AN IV TEST FOR A UNIT ROOT IN GENERALLY TRENDING AND CORRELATED PANELS *

Joakim Westerlund†
Deakin University
Australia

January 15, 2014

Abstract

The asymptotic distribution of all unit root test statistics depend on the deterministic specification of the fitted test regression, which need not be equal to the true one. In time series, this implies that different deterministic specifications have their own critical values, whereas in panels, it implies that different specifications have their own mean and variance correction factors. This paper proposes an IV-based panel unit root test that is general enough to accommodate general error serial and cross-section dependence, and a potentially non-linear deterministic trend function. These allowances make the new test one of the most general around. It is also very simple to implement. Indeed, IV statistic is asymptotically invariant to not only the true trend function, but also the deterministic specification of the test regression. This means that, unlike most existing tests, with this test there is no need for any deterministic-specific mean and variance correction factors.

JEL Classification: C12; C13; C33.

Keywords: Unit root test; Panel data; Deterministic trend function; Asymptotic invariance; Recursive detrending; Common factor models.

*Previous versions of the paper were presented at seminars at University of Maastricht, Deakin University, Queensland University of Technology, and Macquarie University. The author would like to thank seminar participants and in particular Joerg Breitung, Jean-Pierre Urbain, Stan Hurn, Mehdi Hosseinkouchack, Adam Clements, Chris Doucouliagos and Prasad Bhattacharya for many constructive comments. Thank you also to the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number P2009-0189:1.

†Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance, Melbourne Burwood Campus, 221 Burwood Highway, VIC 3125, Australia. Telephone: +61 3 924 46973. Fax: +61 3 924 46283. E-mail address: j.westerlund@deakin.edu.au.
1 Introduction

This paper proposes a test statistic of the null hypothesis of a unit root in panel data where the number of time periods, \( T \), and cross-sectional units, \( N \), are large. The framework is general enough to include most deterministic trend functions that are linear in parameters, including polynomial trend functions, trigonometric functions, and models of discrete and smooth transition shifts. The innovations may be both serially and cross-sectionally correlated in a very unrestricted fashion through a dynamic common factor model. A priori knowledge as to the extent of trending, and innovation serial and cross-sectional dependence is not required, provided that the chosen model specification is general enough to encompass the true one. Under this assumption the proposed test statistic, which is based on the instrumental variables (IV) principle, has the practically very useful property that it is asymptotically invariant to not only all nuisance parameters characterizing the dependence of the innovations and the true trend function, but also the deterministic specification of the test regression. The standard requirement of (at most) a linear trend is therefore not needed, and the otherwise so common mean and variance correction factors reflecting the chosen deterministic specification can be completely avoided. In terms of model specification, this means that researchers can proceed just as in the classical regression context. Indeed, all one has to do is to augment the test regression with whatever deterministic specification felt appropriate. The only requirement is that the chosen specification is general enough. Similarly, provided that the augmentation to account for the dependence of the innovations is sufficient, estimates of nuisance parameters, either parametric or nonparametric, are not needed.

The remainder of this paper is organized as follows. Section 2 presents the model and the key idea. Section 3 lays out the assumptions, which are used in Section 4 to derive the asymptotic distribution of the IV statistic under the unit root null hypothesis. Section 5 reports the results of a small-scale Monte Carlo study. Section 6 concludes. Proofs are relegated to Appendix.
This paper considers the following data generating process (DGP):

\[ Y_{i,t} = \beta_i' D_t + U_{i,t}, \quad (1) \]

\[ U_{i,t} = \rho_i U_{i,t-1} + u_{i,t}, \quad (2) \]

\[ \phi_i(L) u_{i,t} = e_{i,t}, \quad (3) \]

\[ e_{i,t} = \lambda_i F_t + \epsilon_{i,t}, \quad (4) \]

where \( t = 1, \ldots, T, \ i = 1, \ldots, N \), \( D_t = (1, D_{2,t}, \ldots, D_{q,t})' \) is a \( q \)-vector (\( q \geq 1 \)) of trend functions that are assumed to satisfy certain smoothness and collinearity conditions, but are otherwise unrestricted, \( \beta_i \) is a conformable vector of trend coefficients, \( \phi_i(L) = 1 - \sum_{n=1}^{p} \phi_{i,n} L^n \) is a polynomial in the lag operator \( L \) that has all its roots outside the unit circle, \( F_t \) is an \( r \)-vector of common factors, \( \lambda_i \) is a vector of factor loadings, and \( e_{i,t} \) is an idiosyncratic error term. \(^1\)

\( F_t \) and \( e_{i,t} \) are mean zero, independent of each other, and independently and identically distributed (iid) across both \( i \) and \( t \). Hence, in the model consider here, the observed variable, \( Y_{i,t} \), can be decomposed into a deterministic component, \( \beta_i' D_t \), and a stochastic component, \( U_{i,t} \), where the innovation, \( u_{i,t} \), admits to a (strict) dynamic common factor structure, making it both serially and cross-sectionally correlated (see, for example, Pesaran, 2007; Phillips and Sul, 2003; Moon and Perron, 2004, for similar DGP specifications).

Our objective is to test whether \( U_{i,t} \) has a unit root, which in the above DGP is equivalent to testing

\[ H_0 : \rho_1 = \ldots = \rho_N = 1. \]

The main problem in practice when testing this hypothesis is that \( U_{i,t} \) is only observable up to an additive deterministic component, \( \beta_i' D_t \), whose form is unknown to the researcher. Valid inference therefore relies on the researcher being able to account for the confounding effects of this component. Of course, valid inference also requires accounting for the serial and cross-section correlation properties of \( u_{i,t} \); however, the first order of business is how to purge the deterministic component. One of the main departures of the current work from previous ones is the consideration of a relatively unrestricted specification of \( D_t \). In this section, however, in order to facilitate discussion, we follow the convention in the literature

\(^1\)The use of a common lag order, \( p \), is just for notational convenience and is not a restriction.
and assume that \( D_t \) contains (at most) a linear trend (see, for example, Moon and Perron, 2004, 2008; Moon et al., 2006, 2007; Westerlund and Larsson, 2012);

\[
\beta'_i D_t = \beta_{1,i} + \beta_{2,i} t.
\]

We further assume that \( \phi_i(L) = 1 \), such that \( u_{i,t} = e_{i,t} \) is serially uncorrelated (although still potentially cross-correlated).

The test statistic that we propose in based on a two-stage IV procedure, where the first and second stages involve transforming \( Y_{i,t} \) so as to eliminate the effects of the deterministic component and the dependence contained in \( u_{i,t} \), respectively. This procedure leads to an estimate (or proxy) of the idiosyncratic part of \( U_{i,t} \), which can then be subjected to a conventional \( t \)-test for a unit root. In what follows we provide the main idea behind the stage 1 and 2 transformations; the actual test statistic is introduced in Section 4.

2.1 IV stage 1

The main difficulty when trying to purge the deterministic component is the presence of bias, which greatly complicates the testing. In order to illustrate the issues involved, suppose first that \( \beta_{2,i} = 0 \), such that \( \beta'_i D_t = \beta_{1,i} \) can be estimated using \( Y_i = T^{-1} \sum_{s=1}^{T} Y_{i,s} \). The single most common approach to the testing of \( H_0 \) in this case is to regress \( (Y_{i,t} - \bar{Y}_i) \) onto \( (Y_{i,t-1} - \bar{Y}_i) \), and to conduct a pooled \( t \)-test for a unit slope. The problem is that \( E(Y_{i,s} u_{i,t}) \neq 0 \) for \( s \geq t \), which means that although \( E(Y_{i,t-1} u_{i,t}) = 0 \), we still have

\[
E[(Y_{i,t-1} - \bar{Y}_i) u_{i,t}] = -E(\bar{Y}_i u_{i,t}) \neq 0;
\]

that is, while \( Y_{i,t} \) is a martingale, \( (Y_{i,t} - \bar{Y}_i) \) is not. The regression error is therefore correlated the regressor, which in turn renders the ordinary least squares (OLS) estimator biased.

In the classical stationary micro panel setting with \( T \) fixed and \( N \to \infty \) the presence of correlation between error and regressor is rather devastating, as in this case OLS is even inconsistent (Nickell, 1981). The problem is made less severe by assuming that \( N, T \to \infty \); however, while consistent, the asymptotic distribution of the OLS estimator is still miscentered, which in turn calls for some kind of correction. When working with stationary data this is quite simple (see Hahn and Kuersteiner, 2002). Unfortunately, this is not the case when working under the unit root.\(^2\) There are (at least) two reasons for this. First, the appropriate

\(^2\)See Westerlund and Breitung (2013) for a detailed treatment of the effects of detrending on panel unit root tests.
correction factors depend on nuisance parameters reflecting the serial and cross-sectional correlation properties of $u_{i,t}$ (see, for example, Levin et al., 2002; Moon and Perron, 2004). Second, even if $u_{i,t}$ is iid, there is still the problem that under $H_0$ the demeaning regression of $Y_{i,t}$ onto a constant is spurious. This makes the asymptotic distribution of the OLS estimator of the regression of $(Y_{i,t} - \bar{Y}_i)$ onto $(Y_{i,t-1} - \bar{Y}_i)$ dependent on the chosen deterministic specification. The correction factors that apply in the case of a constant are therefore not the same as those that apply in the case of a constant and trend, for example. This is important because the complexity of the calculations needed to obtain the required correction factors increases very quickly with the number of trend terms, which is also one of the reasons for why the literature has not yet ventured outside the linear trend environment. What is more, the bias correction causes increased variance. Hence, bias correction alone is not enough; there is also a need for variance correction.

The IV approach considered here not only eliminates the bias, and hence also the above mentioned problems, but leads to a test statistic that is asymptotically invariant with respect to both the true and chosen deterministic specification. The intuition is simple. Consider the case when $D_i = 1$ and $\lambda_i = 0$, such that if we define $y_{i,t} = \Delta Y_{i,t}$, then, under $H_0$, $y_{i,t} = \Delta U_{i,t} = u_{i,t} = \epsilon_{i,t}$. Hence, rather than using OLS demeaning, we use first-differencing to eliminate the constant (see, for example, Anderson and Hsiao, 1981, for a similar approach in the case of stationary dynamic panel data). The proposed estimate of the idiosyncratic part of $U_{i,t}$ (under $H_0$) is therefore given by

$$R_{i,t} = \sum_{s=2}^{t} y_{i,s},$$

which is defined for $t = 2, ..., T$ (see Bai and Ng, 2004, 2010; Westerlund, 2014, for similar suggestions). The main point here is that, unlike $(Y_{i,t} - \bar{Y}_i)$, $R_{i,t}$ is a martingale, and therefore $E(R_{i,t-1}u_{i,t}) = 0$, suggesting that the OLS slope estimator in a regression of $R_{i,t}$ onto $R_{i,t-1}$ is unbiased. Thus, in terms of the usual IV terminology, we use $R_{i,t-1}$ to instrument for $(Y_{i,t-1} - \bar{Y}_i)$.

The above discussion supposes that $\beta_{2,i} = 0$; however, there is a simple trick that can be used to produce the same outcome also when $\beta_{2,i}$ is not restricted in this way. It starts by noting that when $\beta_{2,i} \neq 0$ the first-differenced model under $H_0$ is given by $y_{i,t} = \beta_{2,i} + \epsilon_{i,t}$. However, rather than using $\bar{Y}_i = T^{-1} \sum_{s=2}^{T} y_{i,s}$ to estimate $\beta_{2,i}$, which would make $\sum_{s=2}^{T} (y_{i,s} - \bar{Y}_i)$ a non-martingale (see Bai and Ng, 2004, for an approach based on $\bar{Y}_i$), we use $\bar{Y}_{i,t} = \bar{Y}_i$.\]
The proposed estimate of the idiosyncratic part of $U_{i,t}$ is given by

$$R_{i,t} = \sum_{s=2}^{t} (y_{i,s} - \bar{y}_{i,s}).$$

It should be mentioned that while the martingale property can in principle be retained by using recursive detrending of $Y_{i,t}$, there is a distinctive advantage to recursively demean (and cumulate) $y_{i,t}$; namely, that the detrending regression is not spurious. This means that the estimated trend coefficients converge to constants rather than to random variables, which in turn mitigates the dependence on the chosen deterministic specification.

### 2.2 IV stage 2

Stage 1 takes care of the bias; however, it does not remove the effect of the cross-section correlation in case $\lambda_i \neq 0$. To illustrate the idea behind the second stage, suppose again that $D_{i,t} = 1$, but now $\lambda_i$ is randomly distributed, such that under $H_0$, $y_{i,t} = u_{i,t} = \lambda'_i F_t + \epsilon_{i,t}$. Had $\lambda_i$ been known, an estimator $\hat{F}_t$ of $F_t$ could have been obtained as the OLS slope in a cross-section regression of $y_{i,t}$ onto $\lambda_i$, and the proposed estimate of the idiosyncratic part of $U_{i,t}$ would have been $\sum_{s=2}^{t} (y_{i,s} - \lambda'_i \hat{F}_s)$. However, $\lambda_i$ is not known and we therefore propose the use of a second instrument. This instrument, denoted $Z_i$, is chosen to be uncorrelated with $\epsilon_{i,t}$ but highly correlated with $\lambda_i$, suggesting that the cross-section OLS estimator of $Z_i y_{i,t}$ onto $Z_i \lambda'_i$ is consistent for $F_t$, or, equivalently, $\hat{F}_t = \frac{1}{N-1} \sum_{i=1}^{N} Z_i y_{i,t}$ is consistent for the space spanned by $F_t$. This is the IV estimator of (the space spanned by) $F_t$. The resulting estimate of the idiosyncratic part of $U_{i,t}$ is given by

$$R_{i,t} = \sum_{s=2}^{t} (y_{i,s} - Z'_i \hat{F}_s).$$

The range of permissable candidates for $Z_i$ is very broad. For example, suppose that $r = 1$ and $E(\lambda_i) = \lambda \neq 0$. Since in this case, $E(Z_i \lambda'_i) = E(\lambda'_i) = \lambda' \neq 0$ and $E(Z_i \epsilon_{i,t}) = E(\epsilon_{i,t}) = 0$, $Z_i = 1$ provides a valid instrument, suggesting that $Z'_i \hat{F}_1 = \bar{y}_t = \frac{1}{N-1} \sum_{i=1}^{N} y_{i,t}$. In other words, if $r = Z_i = 1$, then the IV estimator of the common component, $\lambda'_i F_t$, reduces to $\bar{y}_t$, which is similar to the approach of Pesaran (2007). The main difference is that, in analogy to the detrending, while in our case the estimation is done using a regression in (stationary) first-differenced data, Pesaran (2007) uses a spurious regression that involves $Y_{i,t}$ and $\bar{y}_t = \frac{1}{N-1} \sum_{i=1}^{N} Y_{i,t}$, leading to a nonstandard asymptotic distribution theory.

---

Note: The text contains a reference to Pesaran (2007) and the use of a similar approach in the context of a single time series. See, for example, Shin and So (2001, 2002) for similar proposals in the context of a single time series.
Finally, note that while there are other IV-based tests around (see, for example, Chang, 2002, 2012; Demetrescu and Hanck, 2012; Shin and Kang, 2004), the current approach is materially different. In particular, unlike in the present two-step procedure, these other approaches are based on instrumenting \( Y_{i,t-1} \) by a non-linear transformation of itself, which can be shown to remove the effects of cross-section dependence, but only in the case of weak cross-section dependence and at most a linear trend. Hence, not only are these tests based on a different instrumentation scheme, but they are also not as general as the tests considered in the present paper.

3 Assumptions

The DGP is given by (1)–(4). The conditions placed on this DGP are given in Assumptions 1–3, where \( \rightarrow_p \) and \( \lfloor x \rfloor \) signify convergence in probability and the integer part of \( x \), respectively, and \( \text{rk}(A) \), \( \text{tr}(A) \) and \( ||A|| = \sqrt{\text{tr}(A^\prime A)} \) denote the rank, trace and Frobenius (Euclidean) norm, respectively, of the matrix \( A \). Throughout, \( M < \infty \) denotes a generic positive constant.

Assumption 1 characterizes the permissible trend functions (see Rodrigues, 2012; Westerlund, 2014, for similar assumptions). The conditions are placed on \( d_t \), which is such that \( d_t = 0 \) if \( q = 1 \) and \( d_t = G \Delta D_t \) for \( q \geq 2 \), where \( G \) is a \((q - 1) \times q\) selection matrix of zeros and ones removing the first difference of the first element of \( D_t \), which is assumed to be 1.

Assumption 1. For \( q \geq 2 \), \( J_T \rightarrow_d (s) = 1 \) for \( s = 0.5 \), suggesting that the rank condition fails for \( v \in (0, 0.5) \), but not for \( v \in [0.5, 1] \). The trick here is to simply change the deterministic specification halfway through.
into the sample, using no deterministic component for the first \(0.5T\) – 1 observations and \(d_t\) for the rest. Virtually any model of discrete shift can be accommodated in this way.

**Assumption 2.**

(a) \(\epsilon_{i,t}\) is iid across \(i\) and \(t\) with \(E(\epsilon_{i,t}) = 0\), \(E(\epsilon_{i,t}^2) = \sigma_{\epsilon,i}^2 > 0\) and \(E(\epsilon_{i,t}^4) \leq M\);

(b) \(E(U_{i,0}) \leq M\) for all \(i\);

(c) \(F_t\) is iid across \(t\) with \(E(F_t) = 0\) and \(E(F_tF_t') = \Sigma_F > 0\) and \(E(||F_t||^4) \leq M\);

(d) \(\lambda_i\) is either deterministic such that \(||\lambda_i|| \leq M\), or stochastic such that \(E(||\lambda_i||^2) \leq M\);

(e) \(\lambda_i, \epsilon_{j,t}\) and \(F_s\) are mutually independent for all \(i, j, t\) and \(s\).

