Linear Regression Limit Theory for Large Panels with Mixed Stationary and Nonstationary Regressors.

Malvina Marchese
LSE
May 5, 2012

Abstract
We establish the Normality and the rates of convergence of the pooled OLS estimator in large panel models with mixed stationary and nonstationary regressors. All results are derived for joint limits, where $T$ goes to infinity followed by $n$. We find that in a cointegrated model the estimated coefficient of the stationary regressor converges at rate $\sqrt{nT}$ and the estimated coefficient of the nonstationary regressor converges at a faster rate of $T\sqrt{n}$. In the spurious model both converge at rate $\sqrt{nT}$. This constitutes an improvement with respect to nonstationary spurious panel models where the convergence rate is $\sqrt{n}$. We derive the asymptotic variances of the estimator in the cointegrated and in the spurious model.

1 Introduction

The advantages of panel data over cross section and time series data have long been established in econometric research. Panel data sets usually give the researcher a larger number of data points than conventional cross section and time series data, thus increasing the degrees of freedom and reducing multicollinearity among explanatory variables. This results in more reliable parameter estimates and most importantly enables the researcher to specify and test more sophisticated models with less restrictive behavioural assumptions. Panel data allow to identify and measure effects that are simply not detectable in pure cross section or time series and allow to eliminate or reduce the estimation bias.

A panel data set contains observations on a given number of individuals ($n$) across time ($T$). As such it is a double indexed process and any treatment of asymptotics must take this into account. Furthermore, assumptions must be made on both the cross section and the time series dimension of the model.

The initial focus of research has been on identifying and estimating effects from panel models with a large number of cross section and few time series
observations. Asymptotics for the standard panel estimators in this setting are well established since the work of Hsiao(1986) and Chamberlain(1984).

However, starting from the Nineties, empirical work has used panel data sets with a nonstationary time series components and both large $n$ and $T$ available. Examples of this literature range from testing growth convergence theories in macroeconomics to estimating long run relations between international financial series such as relative prices and exchange rates. These works have been enhanced and facilitated by the availability of a number of important panel data sets covering different individuals, regions and countries over a relatively long period of time, such as the Penn World Table. In a context of few time series observations, the non stationarity of the series cannot be addressed properly but with larger data sets it can be explored properly. When the time series component of the model is assumed nonstationary and both large $n$ and $T$ are taken into account, traditional limit theory is no longer valid. Phillips and Moon investigated regressions with nonstationary panel data for which the time series component is an integrated of order one process I (1) with large $n$ and $T$. They show that, under a variety of cointegrating relations between the regressors and the regressand, the pooled OLS estimator is consistent and asymptotically normal. Their seminal paper doesn’t allow for the presence of both stationary and nonstationary regressors in the same model. In practice, however this framework is very relevant. There are many economic models with mixed stationary weekly dependent and unit root processes. For example, the analysis of demand systems where budget shares or quantities are regressed on relative prices and real income for different countries over time. Typically some prices are quite stable, I (0), and some other are trending, I (1). Money demand equations offer a similar mixture of stationary and nonstationary variables, with real income trending over time for most countries but stationary interest rates.

A comprehensive limit theory for the pooled OLS estimator in this framework has not yet been developed. Baltagi, Kao and Liu (2008) develop asymptotics in a simple panel regression model with error component disturbances. They assume that both the regressor and the remainder disturbance term are autoregressive and possibly nonstationary and derive sequential limit theory for the most common panel estimators. They show that in a model with random error component the OLS estimator has a normal asymptotic distribution and different rates of convergence according to the nonstationarity of the regressor and the remainder disturbance. Their results allow for the presence of a stationary and a non stationary term in the same model, however they do not allow for the simultaneous presence of a stationary and a nonstationary regressor in the model. As a consequence asymptotics for the pooled OLS estimator are derived for either a stationary regressor or a nonstationary regressor with stationary or non stationary disturbances. This paper derives the asymptotic distribution of the OLS estimator in a linear panel regression model with mixed stationary and non stationary regressors as $T$ and $n$ increase to infinity. Our results are for sequential limits, with $T$ going to infinity followed by $n$. The paper is organized as follows. Section 2 presents some literature review on the most recent development in non stationary panel data. Section 3 lays out the model and
the assumptions, with discussion. Section 3 presents the main results, mainly the asymptotic normality and the convergence rates of the estimator in the cointegrated and spurious model. Section 4 concludes the paper and points out some further directions of work. All proofs are contained in the appendix. Notation is fairly standard. The symbol " $\rightarrow_{p}$ " signifies convergence in probability. " $\rightarrow_{d}$ " is convergence in distribution. The inequality " $>$ " signifies positive definiteness when applied to matrices and $||\Omega||$ is the Euclidean norm of the matrix $\Omega$. " $(T,n \rightarrow \infty)_{seq}$ " denotes sequential limits where $T$ goes to infinity followed by $n$.

Brownian motions $W(t)$ on $[0,1]$ are usually written as $W$, and stochastic integrals like $\int_{0}^{1} W(t) dW(t)$ as $\int W dW$.

2 Literature review.

Since the beginning of the Nineties there has been much ongoing research on nonstationary panel data. Quah (1994), Levin and Lin (1993) consider unit root time series regressions with nonstationary panel data and propose a test statistic for unit root. Pedroni (1995) studies some properties of cointegration statistics in pooled time series panel. Robertson and Symons (1992) study the bias that is likely to arise in practise with nonstationary panel data. Baltagi and Kramer (1997) and, more recently Kao and Emerson (2004) investigate the case of a panel time trend model.

