Linear hypothesis tests over fixed effects with serially correlated panels

Kevin D. Duncan

November 2019

Very Preliminary, please do not cite.

Please see `http://kdduncan.github.io/papers.html` *for the most current draft.*

Abstract

I develop a joint hypothesis test over fixed effects in large n small T panel data models with symmetric serial correlation among individuals. This enables joint hypothesis tests over inconsistent fixed effects estimates, including the traditional varying intercept model as well as models with individual specific slope coefficients. I establish two different set of assumptions where a feasible tests exist. The first assumption requires that individual errors follow a stationary $AR(p)$ process. Under this assumption all second and fourth cross product moments can be consistently estimated, while allowing for both individual specific hypothesis to be imposed on the data and covariates to vary across individuals and time with individual specific slopes. The second set of assumptions requires the presence of a known group structure of individuals such that there are at least two people in each group, and that the covariate values and coefficient slopes are shared among all individuals in that group. This set of assumptions enable estimation of a completely unconstrained variance-covariance matrix and higher cross product moments for individuals.

1 Introduction

This paper develops a Wald test for joint hypotheses over fixed effect in large n small T panel data models when the error process features symmetric serial correlation among individuals in the sample. I prove the asymptotic behavior of both a Wald tests using either OLS residuals under the assumptions of a known but unconstrained variance covariance matrix and an unspecified set of linear restrictions. I then show the existence of a feasible implementation under two different sets of assumptions. The first set of assumptions allow for a general set of linear hypotheses with individual and time varying covariates if the errors follow a stationary $AR(p)$ process. The alternative assumptions allow for a non-stationary variance-covariance structure at the cost of a known group structure and shared covariate values across individuals when testing hypotheses of the form that all individuals in the group have the same group-varying coefficient.[1] I focus on the following panel data model;

$$y_{it} = x_{it}'\beta + z_{it}'\gamma_i + \epsilon_{it}; \quad i = 1, ..., n; \; t = 1, ..., T \tag{1}$$

$$\mathrm{E}(\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i' \mid \mathbf{X}_i \; \mathbf{Z}_i) = \Omega \quad \forall i \tag{2}$$

$$\mathrm{E}(\epsilon_{it}\epsilon_{js} \mid \mathbf{X}_i \; \mathbf{Z}_i \; X_j \; Z_j) = 0 \quad \forall j \neq i; \; \forall t, s = 1, ..., T \tag{3}$$

where z_{it} is a $L \times 1$ vector of regressors whose coefficient values vary across individuals, and x_{it} is a $K \times 1$ vector of regressors whose coefficients are shared across individuals. This model is quite flexible, and includes both the usual additive fixed effects model when $z_{it}' = 1$ for all individuals, as well as additional regressors.[2] Recently a large literature has focused on estimation and inference of $\overline{\gamma_i}$, the population or sample average varying coefficients. Comparably, this paper provides a method for testing hypotheses of the form,

$$R_n[\beta \; \gamma_1 \; \gamma_2 \; \ldots \; \gamma_n]' = r_n \tag{4}$$

Where R_n is a $q_n \times K + nL$ matrix of linear restrictions, and r_n is a $q_n \times 1$ vector of hypotheses. Statistics of this form allow researchers to test hypotheses on the latent panel data structure even with inconsistent estimators for $\gamma_1, ..., \gamma_n$.[3] Hypotheses of this form can arise in many common economic models, such as testing for homogeneity of returns to education (Heckman and Vytlacil [1998]), teacher value added (Chetty et al. [2014], Kane et al. [2008]), or country production functions (Durlauf et al. [2001]). Many researchers estimate models without allowing for parameter heterogeneity, and the broadest class of models-where parameters vary both across individuals and time cannot be estimated without auxiliary assumptions.[4] When the underlying population features individual or group varying effects single coefficient models are misspecified, and Ordinary Least Squares may not recover the mean effect (Campello et al. [2018], Heckman and Vytlacil [1998]).

Historically these inconsistent parameters were treated as 'nuisance parameters' (Nickell [1981]) and

[1]Even here the test is indifferent to the source of knowledge on the group coefficient. This enables the test to be used with both fixed number of groups or growing number of groups, as long as the number of individuals per group is greater than 2.

[2]Many papers deal with the case where z_{it}' is symmetric across individuals, specifically $z_{it}' = (1, t)'$, see for example Hansen [2007] or Wooldridge [2005]. Under this framework the regressors are non-stochastic, a criteria we can relax under the condition that $\mathrm{E}[\mathbf{Z}_i \mid \mathbf{X}_i] = \mathrm{E}[\mathbf{Z}_i]$. Since γ_i is fixed for each individual, comapred to the Random Coefficients literature we do not need to condition on γ_i.

[3]For example, letting $L = 1$, and $z_{it} = 1$ gives rise to additive fixed effects model of the form $y_{it} = x_{it}'\beta + \gamma_i + \epsilon_{it}$. Under this framework a common hypothesis test would be $H_0 : \gamma_i = 0 \forall i$.

[4]For example, in the teacher value added literature, Chetty et al. [2014] assume that teacher value added in period t can be estimated from class mean test scores in the previous periods. This allows estimation of individual and time varying teacher value added.

swept away by the common within transformation, first differencing, or with a specific focus on trying to estimate average marginal effects. Instead this paper refocuses efforts to test hypotheses on these inconsistent estimates. Joint hypotheses of this form lead to a non-negligible number of regressors relative to the number of observations. Boos and Brownie [1995] study the behavior of ANOVA tests when the number of levels increases with the sample size. Akritas and Papadatos [2004] extend this framework to allow for heteroskedasticity, non-normality with fixed effects (Bathke [2004])). For regression models, Calhoun [2011] develops an F test for non-normal but homoskedastic data under this asymptotic framework. Equivalently, Anatolyev [2012] explores the behavior of the F, LM, and LR tests with homoskedastic Gaussian errors.

Creating valid tests in large n small T panels formalizes the source of persistent estimation error in model residuals that leads to biased estimates of population moments. Swamy [1970] develops tests for random coefficient models, Pesaran and Yamagata [2006] and Blomquist and Westerlund [2013] extends this to a more general class of large n, T panels. Orme and Yamagata [2006] and Orme and Yamagata [2014] develop F-tests over additive individual intercepts for both homoskedastic and GARCH error processes. Though not related in methodology, recent work has empowered researchers to explore latent panel structure in large n, T panels using augmented LASSO methods (Phillips and Hyungsik [1999], Lin and Ng [2012] and Su et al. [2016]). This allows for joint estimation of group assignment and group-specific fixed slopes. My structure allows for researchers with large n small T panels to test joint hypotheses over latent panel structure.

A large portion of this paper is dedicated to finding assumptions under which researchers are able to consistently estimate population moments. Hansen [2007] develops bias correction methods for FGLS estimation for panel data when errors follow a stationary $AR(p)$ process under both fixed and increasing T panels. Hausman and Kuersteiner [2008] show consistent estimation of a non-stationary variance covariance matrix for panels with additive unobserved heterogeneity. One of method I develop takes the differences in individual errors under a known grouping structure to estimate population moments. This technique has been used elsewhere, but most notably in matching estimators Hanson and Sunderam [2012]. Cattaneo et al. [2016] discuss general issues in estimation of general variance-covariance matrices without symmetry across individuals when the number of parameters is growing.

My paper differs from the existing literature in a few notable of ways. Previous papers have traditionally imposed the specific linear hypothesis that all individual additive fixed effects are zero are equal to the pooled OLS constant, where my paper allows for a more general set of plausible linear hypotheses. The resulting class of models that my test can be applied to is also more general, and common among the Correlated Random Coefficients literature (Arellano and Bonhomme [2012], Wooldridge [2005]). Finally, we explicitly develop assumptions under which even higher population moments can be consistently estimated even in the presence of persistent parameter estimation error. This enables us to extend existing literature to specifically allow for joint hypothesis tests over inconsistently estimated parameters even under the presence of serial correlation.

The paper proceeds in the following manner. Section 2 outlines notation and assumptions, and provides a proof for the asymptotic behavior of the infeasible GLS F and centered Wald tests. Section 3 develops a feasible tests under two different sets of assumptions. The first set of assumptions allows for a general set of linear restrictions with the errors following a stationary $AR(p)$ process. The second set of assumptions allows for an unconstrained error structure where under the null there is a known grouping structure of all individuals that explains parameter heterogeneity, and all groups have more than a single individual. Section 4 provides Monte Carlo evidence of the size and power of our proposed tests. Section 5 applies this to an application in teacher value added [to be added] . Section 6 concludes.

2 Assumptions & Notation

This section describes the estimated equation, what hypothesis researchers are interested in testing, and a baseline set of assumptions required to establish the asymptotic behavior for an *infeasible* generalized least squares F and centered Wald tests. For any $B \times T$ matrix b we define the usual within transformation to be, $\ddot{b} = (I_T - Z_i(Z_i'Z_i)^{-1}Z_i')b$, and for any $B \times D$ matrix A with individual terms denoted A_t, define $P_A = A(A'A)^{-1}A'$, such that $P_{A,ts} = A_t(A'A)^{-1}A_s'$ is 1×1. Stacking Equation (1) across time periods, each individual's data generating process follows the equation,

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\gamma_i + \boldsymbol{\epsilon}_i; i = 1, \dots, n$$

where $\mathbf{X}_i$ is a $T \times K$ vector of covariates whose coefficient is shared across all individuals, $\mathbf{Z}_i$ is a $T \times L$ vector of covariates whose coefficient varies across individuals. We assume that for each individual's errors process is symmetric, such that $\mathrm{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' \mid X_i \; Z_i \; \gamma_i) = \Omega$ for all individuals. Further stacking individual equations into a sample matrix representation we get,

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\boldsymbol{\gamma}_n + \boldsymbol{\epsilon} \tag{5}$$

where

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & \mathbf{Z}_n \end{bmatrix}$$

and $\boldsymbol{\gamma}_n = [\gamma_1 \; \gamma_2 \; \dots \; \gamma_n]$. The OLS estimators for β and $\boldsymbol{\gamma}_n$ are,

$$\hat{\beta} = (\ddot{\mathbf{X}}'\ddot{\mathbf{X}})^{-1}\ddot{\mathbf{X}}'\ddot{\mathbf{Y}}$$
$$\hat{\boldsymbol{\gamma}_n} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

The joint hypothesis of interest takes the form,

$$H_0: \; R_n[\beta \; \boldsymbol{\gamma}_n] = r_n \tag{6}$$

Where R_n is a $q_n \times (K + nL)$ matrix, and r_n is $q_n \times 1$. Usually the tests impose testing a single linear restriction for each fixed effect estimate in a sample, e.g.

$$R_n = \begin{bmatrix} 0_{k,n} : & I_n \end{bmatrix} \quad r_n = [0 \dots 0]'$$

With outside information (r.e. model restrictions, latent panel structure, etc), we can typically make r_n non-zero, or allow for a more general set of linear restrictions imposed by R_n. We are now interested in studying the asymptotic behavior of the following Wald test,

$$W_{n,OLS} = \frac{1}{q}[\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\epsilon}^2_{OLS,it,0} - \hat{\epsilon}^2_{OLS,it}] \tag{7}$$

This test is similar to the regular Wald test, using the residual sum of squares from an unconstrained OLS regression of equation (5), and after imposing (6). Feasible implementation requires estimation of the all cross product fourth and second moments. To carry out this analysis, I next impose a standard assumption of independence across individuals.

Assumption 2.1. $\{(\mathbf{X}_i\ \mathbf{Z}_i),\ \boldsymbol{\epsilon}_i\}$ *are iid across* i*.* $\mathrm{E}(\boldsymbol{\epsilon}_i \mid \mathbf{X}_i\ \mathbf{Z}_i\ \gamma_i) = 0$.

The imposed assumptions are standard in the fixed effects literature (see for example Wooldridge [2010] section 11.7.2). This assumption implies independence across individuals along with strict exogeneity of our regressors. This rules out dynamic panel settings. We do not restrict the relationships between (X_i, Z_i, γ_i), leaving us in a "fixed-effects" or "Correlated Random Coefficients" effects setting. Compared to the later, we are not interested in estimating the population mean coefficients, and instead want to do inference over our in-sample fixed effects coefficients. We next impose the traditional moment conditions,

Assumption 2.2. $\mathrm{Rank}(\sum_{t=1}^{T}\mathrm{E}[\ddot{x}_{it}\ddot{x}'_{it}]) = \mathrm{Rank}(\mathrm{E}[\ddot{\mathbf{X}}'_i\ddot{\mathbf{X}}_i]) = K,\quad \mathrm{Rank}(Z'_iZ_i) = L$ *with probability* 1 *and there exists a* Δ *such that* $\mathrm{E}[x^4_{ith}] \leq \Delta < \infty$*, and* $z^4_{ith} \leq \Delta < \infty$ *with probability 1.* $\mathrm{E}(|\epsilon_{it}|^{4+r}) < \infty$

The first part of this assumption requires that $J < T$. For the usual additive intercept this implies $T \geq 2$, and with a time trend implies $T \geq 3$. The rank condition on $\mathbf{Z}_i$ is due to the fixed T asymptotics. Under small T we need that matrices and moments hold with probability 1, rather than in expectation. The higher moments enable us to use variations of the law of large numbers and central limit theorems to estimate our parameters. The rank conditions in Assumption 2.2 implies that there can only be a single time-invariant regressor among $(\mathbf{X}_i,\ \mathbf{Z}_i)$. Finally we impose a general assumption on the error variance structure. We assume this is the same across all individuals in the sample, ruling out cluster or hierarchical effects.

Assumption 2.3.

$$\mathrm{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}'_i \mid \mathbf{X}_i\ \mathbf{Z}_i\ \gamma_i) = \Omega$$

$$\mathrm{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}' \mid \mathbf{X}\ \mathbf{Z}\ \boldsymbol{\gamma}_n) = \Sigma = I_n \otimes \Omega$$

With $\mathrm{E}(\epsilon_{it}\epsilon_{is} \mid \mathbf{X}_i\ \mathbf{Z}_i\ \gamma_i) = \sigma_{t,s}$

$$\mathrm{E}(\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv} \mid \mathbf{X}_i\ \mathbf{Z}_i\ \gamma_i) = \mu_{stuv}$$

This assumption has two parts. The first is that the variance-covariance matrix is symmetric across individuals in the sample. To prove the behavior of the infeasible test as asymptotically valid we do not need to impose further structure on the second moments, and can allow for it to be unconstrained. The second is dealing with the serial correlation present in the higher moments. Throughout the paper I will refer to $\mathrm{E}(\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv} \mid \mathbf{X}_i\ \mathbf{Z}_i\ \gamma_i)$ as the fourth cross-product moments. Under stronger parametric assumptions the exact behavior of this term is known[5], however throughout we allow this to be quite general.

Two immediate results come out of these assumptions. The first is the traditional asymptotic normality for the fixed effects estimator for β, and the second is that under the Frisch-Waugh-Lovell theorem that

[5] When $\boldsymbol{\epsilon}_i \sim N(0, V)$, Isserlis' Theorem states that $\mathrm{E}[\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv} \mid \mathbf{X}_i\ \mathbf{Z}_i\ \gamma_i) = \mathrm{E}[\epsilon_{it}\epsilon_{is} \mid \mathbf{X}_i\ \mathbf{Z}_i\gamma_i]\,\mathrm{E}[\epsilon_{iu}\epsilon_{iv} \mid \mathbf{X}_i\ \mathbf{Z}_i\gamma_i] + \mathrm{E}[\epsilon_{it}\epsilon_{iu} \mid \mathbf{X}_i\ \mathbf{Z}_i\gamma_i]\,\mathrm{E}[\epsilon_{is}\epsilon_{iv} \mid \mathbf{X}_i\ \mathbf{Z}_i\gamma_i] + \mathrm{E}[\epsilon_{it}\epsilon_{iv} \mid \mathbf{X}_i\ \mathbf{Z}_i\gamma_i]\,\mathrm{E}[\epsilon_{is}\epsilon_{iu} \mid \mathbf{X}_i\ \mathbf{Z}_i\gamma_i]$.

the residual sum of squares for model (5) are the same as within-transformed regression. These results hold under the GLS transformation as well.