Assumption 2 (a) is quite common (see, for example, Moon and Perron, 2004; Pesaran, 2007; Phillips and Sul, 2003) and ensures that the serially uncorrelated error \(\epsilon_{i,t}\) has a strict common factor structure. This is more restrictive than the approximate factor model considered by Bai and Ng (2004, 2010), in which the idiosyncratic error is allowed to be “mildly” cross-correlated; however, it is necessary for the proofs. Assumption 2 (d) is, on the other hand, more relaxed than in Bai and Ng (2004, 2010), who require \(E(||\lambda_i||^4) \leq M\) in case \(\lambda_i\) is stochastic. The reason for this is that we do not require consistent estimation of \(\lambda_i\). Also, while the proofs make use of the assumption that \(F_t\) is homoskedastic, this is actually not necessary, and can be relaxed at the expense of added technical complexity. The only requirement is that \(T^{-1} \sum_{t=1}^T E(F_tF_t')\) has a limit such as \(\Sigma_F\).

Define the \(N \times k\) and \(N \times r\) matrices \(Z = (Z_1, ..., Z_N)'\) and \(\lambda = (\lambda_1, ..., \lambda_N)'\), respectively. The columns of \(Z\) can be deterministic and/or stochastic, provided that the following is satisfied.

**Assumption 3.**

(a) \(\overline{C} = N^{-1}\lambda'Z \rightarrow_p C\) as \(N \rightarrow \infty\), where \(\text{rk}(C) = r \leq k\) and \(||C|| \leq M\);

(b) \(Z_i\) is either deterministic such that \(||Z_i|| \leq M\), or stochastic such that \(E(||Z_i||^2) \leq M\);

(c) \(Z_i\) is independent of \(\epsilon_{j,t}\) for all \(i, j\) and \(t\).
As already mentioned, $Z$ might be thought of as providing an instrument for $\lambda$. Assumption 3 can therefore be regarded as an identifying condition with (a) and (c) corresponding to the usual conditions of instrument validity and orthogonality, respectively. As such, it is important to point out that, as long as the instruments are valid on average, Assumption 3 (a) actually do not preclude instrument failure for some units (see Section 3 for a discussion). Moreover, if $Z$ is deterministic, then Assumption 3 (c) is trivially satisfied.

Remark 1. Assumption 3 is materially different from that of Pesaran et al. (2009), who considers a similar multi-factor error structure, which is dealt with by bringing in additional information in the form of other regressors. Similar to Assumptions 3 (a) and (c) they require that the total number of variables (including $Y_{i,t}$) is at least $r$, and that the idiosyncratic error of the regressors is independent of that of $Y_{i,t}$. The main difference is that in Pesaran et al. (2009) the regressors are assumed not only to contain an exact unit root but also to be cointegrated and “cotrended” with each other and with $Y_{i,t}$, which seems unlikely to hold in practice. Another difference is that while here the instruments are effectively cross-section variables, in Pesaran et al. (2009) the regressors must also have a time series dimension. The current approach is therefore less demanding when it comes to the number of (additional) observations that needs to be brought in.

4 The IV test statistic

Define the projection matrix

$$M_A = I_{T-h+1} - A(A' A)^{-1} A',$$

where $h = \max\{q - 1, p + 2\}$ and $A$ is a $(T-h+1)$-rowed matrix. Let us also introduce the following detrended version of the (generic) variable $a_t$, defined for $t = h, ..., T$:

$$a^d_t = a_t - \sum_{n=2}^{t} a_n d_n' \left( \sum_{n=2}^{t} d_n d_n' \right)^{-1} d_t.$$

In this notation, the IV estimator of $F = (F_h, ..., F_T)'$ is given by

$$\hat{F} = \frac{1}{N} \sum_{i=1}^{N} M_{x_i} y^d_i Z_i',$$

where $x_i = (x_{i,h}, ..., x_{i,T})'$ and $y_i = (y_{i,h}, ..., y_{i,T})'$ are $(T-h+1) \times p$ and $(T-h+1) \times 1$, respectively, $x_{i,t} = (y^d_{i,t-1}, ..., y^d_{i,t-p})'$ is $p \times 1$. The way that the IV test statistic is designed is
to infer $H_0$ by testing for a unit root in the idiosyncratic part of the data. It is therefore useful to introduce $r_i = M_{i}^{'}M_{i} y_i^{'} = (r_{i,T}, \ldots, r_{i,T})', \text{ where } r_{i,t}$ can be seen as an estimator of $\epsilon_{i,t}$ (under $H_0$). Our estimator of the idiosyncratic part of the data in levels, henceforth denoted $R_{i,t}$, is simply the cumulative sum of $r_{i,t}$, that is, $R_{i,t} = \sum_{t=1}^{T} r_{i,t}$. A IV-based $t$-statistic for a unit root can now be set up as

$$t\text{-IV} = \frac{\text{tr}(R_{-1}^{'} \hat{\Sigma}_e^{-1} r)}{\sqrt{\text{tr}(R_{-1}^{'} \hat{\Sigma}_e^{-1} R_{-1})}}$$

where $r = (r_{1}, \ldots, r_{N})$ and $R_{-1} = (R_{-1,1}, \ldots, R_{N,-1})$ are $(T - h + 1) \times N$, while $R_{t-1} = (R_{t-1,1}, \ldots, R_{t-1,T-1})'$ is $(T - h + 1) \times 1$. Also, $\hat{\Sigma}_e = \text{diag}(\sigma_{\epsilon_1}^2, \ldots, \sigma_{\epsilon_N}^2)$, where $\sigma_{\epsilon}^2 = T^{-1}r_{i,t}^2$.

**Remark 2.** $t$-IV is similar to the test statistics of Bai and Ng (2010), and Westerlund (2014). The main difference when compared to Bai and Ng (2010) is that they only consider the case with at most a linear trend. Moreover, in contrast to here, the factors are estimated by principal components. Westerlund (2014) allows for a similarly general trend specification; however, in his study $F_t$ is assumed to be a scalar ($r = 1$), which of course need not be the case in practice.

**Lemma 1.** Under $H_0$ and Assumptions 1–3, as $N, T \to \infty$ with $N/T \to 0$,

$$\text{tr}(N^{-1/2} T^{-1} R_{-1}^{'} \hat{\Sigma}_e^{-1} r) \to_d N(0, \Sigma),$$

$$\text{tr}(N^{-1} T^{-2} R_{-1}^{'} \hat{\Sigma}_e^{-1} R_{-1}) \to_p \Sigma,$$

where $\to_d$ signifies convergence in distribution, and

$$\Sigma = \int_{v=0}^{1} b(v)dv,$$

$$b(v) = v - \int_{w=0}^{v} \int_{u=0}^{w} a(u, w)dudw + \int_{w=0}^{v} \int_{u=w}^{v} a(w, u)dudw,$$

$$a(v, w) = d(v)' \left( \int_{u=0}^{w} d(u)d(u)du \right)^{-1} d(w).$$

Thus, not only is the numerator of $t$-IV statistic unbiased, but the numerator and denominator are proportional to the same variance parameter, $\Sigma$, reflecting the choice of $D_I$. The implication is that

$$t\text{-IV} = \frac{\text{tr}(R_{-1}^{'} \hat{\Sigma}_e^{-1} r)}{\sqrt{\text{tr}(R_{-1}^{'} \hat{\Sigma}_e^{-1} R_{-1})}} \to_d N(0, 1)$$
as $N, T \to \infty$ with $N/T \to 0$, showing how the asymptotic null distribution of $t$–IV is invariant to all the parameters of the model, including the choice of $D_t$. Summarizing this, we have the following theorem.

**Theorem 1.** Under the conditions of Lemma, as $N, T \to \infty$ with $N/T \to 0$,

$$t$–IV $\to_d N(0, 1).$$

The first thing to note about Theorem 1 is that (as long as Assumption 1 is satisfied) it makes no assumption regarding the functional form of the deterministic component, suggesting that the asymptotic distribution is the same regardless of the specification of $D_t$. This result stands in sharp contrast to existing results. In fact, most panel unit root tests are biased even in the simple case when $D_t = 1$.\(^4\) $t$–IV is not only unbiased, but is in fact asymptotically invariant with respect to both the true and fitted deterministic specification. The usefulness of this property cannot be overstated. As is now well understood, the presence of deterministic functions in the test regression affects the asymptotic distribution of all unit root test statistics. In panels this means that the appropriate mean and variance correction factors of the tests change with the specification of the trend functions, necessitating the use of different statistical tables according to the precise specification of the fitted model. This dependence can pose serious problems to implementation. If the chosen specification involves shift dummy variables, for example, then the appropriate correction factors depend on the location of the shift. However, there is not only the practical problem of how to tabulate the correction factors; as alluded in Section 2, there is also the analytical problem that the complexity of the calculations involved in obtaining these factors increase very quickly with both the number and non-linearity of the trend terms. Therefore, researchers typically only provide correction factors for the simple case with (at most) a linear trend, thereby constraining the use of the tests to panels that are characterized by similarly simplistic deterministic behavior. Most panel unit root tests also require that chosen deterministic specification is the same for all units. This means that, for tests to be asymptotically invariant to the true trend function, researchers have to adopt a liberal modeling strategy to ensure that the deterministic behaviors of all the units are captured. Hence, even a very small fraction of units

\(^4\)See Westerlund and Breitung (2013, Section 3) for a discussion regarding the impact of different detrending procedures.
exhibiting non-linear behavior is enough to invalidate testing based on correction factors for the case with a linear trend.

**Remark 3.** As alluded in Section 2, \( \hat{F}_t \) is consistent for the space spanned by \( F_t \). In order to provide some intuition behind the working of this result, suppose for simplicity that the researcher knows that \( \phi_i(L) = 1 \), in which case a restricted version of \( \hat{F}_t \) can be constructed as

\[
\hat{F} = N^{-1} \sum_{i=1}^{N} y_d' Z_i' = N^{-1} y_d' Z,
\]

where \( y_d = (y_d', \ldots, y_d'_{N}) \). Hence, making use of the fact that under \( H_0 \), \( y_d = F_d \lambda' + e_d \), where \( e_d = (e_d', \ldots, e_d'_{N}) \), we obtain

\[
\hat{F} = N^{-1} y_d' Z = F_d N^{-1} \lambda' Z + N^{-1} e_d' Z = F_d \bar{C} + N^{-1} e_d' Z = F_d \bar{C} + O_p(N^{-1/2}),
\]

where the order of the remainder follows from the fact that \( e_d'_{i,t} \) is mean zero, uncorrelated with \( Z_i \) and independent across \( i \). Hence, \( \hat{F} \) is consistent, but not for \( F_d \); only for \( F_d \bar{C} \), which is enough for our purposes. The rotation by \( \bar{C} \) here illustrates the need for the rank condition in Assumption 3 (a). Suppose, for example, that \( r = 1 \), but \( \bar{C} = 0 \). In this case there is a single common factor present. However, since \( F_d \bar{C} = 0 \), \( \hat{F} \) will be unable to capture it.

**Remark 4.** The fact that \( t \)-IV is asymptotically \( N(0,1) \) stands in sharp contrast to the related work of Pesaran (2007) and Pesaran et al. (2009), who, unlike most other research in the field based on principal components analysis (see, for example, Moon and Perron, 2004; Bai and Ng, 2004, 2010), use both \( \bar{y}_t \) and \( \bar{Y}_t \) to approximate \( F_t \), leading to asymptotic distributions that depend on the Brownian motion associated with the cumulative sum process of \( F_t \). As alluded in Section 2, this difference is due to the fact that here it is only the first-differenced data that are “defactored” and then accumulated up to levels.

**Remark 5.** While here we focus on the use of \( t \)-statistic, this is not necessary. Indeed, since \( \hat{\Sigma}^{-1/2} r \) is not only (asymptotically) serially and cross-sectionally uncorrelated, but also homoskedastic, any test statistic can in principle be used, and this without the need for any further nuisance parameter corrections. Moreover, because of the martingale property of \( R_{i,t} \), there is no need for any bias correction. To also eliminate the dependence on the chosen deterministic specification, use of Lemma 1 necessitates that the test statistic is of the form of a ratio. One possibility that was recently considered by Westerlund and Larsson (2012) is to use the Lagrange multiplier principle, leading to test statistics that are of the form \( \text{tr}(R'_{-1} \hat{\Sigma}^{-1} r)^2 / \text{tr}(R'_{-1} \hat{\Sigma}^{-1} R_{-1}) \).
Remark 6. Theorem 1 requires that \( N/T \to 0 \) as \( N, T \to \infty \), which in practice means that \( N < T \). The reason for this requirement is the assumed heterogeneity of \( \beta_i \) and \( \phi_i(L) \) across \( i \), whose elimination induces an estimation error in \( T \), which is then aggravated when pooling across \( N \). The condition that \( N/T \to 0 \) prevents this error from having a dominating effect.

Of course, one may argue that in applications the above results are somewhat “idealized”, in the sense that the instruments in \( Z \) might be difficult to find. However, we argue that this criticism need not be too much of a problem. As in classical IV regressions \( Z \) is ideally chosen to be orthogonal to the error term, here represented by \( \epsilon_{it} \), but highly correlated with the quantity it replaces, that is, with \( \lambda \), measuring the exposure of each individual unit to the common shocks. As such, observable proxies for \( \lambda \) are actually not that difficult to find. For example, if \( Y_{it} \) is GDP or inflation, then \( Z \) might be a vector of trade shares, or, if \( Y_{it} \) is stock prices, then \( Z \) might be the book-to-market ratio and/or the earnings-price ratio. In fact, \( Z \) can even include preliminary consistent estimates of (the space spanned by) \( \lambda \). The only requirement is that the rate of consistency must be at least \( \sqrt{N} \), which is sufficiently relaxed to enable estimation by, for example, principal components (see Bai, 2003, Theorem 2). This is illustrated in Section 5.

Deterministic instruments are also not difficult to come by. Again, if \( Y_{it} \) is GDP, then \( Z \) might be a vector of distance rankings, or, if \( Y_{it} \) is stock prices, then \( Z \) might be industrial classification. Pesaran (2007) assumes that \( Z \) is deterministic and sets \( Z_i = 1 \), suggesting that in this case \( \hat{F} = N^{-1} \sum_{i=1}^{N} M_i y_i' Z_i ' = N^{-1} \sum_{i=1}^{N} M_i y_i' \). Hence, \( \hat{F} \) is just the average \( M_i y_i' \), and the cross-sectional average has been shown to be quite effective in mopping up cross-section dependence (see, for example, Chudik et al., 2011). A vector of ones is therefore a good starting point when constructing \( Z \). Of course, even if \( Z \) is assumed to be deterministic, the choice \( Z_i = 1 \) is just one possibility. In fact, if \( r = 1 \) is known the only condition required for identification is that \( \text{rk}(\lambda' Z) = 1 \), which does not rule out other choices of \( Z \), such as dummy variables, alternating sign series, and so on. The current IV approach can therefore be seen as a multi-factor extension of the cross-section average approach of Pesaran (2007).\(^5\)

As for the selection the appropriate number of instruments, \( k \), this is actually completely analogous to the problem of selecting the number of factors in principal components analysis. According to Assumption 3, the assumed number of factors, \( k \), must be at least as large

\(^5\)The approach considered here can also be viewed more broadly as an extension of the common correlated effects (CCE) approach of Pesaran (2006).
as the true number, \( r \). In case of principal components analysis this means that one has to put \( k \) large enough in the estimation of the factors, whereas here it means that the number of instruments in \( Z \) must be large enough. Therefore, in analogy with the usual practice in principal components analysis (see Bai and Ng, 2002), when there is uncertainty over \( k \), we recommend using an information criterion. There are several possibilities, although the criteria developed by Bai and Ng (2002) for the case of principal components estimation seem to work quite well also in the current context; see Section 5.

The above discussion supposes that the instruments are valid, which in practice need not be the case. A complete analysis of the impact of “weak instruments” is beyond the scope of the present study.\(^6\) However, it might be noted that Assumption 3 (a) actually do not preclude instrument failure for some units, as long as the instruments are valid on average. In fact, if we let \( C_i = E(\lambda_i Z_i') \), for Assumption 3 (a) to hold it is enough that the fraction of units for which the condition \( \text{rk}(C_i) = r \leq k \) holds does not go to zero as \( N \to \infty \). If for some units the condition fails, while for others the failure is local in the sense that \( C_i = N^{-\kappa}c \) with \( \text{rk}(c) = r \leq k \) and \( \kappa > 0 \), then the results reported herein continues to hold as long as \( \kappa \in [0, 1/2) \). In the next section we use Monte Carlo simulation to evaluate the effect of the strength of the instruments in small samples.

### 5 Monte Carlo simulations

A small-scale simulation study was conducted to assess performance of the new test in small samples, and in so doing we will focus on the first- and second-stage instrumentations. The DGP is a restricted version of the one given in (1)–(4), and sets \( \rho_i = \rho, r = 1, \lambda_i \sim N(1, 1), U_{i,0} = 0 \) and \((F_i, e_{i,t})' \sim N(0, I_2)\). As expected, the presence of serial correlation did not have any effects on the results, and we therefore set \( \phi_i(L) = 1 \), although we do not assume knowledge of this in the construction of the \( t \)–IV test statistic. The simplicity of the above DGP enables us to focus on the deterministic specification. In the paper we focus on the case when \( \beta_i = (1, ..., 1)' \) (a \( q \)-vector of ones); alternative parameterizations did not affect the results.