Pesaran and Smith (1995) examine the impact of nonstationary variables on cross section regression estimates with a large number of groups ($n$) available and a large number of time periods ($T$). Assuming that the parameter of interest is the average effect of some exogenous variable on a dependent variable, they argue that when $T$ is large enough it is sensible to run separate regression for each cross section group. In particular they examine the impact of nonstationary variables on the cross section estimates. Under some quite strong assumptions such as exogeneity of the regressors and i.i.d disturbances, they show that no spurious correlation will arise between two I(1) variables and that the cross section OLS estimator of the average effect will be consistent for large $T$.

Phillips and Moon (1999) extend the work of Pesaran and Smith to a very general setting and present a fundamental framework for asymptotics of the OLS estimator in large nonstationary panels. They investigate the behaviour of the OLS estimator in panel models where all regressors are nonstationary under four possible cointegration structures: with no cointegration, with heterogeneous cointegration, with homogenous and near-homogenous cointegration. When there is no cointegration between the regressors and the regressand, the model is said to be spurious. If panel observations with both large cross sectional and time series components are available then, even if the noise is quite strong it can be characterised as as independent across individuals. By pooling the cross section and the time series observations, the OLS estimator attenuates the strong effect of the residuals in the regression while retaining the strength of the signal. Phillips
and Moon show the existence of a very interesting long run average relationship between the regressors and the regressand and they prove that the pols estimator of such relation is both $\sqrt{n}$ consistent and asymptotically normal.

When regressors and regressand are cointegrated results for three cointegration structures are provided: heterogeneous, homogenous and near homogenous cointegration. A cointegrating relation exists between regressors and regressand when their conditional long run variance matrix has deficient rank. If different cointegrating relations are allowed across individuals the model is known as an heterogenous cointegration model. When the cointegrating relation is the same across all individuals there is homogenous cointegration and if there is only slightly different cointegration among individuals the model is a nearly homogenous cointegrated. In the first case Phillips and Moon show that the pols estimator consistently estimates the long run average coefficient between regressors and regressand. By the same logic of the spurious regression model, $\sqrt{n}$ consistency and asymptotic normality are obtained since cross section pooling attenuates the strength of the noise relative to the signal of the regression. The estimator is found $n\sqrt{T}$ consistent and asymptotically normal both in the homogenous and in the near homogenous cointegration model.

The development of asymptotic theory for panel data with large $n$ and $T$ requires assumptions on the treatment of the two indexes. Different approaches are possible. One approach is to fix one index and allow the other to pass to infinity giving an intermediate limit. By letting the other index pass infinity subsequently a sequential limit is obtained. A second approach lets the two indexes pass to infinity along a specific diagonal path determined by a monotonically increasing function relation of the type $T = T(n)$ where $n \to \infty$. This approach is known as diagonal path limit therey. A third approach allows both indexes to infinity simultaneously without any restriction on the path of divergence. This approach is know as joint limit theory. Diagonal path limit theory requires assuming a very specific expansion path and may thus fail to provide an appropriate approximation for a given $(N,T)$ situation. Joint limit theory requires stronger conditions than sequential but, on the other hand, sequential limits can give misleading results when both indexes are allowed to pass to infinity simultaneously. Phillips and Moon’s results are valid both for sequential and joint limits. To derive the latter they need to strenghten the assumptions by imposing the rate condition $n/T \to 0$. In practise their limit theory is more likely to be useful when $n$ is moderate and $T$ is large. Such data configuration can be expected in multicountry macroeconomic data for example when attention is restricted to group countries such as OECD nations or developing countries.

Phillips and Moon’s seminal paper provides a very exhaustive investigation into the asymptotics of the pols estimator in large panel data models with nonstationary $I(1)$ regressors. At present the only work in the literature on large panels that considers the simultaneous presence of stationary and nonstationary components is Baltagi, Kao and Liu (2008). This paper studies the asymptotic properties of the most common panel estimators in a simple error component disturbance model. Both the regressors and the remainder term are assumed autoregressive and possibly nonstationary. Baltagi, Kao and Liu present results
for the scalar case but their findings can be easily extended to multiple regression
provided that all regressors are either stationary or nonstationary. They assume
cross section independence and random effect of the individual effect. All their
results are obtained for sequential limits. They consider the following model:

\[ y_{it} = \alpha + \beta x_{it} + u_{it} \]

with \( i = 1, \ldots, n \) and \( t = 1, \ldots, T \) and assume that

\[ u_{it} = \mu_i + \nu_{it} \]

\( \mu_i \sim i.i.dN(0, \sigma^2) \) independent from \( \nu_{it} \) for all \( i \) and \( T \) and (random effect
assumption):

\[ E(\mu_i|x_{it}) = 0 \]

\[ \nu_{it} = \rho \nu_{i,t-1} + \epsilon_{it} \]

with \( |\rho| \leq 1 \) and \( \epsilon_{it} \) white noise.

\[ x_{it} = \lambda x_{i,t-1} + \varepsilon_{it} \]

with \( |\lambda| \leq 1 \) and \( \varepsilon_{it} \) white noise.

