1})$. *Q.E.D.*

PROOF OF LEMMA C.1(ii): Since

$$\|\hat{M}_{ip}^{-1}-\tilde{M}_{ip}^{-1}\| = \left\|\hat{M}_{ip}^{-1}(\tilde{M}_{ip}-\hat{M}_{ip})\tilde{M}_{ip}^{-1}\right\|$$

$$\le \left[\left\|\hat{M}_{ip}^{-1}-\tilde{M}_{ip}^{-1}\right\|+\left\|\tilde{M}_{ip}^{-1}\right\|\right]\left(\left\|\tilde{M}_{ip}-\hat{M}_{ip}\right\|\left\|\tilde{M}_{ip}^{-1}\right\|\right),$$

it follows that

$$\|\hat{M}_{ip}^{-1}-\tilde{M}_{ip}^{-1}\| \le \frac{\|\tilde{M}_{ip}-\hat{M}_{ip}\|(\|\tilde{M}_{ip}^{-1}\|)^2}{1-\|\tilde{M}_{ip}-\hat{M}_{ip}\|\|\tilde{M}_{ip}^{-1}\|}.$$

Now $\|\tilde{M}_{ip}^{-1}\| \le \|\tilde{M}_{ip}^{-1}-M_{ip}^{-1}\|+\|M_{ip}^{-1}\| = o_p(1)+O_p(1)=O_p(1)$. Using Lemma C.1(i),

$$\|\hat{M}_{ip}^{-1}-\tilde{M}_{ip}^{-1}\| \le O_p(1)\cdot\left\|\hat{M}_{ip}-\tilde{M}_{ip}\right\| = O_p\left(\frac{p}{\min[\sqrt{N},\sqrt{T}\,]}\right).$$

This completes the proof of Theorem 1 Part 1. *Q.E.D.*

The proof of Theorem 1 Part 2 uses the identical argument as Part 1 and thus is omitted. We next consider the proof of Part 3.

D. *The $MQ_f(m)$ Test*

If F_t has m common trends, then any rotation of F_t by a full rank $r\times r$ matrix, H, will also have m common trends. Thus, there exist β and $\beta_\perp$, both $r\times m$, such that $\beta' HF_t$ is stationary, $Y_t=\beta_\perp' HF_t$ is I(1), with $\beta_\perp'\beta=0$. The test MQ_f assumes Y_t has a finite VAR(p) representation as in Stock and Watson (1988). The data are filtered to remove serial correlation. Assume $\Pi(L)\Delta Y_t=\eta_t$ with η_t i.i.d. zero mean and $E(\eta_t\eta_t')=\Sigma_0$, where $\Pi(L)$ is a pth order polynomial in the lag operator. Then $y_t=\Pi(L)\beta_\perp' HF_t$ is an m-vector random walks since $\Delta y_t=\eta_t$. Define

$$\text{(D.1)}\qquad \Phi_f(m)=\frac{1}{2}\sum_{t=1}^{T}(y_ty_{t-1}'+y_{t-1}y_t')\left(\sum_{t=1}^{T}y_{t-1}y_{t-1}'\right)^{-1}.$$

Since $\sum_{t=1}^{T}(\Delta y_ty_{t-1}'+y_{t-1}\Delta y_t')=y_Ty_T'-\sum_{t=1}^{T}\Delta y_t\Delta y_t'$, it follows that

$$T\big(\Phi_f(m)-I_m\big)\Rightarrow\frac{1}{2}[\Sigma_0^{1/2}W_m(1)W_m(1)'\Sigma_0^{1/2}-\Sigma_0]\left[\Sigma_0^{1/2}\int_0^1W_m(s)W_m(s)'\,ds\,\Sigma_0^{1/2}\right]^{-1}.$$

The eigenvalues of the right-hand side are the same as those of (matrix A has the same eigenvalues as BAB^{-1})

$$\Phi_* = \frac{1}{2}[W_m(1)W_m(1)' - I_m]\left[\int_0^1 W_m(s)W_m(s)'\,ds\right]^{-1}. \tag{D.2}$$

Let ν^* and ν_f be the eigenvalues of Φ_* and $\Phi_f(m)$, respectively, with $\nu^*(j)$ and $\nu_f(j)$ being the jth ordered (from largest to smallest) element. Then $MQ_f(m) = T\cdot(\nu_f(m) - 1) \xrightarrow{d} \nu^*(m)$ under the null of m unit roots and the statistic diverges to $-\infty$ under the alternative of $m-1$ unit roots, as in the Q_f of Stock and Watson (1988).

Even if F_t was observed, the $MQ_f(m)$ is still not feasible because $\beta_\perp$ and $\Pi(L)$ are not observed. Suppose (i) $\tilde\beta_\perp$ consistently estimates the space spanned by $\beta_\perp$, i.e., $\tilde\beta_\perp \xrightarrow{p} \beta_\perp C_1'$ for some $m\times m$ matrix C_1; and (ii) $\tilde\Pi(L)$ is an estimate of $\Pi(L)$ satisfying $\tilde\Pi(L) \xrightarrow{p} C_1\Pi(L)C_1^{-1}$. By the result of Stock and Watson (1988), if $\tilde Y_t = \tilde\beta_\perp' HF_t$ and $\tilde y_t = \tilde\Pi(L)\tilde Y_t$, then with $\dot C_1 = (C_1\Sigma_0 C_1')^{1/2}$,

$$T\big(\tilde\Phi_f(m) - I_m\big) \Rightarrow \dot C_1 \Phi_* \dot C_1^{-1},$$

where $\tilde\Phi_f(m)$ is as defined in (D.1) with y_t replaced by $\tilde y_t$. Because Φ_* and $\dot C_1\Phi_*\dot C_1^{-1}$ have the same eigenvalues, the limiting distribution of $T[\tilde\nu_f(m) - 1]$ is equal to $\nu^*(m)$. The test is also valid for an intercept and/or a linear trend, with the obvious replacement of the vector Brownian motion by its demeaned and detrended counterpart. In Stock and Watson's implementation, $\tilde\beta_\perp$ is the matrix of eigenvectors associated with the largest m eigenvalues of $T^{-2}H\sum_{t=1}^T F_tF_t'H'$. Stock and Watson (1988) and Harris (1997) proved that the method of principal components consistently estimates the space spanned by $\beta_\perp$, i.e., $\tilde\beta_\perp \xrightarrow{p} \beta_\perp C_1'$ for some C_1. Furthermore, $\tilde\Pi(L)$ obtained by regressing $\Delta\tilde Y_t$ on lags of $\Delta\tilde Y_t$ consistently estimates $\Pi(L)$ as proved by Stock and Watson (1988).

Since F_t is not observed in our setting, our proposed test is based upon $\hat y_t = \hat\Pi(L)\hat Y_t$, where $\hat Y_t = \hat\beta_\perp'\hat F_t$. We need to show that (a) $\hat\beta_\perp$ obtained by applying the method of principal components to $\hat F_t$ satisfies $\hat\beta_\perp \xrightarrow{p} \beta_\perp C'$ for some matrix C; (b) $\hat\Pi(L)$ obtained from regressing $\Delta\hat Y_t$ on its lags is such that $\hat\Pi(L) \xrightarrow{p} C\Pi(L)C^{-1}$; (c) $\bar y_t = \hat\Pi(L)\hat\beta_\perp' HF_t$, and $T(\bar\Phi_f(m) - I_m) \to C\Phi_* C^{-1}$; and (d) $T(\hat\Phi_f(m) - \bar\Phi_f(m)) = o_p(1)$ so that $T(\hat\Phi_f(m) - I_m) \to \dot C\Phi_*\dot C^{-1}$, where $\dot C = (C\Sigma_0 C')^{1/2}$, and $\hat\Phi_f(m)$ and $\bar\Phi_f(m)$ are defined as in (D.1) but with y_t replaced by $\hat y_t$ and $\bar y_t$, respectively.

We begin with (a). Lemma B.2(ii) and continuity of the eigenvector space imply $\hat\beta_\perp - \tilde\beta_\perp C_2' \xrightarrow{p} 0$ for some invertible C_2. But $\tilde\beta_\perp \xrightarrow{p} \beta_\perp C_1'$. Let $C = C_2C_1$; then

$$\hat\beta_\perp \xrightarrow{p} \beta_\perp C_1'C_2' = \beta_\perp C'.$$

To show (b), note first that by definition, $\hat Y_t = \hat\beta_\perp'\hat F_t$, $Y_t = \beta_\perp' HF_t$, and $\hat F_t = HF_t - HF_1 + V_t$. Thus,

$$\hat Y_t = \hat\beta_\perp' HF_t - \hat\beta_\perp' HF_1 + \hat\beta_\perp' V_t = CY_t + (\hat\beta_\perp' - C\beta_\perp')HF_t - \hat\beta_\perp' HF_1 + \hat\beta_\perp' V_t,$$

$$\Delta\hat Y_t = C\Delta Y_t + (\hat\beta_\perp' - C\beta_\perp')H\Delta F_t + \hat\beta_\perp' v_t.$$

From $\hat\beta_\perp - \beta_\perp C' \xrightarrow{p} 0$ and $\|v_t\| = o_p(1)$ by Lemma 1(b) we have $\Delta\tilde Y_t = C\Delta Y_t + o_p(1)$. If $\Pi(L)\Delta Y_t =$ error, then estimation of a VAR in $\Delta\hat Y_t$ yields $\hat\Pi(L)$, with $\hat\Pi(L)\Delta\hat Y_t =$ error. Since $\Delta\hat Y_t = C\Delta Y_t + o_p(1)$, $\hat\Pi(L) \xrightarrow{p} C\Pi(L)C^{-1}$.

