Thailand Statistician
January 2010; 8(1) : 63-80
www.statassoc.or.th
Contributed paper

Moving Average Correction in a Regression Model

Warangkhana Keerativibool
Department of Mathematics, Faculty of Science, Thaksin University, Phattalung, 93110, Thailand.
E-mail: walex27@hotmail.com

Received: 20 August 2009
Accepted: 20 January 2010

Abstract

Some problems in the errors of regression model is an important issue, such as the autocorrelated error, moving average error. When these problems occur, the ordinary least squares (OLS) estimators can not be used because they are not efficient. This paper proposes a transformation matrix to correct the first-order moving average, MA(1), problem and to recover the one lost observation in a regression model. When the errors have the MA(1) problem, the sample mean squared error (MSE) is shown theoretically and empirically to be an overestimate of the MSE after transformation. The results of simulation study confirm that the errors after removing the MA(1) problem are independent and if the MA(1) problem is not corrected, the MSE overestimates the corrected one at the significance level 0.05.

Keywords: first-order moving average, mean squared error, regression model, transformation matrix.

1. Introduction

Regression analysis is a statistical technique for modeling and analyzing the relationship between several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps us understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. Most commonly, regression analysis estimates the conditional expectation of the dependent variable given the independent variables i.e., the average value of the dependent variable when the independent variables are held fixed. Applications of regression can occur in almost every field such as the engineering, physical, chemical, biological, social sciences, economics, and management. The linear regression model can be written conveniently in a matrix form as follows [1]:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\pi} + \mathbf{v} \tag{1}$$

where $\mathbf{y}$ is a $T \times 1$ dependent random vector of observations, $\mathbf{X}$ is a $T \times K$ matrix of independent explanatory variables with full-column rank, $\boldsymbol{\pi}$ is a $K \times 1$ unknown parameter vector of regression coefficients, and $\mathbf{v}$ is a $T \times 1$ error vector.

The classical assumptions for regression analysis are summarized as follows [2-4]. The sample must be representative of the population for the inference prediction. The error is assumed to be a random variable with a mean of zero conditional on the explanatory variables and all of errors are uncorrelated. The variance of the error is constant across observations: homoscedasticity. The independent variables are error-free and they must be linearly independent, i.e., it must not be possible to express any predictor as a linear combination of the others. These are sufficient conditions for the least squares estimator to possess desirable properties, in particular, these assumptions imply that the parameter estimates will be unbiased, consistent, and efficient in the class of linear unbiased estimators.

When the regression model in (1) is fitted, the residuals $\hat{\mathbf{v}}$ may observe a systematic pattern. These residuals may suggest that some essential independent explanatory variables have not been included in the model. Exclusion could be due to inadequate knowledge in the problem and/or lack of accurate data. In this paper, the error generated in the fitted model is assumed in the form of a first-order moving average or MA(1) process,

$$v_t = \varepsilon_t - \theta\varepsilon_{t-1},\ t = 1, 2, \ldots, T \tag{2}$$

where the first-period back error ε_{t-1} is called the first-lag of error ε_t, the parameters θ of the model must satisfy the following condition to ensure the invertibility of the error terms [5],

$$|\theta| < 1 \tag{3}$$

and the error ε_t in (2) is assumed to be normally independent distributed with mean zero and the variance is finite and also time invariant. So that ε_t is an independent identically distributed random variable, obeying

$$\varepsilon_t \sim \mathrm{NID}\left(0,\sigma^2\right),\ t=1,\ 2,\ \ldots,\ T \qquad (4)$$

It is noteworthy that the value of v_1 in the moving average model (2) depend on the value of ε_0 , which is unknown. The recovery of v_1 will be discussed later. The objectives of this paper are two points. Firstly, the transformation matrix is proposed to correct the first-order moving average problem and to recover the one lost observation in a regression model. Secondly, if the moving average problem in errors is ignored, the estimate of sample mean squared error is shown theoretically and empirically to be an overestimate of the corrected one. This paper is divided in four sections. The second proposes the methodology, i.e., the transformation matrix, to correct the first-order moving average problem in a regression model and the third shows the algorithm for simulate the first-order moving average model and shows all results of the simulation study. Conclusions appear in the 4th section..

2. Methodology

Theorem A-1: The $T \times T$ transformation matrix **P**, used to correct the first-order moving average problem in a regression model is defined by

$$\mathbf{P}=\begin{bmatrix} \frac{1}{\sqrt{1+\theta^2}} & 0 & 0 & 0 & \ldots & 0 \\ \theta & 1 & 0 & 0 & \ldots & 0 \\ \theta^2 & \theta & 1 & 0 & \ldots & 0 \\ \theta^3 & \theta^2 & \theta & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta^{T-1} & \theta^{T-2} & \theta^{T-3} & \theta^{T-4} & \ldots & 1 \end{bmatrix}. \qquad (5)$$

The transformation matrix **P** in (5) is used to transform $\mathbf{y}$ and $\mathbf{X}$ in (1) to be $\mathbf{y}^*$ and $\mathbf{X}^*$, respectively, such that the moving average of the errors $\mathbf{v}$ in (1) is eliminated, to give the model

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\pi} + \boldsymbol{\varepsilon} \qquad (6)$$

where $\mathbf{y}^* = \mathbf{Py}$, $\mathbf{X}^* = \mathbf{PX}$, $\mathrm{E}\left(\boldsymbol{\varepsilon} \,\middle|\, \mathbf{X}^*\right) = \mathbf{0}$, and $\mathrm{E}\left(\boldsymbol{\varepsilon\varepsilon}' \,\middle|\, \mathbf{X}^*\right) = \sigma^2 \mathbf{I}_\mathbf{T}$.