Baltagi, Kao and Liu find that the asymptotic properties of the pols estimator depend crucially on the nonstationarity of the regressor and the re-
mainder disturbance. When the erroro component in the disturbance term and
the regressors are both stationary (\( |\rho| < 1 \) and \( |\lambda| < 1 \)) the estimator is
\( \sqrt{nT} \) consistent and asymptotically normal. If the disturbance is I(1) and
the regressor is I(0) (\( \rho = 1 \) and \( |\lambda| < 1 \)) the estimator is still consistent and
asymptotically normal with rate of convergence \( \sqrt{n} \). It is interesting to notice
that, when only few time series observation are available for this case the esti-
mator is inconsistent. When the disturbance is I(0) and the regressor is I(1)
\( (|\rho| < 1 \) and \( \lambda = 1 \)) the models is cointegrated: the estimator is \( \sqrt{nT} \) con-
sistent and asymptotically normal. If both the disturbance and the regressor
are I(1) the models is known as spurious, as there is no cointegrating relation
between the regressor and the regressand. In this case, when only time series
observations are available it is well know that the ols estimator has a nondegen-
erate limit distribution (Phillips 1986). However when panel observations with
large cross section and time series components are available, even if the noise in
the time series regression is strong, by pooling the estimator can still provide a
consistent estimate of some long run regression coefficient. And indeed Baltagi
Kao and Liu find that the estimator is asymptotically normal and \( \sqrt{n} \) consist-
tent.
3 The model and the assumptions.

We consider the following panel regression model:

\[ y_{it} = \alpha + \beta x_{it} + \gamma z_{it} + \eta_{it} \]

Where

\[ x_{it} = x_{it-1} + \varepsilon_{it} \]
\[ z_{it} = \rho z_{i,t-1} + u_{it} \quad \text{with} \quad |\rho| < 1 \]
\[ \eta_{it} = \lambda \eta_{it-1} + v_{it} \quad \text{with} \quad |\lambda| \leq 1 \]

For simplicity of notation we present results for the scalar case, which extends straightforward to the vector case. When the disturbance is I(0) \((|\lambda| < 1)\) the model is cointegrated, when the disturbance is I(1) \((\lambda = 1)\) the regression is spurious, there is no stationary linear combination of the regressor and the regressand.

We wish to estimate the \(3 \times 1\) vector \(\beta_0 = (\alpha, \beta, \gamma)\)' using panel observations on \((x_{it}, z_{it}, y_{it})\) with \(i = 1, \ldots, n\) and \(t = 1, \ldots, T\).

We make the following assumptions:

**ASSUMPTION 1:** All the three processes \(x_{it}, z_{it}, \eta_{it}\) have common initialisation at \(t = 0\) satisfying \(w_i = (x_{i0}, z_{i0}, \eta_{i0})\) i.i.d across \(i\) with \(E|w_i| < \infty\)

**ASSUMPTION 2:** \(w_{it} = \Phi(L)\xi_{it} = \sum_{j=0}^{\infty} \Phi_j \xi_{it-j} \quad \text{with} \quad \sum_{j=0}^{\infty} j^p|\Phi_j| < \infty \quad \text{and} \quad |\Phi(1)| = \sum_{j=0}^{\infty} \Phi_j \neq 0 \text{ for some } p > 1\)

**ASSUMPTION 3** \(\forall i \xi_{it} \text{ is iid with zero mean, finite variance } \Xi, \text{ and finite fourth order cumulants.}\)

**ASSUMPTION 4:** \(\forall t \xi_{it} \text{ and } \xi_{jt} \text{ are independent for } i \neq j\)

**ASSUMPTION 5:** \(\forall i \forall k \quad E(\nu_{it} \varepsilon_{it+k}) = E(\nu_{it} u_{it+k}) = E(\varepsilon_{it} u_{it+k}) = 0\)

**ASSUMPTION 6:** We assume that conditions for exchangeability limits integration hold (Fubini’s Theorem).
**Remark 1** Assumption 1 is made for convenience and could be generalized to allow for remote past initialisation at the cost of some further complication (Phillips and Lee (1996)).

**Remark 2** Assumption 2 assumes that the model disturbances are generated as a vector linear process. Phillips and Moon (1999) assume that the disturbances are generated by random coefficient linear processes and impose summability and moment conditions on the random coefficients, we follow Baltagi Kao and Liu’s approach and impose moment conditions on the $\xi_{it}$ directly with nonrandom coefficients.

**Remark 3** Assumption 2 and Assumption 3 ensure the existence of a BN decomposition that satisfies a multivariate invariance principle (Phillips and Solo (1992)). This implies the applicability of a FCLT to each process separately and to their outer product.

**Remark 4** Assumption 3 is required to hold for fixed $i$ since our results employ only sequential limits, for joint limits this assumption must be reinforced to hold also for $n \to \infty$.

**Remark 5** With Assumption 4 we introduce cross sectional independence in the model. This is a very strong assumption though very common in large panel literature at present. The limitation it imposes on the model are illustrated in the concluding remarks. It is made to ensure the applicability of LLNs in sequential limits, once the intermediate limit for $T$ is obtained, to cross sectional averages of independent random variables.

**Remark 6** Assumption 5 rules out endogeneity which simplifies our asymptotic results as it rules out the need to centre around the asymptotic bias to obtain convergence to Normality, but it does not affect the convergence rates.

### 4 Main results.

In the matrix notation model

$$Y = W' \theta + \eta$$

the Pooled OLS estimator is obtained as

$$\hat{\theta} = (W'W)^{-1} W'Y$$

**Theorem 7** Under assumption $A1,A2,A3,A4,A6$, with $|\lambda| < 1$ for $(T,n \to \infty)_{\text{seq}}$
\[ D^{\frac{1}{2}} \left( \hat{\vartheta} - \vartheta \right) \rightarrow_d N(0, V) \]

where \( D^{\frac{1}{2}} = \text{diag}(\sqrt{nT}, T \sqrt{n}, \sqrt{nT}) \).

