Generalized Least Squares (GLS) Theory, Heteroscedasticity & Autocorrelation

Ba Chu

E-mail: ba_chu@carleton.ca

Web: http://www.carleton.ca/~bchu

(Note that this is a lecture note. Please refer to the textbooks suggested in the course outline for details. Examples will be given and explained in the class)

1 Objectives

Until now we have assumed that \( \operatorname{var}(u) = \sigma^2 I \) but it can happen that the errors have non-constant variance, i.e., \( \operatorname{var}(u_1) \neq \operatorname{var}(u_2) \neq \cdots \neq \operatorname{var}(u_T) \) or are correlated, i.e., \( E(u_t u_s) \neq 0 \) for \( t \neq s \). This assumption about the errors is true for many economic variables.

Suppose instead that \( \operatorname{Var}(u) = \sigma^2 \Omega \), where the matrix \( \Omega \) contains terms for heteroscedasticity and autocorrelation. **First**, we study the GLS estimator and the feasible GLS estimator. **Second**, we apply these techniques to do statistical inference on the regression model with heteroscedastic errors and test for heteroscedasticity. **Third**, we learn to estimate regression models with autocorrelated disturbances and test for serial correlation. **Fourth**, we also learn to use the ML technique to estimate the regression models with autocorrelated disturbances.

2 GLS

Consider the model:

\[
y = X\beta + u,
\]

(2.1)
where $E[u] = 0$ and $E[u'u] = \sigma^2 \Omega$ with $\Omega$ is any matrix, not necessarily diagonal.

### 2.1 Problem with the OLS

If we estimate (2.1) with the OLS, we will obtain $\hat{\beta}_{OLS} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'u$. It is easy to verify that $\hat{\beta}_{OLS}$ is still unbiased. However, since $\text{Var}(\hat{\beta}_{OLS}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} \neq \sigma^2(X'X)^{-1}$, inferences based on the estimator $\hat{\sigma}^2(X'X)^{-1}$, that we have seen, is misleading. Thus, $t$ and $F$ tests are invalid. Moreover, $\hat{\beta}_{OLS}$ is not always consistent. This is the main reason for us to learn the GLS.

### 2.2 Derivation of the GLS

Suppose $\Omega$ has the eigenvalues $\lambda_1, \ldots, \lambda_T$, by Cholesky’s decomposition, we obtain:

$$\Omega = SAS', \tag{2.1}$$

where $\Lambda$ is a diagonal matrix with the diagonal elements $(\lambda_1, \ldots, \lambda_T)$, and $S$ is an orthogonal matrix. Thus,

$$\Omega^{-1} = S^{-1}\Lambda^{-1}S'^{-1} \tag{2.2}$$

$$= S^{-1}\Lambda^{-1/2}\Lambda'^{-1/2}S'^{-1}$$

$$= PP', \tag{2.3}$$

where $P = S^{-1}\Lambda^{-1/2}$ and $\Lambda^{-1/2}$ is a diagonal matrix with the diagonal elements $(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_T})$. It is straight-forward to prove that $P\Omega P' = I_T$. Now, multiplying both sides of (2.1) by $P$ yields

$$Py = PX\beta + Pu. \tag{2.4}$$

Setting $y^0 = Py$, $X^0 = PX$ and $u^0 = Pu$, we end up with the classical linear regression model:

$$y^0 = X^0\beta + u^0.$$
where $E[u^0] = 0$ and $E[u^0 u^0'] = E[PP'P'] = \sigma^2 \Omega P' = \sigma^2 I_T$. The OLS estimate of $\beta$ is given by

$$\hat{\beta}_{GLS} = (X'X)^{-1}X'y$$

$$= (X'P'PX)^{-1}X'P'Py$$

$$= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$$

The last equation is defined as the GLS estimator denoted by $\hat{\beta}_{GLS}$. Note that the subscript $GLS$ is sometimes omitted for brevity.

The unbiased estimate of $\sigma^2$ is given by

$$\hat{\sigma}^2_{GLS} = \frac{1}{T-K} (y^0 - X^0\hat{\beta})'(y^0 - X^0\hat{\beta}) = \frac{1}{T-K} (y - X\hat{\beta})'\Omega^{-1}(y - X\hat{\beta}).$$

### 2.3 Properties of the GLS estimator

- The GLS estimator is unbiased.

