Quantitative Bio-Science 35(1), 7~10(2016)

Remarks on the SEECM

Hee Soo Lee

Department of Business Administration, Sejong University, Seoul 05006, Korea
(Received May 2, 2016; Revised May 9, 2016; Accepted May 10, 2016)

ABSTRACT

In this paper, we calculate mean and variance of least square estimators (LSEs) of the coefficients in the single equation error correction model (SEECM). We compare them with those in the linear regression model (LRM) to assess the usefulness of the SEECM. We consider three cases that assume different settings for LRM. Our results show that the SEECM is preferable for trend plus random walk error but not for trend plus iid error. For random exploratory variable with error, the SEECM is preferable, depending on the correlation structure of random exploratory variable.

Key words : Least square estimator, Linear regression model, Single equation error correction model

1. Introduction

Error correction models (ECMs) were first introduced by Davidson et al. [1] in a study of the relationship between income and consumption in the United Kingdom. ECMs are a category of multiple time series models that directly estimate the speed at which a dependent variable returns to equilibrium after change in an exploratory variable. They are designed to capture both the short-term and long-term effects of one (or more) time series to another. Thus they are often used in analysis of political and economic processes. The ECMs are useful models when dealing with cointegrated data, but can also be used with stationary data. Engle and Granger [2] suggested an appropriate model for two or more time series that are cointegrated and two-step method for estimating the model. Based on the cointegration of two or more time series, Engle and Granger's two-step ECM assumes endogeneity between the cointegrated time series and does not clearly distinguish dependent variables from independent variables. Thus it might be inconsistent with some existing theories in social science.

Under this situation, we are often better off estimating a single equation error correction model (SEECM). The SEECM clearly distinguishes between dependent and exploratory variables and is appropriate for both cointegrated and long-memoried, but stationary data [3]. In other words, it does not require cointegrated variables to provide information about the rate of error correction. It is applicable for long- and short-term effects of exploratory variables on a dependent variable even when the data are stationary. The concepts of error correction, equilibrium, and long-term effects are not unique to cointegrated variables. Furthermore, an SEECM may provide a more useful modelling technique for stationary data than alternative approaches (see [4] for details). Notably, the SEECM is useful when we have long-memoried and stationary data that shows causal relationship of interest. In this paper, we calculate mean and variance of least square estimator (LSE) of the coefficients in the SEECM. We compare them with those in the LRM to demonstrate the usefulness of the SEECM. We consider three cases that assume different regression settings for exploratory variable and error.

2. Main Results

The SEECM is defined as

$$\begin{aligned}\Delta Y_t &= \alpha + \beta_0 \Delta X_t + \beta_1 Y_{t-1} + \beta_2 X_{t-1} + \epsilon_t \\ &= \alpha + \beta_0 \Delta X_t + \beta_1 (Y_{t-1} - \gamma X_{t-1}) + \epsilon_t \qquad (1)\end{aligned}$$

* Correspondence should be addressed to Hee Soo Lee, Department of Business Administration, Sejong University, Seoul 05006, Korea. Tel: +82-2-6935-2450, Fax: +82-2-3408-4310, E-mail: heesoo@sejong.ac.kr

where ϵ_t is iid error. Assuming the LRM

$$Y_t = \alpha + \beta X_t + \epsilon_t,$$

an alternative SEECM model can be written as

$$\begin{aligned}\Delta Y_t &= \beta(X_t - X_{t-1}) + \epsilon_t - \epsilon_{t-1} \\ &= \alpha + \beta \Delta X_t - (Y_{t-1} - \beta X_{t-1}) + \epsilon_t. \end{aligned} \tag{2}$$

In this Section, we will consider three cases for a given data $(X_1, Y_1),\cdots,(X_n, Y_n)$ as follows.