Theorem A-2: Under the regression model in (1), if we ignore the moving average problem and use the ordinary least squares (OLS) method to estimate the unknown parameter vector, then the corrected minimum sample mean squared error from the OLS estimate, yields the expectation of $\hat{\mathbf{v}}'\hat{\mathbf{v}}$ divided by $\mathrm{T}-\mathrm{K}$,

$$\mathrm{E}\left(\frac{\hat{\mathbf{v}}'\hat{\mathbf{v}}}{\mathrm{T}-\mathrm{K}}\right) \approx \sigma^2\left(1+\theta^2\right). \tag{7}$$

Under the transformed model in (6), the least squares residuals can consistently be used to estimate the variance σ^2 with unbiased estimator $\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\mathrm{T}-\mathrm{K}}$ [6-7]. Therefore,

$$\mathrm{E}\left(\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\mathrm{T}-\mathrm{K}}\right) = \sigma^2. \tag{8}$$

Subtracting equation (8) from equation (7), yields

$$\mathrm{E}\left(\frac{\hat{\mathbf{v}}'\hat{\mathbf{v}}}{\mathrm{T}-\mathrm{K}}\right) - \mathrm{E}\left(\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\mathrm{T}-\mathrm{K}}\right) \geq 0. \tag{9}$$

The equality to zero in (9) can occur if, and only if the moving average problem does not exists in the error $\hat{\mathbf{v}}$.

From the inequality in (9), we can conclude that if we ignore the moving average problem and use the residuals from the ordinary least squares method to estimate the sample mean squared error, the estimated value is greater than the one after the moving average problem has been corrected.

3. Simulation study and results

In this simulation study, we consider a multiple linear regression model with four independent explanatory variables (K = 4), denoted by x_{t1}, x_{t2}, x_{t3}, and x_{t4}, where the first element of independent explanatory variables x_{t1} is given as a constant 1 $\left(x_{t1} = 1\right)$. The parameters of regression model are assumed to be

$$\pi_1 = 20,\ \pi_2 = 5,\ \pi_3 = 7,\ \text{and}\ \pi_4 = 12.$$

Therefore, the multiple linear regression model in this simulation takes the form

$$y_t = 20 + 5x_{t2} + 7x_{t3} + 12x_{t4} + v_t \tag{10}$$

for $t = 1, 2, \ldots, 100$.

The error v_t in (10) is in the form of moving average model of order 1, MA(1), where the parameter θ is assumed to be

$$\theta = 0.6 .$$

Therefore, the moving average model in this simulation is in the form

$$v_t = \varepsilon_t - 0.6\varepsilon_{t-1} \quad (11)$$

for $t = 1, 2, \ldots, 100$.

The error ε_t in (11) is an independent identically distributed random variable, obeying

$$\varepsilon_t \sim N(0, 0.49) .$$

Steps in the Simulation Study

Step 1 Generate normal independent random variable with zero mean and unit variance, called r, by the CALL RANNOR routine in SAS version 9.1 about 5,700 observations; the seed number to generate r is arbitrarily given as 34134.

Step 2 Generate normal distribution of ε_t with zero mean and variance equal to 0.49 about 5,700 observations by the following equation

$$\varepsilon_t = \sigma \times r = 0.7 \times r .$$

Step 3 Construct the series of errors v_t in (11) about 5,700 observations, using the normal distribution of ε_t which obtained in Step 2 where ε_0 are arbitrarily given to be zero. Then split of the series of errors v_t in sequence to preserve the relationship in v_t into 57 samples, each of which consists of 100 observations, the first 100 observations go to sample 1, the second 100 observations go to sample 2, and so on.

Step 4 Test the MA(1) properties for the errors generated in Step 3, estimate the parameter θ and also test the normality for the residuals ε_t by the ARIMA and the UNIVARIATE procedure, respectively, in SAS version 9.1. Discard 7 samples that fail the test, and retain 50 samples for further study.

Step 5 Generate the series of independent explanatory variables x_{t2}, x_{t3}, and x_{t4} about 6,000 observations by the UNIFORM function in SAS version 9.1 where

$x_{t2} \sim U(5,10)$, $x_{t3} \sim U(10,20)$, and $x_{t4} \sim U(2,6)$; the seed numbers to generate x_{t2}, x_{t3}, and x_{t4} are arbitrarily given as 789455, 9875244, and 658214, respectively. Split of the series of independent explanatory variables x_{t2}, x_{t3}, and x_{t4} as Step 3 in sequence into 60 samples, each of which consists of 100 observations.

Step 6 Test the multicollinearity problem for the series of independent explanatory variables x_{t2}, x_{t3}, and x_{t4} in Step 5, using the knowledge of correlation matrix and the variance inflation factor (VIF). Discard 10 samples which present the multicollinearity problem, and retain 50 samples for further study.

Step 7 Construct the dependent variable y_t described in (10), using the corresponding independent explanatory variables x_{t2}, x_{t3}, and x_{t4} obtained in Step 6 and the moving average error v_t obtained in Step 4.

Step 8 Construct the estimate of proposed transformation matrix $\mathbf{P}$ in (5) for each sample, using the estimated value of θ which obtained in Step 4 and use it to transform the regression model in (1) to the transformed model as shown in (6).

Step 9 Estimate the parameters π's of the transformed model in Step 8 by the REG procedure in SAS version 9.1. Then test the moving average in the residuals and test the normality of the residuals by the ARIMA and the UNIVARIATE procedures, respectively. The tests confirm that the errors of all 50 transformed samples in Step 8 are white noises. Therefore, we can say that the proposed transformation matrix $\mathbf{P}$ in (5) has a transformation percentage of 100 percent.