- $\text{Var}(\hat{\beta}_{GLS}) = \sigma^2 (X'\Omega X)^{-1}$.

- The GLS estimator is BLUE.

- If $u \sim N(0, \sigma^2 \Omega)$ then $\hat{\beta} \sim N(\beta, \sigma^2 (X'\Omega^{-1}X)^{-1})$. The F-test formula is given by

$$F = \frac{(\hat{R}\hat{\beta} - r)'[\hat{R}(X'\Omega^{-1}X)^{-1}\hat{R}]^{-1}(\hat{R}\hat{\beta} - r)}{q \hat{\sigma}^2_{GLS}}.$$

- If $\lim_{T \to \infty} \frac{X'y}{T} = Q^0 > 0$, then $\hat{\beta} \xrightarrow{p} \beta$ as $T \to \infty$.

- $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^0^{-1})$. 

3
2.4 Feasible GLS

Since the matrix $\Omega$ is unknown, we need to use an estimator, say, $\hat{\Omega}$. We have

$$\hat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y.$$ 

If the following conditions:

$$p \lim_{T \to \infty} (X'\hat{\Omega}^{-1}X/T) = p \lim_{T \to \infty} (X'\Omega^{-1}X/T)$$

$$p \lim_{T \to \infty} (X'\hat{\Omega}^{-1}u/\sqrt{T}) = p \lim_{T \to \infty} (X'\Omega^{-1}u/\sqrt{T})$$

hold, then $\hat{\beta}_{GLS}$ and $\hat{\beta}_{FGLS}$ are asymptotically equivalent.

3 Heteroscedasticity

Let’s consider the model (2.1) in which the matrix $\Omega$ is a diagonal matrix with the diagonal elements $(\omega^2_1, \ldots, \omega^2_T)$. This is a form of heteroscedasticity defined in terms of inconstant variances. We can estimate this model by using either the FGLS or the modified OLS technique proposed by White [1980]. The FGLS procedure is straightforward. We shall describe White’s OLS procedure in details.

White [1980] proposes a consistent estimator $\hat{\Omega}$ with the diagonal elements $(\hat{\omega}^2_1, \ldots, \hat{\omega}^2_t, \ldots, \hat{\omega}^2_T)$ given by

HC-0 $\hat{\omega}^2_t = \hat{u}^2_t$, where $\hat{u}_t$ is the OLS residual.

HC-1 $\hat{\omega}^2_t = \frac{T}{T-R} \hat{u}^2_t$.

HC-2 $\hat{\omega}^2_t = \frac{\hat{\omega}^2_t}{1-h_t}$, where $h_t$ is t-th diagonal element of the matrix $X(X'X)^{-1}X'$.

HC-3 $\hat{\omega}^2_t = \frac{\hat{\omega}^2_t}{(1-h_t)^2}$. 

4
The asymptotic variance-covariance matrix of $\hat{\beta}_{OLS}$ is given by

$$Avar(\hat{\beta}_{OLS}) = \lim_{T \to \infty} (X'X/T) p \lim(\hat{\sigma}^2_{OLS} X' \hat{\Omega} X/T) \lim_{T \to \infty} (X'X/T).$$

Based on this heteroscedasticity-consistent estimator, the general linear hypotheses may be tested by the Wald statistics:

$$W = (R(\hat{\beta}_{OLS} - r)'[RAvar(\hat{\beta}_{OLS})R']^{-1}(R\hat{\beta} - r) \overset{d}{\Rightarrow} \chi^2(q).$$

The White procedure has large-sample validity. It may not work well in finite samples.

We can identify whether or not there exists heteroscedasticity in the noise by using White’s test. The null hypothesis is homoscedasticity. To do White’s test, we proceed by regressing the OLS residuals on a constant, original regressors, the squares of the regressors, and the cross products of the regressors and obtain the $R^2$ value. The test can be constructed by $TR^2 \sim \chi^2(q)$, where $q$ is the number of variables in the auxiliary regression less one. We reject the null if $TR^2$ is sufficiently large.