Lemma 2.1. *Let $\hat{\beta}$ be the ordinary least squares estimate of β. Then if the assumptions hold, $\hat{\beta} - \beta \rightarrow^p 0$ and $\sqrt{N}(\hat{\beta} - \beta) \rightarrow^d N(0, \mathrm{E}(\ddot{X}_i'\ddot{X}_i)^{-1}\ddot{X}_i'\Omega\ddot{X}_i\,\mathrm{E}(\ddot{X}_i'\ddot{X}_i)^{-1})$. Moreover,*

$$\hat{\mathbf{Y}} - \mathbf{Y} = \hat{\ddot{\mathbf{Y}}} - \ddot{\mathbf{Y}} = (1 - \mathbf{P}_{\ddot{\mathbf{X}}_n})\ddot{\epsilon} = \mathbf{M}_{\ddot{\mathbf{X}}_n}\mathbf{M}_{\mathbf{Z}}\boldsymbol{\epsilon}$$

Using the above Lemma, we can also express the statistics (7) as the difference in the residual sum of squares, which generates quadratic forms,

$$W_{n,OLS} = \frac{1}{q}(R_n[\beta\ \boldsymbol{\gamma}_n] - r_n)'(\mathbf{W}'\hat{\Sigma}^{-1}\mathbf{W})(R_n[\beta\ \boldsymbol{\gamma}_n] - r_n)$$

$$=^{H_0} \frac{1}{q}\boldsymbol{\epsilon}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}R_n'[R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n']^{-1}R_n(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\epsilon}$$

Where $=^{H_0}$ denotes that the equality holds under the null. Traditionally proofs of this quadratic term converges to a χ^2_K distribution have used the fact that, $\sqrt{n}R(\hat{\theta} - \theta) \rightarrow^d N(0, R(Q^{-1})R')$ for some positive definite Q. Under this setup, the sum of squares of independent standard Normal variables is distributed as χ^2_K. Under my asymptotic framework only a share of the variables converge to a standard normal distribution, and remaining terms contain persistent squared estimation error. The main driver of our results is letting $q \rightarrow \infty$ as n increases. Imposing a growing joint hypothesis under the null acts as a central limit theorem. Thus the test statistics can be shown to converges to a Normal distribution under some regularity conditions. Finding a limiting distribution requires us to first understand the mean and norm of these quadratic terms.

Lemma 2.2. *Under Assumptions 2.1-2.3 are met. Define, $P_n^* = \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}R_n'[R_n(\mathbf{W}'\mathbf{W})^{-1}R_n']^{-1}R_n(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$. Then, for a balanced panel with n individuals observed T time periods,*

$$\begin{aligned}\mathrm{E}(W_{n,OLS} \mid \mathbf{X}\ \mathbf{Z}) &= \sum_{t,s}\sigma_{t,s}\frac{\sum_i P^*_{n,its}}{q} \\ &= tr(\mathrm{E}(\Sigma P_n^*))/q\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(W_{n,OLS} \mid \mathbf{X}\ \mathbf{Z}) = q^{-2}&\sum_{tsuv}\mu_{tsuv}\sum_i P^*_{n,ii,ts}P^*_{n,ii,uv} \\ &+ q^{-2}\sum_{tsuv}\sigma_{ts}\sigma_{uv}\sum_{i,j\neq i}P^*_{n,ii,ts}P^*_{n,jj,uv} \\ &+ q^{-2}\sum_{tsuv}\sigma_{tu}\sigma_{sv}\sum_{i,j\neq i}P^*_{n,ij,ts}P^*_{n,ij,uv} \\ &+ q^{-2}\sum_{tsuv}\sigma_{tv}\sigma_{su}\sum_{i,j\neq i}P^*_{n,ij,ts}P^*_{n,ji,uv}\end{aligned}$$

This test statistic requires explicit estimation of all cross-product second and fourth moments for the symmetric error process across individual. Note that the tests rely on the true population parameters $\sigma_{t,s}$

and μ_{tsuv} for all pairs of t, s, u, v. This differs from much of the traditional fixed effects literature where it is sufficient to rely on the moments from the within-transformed OLS residuals (Stock and Watson [2008]. This transformation does not recover the population term, and instead of subject to asymptotic bias. See Example (2.1).

Example 2.1. *Consider the panel data model with just varying intercepts.*

$$\begin{aligned} y_{it} &= \mu_i + \epsilon_{it} \\ \hat{\epsilon} &= y_{it} - \hat{\mu}_i \\ \hat{\mu}_i &= \frac{1}{T}\sum_{t=1}^{T} y_{it} \end{aligned}$$

Since we lack $T \to \infty$, the resulting estimators are unbiased but inconsistent.

$$\hat{\mu}_i = \mu_i + \frac{1}{T}\sum_{t=1}^{T} \epsilon_{it}$$

As a result,

$$\begin{aligned} \hat{\epsilon}_{it} &= y_{it} - \mu_i - \frac{1}{T}\sum_{t=1}^{T} \epsilon_{it} \\ &= \epsilon_{it} - \frac{1}{T}\sum_{t=1}^{T} \epsilon_{it} \end{aligned}$$

Under this fixed T framework, with "general" regularity conditions, the estimator

$$\begin{aligned} n^{-1}\sum_i \hat{\epsilon}_{it}\hat{\epsilon}_{is} &= n^{-1}\sum_i (\epsilon_{it} - \frac{1}{T}\sum_{t=1}^{T}\epsilon_{it})(\epsilon_{is} - \frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}) \\ &= n^{-1}\sum_i \left(\epsilon_{it}\epsilon_{is} - \frac{1}{T}\sum_{u=1}^{T}\epsilon_{iu}(\epsilon_{it}+\epsilon_{is}) + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\epsilon_{it}\epsilon_{is}\right) \\ &\to^p \sigma_{ts} - \frac{1}{T}\sum_u(\sigma_{ut}+\sigma_{us}) + \frac{1}{T^2}\sum_{u,v}\sigma_{uv} \end{aligned}$$

Creating a pivotal statistic, one that doesn't depend on the underlying test statistic, requires estimation of $\sigma_{t,s}$ and $\mu_{t,s,u,v}$ for all t, s, u, v. The next section develops asymptotic theory in order to estimate these terms under two different sets of assumptions. We then provide conditions where the conditional mean of the test statistic can be expressed in matrix notation.

3 Estimating Population Moments

Traditional estimators for σ_{ts} and μ_{tsuv} lead to biased moments under fixed effects due to first stage estimation error as shown in Example 2.1. In this section we develop ways of controlling for the impacts of this first stage estimation error to construct consistent estimators for $\sigma_{t,s}$ and $\mu_{t,s,u,v}$ for all t, s, u, v.

We note two plausible feasible strategies. The first assumes a known number of groups $g = 1, ..., G$, where each group has at least two individuals. Moreover, it requires that Z_i is the same across individuals. Under this structure, researchers are able to test the hypothesis that the individual coefficient is equal to the group coefficient, that is, $\gamma_i = \gamma_g \forall i \in g, \forall g$.

This fits most applications of fixed effects testing for latent panel data structure. Generally this takes the form of an individual specific intercept and time trend. Under these non-stochastic regressors testing group structure provides the most benefit to researchers in the majority of applied contexts. In practice this allows researchers to create a "cleaned" model that removes the group specific effect, and independence across individuals allows for identification of population moments. Benefits to this method include that Ω is allowed to be fully non-stationary across individuals.

The second method allows for a more general matrix of linear restrictions R, and Z_i that vary across individuals. Under this framework errors are assumed to be a mean zero, covariance stationary $AR(p)$ process. This parametric assumption provides closed form solutions for the cross-product fourth moments μ_{tsuv} that can be estimated in a two-stage fashion. These set of assumptions allow for z_{it} are allow to vary across individuals and time, and to have linear restrictions and r_n to be unique for each individual in the sample.

These two methods are not the only way to create feasible estimators, but they cover two major cases in joint hypothesis tests for panel data methods. The main downside of both is the requirement that even within a known group structure individuals remain independent of each other. Both methods rely pooling information across individuals in the sample. Unobserved clustering or interdependence renders this technique impossible.

3.1 Feasible Estimation with Known Group Structure

Most studies that impose fixed effects assume that fixed effects enter as an additive varying intercept for each individual, and occasionally a time trend. This structure is used routinely in both the correlated random coefficient (Wooldridge [2005]) and fixed effect (Hansen [2007]) literatures. Under this shared-regressor framework, for each n if researchers assume a known group structure, such that $\gamma_i = \gamma_g$ for all individuals i in group g, then we can construct a cleaned model,

$$y_{igt} - y_{jgt} =^{H_0} (x_{igt} - x_{jgt})'\beta + \epsilon_{it} - \epsilon_{jt}$$

Under this null, note that

$$y_{it} - x'_{it}\beta - y_{jt} - x'_{jt}\beta =^{H_0} \epsilon_{It} - \epsilon_{jt} \tag{8}$$

Since z_{it} is the same across individuals, and γ_i is the same for everyone in a particular group. This has several useful properties under independence across individuals.

$$
\begin{aligned}
\mathrm{E}(\epsilon_{it} - \epsilon_{jt} \mid \mathbf{W}) &= 0 \\
\mathrm{Var}(\epsilon_{it} - \epsilon_{jt} \mid \mathbf{W}) &= 2\sigma_{tt} \\
\mathrm{Cov}(\epsilon_{it} - \epsilon_{jt}, \epsilon_{is} - \epsilon_{js} \mid \mathbf{W}) &= 2\sigma_{ts} \\
\mathrm{E}((\epsilon_{it} - \epsilon_{jt})(\epsilon_{is} - \epsilon_{js})(\epsilon_{iu} - \epsilon_{ju})(\epsilon_{iv} - \epsilon_{jv}) \mid \mathbf{W}) &= (\mathrm{E}(\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv} \mid \mathbf{W}) + \sigma_{t,s}\sigma_{u,v} + \sigma_{t,u}\sigma_{s,v} + \sigma_{t,v}\sigma_{s,u})
\end{aligned}
$$

This differencing preserves the dynamic relationship in the errors and motivates simple to implement sample analogs. It also allows the researcher to make no assumption on the variance-covariance matrix, leaving it completely unconstrained as long as the error process is the same across individuals.[6] The assumptions for feasible estimation under this asymptotic framework can now be rewritten,

Assumption 3.1. *Let the data be generated by Equation (1). We then make the following assumptions,*

(a) $\mathrm{E}(\epsilon_{it} \mid \mathbf{W}) = 0$, $\mathrm{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' \mid \mathbf{W}) = \Omega \quad \forall i$, *is positive definite, where the* t, s *term is denoted* $\sigma_{t,s}$. *There exists an* $r > 0$ *such that* $\mathrm{E}(|\epsilon_{it}|^{4+r}) < \infty$ *for all* i, t.

(b) $\{X_i, , \boldsymbol{\epsilon}_i\}_{i=1}^n$ *are i.i.d.* Z_i *is non-stochastic and shared across everyone in the sample.*

(c) $\mathrm{Rank}(\sum_{t=1}^T \mathrm{E}[\ddot{x}_{it}\ddot{x}_{it}']) = \mathrm{Rank}(\mathrm{E}[\ddot{X}_i\ddot{X}_i']) = K$ *and* $\mathrm{Rank}(Z_i'Z_i) = L, \forall i$.

(d) There exists a Δ *such that* $\mathrm{E}[x_{ith}^4] \leq \Delta < \infty$.

(e) For all n, *there exist a number of groups* G_n, *where* g_i *denotes the group assignment of individual* i, *such that group size is denoted* n_g. *As* $n \to \infty$, $\sum_g n_g \to \infty$.

Most of these assumptions are standard. The first describes the errors as being mean zero, with the same variance-covariance matrix for all individuals, and sufficiently large enough moments for our application. The second imposes independence across individuals, and reiterates that Z_i is shared across everyone in the sample. The third and fourth are common assumptions in the fixed effect literature to ensure the appropriate law of large number limits exist. The last requirement allows for either researchers to impose a fixed number of groups, and assume $\min_g n_g \to \infty$, or to let $G_n \to \infty$ as long as $\min_g n_g \geq 2$.

Now we can think of the null as assuming the individual's DGP is the equation,

$$y_{igt} = x_{it}'\beta + z_{it}'\gamma_g + \epsilon_{it} \tag{9}$$

Now we are interesting in specific hypotheses such that,

$$\gamma_i = \gamma_g \quad \forall i \tag{10}$$

That is, for all individual, their individual varying regressors are actually the same as some deterministic grouping mechanism. This result motivates a subsample estimator. Within each group we can create $\lfloor n_g/2 \rfloor$ pairs. For $i_1, i_2 \in g$ we take

$$\hat{\zeta}_{i_1,i_2,t} = y_{i_1t} - x_{i_1t}'\hat{\beta} - (y_{i_2t} - x_{i_2t}'\beta) \tag{11}$$

Now, sub-sampling without replacement implies that $\hat{\zeta}$ are iid under assumptions 3.1.

[6] Our asymptotics allow for a growing number of groups. If one lets the number of groups be fixed, and that $\min_{g\in\{1,\ldots,G\}} \lim_{n\to\infty} n_g/n > 0$ we can allow for group specific covariance under the null.

Proposition 3.1. *Let assumptions 3.1 hold. For each g, t, pick two individuals without replacement. And define the term*

$$\hat{\zeta}_{g,i_1,i_2,t} = y_{i_1gt} - x'_{i_1t}\hat{\beta}^{FE} - y_{i_2gt} + x'_{i_2t}\hat{\beta}^{FE}$$

$$\frac{1}{2}\left(\sum_g \lfloor n_g/2 \rfloor\right)^{-1} \sum_{g=1}^{G} \sum_{i_1=1}^{\lfloor n_g/2 \rfloor} \hat{\zeta}_{g,i_1,i_2,t}\hat{\zeta}_{g,i_1,i_2,s} \to^p \sigma_{t,s} \tag{12}$$

$$\begin{aligned}\frac{1}{2}\left(\sum_g \lfloor n_g/2 \rfloor\right)^{-1} \sum_{g=1}^{G} \sum_{i_1=1}^{\lfloor n_g/2 \rfloor} &\hat{\zeta}_{i_1,i_2,t}\hat{\zeta}_{i_1,i_2,s}\hat{\zeta}_{i_1,i_2,u}\hat{\zeta}_{i_1,i_2,v} \\ &\to^p 2(\mathrm{E}(\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv} \mid \mathbf{Z}\ \mathbf{X}) + \sigma_{t,s}\sigma_{u,v} \\ &\qquad + \sigma_{t,u}\sigma_{s,v} + \sigma_{t,v}\sigma_{s,u})\end{aligned}$$

Moreover,

$$\frac{1}{2}\left(\sum_g \lfloor n_g/2 \rfloor\right)^{-1} \sum_{g=1}^{G} \sum_{i_1=1}^{\lfloor n_g/2 \rfloor} \hat{\zeta}_{g,i_1,i_2,t}\hat{\zeta}_{g,i_1,i_2,s} = \frac{1}{2}\left(\sum_g \lfloor n_g/2 \rfloor\right)^{-1} \epsilon'[I_{ts,\sum_g \sum_{n_g}} + K_{ts,1} + K_{ts,2}]\epsilon$$

Where $I_{ts,\sum_g \sum_{n_g}}$ is a diagonal matrix with 1's corresponding to observations that were picked in a particular sampling process. $K_{ts,1}$ is an upper diagonal matrix with 1's corresponding to each first matched pair, and $K_{ts,2}$ is a lower diagonal matrix with 1's corresponding to each second matched pair.u

Under Proposition 3.1 we can generate consistent estimators for Σ and $\mathrm{E}(\epsilon^*_{it}\epsilon^*_{is}\epsilon^*_{iu}\epsilon^*_{iv} \mid \mathbf{W})$.