Step 10 Use the residuals $\hat{\varepsilon}_t$ in Step 9 to calculate the sample mean squared error of the transformed model by the proposed transformation matrix $\mathbf{P}$ in (5),

$$\mathrm{MSE} = \frac{\sum_{t=1}^{100} \hat{\varepsilon}_t^2}{100-4} = \frac{\sum_{t=1}^{100} \hat{\varepsilon}_t^2}{96}.$$

Step 11 Estimate the parameters of the regression model in Step 7 by using the REG procedure where the moving average problem is ignored. Using the residuals $\hat{v}_t$ to calculate the sample mean squared error of the model before transformation,

$$\text{MSE} = \frac{\sum_{t=1}^{100} \hat{v}_t^2}{100-4} = \frac{\sum_{t=1}^{100} \hat{v}_t^2}{96}.$$

Step 12 Compare the sample mean squared errors from Steps 10 and 11 by the TTEST procedure in SAS version 9.1. The result shows that, when ignoring the moving average problem in the errors of the model in (10), the sample mean squared error is greater than the one of the transformed model at the level of significance 0.05. The sample mean squared error of all 50 samples from the model before and after transformation are shown in Figure 1. The mean, standard deviation, minimum and maximum values of the sample mean squared errors before and after transformation are shown in Table 1.

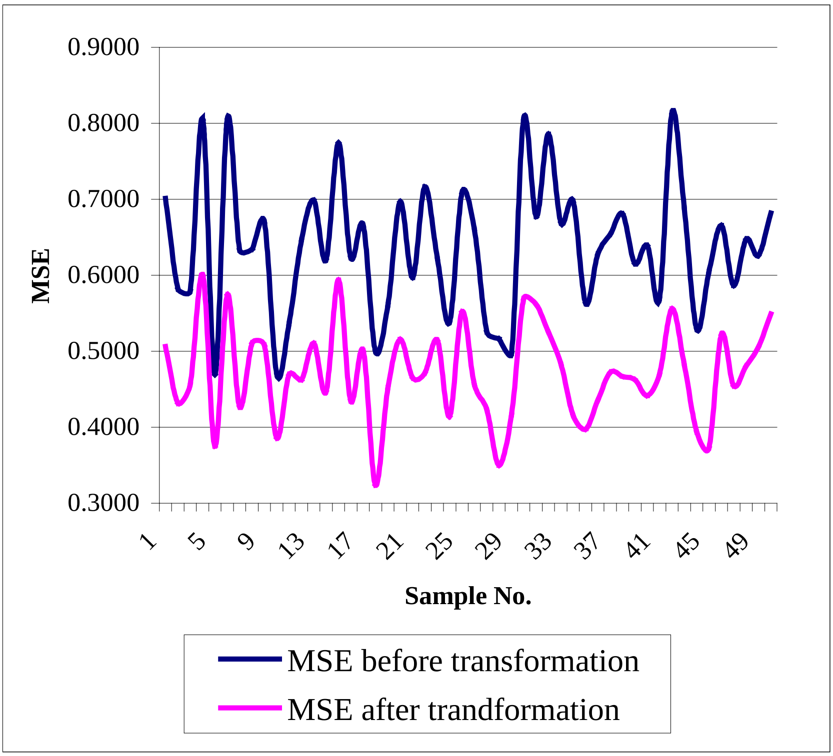

Figure 1. The Sample Mean Squared Error of 50 Samples from the Model before and after Transformation.

Table 1. Mean, Standard Deviation, Minimum and Maximum Values of the Sample Mean Squared Error in 50 Samples with the t and p Values of the Test**.**

Sample no.	MSE	
	Before Transformation	After Transformation
Mean	0.6378	0.4714
SD.	0.0881	0.0626
Min	0.4694	0.3238
Max	0.8142	0.6003
t-test	10.89	
p-value	< 0.001	

4. Conclusions

This paper has presented a transformation matrix in order to correct the first-order moving average problem and to recover the one lost observation in a regression model. When ignoring the first-order moving average, the sample mean squared error is shown to be an overestimate of the sample mean squared error when the moving average has been corrected.

The results of simulation study confirm that the transformed errors of the multiple linear regression model are independent and that before removing the moving average in the errors of the model, the sample mean squared error is greater than the one after the moving average problem has been corrected at the level of significance 0.05.

Acknowledgements

This paper would not have been completed without the help and support of several people in my organization, Thaksin University. I would like to give special thanks to Mrs. Preedaporn Kanjanasomranwong for her comments and suggestions.

References

[1] Montgomery, D.C., Peck, E.A. and Vining, G.G., *Introduction to Linear Regression Analysis,* 4^{th} ed., New York, Wiley, 2006.

[2] Gujarati, D.N., *Basic Econometrics,* 4^{th} ed., Singapore, McGraw-Hill, 2002.

[3] Gujarati, D.N., *Essentials of Econometrics,* 3rd ed., Singapore, McGraw-Hill, 2006.

[4] Kmenta, J., *Elements of Econometrics,* 2nd ed., New York, Macmillan, 1986.

[5] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C., *Time Series Analysis : Forecasting and Control,* 3rd ed., Englewood Cliffs, New Jersey, Prentice Hall, 1994.

[6] Goldberger, A.S., *Econometric Theory,* New York, Wiley, 1964.

[7] Greene, W., *Econometric Analysis,* 6th ed., Upper Saddle River, New Jersey, Prentice-Hall, 2008.

[8] Keerativibool, W., *Selection of a System of Simultaneous Equations Models*, Doctoral Dissertation, School of Applied Statistics, National Institute of Development Administration, 2009.

Appendix

Proof of Theorem A-1.

At the t^{th} observation, the regression model in (1) can be written as follows:

$$y_t = \mathbf{x}_t'\boldsymbol{\pi} + v_t\,,\ t = 1,\ 2,\ \ldots,\ T \tag{A1}$$

where

$$\mathbf{x}_t' = \begin{bmatrix} x_{t1} & x_{t2} & \ldots & x_{tK} \end{bmatrix},$$

$$v_t = \varepsilon_t - \theta\varepsilon_{t-1}\,,\ t = 2,\ 3,\ \ldots,\ T\,. \tag{A2}$$

Replacing v_t in (A2) into (A1),

$$y_t = \mathbf{x}_t'\boldsymbol{\pi} + \varepsilon_t - \theta\varepsilon_{t-1}\,. \tag{A3}$$

Rearrange (A3) in term of ε_t ,

$$\varepsilon_t = y_t - \mathbf{x}_t'\boldsymbol{\pi} + \theta\varepsilon_{t-1}\,. \tag{A4}$$