## 4 Serial Correlation

Note that serial correlation and autocorrelation mean the same thing. Consider the model:

$$y_t = X'_t \beta + u_t, \ \forall \ t = 1, \ldots, T, \quad (4.1)$$

where $E[u_t] = 0$, $E[u_t^2] = \sigma^2$ and $E[u_t u_s] \neq 0$ for $t \neq s$.

For example, we can model serial correlation by an AR(1) model, i.e.,

$$u_t = \rho u_{t-1} + \epsilon_t, \quad (4.2)$$
where $\epsilon_t \sim IID(0, \sigma^2)$ and $|\rho| < 1$. We can compute

\[
E[u_t^2] = \frac{\sigma^2}{1 - \rho^2}, \\
E[u_t u_s] = \frac{\rho|t-s|\sigma^2}{1 - \rho^2}.
\]

More generally, we have

\[
E[u u'] = \frac{\sigma^2}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{T-1} \\
\rho & 1 & \rho & \ldots & \rho^{T-2} \\
\vdots & \vdots & \vdots & \ldots & \vdots \\
\rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \ldots & 1 \end{pmatrix} = \sigma_u^2 \Omega, \tag{4.3}
\]

where $\sigma_u^2$ and $\Omega$ have the obvious meanings.

### 4.1 Estimate (4.1)-(4.2) by FGLS

The FGLS can be done in the following steps:

1. Regress $y_t$ on $X_t$ and obtain $\hat{u}_t = y_t - X_t'\hat{\beta}_{OLS}$.
2. Regress $\hat{u}_t$ on $\hat{u}_{t-1}$ and obtain $\hat{\rho}_{OLS} = \frac{\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^{T} \hat{u}_t^2}$.
3. Construct the FGLS by $\hat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y$, where $\hat{\Omega}$ is obtained from (4.3) by substituting $\rho$ with $\hat{\rho}_{OLS}$.

If $p \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} X_t'u_t = 0$, then $\hat{\beta}_{OLS} \xrightarrow{P} \beta$ and $\hat{\rho}_{OLS} \xrightarrow{P} \rho$. Therefore, $\hat{\beta}_{FGLS}$ is consistent and asymptotically normal.
4.2 Estimate (4.1)-(4.2) by the ML

An application of the conditional probability rules, we have

\[pdf(y_1, y_2) = pdf(y_2|y_1)pdf(y_1),\]
\[pdf(y_1, y_2, y_3) = pdf(y_3|y_1, y_2)pdf(y_1, y_2) = pdf(y_3|y_1, y_2)pdf(y_2|y_1)pdf(y_1),\]

\[\cdots\cdots\cdots\]
\[pdf(y_1, \ldots, y_T) = pdf(y_1) \prod_{t=2}^{T} pdf(y_t|y_{t-1}, \ldots, y_1).\]

Since

\[y_t = X_t'\beta + u_t = X_t'\beta + \rho u_{t-1} + \epsilon_t = X_t'\beta + \rho(y_{t-1} - X_{t-1}'\beta) + \epsilon_t\]

and \(\epsilon_t \sim N(0, \sigma^2_\epsilon),\) we have

\[y_t|y_{t-1} \sim N(\rho y_{t-1} + X_t'\beta - \rho X_{t-1}'\beta, \sigma^2_\epsilon).\]

Moreover, since \(y_1 \sim N(X_1'\beta, \frac{\sigma^2_\epsilon}{1-\rho^2}),\) we can derive the log-likelihood function:

\[\log \mathcal{L}(\beta, \rho, \sigma^2_\epsilon) = \log pdf(y_1, \ldots, y_T)\]
\[= \log pdf(y_1) \prod_{t=2}^{T} pdf(y_t|y_{t-1}).\]

We then maximize this log likelihood function to obtain the MLEs. The asymptotic variance of the MLEs can be obtained from the CR lower bound.
4.3 Estimate (4.1) by the modified OLS

In general, the model (4.1) with general form of heteroscedasticity and autocorrelation can be consistently estimated by the OLS, i.e., \( \hat{\beta}_{OLS} = (X'X)^{-1}X'y \). \( \hat{\beta}_{OLS} \) is AN, i.e.,

\[
\sqrt{T}(\hat{\beta}_{OLS} - \beta) \overset{d}{\to} N(0, \sigma^2 Q^{-1} M Q^{-1}),
\]

where \( Q = \lim_{T \to \infty} X'X / T \) and \( M = p \lim_{T \to \infty} X' \Omega X / T \) under the assumption \( E[uu'] = \sigma^2 \Omega \).