In the implementation, we start by detrending \( y_i \). Three specifications are considered; (i) \( D_t = (1, t)' \) (a linear time trend), (ii) \( D_t = (1, t, t^2)' \) (a quadratic trend) and (iii) \( D_t = (1, s_t)' \),

\(^6\)See Andrews and Stock (2005) for a good overview of the weak instrument literature.
where \( s_t = \frac{1}{1 + \exp(-5(T^{-1}t - 0.5))} \) is the logistic function (a smooth transition break in the level). The detrending gives \( y_t^d \), which is then used as input in the selection of the appropriate lag augmentation order, \( p \). This is done by applying the Schwarz Bayesian information criterion (BIC) to \( M_1 y_t^d \), where \( Y_{i,-1} = (Y_{i,-2}, ..., Y_{i,T-1})' \). The maximum number of lags is set to \( p_{\text{max}} = \lfloor 4(T/100)^{2/9} \rfloor \). The next step is the defactoring. Both deterministic and stochastic choices of \( Z_t \) are considered. In the former case, we set \( Z_t = 1 \), whereas in the latter, we set \( Z_t = \alpha \lambda_i + u_i \), where \( u_i \sim N(0,1) \), and use \( \alpha \) to control the strength of the instrument, as measured by the correlation between \( Z_t \) and \( \lambda_i \), \( \alpha / \sqrt{\alpha^2 + 1} \). Two values of \( \alpha \) are considered, 0.3 and 1, corresponding a correlation of 0.3 and 0.7, respectively. We also consider setting \( Z_t = \hat{\lambda}_i \), where \( \hat{\lambda}_i \) in the principal components estimator of \( \lambda_i \) based on \( M_1 y_t^d \). The fifth and last instrumentation strategy involves using the IC\(_1\) information criterion of Bai and Ng (2002) to select the “best” combination of the four previous candidates. The information criterion is applied to \( M_1 y_t^d \), which means the selection is done conditional on the chosen lag augmentation order. All-in-all, we have five sets of instruments, each of which leads to an estimate of (the space spanned by) \( F_t \), and hence a specific \( t-IV \) test statistic. For brevity, we focus on the size and power at the 5% level when the critical value \(-1.645\) is used. All results are based on 3,000 replications.

The results reported in Table 1 are generally in agreement with theory and can be summarized as follows:

- Except possibly for the test based on the weak instrument, size accuracy is generally quite good, even for samples as small as \( N = 20 \) and \( T = 70 \). Of course, accuracy is not perfect, and some distortions seem to remain; however, things generally improve as \( N \) and \( T \) increases.

- The fact that size accuracy does not change much depending on the specification of \( D_t \) is in accordance with our expectations, as according to Theorem 1 asymptotically there should be no dependence on the trend function. In fact, the size is surprisingly stable, especially given how size accuracy is typically decreasing in the order of the fitted trend function (see, for example, Pesaran et al., 2009).

- Unless the weak instrument is used, there are no major differences in the size results

\(^7\)A number of alternative trend function specifications were considered, ranging from the simple standard specification of (at most) a linear trend, to relatively complex models of single and multiple trend breaks. However, since the conclusions were the same, these results are not reported, but are available upon request.
depending on the treatment of the common factors. The fact that the deterministic instrument seems to lead to relatively good test performance provides credence for our previous proposal of always include in Z a vector of ones (see Section 4).

- In terms of power, there is very little variation in the results across instrumentation strategies, suggesting that even very simple instruments, such as a constant can be used with little or no cost in terms of power. The fact that the test based on the weak instrument seems to perform relatively well, especially among the smaller sample sizes, is to be expected given its distortions under the null.

- As expected, while size accuracy is largely constant across deterministic specifications, power is not. Hence, as usual, the deterministic specification should be carefully chosen to ensure a small number of fitted trend terms yet still capture the essential features of the trending behavior.

- Unreported results on the frequency count suggest that the true number of lags (0) is estimated with very high accuracy. In fact, with $T \geq 100$, accuracy is almost perfect. The BIC therefore seem to work well in the present context. The $IC_1$ seems to be doing a similarly very good job in selecting the appropriate number of instruments/factors (1). In particular, while among the smaller sample sizes there is a tendency to select too many instruments, with $N \geq 20$ the estimated number of instruments is exactly one. The most frequently selected instruments are the deterministic and strong stochastic ones. More importantly, the weak instrument is never selected.

Overall, it appears that our theoretical results provide a quite useful guide to the small-sample behavior of the new test. In particular, they predict very well that under $H_0$ $t$–IV should be asymptotically independent of the both the true and fitted trend functions. Moreover, as argued in Section 4, the need of an instrument to purge the effect of the common factors seems not to be detrimental, as the $IC_1$-based selection procedure considered here leads to good test performance even if some of the instrument candidates are weak. The good performance of the deterministic and principal component-based instruments suggests a very easy way of selecting potential candidates.
6 Conclusion

A new test of the hypothesis of a unit root in panel data was proposed. The test is asymptotically valid in the presence of general serial and cross-section correlation, and require only minimal corrections to that end. It is also asymptotically invariant with respect to a very broad range of deterministic trend functions, which is true even in absence of corrections for the usual dependence on the chosen deterministic specification. This feature not only makes the test statistic easy to compute, but also enables testing in general when the deterministic component of the data is not necessarily made up of a constant and trend.
References


Appendix: Proofs

We start with some notation. Under $H_0$, according to (1)–(4), with $y_{i,t} = \Delta Y_{i,t}$,

$$\phi_i(L)y_{i,t} = \phi_i(L)(\beta'_i \Delta D_t + \Delta U_{i,t}) = \phi_i(L)\beta'_i \Delta D_t + \epsilon_{i,t} = \phi_i(L)\beta'_i \Delta D_t + \lambda'_i F_t + \epsilon_{i,t},$$

giving

$$y_{i,t} = \sum_{k=1}^{p} \phi_{k,i} y_{i,t-k} + \left(\beta'_i \Delta D_t - \sum_{k=1}^{p} \phi_{k,i} \beta'_i \Delta D_{i,t-k} \right) + \lambda'_i F_t + \epsilon_{i,t}$$

$$= \sum_{k=1}^{p} \phi_{k,i} u_{i,t-k} + \beta'_i \Delta D_t + \lambda'_i F_t + \epsilon_{i,t}.$$  

The corresponding expression for $u_{i,t}$ is given by

$$u_{i,t} = \sum_{k=1}^{p} \phi_{k,i} u_{i,t-k} + \lambda'_i F_t + \epsilon_{i,t},$$

suggesting that, since $\beta'_i \Delta D_t = \beta'_i G' \Delta D_t = \beta'_i G' d_t$,

$$y_{i,t} = u_{i,t} + \beta'_i \Delta D_t = u_{i,t} + \beta'_i G' \Delta D_t = u_{i,t} + \beta'_i G' d_t,$$  \hfill (A1)

and so, by the properties of OLS, $y^d_{i,t} = u^d_{i,t}$. Hence, letting $\phi_i = (\phi_{1,i}, ..., \phi_{p,i})'$ and $x_{i,t} = (y^d_{i,t-1}, ..., y^d_{i,t-1})'$, we have

$$y^d_{i,t} = \phi'_i x_{i,t} + \lambda'_i F^d_t + \epsilon^d_{i,t}.$$  \hfill (A2)

This equation can be written in matrix notation as

$$y^d_t = x_t \phi_t + F^d \lambda_t + \epsilon^d_t,$$  \hfill (A3)

where $y_t = (y_{i,t}, ..., y_{T,t})'$ and $\epsilon_t = (\epsilon_{i,t}, ..., \epsilon_{i,T})'$ are $(T - h + 1) \times 1$, $F = (F_{h}, ..., F_T)'$ is $(T - h + 1) \times r$, and $x_t = (x_{i,t}, ..., x_{i,T})'$ is $(T - h + 1) \times p$. Alternatively, we may write the equation for $y^d_{i,t}$ as the following $N$-dimensional system:

$$y^d_t = \phi x_t + F^d \lambda_t + \epsilon^d_t,$$  \hfill (A4)

where $y_t = (y_{1,t}, ..., y_{N,t})'$ and $\epsilon_t = (\epsilon_{1,t}, ..., \epsilon_{N,T})'$ are $N \times 1$, $x_t = (x'_{1,t}, ..., x'_{N,t})'$ is $N \times 1$ and $\phi = \text{diag}(\phi_1', ..., \phi_N')$ is $N \times Np$. The matrix notation

$$y^d = x_\phi + F^d \lambda^d + \epsilon^d$$  \hfill (A5)

will also be used, where $y = (y_{h,T})'$ and $\epsilon = (\epsilon_{h}, ..., \epsilon_{T})'$ are $(T - h + 1) \times N$, while $x = (x_{h}, ..., x_{T})'$ is $(T - h + 1) \times Np$. In what follows the representations in (A2)–(A5) will be used interchangeably.
In what follows it is going to be convenient to let \( P_A = A(A'A)^{-1}A' \), such that \( M_A = I_{T-h+1} - P_A \). In this notation, we have that, under \( H_0 \),

\[
M_{x,i}y_i^T = M_{x,i}(F^d\lambda_i + \epsilon_i^d) = F^d\lambda_i + \epsilon_i^d - P_{x,i}F^d\lambda_i = F^d\lambda_i + v_i, \tag{A6}
\]

where the first equality holds because \( M_{x,x} = 0 \) and

\[
v_i = M_{x,i}e_i^d - P_{x,i}F^d\lambda_i = \epsilon_i^d - P_{x,i}(F^d\lambda_i + e_i^d)
\]

is \( (T-h+1) \times 1 \). In order to capture the fact that \( F_t \) and \( \hat{F}_t \) may be of different dimension, we introduce the \( r \times k \) matrix \( \overline{C} = N^{-1}\lambda'Z = N^{-1}\sum_{i=1}^N\lambda_iZ_i' \). This matrix, whose rank is given by \( r \leq k \) (Assumption 3), will be use to rotate \( F \). In terms of the notational convention of the principal component literature (see, for example, Bai and Ng 2002), we have \( H = \overline{C} \). It follows that

\[
\hat{F} = \frac{1}{N} \sum_{i=1}^N M_{x,i}y_i^T = F^d\frac{1}{N} \sum_{i=1}^N \lambda_iZ_i' + \frac{1}{N} \sum_{i=1}^N v_iZ_i' = F^dH + N^{-1}vZ, \tag{A7}
\]

where \( v = (v_1, ..., v_N)' \) is \( (T-h+1) \times N \). Let us further define \( U = \hat{F} - F^dH \), whose \( t^{th} \) element is given by \( U_t = \hat{F}_t - H'F_t^d \). Direct substitution from (A7) yields

\[
U = \hat{F} - F^dH = N^{-1}vZ,
\]

such that \( U_t = N^{-1}Z'v_t \) with \( v_t \) being the \( t^{th} \) row of \( v \). Let \( H^+ = H(H'H)^{-1} \). In what follows we will make frequent use of the following decomposition of the common component:

\[
F^d\lambda_i = F^dHH'(HH')^{-1}\lambda_i = F^dHH^+\lambda_i = \hat{F}H^+\lambda_i - (\hat{F} - F^dH)H^+\lambda_i
\]

\[
= \hat{F}H^+\lambda_i - UH^+\lambda_i, \tag{A8}
\]

such that under \( H_0 \), \( M_{x,i}y_i^T \) can be written as

\[
M_{x,i}y_i^T = F^d\lambda_i + v_i = \hat{F}H^+\lambda_i - UH^+\lambda_i + v_i. \tag{A9}
\]

Before we come to the proof of Theorem 1 we report some useful lemmas. Here and throughout \( \mathcal{F}_t \) is going to denote a generic sigma-field. Unless otherwise stated, \( \mathcal{F}_t \) is assumed to be generated by \( \{F_s\}_{s=1}^t \) and \( \{\epsilon_s\}_{s=1}^t \). The following \( (T-h+1) \times (T-h+1) \) cumulative sum and lag matrices matrices will also be used:

\[
S = \begin{bmatrix}
1 & 0 & \ldots & 0 \\
1 & 1 & \ddots & \vdots \\
\vdots & \ddots & 0 \\
1 & \ldots & 1
\end{bmatrix},
\]
and

\[ L = \begin{bmatrix}
0 & 0 & \ldots & 0 \\
1 & 0 & \ddots & \vdots \\
0 & 1 & \ddots & 0 \\
\vdots & \ddots & \ddots & 0 \\
0 & \ldots & 0 & 1 & 0
\end{bmatrix}. \]

Note in particular how in this notation, under \( H_0 \), \( R_{i-1} = \text{LSM}_F M_{x_i y_i^d} \).

**Lemma A.1.** It holds that

\[
(M_{F^d H} - M_F) = U(\hat{F}' \hat{F})^{-1} U' + U(\hat{F}' \hat{F})^{-1} H'(F^d)' + F^d H(\hat{F}' \hat{F})^{-1} U' + F^d H[(\hat{F}' \hat{F})^{-1} - (H'(F^d)'F^d H)^{-1}] H'(F^d)'.
\]

**Proof of Lemma A.1.**

The proof is obvious and is therefore omitted.

**Lemma A.2.** Under \( H_0 \) and Assumptions 1–3, uniformly in \( i, j = 1, \ldots, N \),

\[
\begin{align*}
||T^{-1/2}(F^d)'x_i|| &= O_p(1), \\
||T^{-1}x_i'x_i|| &= O_p(1), \\
||T^{-1/2}x_i'e_i^d|| &= O_p(1), \\
||T^{-1}x_i'S'L'e_i^d|| &= O_p(1), \\
||T^{-1}x_i'S'L'x_j|| &= O_p(1), \\
||T^{-2}x_i'S'L'Se_i^d|| &= O_p(1), \\
||T^{-2}(F^d)'S'L'Se_i^d|| &= O_p(1), \\
||T^{-2}x_i'S'L'Sx_j|| &= O_p(1).
\end{align*}
\]

**Proof of Lemma A.2.**

Since \( x_i \) is stationary and ergodic with \( E(e_{i,t} x_{i,t}) = E[E(e_{i,t}|F_{t-1}) x_{i,t}] = 0 \), and similarly \( E(x_{i,t} F_i) = 0 \) (Assumption 2), it is clear that \( ||T^{-1/2}(F^d)'x_i||, ||T^{-1}x_i'x_i|| \) and \( ||T^{-1/2}x_i'e_i^d|| \) are all \( O_p(1) \). The rest of the results follow from application of Lemma 2.1 in Park and Phillips (1989).
Lemma A.3. Under the conditions of Lemma A.2,

\[
\begin{align*}
||T^{-1/2}(F^d)'U|| &= O_p(N^{-1/2}) + O_p(T^{-1/2}), \\
||NT^{-1}U'U|| &= O_p(NT^{-1}) + O_p(1), \\
||T[(F^d)' - (H'(F^d)'F^d)H]^{-1}|| &= O_p((NT)^{-1/2}) + O_p(T^{-1}) + O_p(N^{-1}), \\
||T^{-1/2}U'e^d|| &= O_p(N^{-1}\sqrt{T}) + O_p(N^{-1/2}) + O_p(T^{-1/2}), \\
||T^{-1}U'LSe^d|| &= O_p(N^{-1/2}) + O_p(T^{-1/2}), \\
||T^{-1}U'S'L'e^d|| &= O_p(N^{-1/2}) + O_p(T^{-1/2}), \\
||T^{-1}U'S'L'F|| &= O_p(N^{-1/2}) + O_p(T^{-1/2}), \\
||T^{-1}U'S'L'U|| &= O_p(N^{-1}) + O_p((NT)^{-1/2}) + O_p(T^{-1}), \\
||T^{-2}U'S'L'Se^d|| &= O_p(N^{-1/2}) + O_p(T^{-1/2}), \\
||T^{-2}U'S'L'SF^d|| &= O_p(N^{-1/2}) + O_p(T^{-1/2}).
\end{align*}
\]

Proof of Lemma A.3.

We start with \(T^{-1/2}(F^d)'U\), which, via the definitions of \(U\) and \(v\), can be written as

\[
T^{-1/2}(F^d)'U = T^{-1/2}(F^d)'N^{-1}vZ = \frac{1}{N\sqrt{T}} \sum_{i=1}^{N} (F^d)'[e^d_i - P_{x_i}(F^d\lambda_i + e^d_i)]Z'_i,
\]

where

\[
\frac{1}{N\sqrt{T}} \sum_{i=1}^{N} (F^d)'P_{x_i}(F^d\lambda_i + e^d_i)Z'_i = \frac{1}{N\sqrt{T}} \sum_{i=1}^{N} (F^d)'x_i(x'_i x_i)^{-1}x'_i(F^d\lambda_i + e^d_i)Z'_i.
\]

By Lemma A.2,

\[
\begin{align*}
\left|\frac{1}{N\sqrt{T}} \sum_{i=1}^{N} (F^d)'P_{x_i}(F^d\lambda_i + e^d_i)Z'_i\right| &\leq \frac{1}{\sqrt{T}} \sum_{i=1}^{N} ||T^{-1/2}(F^d)'x_i||||((T^{-1}x'_ix_i)^{-1}||T^{-1/2}x'_i(F^d\lambda_i + e^d_i)||||Z_i||
\end{align*}
\]

Consider \((NT)^{-1/2}\sum_{i=1}^{N} (F^d)'e^d_iZ'_i\). Because \(F_t\) and \(e_{it}\) are independent by Assumption 2, the mean is clearly zero. As for the variance, consider the \(N \times N\) matrix \(E[e^d_i(F^d)'e^d_i]\), defined for \(t = h, ..., T\), which has \(E[(e^d_i)^2]\) along the diagonal and \(E(e^d_i e^d_{ij})\) elsewhere. Let \(a_{kt} = d'_k(\sum_{n=1}^{N} d_n d'_n)^{-1}d_t\). By using the definition of \(e^d_i\) and \(E(e_i e'_i) = \text{diag}(\sigma^2_{e1}, ..., \sigma^2_{eN}) = \Sigma_e\)
(Assumption 2), we get

\[
E[e_t^d (e_t^d)'] = E \left[ \left( e_t - \sum_{k=2}^{t} e_k a_{k,t} \right) \left( e_t - \sum_{s=2}^{t} e_s a_{s,t} \right)' \right]
\]

\[
= E(e_t e_t') - \sum_{k=2}^{t} E(e_k e_k') a_{k,t} - \sum_{s=2}^{t} E(e_s e_s') a_{s,t} + \sum_{k=2}^{t} \sum_{s=2}^{t} E(e_k e_s') a_{k,t} a_{s,t}
\]

\[
= \left( 1 - 2a_{t,t} + \sum_{k=2}^{t} a_{k,t}^2 \right) \Sigma_e
\]

\[
= (1 - a_{t,t}) \Sigma_e,
\]

(A10)

where the last equality holds because \( \sum_{k=2}^{t} a_{k,t}^2 = a_{t,t} \). Moreover, since \( e_t \) is serially uncorrelated with \( \sum_{n=2}^{s} a_{n,t} a_{n,s} = a_{s,t} \), we also have, for \( t > s \),

\[
E[e_t^d (e_s^d)'] = E(e_t e_s') - \sum_{k=2}^{s} E(e_k e_k') a_{k,s} - \sum_{k=2}^{t} E(e_k e_k') a_{k,t} a_{s,t} + \sum_{k=2}^{t} \sum_{n=2}^{s} E(e_k e_n') a_{k,t} a_{n,s}
\]

\[
= -E(e_t e_s') a_{s,t} + \sum_{n=2}^{s} E(e_n e_n') a_{n,t} a_{n,s}
\]

\[
= \left( -a_{s,t} + \sum_{n=2}^{s} a_{n,t} a_{n,s} \right) \Sigma_e = 0.
\]

(A11)

The results in (A10) and (A11) imply that \( E[e_t^d (e_i^d)'] = \sigma_{e_t}^2 \text{diag}((1-a_{h,h}), \ldots, (1-a_{T,T})) = \sigma_{e_t}^2 (I_{T-h+1} - A_{T-h+1}) \), where \( A_{T-h+1} = \text{diag}(a_{h,h}, \ldots, a_{T,T}) \), and \( E[e_i^d (e_j^d)'] = 0 \) for all \( i \neq j \).