For (c), let $\bar y_t = \hat\Pi(L)\hat\beta_\perp' HF_t$, and $\bar\Phi_f(m) = (\sum_{t=2}^T[\bar y_t\bar y_{t-1}' + \bar y_{t-1}\bar y_t'])(\sum_{t=2}^T \bar y_{t-1}\bar y_{t-1}')^{-1}$. Then by the argument of Stock and Watson (1988), $T(\bar\Phi_f(m) - I_m) \Rightarrow \dot C\Phi_*\dot C^{-1}$ and thus $T(\bar\nu_f(m) - 1) \xrightarrow{p} \nu^*(m)$.

For (d), now $\hat{y}_t = \hat{\Pi}(L)\hat{\beta}'_\perp \hat{F}_t$, we will show $T[\hat{\Phi}_f(m) - \bar{\Phi}_f(m)] = o_p(1)$. Lemma B.2(ii) implies $(1/T^2)\sum_{t=2}^T \hat{y}_t\hat{y}'_t = (1/T^2)\sum_{t=2}^T \bar{y}_t\bar{y}'_t + o_p(1)$. It is sufficient to consider the numerator of $\hat{\Phi}_f(m)$ and $\bar{\Phi}_f(m)$. Since $\hat{\Pi}(L) = \hat{\Pi}_0 + \hat{\Pi}_1 L + \cdots + \hat{\Pi}_p L^p$, the numerator of $T(\hat{\Phi}_f(m) - I)$ is equal to

$$\frac{1}{T}\sum_{t=2}^{T}[\Delta\hat{y}_t\hat{y}'_{t-1} + \hat{y}_{t-1}\Delta\hat{y}'_t] = \sum_{j=0}^{p}\sum_{k=0}^{p}\hat{\Pi}_j\hat{\beta}'_\perp\left[\frac{1}{T}\sum_{t=2}^{T}\Delta\hat{F}_{t-j}\hat{F}'_{t-1-k} + \hat{F}_{t-1-j}\hat{\Delta}F'_{t-k}\right]\hat{\beta}_\perp\hat{\Pi}'_k,$$

and the numerator of $T(\bar{\Phi}_f(m) - I)$ is

$$\begin{aligned}&\frac{1}{T}\sum_{t=2}^{T}[\Delta\bar{y}_t\bar{y}'_{t-1} + \bar{y}_{t-1}\Delta\bar{y}'_t]\\&\quad= \sum_{j=0}^{p}\sum_{k=0}^{p}\hat{\Pi}_j\hat{\beta}'_\perp H\left[\frac{1}{T}\sum_{t=2}^{T}\Delta F_{t-j}F'_{t-1-k} + F_{t-1-j}\Delta F'_{t-k}\right]H'\hat{\beta}_\perp\hat{\Pi}'_k.\end{aligned}$$

Lemma D.1 implies that the difference of the two numerators is $o_p(1)$. Thus, $T(\hat{\Phi}_f(m) - \bar{\Phi}_f(m)) = o_p(1)$. Combining (a)–(d), we have $T(\hat{\nu}_f(m) - \bar{\nu}_f(m)) = o_p(1)$, or $T(\hat{\nu}_f(m) - 1) \xrightarrow{d} \nu^*(m)$.

LEMMA D.1: *For all $j, k \geq 0$ and $j, k \leq p$, as $N, T \to \infty$,*

$$\begin{aligned}&T^{-1}\sum_{t=2}^{T}(\Delta\hat{F}_{t-j}\hat{F}'_{t-1-k} + \hat{F}_{t-1-j}\Delta\hat{F}'_{t-k})\\&\quad - T^{-1}H\sum_{t=2}^{T}(\Delta F_{t-j}F'_{t-1-k} + F_{t-1-j}\Delta F'_{t-k})H' \xrightarrow{p} 0.\end{aligned}$$

PROOF: When $j = k = 0$, the lemma is implied by Lemma B.2(iv). For all fixed j, k, by adding and subtracting terms, the above can be turned into the case of $j = k = 0$ plus terms that are $o_p(1)$. For example, when $j = 1$ and $k = 0$,

$$\begin{aligned}&\frac{1}{T}\sum_{t=2}^{T}(\Delta\hat{F}_{t-1}\hat{F}'_{t-1} + \hat{F}_{t-2}\Delta\hat{F}'_t)\\&\quad= \frac{1}{T}\sum_{t=2}^{T}(\Delta\hat{F}_{t-1}\hat{F}'_{t-2} + \hat{F}_{t-1}\Delta\hat{F}'_t) + \frac{1}{T}\sum_{t=2}^{T}\Delta\hat{F}_{t-1}\Delta\hat{F}'_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\Delta\hat{F}_{t-1}\Delta\hat{F}'_t\\&\quad= \frac{1}{T}\sum_{t=2}^{T}(\Delta\hat{F}_t\hat{F}'_{t-1} + \hat{F}_{t-1}\Delta\hat{F}'_t) - \left(\frac{\Delta\hat{F}_T\hat{F}_{T-1}}{T}\right)\\&\qquad + \frac{1}{T}\sum_{t=2}^{T}\Delta\hat{F}_{t-1}\Delta\hat{F}'_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\Delta\hat{F}_{t-1}\Delta\hat{F}'_t.\end{aligned}$$

A similar identify holds for $(1/T)\sum_{t=2}^T(\Delta F_{t-1}F'_{t-1} + F_{t-2}\Delta F'_t)$. The first term on the right-hand side above corresponds to the case of $j = k = 0$. The remaining terms, after subtracting the corresponding terms from $H(1/T)\sum_{t=2}^T[\Delta F_{t-1}F_{t-1} + F_{t-2}\Delta F_{t-1}]H'$, are each $o_p(1)$. *Q.E.D.*

REMARK 1: The validity of the MQ_f test using $\hat{F}_t$ relies on the closeness of $\hat{F}_t$ to HF_t. Lemma B.2 shows that $\hat{F}_t^c$ is close to HF_t^c (demeaned series), and Lemma G.3 shows that $\hat{F}_t^\tau$ is close to HF_t^τ (detrended series). Using analogous arguments, the MQ_f test is also valid for demeaned and detrended $\hat{F}_t$, with the obvious replacement of W_m by W_m^c or W_m^τ. Details are omitted.

E. *The $MQ_c(m)$ Test*

By definition, $\hat{Y}_t = \hat{\beta}_\perp' \hat{F}_t$, where $\hat{\beta}_\perp$ are eigenvectors corresponding to the m largest eigenvalue of $T^{-2}\sum_{t=2}^T \hat{F}_t\hat{F}_t'$. Also recall

$$\hat{Y}_t = CY_t + (\hat{\beta}_\perp - \beta_\perp C)'HF_t - \hat{\beta}_\perp' HF_1 + \hat{\beta}_\perp' V_t,$$

$$\Delta\hat{Y}_t = C\Delta Y_t + (\hat{\beta}_\perp - \beta_\perp C)'Hf_t + \hat{\beta}_\perp' v_t.$$

Let $\xi_t = \Delta Y_t$; then ξ_t is a linear process of the i.i.d. sequence of u_t. Let Ω denote the long-run variance of ξ_t. It is given by $\Omega = \Omega_0 + \Omega_1 + \Omega_1'$, where $\Omega_0 = E\xi_t\xi_t'$ and $\Omega_1 = \sum_{j=1}^\infty E(\xi_0\xi_j')$. Then under the null hypothesis that Y_t has m unit roots,

$$\sum_{t=2}^T Y_{t-1}Y_{t-1}'/T^2 \xrightarrow{d} \Omega^{1/2}\int_0^1 W_m(r)W_m(r)'\,dr\,\Omega^{1/2},$$

denoted by Ξ. In addition,

$$\frac{1}{2}\sum_{t=2}^T (Y_{t-1}\Delta Y_t' + \Delta Y_t' Y_{t-1})/T \xrightarrow{d} \frac{1}{2}[\Omega^{1/2}W_m(1)W_m(1)'\Omega^{1/2} - \Omega_0],$$

denoted by Υ. Since $\Omega_0 \neq \Omega$ for serially correlated ξ_t, the eigenvalues of $\Upsilon\Xi^{-1}$ are not invariant to the nuisance parameter Ω. However, if $\Omega_1 + \Omega_1'$ is subtracted from the expression, so that

$$\frac{1}{2}T^{-1}\sum_{t=2}^T (Y_{t-1}\Delta Y_t' + \Delta Y_t' Y_{t-1} - \Omega_1 - \Omega_1') \Rightarrow \frac{1}{2}[\Omega^{1/2}W_m(1)W_m(1)'\Omega^{1/2} - \Omega] = \Upsilon_1,$$

say, then the eigenvalues of $\Upsilon_1\Xi^{-1}$ do not depend on Ω and are the same as those of Φ_* defined in (D.2).