**Proof.** The proof follows from Lemma 8, Lemma 9, Lemma 10 as a straightforward application of Cramer Theorem. 

**Lemma 8** Under assumption \( A1,A2,A3,A4,A6 \), with \( |\lambda| < 1 \) for \((T, n \rightarrow \infty)_{\text{seq}}\)

\[ \left( D^{-\frac{1}{2}} W' WD^{-\frac{1}{2}} \right) \rightarrow^p \Omega \text{ and } \Omega > 0 \]

**Proof.** See appendix. 

**Lemma 9** Under assumption \( A1,A2,A3,A4,A6 \), with \( |\lambda| < 1 \) for \((T, n \rightarrow \infty)_{\text{seq}}\)

\[ \left( D^{-\frac{1}{2}} W' WD^{-\frac{1}{2}} \right)^{-1} \rightarrow^p \Omega^{-1} \]

**Proof.** From Lemma 8 and CMT. 

**Lemma 10** Under assumption \( A1,A2,A3,A4,A6 \), with \( |\lambda| < 1 \) for \((T, n \rightarrow \infty)_{\text{seq}}\)

\[ D^{-\frac{1}{2}} W' \eta \rightarrow_d N(0, \Sigma) \text{ for } (T, n \rightarrow \infty)_{\text{seq}} \]

**Proof.** See appendix. 

**Theorem 11** Under assumptions \( A1,A2,A3,A4,A5,A6 \), with \( |\lambda| < 1 \) for \((T, n \rightarrow \infty)_{\text{seq}}\) the asymptotic variance of the pols estimator \( V \) is obtained as

\[
V = \sigma^2 \eta \begin{bmatrix}
1 & 0 & 0 \\
0 & \frac{(\sigma^2)^2}{2} & 0 \\
0 & 0 & \frac{\sigma^2}{1-p}
\end{bmatrix}
\]
Proof. See appendix.

Theorem 7 provides asymptotics for the pols estimator in a cointegrated model. Results are coherent with Phillips and Moon (1999). In a cointegrated model with non stationary regressors they find that the rate of convergence of the estimator is $T \sqrt{n}$ as we do, which is not surprising since in presence of a cointegrating relation adding one more stationary regressor does not modify crucially any derivation. It is interesting to observe that the estimators of the coefficients of the stationary regressors converge at a rate of $\sqrt{nT}$ which is the standard convergence rate in large stationary panels. The presence of one non-stationary regressor does not slow down the convergence rate of these estimated coefficients thanks to the existence of a cointegrating relation.

\textbf{Theorem 12} Under assumptions A1, A2, A3, A4, A6, with $|\lambda| = 1$ for $(T, n \to \infty)_{seq}$

$$d^{\frac{1}{2}} \left( \hat{\theta} - \theta \right) \to_d N(0, V^*)$$

where $d^{\frac{1}{2}} = \text{diag} \left( \sqrt{nT}, \sqrt{nT}, \sqrt{nT} \right)$

\textbf{Proof.} The proof follows from Lemma 13, Lemma 14 and Lemma 15 as a straightforward application of Cramer Theorem.

\textbf{Lemma 13} Under assumption A1, A2, A3, A4, A6, with $\lambda = 1$ for $(T, n \to \infty)_{seq}$

$\left( d^{-\frac{1}{2}} W^T W d^{-\frac{1}{2}} \right) \to_p \Xi$ and $\Xi > 0$

\textbf{Proof.} As for Lemma 8.

\textbf{Lemma 14} Under assumption A1, A2, A3, A4, A6, with $\lambda = 1$ for $(T, n \to \infty)_{seq}$

$\left( d^{-\frac{1}{2}} W^T W d^{-\frac{1}{2}} \right)^{-1} \to_p \Xi^{-1}$

\textbf{Proof.} Straightforward application of CMT on Lemma 13.

\textbf{Lemma 15} Under assumption A1, A2, A3, A4, A6, with $\lambda = 1$ for $(T, n \to \infty)_{seq}$

$$d^{-\frac{1}{2}} W^T \eta \to_d N(0, \Phi) \text{ for } (T, n \to \infty)_{seq}$$
Proof. See appendix. ■

Theorem 12 provides asymptotics for the pols estimator in a spurious model. As already noted by Phillips and Moon (1999) this results represent a further instance of the ability of panel data models to identify and estimated effects not detectable with pure cross section or time series data. It is well know that with time series observations of a spurious model the regression coefficient has a nondegenerate limit distribution. The noise is so strong that it dominates the signal, however once large panel observations are available the strong effect of the residuals is attenuated and the regression can provide some consistent estimate.

Theorem 12 shows an improvement from Phillips and Moon and Baltagi, Kao Liu’s results. The presence of stationary regressor speeds up the rate of convergence of the non stationary regressor to $\sqrt{nT}$ whereas in models with only nonstationary regressors the rate is $\sqrt{n}$.

