Introduction to Econometrics II: Recitation 6

Seung-hun Lee*

February 28th, 2022

1 Time series regression

1.1 Building blocks for time series regressions

Since the dataset is no longer IID, there needs to be a separate building blocks for forming the theories for analyzing the time series regressions

- Ergodicity theorem: For this, we want a sequence $\{Y_t\}$ that is strictly stationary and ergodic. Then if $E[Y_t] < \infty$, we have

$$\frac{1}{T}\sum_{t=1}^{T} y_t \xrightarrow{p} E[Y_t]$$

- White noise: An error e_t satisfying $E[e_t] = 0, var[e_t] = \sigma^2$ and $\gamma(k) = 0 \;\forall k \neq 0$

- Linear projection: Y_t can be projected onto the set of the history of the value of Y's up until period $t-1$. Such projection is unique and so is the resulting projection error. Specifically, we can write $Y_t = \mathbb{P}_{t-1}(Y_t) + e_t$. The resulting e_t is serially uncorrelated (and thus WN) and $\mathbb{P}_{t-1}(Y_t)$ is covariance stationary provided that $\{Y_t\}$ is.

- CLT: Let $\{y_t\}$ come from a dependent data satisfying strict stationarity and ergodicity. Specifically, let $y_t = \mu + e_t$ with e_t being a white noise. then, as $T \to \infty$, the following holds,

$$\sqrt{T}\frac{\bar{y}-\mu}{\omega} \sim N(0,1)$$

where $\omega^2 = \sum_{j=-\infty}^{\infty}\gamma(j) = \gamma(0) + 2\sum_{j=1}^{\infty}\gamma(j)$

*Contact me at sl4436@columbia.edu if you spot any errors or have suggestions on improving this note.

1.2 Wold representation: AR(p), MA(q)

The idea behind the Wold representation is that if we have a process $\{Y_t\}$ that is covariance stationary, we can represent this process as an infinite linear function of the projection errors. Formally, Wold representation shows that Y_t can be written as a linear function of the white noise errors and a deterministic part

$$Y_t = \mu_t + \sum_{j=0}^{\infty} \psi_j e_{t-j}$$

If $\mu_t = \mu$, this is purely deterministic.

1.2.1 MA(q) process

A **moving average** process models Y_t as a weighted average of shocks (innovations) $e_t, e_{t-1}, ..., e_{t-q}$. q is the number of non-concurrent error terms e_{t-j} included in the model.

As a primer, we look at the MA(1) model.

$$y_t = \mu + e_t + \theta e_{t-1},\ e_t \sim (0, \sigma^2) \text{ is WN } \forall t$$

Here, e_t is not necessarily normal. What matters is that they are white noise. We can calculate the key moment conditions as follows

$$\begin{aligned}
E[y_t] &= \mu + E[e_t] + \theta E[e_{t-1}] = \mu \\
\gamma(0) &= var(y_t) \\
&= var(\mu + e_t + \theta e_{t-1}) \\
&= var(e_t + \theta e_{t-1}) \ (\because \text{ constants do not affect dispersion}) \\
&= var(e_t) + \theta^2 var(e_{t-1}) = (1+\theta^2)\sigma^2 \\
\gamma(1) &= cov(y_t, y_{t-1}) \\
&= cov(\mu + e_t + \theta e_{t-1}, \mu + e_{t-1} + \theta e_{t-2}) \\
&= \theta cov(e_{t-1}, e_{t-1}) = \theta\sigma^2 \\
\gamma(2) &= cov(y_t, y_{t-2}) \\
&= cov(\mu + e_t + \theta e_{t-1}, \mu + e_{t-2} + \theta e_{t-3}) = 0
\end{aligned}$$

The key is that $\gamma(k) = 0$, when $k \geq 2$. Thus, MA(1) models a shock with a very short-lasting impact.