Because we do not observe ΔY_t, but $\Delta\hat{Y}_t$, which is an estimate of $C\Delta Y_t = C\xi_t$. The long-run variance of $C\xi_t$ is $\Sigma = C\Omega C'$. Let $\Sigma_0 = C\Omega_0 C'$ and $\Sigma_1 = C\Omega_1 C'$. Lemma B.2 (ii) implies $\sum_{t=2}^T \hat{Y}_t\hat{Y}_t'/T^2 \xrightarrow{d} \Sigma^{1/2}\int W_m(r)W_m(r)'\,dr\,\Sigma^{1/2} = \Xi^\dagger$, say, and Lemma B.2(iv) implies $\frac{1}{2}T^{-1}\sum_{t=2}^T(\hat{Y}_{t-1}\Delta\hat{Y}_t' + \Delta\hat{Y}_t'\hat{Y}_{t-1}) \Rightarrow \frac{1}{2}[\Sigma^{1/2}W_m(1)W_m(1)'\Sigma^{1/2} - \Sigma_0] = \Upsilon^\dagger$, say. Again the eigenvalues of $\Upsilon^\dagger\Xi^{\dagger-1}$ depend on nuisance parameters. Let $\hat{\Sigma}_1$ be a consistent estimator of Σ_1. Then

$$\frac{1}{2}\frac{1}{T}\sum_{t=2}^T (\hat{Y}_{t-1}\Delta\hat{Y}_t' + \Delta\hat{Y}_t'\hat{Y}_{t-1} - \hat{\Sigma}_1 - \hat{\Sigma}_1') \Rightarrow \frac{1}{2}[\Sigma^{1/2}W_m(1)W_m(1)'\Sigma^{1/2} - \Sigma].$$

Denote the limit by $\Upsilon_1^\dagger$; then $\Upsilon_1^\dagger\Xi^{\dagger-1}$ have the same eigenvalues as Φ_*.

The objective is to show that Σ_1 is consistently estimable. Or equivalently, $\Sigma = C\Omega C'$ is consistently estimable. Consider the regression of $\hat{Y}_t$ on $\hat{Y}_{t-1}$, and let $\hat{B}$ be the estimated coefficient matrix. Note that $T(\hat{B} - I) = O_p(1)$. From $\hat{\xi}_t = \hat{Y}_t - \hat{B}\hat{Y}_{t-1} = \Delta\hat{Y}_t + (I - \hat{B})\hat{Y}_{t-1}$, we have

$$\hat{\xi}_t = C\xi_t + (\hat{\beta}_\perp - \beta_\perp C)'Hf_t + \hat{\beta}_\perp' v_t + (I - \hat{B})\hat{Y}_{t-1},$$

where $\Delta Y_t = \xi_t$. Letting $w_t = \hat{\beta}_\perp' v_t + (I - \hat{B})\hat{Y}_{t-1}$, we have

$$\hat{\xi}_t = C\xi_t + (\hat{\beta}_\perp - \beta_\perp C)'Hf_t + w_t. \tag{E.1}$$

For arbitrary time series a_t and b_t, define $\tilde{\Gamma}_{ab}(j) = T^{-1}\sum_{t=1}^{T-j} a_t b'_{t+j}$ and let

$$\tilde{M}_{ab} = \tilde{\Gamma}_{ab}(0) + \sum_{j=1}^{J} K(j)[\tilde{\Gamma}_{ab}(j) + \tilde{\Gamma}_{ab}(j)'].$$

We next show $\tilde{M}_{\hat{\xi}\hat{\xi}}$ is consistent for $\Sigma = C\Omega C'$. By (E.1)

$$\begin{aligned}\tilde{M}_{\hat{\xi}\hat{\xi}} - C\tilde{M}_{\xi\xi}C' &= C\tilde{M}_{\xi f}H'(\hat{\beta}_\perp - \beta_\perp C) + (\hat{\beta}_\perp - \beta_\perp C)'H\tilde{M}_{f\xi}C' \\ &\quad + (\hat{\beta}_\perp - \beta_\perp C)'H\tilde{M}_{ff}H'(\hat{\beta}_\perp - \beta_\perp C) + C\tilde{M}_{\xi w} + \tilde{M}_{w\xi}C' \\ &\quad + (\hat{\beta}_\perp - \beta_\perp C)'H\tilde{M}_{fw} + \tilde{M}_{wf}H'(\hat{\beta}_\perp - \beta_\perp C) + \tilde{M}_{ww}.\end{aligned}$$

From $\hat{\beta}_\perp - \beta_\perp C \xrightarrow{p} 0$, $\tilde{M}_{\xi f} = O_p(1)$, and $\tilde{M}_{ff} = O_p(1)$, the first three terms converge to zero. We now show that $\tilde{M}_{\xi w} \xrightarrow{p} 0$, $\tilde{M}_{fw} \xrightarrow{p} 0$, and $\tilde{M}_{ww} \xrightarrow{p} 0$ if $J/\min[\sqrt{T}, \sqrt{N}] \to 0$. We have

$$\begin{aligned}\|\tilde{M}_{\xi w}\| &\le \sum_{j=0}^{J} |K(j)| \Bigg[\left(\frac{1}{T}\sum_{t=1}^{T-j} \|\xi_t\|^2\right)^{1/2} \left(\frac{1}{T}\sum_{t=1}^{T-j} \|w\|_{t+j}^2\right)^{1/2} \\ &\qquad + \left(\frac{1}{T}\sum_{t=1}^{T-j} \|\xi_{t+j}\|^2\right)^{1/2} \left(\frac{1}{T}\sum_{t=1}^{T-j} \|w_t\|^2\right)^{1/2} \Bigg].\end{aligned}$$

From $w_t = \hat{\beta}'_\perp v_t + (I - B)\hat{Y}_{t-1}$, we have $\|w_t\|^2 \le 2\|\hat{\beta}_\perp\|^2 \cdot \|v_t\|^2 + 2\|I - \hat{B}\|^2 \cdot \|\hat{Y}_{t-1}\|^2$, and

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} \|w_t\|^2 &\le 2\|\hat{\beta}_\perp\|^2 \cdot \frac{1}{T}\sum_{t=1}^{T} \|v_t\|^2 + 2O_p(T^{-2}) \cdot \frac{1}{T}\sum_{t=1}^{T} \|\hat{Y}_{t-1}\|^2 \\ &= O_p(C_{NT}^{-2}) + O_p\left(\frac{1}{T}\right) \cdot \frac{1}{T^2}\sum_{t=1}^{T} \|\hat{Y}_{t-1}\|^2 \\ &= O_p(C_{NT}^{-2}) + O_p\left(\frac{1}{T}\right) = O_p(C_{NT}^{-2}).\end{aligned}$$

Thus, using $(1/T)\sum_{t=1}^{T} \|\xi_t\|^2 = O_p(1)$,

$$\begin{aligned}\|\tilde{M}_{\xi w}\| &\le (J+1) \cdot \left[O_p(1)\left(\frac{1}{T}\sum_{t=1}^{T} \|\xi_t\|^2\right)^{1/2} + O_p(1)\left(\frac{1}{T}\sum_{t=1}^{T} \|\xi_t\|^2\right)^{1/2} \right] \cdot O_p(C_{NT}^{-1}) \\ &= (J+1) \cdot O_p(C_{NT}^{-1}),\end{aligned}$$

which converges to zero if $J/\min[\sqrt{N}, \sqrt{T}] \to 0$. Similarly, $\|\tilde{M}_{fw}\| \le (J+1)O_p(C_{NT}^{-1}) \xrightarrow{p} 0$. Next

$$\begin{aligned}\|\tilde{M}_{ww}\| &\le \sum_{j=0}^{J} |K(j)| \left[2\left(\frac{1}{T}\sum_{t=1}^{T-j} \|w_t\|^2\right)^{1/2} \cdot \left(\frac{1}{T}\sum_{t=1}^{T-j} \|w_{t+j}\|^2\right)^{1/2} \right] \\ &\le (J+1) \cdot O_p(C_{NT}^{-2}) \xrightarrow{p} 0.\end{aligned}$$

The above analysis shows $\tilde{M}_{\hat{\xi}\hat{\xi}} - C\tilde{M}_{\xi\xi}C' = o_p(1)$. Since $\tilde{M}_{\xi\xi} \xrightarrow{p} \Omega$ by Newey and West (1987), we have $\tilde{M}_{\hat{\xi}\hat{\xi}} \xrightarrow{p} C\Omega C' = \Sigma$.

F. *Proof of Theorem 2, Consistency*

PROOF OF THEOREM 2: Consider the regression in (C.1): $\Delta e_{it} = \delta_{i0} e_{i,t-1} + \sum_{j=1}^{p} \delta_{ij} \Delta e_{i,t-j} + \epsilon_{i,tp}$. Under the alternative hypothesis, $\delta_{i0} < 0$. Let $e_i = (e_{ip+1}, \dots, e_{iT})'$, $e_{i_{-1}} = (e_{ip}, \dots, e_{iT-1})'$, $\dots, e_{i_{-p-1}} = (e_{i0}, e_{i1}, \dots, e_{iT-p-1})'$. Let $\Delta e_i = e_i - e_{i_{-1}}, \dots, \Delta e_{i_{-p}} = e_{i_{-p}} - e_{i_{-p-1}}$. Finally let $Z_i = (\Delta e_{i_{-1}}, \dots, \Delta e_{i_{-p}})$ and $M_{i,z} = I - Z_i(Z_i'Z)^{-1}Z_i'$. Define $\hat{M}_{i,z}$, $\hat{e}_{i_{-1}}$, and $\Delta \hat{e}_i$ analogously, with $\hat{e}_{it}$ in place of e_{it}. The least squares estimator of δ_{i0} is $\tilde{\delta}_{i0} = (e_{i_{-1}} M_{i,z} e_{i_{-1}})^{-1}(e_{i_{-1}} M_{i,z} \Delta e_i)$. Let $\hat{\delta}_{i0}$ be the counterpart using $\hat{e}_{it}$. Then