5 Conclusions.

This paper has developed a linear regression limit theory for large panels with mixed stationary and nonstationary regressors. Results are consistent with Phillips and Moon’s seminal paper (1999) and confirm the ability of panel data to identify long run average relationships over the cross section even when the original regression is spurious. The most interesting finding is the improved rate of convergence on the estimated coefficient of the nonstationary regressor in the spurious model. The presence of one stationary regressor in the equation increases the convergence to Normality to $\sqrt{nT}$. All the results are derived for sequential limit theory with $T$ going to infinity followed by $n$. Furthermore all proves rely on cross section independence, which is a common assumption in large panel literature but quite restrictive. For instance, multi-country GDP series, exchange rates, and financial asset prices all involve cross section dependence. In general, finding a natural ordering for cross section dependence in economic data is not easy; recent research has attempted to employ either a spatial framework or a factor model framework but still much has to be done to model simultaneously cross section and time series dependence. This paper can be extended in many directions. Under some strengthening of the initial assumptions (higher order moment conditions on the $\xi_i$ and convergence of the ratio $\sqrt{n}/T$ to zero) we could derive joint asymptotics. As identifying individual heterogeneity is one of the main abilities of panel data, the introduction of an individual effect seems highly desirable. It would be of particular interest to work out the asymptotics of the estimator for large $n$ and $T$ under fixed effect assumptions. Baltagi Kao and Liu show the existence of a long run average relation between regressor and regressand in a spurious individual heterogeneity model under random effect assumption. It turns out that this limit depends only on the long run variance matrix of the model disturbances. Further study into the structure of this matrix under random effect assumption seems of some importance as most economic
models imply some correlation between the individual effect and one or more controls. A further step could be to include second order stationary regressors in the model. The simultaneous presence of I(0), I(1), I(2) is common for example in empirical literature on the effect of inequality on growth where there are evidences of second order stationarity in Gini’s coefficient, first order stationarity in Income or PPP and stationarity in other explanatory variables. There has been a long ongoing debate in this literature on the best model to represent the relation between inequality and growth, the standard procedure being to assume a simple linear relationship, the more recent a non linear one. However none of these approaches has taken into account the nonstationarity of some of the regressors which could indeed account for much of the unexplained so far.

6 References:

References


7 Appendix

7.1 Preliminaries.

We start by plugging in the true model the matrix notation model:

\[ \hat{\varphi} - \varphi = (W'W)^{-1} W' \eta \]

\[
(W'W)^{-1} = \begin{bmatrix}
\sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} & \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it}^2 & \sum_{i=1}^{n} \sum_{t=1}^{T} z_{it} x_{it} \\
\sum_{i=1}^{n} \sum_{t=1}^{T} z_{it} & \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} z_{it} & \sum_{i=1}^{n} \sum_{t=1}^{T} z_{it}^2 \\
\end{bmatrix}^{-1}
\]

\[
W' \eta = \begin{bmatrix}
\sum_{i=1}^{n} \sum_{t=1}^{T} \eta_{it} \\
\sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} \eta_{it} \\
\sum_{i=1}^{n} \sum_{t=1}^{T} z_{it} \eta_{it} \\
\end{bmatrix}
\]

Pre multiplying both sides by \( D^{-\frac{1}{2}} \) we obtain

\[
D^{-\frac{1}{2}} \left( \hat{\varphi} - \varphi \right) = \left( D^{-\frac{1}{2}} W' W D^{-\frac{1}{2}} \right)^{-1} D^{-\frac{1}{2}} W' \eta
\]
7.2 Proof of Lemma 8

\[
(D^{-\frac{1}{2}}W'WD^{-\frac{1}{2}}) =
\]

Claim 16

\[
\begin{bmatrix}
\frac{1}{\sqrt{nT}} & 0 & 0 \\
0 & \frac{1}{\sqrt{nT}} & 0 \\
0 & 0 & \frac{1}{\sqrt{nT}}
\end{bmatrix}
\begin{bmatrix}
nT & \sum_{i=1}^{n} x_{it} & \sum_{i=1}^{n} z_{it} \\
\sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} & \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it}^2 & \sum_{i=1}^{n} \sum_{t=1}^{T} z_{it}^2 \\
\sum_{i=1}^{n} \sum_{t=1}^{T} z_{it} & \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} z_{it} & \sum_{i=1}^{n} \sum_{t=1}^{T} z_{it}^2
\end{bmatrix}
\begin{bmatrix}
\frac{1}{\sqrt{nT}} & 0 & 0 \\
0 & \frac{1}{\sqrt{nT}} & 0 \\
0 & 0 & \frac{1}{\sqrt{nT}}
\end{bmatrix}
= \begin{bmatrix}
1 & \frac{1}{nT^2} \sum_{i=1}^{n} x_{it} & \frac{1}{nT^2} \sum_{i=1}^{n} z_{it} \\
\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} & \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it}^2 & \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} z_{it}^2 \\
\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} z_{it} & \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} z_{it} & \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} z_{it}^2
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 \\
0 & \frac{(\sigma_x)^2}{2} & 0 \\
0 & 0 & \frac{\sigma_x^4}{1-p^2}
\end{bmatrix}
= \Omega (T, n \to \infty)
\]

Convergence in probability is proved pointwise, element by element.

(1,1): \( (D^{-\frac{1}{2}}W'WD^{-\frac{1}{2}})_{11} = 1 \overset{p}{\to} 1 \) trivially

(1,2)=(2,1):

\[
(D^{-\frac{1}{2}}W'WD^{-\frac{1}{2}})_{12} = \left(D^{-\frac{1}{2}}W'WD^{-\frac{1}{2}}\right)_{21} = \frac{1}{nT^2} \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} x_{it}
\]
\[
\frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{T^2} \sum_{t=1}^{T} x_{it} \right) \to_d \frac{1}{n} \sum_{i=1}^{n} \left( \sigma_\varepsilon \int_0^{1} W_i(r) dr \right) \quad T \to \infty \forall i \text{ by FCLT}
\]

\[
\frac{1}{n} \sum_{i=1}^{n} \left( \sigma_\varepsilon \int_0^{1} W_i(r) dr \right) \to_p E \left( \sigma_\varepsilon \int_0^{1} W_i(r) dr \right) = 0 \quad n \to \infty \forall T
\]

as \( E \left| \sigma_\varepsilon \int_0^{1} W_i(r) dr \right|^2 \) < \( \infty \) \( \forall T \) and for \( T \to \infty \) by LLN for ind seq