Thus, since the autocovariance structure of \( F_t^d \) is the same as that of \( e_t^d \) with \( \Sigma_e \) replaced by \( \Sigma_F \), we can show that, letting \( V_N = N^{-1} \sum_{i=1}^{N} \sigma_{e_t}^2 Z_i Z_j' = N^{-1} \Sigma e Z \),

\[
E \left[ \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} (F_i^d)' e_i^d Z_i' \right) \left( \frac{1}{\sqrt{NT}} \sum_{j=1}^{N} (F_j^d)' e_j^d Z_j' \right)' \right]
\]

\[
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} E[Z_i' Z_j'(F_i^d)' E(e_i^d (e_j^d)') Z, F] F^d]
\]

\[
= \frac{1}{NT} \sum_{i=1}^{N} E[Z_i' (F_i^d)' E(e_i^d (F_i^d)') Z, F] F^d]
\]

\[
= \frac{1}{NT} \sum_{i=1}^{N} \sigma_{e_t}^2 E[Z_i' (F_i^d)' (I_{T-h+1} - A_{T-h+1}) F^d]
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} \sigma_{e_t}^2 \frac{1}{T} \sum_{t=h}^{T} E[\text{tr}(Z_i Z_i') E(F_i^d (F_i^d)' Z)](1 - a_{t,t})
\]

\[
= E[\text{tr}(V_N)] \Sigma_F \frac{1}{N} \sum_{t=h}^{T} (1 - a_{t,t})^2,
\]

where the first and fourth equalities hold because \( Z_i' Z_j \) is a scalar. Consider \( a_{k,t} \) for \( k, t = \)
This establishes the first result. It follows that $T_{a,k} = d_k |_{T} \left( \int_{T}^{1} \sum_{n=1}^{T} d_n d_n |_{T} \right)^{-1} \int_{T}^{w} d(w) (\int_{u=0}^{w} d(u) d(u) d(u))^{-1} d(w) = a(v,w)$ as $T \to \infty$, which holds uniformly in $(v,w) \in [0,1] \times (0,1)$. If $w \to 0$, then, by the mean value theorem, $a(v,w) \to d(v)(wd(w)d(w))^{-1}d(w)$. Hence, $|a_{k,1}| = O(T^{-1})$, suggesting that $T^{-1} \sum_{i=1}^{T} (1-t_{i,t})^2 = 1 + O(T^{-1})$. But we also have $||V_N|| \leq N^{-1} \sum_{i=1}^{N} \sigma_{e,i}^2 ||Z_i||^2 = O_p(1)$ (Assumption 3), and therefore $||E(\text{tr}(V_N))|| \leq N^{-1} \sum_{i=1}^{T} (1-t_{i,t})^2 || = O(1)$, which in turn implies

$$\left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} (F^d)' e_i^d Z_i' \right| = O_p(1).$$

It follows that

$$||T^{-1/2}(F^d)' U|| = \left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} (F^d)' [e_i^d - P_x(F^d \lambda_i + e_i^d)] Z_i' \right| \leq \frac{1}{\sqrt{N}} \left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} (F^d)' e_i^d Z_i' \right| + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (F^d)' P_x(F^d \lambda_i + e_i^d) Z_i' \right| = O_p(N^{-1/2}) + O_p(T^{-1/2}).$$

This establishes the first result.

Next, consider $NT^{-1} U' U$. By the definitions of $U$ and $v$,

$$NT^{-1} U' U = (NT)^{-1} Z' Z$$

$$= \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i (e_i^d - P_x(F^d \lambda_i + e_i^d)' [e_i^d - P_x(F^d \lambda_j + e_j^d)] Z_j'$$

$$= \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i \left( (e_i^d)' e_i^d - (e_i^d)' P_x(F^d \lambda_j + e_j^d) - (F^d \lambda_i + e_i^d)' P_x(F^d \lambda_j + e_j^d) + (F^d \lambda_i + e_i^d)' P_x(F^d \lambda_j + e_j^d) Z_j' \right)$$

$$= I_1 + I_2 + I_3 + I_4,$$  

(A13)
where

\[ I_1 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i (e_i^d)' e_j^d Z_j' \]

\[ I_2 = -\frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i (e_i^d)' P_{x_j} (F^d \lambda_j + e_j^d) Z_j' \]

\[ I_3 = -\frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i (F^d \lambda_i + e_i^d)' P_{x_j} e_j^d Z_j' \]

\[ I_4 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i (F^d \lambda_i + e_i^d)' P_{x_j} P_{x_j} (F^d \lambda_j + e_j^d) Z_j' \]

Consider \( I_1 \), which we write as

\[ I_1 = \frac{1}{NT} \sum_{i=1}^{N} Z_i (e_i^d)' e_i^d Z_i' + \frac{1}{NT} \sum_{i=1}^{N} \sum_{j \neq i} Z_i (e_i^d)' e_j^d Z_j' \]

Consider the second term on the right, whose mean is clearly zero (by cross-section independence; Assumption 2). As for the variance, we have

\[
E\left( \left| \frac{1}{N} \sum_{i=1}^{N} \sum_{j \neq i} T^{-1} (e_i^d)' e_j^d Z_i' Z_j' \right|^2 \right) = \frac{1}{(NT)^2} \sum_{i=1}^{N} \sum_{j \neq i} \sum_{k=1}^{N} \sum_{l \neq k} E[ (e_i^d)' (e_j^d)' (e_k^d)' e_l^d | Z] \text{tr}(Z_i' Z_k' Z_l' Z_i') \]

\[
= \frac{1}{(NT)^2} \sum_{j=1}^{N} \sum_{n \neq j} E[ (e_n^d)' (e_j^d)' e_n^d | Z] \text{tr}(Z_j' Z_j' Z_j') \]

\[
= \frac{1}{T^2 N^2} \sum_{j=1}^{N} \sum_{n \neq j} \sigma_{e_n}^2 \sigma_{e_j}^2 E[ \text{tr}(Z_n' Z_n' Z_j' Z_j') ] T^{-1} \text{tr}[(I_{T-h+1} - A_{T-h+1})^2] \]

\[
= T^{-1} E[ \text{tr}(V_N^2) ] T^{-1} \text{tr}[(I_{T-h+1} - A_{T-h+1})^2] + O(N^{-1}) \]

\[
= O(T^{-1}) + O(N^{-1}),
\]

where the third equality follows from noting that, by cross-section independence and the cyclical property of the trace,

\[
E[(e_n^d)' (e_j^d)' e_n^d] = \sigma_{e_n}^2 E[(e_n^d)' (I_{T-h+1} - A_{T-h+1}) e_n^d]
\]

\[
= \sigma_{e_n}^2 \text{tr}[E((e_n^d)' (I_{T-h+1} - A_{T-h+1}))] \]

\[
= \sigma_{e_n}^2 \sigma_{e_j}^2 \text{tr}[(I_{T-h+1} - A_{T-h+1})^2],
\]

where \( T^{-1} \text{tr}[(I_{T-h+1} - A_{T-h+1})^2] = T^{-1} \sum_{t=h}^{T} (1 - a_{t})^2 = 1 + O(T^{-1}). \) Thus, since the variance is \( O(T^{-1}) + O(N^{-1}) \), the second term in \( I_1 \) is \( O_p(T^{-1/2}) + O_p(N^{-1/2}) \). Hence, since the
whose order can be deduced from (A12) and (A15). In fact,

\[ ||I_1|| = O_p(1) + O_p(T^{-1/2}) + O_p(N^{-1/2}) = O_p(1), \quad (A14) \]

As for the remaining terms in (A13), we have

\[
||I_2|| \leq \frac{\sqrt{N}}{T} \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} Z_i (\epsilon_i^d)' P_{x_i} (F_d^d \lambda_j + \epsilon^d_j) \right) || Z_j ||
\]

\[
\leq \frac{\sqrt{N}}{T} \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} Z_i (\epsilon_i^d)' x_j \right) \left( \| (T^{-1} x_j')^{-1} \| \| T^{-1/2} x_j'(F_d^d \lambda_j + \epsilon^d_j) \| \right) || Z_j ||
\]

\[ = O_p(\sqrt{NT}^{-1}), \]

with \( ||I_3|| \) being of the same order (in fact, since \( I_3 = I'_2 \), we have \( ||I_3|| = ||I_2|| \)). Also, from

\[
\| (F_d^d \lambda_i + \epsilon_i^d)' P_{x_i} (F_d^d \lambda_j + \epsilon^d_j) \|
\]

\[
= \| (T^{-1/2}(F_d^d \lambda_i + \epsilon_i^d)' x_i) \| (T^{-1} x_j')^{-1} \| T^{-1/2} x_j'(F_d^d \lambda_j + \epsilon^d_j) \|
\]

\[
\times \| (T^{-1} x_j')^{-1} \| \| T^{-1/2} x_j'(F_d^d \lambda_j + \epsilon^d_j) \|
\]

\[ = O_p(1), \]

we get

\[
||I_4|| \leq \frac{N}{T} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \| (F_d^d \lambda_i + \epsilon_i^d)' P_{x_i} (F_d^d \lambda_j + \epsilon^d_j) \| || Z_i || || Z_j || = O_p(NT^{-1}).
\]

which in turn implies, with \( O_p(\sqrt{NT}^{-1}) < O_p(NT^{-1}) \),

\[
||NT^{-1}U'U|| \leq ||I_1|| + ||I_2|| + ||I_3|| + ||I_4|| = O_p(1) + O_p(NT^{-1}). \quad (A15)
\]

Consider \( [(\hat{F}^d \hat{F})^{-1} - (H'(F^d)'F^d H)^{-1}] \). By adding and subtracting terms,

\[
\hat{F}^d \hat{F} - H'(F^d)'F^d H = (\hat{F} - F^d H)' \hat{F} + H'(F^d)'(\hat{F} - F^d H) = U' \hat{F} + H'(F^d)'U,
\]

suggesting

\[
(\hat{F}^d \hat{F})^{-1} - (H'(F^d)'F^d H)^{-1} = (\hat{F}^d \hat{F})^{-1} [\hat{F}^d \hat{F} - H'(F^d)'F^d H] (H'(F^d)'F^d H)^{-1}
\]

\[ = (\hat{F}^d \hat{F})^{-1} [U' \hat{F} + H'(F^d)'U] (H'(F^d)'F^d H)^{-1}, \]

whose order can be deduced from (A12) and (A15). In fact,

\[
||T^{-1/2}U' \hat{F}|| \leq ||T^{-1/2}U'F^d|| + \sqrt{T}N^{-1}||NT^{-1}U'U||
\]

\[
= [O_p(N^{-1/2}) + O_p(T^{-1/2})] + \sqrt{T}N^{-1}[O_p(1) + O_p(NT^{-1})]
\]

\[ = O_p(N^{-1/2}) + O_p(T^{-1/2}) + O_p(\sqrt{TN}^{-1}). \]

28
Therefore, since \(|\|T^{-1} H'(F^d)'F^d H||\) and \(|\|T^{-1} \hat{F}'\hat{F}||\) are \(O_p(1)\), we can show that

\[
T \| (\hat{F}'\hat{F})^{-1} - (H'(F^d)'F^d H)^{-1} \|
\leq T^{-1/2} \| (T^{-1/2} \hat{F}' - T^{-1/2} H'(F^d)'U) \| \| (T^{-1} H'(F^d)'F^d H)^{-1} \|
= T^{-1/2} [O_p(N^{-1/2}) + O_p(T^{-1/2}) + O_p(\sqrt{TN^{-1}})]
= O_p((NT)^{-1/2}) + O_p(T^{-1}) + O_p(N^{-1}),
\]

as required.

Next, consider

\[
T^{-1/2} U' e_i^d = \frac{1}{N\sqrt{T}} \sum_{j=1}^{N} Z_j [e_j^d - P_{x_j} (F^d \lambda_j + e_j^d)] e_i^d.
\]

We begin by noting that

\[
|\| (F^d \lambda_j + e_j^d)' P_{x_j} e_i^d \| \leq |\| T^{-1/2} (F^d \lambda_j + e_j^d)' x_j \| |\| (T^{-1} x_j) (T^{-1})^{-1} \| |\| T^{-1/2} x_j e_i^d \|
= O_p(1),
\]

which is true for all \(i, j = 1, ..., N\). Moreover, by cross-section independence (Assumption 2),

\[
\frac{1}{N\sqrt{T}} \sum_{j=1}^{N} Z_j (e_j^d)' e_i^d = N^{-1} \sqrt{T} T^{-1} Z_i (e_i^d)' e_i^d + \frac{1}{\sqrt{N}} \frac{1}{\sqrt{NT}} \sum_{j\neq i}^{N} Z_j (e_j^d)' e_i^d
= O_p(N^{-1/2}) + O_p(N^{-1/2}),
\]

which we can use to show that

\[
|\| T^{-1/2} U' e_i^d \| = \left| \frac{1}{N\sqrt{T}} \sum_{j=1}^{N} Z_j [e_j^d - P_{x_j} (F^d \lambda_j + e_j^d)] e_i^d \right|
\leq \left| \frac{1}{N\sqrt{T}} \sum_{j=1}^{N} Z_j (e_j^d)' e_i^d \right| + \frac{1}{\sqrt{N}} \frac{1}{\sqrt{NT}} \sum_{j=1}^{N} |Z_j| \| (F^d \lambda_j + e_j^d)' P_{x_j} e_i^d |
\leq O_p(N^{-1/2}) + O_p(N^{-1/2}) + O_p(T^{-1/2}).
\]

The same arguments used for establishing (A17) can be used to show that

\[
\frac{1}{\sqrt{TN}} \sum_{j=1}^{N} |Z_j||T^{-1/2} (F^d \lambda_j + e_j^d)' P_{x_j} LSe_i^d | = O_p(T^{-1/2}),
\]

suggesting

\[
T^{-1} U' LSe_i^d = \frac{1}{NT} \sum_{j=1}^{N} Z_j (e_j^d)' LSe_i^d + O_p(T^{-1/2}),
\]

(A18)

where the first term on the right is \(O_p(N^{-1/2})\).
Consider $T^{-1}U'S'L'e_i^d$, which can be expanded in the following fashion:

$$
T^{-1}U'S'L'e_i^d = \frac{1}{NT} \sum_{j=1}^{N} Z_j (e_j^d - P_j(F^d \lambda_j + e_j^d))^\prime S'L'e_i^d
$$

$$
= \frac{1}{NT} \sum_{j=1}^{N} Z_j (e_j^d)^\prime S'L'e_i^d - \frac{1}{NT} \sum_{j=1}^{N} Z_j (F^d \lambda_j + e_j^d)^\prime P_j S'L'e_i^d,
$$

(A19)

where, by Lemma A.2,

$$
\left| \left| T^{-1}(F^d \lambda_j + e_j^d)^\prime P_j S'L'e_i^d \right| \right| 
\leq T^{-1/2} \left| \left| T^{-1/2}(F^d \lambda_j + e_j^d)^\prime x_j \right| \right| \left| \left| (T^{-1}x_j^{-1})^\prime T^{-1}x_j^\prime S'L'e_i^d \right| \right| 
= O_p(T^{-1/2}).
$$

As for the remaining (first) term, by the definitions of $L$ and $S$,

$$
E \left( \left| \frac{1}{NT} \sum_{j=1}^{N} Z_j (e_j^d)^\prime S'L'e_i^d \right|^2 \right)
$$

$$
= \frac{1}{(NT)^2} \sum_{j=1}^{N} \sum_{k=1}^{N} E[(e_j^d)^\prime S'L'e_i^d(e_k^d)^\prime LSe_k^d \text{tr}(Z_jZ_k^\prime)]
$$

$$
= \frac{1}{(NT)^2} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{T} \sum_{h=1}^{h} \sum_{s=1}^{s} \sum_{t=1}^{t} E[e_j^d e_i^d e_l^d e_{k,m}^d \text{tr}(Z_jZ_k^\prime)]
$$

$$
= \frac{1}{(NT)^2} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{T} \sum_{h=1}^{h} \sum_{s=1}^{s} \sum_{t=1}^{t} E[e_j^d e_i^d | Z, F_{t-1}] e_l^d e_{k,m}^d \text{tr}(Z_jZ_k^\prime)]
$$

$$
+ \frac{2}{(NT)^2} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{T} \sum_{h=1}^{h} \sum_{s=1}^{s} \sum_{t=1}^{t} \sum_{m=1}^{m} E[e_j^d e_i^d | Z, F_{t-1}] e_l^d e_{k,m}^d \text{tr}(Z_jZ_k^\prime)]
$$

$$
= \frac{1}{(NT)^2} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{T} \sum_{h=1}^{h} \sum_{s=1}^{s} \sum_{t=1}^{t} \sigma_{e,i}^2 (1 - a_{t,l}) E[E(e_j^d e_{k,m}^d | Z, F_{t-1}) \text{tr}(Z_jZ_k^\prime)]
$$

$$
= \frac{1}{(NT)^2} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{T} \sum_{h=1}^{h} \sum_{s=1}^{s} \sum_{t=1}^{t} \sigma_{e,i}^2 \sigma_{e,j}^2 (1 - a_{t,l})(1 - a_{m,m}) E[\text{tr}(Z_jZ_k^\prime)]
$$

$$
= \frac{1}{NT^2} \sum_{t=1}^{T} \sigma_{e,i}^2 (1 - a_{t,l})(1 - a_{m,m}) E[\text{tr}(V_N)]
$$

$$
= O(N^{-1}),
$$

where the last result is due to the fact that

$$
\frac{1}{T^2} \sum_{t=1}^{T} \sum_{t=1}^{T} (1 - a_{t,l})(1 - a_{m,m}) = \frac{1}{T^2} \sum_{t=1}^{T} t + O(T^{-1}) \to \int_{v=0}^{1} vdv = \frac{1}{2}
$$

as $T \to \infty$. Hence,

$$
\left| \frac{1}{NT} \sum_{j=1}^{N} Z_j (e_j^d)^\prime S'L'e_i^d \right| = O_p(N^{-1/2}),
$$

30
which is true for all $i = 1, \ldots, N$. It follows that
\[
\|T^{-1}U'S'L'e_i^d\| \leq \left\| \frac{1}{NT} \sum_{j=1}^N Z_j (e_j^d)' S'L'e_i^d \right\| + \frac{1}{N} \sum_{j=1}^N \|Z_j\| \|T^{-1}(F^d\lambda_j + e_j^d)' P_{x_j} S'L'e_i^d\|
\]
\[
= O_p(N^{-1/2}) + O_p(T^{-1/2}). \tag{A20}
\]

The proof of
\[
\|T^{-1}U'S'L'F^d\| = O_p(N^{-1/2}) + O_p(T^{-1/2}) \tag{A21}
\]
is almost identical to the one just given, and is therefore omitted.