$$\hat{\delta}_{i0} = \frac{\hat{e}'_{i_{-1}} \hat{M}_{i,z} \Delta \hat{e}_i}{\hat{e}'_{i_{-1}} \hat{M}_{i,z} \hat{e}_{i_{-1}}}$$

$$= \frac{\tilde{\delta}_{i0} \frac{1}{T}[e'_{i_{-1}} M_{i,z} e_{i_{-1}}] + \frac{1}{T}[\hat{e}'_{i_{-1}} \hat{M}_{i,z} \Delta \hat{e}_i - e'_{i_{-1}} M_{i,z} \Delta e_i]}{\frac{1}{T} \hat{e}'_{i_{-1}} \hat{M}_{i,z} \hat{e}_{i_{-1}}} \tag{F.1}$$

$$= \frac{\tilde{\delta}_{i0}(\frac{1}{T} e'_{i_{-1}} M_{i,z} e_{i_{-1}}) + o_p(1)}{\frac{1}{T} \hat{e}'_{i_{-1}} \hat{M}_{i,z} \hat{e}_{i_{-1}}}, \tag{F.2}$$

where the first equality follows from $(e_{i_{-1}} M_{i,z} \Delta e_i) = \tilde{\delta}_{i0}(e'_{i_{-1}} M_{i,z} e_{i_{-1}})$ and the second equality follows from Lemma F.2. Now $[e'_{i_{-1}} M_{i,z} e_{i_{-1}}/T]$ converges to a positive constant, and $\tilde{\delta}_{i0} \xrightarrow{p} \delta_{i0} < 0$ under the alternative. So the numerator converges to a negative number. The objective here is to show that the ADF diverges under the alternative. The ADF is

$$ADF_{\hat{e}}(i) = \frac{\hat{\delta}_{i0}}{(\hat{\sigma}^2_{\epsilon i}(\hat{e}'_{i_{-1}} M_{i,z} \hat{e}_{i_{-1}})^{-1})^{1/2}},$$

where $\hat{\sigma}_{\epsilon i}$ is the sum of squared residuals divided by $T - p$. We simply note that $\hat{\sigma}^2_{\epsilon i}$ is bounded because $\hat{\sigma}^2_{\epsilon i} \le (1/(T-p)) \sum_{t=p+1}^{T} \Delta e^2_{it} = O_p(1)$. Now

$$T\hat{\delta}_{i0} = \frac{T(\tilde{\delta}_{i0}[\frac{1}{T} \hat{e}'_{i_{-1}} M_{i,z} \hat{e}_{i_{-1}}] + o_p(1))}{\frac{1}{T} \hat{e}'_{i_{-1}} \hat{M}_{i,z} \hat{e}_{i_{-1}}},$$

$$ADF_{\hat{e}}(i) = \frac{\sqrt{T}(\tilde{\delta}_{i0}[\frac{1}{T} e'_{i_{-1}} M_{i,z} e_{i_{-1}}] + o_p(1))}{(\hat{\sigma}^2_{\epsilon i}(\frac{1}{T} \hat{e}'_{i_{-1}} M_{i,z} \hat{e}_{i_{-1}}))^{1/2}}.$$

Consider two cases:

(a) If T/N is bounded, then $\hat{e}'_{i_{-1}} M_{i,z} \hat{e}_{i_{-1}}/T = O_p(1)$ by Lemma F.1, so $T\hat{\delta}_{i0} = O_p(T) \to -\infty$ and $ADF_{\hat{e}}(i) = O_p(\sqrt{T}) \to -\infty$.

(b) If $T/N \to \infty$, then by Lemma F.1

$$T\hat{\delta}_{i0} \le \frac{N(\tilde{\delta}_{i0}[\frac{1}{T} e'_{i_{-1}} M_{i,z} e_{i_{-1}}] + o_p(1))}{O_p(\frac{N}{T}) + O_p(1)} = N\tilde{\delta}_{i0} \cdot O_p(1) = O_p(N) \to -\infty.$$

Similarly, $ADF_{\hat{e}}(i) = O_p(\sqrt{N}) \to -\infty$.
In summary, $ADF_{\hat{e}}(i) = O_p(\min[\sqrt{N}, \sqrt{T}])$ under the alternative $|\rho_i| < 1$. *Q.E.D.*

LEMMA F.1: $(1/T)\hat{e}'_{i_{-1}} M_{i,z} \hat{e}_{i_{-1}} \le O_p(1) + O_p(T/N)$.

PROOF: Note $\hat{e}_{i_{-1}}'\hat{M}_{i,z}\hat{e}_{i_{-1}} \le \hat{e}_{i_{-1}}'\hat{e}_{i_{-1}}$. By (15), $\hat{e}_{i_{-1}} = e_{i_{-1}} - \iota e_{i1} - VH'^{-1}\lambda_i - \hat{F}d_i$, where $\iota = (1,1,\dots,1)'$, $V = (V_p, V_{p+1},\dots,V_{T-1})'$, and $\hat{F} = (\hat{F}_p,\dots,\hat{F}_{T-1})'$. Thus,

$$\hat{e}_{i_{-1}}'\hat{e}_{i_{-1}} \le 4\big[e_{i_{-1}}'e_{i_{-1}} + Te_{i1}^2 + \lambda_i'H^{-1}V'VH^{-1\prime}\lambda_i + d_i'\hat{F}'\hat{F}d_i\big],$$

$$\frac{1}{T}\hat{e}_{i_{-1}}'\hat{e}_{i_{-1}} \le 4\frac{1}{T}e_{i_{-1}}'e_{i_{-1}} + 4e_{i1}^2 + 4\|\lambda_iH^{-1}\|^2\frac{1}{T}\sum_{t=1}^T\|V_t\|^2 + 4\|d_i\|^2\frac{1}{T}\sum_{t=1}^T\|\hat{F}_t\|^2.$$

Now $e_{i_{-1}}'e_{i_{-1}}/T = O_p(1)$ and $\sum_{t=1}^T\|V_t\|^2/T = O_p(T/N)$; see (A.4). Since $\|d_i\|^2 = O_p(1/\min[T, N^2])$ and $\sum_{t=1}^T\|F_t\|^2/T \le O_p(T)$, $\|d_i\|^2\sum_{t=1}^T\|\hat{F}_t\|^2/T \le O_p(T/\min[T,N^2]) \le O_p(1) + O_p(T/N^2)$. *Q.E.D.*

LEMMA F.2: $(1/T)\hat{e}_{i_{-1}}'\hat{M}_{i,z}\Delta\hat{e}_i - \frac{1}{T}e_{i_{-1}}'M_{i,z}\Delta e_i = O_p(p^2/(\min[\sqrt{N},\sqrt{T}]))$.

PROOF: $(1/T)\hat{e}_{i_{-1}}'\hat{M}_{i,z}\Delta\hat{e}_i = (1/T)\hat{e}_{i_{-1}}'\Delta\hat{e}_i - (1/T)\hat{e}_{i_{-1}}'\hat{Z}_i(\hat{Z}_i'\hat{Z}_i/T)^{-1}\hat{Z}_i'\Delta\hat{e}_i/T$. Similarly, $(1/T)e_{i_{-1}}'M_{i,z}\Delta e_i = (1/T)e_{i_{-1}}'\Delta e_i - (1/T)e_{i_{-1}}'Z_i(Z_i'Z_i/T)^{-1}Z_i'\Delta e_i/T$. By Lemma B.1(iv) $(1/T)\hat{e}_{i_{-1}}'\Delta\hat{e}_i - (1/T)e_{i_{-1}}'\Delta e_i = O_p(C_{NT}^{-1})$. Thus, it suffices to show

$$\frac{1}{T}\hat{e}_{i_{-1}}'\hat{Z}_i\left(\frac{1}{T}\hat{Z}_i'\hat{Z}_i\right)^{-1}\frac{1}{T}\hat{Z}_i'\Delta\hat{e}_i - \frac{1}{T}e_{i_{-1}}'Z_i\left(\frac{1}{T}Z_i'Z_i\right)^{-1}\frac{1}{T}Z_i'\Delta e_i = O_p(p^2C_{NT}^{-1}).$$

The above can be written as

$$\begin{aligned}&\frac{1}{T}(\hat{e}_{i_{-1}}'\hat{Z}_i - e_{i_{-1}}'Z_i)\left(\frac{1}{T}\hat{Z}_i'\hat{Z}_i\right)^{-1}\frac{1}{T}\hat{Z}_i'\Delta\hat{e}_i\\ &\quad + \frac{1}{T}e_{i_{-1}}'Z_i\left[\left(\frac{1}{T}\hat{Z}_i'\hat{Z}_i\right)^{-1} - \left(\frac{1}{T}Z_i'Z_i\right)^{-1}\right]\frac{1}{T}\hat{Z}_i'\Delta\hat{e}_i\\ &\quad + \frac{1}{T}e_{i_{-1}}'Z_i\left(\frac{1}{T}Z_i'Z_i\right)^{-1}\frac{1}{T}(\hat{Z}_i'\Delta\hat{e}_i - Z_i'\Delta e_i) = (a)+(b)+(c).\end{aligned}$$