The feature of a short-lasting impact can be generalized to MA(q) models. Write

$$y_t = \mu + \sum_{j=0}^{q} \theta_j e_{t-j},\ e_t \sim (0, \sigma^2) \text{ is WN } \forall t$$

We can obtain the following properties

$$\begin{aligned}
E[y_t] &= \mu + E\left[\sum_{j=0}^{q} \theta_j e_{t-j}\right] = \mu \\
\gamma(0) &= var\left(\sum_{j=0}^{q} \theta_j e_{t-j}\right) = \left(\sum_{j=0}^{q} \theta_j\right) \sigma^2 \\
\gamma(k) &= cov(y_t, y_{t-k}) \\
&= cov\left(\mu + \sum_{j=0}^{q} \theta_j e_{t-j}, \mu + \sum_{j=0}^{q} \theta_j e_{t-k-j}\right) \\
&= cov\left(\theta_k e_{t-k} + ... + \theta_q e_{t-q}, \theta_0 e_{t-k} + \theta_{q-k} e_{t-q}\right) = \left(\sum_{j=0}^{q-k} \theta_{j+k}\theta_j\right) \sigma^2 \\
\gamma(k) &= 0\ \forall k > q
\end{aligned}$$

Key takeaway is that $q+1$ periods and after, the impact of the shock vanishes. This is unlike the AR(p) process where the shock dies out slowly. MA(q) is a strictly stationary and ergodic process.

We can use MA(q) process to do a dynamic causation analysis of the impact of a shock at period t. This is possible since

$$y_t = \sum_{j=0}^{q} \theta_j e_{t-j}$$

and with the white noise error, we can find the impact of e_{t-j} with $\frac{\partial y_t}{\partial e_{t-j}} = \theta_j$. The interpretation of θ_j is the dynamic multiplier of e_{t-j} onto y_t.

1.2.2 AR(p) process

Like MA processes, AR(p) process has a Wold representation. We focus on the AR(1) process, which can be written as

$$y_t = \alpha y_{t-1} + e_t \iff y_t = \frac{e_t}{1 - \alpha L}$$

If we take $t \to \infty$, we can write y_t as a infinite sum of the white noise errors. Specifically

$$y_t = e_t + \alpha e_{t-1} + \alpha^2 e_{t-2} + ... = \sum_{j=0}^{\infty} \alpha^j e_{t-j}$$

So this is equivalent to setting $\theta_j = \alpha^j$ in the above setup. We effectively have MA(∞) process - a process that takes very long to die out. This is represented in these moment conditions.

$$\begin{aligned}
\gamma(0) &= cov\left(\sum_{j=0}^{\infty} \alpha^j e_{t-j}, \sum_{j=0}^{\infty} \alpha^j e_{t-j}\right) = \frac{\sigma^2}{1-\alpha^2} \\
\gamma(1) &= cov\left(\sum_{j=0}^{\infty} \alpha^j e_{t-j}, \sum_{j=1}^{\infty} \alpha^{j-1} e_{t-j}\right) = \frac{\alpha\sigma^2}{1-\alpha^2} \\
\gamma(2) &= cov\left(\sum_{j=0}^{\infty} \alpha^j e_{t-j}, \sum_{j=2}^{\infty} \alpha^{j-2} e_{t-j}\right) = \frac{\alpha^2\sigma^2}{1-\alpha^2} \\
\gamma(k) &= \frac{\alpha^k\sigma^2}{1-\alpha^2}
\end{aligned}$$

So as long as $|\alpha| < 1$, this process does approximate to 0. Depending on the value of α the process can die off quickly, last long, or oscillate around 0.

1.3 Estimating the time series model

We estimate the α coefficient in the following AR(1) model for ergodic stationary process $y_t = \alpha_1 y_{t-1} + e_t$ $(e_t \sim (0, \sigma^2), |\alpha| < 1)$. Define $X_t = [y_{t-1}]$. Then the OLS estimate of α_1 is $\hat{\alpha}_1 = (X_t' X_t)^{-1}(X_t' y_t)$. We want to know if this is consistently estimable. To check this, we write

$$\hat{\alpha}_1 - \alpha_1 = (X_t' X_t)^{-1}(X_t' e_t) \implies \text{plim}(\hat{\alpha}_1 - \alpha_1) = E[X_t' X_t]^{-1} E[X_t' e_t]$$

Since y_t is ergodic stationary, so is $X_t (= y_{t-1})$ and e_t and $X_t' e_t$. If the mean of y_t is finite, we have what it takes to invoke ergodicity theorem to show that the sample mean of $y_{t-1}e_t$ converges to 0. Since $E[y_{t-1}e_t] = 0$ and by ergodicity theorem, we have

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}e_t \xrightarrow{p} E[y_{t-1}e_t] = 0$$

So consistent estimate is possible.