Case (i): Trend plus iid error

Consider an LRM

$$Y_t = \alpha + \beta X_t + \epsilon_t \tag{3}$$

where $X_t = t$ for $t = 1,\cdots,n$ and iid error ϵ_t with $E\epsilon = 0$ and $E\epsilon^2 = \sigma^2$. Least square estimator for model (3) is given by

$$\begin{aligned}\hat{\beta} &= \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) / \sum_{i=1}^{n}(X_i - \bar{X})^2 \\ &= \beta + \frac{12}{n(n^2-1)} \sum_{i=1}^{n}(i - (n+1)/2)\epsilon_i. \end{aligned}$$

Then

$$E(\hat{\beta}) = \beta \text{ and } Var(\hat{\beta}) = 12\sigma^2/(n(n^2-1)). \tag{4}$$

Also consider the alternative SEECM given by (2)

$$Y_t - Y_{t-1} = \beta(X_t - X_{t-1}) + \epsilon_t - \epsilon_{t-1} = \beta(X_t - X_{t-1}) + \eta_t,$$

which may be rewritten as

$$\Delta Y_t = \beta \Delta X_t + \eta_t. \tag{5}$$

LSE for model (5) is given by

$$\hat{\beta} = \sum_{i=1}^{n-1} \Delta X_i \Delta Y_i / \sum_{i=1}^{n-1} (\Delta X_i)^2 = \beta + \frac{1}{n-1} \sum_{i=1}^{n-1} \eta_i.$$

Note here that $\Delta X_i = 1$ and that $E(\eta_t^2) = 2\sigma^2$, $E(\eta_t \eta_{t+1}) = -\sigma^2$ and $E(\eta_t \eta_{t+k}) = 0$ for $k \geq 2$. Then

$$E(\hat{\beta}) = \beta \text{ and } Var(\hat{\beta}) = 2\sigma^2/(n-1)^2. \tag{6}$$

Thus comparing (4) and (6), LSE for LRM of (3) performs better than LSE for SEECM (5).

In order to check the validity of the above calculation, we conduct a small simulation study. We generate data of size n = 100 via the following equation.

$$Y_t = 1.5 + 2X_t + \epsilon_t, X_t = t, \epsilon_t \sim N(0,1)$$

Then we find the LSE of the coefficient $\beta = 2$ in the above LRM and the alternative SEECM with the generated data. This experiment is repeated 100 times to obtain sample mean and sample variance of $\hat{\beta}$.

Table 1. This table reports simulation results for trend plus iid error model (Case (i)).

	LRM	*AlternativeSEECM*
Mean($\hat{\beta}$)	2.0000	2.0027
Var($\hat{\beta}$)	0.0001	0.0002

The simulation results show that $Var(\hat{\beta})$ of LRM is smaller than that of the alternative SEECM as expected.

Case (ii): Trend plus unit root error

Consider an LRM

$$Y_t = \alpha + \beta X_t + u_t \tag{7}$$

where $X_t = t$ for $t = 1,\cdots,n$ and random walk $u_t = \sum_{i=1}^{t} \epsilon_i$ with $E\epsilon = 0$ and $E\epsilon^2 = \sigma^2$. LSE for model (7) is given by

$$\begin{aligned}\hat{\beta} &= \sum_{i=1}^{n-1}(X_i - \bar{X})(Y_i - \bar{Y}) / \sum_{i=1}^{n-1}(X_i - \bar{X})^2 \\ &= \beta + \frac{12}{n(n^2-1)} \sum_{i=1}^{n}(i - (n+1)/2)u_i. \end{aligned}$$

Note that $E(u_t u_s) = s\sigma^2$ if $s \leq t$. When using this, we have

$$E(\hat{\beta}) = \beta \text{ and } Var(\hat{\beta}) = (6\sigma^2/5)n^{-1}. \tag{8}$$

Indeed

$$\begin{aligned}Var(\hat{\beta}) &= 12^2\sigma^2/(n(n^2-1))^2 \left[\sum_{i=1}^{n}(i - (n+1)/2)^2 i \right. \\ &\quad \left. + 2\sum_{j=1}^{n}\sum_{i<j}(i - (n+1)/2)(j - (n+1)/2)i \right] \\ &= 12^2\sigma^2/(n(n^2-1))^2 \left[(n+1)^2 n(n-1)/24 \right. \\ &\quad \left. + 2\sum_{j=1}^{n}\sum_{i<j}(i - (n+1)/2)(j - (n+1)/2)i \right] \\ &= (6\sigma^2/5)n^{-1} + o(n^{-1}). \end{aligned}$$