Then, the i^{th} lag of ε_t can be written as:

$$\varepsilon_{t-i} = y_{t-i} - \mathbf{x}_{t-i}'\boldsymbol{\pi} + \theta\varepsilon_{t-(i+1)}\,. \tag{A5}$$

Use the knowledge of (A5), the equation in (A3) become

$$\begin{aligned}
y_t &= \mathbf{x}_t'\boldsymbol{\pi} + \varepsilon_t - \theta\left(y_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\pi} + \theta\varepsilon_{t-2}\right) \\
y_t + \theta y_{t-1} &= \left(\mathbf{x}_t' + \theta\mathbf{x}_{t-1}'\right)\boldsymbol{\pi} + \varepsilon_t - \theta^2\varepsilon_{t-2} \\
&= \left(\mathbf{x}_t' + \theta\mathbf{x}_{t-1}'\right)\boldsymbol{\pi} + \varepsilon_t - \theta^2\left(y_{t-2} - \mathbf{x}_{t-2}'\boldsymbol{\pi} + \theta\varepsilon_{t-3}\right) \\
y_t + \theta y_{t-1} + \theta^2 y_{t-2} &= \left(\mathbf{x}_t' + \theta\mathbf{x}_{t-1}' + \theta^2\mathbf{x}_{t-2}'\right)\boldsymbol{\pi} + \varepsilon_t - \theta^3\varepsilon_{t-3} \\
&\ \ \vdots
\end{aligned}$$

$$\sum_{i=0}^{T}\theta^{i}y_{t-i} = \sum_{i=0}^{T}\theta^{i}\mathbf{x}'_{\mathbf{t-i}}\boldsymbol{\pi} + \varepsilon_t - \theta^{T+1}\varepsilon_{t-(T+1)}. \quad \text{(A6)}$$

As T becomes large and θ satisfies the invertibility condition, the value of θ^{T+1} in (A6) approach zero. Therefore, (A6) can be reduced to

$$\sum_{i=0}^{T}\theta^{i}y_{t-i} = \sum_{i=0}^{T}\theta^{i}\mathbf{x}'_{\mathbf{t-i}}\boldsymbol{\pi} + \varepsilon_t$$

$$y_t^{*} = \mathbf{x}_{\mathbf{t}}^{*\prime}\boldsymbol{\pi} + \varepsilon_t,\ t = 2, 3, \ldots, T \quad \text{(A7)}$$

where $y_t^{*} = \sum_{i=0}^{T}\theta^{i}y_{t-i}$ and $\mathbf{x}_{\mathbf{t}}^{*\prime} = \sum_{i=0}^{T}\theta^{i}\mathbf{x}'_{\mathbf{t-i}}$.

From (A7) it can be seen that $\mathrm{Var}\left(y_t^{*}\,\middle|\,\mathbf{x}_{\mathbf{t}}^{*}\right) = \mathrm{Var}(\varepsilon_t) = \sigma^2$ for $t = 2, 3, \ldots, T$. In other words, the moving average problem at $t = 2, 3, \ldots, T$ has been corrected. The transformation in (A7) does not include the first observation in (A1). The heteroskedasticity remains unsolved unless the first observation is eliminated. But if the first observation is included in the analysis, the transformation must be extended by the following steps. Firstly, we take the expectation to v_t in (A2),

$$\begin{aligned} E(v_t) &= E(\varepsilon_t) - \theta E(\varepsilon_{t-1}),\ t = 2, 3, \ldots, T \\ &= E(\varepsilon_t) - \theta E(\varepsilon_t) \\ &= (1-\theta)E(\varepsilon_t). \end{aligned}$$

Using the assumption $E(\varepsilon_t) = 0$ in (4), we have the expectation of v_t is equal to zero. Next, from (A1) the variance of y_t given $\mathbf{x}_{\mathbf{t}}$ for $t = 1, 2, \ldots, T$ can be written as

$$\begin{aligned} \mathrm{Var}(y_t\,|\,\mathbf{x}_{\mathbf{t}}) &= \mathrm{Var}(v_t) = E(v_t^2) \\ &= E\left[(\varepsilon_t - \theta\varepsilon_{t-1})^2\right] \\ &= E(\varepsilon_t^2) - 2\theta E(\varepsilon_t\varepsilon_{t-1}) + \theta^2 E(\varepsilon_{t-1}^2) \\ &= E(\varepsilon_t^2) + \theta^2 E(\varepsilon_t^2) \\ &= (1+\theta^2)E(\varepsilon_t^2) \\ \mathrm{Var}(y_t\,|\,\mathbf{x}_{\mathbf{t}}) &= (1+\theta^2)\sigma^2. \end{aligned} \quad \text{(A8)}$$

Hence, the first observation should weighted by $\sqrt{\frac{1}{1+\theta^2}}$, yields the model

$$y_1^* = \mathbf{x}_1^{*\prime}\boldsymbol{\pi} + \varepsilon_1 \tag{A9}$$

where $y_1^* = \sqrt{\frac{1}{1+\theta^2}}\, y_1$ and $\mathbf{x}_1^{*\prime} = \sqrt{\frac{1}{1+\theta^2}}\, \mathbf{x}_1^{\prime}$.

It can be shown that the moving average problem at t = 1 has been corrected,

$$\begin{aligned} \mathrm{Var}\left(y_1^* \,\middle|\, \mathbf{x}_1^*\right) &= \frac{1}{1+\theta^2}\mathrm{Var}\left(y_1 \,\middle|\, \mathbf{x}_1\right) \\ &= \frac{1}{1+\theta^2}\cdot\left(1+\theta^2\right)\sigma^2 \\ &= \sigma^2 . \end{aligned}$$

Combining the results in (A9) and (A7), respectively, the vector of transformed dependent variables $\mathbf{y}^*$ can be written as

$$\mathbf{y}^* = \mathbf{P}\mathbf{y} = \begin{bmatrix} \frac{1}{\sqrt{1+\theta^2}} & 0 & 0 & 0 & \cdots & 0 \\ \theta & 1 & 0 & 0 & \cdots & 0 \\ \theta^2 & \theta & 1 & 0 & \cdots & 0 \\ \theta^3 & \theta^2 & \theta & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta^{T-1} & \theta^{T-2} & \theta^{T-3} & \theta^{T-4} & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \vdots \\ y_T \end{bmatrix}$$

and the matrix of transformed independent variables $\mathbf{X}^* = \mathbf{P}\mathbf{X}$ can be written in similar manner.

