MFE230E Problem Set 4

Due April 21 6:00pm via bCourses

You may NOT use the statsmodels VAR and VECM modules, or any other VAR/VECM package, for this problem set but you may use the statsmodels OLS module.

1. Show that

$$P_t = \mathsf{E}_t \sum_{i=1}^{\infty} \frac{D_{t+i}}{(1+R)^i}$$
$$\implies P_t - \frac{D_t}{R} = \frac{1}{R}\,\mathsf{E}_t \sum_{i=0}^{\infty} \frac{\Delta D_{t+1+i}}{(1+R)^i}.$$

   What does this equation imply for cointegration of prices and dividends?

2. Start from the definition of returns:

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} - 1 \tag{1}$$

   (a) Show that (1) can be written as the approximate identity

$$r_{t+1} \approx k + \rho p_{t+1} + (1-\rho) d_{t+1} - p_t. \tag{2}$$

   (b) Show that (2) implies (ignoring constants)

$$p_t \approx \mathsf{E}_t \sum_{j=0}^{\infty} \rho^j \left[(1-\rho) d_{t+1+j} - r_{t+1+j}\right]$$

   and

$$p_t - d_t \approx E_t \left[\sum_{j=0}^{\infty} \rho^j \left(\Delta d_{t+1+j} - r_{t+1+j}\right)\right].$$

3. Let $\mathbf{X}_t = (x_{1,t}, x_{2,t})'$ and $x_{1,t}, x_{2,t} \sim I(1)$. $x_{1,t}$ and $x_{2,t}$ are cointegrated with cointegration vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$, so that $z_t = \boldsymbol{\alpha}'\mathbf{X}_t \sim I(0)$.

   Consider the VECM(1)

$$\Delta x_{1,t} = \gamma_1\, z_{t-1} + \phi_{11}\Delta x_{1,t-1} + \phi_{12}\Delta x_{2,t-1} + \epsilon_{1,t}$$
$$\Delta x_{2,t} = \gamma_2\, z_{t-1} + \phi_{21}\Delta x_{1,t-1} + \phi_{22}\Delta x_{2,t-1} + \epsilon_{2,t}$$
$$\Leftrightarrow\ \Delta\mathbf{X}_t = \boldsymbol{\gamma}\boldsymbol{\alpha}'\mathbf{X}_{t-1} + \boldsymbol{\Phi}_1\Delta\mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t$$

$$\boldsymbol{\gamma} = \begin{pmatrix}\gamma_1\\ \gamma_2\end{pmatrix},\quad \boldsymbol{\Phi}_1 = \begin{pmatrix}\phi_{11} & \phi_{12}\\ \phi_{21} & \phi_{22}\end{pmatrix}.$$

(a) Write the VECM(1) as a VAR(p):

$$\mathbf{\Psi}(L)\mathbf{X}_t = \boldsymbol{\epsilon}_t$$

Solve for $\mathbf{\Psi}(L)$ in terms of the VECM coefficients $\boldsymbol{\gamma}$ and $\mathbf{\Phi}_1$. What is the lag order p of the VAR?

(b) Now consider the special case $\boldsymbol{\alpha} = (1, -1)'$. What restrictions of the VAR coefficients are implied by the VECM(1)?

4. Next, let's run some simulations. Fix

$$\boldsymbol{\alpha} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \mathbf{\Phi}_1 = \begin{pmatrix} 0.2 & -0.1 \\ 0 & -0.25 \end{pmatrix}$$

for all simulations. Consider four cases for $\boldsymbol{\gamma}$:

$$\begin{aligned} \boldsymbol{\gamma}_1 &= (0, 0.3)' \\ \boldsymbol{\gamma}_2 &= (0, 0.03)' \\ \boldsymbol{\gamma}_3 &= (-0.25, 0.1)' \\ \boldsymbol{\gamma}_4 &= (-0, 0)' \end{aligned}$$

and $T_1 = 250$ and $T_2 = 2500$. For simplicity, let's analyze a single simulation for each combination of $\boldsymbol{\gamma}_i$ and T_j.

Note: Decide for yourself how you might run the simulations in order to answer the questions below (e.g. how many simulations to run in total, ...).

(a) For each combination of $\boldsymbol{\gamma}_i$ and T_j, estimate a

i. VECM(1)

ii. VAR in $\mathbf{X}_t$ with the optimal lag length according to the BIC

iii. VAR in $\Delta\mathbf{X}_t$ with the optimal lag length according to the BIC.

Which estimation specification(s) are most appropriate for the different cases? What are the tradeoffs of the different specifications?

(b) Comment on the estimation results. Are the VAR restrictions that you derived in 3(b) satisfied? Discuss the implications for the cointegration mechanism implied by the different estimation methods. How do $\boldsymbol{\gamma}$ and T affect the results? (You do not have to report complete estimation outputs, just the "important" information.)

(c) Run long-horizon regressions to explore the forecastability of $x_{1,t}$ and $x_{2,t}$ that is a result of cointegration. Select the appropriate horizons for the long-horizon regressions.

(d) Study the serial correlation patterns of the errors of the long-horizon regressions.

(e) Compute different specifications of the variance-covariance matrix. Which variance-covariance matrix/matrices are most appropriate?

5. Download the file `MFE230E_PS4_data.csv` from bCourses. The file contains annual data for S&P500 prices and earnings.

   (a) Test whether log prices and log earnings are stationary or not.

   (b) Construct the log price-earnings ratio $p_t - e_t$ and test whether it is stationary or not.

   (c) Estimate a VECM with one lags and report the results. Describe and interpret the results. Which specification(s) of the variance-covariance of the estimated parameters are appropriate or inappropriate?

   (d) Compute long-horizon regressions for $k = 1, 2, 3, 4, 5$ for the regressions

$$\Delta p_{t+1} + \cdots + \Delta p_{t+k} = \alpha_k + \beta_k (p_t - e_t) + u_{t+k,k}$$
$$\Delta e_{t+1} + \cdots + \Delta e_{t+k} = \alpha_k + \beta_k (p_t - e_t) + u_{t+k,k}$$

   (e) Study the heteroskedasticity and serial correlation patterns of the errors of the long-horizon regressions.

   (f) Compute the standard OLS standard errors as well as the White, Newey-West and Hansen-Hodrick standard error corrections. Which specification is most appropriate?

   (g) Summarize your results. What conclusion can you draw about the behavior of returns and earnings?