Moreover, since the draws for the grouping is independent, we can draw sequences of independent pairs in order to further refine this estimator. This is similar to a bootstrap, but the initial procedure is done without replacement versus with. For our empirical section, we implement this estimator with 100 repeated independent pairing sequences.

3.2 General R, Varying Z_i

The above method meets with the main uses of fixed effects, and the resulting hypotheses researchers are interested in. However in many cases researchers may face time varying regressors within each individual. Under this condition the previous results do not hold, since

$$\hat{\zeta}_{i_1,i_2,t} = y_{i_1t} - y_{i_2t} - (x_{i_1t} - x_{i_2t})'\hat{\beta} - (z_{i_1t} - z_{i_2t})'\hat{\gamma}_g \tag{13}$$

Under the null $\hat{\gamma}_g \to^p \gamma_g$ if $\min_g n_g \to \infty$. This is met in fixed G asymptotics, but if the researcher further believes that the number of groups increases with the sample size will not hold. The following assumption clarifies the structural assumption on the error's data generating process.

Assumption 3.2. *$\epsilon_{it} = \epsilon^{-\prime}_{it}\alpha + \eta_{it}$ where η_{it} is strictly stationary in t for each i, $\mathrm{E}[\eta^2_{it}] = \sigma^2_\eta$, $\mathrm{E}(\eta^4_{it}) = \mu_{4,\eta}$, $\mathrm{E}[\eta_{it}\eta_{i\tau}] = 0$ for all $t \neq \tau$, $\mathrm{E}(\eta^{4+r}_{it}) < \infty$ for some $r > 0$, and the roots of $1 - \alpha_1\xi - \alpha_2\xi^2 - \cdots - \alpha_p\xi^p = 0$ have modulus greater than 1. We also have $T > p/2$ such that*

$$\mathrm{E}(\boldsymbol{\epsilon}_i \mid X_i\ Z_i) = 0,\ \mathrm{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}'_i \mid X_i\ Z_i) = \Gamma(\alpha)$$

and $MA(\infty)$ representation,

$$\epsilon_{it} = \sum_{d=0}^{\infty} \psi_d \eta_{i(t-d)}$$

This assumption states that the errors follow a mean zero $AR(p)$ process, where the innovation process has $4+r$ moments. This assumption effectively states that the individual error process is "block homogeneous" in time. The stationarity assumption guarantees the existence of an invertible $MA(\infty)$ representation of each individual's error process with absolutely summable coefficients. The following Lemma shows that the $AR(p)$'s covariances can be represented as a function of α and σ_η^2, and the cross-product fourth moments as a function of $\alpha,\ \sigma_\eta^2,\ \mu_{4,\eta}$.

Lemma 3.1. *Let Assumption 3.2 hold. Then, for any t, and $j,k,l \in \mathbb{Z}$ we have*

$$\mathrm{E}(\epsilon_{it}\epsilon_{i(t-j)}) = \sigma^2 \sum_{d=0}^{\infty} \psi_d \psi_{d+j}$$

and

$$\begin{aligned}
\mathrm{E}(\epsilon_{it}\epsilon_{i(t-j)}\epsilon_{i(t-k)}\epsilon_{i(t-l)}) = (\mu_{4,\eta} - 3\sigma_\eta^4) &\sum_{d=0}^{\infty} \psi_{d+|l|}\psi_{d+|l-j|}\psi_{d+|l-k|}\psi_d \\
&+ \sigma_\eta^4 \sum_{d=0}^{\infty}\sum_{c\neq d}^{\infty} \psi_{d+|k|}\psi_{c+|l-j|}\psi_c\psi_d + \sigma_\eta^4 \sum_{d=0}^{\infty}\sum_{c\neq d}^{\infty} \psi_{c+|j|}\psi_{d+|k-i|}\psi_c\psi_d \\
&+ \sigma_\eta^4 \sum_{d=0}^{\infty}\sum_{b\neq d}^{\infty} \psi_{b+|i|}\psi_b\psi_{d+|k-j|}\psi_d
\end{aligned}$$

This Lemma shows that creating a feasible versions of Theorem 4.1 requires creating a consistent estimator for α, σ_η^2, and $\mu_{4,\eta}$. The three estimators proceed in similar, but slightly different methods. In all three cases we generate a first stage estimator where we explicitly take into account for the role of first stage estimation error, due to not having large T asymptotics within each individual. Using this asymptotic expansion, we can correct for this bias to recover the population parameter of interest. We first generate a consistent estimator for α that requires no knowledge of the innovation distribution. This is the same estimator suggested in Hansen [2007], and our change in the design matrix leads to a mostly notational change in their proofs (see Technical Appendix). Define the OLS estimator for α to be,

$$\hat{\alpha} = \left(\frac{1}{n(T-p)} \sum_{i=1}^{n} \sum_{t=p+1}^{T} \hat{\epsilon}_{it}^{-} \hat{\epsilon}_{it}^{-\prime} \right)^{-1} \left(\frac{1}{n(T-p)} \sum_{i} \sum_{t=p+1}^{T} \hat{\epsilon}_{it}^{-} \hat{\epsilon}_{it} \right) \tag{14}$$

Where $\hat{\epsilon}_{it}^{-\prime} = [\hat{\epsilon}_{i(t-p)}, \ldots, \hat{\epsilon}_{i(t-1)}]$. Under these asymptotics, we can show $\hat{\alpha} \to^p \alpha_T(\alpha) = (\Gamma(\alpha) + \frac{1}{T-p}\Delta_\Gamma(\alpha))^{-1}(A(\alpha) + \frac{1}{T-p}\Delta_A(\alpha))$, with $A(\alpha) = [\gamma_1\ \ldots\ \gamma_p]'$, Δ_Γ is a $p \times p$ matrix with

$$[\Delta_\Gamma(\alpha)]_{k,j} = \mathrm{tr}(\Gamma(\alpha)\frac{1}{n}\sum_{i=1}^{n} Z_i(Z_i'Z_i)^{-1}Z_{i,-k}'Z_{i,-j}(Z_i'Z_i)^{-1}Z_i')$$

$$-\,\mathrm{tr}(\Gamma_{-k}(\alpha)\frac{1}{n}\sum_{i=1}^{n} Z_{i,-j}(Z_i'Z_i)^{-1}Z_i') - \mathrm{tr}(\Gamma_{-j}(\alpha)\frac{1}{n}\sum_{i=1}^{n} Z_{i,-k}(Z_i'Z_i)^{-1}Z_i')$$

and $\Delta_A(\alpha)$ is a $p \times 1$ matrix with

$$[\Delta_A(\alpha)]_{i,1} = \mathrm{tr}(\Gamma(\alpha)\frac{1}{n}\sum_{i=1}^{n} Z_i(Z_i'Z_i)^{-1}Z_{i,-k}'Z_{i,-0}(Z_i'Z_i)^{-1}Z_i')$$

$$-\,\mathrm{tr}(\Gamma_{-k}(\alpha)\frac{1}{n}\sum_{i=1}^{n} Z_{i,-0}(Z_i'Z_i)^{-1}Z_i') - \mathrm{tr}(\Gamma_{-0}(\alpha)\frac{1}{n}\sum_{i=1}^{n} Z_{i,-k}(Z_i'Z_i)^{-1}Z_i')$$

with $\Gamma_{-k}(\alpha) = \mathrm{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_{i,-k}' \mid \mathbf{W})$, $\boldsymbol{\epsilon}_{i,-k}' = [\boldsymbol{\epsilon}_{i(p+1-k)},\ \boldsymbol{\epsilon}_{i(p+2-k)}, \ldots,\ \boldsymbol{\epsilon}_{i(T-k)}]$, and $Z_{i,-k}$ defined equivalently. Without estimation error the OLS estimator would just be equal to $\Gamma(\alpha)^{-1}A(\alpha)$, and these remaining terms represent the role of first stage estimation error on the expected value of the estimator. Note that the resulting value $\alpha_T(\alpha)$, a function of the true underlying parameter. Therefore in sufficiently large samples this implies if $\alpha_T(\alpha)$ is invertible, we can generate a consistent estimator for α by taking the inverse of α_T^{-1} around $\hat{\alpha}$. That is,

$$\hat{\alpha}^\infty = \alpha_T^{-1}(\hat{\alpha}) = \alpha_T^{-1}(\alpha_T(\alpha)) = \alpha \tag{15}$$

Even more information about this estimator can be recovered.

Proposition 3.2. *Suppose $\alpha_T(\alpha)$ is continuously differentiable in α and that the derivative matrix of $\alpha_T(\alpha)$ in α, $H = D\alpha_T(\alpha)$, is invertible for all α such that Assumption 3.2 is satisfied, and $D\alpha_T(\alpha)$ is the derivative matrix of $\alpha_T(\alpha)$ in α . Then, $\hat{\alpha}^\infty - \alpha \to^p 0$ and*

$$\sqrt{N}(\hat{\alpha}^\infty - \alpha) \to^d \frac{1}{T-p}H^{-1}(\Gamma_p(\alpha) + \frac{1}{T-p}\Delta_{\gamma(\alpha)})^{-1}\chi$$

where $\chi = N(0, \Xi_T)$, and

$$\Xi_T = \mathrm{E}[\sum_{t_1=p+1}^{T}\sum_{t_2=p+1}^{T} \ddot{\epsilon}_{it_1}\ddot{\mu}_{it_1}\ddot{\mu}_{it_2}\ddot{\epsilon_{it_2}}]$$

and

$$\ddot{\mu_{it}} = \ddot{\epsilon}_{it} - \ddot{\epsilon}_{it}^{-\prime}\alpha_T(\alpha)$$

The above proposition shows shows the proposed method generates a consistent estimation of α without any distributional assumptions on the innovation process outside of the usual exclusion restrictions for OLS. Most importantly it requires that the polynomial lag degree of α is known precisely by the researcher.

We now show that under Assumption 3.2 we can estimate Σ and $\mathrm{E}(\epsilon_{it}^*\epsilon_{is}^*\epsilon_{iu}^*\epsilon_{iv}^* \mid \mathbf{W})$ through estimation of σ_η^2 and $\mu_{4,\eta}$. These estimators work by constructing composite errors that recover the innovation plus some nuisance terms we want to remove from the expected value in sufficiently large sample.

Note that

$$
\begin{aligned}
y_{it}-&x_{it}'\hat{\beta} - z_{it}\hat{\gamma}_i - \hat{\alpha}(y_{i(t-1)} - x_{i(t-1)}'\hat{\beta} - z_{i(t-1)}\hat{\gamma}_i) \\
&= \eta_{it} + (x_{it} - \alpha x_{i(t-1)})'(\beta - \hat{\beta}) + (z_{it} - \alpha z_{i(t-1)})'(\gamma_i - \hat{\gamma}_i)
\end{aligned}
$$

Under our assumptions, we know that the usual fixed effects estimator $\hat{\beta}^{FE}$ converges in probability to β. Therefore asymptotically we just have to control for the behavior of the right hand side of

$$
y_{it} - x_{it}'\hat{\beta} - z_{it}\hat{\gamma}_i - \hat{\alpha}(y_{i(t-1)} - x_{i(t-1)}'\hat{\beta} - z_{i(t-1)}\hat{\gamma}_i) \to^p \eta_{it} + (z_{it} - \alpha z_{i(t-1)})'(\gamma_i - \hat{\gamma}_i)
$$

From this large sample behavior, we can now expand the LHS to the second and fourth powers. This generates the following pair of estimators.

Proposition 3.3. *Let the assumptions hold. Define,* $P_{Z_i,st}^L = \hat{\phi}(L)P_{Z_i,st} = (\phi(L')z_{it})'(Z_i'Z_i)^{-1}z_{is})$*. Then,*

$$
\begin{aligned}
&(n(T-p))^{-1}\sum_i\sum_{t>p}\hat{\phi}(L)\hat{\epsilon}_{it}\hat{\phi}(L')\hat{\epsilon}_{it}/\omega_1 \to^p \sigma_\eta^2 \\
&\omega_1 = (1 - 2\frac{\sum_{i,t>p}\sum_s P_{Z_i,st}^L\hat{\psi}_{t-s}}{n(T-p)} + \frac{\sum_{i,t>p}\sum_{s,u}P_{Z_i,st}^L P_{Z_i,ut}^L\hat{\sigma}_{us}}{n(T-p)})
\end{aligned}
$$

$$
\begin{aligned}
&(n(T-p))^{-1}\sum_i\sum_{t>p}(\hat{\phi}(L)\hat{\epsilon}_{it}\hat{\phi}(L)\hat{\epsilon}_{it}\hat{\phi}(L)\hat{\epsilon}_{it}\hat{\phi}(L)\hat{\epsilon}_{it} - (\hat{\sigma}^2)^2\omega_1)/\omega_2 \to^p \mu_{4,\eta} \\
\omega_2 &= \frac{1}{n(T-p)}\sum_{i,t>p}(6\sum_{s,u}P_{Z_i,ts}^L P_{Z_i,tu}^L(\sum_{d=0}^{\infty}\hat{\psi}_d\hat{\psi}_{d+|s-u|} - \hat{\psi}_{u-t}\hat{\psi}_{s-t}) \\
&\quad - 4\sum_{s,u,v}P_{Z_i,ts}^L P_{Z_i,tu}^L P_{Z_i,tv}^L(\hat{\psi}_{s-t}\sum_{d=0}^{\infty}\hat{\psi}_d\hat{\psi}_{d+|u-v|} \\
&+ \hat{\psi}_{u-t}\sum_{d=0}^{\infty}\hat{\psi}_d\hat{\psi}_{|s-v|+d} + \hat{\psi}_{v-t}\sum_{d=0}^{\infty}\hat{\psi}_d\hat{\psi}_{|u-s|+d} - 3\hat{\psi}_{v-t}\hat{\psi}_{u-t}\hat{\psi}_{s-t}) \\
\omega_3 &= \frac{1}{n(T-p)}\sum_{i,t>p}(1 - 4\sum_s\hat{\psi}_{s-t}P_{Z_i,ts}^L + 6\sum_{s,u}P_{Z_i,ts}^L P_{Z_i,tu}^L\hat{\psi}_{u-t}\hat{\psi}_{|u-t|+|s-u|} \\
&\quad - 4\sum_{s,u,v}\hat{\psi}_{v-t}\hat{\psi}_{u-t}\hat{\psi}_{s-t}P_{Z_i,ts}^L P_{Z_i,tu}^L P_{Z_i,tv}^L + \sum_{s,u,v,w}\hat{\psi}_{v-t}\hat{\psi}_{u-t}\hat{\psi}_{s-t}\hat{\psi}_{w-t}P_{Z_i,ts}^L P_{Z_i,tu}^L P_{Z_i,tv}^L P_{Z_i,tw}^L)
\end{aligned}
$$

Moreover,

$$
(n(T-p))^{-1}\sum_i\sum_{t>p}\hat{\phi}(L)\hat{\epsilon}_{it}\hat{\phi}(L')\hat{\epsilon}_{it}/\omega_1 = (n(T-p))^{-1}\epsilon'(I_n - P_{\mathbf{W}})I_{nT}(I_n - P_{\mathbf{W}})\epsilon/\omega_1
$$

With $\hat{\psi}_{-j} = 0 \ \forall \ j > 0$*. Under the null, then we can use the residuals from the restricted model* $\hat{\epsilon}_{0,it}$ *and* $P_{C^{-1\prime}\mathbf{W}',ii,st}^L - P_{C^{-1\prime}\mathbf{W}',ii,st}^{L*}$ *replaces* $P_{Z_i,st}^L$

Proposition 3.3 implies that we can construct consistent estimators for σ_η^2 and $\mu_{\eta,4}$ using the asymptotic expansion around $y_{it} - x_{it}'\hat{\beta} - z_{it}\hat{\gamma}_i - \hat{\alpha}(y_{i(t-1)} - x_{i(t-1)}'\hat{\beta} - z_{i(t-1)}\hat{\gamma}_i)$, in this case the second moment of the underlying $AR(p)$ process. Under both the alternative and null, we know an analytical correction exists. Again, we are careful here to denote that under the null q can be increasing in a way to make the total number of regressors still be non-negligible to the sample size in the limit. Under this framework the correction will still be needed. As noted in Calhoun [2011], the null usually has fewer parameters and thus is the preferred method for estimating the sample moments.