\( Q \) can be estimated by \( \hat{Q} = X'X / T \); \( \sigma^2 M \) can be consistently estimated by the Newey-West heteroscedasticity and autocorrelation consistent covariance (HAC) matrix estimator:

\[
HAC = \hat{\Gamma}_0 + \sum_{j=1}^{m} w(j, m)(\hat{\Gamma}_j + \hat{\Gamma}'_j),
\]

where

\[
\hat{\Gamma}_j = \frac{1}{T} \sum_{t=j+1}^{T} X_t \hat{u}_t \hat{u}_{t-j} X'_{t-j}, \quad \forall \ j = 1, \ldots, m,
\]

where \( w(j, m) \) is a lag window, \( m \) is a lag truncation parameter. If \( w(j, m) = 1 - j/(m+1) \), then \( w(j, m) \) is called the Barlett window.

Hence, the asymptotic covariance matrix of \( \hat{\beta}_{OLS} \) can be estimated by \( \hat{Q}^{-1} HAC \hat{Q}^{-1} \).

4.4 Test for serial correlation

- The Durbin-Watson test

\[
DW = \frac{\sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T} \hat{u}_t^2},
\]

where \( \hat{u}_t \) is the OLS residual, is used to test for the first-order serial correlation (i.e., AR(1)).

- For \( H_0 : \rho = 0 \) vs. \( H_1 : \rho > 0 \), reject \( H_0 \) if \( DW \leq DW_\ell \), do not reject if \( DW > DW_u \), and inconclusive if \( DW_\ell < DW < DW_u \) with \( DW_\ell \) is the lower critical value and \( DW_u \) is upper critical value.

- For \( H_0 : \rho = 0 \) vs. \( H_1 : \rho < 0 \), reject \( H_0 \) if \( DW \geq 4 - DW_\ell \), accept if \( DW < 4 - DW_u \).
• The Breusch-Godfrey LM test is used to test for high-order serial correlation (i.e., the AR\(^p\)) model as given by

\[ u_t = \sum_{i=1}^{p} \phi_i u_{t-i} + \epsilon_t. \]

The test is carried out by doing the OLS regression: \( \widehat{u}_t = X'_t \gamma + \sum_{i=1}^{p} b_i \widehat{u}_{t-i} + \text{error}. \) Next, we do either the LM test \( TR^2 = d \chi^2(q) \) under the null hypothesis of no serial correlation or the F-test for \( b_1 = b_2 = \cdots = b_p = 0. \)

5 Autoregressive Conditional Heteroscedasticity (ARCH)

Engle [1982] suggests that heteroscedasticity may occur in time-series context. In speculative markets such as exchange rates and stock market returns, one often observes the phenomenon called volatility clustering – large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes. Observations of this phenomenon in financial time series have led to the use of ARCH models. Engle [1982] formulates the notion that the recent past might give information about the conditional disturbance variance. He postulates the relation:

\[ y_t = X'_t \beta + \sigma_t \epsilon_t, \text{ where } \epsilon_t \sim IID(0, 1), \]
\[ \sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_p u_{t-p}^2, \text{ where } u_t = y_t - X'_t \beta. \]

Testing for ARCH can be done in the following steps:

1. Fit \( y \) to \( X \) by OLS and obtain the residuals \( \{\epsilon_t\}. \)

2. Compute the OLS regression, \( \epsilon_t^2 = \widehat{\alpha}_0 + \widehat{\alpha}_1 \epsilon_{t-1}^2 + \cdots + \widehat{\alpha}_p \epsilon_{t-p}^2 + \text{error}. \)

3. Test the joint significance of \( \widehat{\alpha}_0, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_p. \)

6 Exercises

1. Problems 6.1, 6.2, 6.3 and 6.7 (pp. 202-203, JD97).
2. Consider the model

\[ y_t = \beta y_{t-1} + u_t, \quad |\beta| < 1, \]
\[ u_t = \epsilon_t + \theta \epsilon_{t-1}, \quad |\theta| < 1, \]

where \( t = 1, \ldots, T \) and \( \epsilon_t \sim IID(0, \sigma^2) \).