Consider $T^{-1}U'S'L'U$. Write
\[
T^{-1}U'S'L'U = \frac{1}{N^2T} \sum_{j=1}^N \sum_{i=1}^N Z_j (e_j^d)' P_{x_j} (F^d\lambda_j + e_j^d)' S'L' (e_i^d)' P_{x_i} (F^d\lambda_i + e_i^d) Z_i',
\]
\[
= J_1 + J_2 + J_3 + J_4,
\]
where
\[
J_1 = \frac{1}{N^2T} \sum_{j=1}^N \sum_{i=1}^N Z_j (e_j^d)' S'L'e_i^d Z_i',
\]
\[
J_2 = -\frac{1}{N^2T} \sum_{j=1}^N \sum_{i=1}^N Z_j (e_j^d)' P_{x_j} (F^d\lambda_j + e_j^d)' S'L'e_i^d Z_i',
\]
\[
J_3 = -\frac{1}{N^2T} \sum_{j=1}^N \sum_{i=1}^N Z_j (F^d\lambda_j + e_j^d)' P_{x_j} S'L'e_i^d Z_i',
\]
\[
J_4 = \frac{1}{N^2T} \sum_{j=1}^N \sum_{i=1}^N Z_j (F^d\lambda_j + e_j^d)' P_{x_j} S'L'P_{x_i} (F^d\lambda_i + e_i^d) Z_i'.
\]
We start with $J_1$. By using the same steps as in the proof for $||T^{-1}U'S'L'e_i^d||$, 
\[
E \left( \left| \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i(e^d_j)'S'L'e_i^d \right| \right|^2 \right) 
\]
\[
= \frac{1}{(NT)^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{m=1}^{N} \sum_{l=1}^{N} E[Z_i(e^d_j)'S'L'e_i^d (e^d_j)'LSe_m^d tr(Z_i'Z_m'Z_i)] 
\]
\[
= \frac{1}{(NT)^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{m=1}^{N} \sum_{l=1}^{N} T^{-1} s^{-1} E[e^d_{i,k} e^d_{j,l} e^d_{m,n} tr(Z_i'Z_m'Z_i)] 
\]
\[
= \frac{1}{(NT)^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{m=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} s^{-1} \sum_{t'=1}^{T} s^{-1} e^2_{i,j} (1-a_{i,t})E[e^d_{i,k} e^d_{m,n} tr(Z_i'Z_m'Z_i)] 
\]
\[
= \frac{1}{T^2} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} (1-a_{i,t})E[tr(V_N^2)], 
\]
which is $O_p(1)$, and therefore
\[
||J_1|| = \frac{1}{N} \left| \left| \frac{1}{NT} \sum_{i=1}^{N} Z_i(e^d_j)'S'L'e_i^d \right| \right| = O_p(N^{-1}). 
\]
Similar calculations reveal that
\[
||J_2|| = - \frac{1}{\sqrt{NT}} \frac{1}{N} \sum_{j=1}^{N} \left| \left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} Z_i(e^d_j)'S'L'x_i \right| \right| |(T^{-1}x'_i x_i)_{-1}^{-1}| \times ||T^{-1/2}x'_i (F^d \lambda_i + e^d_i) || ||Z_i|| = O_p((NT)^{-1/2}), 
\]
with $||J_3||$ being of the same order. As for $J_4$, making use of Lemma A.2, we have
\[
||T^{-1}(F^d \lambda_i + e^d_i)'P \lambda_i S'L'P \lambda_i (F^d \lambda_i + e^d_i)'Z_i|| 
\]
\[
\leq T^{-1} ||T^{-1/2}(F^d \lambda_i + e^d_i)'x_i|| |((T^{-1}x'_i x_i)_{-1}^{-1}||T^{-1}x'_iS'L'x_i||\times ||(T^{-1}x'_i x_i)_{-1}^{-1}||T^{-1/2}x'_i (F^d \lambda_i + e^d_i)|| = O_p(T^{-1}), 
\]
with which we obtain
\[
||J_4|| \leq \frac{1}{N^2} \sum_{j=1}^{N} \sum_{i=1}^{N} ||Z_i|| |T^{-1}(F^d \lambda_i + e^d_i)'P \lambda_i S'L'P \lambda_i (F^d \lambda_i + e^d_i)|| ||Z_i|| = O_p(T^{-1}). 
\]
Hence, by adding the terms,
\[
\|T^{-1}U'S'L'U\| \leq \|J_1\| + \|J_2\| + \|J_3\| + \|J_4\| = O_p(N^{-1}) + O_p((NT)^{-1/2}) + O_p(T^{-1}),
\] (A22)
which establishes the required result for \(\|T^{-1}U'S'L'U\|\).

Next, consider \(T^{-2}U'S'L'S\epsilon_i^d\). By using calculations similar to those used for evaluating \(J_1\) before, we can show that
\[
E\left(\left\|\frac{1}{\sqrt{NT^2}} \sum_{j=1}^{N} Z_j(e_j^d)'S'L'S\epsilon_i^d\right\|^2\right) = O_p(1),
\]
suggesting
\[
\left\|\frac{1}{NT^2} \sum_{j=1}^{N} Z_j(e_j^d)'S'L'S\epsilon_i^d\right\| = \left\|\frac{1}{\sqrt{N}} \sum_{j=1}^{N} Z_j(e_j^d)'S'L'S\epsilon_i^d\right\| = O_p(N^{-1/2}).
\]
Moreover, by using Lemma A.2, we obtain
\[
\left\|\frac{1}{NT^2} \sum_{j=1}^{N} Z_j\lambda_j (F^d\lambda_j + e_j^d)'P_x'S'L'S\epsilon_i^d\right\|
\leq \frac{1}{\sqrt{T}} \sum_{j=1}^{N} ||Z_j||||\lambda_j||\left\|T^{-1/2}(F^d\lambda_j + e_j^d)'x_j\right\||\left\|(T^{-1}x_jx_j)^{-1}\right\||\left\|T^{-2}x_j'S'L'S\epsilon_i^d\right\|
= O_p(T^{-1/2}),
\]
which in turn implies
\[
\left\|T^{-2}U'S'L'S\epsilon_i^d\right\|
= \left\|\frac{1}{NT^2} \sum_{j=1}^{N} Z_j(e_j^d - P_x(F^d\lambda_j + e_j^d))'S'L'S\epsilon_i^d\right\|
\leq \left\|\frac{1}{NT^2} \sum_{j=1}^{N} Z_j(e_j^d)'S'L'S\epsilon_i^d\right\| + \left\|\frac{1}{NT^2} \sum_{j=1}^{N} Z_j(F^d\lambda_j + e_j^d)'P_x'S'L'S\epsilon_i^d\right\|
= O_p(N^{-1/2}) + O_p(T^{-1/2}).
\] (A23)

The proof of
\[
\left\|T^{-2}U'S'L'SF^d\right\| = O_p(N^{-1/2}) + O_p(T^{-1/2})
\] (A24)
is almost identical to the previous one. It is therefore omitted. This completes the proof of the lemma. ■
**Lemma A.4.** Under the conditions of Lemma A.2,

\[
||T^{-1}R'_{i-1}U|| = O_p(N^{-1}) + O_p(T^{-1/2}),
\]

\[
||T^{-1}R'_{i-1}F^d|| = O_p(1).
\]

**Proof of Lemma A.4.**

We begin with \(T^{-1}R'_{i-1}U\). Insertion of \(R_{i-1} = LSM_T M_x (F^d \lambda_i + \epsilon_i)\) yields

\[
R'_{i-1}U = (F^d \lambda_i + \epsilon_i^d)' M_x M_{fT} S'L'U
\]

\[
= (F^d \lambda_i + \epsilon_i^d)' M_{fT} S'L'U - (F^d \lambda_i + \epsilon_i^d)' P_{x_i} M_{fT} S'L'U
\]

\[
= (F^d \lambda_i + \epsilon_i^d)' M_{fT} S'L'U + (F^d \lambda_i + \epsilon_i^d)' (M_{fT} - M_{fT}H) S'L'U
\]

\[
- (F^d \lambda_i + \epsilon_i^d)' P_{x_i} M_{fT} S'L'U
\]

\[
= (\epsilon_i^d)' S'L'U - (\epsilon_i^d)' P_{fT} H S'L'U + (F^d \lambda_i + \epsilon_i^d)' (M_{fT} - M_{fT}H) S'L'U
\]

\[
- (F^d \lambda_i + \epsilon_i^d)' P_{x_i} M_{fT} S'L'U. \tag{A25}
\]

By Assumption 3, \(H \rightarrow_p C\) and \(||C|| = O(1)\). This, together with Lemmas A.2 and A.3, implies

\[
||T^{-1}(\epsilon_i^d)' P_{fT} H S'L'U||
\]

\[
\leq T^{-1/2}||H||^2 ||T^{-1/2}(\epsilon_i^d)'F^d|| ||(T^{-1}H'(F^d)' F^d H)^{-1}|| ||T^{-1}(F^d)' S'L'U||
\]

\[
= T^{-1/2}[O_p(N^{-1/2}) + O_p(T^{-1/2})] = O_p((NT)^{-1/2}) + O_p(T^{-1}).
\]

and

\[
||T^{-1}(F^d \lambda_i + \epsilon_i^d)' P_{x_i} M_{fT} S'L'U||
\]

\[
\leq ||T^{-1}(F^d \lambda_i + \epsilon_i^d)' P_{x_i} S'L'U||
\]

\[
= T^{-1/2}||T^{-1/2}(F^d \lambda_i + \epsilon_i^d)' x_i|| ||(T^{-1} x_i' x_i)^{-1}|| ||T^{-1} x_i' S'L'U||
\]

\[
= O_p(T^{-1/2}).
\]
Moreover, applying in turn Lemmas A.1, and then Lemmas A.2 and A.3,

\[ \|T^{-1}(F^d \lambda_i + \epsilon^d_i)'(M_{\mathcal{F} \mathcal{H}} - M_{\mathcal{F} \mathcal{H}})S'L'U\| \]
\[ \leq T^{-1/2}\|T^{-1/2}(F^d \lambda_i + \epsilon^d_i)'U\|\|T^{-1}F\hat{F}^{-1}\|\|T^{-1}U'S'L'U\| \]
\[ + T^{-1/2}\|H\|\|T^{-1/2}(F^d \lambda_i + \epsilon^d_i)'U\|\|T^{-1}\hat{F}^{-1}\|\|T^{-1}(F^d)'S'L'U\| \]
\[ + \|H\|\|T^{-1}(F^d \lambda_i + \epsilon^d_i)'F^d\|\|T^{-1}\hat{F}^{-1}\|\|T^{-1}U'S'L'U\| \]
\[ + \|H\|^2\|T^{-1}(F^d \lambda_i + \epsilon^d_i)'F^d\|\|T((\hat{F}F)^{-1} - (H'(F^d)'F^dH)^{-1})\|\|T^{-1}(F^d)'S'L'U\| \]
\[ = T^{-1/2}[O_p(N^{-1}\sqrt{T}) + O_p(N^{-1/2}) + O_p(T^{-1/2})] \]
\[ \times [O_p(N^{-1}) + O_p((NT)^{-1/2}) + O_p(T^{-1})] \]
\[ + T^{-1/2}[O_p(N^{-1}\sqrt{T}) + O_p(N^{-1/2}) + O_p(T^{-1/2})][O_p(N^{-1/2}) + O_p(T^{-1/2})] \]
\[ + [O_p(N^{-1}) + O_p((NT)^{-1/2}) + O_p(T^{-1})] \]
\[ + [O_p((NT)^{-1/2}) + O_p(T^{-1})][O_p(N^{-1/2}) + O_p(T^{-1/2})] \]
\[ = O_p(N^{-1}) + O_p((NT)^{-1/2}) + O_p(T^{-1}). \]

Therefore, since form Lemma A.3,

\[ T^{-1}(\epsilon^d_i)'S'L'U = T^{-1}(\epsilon^d_i)'S'L' \frac{1}{N} \sum_{j=1}^{N} \epsilon^d_i Z_j + O_p(T^{-1/2}) = O_p(N^{-1/2}) + O_p(T^{-1/2}), \]

we obtain

\[ \|T^{-1}R_{i,-1}^d U\| = O_p(N^{-1/2}) + O_p(T^{-1/2}), \quad \text{(A26)} \]

as was to be shown.

Consider \( T^{-1}R_{i,-1}^d F^d \). By using the same arguments leading up to (A26) we obtain

\[ \|T^{-1}R_{i,-1}^d F^d\| \]
\[ = \|T^{-1}(F^d \lambda_i + \epsilon^d_i)'M_{\mathcal{F} \mathcal{H}}S'L'F^d\| \]
\[ \leq \|T^{-1}(F^d \lambda_i + \epsilon^d_i)'M_{\mathcal{F} \mathcal{H}}S'L'F^d\| \]
\[ \leq \|T^{-1}(F^d \lambda_i + \epsilon^d_i)'M_{\mathcal{F} \mathcal{H}}S'L'F^d\| + \|T^{-1}(F^d \lambda_i + \epsilon^d_i)'(M_{\mathcal{F} \mathcal{H}} - M_{\mathcal{F} \mathcal{H}})S'L'F^d\| \]
\[ = \|T^{-1}(\epsilon^d_i)'M_{\mathcal{F} \mathcal{H}}S'L'F^d\| + \|T^{-1}(F^d \lambda_i + \epsilon^d_i)'(M_{\mathcal{F} \mathcal{H}} - M_{\mathcal{F} \mathcal{H}})S'L'F^d\|. \quad \text{(A27)} \]

The first term on the right is

\[ \|T^{-1}(\epsilon^d_i)'M_{\mathcal{F} \mathcal{H}}S'L'F^d\| \leq \|T^{-1}(\epsilon^d_i)'S'L'F^d\| = O_p(1), \]
whereas the second is
\[
\|T^{-1}(F^d \lambda_i + e^d_i)'(M_\ell - M_{\ell \ell} H)S'L'F^d\|
\leq \|T^{-1}(F^d \lambda_i + e^d_i)'U\|\|(T^{-1} \hat{F}' \hat{F})^{-1}\|\|T^{-1}U'S'L'F^d\|
+ T^{-1/2}\|H\|\|T^{-1/2}(F^d \lambda_i + e^d_i)'U\|\|(T^{-1} \hat{F}' \hat{F})^{-1}\|\|T^{-1}(F^d)'S'L'F^d\|
+ T^{-1/2}\|H\|\|T^{-1/2}(F^d \lambda_i + e^d_i)'F^d\|\|(T^{-1} \hat{F}' \hat{F})^{-1}\|\|T^{-1}U'S'L'F^d\|
+ \|H\|^2\|T^{-1}(F^d \lambda_i + e^d_i)'F^d\|\|T(\hat{F}' \hat{F})^{-1} - \left(H'(F^d)'F^d H\right)^{-1}\|\|T^{-1}(F^d)'S'L'F^d\|
= T^{-1/2}[O_p(N^{-1/2}) + O_p(N^{-1/2}) + O_p(T^{-1/2})] + O_p(T^{-1/2})
+ [O_p(N^{-1/2}) + O_p(T^{-1/2})] + [O_p((NT)^{-1/2}) + O_p(T^{-1}) + O_p(N^{-1})]
= O_p(N^{-1}) + O_p((NT)^{-1/2}) + O_p(T^{-1}).
\]

It follows that
\[
\|T^{-1}R'_{i-1}F^d\| = O_p(1), \quad (A28)
\]
and so the proof of the lemma is complete. 

\textbf{Proof of Lemma 1.}

We start by assuming that $\Sigma_e$ is known. We then show that the asymptotic results are unaffected if $\Sigma_e$ is replaced by $\hat{\Sigma}_e$.