However this does not generate appropriate variance estimates. The estimate for the conditional mean becomes

$$\begin{aligned}
\mathrm{tr}(\hat{\Sigma}P_n^*) &= \sum_{s,u} \hat{\sigma}_{s,u} \sum_{i,j} P_{ij,su} \\
&= \hat{\sigma} \sum_{s,u} \phi_d \phi_{d+|s-u|} \sum_{i,j} P_{ij,su} \\
&= \epsilon'[(I_{nt} - K_{nt}) * (n * t)^{-1} \sum_{s,u} \sum_{d=0}^{\infty} \hat{\phi}_d \hat{\phi}_{d+|s-u|} \sum_{i,j} P_{ij,su}]\epsilon
\end{aligned}$$

Note here that $\hat{p}hi_d$ is a function of ϵ for all d, thus $\mathrm{E}[\epsilon'[(I_{nt} - K_{nt}) * (n*t)^{-1} \sum_{s,u} \sum_{d=0}^{\infty} \hat{\phi}_d \hat{\phi}_{d+|s-u|} \sum_{i,j} P_{ij,su}]\epsilon \mid X] \neq \mathrm{E}[tr(\epsilon'\epsilon[(I_{nt} - K_{nt}) * (n*t)^{-1} \sum_{s,u} \sum_{d=0}^{\infty} \hat{\phi}_d \hat{\phi}_{d+|s-u|} \sum_{i,j} P_{ij,su} \mid X]$. To estimate this term consistently a bootstrap procedure is carried out.

The next section we develop asymptotic theory for the infeasible OLS and GLS estimators. Section 4 deals with developing assumptions under which the true population moments can be recovered for a feasible implementation of these tests.

4 A Feasible Joint Hypothesis Test

We develop two GLS hypothesis tests for when the number of restrictions grows with the sample size. While infeasible these tests motivate separate paths to generate feasible tests. In practice both are very similar to each other, one being a traditional F-test with GLS terms, and the second a centered Wald test. In both cases the arguments are largely identical, and that we can generate pivotal test statistics that converge to a standard Normal distribution.

Under this structure, we can reformulate our test statitics into

$$\hat{G}_{ar} =^{H_0} \frac{1}{q} \frac{\boldsymbol{\epsilon}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}R_n'[R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n']^{-1}R_n(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\epsilon}}{\epsilon'[(I_{nt} - K_{nt}) * (n * t)^{-1} \sum_{s,u} \sum_{d=0}^{\infty} \hat{\phi}_d \hat{\phi}_{d+|s-u|} \sum_{i,j} P_{ij,su}]\epsilon} - 1$$

$$\hat{G}_{n,gr} =^{H_0} \frac{1}{q} \frac{\boldsymbol{\epsilon}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}R_n'[R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n']^{-1}R_n(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\epsilon}}{\frac{1}{2}\left(\sum_g \lfloor n_g/2 \rfloor\right)^{-1} \epsilon'[I_{ts,\sum_g \sum_{n_g}} + K_{ts,1} + K_{ts,2}]\epsilon} - 1$$

Under this transformation these test statistics is now mean zero. This helps generate the following main result.

Theorem 4.1. *Let Assumptions 2.1-2.2 hold and* $\mathrm{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' \mid X_i\ Z_i) = \Sigma$ *is known. Then,*

$$\frac{q^{1/2}}{\nu_n}\left(\frac{\boldsymbol{\epsilon}^{*\prime} P^*_{C^{-1\prime}\mathbf{W}'}\boldsymbol{\epsilon}^*/q}{\boldsymbol{\epsilon}^{*\prime} P_{C^{-1\prime}\mathbf{W}'}\boldsymbol{\epsilon}^*/(n(T-L)-K)} - 1\right) \Rightarrow N(0,1) \tag{16}$$

Where

$$\nu_F^2 = q^{-1}\sum_{s,t,u,v} K_{stuv} - q^{-1}\left(\sum_i\left(\sum_{t,s} P^*_{ii,tt}P^*_{ii,ss} + 2\sum_{t,s} P^{*2}_{ii,ts}\right)\right) + 2(1+d_n)$$

and

$$\begin{aligned}
\mathbf{W} &= [\mathbf{X}\ \mathbf{Z}] \\
P^*_{C^{-1\prime}\mathbf{W}'} &= C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n'(R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}C^{-1} \\
P_{C^{-1\prime}\mathbf{W}'} &= C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}C^{-1} \\
d &= q/(n(T-L)-K) \\
P^* &= (P^*_{C^{-1\prime}\mathbf{W}'} + dP_{C^{-1\prime}\mathbf{W}'} - d) \\
K_{stuv} &= \mathrm{E}(\epsilon^*_{it}\epsilon^*_{is}\epsilon^*_{iu}\epsilon^*_{iv} \mid \mathbf{W})\sum_i P^*_{ii,ts}P^*_{ii,uv}
\end{aligned}$$

This theorem follows as ν_F^2 is the variance of the quadratic form that makes up the numerator. As in (Calhoun [2011]), the regular F or Wald tests will only be asymptotically valid if the leading two terms converge to zero.

5 Monte Carlo

This section compares various feasible implementations of our test statistics under conditions where both the grouping assumptions and the parametric assumptions of the error process are valid. I first provide Monte Carlo evidence that Lemma 7.8 is correct in Table 1. This is to check that the expression for the fourth moments is correct. Data is drawn from two sources, I either draw directly from an $AR(1)$ process with parameter a that has a burn in of 1000 periods and a draw of 10000. Columns calculated using this method are denoted ".ar.". The second method recreates a panel with n individuals with t time periods, and calculates all second and fourth cross product moments. Columns using this method are denoted '.indv.". This better represents how my sample is done. Asymptotically all estimates should be the same between the two methods, and ensure that the panel data generating function I wrote works as intended.

I take the mean difference between the sample innovation processes second and fourth moments to their true values in columns sigma.innov and mu.innov respectively. I then calculate all unique second and fourth cross product moments using both the '.ar.' and '.indv.' methods, and then calculate the total mean error in each column. Currently this is done with 1000 simulations for each row. For reference, the second moment of the Student's t-distribution (denoted rt) with 8 degrees of freedom is 8/6, and the fourth moment is 8. The centered exponential distribution (denoted re) has variance of 1, and fourth moment of 8.[7]

[7]Usually these are expressed in terms of excess kurtosis, but since the raw moments appear in my test statistic, I tend to

Overall, the raw estimates are very close. With low n appropriately targeting the full vector of fourth moments can still be subject to considerable bias. Even without any serial correlation, the exponential distribution with only 250 individuals over 1000 draws is still mean off by quite a bit. Equivalent and equal sized errors are found on the Student's t-distribution with 8 degrees of freedom with only 250 individuals in the sample.

For our actual test statistics, I estimate models of the form

$$y_{it} = \gamma_i + \epsilon_{it}$$

and test the null $\gamma_i = 0, \forall i$, and ϵ_{it} follows a covariance-stationary $AR(p)$ process. This allows us to sample $AR(p)$ structures of different length, coefficient signs, and innovation structures. This format also allows us to compare the test behavior when varying the sample size, n. Under this framework we assume all individuals are in the same group, and that group coefficient should be zero. The error process is covariance stationary, and thus Assumptions 3.1 and 3.2 are satisfied.

Table 2 provides evidence of the performance for our estimator for α, and comparisons between our two estimators for the covariance and cross product fourth moments. We let n be either small, with $n = 250$, or large, with $n = 1000$. We fix $t = 3$. This is done for a few reasons, at $t = 2$, the model is not identified. For each individual $\hat{\epsilon}_{i1} = -\hat{\epsilon}_{i2}$. At $t = 3$, there are 6 moments needed to be estimated in Σ, and 15 unique cross product fourth moments. When t increases to say, 5, the number of unique second moments becomes 15, and the cross product fourth moments rise to 70. Thus this keeps the estimation dimensionality as small as possible.

Our errors are drawn from an $AR(1)$ process, where $\alpha \in \{-.1, .5, .8\}$ and with innovations drawn from a student t distribution with 8 or 25 degrees of freedom, or a centered exponential distribution. These values of α show cover situations when α is negative, slow to zero, close to 1, and in between. The varying innovation processes cover situations where the underlying process has very fat tails, to approximately normal. For each value we report the mean and variance relative to the true parameter. As above, I reduce the dimensionality of the reported statistic for the second and cross product fourth moments by summing across all the values.

report those values instead.

Table 1: Verification of Lemma 7.8

n	t	dist	edf	a	sigma.innov	mu.innov	snd.ar.sum	snd.indv.sum	frth.ar.sum	frth.indv.sum
250	3	rt	8	0.0	0.0006666692	0.04320793	0.0048008307	-0.0089339876	0.1561580	-0.1157353
500	3	rt	8	0.0	0.0009460892	0.09668527	0.0015257145	-0.0003052253	0.2714698	-0.3562674
250	3	re	NA	0.0	-0.0010507308	-0.03681347	-0.0044253251	-0.0409681523	-0.1698573	-0.9245863
500	3	re	NA	0.0	-0.0014467846	-0.04773592	-0.0029843601	0.0140731601	-0.1600662	0.3637025
250	3	rt	8	0.5	0.0006030620	-0.01338317	-0.0032190118	0.0209597549	-0.1276685	1.2501550
500	3	rt	8	0.5	-0.0002940331	-0.07004805	-0.0003692341	0.0091787952	-0.4698096	0.1516936
250	3	NA	0.5	-0.0004597375	-0.04517766	-0.0039958334	-0.0075412543	-0.2701743	0.8794044	
500	3	NA	0.5	0.0002254799	0.02067541	0.0011892611	0.0159095453	0.1600038	0.3479310	

Table 2: Performance of Component Parts Across Monte Carlo Design[a]

	n	t	e	edf	a	a.mean	a.var	snd sub mean	snd sub var	snd analytic mean	snd analytic var	frth sub mean	frth sub var	frth analytic mean	frth analytic var
1	250.00	3.00	rt	8.00	-0.10	0.00	0.01	-0.01	0.03	-0.02	0.01	0.05	8.55	0.12	5.31
2	1000.00	3.00	rt	8.00	-0.10	-0.00	0.00	0.00	0.01	-0.02	0.00	-0.40	3.84	-0.45	2.87
3	250.00	3.00	re2		-0.10	0.00	0.02	-0.00	0.02	-0.00	0.01	-0.20	13.88	-0.11	7.76
4	1000.00	3.00	re2		-0.10	-0.00	0.01	0.00	0.01	-0.01	0.00	-0.04	4.81	-0.12	2.80
5	250.00	3.00	rt	25.00	-0.10	0.00	0.01	-0.01	0.02	-0.02	0.01	-0.06	0.91	-0.04	0.44
6	1000.00	3.00	rt	25.00	-0.10	-0.00	0.00	0.01	0.00	-0.01	0.00	0.01	0.23	-0.04	0.10
7	250.00	3.00	rt	8.00	0.50	0.00	0.02	0.05	0.05	0.64	0.61	-0.60	21.62	14.05	1415.64
8	1000.00	3.00	rt	8.00	0.50	0.00	0.00	-0.02	0.01	0.57	0.42	-0.65	5.68	6.61	97.77
9	250.00	3.00	re2		0.50	0.01	0.02	-0.02	0.04	0.43	0.31	0.85	27.25	12.35	1272.36
10	1000.00	3.00	re2		0.50	0.00	0.01	0.03	0.01	0.46	0.28	-0.08	6.26	4.91	59.18
11	250.00	3.00	rt	25.00	0.50	0.00	0.02	0.05	0.03	0.53	0.41	0.03	3.05	10.35	2900.41
12	1000.00	3.00	rt	25.00	0.50	0.00	0.00	-0.00	0.01	0.47	0.28	-0.10	0.79	4.37	37.70
13	250.00	3.00	rt	8.00	0.80	-0.02	0.01	-0.07	0.21	2.37	6.43	1.34	264.83	789.46	16977550.76
14	1000.00	3.00	rt	8.00	0.80	-0.00	0.00	0.21	0.10	2.77	8.03	0.41	69.40	184.26	692336.17
15	250.00	3.00	re2		0.80	-0.02	0.01	-0.12	0.16	1.72	3.55	-3.31	242.40	509.90	6856951.24
16	1000.00	3.00	re2		0.80	-0.00	0.00	0.02	0.04	1.93	3.96	2.45	58.72	142.55	474678.79
17	250.00	3.00	rt	25.00	0.80	-0.02	0.01	-0.04	0.14	1.98	4.48	-3.04	90.79	550.24	8561856.31
18	1000.00	3.00	rt	25.00	0.80	-0.00	0.00	0.04	0.03	2.13	4.76	0.31	20.51	112.65	95179.37

[a] 1000 Simulations for each row. For high values of α the estimator can fail to find under particular draws. As a result, the program will create a new Monte Carlo pull until a valid estimate is generated for a given data set.

Figure 1: Sum of Square Errors Second Moments for AR Estimator

Figure 2: Sum of Square Errors Second Moments for GR Estimator

OVerall, the sub-sampling method appears to work great in estimating the underlying population moments. Comparably, the analytical moments appear catastrophically off.

5.1 Test Statistics Behavior

In this section I compare a series of closely related tests to 7 using either the analytical moments calculations, or the group subsampling method. I simulate over 5000 draws, and use the very non-normal and censored exponential distribution.

Figures 1 and 2 compare the sum of square errors in estimating the second moments between the grouping subsample and analytical expressions. Under the chosen data generating process, the sub-sampling method does particularly well.

Figure 2 show sample distributions for the case when $n = 250, t = 3, a = 0.3$ and the innovations follow a centered exponential distribution. Surprisingly the Wald GLS does not behave accordingly at all, and seems to mirror the underlying innovation distribution directly. The feasible GLS implementations seem to exhibit general skewness, however in both the case of the FLGS grouping Wald, and the OLS analytical Wald, this skewness is in the opposite direction of the exponential distribution. The Wald OLS test statistic with the grouping estimator performs particularly well.

A final note is that the variance for the grouping estimator based statistics is significantly less than the analytical one. Again, this fact surprises me, since I figured the parametric assumptions would have given additional identification power to these parameters. The high level of underlying non-linearities in the estimators, and the multiple steps required to get the final output, probably contributes to these underlying efficiency losses.

6 Conclusion

This paper develops joint hypothesis tests over fixed effects for large n small T panels. We show the existence of feasible tests under two sets of assumptions. The first assumes a known grouping structure, where the regressors with individual varying coefficients are the same for each individual in each group. Researchers in this framework are interested in tests for the individual fixed effects being equal to the group fixed effect. The second set of assumptions allows for covariates with individual varying fixed effects, and a set of linear restrictions that can also vary at the individual level.

I show that the underlying methods require to create feasible versions of my tests behave well to be used in a general setting even with only moderate number of observations, and that in most cases that the estimator using a known grouping structure outperforms the version leveraging parametric assumptions on the error process.

Figure 3: Asymptotic Distribution of Wald OLS with AR Correction

Figure 4: Asymptotic Distribution of Wald OLS with GR Correction

Formal size and power tests of the tests still need to be carried out, and for the most recent iterations I am working on the asymptotic variance to normalize the tests to be standard normal. I also plan on adding in an application towards teacher value added subject to getting the data.