(1,3)=(3,1):

\[
\left( D^{-\frac{1}{2}}W'WD^{-\frac{1}{2}} \right)_{13} = \left( D^{-\frac{1}{2}}W'WD^{-\frac{1}{2}} \right)_{31} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} z_{it} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T} \sum_{t=1}^{T} z_{it}
\]

\[
\frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{T} \sum_{t=1}^{T} z_{it} \right) \to_d \frac{1}{n} \sum_{i=1}^{n} E(z_i) \quad T \to \infty \forall i \text{ by Ergodic Th}
\]

\[
\frac{1}{n} \sum_{i=1}^{n} E(z_i) \to_P E(z_i) = 0
\]

as \( E \left| \frac{1}{T} \sum_{t=1}^{T} z_{it} \right|^2 \) < \( \infty \) \( \forall T \) and for \( T \to \infty \) by LLN for ind seq

(2,2):

\[
\left( D^{-\frac{1}{2}}W'WD^{-\frac{1}{2}} \right)_{22} = \frac{1}{T^2n} \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it}^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} x_{it}^2
\]

\[
\frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{T^2} \sum_{t=1}^{T} x_{it}^2 \right) \to_d \frac{1}{n} \sum_{i=1}^{n} \left( \sigma_\varepsilon \int_0^{1} W_i^2(r) dr \right) \quad T \to \infty \forall i \text{ FCLT}
\]

\[
\frac{1}{n} \sum_{i=1}^{n} \left( \sigma_\varepsilon \int_0^{1} W_i^2(r) dr \right) \to_P E \left( \sigma_\varepsilon \int_0^{1} W_i^2(r) \right) = \frac{(\sigma_\varepsilon)^2}{2} \quad n \to \infty \forall T
\]
\[
\text{as } E \left| \sigma_z \int_0^1 W_i^2(r)dr \right|^2 < \infty \forall T \text{ and for } T \to \infty \text{ by LLN for ind seq}
\]

(2.3)\text{=(3.2):}

\[
\left( D^{-\frac{1}{2}} W^t W D^{-\frac{1}{2}} \right)_{23} = \left( D^{-\frac{1}{2}} W^t W D^{-\frac{1}{2}} \right)_{32} = \frac{1}{nT^{\frac{1}{2}}} \sum_{i=1}^n \sum_{t=1}^T x_{it} z_{it} = \frac{1}{n} \sum_{i=1}^n \frac{1}{T^{\frac{1}{2}}} \sum_{t=1}^T x_{it} z_{it}
\]

\[
\frac{1}{n} \sum_{i=1}^n \left( \frac{1}{T^{\frac{1}{2}}} \sum_{t=1}^T x_{it} z_{it} \right) \rightarrow_d \frac{1}{n} \sum_{i=1}^n (\sigma_z \sigma_u) \int_0^1 W_i^2 dW_i^2 \text{ as } T \to \infty \quad \forall i
\]

(Phillips and Solo (1992))

\[
\frac{1}{n} \sum_{i=1}^n (\sigma_z \sigma_u) \int_0^1 W_i^2 dW_i^2 \rightarrow_p E \left| \left( \sigma_z^2 \sigma_u^2 \right) \int_0^1 W_i^2 dW_i^2 \right| = 0 \quad n \to \infty
\]

\[
\text{as } E \left| \left( \sigma_z^2 \sigma_u^2 \right) \int_0^1 W_i^2 dW_i^2 \right| < \infty \forall T \text{ and for } T \to \infty \text{ by LLN for ind seq}
\]

(3.3):

\[
\left( D^{-\frac{1}{2}} W^t W D^{-\frac{1}{2}} \right)_{33} = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T z_{it}^2 = \frac{1}{n} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T z_{it}^2
\]

\[
\frac{1}{n} \sum_{i=1}^n \left( \frac{1}{T} \sum_{t=1}^T z_{it}^2 \right) \rightarrow_p \frac{1}{n} \sum_{i=1}^n E(z_{it}^2) = \frac{1}{n} \sum_{i=1}^n \frac{\sigma_z^2}{1-\rho^2} T \rightarrow \infty \quad \forall i \text{ by ErgodicTh}
\]

\[
\frac{1}{n} \sum_{i=1}^n \frac{\sigma_z^2}{1-\rho^2} \rightarrow_p \frac{\sigma_z^2}{1-\rho^2}
\]

as \( \max_i E \left\{ |X_i|^{1+\eta} \right\} < \infty \) for some \( \eta > 0 \) \forall T \text{ and for } T \to \infty \text{ by LLN for ind seq}
7.3 Proof of Lemma 9.

\[ D^{-\frac{1}{2}} W' \eta = \begin{pmatrix} 1 \sqrt{T} \sum_{i=1}^{n} \sum_{t=1}^{T} \eta_{it} \\ T \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} \eta_{it} \\ \sqrt{n} \sum_{i=1}^{n} \sum_{t=1}^{T} z_{it} \eta_{it} \end{pmatrix} = \begin{pmatrix} 1 \sqrt{n} \sum_{i=1}^{n} \eta_{it} \\ \frac{1}{T} \sum_{t=1}^{T} x_{it} \eta_{it} \\ \frac{1}{T} \sum_{t=1}^{T} z_{it} \eta_{it} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{i,T} \\ \frac{1}{\sqrt{T}} \sum_{i=1}^{T} x_{it} \eta_{it} \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_{it} \eta_{it} \end{pmatrix} \]