(a) Show that the OLS estimator of \( \beta \) is *inconsistent* when \( \theta \neq 0 \).

(b) Explain how you would test \( H_0 : \theta = 0 \) vs. \( H_1 : \theta \neq 0 \) using the Lagrangian multiplier test principle. Is such test likely to be superior to the Durbin-Watson test?.

3. Consider the model

\[ y_t = \beta x_t + u_t, \]
\[ u_t = \rho u_{t-1} + \epsilon_t, \quad |\rho| < 1, \]

where \( x_t \) is non-stochastic and \( \epsilon_t \sim IID - N(0, \sigma^2) \).

(a) Given a sample \( \{x_t, y_t\}_{t=1}^T \), construct the log-likelihood function.

(b) Estimate the parameters \( \beta, \rho \) and \( \sigma^2 \) by the ML method.

(c) Derive the asymptotic variances of these MLEs.

(d) Given \( x_{T+1} \), what do you think is the best predictor of \( y_{T+1} \)?

4. Consider the linear regression model \( \mathbf{y} = X \beta + \mathbf{u} \), where \( \mathbf{y} \) is a \( T \times 1 \) vector of observations on the dependent variable, \( X \) is a \( T \times K \) matrix of observations on \( K \) non-stochastic explanatory variables with \( \text{rank}(X) = K < T \), \( \beta \) is a \( K \times 1 \) vector of unknown coefficients, and \( \mathbf{u} \) is a \( T \times 1 \) vector of unobserved disturbances such that \( E[\mathbf{u}] = 0 \) and \( E[\mathbf{uu}'] = \sigma^2 \Omega \), with \( \Omega \) positive definite and known.

(a) Derive the GLS estimator \( \hat{\beta}_{GLS} \) of \( \beta \).
(b) Assuming that $T^{-1/2}X'\Omega^{-1}u \overset{d}{=} N(0, \sigma^2 \lim_{T \to \infty} X'\Omega^{-1}X/T)$, derive the limiting distribution of $\hat{\beta}_{GLS}$ as $T \to \infty$. State clearly any further assumptions you make and any statistical results you use.

5. In the generalized regression model, suppose that $\Omega$ is known.

(a) What is the covariance matrix of the OLS and GLS estimators of $\beta$?

(b) What is the covariance matrix of the OLS residual vector $\hat{u}_{OLS} = y - X\hat{\beta}_{OLS}$?

(c) What is the covariance matrix of the GLS residual $\hat{u}_{GLS} = y - X\hat{\beta}_{GLS}$?

(d) What is the covariance matrix of the OLS and GLS residual vectors?

6. Suppose that $y$ has the pdf $pdf(y_t|X_t) = 1/(X_t'\beta_{i=1}^K\beta_{j=1}^K) \exp\{-y/((X_t'\beta))\}$, $y > 0$. Then $E[y_t|X_t] = \beta'X_t$ and $Var[y_t|X_t] = (\beta'X_t)^2$. For this model, prove that the GLS and MLE are the same, even though this distribution involves the same parameters in the conditional mean function and the disturbance variance.

7. Suppose that the regression model is $y_t = \mu + \epsilon_t$, where $\epsilon$ has a zero mean, constant variance, and equal correlation $\rho$ across observations. Then $Cov(\epsilon_t, \epsilon_s) = \sigma^2 \rho$ if $t \neq s$. Prove that the OLS estimator of $\mu$ is inconsistent.

8. Suppose that the regression model is $y_t = \mu + \epsilon_t$, where

$$E[\epsilon_t|x_t] = 0, \ Cov(\epsilon_t, \epsilon_s|x_t, x_s) = 0 \text{ for } i = j, \text{ but } Var[\epsilon_t|x_t] = \sigma^2 x_t^2, \ x_t > 0.$$ 

(a) Given a sample of observations on $y_t$ and $x_t$, what is the most efficient estimator of $\mu$?, What is its variance?.

(b) What is the OLS estimator of $\mu$?, and what is its variance?.