Bibliography

M. Akritas and N. Papadatos. heteroscedastic one-way anova and lack-of-fit tests. *Journal of the American Statistical Association*, 99(466):368–382, 2004.

S. Anatolyev. Inference in regression models with many predictors. *Jounral of Econometrics*, 170(2):368–382, 2012.

M. Arellano and S. Bonhomme. Identifying distribution characteristics in random coefficients panel data models. *The Review of Economic Studies*, 79:987–1020, 2012.

A. Bathke. The anova f test can still be used in some balanced designs with unequal variances and nonnormal data. *Journal of Statistical Planning and Inference*, 126(2):413–422, December 2004.

J Blomquist and J Westerlund. Testing slope homogeneity in large panels with serial correlation. *Economics Letters*, 121:374–378, 2013.

D. D. Boos and C. Brownie. ANOVA and rank tests when the number of treatments is large. *Statistics & Probability Letters*, 23:183–191, May 1995.

G Calhoun. Hypothesis testing in linear regression when k/n is large. *Journal of Econometrics*, 165(2): 163–174, 2011.

M Campello, Galvao AF, and Juhl T. Testing for slope heterogeneity bias in panel data models. *Journal of Business & Economic Statistics*, Forthcoming, 2018.

MD Cattaneo, M Jansson, and WK Newey. Inference in linear regression models with many covariates and heteroskedasticity. *working paper*, 2016.

R. Chetty, JN. Friedman, and JE. Rockoff. Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632, 2014.

P. de Jong. A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields*, 75:261–277, 1987.

SN Durlauf, A. Kourtellos, and A Minkin. The local solow growth model. *European Economic Review*, 45: 928–940, 2001.

CB. Hansen. Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects. *Journal of Econometrics*, 140(2):670–694, October 2007.

S G. Hanson and A Sunderam. The variance of non-parametric treatment effect estimators in the presence of clustering. *The Review of Economics and Statistics*, 94(4):1197–1201, 2012.

J Hausman and G. Kuersteiner. Difference in difference meets generalized least squares: Higher order properties of hypotheses tests. *Journal of Econometrics*, 144:371–391, 2008.

J. Heckman and E. Vytlacil. Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. *Journal of Human Resources*, 33(4):974–987, 1998.

TJ Kane, JE Rockoff, and DO Staiger. What does certification tell us about teacher effectiveness? evidence from new york city. *Economics of Education Review*, 27:615–631, 2008.

CC Lin and S. Ng. Estimation of panel data models with parameter heterogeneity when group membership is unknown. *Journal of Econometric Methods*, 1:42–55, 2012.

S Nickell. Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426, 1981.

C.D. Orme and T. Yamagata. The asymptotic distribution of of the f-test statistic for individual effects. *Journal of Econometrics*, 9:404–422, 2006.

C.D. Orme and T. Yamagata. A heteroskedasticity-robust f-test statistic for individual effects. *Economic Reviews*, 33:431–471, 2014.

M.H. Pesaran and T. Yamagata. Testing slope homogeneity in large panels. *Journal of Econometrics*, 142: 50–93, 2006.

P.C.B Phillips and M.R Hyungsik. Linear regression limit theory for nonstationary panel data. *Econometrica*, 67:1057–1111, 1999.

W. Stock and M.W. Watson. Heteroskedasticity-robust standard errors for fixed effects panel data regression. 76:155–174, 2008.

L. Su, Z. Shi, and P.C.B. Phillips. Identifying latent structures in panel data. *Econometrica*, 84:2215–2264, 2016.

PAVB Swamy. Efficient inference in random coefficient regression models. *Econometrica*, 38:311–323, 1970.

J.M. Wooldridge. Fixed-effects and related estimators for correlated-random coefficient and treatment-effect models. *The Review of Economics and Statistics*, 87:385–390, 2005.

JM. Wooldridge. *Econometric analysis of cross section and panel data*. Massachusetts Institute of Technology, 2010.

7 Technical Appendix

7.1 Fixed Effect Estimation Lemmas

Proof of Lemma 2.1. The proof is almost identical to Hansen (2004) Lemma 1.8.1 except that the Z_i's are allowed to vary across individuals.

We know $||AB|| \leq ||A||||B||$ by the Cauchy-Schwarz inequality, and $I - Z_i(Z_i'Z_i)^{-1}Z_i'$ is positive semi-definite for all i,

$$\begin{aligned}
\mathrm{E}\,||\ddot{X}_i{}'\ddot{X}_i|| &\leq \mathrm{E}\,||\ddot{X}_i||||\ddot{X}_i|| = \mathrm{E}\,||\ddot{X}_i||^2 \\
&= \mathrm{E}(\mathrm{tr}(X_i'X_i - X_i'Z_i(Z_i'Z_i)^{-1}Z_i'Z_i)) \\
&\leq \mathrm{tr}(\mathrm{E}(X_i'X_i)) = \mathrm{tr}(\sum_{t=1}^{T}\mathrm{E}(x_{it}x_{it}')) < \infty
\end{aligned}$$

Next, define $\ddot{\boldsymbol{\epsilon}}_i = \boldsymbol{\epsilon}_i - Z_i(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i$. Then, $||\ddot{X}_i{}'\ddot{\boldsymbol{\epsilon}}_i|| \leq (\mathrm{E}(||\ddot{X}_i||^2)\,\mathrm{E}(\ddot{\boldsymbol{\epsilon}}_i)^2)^{1/2}$, and the same arguments as before hold. By the weak law of large numbers,

$$\frac{1}{nT}\sum_i \ddot{X}_i{}'\ddot{X}_i \to^p \mathrm{E}(\ddot{X}_i{}'\ddot{X}_i)$$

and

$$\frac{1}{nT}\sum_i \ddot{X}_i{}'\ddot{\boldsymbol{\epsilon}}_i \to^p 0$$

Then $\hat{\beta} - \beta \to^p 0$.

For asymptotic normality, we know by 2.2 $\ddot{X}_i'\ddot{\boldsymbol{\epsilon}}_i$ is iid and has mean zero. Then, $\mathrm{E}(||\ddot{X}_i'\ddot{\boldsymbol{\epsilon}}_i\ddot{\boldsymbol{\epsilon}}_i'\ddot{X}_i||) \leq (2\,\mathrm{E}(||X_i||^4)\,\mathrm{E}(||\boldsymbol{\epsilon}_i||^4)^{1/2} < \infty$, by ..., the Cauchy-Schwarz inequality, and

$$\begin{aligned}
\mathrm{E}(||\ddot{X}_i') &= \mathrm{E}((\mathrm{tr}(X_i'X_i))^2 - 2\,\mathrm{tr}(X_i'X_i)\,\mathrm{tr}(X_i'Z_i(Z_i'Z_i)^{-1}Z_i'X_i) + (\mathrm{tr}(X_i'Z_i(Z_i'Z_i)^{-1}Z_i'X_i))^2) \\
&\leq \mathrm{E}(2\,\mathrm{tr}()^2) = 2\,\mathrm{E}(||X_i||^4)
\end{aligned}$$

where in the inequality follows from $X_i'X_i$, $X_i'Z_i(Z_i'Z_i)^{-1}Z_i'X_i$ and $I_T - Z_i(Z_i'Z_i)^{-1}Z_i$ positive semi-definite. It then follows from the Lindberg-Levy CLT that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\ddot{X}_i'\ddot{\boldsymbol{\epsilon}}_i \to^d N(0,\Omega)$ since $\mathrm{E}(\ddot{X}_i'\ddot{\boldsymbol{\epsilon}}_i\ddot{\boldsymbol{\epsilon}}_i'\ddot{X}_i) = \mathrm{E}(\ddot{X}_i'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'\ddot{X}_i) = \mathrm{E}(\ddot{X}_i'\Gamma(\alpha)\ddot{X}_i)$, from which $\sqrt{N}(\hat{\beta} - \beta) \to^d N(0, M^{-1}\Omega M^{-1})$ is obtained. □

Lemma 7.1. *Let the assumptions hold. Then define* $\hat{\alpha} = \left(\frac{1}{N}\sum_{i=1}^{n}\sum_{t=p+1}^{T}\tilde{e}_{it}^{-}\tilde{e}_{it}^{-\prime}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{n}\sum_{t=p+1}^{T}\tilde{e}_{it}^{-}\tilde{e}_{it}'\right)$ *be the least squares estimate of* α *using the lest squares residuals* $\tilde{e}_{it}$ *from estimating* β_1*. Then,*

$$\hat{\alpha} = \left(\frac{1}{N}\sum_{i=1}^{n}\sum_{t=p+1}^{T}\ddot{e}it^{-}\ddot{e}_{it}^{-\prime}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{n}\sum_{t=p+1}^{T}\ddot{e}_{it}^{-}\ddot{e}_{it}'\right) + o_p(N^{-1/2})$$

Proof of 7.1. Same as **?** □

Lemma 7.2. *Define* $\tilde{e}_{it}$ *to be the residual from least squares regression, i.e.* $\tilde{e}_{it} = y_{it} - x_{it}'\beta_1 - z_{it}'\beta_2^i = v_{it} - x_{it}'(\hat{\beta}_1 - \beta) - z_{it}'(\hat{\beta}_2^i - \beta_2)$*, where* $\hat{\beta}_1$ *and* $\hat{\beta}_2^i$ *are least squares estimates of* β_1 *and* β_2^i*. Then under our Assumptions,* $N^{-1}\sum_i\sum_{t=p+1}^{T}\tilde{e}_{it}^{-}\tilde{e}_{it}^{-\prime} = N^{-1}\sum_i\sum_{t=p+1}^{T}\ddot{e}_{it}^{-}\ddot{e}_{it}^{-\prime} + o_p(N^{-1/2})$ *and* $N^{-1}\sum_i\sum_{t=p+1}^{T}\tilde{e}_{it}^{-}\tilde{e}_{it}' = N^{-1}\sum_i\sum_{t=p+1}^{T}\ddot{e}_{it}^{-}\ddot{e}_{it}' + o_p(N^{-1/2})$

7.2 Supplementary Lemmas

These Lemma's are useful for establishing some of the results, but are not a direct interest of the paper.

Proof of Lemma 3.1. Since Y_t is a covariance stationary pth-order autoregressive process, there exists a $MA(\infty)$ representation of the form,

$$\epsilon_{it} = \alpha(L)\eta_{it} \tag{17}$$

where $\alpha(L) = (1 - \alpha_1 L - \cdots - \alpha_p L^p)^{-1}$, where the inverted $MA(\infty)$ representation generates the infinite sequence $\{\psi_j\}_{j=1}^{\infty}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Then, for any t, s, u, v, such that we can rewrite this as $s = t - i,\ u = t - j,\ v = t - k$ with $i, j, k \in \mathbb{Z}$. Then, we have

$$\begin{aligned}
\mathrm{E}(\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv}) &= \mathrm{E}(\epsilon_{it}\epsilon_{i(t-j)}\epsilon_{i(t-k)}\epsilon_{i(t-l)}) && (18)\\
&= \mathrm{E}(\sum_{a=0}^{\infty}\sum_{b=0}^{\infty}\sum_{c=0}^{\infty}\sum_{d=0}^{\infty} \psi_a\psi_b\psi_c\psi_d\eta_{i(t-a)}\eta_{i(t-j-b)}\eta_{i(t-k-c)}\eta_{i(t-l-d)}) && (19)\\
&= \sum_{a=0}^{\infty}\sum_{b=0}^{\infty}\sum_{c=0}^{\infty}\sum_{d=0}^{\infty} \psi_a\psi_b\psi_c\psi_d\,\mathrm{E}(\eta_{i(t-a)}\eta_{i(t-j-b)}\eta_{i(t-k-c)}\eta_{i(t-l-d)}) && (20)
\end{aligned}$$

Now we know that η_{it} is an iid process. Such that,

$$\mathrm{E}(\eta_{i(t-a)}\eta_{i(t-j-b)}\eta_{i(t-k-c)}\eta_{i(t-l-d)}) = \begin{cases} \sigma^4 & t-a = t-j-b \neq t-k-c = t-l-d \\ & t-a = t-k-c \neq t-j-b = t-l-d \\ & t-a = t-l-d \neq t-j-b = t-k-c \\ \mu_4 & t-a = t-j-b = t-k-c = t-l-d \\ 0 & \text{otherwise} \end{cases}$$

Carrying out expectations, we get,

$$\begin{aligned}
\mathrm{E}(\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv}) &= \sum_{a=0}^{\infty}\sum_{b=0}^{\infty}\sum_{c=0}^{\infty}\sum_{d=0}^{\infty} \psi_a\psi_b\psi_c\psi_d\,\mathrm{E}(\eta_{i(t-a)}\eta_{i(t-i-b)}\eta_{i(t-j-c)}\eta_{i(t-k-d)})\\
&= (\mu_4 - 3\sigma^4)\sum_{d=0}^{\infty} \psi_{d+|k|}\psi_{d+|k-i|}\psi_{d+|k-j|}\psi_d\\
&\quad + \sigma^4 \sum_{d=0}^{\infty}\sum_{c\neq d}^{\infty} \psi_{d+|k|}\psi_{c+|j-i|}\psi_c\psi_d\\
&\quad + \sigma^4 \sum_{d}\sum_{c} \psi_{c+|j|}\psi_{d+|k-i|}\psi_c\psi_d\\
&\quad + \sigma^4 \sum_{d}\sum_{b} \psi_{b+|i|}\psi_b\psi_{d+|k-j|}\psi_d
\end{aligned}$$

$\square$

Lemma 7.3. *If the Cholesky decomposition of* $\Gamma(\alpha) = LL'$*, and* $\Sigma = I_n \otimes \Gamma(\alpha)$ *then* $C = I_n \otimes L$

Proof of Lemma 7.3. We know that both I_n and $\Gamma(\alpha)$ are symmetric and positive definite matrices. Therefore, for reach we know there exists a lower triangular matrix such that

$$\begin{aligned} I_n &= I_n I_n' \\ \Gamma(\alpha) &= LL' \end{aligned}$$

This implies

$$\begin{aligned} I_n \otimes \Gamma(\alpha) &= (I_n I_n') \otimes LL' \\ &= (I_n \otimes L)(I_n \otimes L)' \end{aligned}$$

□

Lemma 7.4. *Let the assumptions hold. Then,*

$$\mathrm{E}(\epsilon^*_{it}\epsilon^*_{is}\epsilon^*_{iu}\epsilon^*_{iv} \mid \mathbf{W}) = \sum_{e,d,f,g} \mathrm{E}(\epsilon_{ie}\epsilon_{id}\epsilon_{if}\epsilon_{ig} \mid \mathbf{W}) C_{i(T-1)+t,e} C_{i(T-1)+s,d} C_{i(T-1)+u,f} C_{i(T-1)+v,g} \tag{21}$$

Proof of Lemma 7.4. since $\boldsymbol{\epsilon}^* = C^{-1}\boldsymbol{\epsilon}_i$, we know that $\epsilon^*_{it} = C^{-1}_{i(T-1)+t,\cdot}\boldsymbol{\epsilon}_i$, where $C^{-1}_{i(T-1)+t,\cdot}$ is the $1 \times nT$ row of C^{-1} . Then,

$$\mathrm{E}(\epsilon^*_{it}\epsilon^*_{is}\epsilon^*_{iu}\epsilon^*_{iv} \mid \mathbf{W})\,\mathrm{E}(C^{-1}_{i(T-1)+t,\cdot}\boldsymbol{\epsilon}_i C^{-1}_{i(T-1)+s,\cdot}\boldsymbol{\epsilon}_i C^{-1}_{i(T-1)+u,\cdot}\boldsymbol{\epsilon}_i C^{-1}_{i(T-1)+v,\cdot}\boldsymbol{\epsilon}_i \mid \mathbf{W})$$

Since we know the $AR(p)$ structure, C^{-1} is a known matrix of constants. Applying expectations,

$$\mathrm{E}(\epsilon^*_{it}\epsilon^*_{is}\epsilon^*_{iu}\epsilon^*_{iv} \mid \mathbf{W}) = \sum_{e,d,f,g} \mathrm{E}(\epsilon_{ie}\epsilon_{id}\epsilon_{if}\epsilon_{ig} \mid \mathbf{W}) C_{i(T-1)+t,e} C_{i(T-1)+s,d} C_{i(T-1)+u,f} C_{i(T-1)+v,g} \tag{22}$$

□

7.3 GLS F Test

These results are used to establish the asymptotic distribution of the centered GLS F-test.