However, this does not mean that the OLS estimate is unbiased. To see this, we need to re-evaluate the strict exogeneity condition used to justify unbiasedness, $E[e_t|X_1, ..., X_T] = 0$. This means that e_t is uncorrelated with all possible values of X. In time series setting, this may be unrealistic. Check that

$$\begin{aligned} E[e_t X_{t-1}] &= E[e_t y_{t-2}] = 0 \\ E[e_t X_t] &= E[e_t y_{t-1}] = 0 \\ E[e_t X_{t+1}] &= E[e_t y_t] = \sigma^2 (\neq 0) \end{aligned}$$

Thus, strict exogeneity may not be satisfied for most of the time. At most we can satisfy predetermined regressor conditions. Key difference with the consistency condition is that we only needed contemporary uncorrelatdness for a consistency whereas strict exogeneity (a stronger condition) is required for unbiasedness.

As for the asymptotic variances of this estimator, we need a different approach that takes into account a time dependent structure of the data. Let $Q = E[X'X]$. Then we can write

$$\sqrt{T}(\hat{\alpha} - \alpha) \to N(0, Q^{-1}\Omega Q^{-1})$$

where Ω comes from $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_t = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_t e_t \to N(0, \Omega)$. If we were to assume homoskedasticity in the sense that $\Omega = \sigma^2 Q = \sigma^2 \frac{\sigma^2}{1-\alpha^2}$. Then the asymptotic distribution is

$$\sqrt{T}(\hat{\alpha} - \alpha) \to N(0, 1 - \alpha^2)$$

and this is where $|\alpha| < 1$ condition is essential. If homoskedasticity is not guaranteed, the alternative we can take is the **HAC (heteroskedasticity and autocorrelation consistent) standard errors**. We first build this standard errors by calculating the variance.

$$\Omega = \sum_{j=-\infty}^{\infty} \Gamma_g(j) = \Gamma_g(0) + \sum_{j=1}^{\infty} \left(\Gamma_g(j) + \Gamma_g(j)' \right)$$

where $\Gamma_g(k) = E[g_t g_{t-k}]$. Since we are working on a stationary and ergodic process, this converges to 0 at some point. The question is to find out how much to sum over. Newey and West (1987) proposed this type of weighted sum of autocovariances as the estimator for Ω

$$\widehat{\Omega} = \widehat{\Gamma}_g(0) + \sum_{j=1}^{M} \left(1 - \frac{k}{M+1} \right) \left(\widehat{\Gamma}_g(j) + \widehat{\Gamma}_g(j)' \right)$$

where $M/T \to 0$. There are many versions of M, one of the most frequently used version being $M = 0.75 \times T^{1/3}$.

Similar structure holds generally. Take the following autoregressive distributed lag model (ADL(p,q)) model

$$y_t = \sum_{j=1}^{p} \alpha_j y_{t-j} + \sum_{k=1}^{q} \delta_k z_{t-k} + e_t$$

What we can do is to define $X_t = [y_{t-1}, .., y_{t-p}, z_{t-1}, ..., z_{t-q}]$ and run a similar process.

Another approach to regressing the AR(1) model is to use a GLS type of approach. Recall that even if the spherical variances assumptions for e is broken down, if we find P matrix such that $D^{-1} = PP'$, where $D = E[ee'|X]$, spherical variances can be obtained by multiplying P' to the original DGP. In heteroskedastic settings, we just had to reweigh the observation by the inverse of the variance. There are more complications in a serially correlated errors setup but still workable.