Consider $\text{tr}(N^{-1/2}T^{-1}R'_{i-1}\Sigma_{-1}^e r)$. Making use of (A9) to substitute for $M_{x_i y_i^d}$ in $r_i = M_\ell M_{x_i} y_i^d$, since $M_\ell \hat{F} = 0$,
\[
\text{tr}(N^{-1/2}T^{-1}R'_{i-1}\Sigma_{-1}^e r) = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{\epsilon i}^{-2} R_i'_{i-1} r_i
\]
\[
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{\epsilon i}^{-2} R_i'_{i-1} M_\ell M_{x_i} y_i^d
\]
\[
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{\epsilon i}^{-2} R_i'_{i-1} M_\ell [\hat{F} H^+ \lambda_i - U H^+ \lambda_i + e^d_i - P_{x_i} (F^d \lambda_i + e^d_i)]
\]
\[
= K_1 + K_2 + K_3, \quad (A29)
\]
where

\[ K_1 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{\epsilon_i}^{-2} R_{i-1}^\prime M_T \epsilon_i^d, \]

\[ K_2 = -\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{\epsilon_i}^{-2} R_{i-1}^\prime M_T U H^+ \lambda_i, \]

\[ K_3 = -\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{\epsilon_i}^{-2} R_{i-1}^\prime M_T P_i (F^d \lambda_i + \epsilon_i^d). \]

We begin by considering \( K_2 \). Clearly,

\[ T^{-1} R_{i-1}^\prime M_T U = T^{-1} R_{i-1}^\prime M_{F^d H} U + T^{-1} R_{i-1}^\prime (M_{F^d H} - M_T) U, \]

where the second term can be evaluated in the usual way, by using Lemmas A.1–A.4;

\[ ||T^{-1} R_{i-1}^\prime (M_{F^d H} - M_T) U|| \]

\[ \leq N^{-1} ||T^{-1} R_{i-1}^\prime U|| ||(T^{-1} \hat{F}^d \hat{F})^{-1}|| ||NT^{-1} U|| \]

\[ + T^{-1/2} ||H|| ||T^{-1} R_{i-1}^\prime|| ||(T^{-1} \hat{F}^d \hat{F})^{-1}|| ||NT^{-1} U|| \]

\[ + N^{-1} ||H|| ||T^{-1} R_{i-1}^\prime F^d|| ||(T^{-1} \hat{F}^d \hat{F})^{-1}|| ||NT^{-1} U|| \]

\[ + T^{-1/2} ||H|| ||T^{-1} R_{i-1}^\prime F^d|| ||T[(\hat{F}^d \hat{F})^{-1} - (H'(F^d)'F^d H)^{-1}]|| ||T^{-1/2} (F^d)' U|| \]

\[ = N^{-1} [O_p(N^{-1/2}) + O_p(T^{-1/2})] [O_p(NT^{-1}) + O_p(1)] \]

\[ + T^{-1/2} [O_p(N^{-1/2}) + O_p(T^{-1/2})]^2 + N^{-1} [O_p(NT^{-1}) + O_p(1)] \]

\[ + T^{-1/2} [O_p((NT)^{-1/2}) + O_p(T^{-1}) + O_p(N^{-1})] [O_p(NT^{-1/2}) + O_p(T^{-1/2})] \]

\[ = O_p(N^{-1}) + O_p((NT)^{-1/2}) + O_p(T^{-1}). \]

Also, by Lemmas A.3 and A.4,

\[ T^{-1} R_{i-1}^\prime M_{F^d H} U = T^{-1} R_{i-1}^\prime U - T^{-1/2} T^{-1} R_{i-1}^\prime F^d H (T^{-1} H' (F^d)' F^d H)^{-1} T^{-1/2} H' (F^d)' U \]

\[ = T^{-1} R_{i-1}^\prime U + T^{-1/2} [O_p(N^{-1/2}) + O_p(T^{-1/2})] \]

\[ = T^{-1} R_{i-1}^\prime U + O_p((TN)^{-1/2}) + O_p(T^{-1}), \]

from which it follows that

\[ T^{-1} R_{i-1}^\prime M_T U = T^{-1} R_{i-1}^\prime U + O_p(N^{-1}) + O_p((TN)^{-1/2}) + O_p(T^{-1}). \]
\[
\left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{ij}^{-2} (R'_{i-1} M_{i} U - R'_{i-1} U) H^+ \lambda_i \right|
\leq \sqrt{N} \frac{1}{N} \sum_{i=1}^{N} \sigma_{ij}^{-2} ||T^{-1} (R'_{i-1} M_{i} U - R'_{i-1} U)||||H^+|||\lambda_i||
\leq \sqrt{N} [O_p(N^{-1}) + O_p((TN)^{-1/2}) + O_p(T^{-1})]
\leq O_p(N^{-1/2}) + O_p(T^{-1/2}) + O_p(\sqrt{NT}^{-1}),
\]

we obtain
\[
K_2 = - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{ij}^{-2} R'_{i-1} U H^+ \lambda_i - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{ij}^{-2} (R'_{i-1} M_{i} U - R'_{i-1} U) H^+ \lambda_i
= - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_{ij}^{-2} T^{-1} R'_{i-1} U H^+ \lambda_i + O_p(N^{-1/2}) + O_p(T^{-1/2}) + O_p(\sqrt{NT}^{-1}). \quad (A30)
\]

As for the remaining term, a careful inspection of the proofs of Lemmas A.3 and A.4 reveals that
\[
T^{-1} R'_{i-1} U = T^{-1} (e'_d)' S' L' U + O_p(T^{-1/2})
= T^{-1} (e'_d)' S' L' \frac{1}{N} \sum_{j=1}^{N} e'_j Z'_j + O_p(T^{-1/2}). \quad (A31)
\]

where the first term on the right is \(O_p(N^{-1/2})\). Therefore, by cross-section independence,
\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_{ij}^{-2} T^{-1} R'_{i-1} U H^+ \lambda_i
= \frac{1}{\sqrt{N}} \frac{1}{NT} \sum_{i=1}^{N} \sigma_{ij}^{-2} (e'_d)' S' L' e'_d H^+ \lambda_i + O_p(\sqrt{NT}^{-1/2})
= O_p(N^{-1/2}) + O_p(\sqrt{NT}^{-1/2}).
\]

The implication is that, since \(O_p(\sqrt{NT}^{-1/2}) > O_p(\sqrt{NT}^{-1})\),
\[
||K_2|| = O_p(N^{-1/2}) + O_p(T^{-1/2}) + O_p(\sqrt{NT}^{-1/2}). \quad (A32)
\]

As for \(K_3\), since
\[
||T^{-1} R'_{i-1} M_{i} P_{i}(F^d \lambda_i + e'_d)||
\leq ||T^{-1} R'_{i-1} P_{i}(F^d \lambda_i + e'_d)||
\leq T^{-1/2} ||T^{-1} R'_{i-1} x_i|| ||(T^{-1} x_i)_{-1}^{-1}|| ||T^{-1/2} x_i(F^d \lambda_i + e'_d)||
= O_p(T^{-1/2}),
\]

38
we obtain

$$|K_3| \leq \sqrt{N} \frac{1}{N} \sum_{i=1}^{N} \sigma_{i,j}^{-2} \| T^{-1} R'_{i-1} M F P_x (F^d \lambda_i + e_i^d) \| \| \lambda_i \| = O_p(\sqrt{NT^{-1/2}}).$$  \hspace{1cm} (A33)

$K_1$ requires more work. We begin by expanding it in the following way:

$$K_1 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{i,j}^{-2} R'_{i-1} M F M_x e_i^d$$

$$= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{i,j}^{-2} R'_{i-1} (M F \hat{H} M_x e_i^d) - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{i,j}^{-2} R'_{i-1} (M F \hat{H} - M \hat{F}) M_x e_i^d$$

$$- \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{i,j}^{-2} R'_{i-1} (M \hat{F} \hat{H} - M \hat{F}) M_x e_i^d.$$  \hspace{1cm} (A34)

The summand of the last term can be evaluated as usual, giving

$$|T^{-1} R'_{i-1} (M \hat{F} \hat{H} - M \hat{F}) M_x e_i^d|$$

$$\leq |T^{-1} R'_{i-1} (M \hat{F} \hat{H} - M \hat{F}) e_i^d|$$

$$\leq T^{-1/2} \|T^{-1} R'_{i-1} U \| \| (T^{-1} \hat{F} \hat{F})^{-1} \| \|T^{-1/2} U e_i^d\|$$

$$+ T^{-1/2} \|H \| \|T^{-1} R'_{i-1} U \| \| (T^{-1} \hat{F} \hat{F})^{-1} \| \|T^{-1/2} (F^d) e_i^d\|$$

$$+ T^{-1/2} \|H \| \|T^{-1} R'_{i-1} F^d \| \| (T^{-1} \hat{F} \hat{F})^{-1} \| \|T^{-1/2} U e_i^d\|$$

$$+ T^{-1/2} \|H \| \|T^{-1} R'_{i-1} F^d \| \| (T^{-1} \hat{F} \hat{F})^{-1} \| \|T^{-1/2} (F^d) e_i^d\|$$

$$= T^{-1/2} [O_p(N^{-1/2}) + O_p(T^{-1/2})] [O_p(N^{-1} \sqrt{T}) + O_p(N^{-1/2}) + O_p(T^{-1/2})]$$

$$+ T^{-1/2} [O_p(N^{-1/2}) + O_p(T^{-1/2})]$$

$$+ T^{-1/2} [O_p(N^{-1} \sqrt{T}) + O_p(N^{-1/2}) + O_p(T^{-1/2})]$$

$$+ T^{-1/2} [O_p((N T)^{-1/2}) + O_p(T^{-1}) + O_p(N^{-1})]$$

$$= O_p(N^{-1}) + O_p((N T)^{-1/2}) + O_p(T^{-1}),$$

suggesting that

$$\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{i,j}^{-2} R'_{i-1} (M \hat{F} \hat{H} - M \hat{F}) M_x e_i^d \right\|$$

$$\leq \sqrt{N} \frac{1}{N} \sum_{i=1}^{N} \sigma_{i,j}^{-2} \| T^{-1} R'_{i-1} (M \hat{F} \hat{H} - M \hat{F}) M_x e_i^d \|$$

$$= \sqrt{N} [O_p(N^{-1}) + O_p((N T)^{-1/2}) + O_p(T^{-1})]$$

$$= O_p(N^{-1/2}) + O_p(T^{-1/2}) + O_p(\sqrt{NT^{-1}}).$$ \hspace{1cm} (A35)
Moreover, from

\[
|T^{-1} R_{i-1} \epsilon_i^d| \leq |T^{-1} R_{i-1}^T P_s \epsilon_i^d| \\
\leq T^{-1/2} |T^{-1} R_{i-1}^T x_i| |(T^{-1} x_i^T x_i)^{-1}||T^{-1/2} x_i^T \epsilon_i^d| \\
= O_p(T^{-1/2}),
\]

we obtain

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sigma_{\epsilon_i^d}^{-2} R_{i-1}^T M_{F^2 H} P_s \epsilon_i^d \\
= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sigma_{\epsilon_i^d}^{-2} R_{i-1}^T F^d H (H'(F^d)' F^d H)^{-1} H'(F^d)' \epsilon_i^d \\
= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sigma_{\epsilon_i^d}^{-2} R_{i-1}^T \epsilon_i^d + O_p(\sqrt{NT^{-1/2}}) + O_p(N^{-1/2}), \tag{A36}
\]

It remains to consider the first term on the right-hand side of (A34). It is given by

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sigma_{\epsilon_i^d}^{-2} R_{i-1}^T F^d H (H'(F^d)' F^d H)^{-1} H'(F^d)' \epsilon_i^d \\
\leq \frac{\sqrt{N}}{\sqrt{T}} \sum_{i=1}^N \sigma_{\epsilon_i^d}^{-2} |H|^2 |T^{-1} R_{i-1}^T F^d||T^{-1} H'(F^d)' F^d H)^{-1}||T^{-1/2} (F^d)' \epsilon_i^d| \\
= \sqrt{NT^{-1/2}} [O_p(1) + O_p(N^{-1/2} \sqrt{T})] = O_p(\sqrt{NT^{-1/2}}) + O_p(N^{-1/2}).
\]

But we also have

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sigma_{\epsilon_i^d}^{-2} R_{i-1}^T \epsilon_i^d = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sigma_{\epsilon_i^d}^{-2} (F^d \lambda_i + \epsilon_i^d)' M_s (M_F - M_{F^2 H}) S' L' \epsilon_i^d \\
= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sigma_{\epsilon_i^d}^{-2} (F^d \lambda_i + \epsilon_i^d)' M_s (M_F - M_{F^2 H}) S' L' \epsilon_i^d \\
+ \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sigma_{\epsilon_i^d}^{-2} (F^d \lambda_i + \epsilon_i^d)' M_s (M_F - M_{F^2 H}) S' L' \epsilon_i^d, \tag{A38}
\]

where, by using the same steps leading up to (A28),

\[
|T^{-1}(F^d \lambda_i + \epsilon_i^d)' M_s (M_F - M_{F^2 H}) S' L' \epsilon_i^d| \leq |T^{-1}(F^d \lambda_i + \epsilon_i^d)' (M_F - M_{F^2 H}) S' L' \epsilon_i^d| \\
= O_p(N^{-1}) + O_p((NT)^{-1/2}) + O_p(T^{-1}),
\]
and therefore

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{e_i}^{-2} (F^d \lambda_i + e_i^d)' M_x, (M_F - M_{F\theta_H}) S'L' e_i^d
\]

\[
\leq \sqrt{N} \frac{1}{N} \sum_{i=1}^{N} \sigma_{e_i}^{-2} |T^{-1} (F^d \lambda_i + e_i^d)' M_x, (M_F - M_{F\theta_H}) S'L' e_i^d|
\]

\[
= \sqrt{N} |O_p(N^{-1}) + O_p((NT)^{-1/2}) + O_p(T^{-1})|
\]

\[
= O_p(N^{-1/2}) + O_p(T^{-1/2}) + O_p(\sqrt{NT}^{-1}).
\]

Moreover, since

\[
|T^{-1} (F^d \lambda_i + e_i^d)' P_{x}, M_{F\theta_H} S'L' e_i^d|
\]

\[
\leq |T^{-1} (F^d \lambda_i + e_i^d)' P_{x} S'L' e_i^d|
\]

\[
\leq T^{-1/2} ||T^{-1/2} (F^d \lambda_i + e_i^d)' x_i || ||T^{-1/2} (F^d x_i)' || ||T^{-1/2} S'L' e_i^d||
\]

\[
= O_p(T^{-1/2}),
\]

and

\[
|T^{-1} (e_i^d)' P_{F\theta_H} S'L' e_i^d|
\]

\[
\leq T^{-1/2} ||H||^2 ||T^{-1/2} (e_i^d)' F^d|| ||T^{-1} (F^d H)' (F^d)' F^d H|| ||T^{-1} (F^d)' S'L' e_i^d||
\]

\[
= O_p(T^{-1/2}),
\]

we can show that

\[
T^{-1} (F^d \lambda_i + e_i^d)' M_x, M_{F\theta_H} S'L' e_i^d
\]

\[
= T^{-1} (F^d \lambda_i + e_i^d)' M_{F\theta_H} S'L' e_i^d - T^{-1} (F^d \lambda_i + e_i^d)' P_{F\theta_H} S'L' e_i^d
\]

\[
= T^{-1} (F^d \lambda_i + e_i^d)' M_{F\theta_H} S'L' e_i^d + O_p(T^{-1/2})
\]

\[
= T^{-1} (e_i^d)' M_{F\theta_H} S'L' e_i^d + O_p(T^{-1/2})
\]

\[
= T^{-1} (e_i^d)' S'L' e_i^d + O_p(T^{-1/2}),
\]

suggesting

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_{e_i}^{-2} T^{-1} (F^d \lambda_i + e_i^d)' M_x, M_{F\theta_H} S'L' e_i^d
\]

\[
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_{e_i}^{-2} T^{-1} (e_i^d)' S'L' e_i^d + O_p(\sqrt{NT}^{-1/2}).
\]
Direct substitution into (A38) now yields
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{\varepsilon_i}^{-2} R_i^{\prime} R_i^{-1} e_i^d
\]
\[
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_{\varepsilon_i}^{-2} T^{-1} (T^d \lambda_i + e_i^d)^\prime M \sigma_i (M_F - M_P) S' L' e_i^d
\]
\[
+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_{\varepsilon_i}^{-2} T^{-1} (T^d \lambda_i + e_i^d)^\prime M \sigma_i (M_F - M_P) S' L' e_i^d
\]
\[
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_{\varepsilon_i}^{-2} T^{-1} (e_i^d)^\prime S' L' e_i^d + O_p(N^{-1/2}) + O_p(\sqrt{NT^{-1/2}}),
\]
where the order of the remainder follows from using $O_p(T^{-1/2}) < O_p(\sqrt{NT^{-1/2}})$ and $O_p(\sqrt{NT^{-1/2}}) > O_p(\sqrt{NT^{-1}})$.