Lemma 7.5. *Under Assumptions 3.2-2.2 are met. Then for any* n*,* T*,*

$$\mathrm{E}(F_n \mid \mathbf{W}) = 1$$

Proof of Proposition 7.5. I first prove results for the GLS Wald statistic, $W_{n,GLS}$. The numerator and denominator are independent as.

$$
\begin{aligned}
&C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n'(R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}'C^{-1\prime}(I_{nT}-C^{-1}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}C^{-1})\\
&\qquad = C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n'(R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}'C^{-1}\\
&\qquad - C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n'(R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}'C^{-1}C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}C^{-1}\\
&\qquad = C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n'(R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}'C^{-1}\\
&\qquad - C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n'(R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}'C^{-1}\\
&\qquad = 0
\end{aligned}
$$

We know $\mathrm{E}(\boldsymbol{\epsilon}^*) = 0$, by our GLS parameterization the resulting error structure is equivalent to the identity matrix, $\mathrm{E}(\boldsymbol{\epsilon}^*\boldsymbol{\epsilon}^{*\prime}) = C^{-1}(I_n \otimes \Gamma(\alpha))C^{-1\prime} = I_{nT}$, and by usual expectations of a quadratic form, we have

$$
\begin{aligned}
&\mathrm{E}(\boldsymbol{\epsilon}^{*\prime}C^{-1}\mathbf{W}(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n'(R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}'C^{-1\prime}\boldsymbol{\epsilon}^*/q_n \mid \mathbf{W})\\
&\quad = \mathrm{tr}(C^{-1}\mathbf{W}(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n'(R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}'C^{-1\prime})/q_n\\
&\quad = \mathrm{tr}((R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}'C^{-1\prime}C^{-1}\mathbf{W}(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')/q_n\\
&\quad = \mathrm{tr}((R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')/q_n = 1
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathrm{E}(\boldsymbol{\epsilon}^{*\prime}(I_{nT}-P_{C\mathbf{W}})\boldsymbol{\epsilon}^*/(n(T-L)-K) \mid \mathbf{W}) &= \mathrm{tr}(C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}C^{-1})/(n(T-L)-K)\\
&= \mathrm{tr}((C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}C^{-1})I_n \otimes \Gamma(\alpha))/(n(T-L)-K)\\
&= 1
\end{aligned}
$$

□

Lemma 7.6. *Let Assumptions 3.2-2.2 hold. Then, for a balanced panel with n individuals observed T time periods,*

$$
\mathrm{Var}(q^{-1/2}\boldsymbol{\epsilon}^{*\prime}P^*\boldsymbol{\epsilon}^* \mid \mathbf{W}) = q^{-1}\sum_{s,t,u,v} K_{stuv} - q^{-1}\left(\sum_i \left(\sum_{t,s} P^*_{ii,tt}P^*_{ii,ss} + 2\sum_{t,s} P^{*2}_{ii,ts}\right)\right) + 2(1+d_n)
$$

Proof of Lemma 7.6. The usual variance formula is,

$$
\mathrm{Var}(\boldsymbol{\epsilon}^{*\prime}P^*\boldsymbol{\epsilon}^* \mid \mathbf{W}) = \mathrm{E}\left((\boldsymbol{\epsilon}^{*\prime}A_n^*\boldsymbol{\epsilon}^*)^2 \mid \mathbf{W}\right) - \mathrm{E}\left(\boldsymbol{\epsilon}^*A_n^*\boldsymbol{\epsilon}^{*\prime} \mid \mathbf{W}\right)^2
$$

Since $d_n = q/(n(T-L)-K)$,

$$
\begin{aligned}
\mathrm{E}\left(\boldsymbol{\epsilon}^{*\prime}P^*\boldsymbol{\epsilon}^* \mid \mathbf{W}\right) &= \mathrm{tr}(P^*I_{nT})\\
&= \mathrm{tr}(P^*) = \mathrm{tr}(P^*_{C^{-1\prime}\mathbf{W}'} - d_n(I_{nT} - P_{C^{-1\prime}\mathbf{W}'}))\\
&= \mathrm{tr}(P^*_{C^{-1\prime}\mathbf{W}'}) - d_n\,\mathrm{tr}(I_{nT} - P_{C^{-1\prime}\mathbf{W}'}) = q_n - d_n(n(T-L)-K) = q_n - q_n = 0
\end{aligned}
$$

By symmetry of the errors across individuals,

$$\begin{aligned}
\mathrm{E}\left((\boldsymbol{\epsilon}^{*\prime} P^* \boldsymbol{\epsilon}^*)^2 \mid \mathbf{W}\right) &= \sum_i \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} P_{ii}^* \boldsymbol{\epsilon}_i^* \boldsymbol{\epsilon}_i^{*\prime} P_{ii}^* \boldsymbol{\epsilon}_i^* \mid \mathbf{W}) + 2\sum_{i=2}^{n}\sum_{j<i} \mathrm{E}(\boldsymbol{\epsilon}_i^* P_{ij}^* \boldsymbol{\epsilon}_j^{*\prime} \boldsymbol{\epsilon}_i^* P_{ij}^* \boldsymbol{\epsilon}_j^{*\prime} \mid \mathbf{W}) \\
&= \sum_{t,s,u,v} \mathrm{E}(\epsilon_{it}^* \epsilon_{is}^* \epsilon_{iu}^* \epsilon_{iv}^* \mid \mathbf{W}) \sum_i P_{ii,ts}^* P_{ii,uv}^* + \sum_{i,j<i} \left(\sum_{t,s} P_{ii,tt}^* P_{jj,ss}^* + 2\sum_{t,s} P_{ij,ts}^{2*} \right) \\
&= \sum_{t,s,u,v} \mathrm{E}(\epsilon_{it}^* \epsilon_{is}^* \epsilon_{iu}^* \epsilon_{iv}^* \mid \mathbf{W}) \sum_i P_{ii,ts}^* P_{ii,uv}^* - \left(\sum_i \left(\sum_{t,s} P_{ii,tt}^* P_{ii,ss}^* + 2\sum_{t,s} P_{ii,ts}^{*2} \right)\right) \\
&\quad + \mathrm{tr}(P^*)^2 + 2\,\mathrm{tr}(P^{*2}) \\
&= \sum_{t,s,u,v} \mathrm{E}(\epsilon_{it}^* \epsilon_{is}^* \epsilon_{iu}^* \epsilon_{iv}^* \mid \mathbf{W}) \sum_i P_{ii,ts}^* P_{ii,uv}^* - \left(\sum_i \left(\sum_{t,s} P_{ii,tt}^* P_{ii,ss}^* + 2\sum_{t,s} P_{ii,ts}^{*2} \right)\right) \\
&\quad + 2(1+d_n)
\end{aligned}$$

Dividing by q gives the result. $\square$

Lemma 7.7. *Let Assumptions 3.2-2.2 hold, then as $n, q \to \infty$, T fixed, then,*

$$\frac{\boldsymbol{\epsilon}^{*\prime} P_{C^{-1\prime}\mathbf{W}'}^* \boldsymbol{\epsilon}^* - q}{\sqrt{\mathrm{Var}(\boldsymbol{\epsilon}^{*\prime} P_{C^{-1\prime}\mathbf{W}'}^* \boldsymbol{\epsilon}^* \mid \mathbf{W})}} \Rightarrow N(0,1)$$

$$\boldsymbol{\epsilon}^{*\prime}(I_{nT} - P_{C\mathbf{W}})\boldsymbol{\epsilon}^*/(n(T-L)-K) \to^p 1$$

Proof of Lemma 7.7. The two proofs are nearly identical, so for brevity we show only the first one. By symmetry of the orthogonal projection matrix we have,

$$q^{-1/2}(\boldsymbol{\epsilon}^{*\prime} P_{C^{-1\prime}\mathbf{W}'}^* \boldsymbol{\epsilon}^* - q) = q^{-1/2}\sum_i (\boldsymbol{\epsilon}_i^{*\prime} P_{C^{-1\prime}\mathbf{W}',ii}^* \boldsymbol{\epsilon}_i^* - \sum_t P_{C^{-1\prime}\mathbf{W}',ii,tt}^*) + 2q^{-1/2}\sum_{i=2}^{n}\sum_{j<i-1} \boldsymbol{\epsilon}_i^{*\prime} P_{C^{-1\prime}\mathbf{W}',ij}^* \boldsymbol{\epsilon}_j^* \tag{23}$$

The first and second summation are mean zero processes, and by Assumption 2.1, the two are uncorrelated. Therefore, we show each part follows a central limit theorem separately. The first term is a mean zero process as,

$$q^{-1/2}\sum_i \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} P_{C^{-1\prime}\mathbf{W}',ii}^* \boldsymbol{\epsilon}_i^* - \sum_t P_{C^{-1\prime}\mathbf{W}',ii,tt}^*) = \mathrm{tr}(P_{C^{-1\prime}\mathbf{W}'}^*) - \mathrm{tr}(P_{C^{-1\prime}\mathbf{W}'}^*) = q - q = 0$$

Define the Frobenious Norm of a $nT \times nT$ matrix A to be $||A|| = \sqrt{\mathrm{tr}(AA')}$. Therefore, by applying Cauchy-Schwarz and definition of the Frobenious Norm we have,

$$\begin{aligned}
|\boldsymbol{\epsilon}_i^{*\prime} P_{C^{-1\prime}\mathbf{W}',ii}^* \boldsymbol{\epsilon}_i^*| &\leq ||\boldsymbol{\epsilon}_i^*|| ||P_{C^{-1\prime}\mathbf{W}',ii}^* \boldsymbol{\epsilon}_i^*|| \\
&\leq ||\boldsymbol{\epsilon}_i^*|| ||P_{C^{-1\prime}\mathbf{W}',ii}^*|| ||\boldsymbol{\epsilon}_i^*|| = ||\boldsymbol{\epsilon}_i^*||^2 ||P_{C^{-1\prime}\mathbf{W}',ii}^*||
\end{aligned}$$

We know $P_{C^{-1\prime}\mathbf{W}',ii}^*$ is a sub-matrix of $P_{C^{-1\prime}\mathbf{W}'}^*$, which is an orthogonal projection matrix, which by construction has Frobenious norm 1. The Frobenious norm is also weakly increasing in the number of cell blocks, such that $||P^*_{C^{-1\prime}\mathbf{W}',ii}|| \leq ||P_{C^{-1\prime}\mathbf{W}'}|| < 1$. This implies $|\boldsymbol{\epsilon}_i^{*\prime} P^*_{C^{-1\prime}\mathbf{W}',ii}\boldsymbol{\epsilon}_i^*| \leq ||\boldsymbol{\epsilon}_i^*||^2$. Then by Minkowski's Inequality and Assumption 3.2

$$\mathrm{E}\,|\boldsymbol{\epsilon}_i^{*\prime} P^*_{C^{-1\prime}\mathbf{W}',ii}\boldsymbol{\epsilon}_i^*|^{2+r} \leq [\sum_t \{\mathrm{E}(|\boldsymbol{\epsilon}_{it}^{*2}|)^{2+r}\}^{\frac{1}{2+r}}]^{2+r} < \infty$$

And the first term converges to a standard Normal distribution by the Lindeberg-Feller CLT. For the second term, we define Λ_n to be a block diagonal matrix with the $T \times T$ sub-matrices along $P^*_{C^{-1\prime}\mathbf{W}'}$'s diagonal. Then, let $\tilde{P}^*_{C^{-1\prime}\mathbf{W}'} = P^*_{C^{-1\prime}\mathbf{W}'} - \Lambda_n$. We define the following martingale difference series.

$$U(n) = \sigma(n)^{-1} \sum_{i \leq n} \sum_{j \leq n} \boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*$$

With $\sigma(n)^2 = \sum_{ij} \mathrm{E}((\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*)^2 \mid \mathbf{W})$. From de Jong [1987] Proposition 3.2 it suffices to show that the following terms are all lower order than $\sigma(n)^4$.

$$\begin{aligned}
DJ_I &= \sum_{1\leq i<j\leq n} \mathrm{E}[(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*)^4] \\
DJ_{II} &= \sum_{1\leq i<j<k\leq n} \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*)^2(\boldsymbol{\epsilon}_i^{*} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ik}\boldsymbol{\epsilon}_k^*)^2 + \mathrm{E}(\boldsymbol{\epsilon}_j^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ji}\boldsymbol{\epsilon}_i^*)^2(\boldsymbol{\epsilon}_j^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',jk}\boldsymbol{\epsilon}_k^*)^2 \\
&\quad + \mathrm{E}(\boldsymbol{\epsilon}_k^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ki}\boldsymbol{\epsilon}_i^*)^2(\boldsymbol{\epsilon}_k^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',kj}\boldsymbol{\epsilon}_j^*)^2 \\
DJ_{IV} &= \sum_{1\leq i<j<k<l\leq n} \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ik}\boldsymbol{\epsilon}_k^*)(\boldsymbol{\epsilon}_l^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',lj}\boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_l^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',lk}\boldsymbol{\epsilon}_k^*) \\
&\quad + \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',il}\boldsymbol{\epsilon}_l^*)(\boldsymbol{\epsilon}_k^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',kj}\boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_k^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',kl}\boldsymbol{\epsilon}_l^*) \\
&\quad + \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ik}\boldsymbol{\epsilon}_k^*)(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',il}\boldsymbol{\epsilon}_l^*)(\boldsymbol{\epsilon}_j^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',jk}\boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_j^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',jl}\boldsymbol{\epsilon}_k^*)
\end{aligned}$$

We now show that DJ_I and DJ_{II} are of lower order than $\sigma(n)^4$. Let us have some sequence $K(n)$ where as $n \to \infty, K(n) \to \infty$. We first show that these terms follow a Lindberg-style condition.

$$\begin{aligned}
&\max_{1\leq i<j\leq n} \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^* I_{|\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*| > K(n)\sum_{t,s} P^{*2}_{C^{-1\prime}\mathbf{W}',ij,ts}} \mid \mathbf{W})) \\
&\quad = \max_{1\leq i<j\leq n} \mathrm{E}(\mathrm{tr}(\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*\boldsymbol{\epsilon}_i^{*\prime}) I_{|\,\mathrm{tr}(\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*\boldsymbol{\epsilon}_i^{*\prime})| > K(n)\sum_{t,s} P^{*2}_{C^{-1\prime}\mathbf{W}',ij,ts}} \mid \mathbf{W})) \\
&\quad = \max_{1\leq i<j\leq n} \mathrm{tr}(\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\, \mathrm{E}(\boldsymbol{\epsilon}_j^*\boldsymbol{\epsilon}_i^{*\prime} I_{|\,\mathrm{tr}(\boldsymbol{\epsilon}_j^*\boldsymbol{\epsilon}_i^{*\prime})| > K(n)} \mid \mathbf{W})/(\sum_{t,s} P^{*2}_{C^{-1\prime}\mathbf{W}',ij,ts})) \to^p 0
\end{aligned}$$