Here note that

$$\sum_{j=1}^{n}\sum_{i<j}^{n} i^2 j = \sum_{j=1}^{n}(2j^4 - 3j^3 + j^2)/6 = n^5/15 + o(n^5)$$

$$(n+1)/2 \sum_{j=1}^{n}\sum_{i<j}^{n} i j = (n+1)/2 \sum_{j=1}^{n}(j^3 - j^2)/2$$

$$= n^5/16 + o(n^5)$$

$$(n+1)/2 \sum_{j=1}^{n} \sum_{i<j}^{n} i^2 = (n+1)/2 \sum_{j=1}^{n} (2j^3 - 3j^2 + j)/6$$
$$= n^5/24 + o(n^5)$$

$$(n+1)^2/4 \sum_{j=1}^{n} \sum_{i<j}^{n} i = (n+1)^2/4 \sum_{j=1}^{n} (j^2 - j)/2$$
$$= n^5/24 + o(n^5).$$

Next consider an alternative SEECM

$$\Delta Y_t = \beta \Delta X_t + \epsilon_t. \quad (9)$$

LSE for model (9) is given by

$$\hat{\beta} = \sum_{i=1}^{n-1} \Delta X_i \Delta Y_i / \sum_{i=1}^{n-1} (\Delta X_i)^2 = \beta + \frac{1}{n-1} \sum_{i=1}^{n-1} \epsilon_i .$$

Then

$$E(\hat{\beta}) = \beta \text{ and } Var(\hat{\beta}) = \sigma^2/(n-1). \quad (10)$$

Comparing (8) and (10), the alternative SEECM performs better than the LRM.

In order to check the validity of this expectation, we conduct a small simulation study. We generate data of size $n = 100$ from the following equation.

$$Y_t = 1.5 + 2X_t + u_t, X_t = t, u_t = \sum_{i=1}^{t} \epsilon_i , \epsilon_t \sim N(0,1)$$

Then we estimate the coefficient of $\beta = 2$ in the LRM and the alternative SEECM with the generated data. This experiment is repeated 100 times to obtain sample mean and sample variance of $\hat{\beta}$. Table 2 reports the simulation results.

The simulation results show that the $Var(\hat{\beta})$ of the alternative SEECM is smaller than that of the LRM as expected.

Case (iii): Stationary X_t plus iid error model

Consider an LRM

$$Y_t = \alpha + \beta X_t + \epsilon_t \quad (11)$$

where X_t is stationary (μ, σ_X^2) and iid error ϵ_t with $E\epsilon = 0$ and $E\epsilon^2 = \sigma^2$. X_t and ϵ_t are independent. LSE for model (11) is given by

$$\hat{\beta} = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) / \sum_{i=1}^{n} (X_i - \bar{X})^2$$
$$= \beta + \sum_{i=1}^{n} (X_i - \bar{X})(\epsilon_i - \bar{\epsilon}) / \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Then

$$E(\hat{\beta}) = \beta \text{ and } Var(\hat{\beta}) = E\left[\sigma^2/(\sum_{i=1}^{n} (X_i - \bar{X})^2)\right]$$
$$= \sigma^2/(n\sigma_X^2) + o(1). \quad (12)$$

We assume here that

$$E\left[n/(\sum_{i=1}^{n} (X_i - \bar{X})^2)\right] \to 1/\sigma_X^2.$$

Also consider an alternative SEECM

$$Y_t - Y_{t-1} = \beta(X_t - X_{t-1}) + \epsilon_t - \epsilon_{t-1}$$
$$= \beta(X_t - X_{t-1}) + \eta_t.$$

This may be rewritten as

$$\Delta Y_t = \beta \Delta X_t + \eta_t. \quad (13)$$

LSE for model (13) is given by

$$\hat{\beta} = \sum_{i=2}^{n} \Delta X_i \Delta Y_i / \sum_{i=2}^{n} (\Delta X_i)^2 = \beta + \frac{\sum_{i=2}^{n} \Delta X_i \eta_i}{\sum_{i=2}^{n} \Delta X_i^2}.$$