If the error term is AR(1), $e_t = \rho e_{t-1} + u_t$, the $D = E[ee'|X]$ and P (such that $PP' = D^{-1}$) matrix are defined as (all of them are T-dimensional square matrices)

$$D = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & .... \\ \gamma(1) & \gamma(0) & \gamma(1) & .... \\ .... & ... & ... & .... \\ \gamma(T-1) & ... & ... & \gamma(0) \end{pmatrix} = \frac{\gamma(0)}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & .... \\ \rho & 1 & \rho & .... \\ .... & ... & ... & .... \\ \rho^{T-1} & ... & ... & 1 \end{pmatrix}, \ P' = \frac{1}{\sigma_u} \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & ... \\ -\rho & 1 & 0 & .... \\ .... & ... & ... & .... \\ 0 & ... & -\rho & 1 \end{pmatrix}$$

using the relation that $\gamma(j) = \frac{\rho^j \gamma(0)}{1-\rho^2}$. Effectively, we transform (or quasi-difference) the DGP $y_t = x_t'\beta + e_t$ by

$$y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})'\beta + \underbrace{e_t - \rho e_{t-1}}_{u_t}$$

whose error u_t is spherical. For a consistent estimation, we need to check whether a stronger condition (since it involves x_{t-1}, not just x_t) $E[(x_t - \rho x_{t-1})u_t] = 0$ holds.

Lastly, how many lags do we need? This is where the **information criterion** comes in. It gives the number that sums the measure of fitness of the model and the penalty for a complex model. They are calculated as

$$IC(k) = \log \hat{\sigma}^2 + k\frac{C_T}{T}$$

where there are k total regressors. $\log \hat{\sigma}^2$ captures the (lack of) fitness[1] - higher values indicate loose model in terms of how much of the variation is explained by the regressors. $k\frac{C_T}{T}$ captures the penalty for making the model too large - it increases with k. The number of re-

[1]Note that $\hat{\sigma}^2$ is a function of k, in that more regressors usually lead to decrease in residuals.

gressors (lags) can be selected by the k values that makes $IC(k)$ lowest. There are two types of information criterion

- Akaike information criterion: $C_t = 2$
- Bayesian information criterion: $C_T = \log T$

2 Panel regression

To motivate the discussion going forward, consider the following setting. Let y and $x = (x_1, x_2, ..., x_k)$ be the observable factors. Denote α as an unobservable random variable that is incorporated into the data generating process additively. Then, we can write $E[y|x, \alpha]$ as

$$E[y|x, \alpha] = x'\beta + \alpha$$

We are interested in estimating β. To see whether we can find a consistent estimator for β, we need to see how α is correlated with x. If α is independent from x, then it is not different from the idiosyncratic error and β is consistently estimable. However, in most cases, α could be correlated with x. If this is the case, then we cannot find a consistent estimator for β unless we find a perfect proxy for α. However, if α is fixed across time for an individual and we have access to panel data, we can address this issue. Most discussion of panel data in class will focus on α representing an unobservable trait inherent for individuals.

2.1 Setting up a (static) panel model

In a panel data setting, the data now has two dimensions - dimensions across different unit of observation i and across time t. Our data generating process is

$$y_{it} = x_{it}'\beta + v_{it} \iff y_i = X_i\beta + v_i$$

where $x_{it} \in \mathbb{R}^k, y_{it} \in \mathbb{R}$, and v_{it} is a 'composite' error term. We can stack each individual-time cell observation for each individual to get $y_i \in \mathbb{R}^T, X_i \in \mathbb{R}^{T\times k}, v_i \in \mathbb{R}^T$. If you'd like, you can stack all observations to get a matrix notation $y \in \mathbb{R}^{nT\times 1}, X \in \mathbb{R}^{nT\times k}$, and $v \in \mathbb{R}^{nT\times 1}$ and write

$$y = X\beta + e$$

if you can keep in mind the different dimensional restrictions on each variables.