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{i,T} \rightarrow_d N(0, \Sigma) \iff \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \lambda' Z_{i,T} \rightarrow_d N(0, \lambda' \Sigma \lambda) \text{ for } \lambda \neq 0 \]

\[ \lambda' Z_{i,T} = \frac{\lambda_1}{\sqrt{T}} \sum_{t=1}^{T} \eta_{it} + \frac{\lambda_2}{T} \sum_{t=1}^{T} x_{it} \eta_{it} + \frac{\lambda_3}{\sqrt{T}} \sum_{t=1}^{T} z_{it} \eta_{it} \]

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \lambda' Z_{i,T} \rightarrow_d N(0, \lambda' \Sigma \lambda) \iff \text{E} |\lambda' Z_{i,T}|^2 < \infty \text{ for } t \text{ and for } T \rightarrow \infty \]

\[ \lim_{T \rightarrow \infty} \text{E} |\lambda' Z_{i,T}|^2 = \lim_{T \rightarrow \infty} \left( \left| \frac{\lambda_1}{\sqrt{T}} \sum_{t=1}^{T} \eta_{it} \right|^2 + c_r \left| \frac{\lambda_2}{T} \sum_{t=1}^{T} x_{it} \eta_{it} \right|^2 + c_r \left| \frac{\lambda_3}{\sqrt{T}} \sum_{t=1}^{T} z_{it} \eta_{it} \right|^2 \right) \leq \lim_{T \rightarrow \infty} c_r \left( \left| \frac{\lambda_1}{\sqrt{T}} \sum_{t=1}^{T} \eta_{it} \right|^2 + c_r \left| \frac{\lambda_2}{T} \sum_{t=1}^{T} x_{it} \eta_{it} \right|^2 + c_r \left| \frac{\lambda_3}{\sqrt{T}} \sum_{t=1}^{T} z_{it} \eta_{it} \right|^2 \right) \]

1:

\[ \lim_{T \rightarrow \infty} \text{E} \left| \frac{\lambda_1}{\sqrt{T}} \sum_{t=1}^{T} \eta_{it} \right|^2 < \infty \]

\[ \frac{\lambda_1}{\sqrt{T}} \sum_{t=1}^{T} \eta_{it} \rightarrow_d N(0, \lambda_1^2 \sigma_{\eta}^2) \text{ for } T \rightarrow \infty \]

\[ \lim_{T \rightarrow \infty} \text{E} \left| \frac{\lambda_1}{\sqrt{T}} \sum_{t=1}^{T} \eta_{it} \right|^2 = \lambda_1^2 \sigma_{\eta}^2 < \infty \text{ by Fubini's.} \]
2: 

\[ \lim_{T \to \infty} E \left| \frac{\lambda_2}{T} \sum_{t=1}^{T} x_{it} \eta_{it} \right|^2 < \infty \]

\[ \frac{\lambda_2}{T} \sum_{t=1}^{T} x_{it} \eta_{it} \Rightarrow \frac{\lambda_2}{2} \sigma_d^2 (W(1)^2 - 1) < \infty \text{ for } T \to \infty \]

3: 

\[ \lim_{T \to \infty} E \left| \frac{\lambda_3}{\sqrt{T}} \sum_{t=1}^{T} z_{it} \eta_{it} \right|^2 < \infty \]

\[ \frac{\lambda_3}{\sqrt{T}} \sum_{t=1}^{T} z_{it} \eta_{it} \rightarrow_d N(0, \sigma_\eta^2 \sigma_z^2) \text{ by Ergodic th as } T \to \infty \]

\[ \lim_{T \to \infty} E \left| \frac{\lambda_3}{\sqrt{T}} \sum_{t=1}^{T} z_{it} \eta_{it} \right|^2 = \lambda_3^2 \sigma_\eta^2 \sigma_z^2 < \infty \text{ by Fubini} \]

7.4 Proof of Theorem 11:

\[ D^{-\frac{1}{2}} W' \eta \rightarrow_d N(0, \Sigma) \text{ (T, } n \rightarrow \infty) \text{ seq} \]

And

\[ (D^{-\frac{1}{2}} W' W D^{-\frac{1}{2}})^{-1} \rightarrow_p \Omega^{-1} (T, n \rightarrow \infty) \text{ seq} \]

By Cramer’s theorem

\[ D^{\frac{1}{2}} \left( \hat{\theta} - \theta \right) \rightarrow_d N(0, V) \text{ as } (T, n \rightarrow \infty) \text{ seq} V = \Omega^{-1} \Sigma \Omega^{-1} \]

\[ Var \left( D^{\frac{1}{2}} \left( \hat{\theta} - \theta \right) \right) = D^{\frac{1}{2}} Var((\hat{\theta} - \theta)) D^{\frac{1}{2}}' = D^{\frac{1}{2}} E \left[ (W' W)^{-1} W' \eta' W (W' W)^{-1} \right] D^{\frac{1}{2}} = D^{\frac{1}{2}} E \left[ (W' W)^{-1} W' E(\eta' | W) W (W' W)^{-1} \right] D^{\frac{1}{2}} = \sigma_n^2 E \left( D^{\frac{1}{2}} (W' W)^{-1} D^{\frac{1}{2}} \right) = \sigma^2 E \left( D^{-\frac{1}{2}} (W' W) D^{-\frac{1}{2}} \right)^{-1} \]