Where convergence in probability comes from Assumption 3.2 such that $\mathrm{E}(\boldsymbol{\epsilon}_j^*\boldsymbol{\epsilon}_i^{*\prime} I_{|\,\mathrm{tr}(\boldsymbol{\epsilon}_j^*\boldsymbol{\epsilon}_i^{*\prime})| > K(n)} \mid \mathbf{W}) \to^p 0$. This condition ensures that a tail-truncated series converges to the full sequence in L^2.

$$
\begin{aligned}
\mathrm{Var}(\sum_{ij} \boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^* &- \boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^* I_{|\boldsymbol{\epsilon}_i^{*\prime}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*|<K(n)\sum_{t,s} P^{*2}_{C^{-1\prime}\mathbf{W}',ij,ts}} \mid \mathbf{W}) \\
&\leq \mathrm{Var}(\sum_{ij} \boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^* I_{|\boldsymbol{\epsilon}_i^{*\prime}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*|>K(n)\sum_{t,s} P^{*2}_{C^{-1\prime}\mathbf{W}',ij,ts}} \mid \mathbf{W}) \\
&\leq \sum_{ij} \sigma_{ij}^2 (\max_{1\leq i<j\leq n} \sigma_{ij}^{-2}\, \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^* I_{|\boldsymbol{\epsilon}_i^{*\prime}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*|>K(n)\sum_{t,s} P^{*2}_{C^{-1\prime}\mathbf{W}',ij,ts}} \mid \mathbf{W})) \\
&= o(\sigma(n))
\end{aligned}
$$

Since the truncated sequence converges to the full one in L^2, it suffices to show that $DJ'_I,\ DJ'_{II},\ DJ'_{III}$ are all lower order than $\sigma(n)^4$. From above we know

$$
\max_{1\leq i<j\leq n} \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^* I_{|\boldsymbol{\epsilon}_i^{*\prime}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*|>K(n)\sum_{t,s} P^{*}_{C^{-1\prime}\mathbf{W}',ij,ts}} \mid \mathbf{W})) \to^p 0
$$

Which in turn implies

$$
\mathrm{E}((\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*)^2 I_{|\boldsymbol{\epsilon}_i^{*\prime}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*|\leq K(n)\sum_{t,s} P^{*}_{C^{-1\prime}\mathbf{W}',ij,ts}} \mid \mathbf{W})) \leq K(n)(\sum_{t,s} P^*_{C^{-1\prime}\mathbf{W}',ij,ts})^2
$$

But this implies

$$
\mathrm{E}((\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*)^4 I_{|\boldsymbol{\epsilon}_i^{*\prime}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*|\leq K(n)\sum_{t,s} P^{*}_{C^{-1\prime}\mathbf{W}',ij,ts}} \mid \mathbf{W})) \leq K(n)^2(\sum_{t,s} P^{*2}_{C^{-1\prime}\mathbf{W}',ij,ts} hu)^4
$$

But this implies through the Cauchy-Schwarz theorem, that $DJ'_I,\ DJ'_{II}$ are of lower order than $\sigma(n)^4$. Finally, for DJ'_{IV} we know,

$$
\begin{gathered}
\mathrm{E}((\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ik}\boldsymbol{\epsilon}_k^*)(\boldsymbol{\epsilon}_l^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',lj}\boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_l^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',lk}\boldsymbol{\epsilon}_k^*) \mid \mathbf{W}) = \\
\sum_{t,s,u,v,w,x,y,z} \mathrm{E}(\epsilon^*_{it}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij,ts}\epsilon^*_{js}\epsilon^*_{iu}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ik,uv}\epsilon^*_{kv}\epsilon^*_{lw}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',lj,wx}\epsilon^*_{jx}\epsilon^*_{ly}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',lk,yz}\epsilon^*_{kz} \mid \mathbf{W})
\end{gathered}
$$

Where,

$\mathrm{E}(\epsilon^*_{it}\epsilon^*_{i\tau} \mid \mathbf{W}) = \begin{cases} 1 & t=\tau \\ 0 & \text{otherwise} \end{cases}$

Implies

$$
\begin{aligned}
\mathrm{E}((\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*)&(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ik}\boldsymbol{\epsilon}_k^*)(\boldsymbol{\epsilon}_l^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',lj}\boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_l^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',lk}\boldsymbol{\epsilon}_k^*) \mid \mathbf{W}) \\
&= \sum_{t,s,u,v} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij,ts}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ik,tu}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',lj,vs}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',lk,vu}
\end{aligned}
$$

But then again we know that the eigenvalues of $\lambda_{\max}(\tilde{P}^*_{C^{-1\prime}\mathbf{W}'}) \leq |\lambda_{\max}(P^*_{C^{-1\prime}\mathbf{W}'}) - \lambda_{\min}(\Lambda)| \leq 1$, and the matrix being idempotent implies that again these sums are bounded by similar logic as above. As a result we know these sums are bounded, and are lower order than $\sigma(n)^4$. Thus G'_{IV} is of lower order than $\sigma(n)^4$, completing the proof.

□

Proof of Theorem 4.1. From Lemma (7.6) we know that the numerator is conditionally mean zero, with variance η^2. Lemma (7.7) shows that the numerator is asymptotically normal, and that the denominator converges in probability to 1, completing the proof. □

7.4 Error Correction

Proof of Lemma 3.3. Define $\phi(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p$, where L is the lag operator. Under Assumptions 3.2-2.2,

$$\phi(L)\epsilon_{it} = \eta_{it} \tag{24}$$

Since $\phi(L)$ has roots outside the unit circle, there exists an invertible MA(∞) representation,

$$\epsilon_{it} = \sum_{j=0}^{\infty} \psi_j \eta_{i(t-j)} \tag{25}$$

Such that $\mathrm{E}(\epsilon_{it}\epsilon_{i(t-s)} \mid \mathbf{W}) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|s|}$. For $\boldsymbol{\epsilon}_i = [\epsilon_{i1} \ \ldots \ \epsilon_{iT}]'$, we know

$$\mathrm{E}(\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i' \mid \mathbf{W}) = \sigma^2 \Omega \tag{26}$$

Then, for $\ddot{\epsilon}_{it} = \epsilon_{it} - z_{it}'(Z_i'Z_i)^{-1}Z_i'\epsilon_i$,

$$\phi(L)\ddot{\epsilon}_{it} = \eta_{it} - \phi(L)z_{it}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i = \eta_{it} - \phi(L)z_{it}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i \tag{27}$$

Then,

$$\begin{aligned}
\mathrm{E}[\phi(L)\ddot{\epsilon}_{it}\phi(L')\ddot{\epsilon}_{it} \mid \mathbf{W}] &= \mathrm{E}((\eta_{it} - \phi(L)z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i)(\eta_{it} - \phi(L')z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i)' \mid \mathbf{W}) \\
&= \mathrm{E}(\eta_{it}^2 \mid \mathbf{W}) - 2\phi(L)z_{it}'(Z_i'Z_i)^{-1}Z_i' \,\mathrm{E}(\boldsymbol{\epsilon}_i \eta_{it} \mid \mathbf{W}) \\
&\quad + \mathrm{E}((\phi(L)z_{it})'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'Z_i(Z_i'Z_i)^{-1}\phi(L')z_{it} \mid \mathbf{W})
\end{aligned}$$

Now note that

$$\mathrm{E}(\epsilon_{is}\eta_{it} \mid \mathbf{W}) = \begin{cases} 0 & t > s \\ \psi_{|t-s|}\sigma^2 & \text{otherwise} \end{cases}$$

Such that we can define $\boldsymbol{\psi}_t = [1_{t\le 1}\psi_{|1-t|} \ldots 1_{t\le T}\psi_{|T-t|}]$. Thus, $\mathrm{E}(\boldsymbol{\epsilon}_i \eta_{it} \mid \mathbf{W}) = \sigma^2 \boldsymbol{\psi}_t'$. Thus,

$$
\begin{aligned}
\mathrm{E}[\phi(L)\ddot{\epsilon}_{it}\phi(L')\ddot{\epsilon}_{it} \mid \mathbf{W}] &= \sigma_\eta^2 - 2\sigma_\eta^2\phi(L)z_{iT}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\psi}_t' \\
&\quad + \phi(L)\,\mathrm{E}(\mathrm{tr}((\phi(L)z_{iT})'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'Z_i(Z_i'Z_i)^{-1}\phi(L)z_{iT} \mid \mathbf{W}) \\
&= \sigma_\eta^2 - 2\phi(L)\sigma_\eta^2 z_{iT}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\psi}_t' + \sigma^2\,\mathrm{tr}(\Omega Z_i(Z_i'Z_i)^{-1}\phi(L')z_{it}(\phi(L)z_{it})'(Z_i'Z_i)^{-1}Z_i') \\
&= \sigma_\eta^2(1 - 2\phi(L)z_{iT}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\psi}_t' + \mathrm{tr}(\Omega Z_i(Z_i'Z_i)^{-1}\phi(L')z_{iT}(\phi(L)z_{iT})'(Z_i'Z_i)^{-1}Z_i') \\
&< \infty
\end{aligned}
$$

Then the results follow follow from the Khintchine's LLN as,

- $\phi(L)\ddot{\epsilon}_{it}\phi(L')\ddot{\epsilon}_{it}$ are iid across individuals by Assumption 2.1
- $\mathrm{E} \mid \eta_{it}\eta_{it} \mid < \infty$
- $\mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\eta_{it}| \mid \mathbf{W}) < \left(\mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i|^2\mathbf{W})\,\mathrm{E}(|\eta_{it}^2| \mid \mathbf{W})\right)^{1/2}$
- $\mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'Z_i(Z_i'Z_i)^{-1}z_{it}| \mid \mathbf{W}) \le \left(\mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i|^2\mathbf{W})\,\mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i|^2\mathbf{W})\right)^{1/2}$

With

$$
\begin{aligned}
\mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i|^2\mathbf{W}) &= \mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'Z_i(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i|^2\mathbf{W}) \\
&\le ||z_{it}'(Z_i'Z_i)^{-1}Z_i'||^2\,\mathrm{E}(||z_{iT}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i||^2 \mid \mathbf{W}) \\
&= \mathrm{tr}(z_{it}(Z_i'Z_i)^{-1}z_{it}) \times \mathrm{E}(\mathrm{tr}(\boldsymbol{\epsilon}_i'Z_i(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i) \mid \mathbf{W}) \\
&\le L\,\mathrm{E}(\mathrm{tr}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i') \mid \mathbf{W}) = L\,\mathrm{E}(\sum_t^T \eta_{it} \mid \mathbf{W}) < \infty
\end{aligned}
$$

The final inequality follows as $\mathrm{tr}(z_{it}'(Z_i'Z_i)^{-1}z_{it}) < \mathrm{tr}(Z_i(Z_i'Z_i)^{-1}Z_i') = L$ and $\mathrm{tr}(\boldsymbol{\epsilon}_i'Z_i(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i) \le \mathrm{tr}(\boldsymbol{\epsilon}_i'\boldsymbol{\epsilon}_i)$.

For the second set of error corrections, Define $P_{ts} = \phi(L)z_{it}(Z_i'Z_i)^{-1}z_{is}'$

$$
\begin{aligned}
\mathrm{E}[(\phi(L)\hat{\epsilon}_{it})^4 \mid \mathbf{W}] &= \mathrm{E}((\eta_{it} - \phi(L)z_{it}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i)(\eta_{it} - \phi(L)z_{it}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i)'\times \\
&\qquad (\eta_{it} - \phi(L)z_{it}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i)(\eta_{it} - \phi(L)z_{it}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i)' \mid \mathbf{W}) \\
&= \mu_4 - 4\sum_s \mathrm{E}(\eta_{it}^3\epsilon_{is} \mid \mathbf{W})P_{ts} \\
&\quad + 6\sum_{s,u} \mathrm{E}(\eta_{it}^2\epsilon_{is}\epsilon_{iu}) \mid \mathbf{W})P_{ts}P_{tu} \\
&\quad - 4\sum_{s,u,v} \mathrm{E}(\eta_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv}) \mid \mathbf{W})P_{ts}P_{tu}P_{tv} \\
&\quad + \sum_{s,u,v,w} \mathrm{E}(\epsilon_{is}\epsilon_{iu}\epsilon_{iv}\epsilon_{iw}) \mid \mathbf{W})P_{ts}P_{tu}P_{tv}P_{tv}
\end{aligned}
$$

As above, we can now generate the expectations of each term. Let $\psi_{-j} = 0, \forall j > 0$. This simplifies the notiation quite a bit, and allows us to express the expectation as

$$
\begin{aligned}
\mathrm{E}(\eta_{it}^3 \epsilon_{is} \mid \mathbf{W}) &= \mathrm{E}(\eta_{it}^3 \sum_{d=0}^{\infty} \psi_d \eta_{i(s-d)} \mid \mathbf{W}) \\
&= \begin{cases} \psi_{|s-t|}\mu_4 & s \geq t \\ 0 & \text{otherwise} \end{cases} \\
\mathrm{E}(\eta_{it}^2 \epsilon_{is}\epsilon_{iu} \mid \mathbf{W}) &= \mathrm{E}(\eta_{it}^2 (\sum_{d=0}^{\infty} \psi_d \eta_{i(s-d)})(\sum_{d=0}^{\infty} \psi_d \eta_{i(u-d)}) \mid \mathbf{W}) \\
&= \begin{cases} \sigma^4(\sum_{d=0}^{\infty} \psi_d \psi_{d+|s-u|} - \psi_{u-t}\psi_{|u-t|+|s-u|}) & \\ \quad + \mu_4 \psi_{u-t}\psi_{|u-t|+|s-u|} & s \geq u \geq t \\ \sigma^4 \sum_{d=0}^{\infty} \psi_d \psi_{d+|s-u|} & s > t > u \\ 0 & \text{otherwise} \end{cases} \\
\mathrm{E}(\eta_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv} \mid \mathbf{W}) &= \mathrm{E}(\eta_{it}(\sum_{d=0}^{\infty} \psi_d \eta_{i(s-d)})(\sum_{d=0}^{\infty} \psi_d \eta_{i(u-d)})(\sum_{d=0}^{\infty} \psi_d \eta_{i(v-d)}) \mid \mathbf{W}) \\
&= \begin{cases} \mu_4 \psi_{v-t}\psi_{u-t}\psi_{s-t} & \\ \quad + \sigma^4(\psi_{s-t} \sum_{d=0}^{\infty} \psi_d \psi_{|u-v|+f} - \psi_{v-t}\psi_{u-t}\psi_{s-t}) & \\ \quad + \sigma^4(\psi_{u-t} \sum_{d=0}^{\infty} \psi_d \psi_{|s-v|+d} - \psi_{v-t}\psi_{u-t}\psi_{s-t}) & s \geq u \geq v \geq t \\ \quad + \sigma^4(\psi_{v-t} \sum_{d=0}^{\infty} \psi_d \psi_{|u-s|+d} - \psi_{v-t}\psi_{u-t}\psi_{s-t}) & \\ \psi_{s-t} \sum_d \psi_d \psi_{d+|u-v|} + \psi_{u-t} \sum_d \psi_d \psi_{d+|s-v|} & s > u > t > v \\ \psi_{s-t} \sum_d \psi_d \psi_{d+|u-v|} & s > t > u > v \\ 0 & \text{otherwise} \end{cases} \\
\mathrm{E}(\epsilon_{is}\epsilon_{iu}\epsilon_{iv}\epsilon_{iw} \mid \mathbf{W}) &= (\mu_4 - 3\sigma^4) \sum_{d=0}^{\infty} \psi_{d+|k|}\psi_{d+|k-i|}\psi_{d+|k-j|}\psi_d + \sigma^4 \sum_{d=0}^{\infty} \sum_{c \neq d}^{\infty} \psi_{d+|k|}\psi_{c+|j-i|}\psi_c \psi_d \\
&\quad + \sigma^4 \sum_d \sum_c \psi_{c+|j|}\psi_{d+|k-i|}\psi_c\psi_d + \sigma^4 \sum_d \sum_b \psi_{b+|i|}\psi_b \psi_{d+|k-j|}\psi_d \\
&= \mu_4 \pi_{1,suvw} + \sigma^4 \pi_{2,suvw}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}[(\phi(L)\hat{\epsilon}_{it})^4 \mid \mathbf{W}] = \mu_4 - 4\mu_4 \sum_{s\geq t} \psi_{s-t} P_{ts} \\
&+ 6\sum_{s,u}(\sigma^4 \sum_{d=0}^{u-t-1} \psi_d \psi_{d+|s-u|} + \mu_4 \psi_{u-t}\psi_{|u-t|+|s-u|}) P_{ts}P_{tu} \\
&- 4\sum_{s,u,v}(\mu_4 \psi_{v-t}\psi_{u-t}\psi_{s-t} + \sigma^4 \psi_{s-t}\sum_{d=0}^{\infty}\psi_d\psi_{|u-v|+f} \\
&+ \sigma^4 \psi_{u-t}\sum_{d=0}^{\infty}\psi_d \psi_{|s-v|+d} + \sigma^4\psi_{v-t}\sum_{d=0}^{\infty}\psi_d\psi_{|u-s|+d}) P_{ts}P_{tu}P_{tv} \\
&\sum_{s,u,v,w}(\mu_4 \pi_{1,suvw} + \sigma^4 \pi_{2,suvw}) P_{ts}P_{tu}P_{tv}P_{tw} \\
&= \mu_4(1 - 4\sum_{s}\psi_{s-t}P_{ts} + 6\sum_{s,u}\psi_{u-t}\psi_{|u-t|+|s-u|}P_{ts}P_{tu} \\
&- 4\sum_{s,u,v}\psi_{v-t}\psi_{u-t}\psi_{s-t}P_{ts}P_{tu}P_{tv} + \sum_{s,u,v,w}\pi_{1,suvw}) \\
&+ \sigma^4(6\sum_{s,u}\sum_{d=0}^{u-t-1}\psi_d\psi_{d+|s-u|}P_{ts}P_{tu} \\
&- 4\sum_{s,u,v}(\psi_{s-t}\sum_{d=0}^{\infty}\psi_d\psi_{|u-v|+f} + \psi_{u-t}\sum_{d=0}^{\infty}\psi_d\psi_{|s-v|+d} + \psi_{v-t}\sum_{d=0}^{\infty}\psi_d\psi_{|u-s|+d})P_{ts}P_{tu}P_{tv} \\
&+ \sum_{s,u,v,w}\pi_{2,suvw}P_{ts}P_{tu}P_{tv}P_{tw})
\end{aligned}
$$