Consider (a). Let $\xi_i = (\hat{Z}_i'\hat{Z}_i/T)^{-1}\hat{Z}_i'\Delta\hat{e}_i/T$. Then ξ_i is $p\times 1$ and $\|\xi_i\| = O_p(p^{1/2})$. Next,

$$\begin{aligned}\frac{1}{T}(\hat{e}_{i_{-1}}'\hat{Z}_i - e_{i_{-1}}'Z_i) = \Bigg(&\frac{1}{T}\sum_{t=p+1}^T(\hat{e}_{it-1}\Delta\hat{e}_{it-1} - e_{it-1}\Delta e_{it-1}),\\ &\dots,\frac{1}{T}\sum_{t=p+1}^T(\hat{e}_{it-1}\Delta\hat{e}_{it-p} - e_{it-1}\Delta e_{it-p})\Bigg).\end{aligned}$$

Thus $\|(\hat{e}_{i_{-1}}'\hat{Z}_i - e_{i_{-1}}'Z_i)/T\|^2 = \sum_{k=1}^p(\sum_{t=p+1}^T[\hat{e}_{it-1}\Delta\hat{e}_{it-k} - e_{it-1}\Delta e_{it-k}]/T)^2$. But for each $k\ge 1$, from $\hat{e}_{it-1} = \hat{e}_{it-k-1} + \Delta\hat{e}_{it-k} + \cdots + \Delta\hat{e}_{it-1}$, it follows that

$$\begin{aligned}\frac{1}{T}\sum_{t=p+1}^T(\hat{e}_{it-1}\Delta\hat{e}_{it-k} - e_{it-1}\Delta e_{it-k}) &= \frac{1}{T}\sum_{t=p+1}^T(\hat{e}_{it-k-1}\Delta\hat{e}_{it-k} - e_{it-k-1}\Delta e_{it-k})\\ &\quad + \sum_{h=1}^k\frac{1}{T}\sum_{t=p+1}^T(\Delta\hat{e}_{it-h}\Delta\hat{e}_{it-k} - \Delta e_{it-h}\Delta e_{it-k}).\end{aligned}$$

The first term on the right-hand side is $O_p(C_{NT}^{-1})$ by Lemma B.1(iv). The second term is $kO_p(C_{NT}^{-1})$ following the argument in proving Lemma B.1(iii). Thus

$$\frac{1}{T}\sum_{t=p+1}^{T}(\hat{e}_{it-1}\Delta\hat{e}_{it-k}-e_{it-1}\Delta e_{it-k})=(k+1)O_p(C_{NT}^{-1}),$$

and $\|T^{-1}(\hat{e}_{i_{-1}}'\hat{Z}_i-e_{i_{-1}}'Z_i)\|^2=\sum_{k=1}^{p}(k+1)^2O_p(C_{NT}^{-2})=p^3O_p(C_{NT}^{-2})$. So $\|a\|\le\|T^{-1}(\hat{e}_{i_{-1}}'\hat{Z}_i-e_{i_{-1}}'Z_i)\|\|\xi_i\|=p^{3/2}O_p(C_{NT}^{-1})\sqrt{p}=p^2O_p(C_{NT}^{-1})$.

Consider (b), which we can write as $\eta_i'(T^{-1}Z_i'Z_i-T^{-1}\hat{Z}_i'\hat{Z}_i)\xi_i$, where $\eta_i=T^{-1}e_{i_{-1}}'Z_i(Z_i'Z_i/T)^{-1}$ and ξ_i is defined earlier. Note that $\|\eta_i\|=O_p(\sqrt{p})$. It is proved in (C.3) that $\|T^{-1}Z_i'Z_i-T^{-1}\hat{Z}_i'\hat{Z}_i\|=pO_p(C_{NT}^{-1})$. Thus $\|b\|=p^2O_p(C_{NT}^{-1})$. Next consider (c), which is equal to $\eta_iT^{-1}(\hat{Z}_i'\Delta\hat{e}_i-Z_i'\Delta e_i)=pO_p(C_{NT}^{-1})$ because $\|T^{-1}(\hat{Z}_i'\Delta\hat{e}_i-Z_i'\Delta e_i)\|\le\{\sum_{j=1}^{p}[\frac{1}{T}\sum_{t=p+1}^{T}(\Delta\hat{e}_{it-j}\Delta\hat{e}_{it}-\Delta e_{it-j}\Delta e_{it})]^2\}^{1/2}\le\sqrt{p}O_p(C_{NT}^{-1})$. *Q.E.D.*

G. *Preliminaries for Theorem 3*

Introduce $\tilde{e}_{it}=e_{it}-e_{i1}-(e_{iT}-e_{i1})(t-1)/(T-1)$; then by (17)

$$\hat{e}_{it}=\tilde{e}_{it}-\lambda_i'H^{-1}\sum_{s=2}^{t}v_s-d_i\sum_{s=2}^{t}\hat{f}_s, \tag{G.1}$$

$$\Delta\hat{e}_{it}=\Delta\tilde{e}_{it}-\lambda_i'H^{-1}v_t-d_i'\hat{f}_t. \tag{G.2}$$

LEMMA G.1: *For* $\rho_i=1$ *or* $|\rho_i|<1$:
(i) $(1/\sqrt{T})\hat{e}_{it}=(1/\sqrt{T})\tilde{e}_{it}+O_p(C_{NT}^{-1})$, *uniformly in* $t\in[1,T]$;
(ii) $(1/T^2)\sum_{t=2}^{T}\hat{e}_{it}^2=(1/T^2)\sum_{t=2}^{T}\tilde{e}_{it}^2+O_p(C_{NT}^{-1})$;
(iii) $(1/T)\sum_{t=2}^{T}(\Delta\hat{e}_{it})^2=(1/T)\sum_{t=2}^{T}(\Delta\tilde{e}_{it})^2+O_p(C_{NT}^{-1})=(1/T)\sum_{t=2}^{T}(\Delta e_{it})^2+O_p(C_{NT}^{-1})$;
(iv) $(1/T)\sum_{t=2}^{T}\hat{e}_{it-1}\Delta\hat{e}_{it}=(1/T)\sum_{t=2}^{T}\tilde{e}_{it-1}\Delta\tilde{e}_{it}+O_p(C_{NT}^{-1})$.

PROOF: The proof is similar to that of Lemma B.1. Consider (i). This follows from (G.1), $T^{-1/2}\|\sum_{s=2}^{t}v_s\|=O_p(C_{NT}^{-1})$, $\|d_i\|=O_p(1/\min[N,\sqrt{T}])=O_p(C_{NT}^{-1})$ and $T^{-1/2}\|\sum_{s=2}^{t}\hat{f}_s\|=O_p(1)$. Result (ii) is an immediate consequence of (i).

The first equality of (iii) follows from (G.2), $\sum_{t=2}^{T}\|v_t\|^2/T=O_p(C_{NT}^{-2})$, $\sum_{t=2}^{T}\|\hat{f}_t\|^2/T=O_p(1)$, and $\|d_i\|^2=O_p(C_{NT}^{-2})$. To prove the second equality, note that $\Delta\tilde{e}_{it}=\Delta e_{it}-\overline{\Delta e_i}$, but $\overline{\Delta e_i}=(e_{iT}-e_{i1})/(T-1)=O_p(T^{-1/2})$, implying the second equality of (iii).

For (iv), consider the two identities:

$$\frac{1}{T}\sum_{t=2}^{T}\hat{e}_{it-1}\Delta\hat{e}_{it}=\frac{1}{2T}(\hat{e}_{iT})^2-\frac{1}{2T}\hat{e}_{i1}^2-\frac{1}{2T}\sum_{t=2}^{T}(\Delta\hat{e}_{it})^2, \tag{G.3}$$

$$\frac{1}{T}\sum_{t=2}^{T}\tilde{e}_{it-1}\Delta\tilde{e}_{it}=\frac{1}{2T}(\tilde{e}_{iT})^2-\frac{1}{2T}\tilde{e}_{i1}^2-\frac{1}{2T}\sum_{t=2}^{T}(\Delta\tilde{e}_{it})^2. \tag{G.4}$$

Applying (i) with $t=T$ and $t=1$, respectively, we get $T^{-1}[(\hat{e}_{iT})^2-(\tilde{e}_{iT})^2]=O_p(C_{NT}^{-2})$ and $T^{-1}[(\hat{e}_{i1})^2-(\tilde{e}_{i1})^2]=O_p(C_{NT}^{-2})$. These results together with (iii) imply (iv). *Q.E.D.*

Next consider the properties of estimated common factors $\hat{F}_t$. From $\hat{f}_t=Hf_t+v_t$,

$$\hat{F}_t=\sum_{s=2}^{t}\hat{f}_s=H\sum_{s=2}^{t}f_s+\sum_{s=2}^{t}v_s=H\sum_{s=2}^{t}(\Delta F_s-\overline{\Delta F})+\sum_{s=2}^{t}v_s \tag{G.5}$$

$$= H\left[F_t - F_1 - \frac{F_T - F_1}{T-1}(t-1)\right] + V_t,$$

where $V_t = \sum_{s=2}^{t} v_s$. For a sequence y_t, let y_t^τ denote the residual from regressing $\{y_t\}$ on $[1, t]$ $(t = 2, \dots, T)$. That is, y_t^τ is the demeaned and detrended series. Then from (G.6),

(G.6) $\qquad \hat{F}_t^\tau = HF_t^\tau + V_t^\tau$

(clearly, demeaning and detrending will remove $F_1 + (F_T - F_1)(t-1)/(T-1)$).

LEMMA G.2: $\max_{2\le t\le T}(1/\sqrt{T})\|V_t^\tau\| = O_p(C_{NT}^{-1})$.