Depending on the setup, there are three types of composite error in this case

- One-way FE: $v_{it} = c_i + e_{it}$
- Two-way FE (TWFE): $v_{it} = c_i + \delta_t + e_{it}$
- Interacted FE (IFE): $v_{it} = \lambda_i f_t + e_{it}$

We limit ourself to a one-way fixed effects setting. Which requires us to set assumptions on c_i and e_{it}. The assumption on c_i determines the method of panel regression we will use

Definition 2.1 (Pooled OLS, random effects, fixed effects)**.** There are three types of estimators we study in (static) panel regression, determined by assumptions on $E[c_i|X_i]$

- Is c_i constant $\forall i$? Then **pooled OLS (POLS) estimator** is consistent and efficient
- Is c_i latent (in that it is unobserved and possibly different for each i) but uncorrelated with X_i? Then POLS is consistent but inefficient. We use **random effects (RE) GLS estimator** here
- Is c_i latent and possibly correlated with X_i? Then POLS and RE are inconsistent. We use **fixed effects (FE) estimator** - which are within estimation (WE), least squares dummy variables (LSDV), and in case we have $T = 2$, first-differencing (FD)

The assumptions on e_{it} the degree of exogeneity we are willing to assume. This has some bearings on the consistency of some estimators.

Assumption 2.1 (Strict exogeneity, sequential exogeneity)**.** *The following are the different assumptions we set on e_{it} with respect to X_i*

- ***Strict exogeneity****: $E[e_{it}|x_{i1}, ..., x_{iT}, c_i] = 0$. Or that the past, current, or future values of x_{it} cannot predict e_{it}*
  - *A weaker version: $E[x_{is}e_{it}] = 0$ for $s = 1, ..., T$. This implies that e_{it} cannot be predicted by x_{is}*
- ***Sequential exogeneity****: $E[e_{it}|x_{i1}, ..., x_{it}, c_i] = 0$ for each t. Past and present value of x_{it} cannot predict e_{it}. It is silent on whether future values of x_{it} predicts e_{it}*

2.2 Pooled OLS

Here, we assume that c_i is a constant. Then the data generating process is written

$$y_{it} = x'_{it}\beta + c + e_{it}$$

Thus, c_i practically becomes an overall intercept. If x_{it} already contains a constant, then the data generating process becomes a simple OLS with both time and entity dimensions. If we assume that $E[x_{it}e_{it}] = 0$, we end up with the following pooled OLS estimate

$$\begin{aligned}
\hat{\beta}_{POLS} &= \left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x'_{it}\right)^{-1}\left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}y_{it}\right) \\
&= \left(\sum_{i=1}^{n} X'_i X_i\right)^{-1}\left(\sum_{i=1}^{n} X'_i y_i\right)
\end{aligned}$$

We can see the unbiasedness, sample variance, and the asymptotic distribution can be derived as

$$\begin{aligned}
E[\hat{\beta}_{POLS}|X_i] &= \beta + \left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x'_{it}\right)^{-1}\left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}E[e_{it}|x_{i1},....,x_{iT},c]\right) = \beta \\
\operatorname{plim}\hat{\beta}_{OLS} &= \beta + \operatorname{plim}\left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x'_{it}\right)^{-1}\left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}e_{it}\right) = \beta \\
\widehat{V}_{\hat{\beta}_{POLS}} &= (X'X)^{-1}\left(\sum_{i=1}^{n} X'_i\hat{e}_i\hat{e}'_i X_i\right)(X'X)^{-1} \\
\sqrt{nT}(\hat{\beta}_{POLS} - \beta) &\xrightarrow{d} N(0, V_\beta)
\end{aligned}$$

where the strict exogeneity is used for unbiasedness. For asymptotics, fix T and let $n \to \infty$. As usual, $V_\beta = Q^{-1}\Omega Q^{-1}$ where $Q^{-1} = E[X'X]^{-1}$ and $\Omega = E\left[\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x'_{it}e^2_{it}\right]$. Since we have more observations to determine the estimate of β (n vs. nT), the estimator converges to the true value quicker!

We can also see the weakness of the POLS estimator. We were able to ignore potential correlation between c_i and x_{it} by assuming c_i is a constant. If they are correlated, then it can be shown that there will be a bias and inconsistencies in POLS estimates.

2.3 Random effects estimation

Here, we assume that c_i is latent, but uncorrelated with x_i. We can show that POLS will be consistent. However, POLS is not the most efficient estimator. We can find why this is the case and the better alternative using these assumptions.