(c) Prove that the estimator in part (a) is at least as efficient as the estimator in part (b).
References


\[ u_t = u_{t-1} + \varepsilon_t \]

**Derivation:**

Since \( E[u_t^2] = E[(u_{t-1} + \varepsilon_t)^2] \)

\[ E[u_t^2] = E[u_{t-1}^2] + 2E[u_{t-1}]E[\varepsilon_t] + E[\varepsilon_t^2] \]

\[ u_{t-1} \text{ is a constant,} \quad E[u_{t-1}] = 0, \quad E[\varepsilon_t^2] \text{ is a constant,} \]

\[ E[u_t^2] = E[u_{t-1}^2] + E[\varepsilon_t^2] = \sigma^2 \]

Since \( E[u_t^2] = \sigma^2/\text{(no heteroscedasticity)}, \) we have \( \sigma^2 = \sigma^2 E[u_{t-1}]E[\varepsilon_t] \)

\[ \sigma^2 = \frac{E[\varepsilon_t^2]}{1 - p^2} \]

**Dividing:**

\[ u_t = u_{t-1} + \varepsilon_t \]

\[ \begin{align*}
    u_t &= p u_{t-1} + \varepsilon_t \\
    u_{t-1} &= pu_{t-2} + \varepsilon_{t-1} \\
    \vdots \\
    u_1 &= u_0 + \varepsilon_1 \\
    u_s &= pu_s + \varepsilon_s \\
    u_t u_s &= (pu_{t-1} + \varepsilon_t) (pu_{s-1} + \varepsilon_s) \\
    E[u_t u_s] &= p^2 E[u_{t-1} u_{s-1}] + p E[u_{t-1} \varepsilon_s] + p E[\varepsilon_t \varepsilon_s] \\
    E[\varepsilon_t \varepsilon_s] &= 0 \\
    \end{align*} \]

For \( s < t-1, \) we have

\[ \Theta E[\varepsilon_s u_{t-1}] = E\left[ \varepsilon_s \sum_{i=-\infty}^{t-1} p^{t-1-i} \varepsilon_i \right] = E\left[ \varepsilon_s (p^{t-1} \varepsilon_0 + \cdots + p^{i-1} \varepsilon_i + \cdots) \right] \]

\[ = p^{t-s-1} E[\varepsilon_s^2] \quad \text{because} \quad E[\varepsilon_t \varepsilon_s] = 0 \quad \forall \ t \neq s. \quad (1.2) \]

(0+) the proof \( g(x_2) \) is beyond the scope of this course.
(2.1) 
\[ E[u_{S-1}^2] = E\left[ E_t \sum_{i=-\infty}^{S-1} P^{s-1-i} \epsilon_c \right]. \]

\[ = E\left[ E_t \left( P^{S-1} \epsilon_{-\infty} + \ldots + P^0 \epsilon_{S-1} \right) \right]. \]

Since \( t > s_i > S \) (so there is no overlapping).

Else, we have

\[ E[u_{S-1}^2] = E\left[ E_t \sum_{i=-\infty}^{S-1} P^{s-1-i} \epsilon_c \right] = \]

\[ = E\left[ E_t \left( P^{S-1} \epsilon_{-\infty} + \ldots + P^0 \epsilon_{S-1} \right) \right]. \]

Since \( S > t+1 \) so there is no overlapping.

\[ E[u_{S-1}^2] = E\left[ E_t \sum_{i=-\infty}^{S-1} P^{s-1-i} \epsilon_c \right] = \]

\[ = E\left[ E_t \left( P^{S-1} \epsilon_{-\infty} + \ldots + P^0 \epsilon_{S-1} \right) \right]. \]

Note that \( E[u_{S-1}^2] = E[u_{S-1}^2] \) a.e., we have

\[ E[u_{S-1}^2] = E[u_{S-1}^2]. \]

If \( S < t+1 \), in view of (1.2) \& (2.1), we obtain:

(2.4) 
\[ (1-P^2) E[u_i u_s] = P^{t-S} \delta_{i}^2. \]

If \( S > t \), in view of (1.2), (2.2), \& (2.3), we obtain:

(2.5) 
\[ (1-P^2) E[u_i u_s] = \frac{P^{t-S} \delta_{i}^2}{1-P^2}. \]
We can combine (2.4) and (2.5) as:

\[
E[\mu \nu] = \frac{6 \mu^2 \rho L}{1 - \rho^2}
\]