Q.E.D.

Proof of Theorem A-2.

Under the regression model in (1) and the moving average problem is ignored, the OLS estimator of unknown parameter vector $\boldsymbol{\pi}$ is

$$\hat{\boldsymbol{\pi}} = \left(\mathbf{X}^{\prime}\mathbf{X}\right)^{-1}\mathbf{X}^{\prime}\mathbf{y} . \tag{A10}$$

Replacing $\hat{\boldsymbol{\pi}}$ in (A10) into the regression model in (1), yields the estimated error $\hat{\mathbf{v}}$ as follows:

$$\hat{\mathbf{v}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\pi}} = \mathbf{y} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \left[\mathbf{I}_{\mathrm{T}} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]\mathbf{y}\,. \tag{A11}$$

The expectation of $\hat{\mathbf{v}}'\hat{\mathbf{v}}$ can be written as

$$\begin{aligned}\mathrm{E}\left(\hat{\mathbf{v}}'\hat{\mathbf{v}}\right) &= \mathrm{E}\left\{\mathbf{y}'\left[\mathbf{I}_{\mathrm{T}} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]\mathbf{y}\right\} \\ &= \mathrm{tr}\left\{\left[\mathbf{I}_{\mathrm{T}} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]\boldsymbol{\Phi}\right\} + \left(\mathbf{X}\boldsymbol{\pi}\right)'\left[\mathbf{I}_{\mathrm{T}} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]\left(\mathbf{X}\boldsymbol{\pi}\right)\end{aligned} \tag{A12}$$

where $\mathbf{X}\boldsymbol{\pi}$ denotes the expectation of $\mathbf{y}$ and $\boldsymbol{\Phi}$ denotes the conditional variance of $\mathbf{y}$ given $\mathbf{X}$.

Consider the last term in (A12),

$$\begin{aligned}\left(\mathbf{X}\boldsymbol{\pi}\right)'\left[\mathbf{I}_{\mathrm{T}} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]\left(\mathbf{X}\boldsymbol{\pi}\right) &= \boldsymbol{\pi}'\mathbf{X}'\mathbf{X}\boldsymbol{\pi} - \boldsymbol{\pi}'\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\pi} \\ &= \boldsymbol{\pi}'\mathbf{X}'\mathbf{X}\boldsymbol{\pi} - \boldsymbol{\pi}'\mathbf{X}'\mathbf{X}\boldsymbol{\pi} = 0\,.\end{aligned}$$

Therefore, the expectation of $\hat{\mathbf{v}}'\hat{\mathbf{v}}$ in (A12) is equal to

$$\mathrm{E}\left(\hat{\mathbf{v}}'\hat{\mathbf{v}}\right) = \mathrm{tr}\left\{\left[\mathbf{I}_{\mathrm{T}} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]\boldsymbol{\Phi}\right\} = \mathrm{tr}\left(\boldsymbol{\Phi}\right) - \mathrm{tr}\left[\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Phi}\right]. \tag{A13}$$

The conditional variance of $\mathbf{y}$ given $\mathbf{X}$, $\boldsymbol{\Phi}$, in terms of the autocovariance function of error $\mathbf{v}$ can be rewritten in terms of the autocorrelation function of error $\mathbf{v}$ at lag $1, 2, \ldots,\ \mathrm{T}-1$ as shown in [8],

$$\boldsymbol{\Phi} = \mathrm{Var}\left(\mathbf{y}\,|\,\mathbf{X}\right) = \mathrm{Var}\left(\mathbf{v}\right) = \mathrm{E}\left(\mathbf{v}\mathbf{v}'\right)$$

$$= \sigma^2\phi(0)\begin{bmatrix} 1 & \frac{\phi(1)}{\phi(0)} & \frac{\phi(2)}{\phi(0)} & \frac{\phi(3)}{\phi(0)} & \cdots & \frac{\phi(\mathrm{T}-2)}{\phi(0)} & \frac{\phi(\mathrm{T}-1)}{\phi(0)} \\ \frac{\phi(1)}{\phi(0)} & 1 & \frac{\phi(1)}{\phi(0)} & \frac{\phi(2)}{\phi(0)} & \cdots & \frac{\phi(\mathrm{T}-3)}{\phi(0)} & \frac{\phi(\mathrm{T}-2)}{\phi(0)} \\ \frac{\phi(2)}{\phi(0)} & \frac{\phi(1)}{\phi(0)} & 1 & \frac{\phi(1)}{\phi(0)} & \cdots & \frac{\phi(\mathrm{T}-4)}{\phi(0)} & \frac{\phi(\mathrm{T}-3)}{\phi(0)} \\ \frac{\phi(3)}{\phi(0)} & \frac{\phi(2)}{\phi(0)} & \frac{\phi(1)}{\phi(0)} & 1 & \cdots & \frac{\phi(\mathrm{T}-5)}{\phi(0)} & \frac{\phi(\mathrm{T}-4)}{\phi(0)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\phi(\mathrm{T}-2)}{\phi(0)} & \frac{\phi(\mathrm{T}-3)}{\phi(0)} & \frac{\phi(\mathrm{T}-4)}{\phi(0)} & \frac{\phi(\mathrm{T}-5)}{\phi(0)} & \cdots & 1 & \frac{\phi(1)}{\phi(0)} \\ \frac{\phi(\mathrm{T}-1)}{\phi(0)} & \frac{\phi(\mathrm{T}-2)}{\phi(0)} & \frac{\phi(\mathrm{T}-3)}{\phi(0)} & \frac{\phi(\mathrm{T}-4)}{\phi(0)} & \cdots & \frac{\phi(1)}{\phi(0)} & 1 \end{bmatrix} \tag{A14}$$