Let $P_{i,t} = T^{-1} \sigma_{\varepsilon_i}^{-2} (e_i^d)^\prime S' L' e_i^d$, such that
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{\varepsilon_i}^{-2} (e_i^d)^\prime S' L' e_i^d = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} P_{i,t}.
\]
We now use the same steps as in Phillips and Moon (1999, pages 1101–1103) to verify that $P_{i,t}$ satisfies conditions (i)–(iv) of their central limit theorem, Theorem 3. We begin by considering the mean of $P_{i,t}$. Since $E[e_i^d(e_i^d)^\prime] = 0$ for all $t > s = h, ..., T$,
\[
\sum_{k=2}^{t-1} \sum_{j=h}^{t-1} E(e_k e_j^\prime) a_{k,l} = \sum_{k=h}^{t-1} a_{k,t} \Sigma_e,
\]
\[
\sum_{k=2}^{t-1} \sum_{s=h}^{t-1} \sum_{n=2}^{s} E(e_n e_k^\prime) a_{k,l} a_{s,t-1} = \sum_{k=h}^{t-1} \sum_{s=h}^{t-1} \sum_{n=2}^{s} E(e_n e_k^\prime) a_{k,t} a_{n,s} = \sum_{s=h}^{t-1} \sum_{a_{n,t} a_{n,s} = a_{s,t}}.
\]
By using this, $(e_i^d)^\prime S' L' e_i^d = \sum_{k=h}^{T} \sum_{s=h}^{T} \sum_{n=2}^{s} e_i^d e_i^d$ and $\sum_{n=2}^{s} a_{n,t} a_{n,s} = a_{s,t}$, we can show that
\[
E(P_{i,t}) = T^{-1} \sigma_{\varepsilon_i}^{-2} E[(e_i^d)^\prime S' L' e_i^d] = \frac{1}{T} \sum_{t=1}^{T-1} \sum_{l=k}^{h} \sigma_{\varepsilon_i}^{-2} E(e_i^d e_i^d)
\]
\[
= \frac{1}{T} \sum_{l=s}^{t-1} \sum_{h}^{s} a_{n,t} a_{n,s} - a_{s,t}) = 0.
\]
The variance can be evaluated in the following fashion:
\[
E(P_{i,t}^2) = E[(T^{-1} \sigma_{\varepsilon_i}^{-2} (e_i^d)^\prime S' L' e_i^d)^2]
\]
\[
= \frac{1}{T^2} \sum_{t=1}^{T-1} \sum_{l=k}^{h} \sum_{n=2}^{s} \sigma_{\varepsilon_i}^{-4} E(e_i^d e_i^d E((e_i^d)^2) | T_{t-1})]
\]
\[
+ \frac{2}{T^2} \sum_{t=1}^{T-1} \sum_{l=k}^{h} \sum_{n=2}^{s} \sigma_{\varepsilon_i}^{-4} E(e_i^d e_i^d E(e_i^d) | T_{t-1})]
\]
\[
= \frac{1}{T^2} \sum_{t=1}^{T-1} \sum_{l=k}^{h} \sum_{n=2}^{s} (1 - a_{n,t} a_{n,s} = a_{s,t}) \sigma_{\varepsilon_i}^{-2} E(e_i^d e_i^d).
\]
where the first equality holds because of symmetry, while the second uses \(E[(\epsilon_{t,j}^d)^2] = (1 - a_{t,j})\sigma^2_{\epsilon j} \). Consider \(\sum_{k=h}^t \sum_{j=h}^l E[\epsilon_t^d (\epsilon_j^d)']\), which for \(t \geq l = h, \ldots, T\) can be written as

\[
\sum_{k=h}^t \sum_{j=h}^l E[\epsilon_t^d (\epsilon_j^d)'] = \sum_{k=h}^t \sum_{j=h}^l E(\epsilon_t^d) E(\epsilon_j^d)' - \sum_{k=h}^t \sum_{j=h}^l E(\epsilon_t^d \epsilon_j^d) a_{j,k} - \sum_{k=h}^t \sum_{j=h}^l E(\epsilon_t^d) E(\epsilon_j^d) a_{j,k}
\]

\[+ \sum_{k=h}^t \sum_{j=h}^l \sum_{s=2}^k \sum_{l=1}^s E(\epsilon_t^d \epsilon_j^d) a_{l,k} a_{n,s},\]

where, since \(\epsilon_t\) is serially uncorrelated with \(E(\epsilon_t^d) = \Sigma_{\epsilon}\) (Assumption 2),

\[
\sum_{k=h}^t \sum_{j=h}^l E(\epsilon_t^d) = [l - (p + 1)] \Sigma_{\epsilon},
\]

\[
\sum_{k=h}^t \sum_{j=h}^l E(\epsilon_t^d) a_{j,k} = \sum_{k=h}^t \sum_{j=h}^l a_{j,k} \Sigma_{\epsilon},
\]

\[
\sum_{k=h}^t \sum_{j=h}^l \sum_{s=2}^k \sum_{l=1}^s E(\epsilon_t^d) a_{s,k} = \sum_{k=h}^t \sum_{j=h}^l \sum_{s=2}^k \sum_{l=1}^s \sum_{j=2}^k E(\epsilon_t^d) a_{s,k} + \sum_{k=h}^t \sum_{j=h}^l \sum_{s=2}^k \sum_{l=1}^s E(\epsilon_t^d) a_{s,k}
\]

\[= \left( \sum_{k=h}^t \sum_{s=2}^k a_{s,k} + \sum_{k=l+1}^t \sum_{s=2}^k a_{s,k} \right) \Sigma_{\epsilon}.
\]

Also,

\[
\sum_{k=h}^t \sum_{s=h}^l \sum_{m=2}^k \sum_{n=2}^s E(\epsilon_t^d) a_{m,k} a_{n,s} = \sum_{k=h}^t \sum_{s=h}^l \sum_{m=2}^k \sum_{n=2}^s E(\epsilon_t^d) a_{m,k} a_{n,s}
\]

\[+ \sum_{k=l+1}^t \sum_{s=h}^l \sum_{m=2}^k \sum_{n=2}^s E(\epsilon_t^d) a_{m,k} a_{n,s},\]

where

\[
\sum_{k=h}^t \sum_{s=h}^l \sum_{m=2}^k \sum_{n=2}^s E(\epsilon_t^d) a_{m,k} a_{n,s}
\]

\[= \sum_{k=h}^t \sum_{s=h}^l \sum_{m=2}^k \sum_{n=2}^s E(\epsilon_t^d) a_{m,k} a_{n,s} + \sum_{k=h}^t \sum_{s=h}^l \sum_{m=2}^k \sum_{n=2}^s E(\epsilon_t^d) a_{m,k} a_{n,s}
\]

\[= \sum_{k=h}^t \left( \sum_{s=h}^l a_{n,k} a_{n,s} + \sum_{s=k+1}^l a_{n,k} a_{n,s} \right) \Sigma_{\epsilon}
\]

\[= \sum_{k=h}^t \left( \sum_{s=h}^l a_{s,k} + \sum_{s=k+1}^l a_{k,s} \right) \Sigma_{\epsilon},
\]

and

\[
\sum_{k=l+1}^t \sum_{s=h}^l \sum_{m=2}^k \sum_{n=2}^s E(\epsilon_t^d) a_{m,k} a_{n,s} = \sum_{k=l+1}^t \sum_{s=h}^l \sum_{m=2}^k \sum_{n=2}^s a_{n,k} a_{n,s} \Sigma_{\epsilon} = \sum_{k=l+1}^t \sum_{s=h}^l a_{s,k} \Sigma_{\epsilon},
\]

43
suggesting that
\[
\sum_{k=h}^{t} \sum_{j=h}^{l} E[e_{i,k}^d (e_{j,t}^d)'] = [l - (p + 1)] \Sigma_{e} - \sum_{k=h}^{t} \sum_{j=h}^{l} a_{j,k} \Sigma_{e} - \left( \sum_{k=h}^{t} \sum_{s=h}^{k} a_{s,k} + \sum_{k=l+1}^{t} \sum_{s=k+1}^{l} a_{s,k} \right) \Sigma_{e}
\]
\[
+ \sum_{k=h}^{t} \left( \sum_{s=h}^{k} a_{s,k} + \sum_{s=k+1}^{l} a_{s,k} \right) \Sigma_{e} + \sum_{k=l+1}^{t} \sum_{s=k+1}^{l} a_{s,k} \Sigma_{e}
\]
\[
= \left( [l - (p + 1)] - \sum_{k=h}^{t} \sum_{s=h}^{k} a_{s,k} + \sum_{k=h}^{t} \sum_{s=k+1}^{l} a_{s,k} \right) \Sigma_{e} = b_{j} \Sigma_{e},
\]

with an implicit definition of \( b_{j} \). It follows that
\[
E(P_{i,T}^2) = \frac{1}{T^2} \sum_{t=h}^{T} \sum_{k=h}^{t-1} \sum_{l=0}^{t-1} (1 - a_{t,l}) \sigma_{i,j} E(e_{i,k}^d e_{j,t}^d) = \frac{1}{T^2} \sum_{t=h}^{T} (1 - a_{t,l}) b_{t-1} = \Sigma_{T},
\]
whose limit as \( T \to \infty \) can be deduced by noting that by the Cauchy–Schwarz inequality, \( |T^{-2} \sum_{t=h}^{T} a_{t,l} b_{t-1}| = O(T^{-1}) \). Also,
\[
T^{-1} b_{[vT]} \to v - \int_{w=0}^{v} \int_{u=0}^{w} a(u, w) du dw + \int_{w=0}^{v} \int_{w=0}^{v} a(w, u) du dw = b(v),
\]
from which it follows that
\[
\Sigma_{T} = \frac{1}{T^2} \sum_{t=h}^{T} (1 - a_{t,l}) b_{t-1} = \frac{1}{T^2} \sum_{t=h}^{T} b_{t-1} + O(T^{-1}) \to \int_{v=0}^{1} b(v) dv = \Sigma
\] as \( T \to \infty \). Hence, \( P_{i,T} \) is not only iid across \( i \), but also mean zero and with variance \( \Sigma_{T} \to \Sigma \). Since the scaling of \( P_{i,T} \) is just one, this implies that conditions (i), (ii) and (iv) of Theorem 3 of Phillips and Moon (1999) are satisfied.\(^8\) Verifying condition (iii) means showing that \( P_{i,T}^2 \) is uniformly integrable (over \( T \)), which holds if \( P_{i,T}^2 \to P_{i}^2 \) and \( E(P_{i,T}^2) \to E(P_{i}^2) \) (see Moon and Phillips, 2000, page 971). Note first that
\[
\frac{1}{\sqrt{T}} \sum_{s=1}^{t} e_{i,s} = \frac{1}{\sqrt{T}} \sum_{s=1}^{t} \left( e_{i,s} - \sum_{k=1}^{s} e_{i,k} a_{k,t} \right) = \frac{1}{\sqrt{T}} \sum_{s=1}^{t} \left( 1 - (t + 1 - s) a_{s,t} \right) e_{i,s},
\]
where the last equality holds because \( \sum_{s=1}^{t} \sum_{k=1}^{s} e_{i,k} a_{k,t} = \sum_{s=1}^{t+1} (t + 1 - s) e_{i,s} a_{s,t} = \sum_{s=1}^{t} (t + 1 - s) e_{i,s} a_{s,t} \). Here, letting \( t = [vT] \) and \( v \in [0, 1],
\[
\frac{1}{\sqrt{T}} \sum_{s=1}^{t} \left[ 1 - T^{-1} (t + 1 - s) a_{s,t} \right] e_{i,s} \to_{d} \sigma_{e,i} \int_{u=0}^{v} \left[ 1 - (v - u) a(u, v) \right] dW_{e,i}(u) = \sigma_{e,i} B_{e,i}(v)
\]
\(^8\)In the notation of Phillips and Moon (1999), we have \( Y_{i,T} = C_{i} Q_{i,T} \), where \( C_{i} = 1 \) and \( Q_{i,T} = P_{i,T} \).
as $T \to \infty$, where $W_{\epsilon, i}(v)$ is a standard scalar Brownian motion and $B_{\epsilon, i}(v)$ is implicitly defined. Application of the continuous mapping theorem now yields

$$P_{i,T} = \frac{1}{T} \sum_{t=1}^{T} \sum_{k=1}^{\infty} \sigma_{\epsilon, i}^2 e_{i,k} e_{i,k}^d \to_d \int_{v=0}^{1} B_{\epsilon, i}(v) dB_{\epsilon, i}(v) = P_i$$

as $T \to \infty$, and so $P_{i,T}^2 \to_d P_i^2$ (via continuous mapping). It is not difficult to show that $E(P_i^2) \to \Sigma$, and therefore $E(P_{i,T}^2) \to E(P_i^2)$, verifying that $P_{i,T}^2$ is uniformly integrable. The conditions of Theorem 3 of Phillips and Moon (1999) are therefore satisfied. It follows that

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{\epsilon, i}^{-2} (e_{i,i}^d)' S' L' e_{i,i}^d = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{k=1}^{\infty} \sigma_{\epsilon, i}^{-2} e_{i,k} e_{i,k}^d$$

$$\to_d \frac{1}{\sqrt{N}} \sum_{i=1}^{N} P_{i,T} N(0, \Sigma) \quad \text{(A40)}$$

as $N, T \to \infty$. Hence, provided that $N/T = o(1)$,

$$K_1 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{\epsilon, i}^{-2} R_{i,-1}' M_{F} M_{\xi} e_{i,i}^d$$

$$= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{\epsilon, i}^{-2} R_{i,-1}' M_{F/H} e_{i,i}^d - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{\epsilon, i}^{-2} R_{i,-1}' M_{F/H} P_{\xi} e_{i,i}^d$$

$$- \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{\epsilon, i}^{-2} R_{i,-1}' (M_{F/H} - M_{F}) M_{\xi} e_{i,i}^d$$

$$= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{\epsilon, i}^{-2} (e_{i,i}^d)' S' L' e_{i,i}^d + O_p(N^{-1/2}) + O_p(T^{-1/2}) + O_p(\sqrt{NT}^{-1/2})$$

$$\to_d N(0, \Sigma). \quad \text{(A41)}$$

Thus, by combining (A32), (A33) and (A41), (A29) becomes

$$\text{tr}(N^{-1/2} T^{-1} R_{-1}' \Sigma_{\epsilon}^{-1} r)$$

$$= K_1 + K_2 + K_3$$

$$= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{\epsilon, i}^{-2} (e_{i,i}^d)' S' L' e_{i,i}^d + O_p(N^{-1/2}) + O_p(T^{-1/2}) + O_p(\sqrt{NT}^{-1/2})$$

$$\to_d N(0, \Sigma). \quad \text{(A42)}$$

as $N, T \to \infty$ with $N/T \to 0$.

Let us now consider $\text{tr}(N^{-1} T^{-2} R_{-1}' \Sigma_{\epsilon}^{-1} R_{-1})$. Clearly,

$$\text{tr}(N^{-1} T^{-2} R_{-1}' \Sigma_{\epsilon}^{-1} R_{-1})$$

$$= \frac{1}{NT^2} \sum_{i=1}^{N} \sigma_{\epsilon, i}^{-2} R_{i,-1}' R_{i,-1}$$

$$= \frac{1}{NT^2} \sum_{i=1}^{N} \sigma_{\epsilon, i}^{-2} (F_{\xi} \lambda_{i} + e_{i,i}^d)' M_{F} S' L S M_{F} M_{\xi} (F_{\xi} \lambda_{i} + e_{i,i}^d), \quad \text{(A43)}$$

45
where

\[
(F^d \lambda_i + \epsilon_i^d)' M_x, M_{\mathcal{F}} S'L' L S M_{\mathcal{F}} M_x (F^d \lambda_i + \epsilon_i^d) \\
= (F^d \lambda_i + \epsilon_i^d)' M_x, M_{\mathcal{F}} S'L' L S M_{\mathcal{F}} M_x (F^d \lambda_i + \epsilon_i^d) \\
+ (F^d \lambda_i + \epsilon_i^d)' M_x, (M_{\mathcal{F}} - M_{\mathcal{F}}} \hat{\mathcal{H}}) S'L' L S M_{\mathcal{F}} M_x (F^d \lambda_i + \epsilon_i^d) \\
+ (F^d \lambda_i + \epsilon_i^d)' M_x, M_{\mathcal{F}} S'L' L S (M_{\mathcal{F}} - M_{\mathcal{F}}} \hat{\mathcal{H}}) M_x (F^d \lambda_i + \epsilon_i^d) \\
+ (F^d \lambda_i + \epsilon_i^d)' M_x, (M_{\mathcal{F}} - M_{\mathcal{F}}} \hat{\mathcal{H}}) S'L' L S (M_{\mathcal{F}} - M_{\mathcal{F}}} \hat{\mathcal{H}}) M_x (F^d \lambda_i + \epsilon_i^d) \\
= L_1 + L_2 + L_3, \quad (A44)
\]

with

\[
L_1 = (\epsilon_i^d)' M_x, M_{\mathcal{F}} S'L' L S M_{\mathcal{F}} M_x \epsilon_i^d, \\
L_2 = 2(F^d \lambda_i + \epsilon_i^d)' M_x, (M_{\mathcal{F}} - M_{\mathcal{F}}} \hat{\mathcal{H}}) S'L' L S M_{\mathcal{F}} M_x \epsilon_i^d, \\
L_3 = (F^d \lambda_i + \epsilon_i^d)' M_x, (M_{\mathcal{F}} - M_{\mathcal{F}}} \hat{\mathcal{H}}) S'L' L S (M_{\mathcal{F}} - M_{\mathcal{F}}} \hat{\mathcal{H}}) M_x (F^d \lambda_i + \epsilon_i^d).
\]

Note how the dependence of $L_1$, $L_2$ and $L_3$ on $i$ has been suppressed for simplicity. We start with $L_2$. By Lemmas A.1–A.4,

\[
|L_2| = 2|T^{-2}(F^d \lambda_i + \epsilon_i^d)' M_x, (M_{\mathcal{F}} - M_{\mathcal{F}}} \hat{\mathcal{H}}) S'L' L S M_{\mathcal{F}} M_x \epsilon_i^d|
\leq 2|T^{-2}(F^d \lambda_i + \epsilon_i^d)' (M_{\mathcal{F}} - M_{\mathcal{F}}} \hat{\mathcal{H}}) S'L' L S \epsilon_i^d|
\leq 2T^{-1/2}||T^{-1/2}(F^d \lambda_i + \epsilon_i^d)'U|||||T^{-1}(\hat{\mathcal{F}}\hat{\mathcal{F}})^{-1}||||T^{-2}U'S'L'S \epsilon_i^d||
\leq 2T^{-1/2}||H||||T^{-1/2}(F^d \lambda_i + \epsilon_i^d)'U|||||T^{-1}(\hat{\mathcal{F}}\hat{\mathcal{F}})^{-1}||||T^{-2}F'S'L'S \epsilon_i^d||
\leq 2||H||T^{-1}(F^d \lambda_i + \epsilon_i^d)'F|||||T^{-1}(\hat{\mathcal{F}}\hat{\mathcal{F}})^{-1}||||T^{-2}U'S'L'S \epsilon_i^d||
\leq 2||H||^2||T^{-1}(F^d \lambda_i + \epsilon_i^d)'F||||T[(\hat{\mathcal{F}}\hat{\mathcal{F}})^{-1} - (H'(\hat{\mathcal{F}}'H)^{-1})]||||T^{-2}F'S'L'S \epsilon_i^d||
= T^{-1/2}[O_p(N^{-1/2}) + O_p(N^{-1/2}) + O_p(T^{-1/2})][O_p(N^{-1/2}) + O_p(T^{-1/2})]
+ T^{-1/2}[O_p(N^{-1/2}) + O_p(N^{-1/2}) + O_p(T^{-1/2})] + [O_p(N^{-1/2}) + O_p(T^{-1/2})]
\leq [O_p((NT)^{-1/2}) + O_p(T^{-1}) + O_p(N^{-1})]
= O_p(N^{-1/2}) + O_p(T^{-1/2}). \quad (A45)
\]