Note here that ΔX_i is stationary $(0, \sigma_{\Delta X}^2)$ with $\sigma_{\Delta X}^2 = 2\sigma_X^2 - 2E(X_0 X_1)$ and that $E(\eta_t^2) = 2\sigma^2$, $E(\eta_t \eta_{t+1}) = -\sigma^2$ and $E(\eta_t \eta_{t+k}) = 0$ for $k \geq 2$. Then $E(\hat{\beta}) = \beta$ and

$$Var(\hat{\beta}) = n\sigma^2(2\sigma_{\Delta X}^2 - 2Cov(\Delta X_2, \Delta X_1))/(n\sigma_{\Delta X}^2)^2 + o(1)$$
$$= \sigma^2(1 - Corr(\Delta X_2, \Delta X_1))/[n\sigma_X^2(1 - Corr(X_0, X_1))]. \quad (14)$$

Note that

$$E(\sum_{i=2}^{n} \Delta X_i \eta_i)^2 = \sum_{i=2}^{n} E(\Delta X_i \eta_i)^2 + \sum_{i=2}^{n} \sum_{j \neq i} E(\Delta X_i \eta_i \Delta X_j \eta_j)$$
$$= \sum_{i=2}^{n} E(\Delta X_i)^2 E\eta_i^2 + \sum_{i=2}^{n} \sum_{j \neq i} E(\Delta X_i \Delta X_j) E(\eta_i \eta_j)$$
$$= 2(n-1)\sigma^2 \sigma_{\Delta X}^2 + 2 \sum_{i=2}^{n} E(\Delta X_i \Delta X_{i+1}) E(\eta_i \eta_{i+1})$$
$$= 2(n-1)\sigma^2 \sigma_{\Delta X}^2 - 2(n-1)Cov(\Delta X_2, \Delta X_1)\sigma^2$$

Table 2. This table reports simulation results for trend plus unit error model (Case (ii)).

	LRM	*AlternativeSEECM*
Mean($\hat{\beta}$)	1.9913	1.9964
Var($\hat{\beta}$)	0.0138	0.0110

Table 3. This table reports simulation results for stationary X_t plus iid error (Case (iii)).

	LRM	*AlternativeSEECM*
Mean$(\hat{\beta})$	2.0142	2.0152
Var$(\hat{\beta})$	0.0110	0.0182

and that

$Cov(\Delta X_2, \Delta X_1) = 2E(X_1X_0) - E(X_i^2) - E(X_2X_0)$ and $\sigma_{\Delta X}^2 = 2[\sigma_X^2 - E(X_0X_1)]$.

If $E(X_0X_j) = \gamma_j$, we have

$$Var(\hat{\beta}) = \sigma^2(n\sigma_X^2)^{-1}(3 - 4\gamma_1 + \gamma_2)/[2(1-\gamma_1)^2] + o(1) \tag{15}$$

Thus comparing (12) and (15), the alternative SEECM performs better than the LRM if

$$2\gamma_1^2 - \gamma_2 - 1 > 0.$$

For instance, if $\gamma_1 > 1/\sqrt{2}$ and $\gamma_2 = 0$, then the alternative SEECM performs better than the LRM. If X_t's are iid with $\gamma_1 = \gamma_2 = 0$, the LRM performs better than the alternative SEECM.

In order to check the validity of the above calculation, we conduct a small simulation study. We generate data of size n = 100 via the following equation.

$$Y_t = 1.5 + 2X_t + u_t, X_t \sim N(0,1), \epsilon_t \sim iidN(0,1).$$

Then we estimate the coefficient of $\beta = 2$ in the LRM and the alternative SEECM with the generated data. This experiment is repeated 100 times to calculate sample mean and sample variance of $\hat{\beta}$. Table 3 reports the simulation results.

The simulation results show that $Var(\hat{\beta})$ of the LRM is smaller than that of the alternative SEECM, which is expected in a certain condition.

In sum, alternative SEECM is preferable when trend plus random walk error (Case (ii)) and LRM performs better than alternative SEECM when trend plus iid error. On the contrary, LRM or alternative SEECM had better be employed depending on correlation condition of γ_1 and γ_2 (Case (iii)).

References

1. Davidson JD, Hendry D, Srba F, Yeo S. Econometric modelling of the aggregate time series relationship between consumers' expenditures and income in the United Kingdom. Economic Journal 1978;88:661-692.
2. Engle RF, Granger CW. Co-integration and error correction: representation, estimation and testing. Econometrica 1987;55:251-276.
3. Beck N. Comparing dynamic specifications: the case of presidential approval. Political Analysis 1991;3:51-87.
4. Durr RH. An essay on cointegration and error correction models. Political Analysis 1992;4:185-228.