$$= \sigma^2 \phi(0) \times \mathbf{R} = \sigma^2 \phi(0) \times (\mathbf{I_T} + \mathbf{F}) \tag{A15}$$

where $\phi(0)$ in (A14) can be written in the term of variance of error v_t, $E(v_t^2)$, as follows:

$$\phi(0) = \frac{E(v_t^2)}{\sigma^2} = 1 + \theta^2 , \tag{A16}$$

$\phi(1)$ in (A14) can be written in the term of autocovariance of error v_t at lag 1, $E(v_t v_{t-1})$, as follows:

$$\begin{aligned}
\phi(1) &= \frac{E(v_t v_{t-1})}{\sigma^2} = \frac{1}{\sigma^2} E\left[(\varepsilon_t - \theta\varepsilon_{t-1})(\varepsilon_{t-1} - \theta\varepsilon_{t-2})\right] \\
&= \frac{1}{\sigma^2}\left[E(\varepsilon_t\varepsilon_{t-1}) - \theta E(\varepsilon_t\varepsilon_{t-2}) - \theta E(\varepsilon_{t-1}^2) + \theta^2 E(\varepsilon_{t-1}\varepsilon_{t-2})\right] \\
&= \frac{-\theta E(\varepsilon_{t-1}^2)}{\sigma^2} \\
&= \frac{-\theta E(\varepsilon_t^2)}{\sigma^2} \\
&= -\theta ,
\end{aligned}$$

$\phi(s)$, $s = 2, 3, \ldots, T-1$ in (A14) can be written in the term of autocovariance of error v_t at lag s, $E(v_t v_{t-s})$, as follows:

$$\begin{aligned}
\phi(s) &= \frac{E(v_t v_{t-s})}{\sigma^2} = \frac{1}{\sigma^2} E\left[(\varepsilon_t - \theta\varepsilon_{t-1})(\varepsilon_{t-s} - \theta\varepsilon_{t-(s+1)})\right] \\
&= \frac{1}{\sigma^2}\left[E(\varepsilon_t\varepsilon_{t-s}) - \theta E(\varepsilon_t\varepsilon_{t-(s+1)}) - \theta E(\varepsilon_{t-1}\varepsilon_{t-s}) + \theta^2 E(\varepsilon_{t-1}\varepsilon_{t-(s+1)})\right] \\
&= 0 ,
\end{aligned}$$

$$\mathbf{F}=\begin{bmatrix} 0 & \frac{\phi(1)}{\phi(0)} & \frac{\phi(2)}{\phi(0)} & \frac{\phi(3)}{\phi(0)} & \cdots & \frac{\phi(T-2)}{\phi(0)} & \frac{\phi(T-1)}{\phi(0)} \\ \frac{\phi(1)}{\phi(0)} & 0 & \frac{\phi(1)}{\phi(0)} & \frac{\phi(2)}{\phi(0)} & \cdots & \frac{\phi(T-3)}{\phi(0)} & \frac{\phi(T-2)}{\phi(0)} \\ \frac{\phi(2)}{\phi(0)} & \frac{\phi(1)}{\phi(0)} & 0 & \frac{\phi(1)}{\phi(0)} & \cdots & \frac{\phi(T-4)}{\phi(0)} & \frac{\phi(T-3)}{\phi(0)} \\ \frac{\phi(3)}{\phi(0)} & \frac{\phi(2)}{\phi(0)} & \frac{\phi(1)}{\phi_j(0)} & 0 & \cdots & \frac{\phi(T-5)}{\phi(0)} & \frac{\phi(T-4)}{\phi(0)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\phi(T-2)}{\phi(0)} & \frac{\phi(T-3)}{\phi(0)} & \frac{\phi(T-4)}{\phi(0)} & \frac{\phi(T-5)}{\phi(0)} & \cdots & 0 & \frac{\phi(1)}{\phi(0)} \\ \frac{\phi(T-1)}{\phi(0)} & \frac{\phi(T-2)}{\phi(0)} & \frac{\phi(T-3)}{\phi(0)} & \frac{\phi(T-4)}{\phi(0)} & \cdots & \frac{\phi(1)}{\phi(0)} & 0 \end{bmatrix}$$

$$=\begin{bmatrix} 0 & f(1) & 0 & 0 & \cdots & 0 & 0 \\ f(1) & 0 & f(1) & 0 & \cdots & 0 & 0 \\ 0 & f(1) & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & f(1) \\ 0 & 0 & 0 & 0 & \cdots & f(1) & 0 \end{bmatrix}, \qquad \text{(A17)}$$

and $f(1)$ in (A17) represents the autocorrelation function of the error v_t at lag 1,

$$f(1)=\frac{\phi(1)}{\phi(0)}=-\frac{\theta}{1+\theta^2} \qquad \text{(A18)}$$

Therefore, the expectation of $\hat{\mathbf{v}}'\hat{\mathbf{v}}$ in (A13) can be written as

$$\begin{aligned} E(\hat{\mathbf{v}}'\hat{\mathbf{v}}) &= \operatorname{tr}\left[\sigma^2\phi(0)\times\mathbf{R}\right]-\operatorname{tr}\left\{\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\left[\sigma^2\phi(0)\times\mathbf{R}\right]\right\} \\ &= \sigma^2\phi(0)\left\{\operatorname{tr}(\mathbf{R})-\operatorname{tr}\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}\right]\right\} \\ &= \sigma^2\phi(0)\left\{T-\operatorname{tr}\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I_T}+\mathbf{F})\right]\right\} \\ &= \sigma^2\phi(0)\left\{T-\operatorname{tr}\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]-\operatorname{tr}\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{F}\right]\right\} \\ &= \sigma^2\phi(0)\left\{(T-K)-\operatorname{tr}\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{F}\right]\right\}. \end{aligned} \qquad \text{(A19)}$$