Assumption 2.2 (Random effects assumptions)**.** *The following are the assumptions for the random effects models*

RE1 *We assume* $E[c_i|X_i] = 0, E[c_i] = 0$ *and strict exogeneity* $E[e_{it}|X_i, c_i] = 0$

RE2 *rank* $\left(E[X_i'\Omega^{-1}X_i]\right) = k$ *(full column rank)*

RE3 *Conditionally spherical variance matrix:* $E[e_ie_i'|X_i, c_i] = \sigma_e^2 I_T$ *and* $E[c_i^2|X_i] = \sigma_c^2$

Assumption **RE3** implies that the covariance of e_{it} is

$$E[e_{it}e_{is}] = \begin{cases} 0 & t \neq s \\ \sigma_e^2 & t = s \end{cases}$$

The covariance of v_{it} with this assumption is

$$E[v_{it}v_{is}] = \begin{cases} \sigma_c^2 & t \neq s \\ \sigma_e^2 + \sigma_c^2 & t = s \end{cases}$$

So for a given entity i, there is a serial correlation relationship across the error terms. This also implies that as we did for AR(1) regression, a GLS estimation can do better in terms of efficiency.

For this, we write $v_i = \begin{pmatrix} v_{i1} \\ ... \\ v_{iT} \end{pmatrix}$, $e_i = \begin{pmatrix} e_{i1} \\ ... \\ e_{iT} \end{pmatrix}$, and let 1_T be a T-dimensional column vector of 1's. Then, we can write

$$v_i = 1_T c_i + e_i$$

Using this notation, we can define

$$\begin{aligned} E[v_iv_i'] &= E\left[(1_Tc_i + e_i)(1_Tc_i + e_i)'\right] = E[1_T1_T'c_i^2 + e_ie_i'] \\ &= \sigma_c^2 1_T 1_T' + \sigma_e^2 I_T \end{aligned}$$

Using the fact that $1_T 1_T'$ is a $T \times T$ matrix of $1'$s as their elements. $E[v_i v_i']$ has the following matrix representation

$$E[v_i v_i'] = \begin{pmatrix} \sigma_c^2 + \sigma_e^2 & \sigma_c^2 & ... & .... \\ \sigma_c^2 & \sigma_c^2 + \sigma_e^2 & ... & ... \\ ... & ... & ... & ... \\ ... & ... & \sigma_c^2 & \sigma_c^2 + \sigma_e^2 \end{pmatrix} = \Omega$$

Then, the random effects GLS estimation can be obtained by

$$\hat{\beta}_{RE} = \left(\sum_{i=1}^{n} X_i' \widehat{\Omega}^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{n} X_i' \widehat{\Omega}^{-1} y_i \right)$$

with the sample variance of $\left(\sum_{i=1}^{n} X_i' \widehat{\Omega}^{-1} X_i \right)^{-1}$. The $\widehat{\Omega}$ can be obtained by estimating each component of the Ω matrix. σ_e^2 can be obtained by the unbiased estimate of $E[e_{it}^2]$ and σ_c^2 from that for $E[v_{it} v_{is}]$ for $t \neq s$. Thus,

$$\hat{\sigma}_e^2 = \frac{1}{nT - k} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{e}_{it}^2$$
$$\hat{\sigma}_c^2 = \frac{1}{nT(T-1)/2 - k} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{s=t+1}^{T} \hat{v}_{it} \hat{v}_{is}$$

The random effects can be considered as a quasi-demeaned estimator. Note that

$$E[v_i v_i'] = \sigma_c^2 1_T 1_T' + \sigma_e^2 I_T = T\sigma_c^2 1_T (\underbrace{1_T' 1_T}_{T})^{-1} 1_T' + \sigma_e^2 I_T = T\sigma_c^2 P_{1_T} + \sigma_e^2 I_T$$

So the square root matrix of Ω, which I write P (s.t. $PP' = \Omega^{-1}$) is

$$P' = \frac{1}{\sigma_e} [I_T - \rho P_T] \quad \left(\rho = 1 - \sqrt{\frac{\sigma_e^2}{T\sigma_c^2 + \sigma_e^2}} \right)$$

Thus, the transformed data generating process $\tilde{y}_{it} = \tilde{x}_{it}' \beta + \tilde{v}_{it}$, where $\tilde{y}_{it} = y_{it} - \rho \bar{y}_i$ has a spherical variances involving $\tilde{v}_{it}$ term.