Similarly, as above, we know the results follow follow from the Khintchine's LLN as,

- $\phi(L)\ddot{\epsilon}_{it}\phi(L')\ddot{\epsilon}_{it}\phi(L'')\ddot{\epsilon}_{it}\phi(L''')\ddot{\epsilon}_{it}$ are iid across individuals by Assumption 2.1
- $\mathrm{E} \mid \eta_{it}\eta_{it} \mid < \infty$
- $\mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\eta_{it}| \mid \mathbf{W}) < \big(\mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i|^2\mathbf{W})\,\mathrm{E}(|\eta_{it}^2| \mid \mathbf{W})\big)^{1/2}$
- $\mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'Z_i(Z_i'Z_i)^{-1}z_{it}| \mid \mathbf{W}) \leq \big(\mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i|^2\mathbf{W})\,\mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i|^2\mathbf{W})\big)^{1/2}$

Then, we can calculate the terms

$$
\begin{aligned}
\omega_2 = \frac{1}{n(T-p)}\sum_{i,t>p}(6\sum_{s,u}P_{ts}P_{tu}(\sum_{d=0}^{\infty}\psi_d\psi_{d+|s-u|} - \psi_{u-t}\psi_{|u-t|+|s-u|}) \\
&- 4\sum_{s,u,v}P_{ts}P_{tur}P_{tv}(\psi_{s-t}\sum_{d=0}^{\infty}\psi_d\psi_{|u-v|+f} + \psi_{u-t}\sum_{d=0}^{\infty}\psi_d\psi_{|s-v|+d} + \psi_{v-t}\sum_{d=0}^{\infty}\psi_d\psi_{|u-s|+d} - 3\psi_{v-t}\psi_{u-t}\psi_{s-t}) \\
\omega_3 = \frac{1}{n(T-p)}\sum_{i,t>p}(1 - 4\sum_{s}\psi_{s-t}P_{ts} + 6\sum_{s,u}P_{ts}P_{tu}\psi_{u-t}\psi_{|u-t|+|s-u|} \\
&- 4\sum_{s,u,v}\psi_{v-t}\psi_{u-t}\psi_{s-t}P_{ts}P_{tu}P_{tv} + \sum_{s,u,v,w}\pi_{1,suvw})
\end{aligned}
$$

As a result, we get the following system of equations,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1/\omega_1 \end{bmatrix} \begin{bmatrix} (n(T-p))^{-1} \sum_{i,t>p} \hat{\epsilon}^4 \\ (n(T-p))^{-1} \sum_{i,t>p} \hat{\epsilon}^2 \end{bmatrix} \to^p \begin{bmatrix} w_2 & w_3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_4 \\ \sigma^4 \end{bmatrix}$$

$$\begin{bmatrix} 1/\omega_2 & -\omega_3/\omega_2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1/\omega_1 \end{bmatrix} \begin{bmatrix} (n(T-p))^{-1} \sum_{i,t>p} \hat{\epsilon}^4 \\ (n(T-p))^{-1} \sum_{i,t>p} \hat{\epsilon}^2 \end{bmatrix} \to^p \begin{bmatrix} \mu_4 \\ \sigma^4 \end{bmatrix}$$

$$\begin{bmatrix} 1/\omega_2 & -\omega_3/(\omega_1\omega_2) \\ 0 & 1/\omega_1 \end{bmatrix} \begin{bmatrix} (n(T-p))^{-1} \sum_{i,t>p} \hat{\epsilon}^4 \\ (n(T-p))^{-1} \sum_{i,t>p} \hat{\epsilon}^2 \end{bmatrix} \to^p \begin{bmatrix} \mu_4 \\ \sigma^4 \end{bmatrix}$$

□

Lemma 7.8. *Under our Assumptions,*

$$N^{-1} \sum_i \frac{1}{T-p} \sum_{t=p+1}^{T} \ddot{e}_{it}^{-} \ddot{e}_{it}^{-\prime} = \mathrm{E}(\sum_{t=p+1}^{T} \ddot{\epsilon}_{it}^{-} \ddot{\epsilon}_{it}^{-\prime}) = (T-p)(\Gamma_p(\alpha) + \frac{1}{T-p} \Delta_\Gamma(\alpha))$$

$$N^{-1} \sum_i \sum_{t=p+1}^{T} \ddot{e}_{it}^{-} \ddot{e}_{it} = (T-p)(A(\alpha) + \frac{1}{T-p} A_\Gamma(\alpha))$$

where $\Gamma_p(\alpha) + \frac{1}{T-p}\Delta_\Gamma(\alpha)$ *is a* $p \times p$ *matrix with* k, j *element*

$$\mathrm{E}(\frac{1}{T-p} \sum_{t=p+1}^{T} \ddot{\epsilon}_{it}^{-} \ddot{\epsilon}_{it}^{-\prime}) = \gamma_{|i-j|}(\alpha) - \frac{1}{T-p} \operatorname{tr}(\Gamma_{-k}(\alpha) Z_{i,-j} (Z_i' Z_i)^{-1} Z_i)$$
$$- \frac{1}{T-p} \operatorname{tr}(\Gamma_{-j}(\alpha) Z_{i,-k} (Z_i' Z_i)^{-1} Z_i) + \frac{1}{T-p} \operatorname{tr}(\Gamma(\alpha) Z_i (Z_i' Z_i)^{-1} Z_{i,-k}' Z_{i,-j} (Z_i' Z_i)^{-1} Z_i')$$

$A(\alpha) + \frac{1}{T-p}\Delta_A(\alpha)$ *is a* $p \times 1$ *vector with ith element*

$$\mathrm{E}(\frac{1}{T-p} \sum_{t=p+1}^{T} \ddot{\epsilon}_{it}^{-} \ddot{\epsilon}_{it}') = \gamma_i(\alpha) - \frac{1}{T-p} \operatorname{tr}(\Gamma_{-k}(\alpha) Z_{i,-0} (Z_i' Z_i)^{-1} Z_i)$$
$$- \frac{1}{T-p} \operatorname{tr}(\Gamma_{-0}(\alpha) Z_{i,-k} (Z_i' Z_i)^{-1} Z_i) + \frac{1}{T-p} \operatorname{tr}(\Gamma(\alpha) Z_i (Z_i' Z_i)^{-1} Z_{i,-k}' Z_{i,-0} (Z_i' Z_i)^{-1} Z_i')$$

Lemma 7.9. *Let* $\hat{\alpha} = (\frac{1}{N}\sum_{i=1}^{N}\sum_{t=p+1}^{T} \hat{\epsilon}_{it}^{-}\hat{\epsilon}_{it}^{-\prime})^{-1}(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=p+1}^{T} \hat{\epsilon}_{it}^{-}\hat{\epsilon}_{it})$ *be the least squares estimate of* α *using the least squares residuals,* $\hat{\epsilon}_{it}$ *from estimating* β*. If the Assumptions hold,*

$$\hat{\alpha} = (\frac{1}{N} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \ddot{\epsilon}_{it}^{-} \ddot{\epsilon}_{it}^{-\prime})^{-1} (\frac{1}{N} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \ddot{\epsilon}_{it}^{-} \ddot{\epsilon}_{it}) + o_o(N^{-1/2})$$

Lemma 7.10. *Define* $\ddot{\mu}_{it} = \ddot{e}_{it} - \ddot{\epsilon}_{it}^{-\prime} \alpha_T(\alpha)$*. If the Assumptions hold,*

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\ddot{\epsilon}_{it}^{-}\ddot{\mu}_{it} \to^{D} N(0,\Xi)$$

where $\Xi = \mathrm{E}(\sum_{t_1=p+1}^{T}\sum_{t_2=p+1}^{T}\ddot{\epsilon}_{it_1}\ddot{\mu}_{it_1}\ddot{\mu}_{it_2}\ddot{\epsilon}_{it_2}]$

Proposition 7.1. *If the Assumptions hold then* $\hat{\alpha} \to^{p} \alpha_T(\alpha)$*, where*

$$\alpha_T(\alpha) = \mathrm{E}\left[\ddot{e}_{it}^{-}\ddot{e}_{it}^{-\prime}\right]\mathrm{E}\left[\sum_{t=p+1}^{T}\ddot{e}_{it}^{-}\ddot{e}_{it}\right] = (\Gamma_p(\alpha) + \frac{1}{T-p}\Delta_\Gamma(\alpha))^{-1}(A(\alpha) + \frac{1}{T-p}\alpha_A(\alpha))$$

Proof of Lemma 7.8. Again, the proof is almost identical to Hansen (2004) Lemma 1.8.3.

$$\begin{aligned}
\left[N^{-1}\sum_{i}\sum_{t=p+1}^{T}\ddot{e}_{it}^{-}\ddot{e}_{it}^{-\prime}\right]_{k,j} &= N^{-1}\sum_{i}\frac{1}{T-p}\sum_{t=p+1}^{T}\epsilon_{(t-k)}\epsilon_{i(t-j)} \\
&\quad + N^{-1}\sum_{i}\frac{1}{T-p}\sum_{t=p+1}^{T}\epsilon_{i(t-k)}z'_{i(t-j)}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i \\
&\quad + N^{-1}\sum_{i}\frac{1}{T-p}\sum_{t=p+1}^{T}\epsilon_{i(t-j)}z'_{i(t-k)}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i \\
&\quad + N^{-1}\sum_{i}\frac{1}{T-p}\sum_{t=p+1}^{T}z'_{i(t-k)}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'Z_i(Z_i'Z_i)^{-1}z_{i(t-j)} \\
&= N^{-1}\sum_{i}\frac{1}{T-p}\sum_{t=p+1}^{T}\epsilon_{(t-k)}\epsilon_{i(t-j)} \\
&\quad + N^{-1}\sum_{i}\frac{1}{T-p}Z'_{i,-j}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_{i,-k} \\
&\quad + N^{-1}\sum_{i}\frac{1}{T-p}Z'_{i,-k}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_{i,-j} \\
&\quad + N^{-1}\sum_{i}\frac{1}{T-p}Z'_{i,-k}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'Z_i(Z_i'Z_i)^{-1}Z_{i,-j} \\
&= \sigma_{k-j} + ((N(T-p))^{-1}\sum_{i}Z'_{i,-j}(Z_i'Z_i)^{-1}Z_i')\Gamma_{-k}(\alpha) \\
&\quad + ((N(T-p))^{-1}\sum_{i}Z'_{i,-k}(Z_i'Z_i)^{-1}Z_i')\Gamma_{-j}(\alpha) \\
&\quad + \mathrm{tr}(\Gamma(\alpha)(N(T-p))^{-1}\sum_{i}Z_i(Z_i'Z_i)^{-1}Z_{i,-j}Z'_{i,-k}(Z_i'Z_i)^{-1}Z_i')
\end{aligned}$$

Denote E^* to be the conditional expectation given X_i and Z_i. Then the results follow from the Khinchin Law of Large Numbers, repeated application of the triangle and Cauchy-Schwarz inequalities since,

- $\left[\sum_{t=p+1}^{T}\ddot{e}_{it}^{-}\ddot{e}_{it}^{-\prime}\right]_{k,j}$ is i.i.d across individuals under Assumption 2.1.
- $\mathrm{E}^*[\epsilon_{(t-k)}\epsilon_{i(t-j)}] < \infty$ for all k, j.

- $\mathrm{E}^*[\epsilon_{i(t-k)} z'_{i(t-j)}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i] \leq \Big(\mathrm{E}^*(|z'_{i(t-j)}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i|^2)\,\mathrm{E}^*(|\epsilon_{i(t-k)}|^2)\Big)^{1/2}$
- $\mathrm{E}^*(z'_{i(t-k)}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}'_iZ_i(Z'_iZ_i)^{-1}z_{i(t-j)}) \leq \Big(\mathrm{E}^*(|z'_{i(t-j)}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i|^2)\,\mathrm{E}^*(|z'_{i(t-j)}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i|^2)\Big)^{1/2}$
-

$$\begin{aligned}
\mathrm{E}^*(|z'_{i(t-j)}(Z'_iZ_i)^{-1}\boldsymbol{\epsilon}_i|^2) &= \mathrm{E}^*(|z'_{i(t-j)}(Z'_iZ_i)^{-1}(Z'_iZ_i)(Z'_iZ_i)^{-1}Z_i\boldsymbol{\epsilon}_i|^2)\\
&\leq ||z'_{i(t-j)}(Z'_iZ_i)^{-1}Z_i||^2\,\mathrm{E}^*(||Z_i(Z'_iZ_i)^{-1}Z_i\boldsymbol{\epsilon}_i||^2)\\
&= trace(z_{i(t-j)}(Z'_iZ_i)^{-1}z_{i(t-j)}) \times \mathrm{E}^*[trace(\boldsymbol{\epsilon}'_iZ_i(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i)]\\
&\leq L\,\mathrm{E}^*(\mathrm{tr}(\boldsymbol{\epsilon}'_i\boldsymbol{\epsilon}_i)) < \infty
\end{aligned}$$

Then, by the Law of Iterated Expectations, it follows that, $\mathrm{E}\left[\frac{1}{T-p}\sum_{t=p+1}^{T}\ddot{e}^{-}_{it}\ddot{e}^{-\prime}_{it}\right] < \infty$

□

Proof of Lemma 7.9. Same as Hansen (2004) Lemma 1.8.4. □

Proof of Lemma 7.10. Same as Hansen (2004) 1.8.5. □