We can normalized the independent variable x_{tk} to be $\tilde{x}_{tk} = x_{tk} - \bar{x}_k$. Hence, we have

$$\sum_{t=1}^{T} \tilde{x}_{tk} = \sum_{t=1}^{T} \left(x_{tk} - \bar{x}_k \right) = 0,\ k = 2, 3, \ldots, K$$

then the matrix of independent variables becomes

$$\mathbf{X} = \begin{bmatrix} 1 & \tilde{x}_{12} & \tilde{x}_{13} & \cdots & \tilde{x}_{1K} \\ 1 & \tilde{x}_{22} & \tilde{x}_{23} & \cdots & \tilde{x}_{2K} \\ 1 & \tilde{x}_{32} & \tilde{x}_{33} & \cdots & \tilde{x}_{3K} \\ 1 & \tilde{x}_{42} & \tilde{x}_{43} & \cdots & \tilde{x}_{4K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \tilde{x}_{T-3,2} & \tilde{x}_{T-3,3} & \cdots & \tilde{x}_{T-3,K} \\ 1 & \tilde{x}_{T-2,2} & \tilde{x}_{T-2,3} & \cdots & \tilde{x}_{T-2,K} \\ 1 & \tilde{x}_{T-1,2} & \tilde{x}_{T-1,3} & \cdots & \tilde{x}_{T-1,K} \\ 1 & \tilde{x}_{T2} & \tilde{x}_{T3} & \cdots & \tilde{x}_{TK} \end{bmatrix}. \tag{A20}$$

By the usual assumption of the regression model, two independent variables $\tilde{x}_{tk}$ and $\tilde{x}_{tk'}$, $k \neq k'$ are uncorrelated. That is,

$$\sum_{t=1}^{T} \tilde{x}_{tk}\tilde{x}_{tk'} = \sum_{t=1}^{T} \left(x_{tk} - \bar{x}_k \right)\left(x_{tk'} - \bar{x}_{k'} \right)$$

is insignificantly different from zero. Therefore, the inverse of matrix $\mathbf{X}'\mathbf{X}$ in (A20) can be reduced to the form of diagonal matrix,

$$\left(\mathbf{X}'\mathbf{X} \right)^{-1} = \operatorname{diag}\left\{ \frac{1}{T}, \frac{1}{\sum_{t=1}^{T} \tilde{x}_{t2}^2}, \frac{1}{\sum_{t=1}^{T} \tilde{x}_{t3}^2}, \ldots, \frac{1}{\sum_{t=1}^{T} \tilde{x}_{tK}^2} \right\}. \tag{A21}$$

The value of $\operatorname{tr}\left[\mathbf{X}\left(\mathbf{X}'\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{F} \right]$ in (A19) is equivalent to the value of

$$\operatorname{tr}\left[\left(\mathbf{X}'\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{F}\mathbf{X} \right]. \tag{A22}$$

Post-multiply the transpose of matrix $\mathbf{X}$ in (A20) by the matrix $\mathbf{F}$ in (A17), yields

$$\mathbf{X}'\mathbf{F} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & \cdots & a_{1,T-3} & a_{1,T-2} & a_{1,T-1} & a_{1T} \\ a_{21} & a_{22} & a_{23} & a_{24} & \cdots & a_{2,T-3} & a_{2,T-2} & a_{2,T-1} & a_{2T} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ a_{K1} & a_{K2} & a_{K3} & a_{K4} & \cdots & a_{K,T-3} & a_{K,T-2} & a_{K,T-1} & a_{KT} \end{bmatrix} \tag{A23}$$

where

$$a_{11} = a_{1T} = f(1),$$
$$a_{12} = a_{13} = a_{14} = \ldots = a_{1,T-3} = a_{1,T-2} = a_{1,T-1} = 2f(1),$$

and for $k = 2, 3, \ldots, K$

$$a_{k1} = \tilde{x}_{2k} f(1),$$
$$a_{k2} = (\tilde{x}_{1k} + \tilde{x}_{3k}) f(1),$$
$$a_{k3} = (\tilde{x}_{2k} + \tilde{x}_{4k}) f(1),$$
$$a_{k4} = (\tilde{x}_{3k} + \tilde{x}_{5k}) f(1),$$
$$\vdots$$
$$a_{k,T-3} = (\tilde{x}_{T-4,k} + \tilde{x}_{T-2,k}) f(1),$$
$$a_{k,T-2} = (\tilde{x}_{T-3,k} + \tilde{x}_{T-1,k}) f(1),$$
$$a_{k,T-1} = (\tilde{x}_{T-2,k} + \tilde{x}_{Tk}) f(1),$$
$$a_{kT} = \tilde{x}_{T-1,k} f(1).$$

Post-multiply the matrix $\mathbf{X}'\mathbf{F}_j$ in (A23) by the matrix $\mathbf{X}$ in (A20) will produce the matrix $\mathbf{X}'\mathbf{F}_j\mathbf{X}$ which takes the form

$$\mathbf{X}'\mathbf{F}_j\mathbf{X} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1K} \\ b_{21} & b_{22} & \cdots & b_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ b_{K1} & b_{K2} & \cdots & b_{KK} \end{bmatrix} \tag{A24}$$

where the diagonal elements of the matrix $\mathbf{X}'\mathbf{F}_j\mathbf{X}$ in (A24) is given by