PROOF: This simply follows from $T^{-1/2}\|V_t\| = T^{-1/2}\|\sum_{s=2}^{t} v_s\| = O_p(C_{NT}^{-1})$ uniformly in t. To see this, let $V = (V_2, V_3, \dots, V_T)'$, $V^\tau = (V_2^\tau, V_3^\tau, \dots, V_T^\tau)'$, and let $Z = (Z_2, Z_3, \dots, Z_T)'$, where $Z_t = (1, t)$. Then $V^\tau = M_z V$, where $M_z = I - Z(Z'Z)^{-1}Z'$. Then, it is easy to show that V_t^τ can be written as

$$V_t^\tau = V_t + a_T \frac{1}{T}\sum_{j=2}^{T} V_j + b_T(t/T)\frac{1}{T^2}\sum_{j=2}^{T} jV_j,$$

where a_T and b_T are bounded numbers. It follows that

$$\frac{1}{\sqrt{T}}V_t^\tau = \frac{1}{\sqrt{T}}V_t + a_T \frac{1}{T^{3/2}}\sum_{j=2}^{T} V_j + b_T(t/T)\frac{1}{T^{5/2}}\sum_{j=2}^{T} jV_j.$$

However, $\|T^{-3/2}\sum_{j=2}^{T} V_j\| \le \max_t T^{-1/2}\|V_t\|$. Similarly, $T^{-5/2}\|\sum_{j=2}^{T} jV_j\| \le C\max_t T^{-1/2}\|V_t\|$. *Q.E.D.*

LEMMA G.3: *As* $N, T \to \infty$:
(i) $(1/\sqrt{T})\hat{F}_t^\tau = H(1/\sqrt{T})F_t^\tau + O_p(C_{NT}^{-1})$, *uniformly in* $t \in [1, T]$;
(ii) $T^{-2}\sum_{t=2}^{T}\hat{F}_{t-1}^\tau\hat{F}_{t-1}^{\tau\prime} = HT^{-2}\sum_{t=2}^{T} F_{t-1}^\tau F_{t-1}^{\tau\prime}H' + O_p(C_{NT}^{-1})$;
(iii) $(1/T)\sum_{t=2}^{T}\Delta\hat{F}_t^\tau\Delta\hat{F}_t^{\tau\prime} = H(1/T)\sum_{t=2}^{T}\Delta F_t\Delta F_t'H' + O_p(C_{NT}^{-1})$;
(iv) $(1/T)\sum_{t=2}^{T}(\hat{F}_{t-1}^\tau\Delta\hat{F}_t' + \Delta\hat{F}_t\hat{F}_{t-1}^{\tau\prime}) = (1/T)H\sum_{t=2}^{T}(F_{t-1}^\tau\Delta F_t' + \Delta F_t F_{t-1}^{\tau\prime})H' + O_p(C_{NT}^{-1})$.

PROOF: (i) follows from (G.6) and Lemma G.2. (ii) is an immediate consequence of (i). For (iii) write $\hat{F}_t^\tau = \hat{F}_t - \hat{a} - \hat{b}t$ for some $\hat{a}$ and $\hat{b}$. This is possible because $\hat{F}_t^\tau$ is the projection residual of $\hat{F}_t$. Thus, $\Delta\hat{F}_t^\tau = \Delta\hat{F}_t - \hat{b}$. Note that the slope coefficient satisfies $\|b\| = O_p(T^{-1/2})$ because $T^{-1/2}\hat{F}_t = O_p(1)$. Furthermore, $\Delta\hat{F}_t = H(\Delta F_t - \overline{\Delta F}) + v_t = H\Delta F_t + O_p(T^{-1/2}) + v_t$ because $\overline{\Delta F} = O_p(T^{-1/2})$ by assumption that ΔF_t has zero mean. Thus $\Delta\hat{F}_t^\tau = H\Delta F_t + O_p(T^{-1/2}) + v_t$, from which (iii) follows easily. Next, consider (iv). First note that $\Delta\hat{F}_t$ can be replaced by $\Delta\hat{F}_t^\tau$. This is because $\Delta\hat{F}_t = \Delta\hat{F}_t^\tau + \hat{b}$ and $\sum_{t=2}^{T}\hat{F}_{t-1}^\tau = 0$ (normal equation). Then we have the identity

(G.7) $$\frac{1}{T}\sum_{t=2}^{T}(\hat{F}_{t-1}^\tau\Delta\hat{F}_t^{\tau\prime} + \Delta\hat{F}_t^\tau\hat{F}_{t-1}^{\tau\prime}) = \frac{\hat{F}_T^\tau\hat{F}_T^{\tau\prime}}{T} - \frac{\hat{F}_1^\tau\hat{F}_1^{\tau\prime}}{T} - \frac{1}{T}\sum_{t=2}^{T}\Delta\hat{F}_t^\tau\Delta\hat{F}_t^{\tau\prime}$$

and

$$H\frac{1}{T}\left[\sum_{t=2}^{T}\Delta F_t^\tau F_{t-1}^{\tau\prime} + F_{t-1}^\tau\Delta F_t^{\tau\prime}\right]H'$$

$$= H\frac{F_T^\tau F_T^{\tau\prime}}{T}H' - H\frac{F_1^\tau F_1^{\tau\prime}}{T}H' - H\frac{1}{T}\sum_{t=2}^{T}\Delta F_t^\tau\Delta F_t^{\tau\prime}H'.$$

From part (i) with $t = T$, $\hat{F}_T^\tau \hat{F}_T^{\tau\prime}/T - H(F_T^\tau F_T^{\tau\prime}/T)H' = O_p(C_{NT}^{-1})$, and with $t = 1$, $\hat{F}_1^\tau \hat{F}_1^{\tau\prime}/T - H(F_1^\tau F_1^{\tau\prime}/T)H' = O_p(C_{NT}^{-1})$. From $\Delta F^\tau = \Delta F_t + O_p(T^{-1/2})$, $T^{-1}\sum_{t=2}^T \Delta F_t^\tau \Delta F_t^{\tau\prime} - T^{-1}\sum_{t=2}^T \Delta F_t \times \Delta F_t' = O_p(T^{-1/2})$. This, together with part (iii), yields

$$\frac{1}{T}\sum_{t=2}^{T}\Delta\hat{F}_t^\tau \Delta\hat{F}_t^{\tau\prime} - H\frac{1}{T}\sum_{t=2}^{T}\Delta F_t^\tau \Delta F_t^{\tau\prime} H' = O_p(C_{NT}^{-1}).$$

Combining these results leads to (iv). *Q.E.D.*

Before proving the theorem for serially correlated disturbances, we first prove the theorem for i.i.d. disturbances, which provides a substantial insight with a very simple proof.

The DF statistic is

$$DF_{\hat{e}}^\tau(i) = \frac{T^{-1}\sum_{t=2}^T \hat{e}_{it-1}\Delta\hat{e}_{it}}{(\hat{\sigma}_{\epsilon i}^2 T^{-2}\sum_{t=2}^T \hat{e}_{it-1}^2)^{1/2}},$$

where $\hat{\sigma}_{\epsilon i}^2 = T^{-1}\sum_{t=2}^T(\Delta\hat{e}_{it} - \hat{a}_i\hat{e}_{it-1})^2$, which converges to $\sigma_{\epsilon i}^2$.

PROPOSITION 3: *Suppose the assumptions of Theorem* 3 *hold. If* $D_i(L) = 1$, *i.e.*, $e_{it} = \rho_i e_{it-1} + \epsilon_{it}$ *with* ϵ_{it} *being i.i.d.* $(0, \sigma_{\epsilon i}^2)$, *then under* $\rho_i = 1$, *as* $N, T \to \infty$,

$$DF_{\hat{e}}^\tau(i) \Rightarrow -\frac{1}{2}\left(\int_0^1 V_{\epsilon i}(r)\right)^{-1/2}.$$

Proposition 3 is implied by the lemma below.

LEMMA G.4: *Under the assumptions of Proposition* 3 *with* $\rho_i = 1$, *for* $t = [\mathrm{Tr}]$, *as* $N, T \to \infty$:

(i) $(1/\sqrt{T})\hat{e}_{it} \Rightarrow [W_{\epsilon i}(r) - rW_{\epsilon i}(1)]\sigma_{\epsilon i} \equiv V_{\epsilon i}(r)\sigma_{\epsilon i}$;

(ii) $T^{-1}\sum_{t=2}^T \hat{e}_{it-1}^2 \Rightarrow \sigma_{\epsilon i}^2 \int_0^1 V_{\epsilon i}(r)^2\,dr$;

(iii) $(1/T)\sum_{t=2}^T(\Delta\hat{e}_{it})^2 \xrightarrow{p} \sigma_{\epsilon i}^2$;

(iv) $(1/T)\sum_{t=2}^T \hat{e}_{it-1}\Delta\hat{e}_{it} \Rightarrow -\sigma_{\epsilon i}^2/2$.