It is easy to show that $|L_3|$ is of even lower order of magnitude than this. The last two terms

46
in (A44) are therefore $O_p(1)$. It remains to consider the first term, $L_1$, which we expand as

$$L_1 = \left(e_i^d\right)'M_iM_{F^dH}S'L'LSM_{F^dH}M_i\epsilon_i^d$$

$$= T^{-2}\left(e_i^d\right)'M_iS'L'LSM_i\epsilon_i^d - 2T^{-2}\left(e_i^d\right)'M_iP_{F^dH}S'L'LSM_{F^dH}M_i\epsilon_i^d$$

$$+ T^{-2}\left(e_i^d\right)'M_iP_{F^dH}S'L'LSM_{F^dH}M_i\epsilon_i^d$$

$$= T^{-2}\left(e_i^d\right)'S'L'Se_i^d - 2T^{-2}\left(e_i^d\right)'P_{F^dH}S'L'LSM_i\epsilon_i^d + T^{-2}\left(e_i^d\right)'P_{F^dH}S'L'LSM_i\epsilon_i^d$$

$$- 2T^{-2}\left(e_i^d\right)'M_iP_{F^dH}S'L'LSM_{F^dH}M_i\epsilon_i^d$$

$$+ T^{-2}\left(e_i^d\right)'M_iP_{F^dH}S'L'LSM_{F^dH}M_i\epsilon_i^d,$$  \hspace{1cm} (A46)

where the second and third terms on the right are (by Lemma A.2)

$$|T^{-2}\left(e_i^d\right)'P_{F^dH}S'L'LSM_i\epsilon_i^d|$$

$$\leq |T^{-2}\left(e_i^d\right)'P_{F^dH}S'L'Se_i^d|$$

$$\leq T^{-1/2}\|T^{-1/2}\left(e_i^d\right)'x_i\|\|(T^{-1}x_i')^{-1}\|\|T^{-2}x_i'S'L'Se_i^d\|$$

$$= O_p(T^{-1/2}),$$

and

$$|T^{-2}\left(e_i^d\right)'P_{F^dH}S'L'LSM_i\epsilon_i^d|$$

$$\leq T^{-1}\|T^{-1/2}\left(e_i^d\right)'x_i\|^2\|(T^{-1}x_i')^{-1}\|^2\|T^{-2}x_i'S'L'Sx_i\|$$

$$= O_p(T^{-1}).$$

The third and fourth terms are of the same orders of magnitude as the second and third, respectively. Hence,

$$L_1 = T^{-2}\left(e_i^d\right)'S'L'Se_i^d + O_p(T^{-1/2}),$$  \hspace{1cm} (A47)

giving

$$(F^d\lambda_i + e_i^d)'M_iM_{F^dH}S'L'LSM_{F^dH}M_i(F^d\lambda_i + e_i^d)$$

$$= L_1 + L_2 + L_3$$

$$= T^{-2}\left(e_i^d\right)'S'L'Se_i^d + O_p(N^{-1/2}) + O_p(T^{-1/2}),$$  \hspace{1cm} (A48)
which in turn implies that (A43) simplifies to
\[
\text{tr}(N^{-1}T^{-2}R'_{-1}\Sigma^{-1}_e R_{-1})
\]
\[
= \frac{1}{N} \sum_{i=1}^{N} \sigma_{ei}^{-2}(F^d \lambda_i + \epsilon_i^d)'M_{xi}M_{x}S'LSM_{x}M_{x}^'(F^d \lambda_i + \epsilon_i^d)
\]
\[
= \frac{1}{N} \sum_{i=1}^{N} \sigma_{ei}^{-2}T^{-2}(\epsilon_i^d)'S'L'S\epsilon_i^d + O_p(N^{-1/2}) + O_p(T^{-1/2}).
\] (A49)

In order to obtain the probability limit of the remaining term we apply the law of large numbers given in Corollary 1 of Phillips and Moon (1999). In their notation, we have $Y_{i,T} = C_i Q_{i,T}$, where $C_i = 1$ and $Q_{i,T} = \sigma_{ei}^{-2}T^{-2}(\epsilon_i^d)'S'L'S\epsilon_i^d$. Since $Q_{i,T}$ is iid, the conditions for Corollary 1 are that (i) $|Q_{i,T}|$ is uniformly integrable over $T$ for all $i$ and (ii) $\sup_{i=1,\ldots,N} |C_i| \leq \infty$, where the latter is obviously satisfied in view of the fact that $C_i = 1$. As for (ii), making use of some of our previous results,

\[
Q_{i,T} = \sigma_{ei}^{-2}T^{-2}(\epsilon_i^d)'S'L'S\epsilon_i^d = \frac{1}{T^2} \sum_{t=h}^{T} \sum_{t=h}^{T} \sum_{t=h}^{T} \sigma_{ei}^{-2} \epsilon_{i,k} \epsilon_{i,n} \rightarrow_d \int_{v=0}^{1} B_{ei,v}^2 dv = Q_i
\]
as $T \rightarrow \infty$. Moreover,

\[
E(Q_i) = \frac{1}{T^2} \sum_{t=h}^{T} \sum_{t=h}^{T} \sigma_{ei}^{-2} E(\epsilon_{i,k} \epsilon_{i,n}^d) = \frac{1}{T^2} \sum_{i=1}^{T} \sum_{k=1}^{T} b_{i,-1} \rightarrow \int_{v=0}^{1} b(v)dv = \Sigma,
\]

and it is easy to show that $E(Q_i) = \Sigma$. Therefore, since $Q_{i,T} \rightarrow_d Q_i$ and $E(Q_{i,T}) \rightarrow E(Q_i)$, we have that $|Q_{i,T}|$ is uniformly integrable (see Moon and Phillips, 2000, page 971). It follows that

\[
\frac{1}{N} \sum_{i=1}^{N} Q_{i,T} = \frac{1}{N} \sum_{i=1}^{N} \sigma_{ei}^{-2}T^{-2}(\epsilon_i^d)'S'L'S\epsilon_i^d \rightarrow_p \Sigma
\] (A50)
as $N, T \rightarrow \infty$, and so we obtain

\[
\text{tr}(N^{-1}T^{-2}R'_{-1}\Sigma^{-1}_e R_{-1}) = \frac{1}{N} \sum_{i=1}^{N} \sigma_{ei}^{-2}T^{-2}(\epsilon_i^d)'S'L'S\epsilon_i^d + O_p(N^{-1/2}) + O_p(T^{-1/2})
\]
\[
\rightarrow_p \Sigma.
\] (A51)

It remains to consider $\hat{\Sigma}_e$. The $i^{th}$ element along the main diagonal is given by

\[
\hat{\sigma}_{ei}^2 = T^{-1}r_i'\hat{r}_i
\]
\[
= T^{-1}(M_{x}M_{x}y_i^d)'M_{x}y_i^d
\]
\[
= T^{-1}(F^d \lambda_i + \nu_i)'M_{x}(F^d \lambda_i + \nu_i)
\]
\[
= M_1 + M_2,
\] (A52)
where (again ignoring the dependence on $i$)

\[ M_1 = T^{-1}(F^d \lambda_i + v_i) M_{F^{dH}} (F^d \lambda_i + v_i), \]
\[ M_2 = T^{-1}(F^d \lambda_i + v_i) (M_{F^{dH}} - M_{F^{dH}}) (F^d \lambda_i + v_i). \]

Consider $M_1$. By the definition of $v_i$,

\[ M_1 = T^{-1}(F^d \lambda_i + v_i) M_{F^{dH}} (F^d \lambda_i + v_i) = T^{-1} v_i^j M_{F^{dH}} v_i \]
\[ = T^{-1} [e_i^d - P_{x_i} (F^d \lambda_i + e_i^d)]' M_{F^{dH}} [e_i^d - P_{x_i} (F^d \lambda_i + e_i^d)] \]
\[ = T^{-1} (e_i^d)' M_{F^{dH}} e_i^d - 2T^{-1} (e_i^d)' M_{F^{dH}} P_x (F^d \lambda_i + e_i^d) \]
\[ + T^{-1} \lambda_i' (F^d \lambda_i + e_i^d)' P_x M_{F^{dH}} P_x (F^d \lambda_i + e_i^d) \lambda_i, \]

where the order of the second term on the right is given by

\[ |T^{-1} (e_i^d)' M_{F^{dH}} P_x (F^d \lambda_i + e_i^d)| \]
\[ \leq |T^{-1} (e_i^d)' P_x (F^d \lambda_i + e_i^d)| \]
\[ \leq T^{-1} ||T^{-1/2} (e_i^d)' x_i|| (T^{-1} x_i x_i)^{1/2} ||T^{-1/2} x_i' (F^d \lambda_i + e_i^d)|| \]
\[ = O_p(T^{-1}). \]

The order of the third term is lower than this. It remains to consider the first term, which we can write as

\[ T^{-1} (e_i^d)' M_{F^{dH}} e_i^d = T^{-1} (e_i^d)' e_i^d - T^{-1} T^{-1/2} (e_i^d)' F^d H (T^{-1} H' (F^d)' F^d H)^{-1} T^{-1/2} H' (F^d)' e_i^d \]
\[ = T^{-1} (e_i^d)' e_i^d + O_p(T^{-1}) \]
\[ = \sigma_{e,i}^2 T^{-1} \text{tr}(I_{T-h+1}) + T^{-1/2} T^{-1/2} \text{tr} [e_i^d (e_i^d)'] - \sigma_{e,i}^2 T^{-1} h_{I_{T-h+1}} + O_p(T^{-1}) \]
\[ = \sigma_{e,i}^2 + O_p(T^{-1/2}), \]

where the third equality follows from writing $(e_i^d)' e_i^d = \text{tr} [e_i^d (e_i^d)']$ to which we then add and subtract $\sigma_{e,i}^2 I_{T-h+1}$. As for the fourth and last equality, making use of the fact that

\[ E[e_i^d (e_i^d)'] = \sigma_{e,i}^2 (I_{T-h+1} - A_{T-h+1}), \text{tr}(A_{T-h+1}) = \sum_{t=h}^T a_{t,t} = T^{-1} \sum_{t=h}^T T a_{t,t} = O(1) \]

and

\[ E[(e_i^d)^4] = E(e_i^4) + o(1) = O(1) \]

(details are available upon request), it is possible to show
that, by a central limit theorem,
\[
T^{-1/2} \text{tr}[\epsilon_i^d (\epsilon_i^d)' - \sigma_{\epsilon,i}^2 I_{T-h+1}] = T^{-1/2} \text{tr}[\epsilon_i^d (\epsilon_i^d)' - \sigma_{\epsilon,i}^2 (I_{T-h+1} - A_{T-h+1})] - \sigma_{\epsilon,i}^2 T^{-1/2} \text{tr}(A_{T-h+1})
\]
\[
= T^{-1/2} \text{tr}[\epsilon_i^d (\epsilon_i^d)' - E(\epsilon_i^d (\epsilon_i^d)')] - \sigma_{\epsilon,i}^2 T^{-1/2} \text{tr}(A_{T-h+1})
\]
\[
= O_p(1) + O_p(T^{-1/2}).
\]

The expression for \( M_1 \) therefore simplifies to
\[
M_1 = T^{-1}(F^d \lambda_i + v_i)'(M_p - M_{FH})(F^d \lambda_i + v_i) = \sigma_{\epsilon,i}^2 + O_p(T^{-1/2}). \tag{A53}
\]

As for \( M_2 \), by the definition of \( v_i \),
\[
F^d \lambda_i + v_i = F^d \lambda_i + \epsilon_i^d - P_h(F^d \lambda_i + \epsilon_i^d) = M_h(F^d \lambda_i + \epsilon_i^d),
\]
which we can use to show that
\[
|M_2| = |T^{-1}(F^d \lambda_i + v_i)'(M_p - M_{FH})(F^d \lambda_i + v_i)|
\]
\[
= |T^{-1}(F^d \lambda_i + \epsilon_i^d)'M_h(M_p - M_{FH})M_h(F^d \lambda_i + \epsilon_i^d)|
\]
\[
\leq |T^{-1}(F^d \lambda_i + \epsilon_i^d)'(M_p - M_{FH})(F^d \lambda_i + \epsilon_i^d)|
\]
\[
\leq T^{-1}\|| T^{-1/2}(F^d \lambda_i + \epsilon_i^d)'U||^2||| (T^{-1}(F)'F)_{-1}||
\]
\[
+ 2T^{-1/2}||H|||T^{-1/2}(F^d \lambda_i + \epsilon_i^d)'U|||(T^{-1}(F)'F)_{-1}||| |T^{-1}(F)^d)'(F^d \lambda_i + \epsilon_i^d)||
\]
\[
+ ||H||^2||T^{-1}(F^d \lambda_i + \epsilon_i^d)'F||^2|||((F)'F)_{-1} - (H'(F)'F^d)_{-1}||
\]
\[
= T^{-1}[O_p(N^{-1}\sqrt{T}) + O_p(N^{-1/2}) + O_p(T^{-1/2})]^2
\]
\[
+ T^{-1/2}[O_p(N^{-1}\sqrt{T}) + O_p(N^{-1/2}) + O_p(T^{-1/2})]
\]
\[
+ [O_p((NT)^{-1/2}) + O_p(T^{-1}) + O_p(N^{-1})]
\]
\[
= O_p((NT)^{-1/2}) + O_p(T^{-1}) + O_p(N^{-1}), \tag{A54}
\]

which in turn implies
\[
\hat{\sigma}_{\epsilon,i}^2 = M_1 + M_2 = \sigma_{\epsilon,i}^2 + O_p((NT)^{-1/2}) + O_p(T^{-1/2}) + O_p(N^{-1}). \tag{A55}
\]

Taylor expansion now yields
\[
\text{tr}(N^{-1/2}T^{-1}R_{-1}' \Sigma_{\epsilon}^{-1} \Sigma_{\epsilon}^{-1} r) = \text{tr}(N^{-1/2}T^{-1}R_{-1}' \Sigma_{\epsilon}^{-1} \Sigma_{\epsilon}^{-1} r) + O_p(T^{-1/2}) + O_p(\sqrt{NT^{-1/2}})
\]
\[
+ O_p(N^{-1/2}), \tag{A56}
\]
\[
(\text{tr}(N^{-1/2}T^{-2}R_{-1}' \Sigma_{\epsilon}^{-1} R_{-1}))^{-1/2} = (\text{tr}(N^{-1/2}T^{-2}R_{-1}' \Sigma_{\epsilon}^{-1} R_{-1}))^{-1/2} + O_p((NT)^{-1/2})
\]
\[
+ O_p(T^{-1/2}) + O_p(N^{-1}). \tag{A57}
\]

This completes the proof of the lemma. \( \square \)
Table 1: Simulated size and power at the 5% level.

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>( \rho = 1 ) (size)</th>
<th>( \rho = 0.9 ) (power)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( t-\text{IV}_1 )</td>
<td>( t-\text{IV}_2 )</td>
</tr>
<tr>
<td>10</td>
<td>70</td>
<td>10.1</td>
<td>14.4</td>
</tr>
<tr>
<td>20</td>
<td>70</td>
<td>7.6</td>
<td>12.0</td>
</tr>
<tr>
<td>30</td>
<td>70</td>
<td>7.1</td>
<td>11.8</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>9.1</td>
<td>12.9</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>7.4</td>
<td>12.1</td>
</tr>
<tr>
<td>30</td>
<td>100</td>
<td>6.5</td>
<td>10.6</td>
</tr>
</tbody>
</table>

Linear trend

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>( \rho = 1 ) (size)</th>
<th>( \rho = 0.9 ) (power)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>70</td>
<td>9.7</td>
<td>13.8</td>
</tr>
<tr>
<td>20</td>
<td>70</td>
<td>8.2</td>
<td>13.3</td>
</tr>
<tr>
<td>30</td>
<td>70</td>
<td>7.9</td>
<td>11.9</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>10.1</td>
<td>14.2</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>7.7</td>
<td>13.0</td>
</tr>
<tr>
<td>30</td>
<td>100</td>
<td>8.0</td>
<td>11.2</td>
</tr>
</tbody>
</table>

Quadratic trend

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>( \rho = 1 ) (size)</th>
<th>( \rho = 0.9 ) (power)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>70</td>
<td>10.3</td>
<td>14.2</td>
</tr>
<tr>
<td>20</td>
<td>70</td>
<td>7.2</td>
<td>12.3</td>
</tr>
<tr>
<td>30</td>
<td>70</td>
<td>7.4</td>
<td>11.7</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>9.6</td>
<td>12.7</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>8.0</td>
<td>12.0</td>
</tr>
<tr>
<td>30</td>
<td>100</td>
<td>6.4</td>
<td>10.4</td>
</tr>
</tbody>
</table>

Smooth transition level break

Notes: \( t-\text{IV}_1 \)–\( t-\text{IV}_4 \) refer to the \( t-\text{IV} \) test statistic based on the deterministic, the weak stochastic, the strong stochastic, and the principal component instrument, respectively. \( t-\text{IV}_5 \) refers to the \( t-\text{IV} \) test statistic based on a data driven choice of instrument.