$$b_{11} = \sum_{t=1}^{T} a_{1t} = 2f(1) + 2(T-2)f(1) = 2(T-1)f(1)$$

and for $k = 2, 3, \ldots, K$

$$
\begin{aligned}
b_{kk} &= \sum_{t=1}^{T} a_{kt}\tilde{x}_{tk} \\
&= a_{k1}\tilde{x}_{1k} + a_{k2}\tilde{x}_{2k} + a_{k3}\tilde{x}_{3k} + a_{k4}\tilde{x}_{4k} + \ldots \\
&\quad + a_{k,T-3}\tilde{x}_{T-3,k} + a_{k,T-2}\tilde{x}_{T-2,k} + a_{k,T-1}\tilde{x}_{T-1,k} + a_{kT}\tilde{x}_{Tk} \\
&= \tilde{x}_{1k}\tilde{x}_{2k}f(1) + \tilde{x}_{2k}\left(\tilde{x}_{1k} + \tilde{x}_{3k}\right)f(1) + \tilde{x}_{3k}\left(\tilde{x}_{2k} + \tilde{x}_{4k}\right)f(1) + \tilde{x}_{4k}\left(\tilde{x}_{3k} + \tilde{x}_{5k}\right)f(1) \\
&\quad + \ldots + \tilde{x}_{T-3,k}\left(\tilde{x}_{T-4,k} + \tilde{x}_{T-2,k}\right)f(1) + \tilde{x}_{T-2,k}\left(\tilde{x}_{T-3,k} + \tilde{x}_{T-1,k}\right)f(1) \\
&\quad + \tilde{x}_{T-1,k}\left(\tilde{x}_{T-2,k} + \tilde{x}_{Tk}\right)f(1) + \tilde{x}_{Tk}\tilde{x}_{T-1,k}f(1) \\
&= 2f(1)\left\{\tilde{x}_{1k}\tilde{x}_{2k} + \tilde{x}_{2k}\tilde{x}_{3k} + \tilde{x}_{3k}\tilde{x}_{4k} + \tilde{x}_{4k}\tilde{x}_{5k} + \ldots \right. \\
&\qquad \left. + \tilde{x}_{T-4,k}\tilde{x}_{T-3,k} + \tilde{x}_{T-3,k}\tilde{x}_{T-2,k} + \tilde{x}_{T-2,k}\tilde{x}_{T-1,k} + \tilde{x}_{T-1,k}\tilde{x}_{Tk}\right\} \\
&= 2f(1)\sum_{t=2}^{T} \tilde{x}_{t-1,k}\tilde{x}_{tk}.
\end{aligned}
$$

The off-diagonal elements of the matrix $\mathbf{X'F_jX}$ in (A24) is not given here because under the condition is that the inverse of matrix $\mathbf{X'X}$ in (A21) is diagonal matrix, these elements are not necessary to calculate the value of $\mathrm{tr}\left[(\mathbf{X'X})^{-1}\mathbf{X'F_jX}\right]$ in (A19).

Therefore, the value of $\mathrm{tr}\left[(\mathbf{X'X})^{-1}\mathbf{X'F_jX}\right]$ in (A19) is equal to

$$
2\left\{\frac{T-1}{T}f(1) + f(1)\frac{\sum_{t=2}^{T}\tilde{x}_{t-1,2}\tilde{x}_{t2}}{\sum_{t=1}^{T}\tilde{x}_{t2}^{2}} + f(1)\frac{\sum_{t=2}^{T}\tilde{x}_{t-1,3}\tilde{x}_{t3}}{\sum_{t=1}^{T}\tilde{x}_{t3}^{2}} + \ldots + f(1)\frac{\sum_{t=2}^{T}\tilde{x}_{t-1,K}\tilde{x}_{tK}}{\sum_{t=1}^{T}\tilde{x}_{tK}^{2}}\right\}. \qquad \text{(A25)}
$$

Under the assumption of nonautocorrelation in the independent variable $\tilde{x}_{tk}$, we have

$$
\frac{\sum_{t=2}^{T}\tilde{x}_{t-1,k}\tilde{x}_{tk}}{\sum_{t=1}^{T}\tilde{x}_{tk}^{2}},\ k = 2, 3, \ldots, K
$$

is insignificantly different from zero. Therefore, the value of $\mathrm{tr}\left[(\mathbf{X'X})^{-1}\mathbf{X'F_jX}\right]$ in (A25) can be approximated by

$$
2\frac{T-1}{T}f(1) = 2f(1) - \frac{2}{T}f(1). \qquad \text{(A26)}
$$

As T becomes large and the absolute value of $f(1)$ is less than one, the second term, $\frac{2}{T}f(1)$, in (A26) can be dropped. Therefore, (A26) can be approximately reduced to

$$\mathrm{tr}\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{F}_j\right] = \mathrm{tr}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{F}_j\mathbf{X}\right] \approx 2f(1). \tag{A27}$$

Substituting the value of $f(1)$ in (A18) into (A27) will provide

$$\mathrm{tr}\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{F}_j\right] \approx -\frac{2\theta}{1+\theta^2}. \tag{A28}$$

Replacing $\phi(0)$ in (A16) and $\mathrm{tr}\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{F}_j\right]$ in (A28) into the expected value of $\hat{\mathbf{v}}'\hat{\mathbf{v}}$ in (A19), yields the *approximated* value of the expectation of $\hat{\mathbf{v}}'\hat{\mathbf{v}}$

$$E(\hat{\mathbf{v}}'\hat{\mathbf{v}}) \approx \sigma^2\left\{(1+\theta^2)\left[(T-K)+\frac{2\theta}{1+\theta^2}\right]\right\}. \tag{A29}$$

Divided the expectation of $\hat{\mathbf{v}}'\hat{\mathbf{v}}$ in (A29) by $T-K$,

$$E\left(\frac{\hat{\mathbf{v}}'\hat{\mathbf{v}}}{T-K}\right) \approx \sigma^2\left\{(1+\theta^2)\left[1+\frac{\frac{2\theta}{1+\theta^2}}{T-K}\right]\right\}. \tag{A30}$$

As T becomes large, (A30) can be reduced to

$$E\left(\frac{\hat{\mathbf{v}}'\hat{\mathbf{v}}}{T-K}\right) \approx \sigma^2(1+\theta^2). \tag{A31}$$

Under the invertibility conditions in (3), we have $1<1+\theta^2<2$. Therefore, (A31) is always greater than σ^2.

Q.E.D.