17
\[
V = \lim_{(T,n \to \infty)_\text{seq}} \text{Var}(D^{\frac{1}{2}}(\hat{\theta} - \theta)).
\]

\[
\sigma^2 \mathbb{E} \left( D^{\frac{1}{2}}(W'W)D^{\frac{1}{2}} \right) \overset{p}{\to} \sigma^2 \begin{bmatrix}
1 & 0 & 0 \\
0 & (\sigma_2^2) & 0 \\
0 & 0 & \frac{\sigma^2}{1-p^2}
\end{bmatrix} \quad (T,n \to \infty)_\text{seq} \text{ by Lemma 8}
\]

\[
V = \sigma^2 \begin{bmatrix}
1 & 0 & 0 \\
0 & (\sigma_2^2) & 0 \\
0 & 0 & \frac{\sigma^2}{1-p^2}
\end{bmatrix}
\]

7.5 Proof of Lemma 13

Claim 17

\[
\left( d^{\frac{1}{2}}W'Wd^{-\frac{1}{2}} \right) =
\]

\[
\frac{1}{\sqrt{nT}} \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & \frac{1}{\sqrt{nT}}
\end{bmatrix} \begin{bmatrix}
nT & \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} & \sum_{i=1}^{n} \sum_{t=1}^{T} z_{it} \\
\sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} & \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it}^2 & \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} z_{it} \\
\sum_{i=1}^{n} \sum_{t=1}^{T} z_{it} & \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} z_{it} & \sum_{i=1}^{n} \sum_{t=1}^{T} z_{it}^2
\end{bmatrix} \frac{1}{\sqrt{nT}} \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & \frac{1}{\sqrt{nT}}
\end{bmatrix} =
\]

\[
\left[ \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} \quad \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it}^2 \quad \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} z_{it} \right] \overset{p}{\to} \begin{bmatrix}
1 & 0 & 0 \\
0 & \sigma & 0 \\
0 & 0 & \frac{\sigma^2}{1-p^2}
\end{bmatrix} \quad \text{as}(T,n \to \infty)_\text{seq}
\]

Proof. as in the previous claim. ■
7.6 Proof of Lemma 15

\[ d^{-\frac{1}{2}}W'\eta = \begin{pmatrix} \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \eta_{it} \\ \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} \eta_{it} \\ \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} z_{it} \eta_{it} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta_{it} \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \eta_{it} \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_{it} \eta_{it} \end{pmatrix} \]

By Cramer Wold Device

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{i,T} \rightarrow_d N(0, \Sigma) \text{ if } f \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \lambda' Z_{i,T} \rightarrow_d N(0, \lambda' \Phi \lambda) \quad \forall \lambda \neq 0 \]

We claim

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \lambda' Z_{i,T} \rightarrow_d N(0, \lambda' \Phi \lambda) \quad \forall \lambda \neq 0 \text{ for } (T, n \rightarrow \infty)_{seq} \]

\[ \lambda' Z_{i,T} = \frac{\lambda_1}{\sqrt{T}} \sum_{t=1}^{T} \eta_{it} + \frac{\lambda_2}{\sqrt{T}} \sum_{t=1}^{T} x_{it} \eta_{it} + \frac{\lambda_3}{\sqrt{T}} \sum_{t=1}^{T} z_{it} \eta_{it} \]

By Linderberg-Levy CLT

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \lambda' Z_{i,T} \rightarrow_d N(0, \lambda' \Phi \lambda) \text{ iff } \mathbb{E} \left| \lambda' Z_{i,T} \right|^2 < \infty \quad \forall t \text{ and for } T \rightarrow \infty \]

1':

\[ \lim_{T \rightarrow \infty} \mathbb{E} \left| \frac{\lambda_1}{\sqrt{T}} \sum_{t=1}^{T} \eta_{it} \right|^2 < \infty \]

since

\[ \frac{\lambda_1}{\sqrt{T}} \sum_{i=1}^{T} \eta_{it} \rightarrow_d N(0, \lambda_1^2 \sigma^2_{\eta i}) \text{ by Ergodic Th as } T \rightarrow \infty \]

\[ \lim_{T \rightarrow \infty} \mathbb{E} \left| \frac{\lambda_3}{\sqrt{T}} \sum_{t=1}^{T} \eta_{it} \right|^2 = \lambda_3^2 \sigma^2_{\eta i} < \infty \text{ by Fubini's.} \]
2':

\[ \lim_{T \to \infty} E \left( \frac{\lambda_2}{\sqrt{T}} \sum_{t=1}^{T} x_{it} \eta_{it} \right)^2 < \infty \]

3':

\[ \lim_{T \to \infty} E \left( \frac{\lambda_3}{\sqrt{T}} \sum_{t=1}^{T} z_{it} \eta_{it} \right)^2 < \infty \]

since

\[ \frac{\lambda_1}{\sqrt{T}} \sum_{t=1}^{T} z_{it} \eta_{it} \to_d N(0, \sigma_{\eta_{it}}^2 \sigma_{z_{it}}^2) \text{by } T \to \infty \]

\[ \lim_{T \to \infty} E \left( \frac{\lambda_3}{\sqrt{T}} \sum_{t=1}^{T} z_{it} \eta_{it} \right)^2 = \frac{\lambda_3^2 \sigma_{\eta_{it}}^2 \sigma_{z_{it}}^2}{\sqrt{T}} < \infty \text{ by Fubini's} \]

Then

\[ d^{-\frac{1}{2}} W' \eta \to_d N(0, \Phi) \text{ as } (T, n \to \infty)_{\text{seq}} \]

and

\[ \left( d^{-\frac{1}{2}} W' W d^{-\frac{1}{2}} \right)^{-1} \to_p \Xi^{-1} \text{ as } (T, n \to \infty)_{\text{seq}} \]

by Cramer's Theorem

\[ D^{\frac{1}{2}} \left( \hat{\theta} - \theta \right) \to_d N(0, \Omega^*) \text{ with } V = \Xi^{-1} \Phi \Xi^{-1} \]