PROOF: (i): By Lemma G.1(i), it suffices to show $T^{-1/2}\tilde{e}_{it}$ has the said limit. But $T^{-1/2}\tilde{e}_{it} = T^{-1/2}e_{it} - T^{-1/2}e_{iT}(t-1)/(T-1) - T^{-1/2}e_{i1}$. By the invariance principle, $T^{-1/2}e_{it} \Rightarrow W_{\epsilon i}(r)\sigma_{\epsilon i}$, and $T^{-1/2}e_{iT}(t-1)/(T-1) \Rightarrow rW_{\epsilon i}(1)\sigma_{\epsilon i}$. Furthermore, $T^{-1/2}e_{i1} \to 0$, proving (i). Result (ii) follows from (i) and the continuous mapping theorem. Under $\rho_i = 1$, $\Delta e_{it} = \epsilon_{it}$, and so $T^{-1}\sum_{t=2}^T \Delta e_{it}^2 \xrightarrow{p} \sigma_{\epsilon i}^2$. This implies (iii) in view of Lemma G.1(iii). Finally, (iv) follows from (G.3), part (i) and (iii). To see this, by (i), $T^{-1}\hat{e}_{iT}^2 \Rightarrow \sigma_{\epsilon i}^2 V_{\epsilon i}(1)^2 = 0$. Furthermore, $T^{-1}\hat{e}_{i1}^2 = 0$. Thus the right-hand side of (G.3) converges to $-\sigma_{\epsilon i}^2/2$, proving (iv). *Q.E.D.*

Likewise, when F_t is I(1) and is driven by i.i.d. errors, the test is very simple. The DF statistic for the series $\hat{F}_t$ with demeaning and detrending is numerically equal to (see, e.g., Hayashi (2000, p. 608))

$$DF_{\hat{F}}^\tau = \frac{T^{-1}\sum_{t=2}^T \hat{F}_{t-1}^\tau \Delta\hat{F}_t}{(\hat{\sigma}_u^2 T^{-2}\sum_{t=2}^T(\hat{F}_{t-1}^\tau)^2)^{1/2}}, \tag{G.8}$$

where $\hat{\sigma}_u^2 = \sum_{t=2}^T(\Delta\hat{F}_t - \hat{a} - \hat{b}t - \hat{c}\hat{F}_{t-1})^2/(T-2)$ with $(\hat{a}, \hat{b}, \hat{c})$ being the OLS estimate when regressing $\Delta\hat{F}_t$ on $[1, t, \hat{F}_{t-1}]$. It is easy to show that $\hat{\sigma}_u^2 \xrightarrow{p} H^2\sigma_u^2$.

PROPOSITION 4: *Suppose the assumptions of Theorem* 3 *hold. If* $C(L) = 1$, *i.e.*, $F_t = F_{t-1} + u_t$ *with* u_t *being i.i.d.* $(0, \sigma_u^2)$, *then*

$$DF_{\hat{F}}^{\tau} \Rightarrow \frac{\int_0^1 W^{\tau}(s)\, dW(s)}{(\int_0^1 W^{\tau}(s)^2\, ds)^{1/2}}.$$

This proposition is implied by the following lemma.

LEMMA G.5: *Under the assumptions of Proposition* 4, *as* $N, T \to \infty$:
(i) $(1/\sqrt{T})\hat{F}_t^{\tau} \Rightarrow H\sigma_u W^{\tau}(r)$;
(ii) $T^{-2}\sum_{t=2}^{T}\hat{F}_{t-1}^{\tau}\hat{F}_{t-1}^{\tau\prime} \Rightarrow H^2\sigma_u^2\int_0^1 W_u^{\tau}(r)^2\, dr$;
(iii) $(1/T)\sum_{t=2}^{T}\hat{F}_{t-1}^{\tau}\Delta\hat{F}_t \Rightarrow H^2\sigma_u^2\int_0^1 W_u^{\tau}(r)\, dW_u(r)$.

PROOF: The results follow from Lemma G.3 and the corresponding weak convergence for the series F_t^{τ}. For example, $T^{-1/2}F_t^{\tau} \Rightarrow \sigma_u W^{\tau}(r)$. Result (ii) follows from (i) and the continuous mapping theorem. Result (iii) follows from Lemma G.3(iv) and $T^{-1}\sum_{t=2}^{T}F_{t-1}^{\tau}\Delta F_t \Rightarrow \sigma_u^2\int_0^1 W_u^{\tau}(r)\, dW_u(r)$. *Q.E.D.*

H. *Testing* $\hat{e}_{it}$ *Using the ADF, Linear Trend Case*

Recall

$$\text{(H.1)} \qquad \hat{e}_{it} = \tilde{e}_{it} - \lambda_i' H^{-1}\sum_{s=2}^{t} v_s - d_i\sum_{s=2}^{t}\hat{f}_s,$$

$$\text{(H.2)} \qquad \Delta\hat{e}_{it} = \Delta\tilde{e}_{it} - \lambda_i' H^{-1} v_t - d_i'\hat{f}_t,$$

where $\tilde{e}_{it} = e_{it} - e_{i1} - (e_{iT} - e_{i1})(t-1)/(T-1)$, and $\Delta\tilde{e}_{it} = \Delta e_{it} - \overline{\Delta e_i}$. Also note $\tilde{e}_{it} = \sum_{s=2}^{t}\Delta\tilde{e}_{it}$. The proof consists of two steps. The first is to show the ADF test based on $\tilde{e}_{it}$ has the desired limiting distribution, i.e., $ADF_{\tilde{e}}(i) \Rightarrow (-1/2)(\int_0^1 V_{i,\epsilon}(r)^2\, dr)^{-1}$. The second is to show the ADF test based on $\hat{e}_{it}$ is asymptotically the same as that based on $\tilde{e}_{it}$, i.e., $ADF_{\hat{e}}(i) - ADF_{\tilde{e}}(i) = o_p(1)$, as $N, T \to \infty$.

The ADF test in this linear trend case has the same distribution as the test considered in Schmidt and Lee (1991), a modified version of the LM test for the presence of a unit root around a linear trend developed in Schmidt and Phillips (1992). Note that even when e_{it} is observable, the first step has not been explicitly stated in the literature. Schmidt and Phillips (1992) considered nonparametric correction of serial correlation, but not via ADF regression.

LEMMA H.1: *Assume* $\Delta e_{it} = D_i(L)\epsilon_{it}$ *satisfies Assumption* C(i). *Consider the regression* $\Delta\tilde{e}_{it} = \delta_{i0}\tilde{e}_{i,t-1} + \sum_{j=1}^{p}\delta_{ij}\Delta\tilde{e}_{i,t-j} + error$. *Let* $ADF_{\tilde{e}}(i)$ *be the t statistic for testing* $\delta_{i0} = 0$. *If* p *is chosen such that* $p \to \infty$ *with* $p^3/T \to 0$, *then* $ADF_{\tilde{e}}(i) \Rightarrow (-1/2)(\int_0^1 V_{i,\epsilon}(r)^2\, dr)^{-1}$.

The proof uses a similar argument to that used in Said and Dickey (1984) and Berk (1974). Since the argument is tedious, the detail is omitted. Instead, we outline the proof for fixed p, which is drastically simpler. For this, we assume $\Delta e_{it} = D_i(L)\epsilon_{it}$ has a finite AR(p) representation:

$$\Delta e_{it} = \sum_{j=1}^{p}\delta_{ij}\Delta e_{i,t-j} + \epsilon_{it},$$

where ϵ_{it} are i.i.d. $(0, \sigma^2_{\epsilon,i})$. It can be shown that, when replacing Δe_{it} by $\Delta\tilde{e}_{it}$,

$$\Delta\tilde{e}_{it} = \sum_{j=1}^{p} \delta_{ij}\Delta\tilde{e}_{i,t-j} + \epsilon_{it} - \bar{\epsilon}_i + O_p(T^{-1}), \tag{H.3}$$

where $\bar{\epsilon}_i = \sum_{t=2}^{T} \epsilon_{it}/(T-1)$. Rewrite the regression equation stated in the lemma in matrix form:

$$\Delta\tilde{e}_i = \delta_{i0}\tilde{e}_{i_{-1}} + Z\phi + \text{error},$$

where Z is a matrix consisting of lags in $\Delta\tilde{e}_{i,t}$ and $\phi = (\delta_{i1}, \ldots, \delta_{ip})'$. The least squares estimator of δ_{i0} satisfies

$$T\hat{\delta}_{i0} = \frac{T^{-1}\tilde{e}'_{i_{-1}} M_Z \Delta\tilde{e}_i}{T^{-2}\tilde{e}'_{i_{-1}} M_Z \tilde{e}_{i_{-1}}} = \frac{T^{-1}\tilde{e}'_{i_{-1}}(\epsilon_i - \bar{\epsilon}_i) + o_p(1)}{T^{-2}\tilde{e}'_{i_{-1}} M_Z \tilde{e}_{i_{-1}}} = \frac{T^{-1}\tilde{e}'_{i_{-1}}(\epsilon_i - \bar{\epsilon}_i) + o_p(1)}{T^{-2}\tilde{e}'_{i_{-1}}\tilde{e}_{i_{-1}}}.$$

The second equality follows from $\Delta\tilde{e}_i = Z\phi + \epsilon_i - \bar{\epsilon}_i + O_p(T^{-1})$ by (H.3) and $M_Z Z = 0$, and the last equality follows from $T^{-2}\tilde{e}'_{i_{-1}} M_Z \tilde{e}_{i_{-1}} = T^{-2}\tilde{e}'_{i_{-1}}\tilde{e}_{i_{-1}} + o_p(1)$.

We will show $T\hat{\delta}_{i0} \Rightarrow -[2D_i(1)\int_0^1 V_i(r)^2\,dr]^{-1}$. Now $T^{-2}\tilde{e}'_{i_{-1}}\tilde{e}_{i_{-1}} = T^{-2}\sum_{t=p}^{T}\tilde{e}_{it}^2 \Rightarrow D_i(1)^2\sigma^2_{\epsilon i} \times \int_0^1 V_i(r)^2\,dr$ simply follows from the relationship between $\tilde{e}_{it}$ and e_{it} and the weak convergence of $T^{-1/2}e_{it}$ to $D_i(1)\sigma_{\epsilon i}W_i(r)$ for $t = [Tr]$. But the limit of the numerator requires an extra argument. By the Beveridge–Nelson decomposition, $\Delta e_{it} = D_i(1)\epsilon_{it} + \eta_{it-1} - \eta_{it}$ where $\eta_{it} = \sum_{j=0}^{\infty}(\sum_{k=j+1}^{\infty} d_{ik})\epsilon_{it-j}$. This leads to

$$\Delta\tilde{e}_{it} = \Delta e_{it} - \overline{\Delta e_i} = D_i(1)(\epsilon_{it} - \bar{\epsilon}_i) + \eta_{it-1} - \eta_{it} + T^{-1}(\eta_{iT} - \eta_{i1}).$$

Cumulating $\Delta\tilde{e}_{it}$ gives

$$\tilde{e}_{it} = \sum_{s=2}^{t}\Delta\tilde{e}_{it} = D_i(1)\sum_{s=2}^{t}(\epsilon_{is} - \bar{\epsilon}_i) + \eta_{i1} - \eta_{it} + T^{-1}(\eta_{i1} - \eta_{iT})(t-1).$$

Thus

$$\begin{aligned} T^{-1}\tilde{e}'_{i_{-1}}(\epsilon_i - \bar{\epsilon}_i) &= T^{-1}\sum_{t=p}^{T}\tilde{e}_{it-1}(\epsilon_{it} - \bar{\epsilon}_i) \\ &= D_i(1)\frac{1}{T}\sum_{t=p}^{T}\sum_{s=2}^{t-1}(\epsilon_{is} - \bar{\epsilon}_i)(\epsilon_{it} - \bar{\epsilon}_i) + \frac{1}{T}\sum_{t=p}^{T}(\eta_{i1} - \eta_{it-1})(\epsilon_{it} - \bar{\epsilon}_i) \\ &\quad + (\eta_{i1} - \eta_{iT})\frac{1}{T^2}\sum_{t=p}^{T}(t-2)(\epsilon_{it} - \bar{\epsilon}_i). \end{aligned}$$

The last two terms are each $o_p(1)$ and the first term converges to $-1/2D_i(1)\sigma^2_{\epsilon i}$, which can be proved as in Lemma G.4(iv) since the term can be rewritten as $D_i(1)T^{-1}\sum_{t=p}^{T} g_{t-1}\Delta g_t$, with $g_t = \sum_{s=2}^{t}(\epsilon_{is} - \bar{\epsilon}_i)$. The limit of $T\hat{\delta}_{i0}$ is thus obtained, which depends on nuisance parameter $D_i(1)$. But the t-statistic eliminates $D_i(1)$, as is well known in the standard ADF test.

Given Lemma H.1 and Lemma G.1, the proof that $ADF_{\hat{e}}(i) - ADF_{\tilde{e}}(i) = o_p(1)$ is almost identical to the proof in Theorem 2. Thus the detail is omitted. For insight, see the proof of Proposition 3 when the disturbances are i.i.d.

The proof of part 2 and part 3 for serially correlated disturbances is also omitted as the proof is almost the same as in Theorem 2, given Lemma G.2 and Lemma G.3. Also see Remark 1 in Appendix D. For insight, see the proof of Proposition 4 when the disturbances are i.i.d.

PROOF OF THEOREM 4: In this theorem, e_{it} are assumed to be cross-sectionally independent. Thus, the test statistics $ADF_e^c(i)$ $(i = 1, 2, \dots, N)$ based on the true series e_{it} are independent over i. Theorem 1 shows that $ADF_{\hat{e}}^c(i) - ADF_e^c(i) = o_p(1)$, that is, $ADF_{\hat{e}}^c(i)$ not only has the same asymptotic distribution as $ADF_e^c(i)$, but they are asymptotically equivalent. This implies asymptotic independence of $ADF_{\hat{e}}^c(i)$. The same is true for the linear trend model $ADF_{\hat{e}}^\tau(i)$ over i. The analysis of the pooled test then proceeds following the arguments of Choi (2001). *Q.E.D.*

REFERENCES

AZNAR, A., AND M. SALVADOR (2002): "Selecting the Rank of the Cointegration Space and the Form of the Intercept Using an Information Criteria," *Econometric Theory*, 18, 926–947.

BAI, J. (2003): "Inference on Factor Models of Large Dimensions," *Econometrica*, 71, 135–172.

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

BALTAGI, B., AND C. KAO (2001): "Nonstationary Panels, Cointegration in Panels, and Dynamic Panels," Center for Policy Research, Syracuse University.

BANERJEE, A., M. MARCELLINO, AND C. OSBAT (2001): "Testing for PPP: Should We Use Panel Methods?" Mimeo, European University Institute.

BERK, K. N. (1974): "Consistent Autoregressive Spectral Estimates," *The Annals of Statistics*, 2, 489–502.

CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): "Arbitrage, Factor Structure and Mean–Variance Analysis in Large Asset Markets," *Econometrica*, 51, 1305–1324.

CHANG, Y. (2002): "Nonlinear IV Unit Root Tests in Panels with Cross-Section Dependency," *Journal of Econometrics*, 110, 261–292.

CHANG, Y., AND W. SONG (2002): "Panel Unit Root Tests in the Presence of Cross-Section Heterogeneity," Unpublished manuscript, Rice University.

CHAO, J., AND P. PHILLIPS (1999): "Model Selection in Partially Nonstationary Vector Autoregressive Processes with Reduced Rank Structure," *Journal of Econometrics*, 91, 227–272.

CHOI, I. (2001): "Unit Root Tests for Panel Data," *Journal of International Money and Finance*, 20, 249–272.

——— (2002): "Combination Unit Root Tests for Cross-Sectionally Correlated Panels," Unpublished manuscript.

CONNOR, G., AND R. KORAJZCYK (1986): "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15, 373–394.

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

ELLIOTT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.

ENGLE, R. F., AND S. KOZICKI (1993): "Testing for Common Features," *Journal of Business and Economic Statistics*, 11, 369–379.

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.

GONZALO, J., AND C. GRANGER (1995): "Estimation of Common Long-Memory Components in Cointegrated Systems," *Journal of Business and Economic Statistics*, 13, 27–35.

HARRIS, D. (1997): "Principal Components Analysis of Cointegrated Time Series," *Econometric Theory*, 13, 529–557.

HAYASHI, F. (2000): *Econometrics*. Princeton, NJ: Princeton University Press.

IM, K., M. PESARAN, AND Y. SHIN (2003): "Testing for Unit Roots in Heterogeneous Panels," *Journal of Econometrics*, 115, 53–74.

JOHANSEN, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

——— (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

KING, R. G., C. PLOSSER, J. STOCK, AND M. WATSON (1991): "Stochastic Trends and Economic Fluctuations," *American Economic Review*, 81, 819–840.

LEVIN, A., C. F. LIN, AND J. CHU (2002): "Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties," *Journal of Econometrics*, 98, 1–24.

MACKINNON, J. (1994): "Approximate Asymptotic Distribution Functions for Unit-Root and Cointegration Tests," *Journal of Business and Economic Statistics*, 12, 167–177.

MADDALA, G. S., AND S. WU (1999): "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test," *Oxford Bulletin of Economics and Statistics*, Special Issue, 631–652.

MOON, R., AND B. PERRON (2003): "Testing for a Unit Root in Panels with Dynamic Factors," forthcoming in the *Journal of Econometrics*.

NEWEY, W., AND K. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

NG, S., AND P. PERRON (2001): "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," *Econometrica*, 69, 1519–1554.

O'CONNELL, P. (1998): "The Overvaluation of Purchasing Power Parity," *Journal of International Economics*, 44, 1–19.

PANTULA, S. G. (1991): "Asymptotic Distributions of Unit Root Tests when the Process is Nearly Stationary," *Journal of Business and Economic Statistics*, 9, 63–71.

PEDRONI, P. (1995): "Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis: New Results," Indiana University Working Paper 95-013.

PESARAN, H., AND R. SMITH (1995): "Estimating Long Run Relationships from Dynamic Heterogeneous Panels," *Journal of Econometrics*, 68, 79–113.

PHILLIPS, P. C. B., AND S. N. DURLAUF (1986): "Multiple Regression with Integrated Processes," *Review of Economic Studies*, 53, 473–496.

PHILLIPS, P. C. B., AND R. MOON (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057–1111.

PHILLIPS, P. C. B., AND D. SUL (2003): "Dynamic Panel Estimation and Homogeneity Testing Under Cross-Section Dependence," *Econometrics Journal*, 6, 217–259.

QUAH, D. (1994): "Exploiting Cross-Section Variations for Unit Root Inference in Dynamic Panels," *Economics Letters*, 44, 1–9.

SAID, S. E., AND D. A. DICKEY (1984): "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71, 599–607.

SCHMIDT, P., AND J. LEE (1991): "A Modification of Schmidt-Phillips Unit Root Test," *Economics Letters*, 36, 285–289.

SCHMIDT, P., AND P. PHILLIPS (1992): "LM Tests for a Unit Root in the Presence of Deterministic Trends," *Oxford Bulletin of Economics and Statistics*, 54, 257–287.

SCHWERT, G. W. (1989): "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 7, 147–160.

STOCK, J. H., AND M. W. WATSON (1988): "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097–1107.

——— (2002): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.