Lecture 9. Multivariate Linear Microstructure Models

1. Random-Walk Decomposition

Let $X_t = \left[\begin{array}{cc} \Delta p_t & x_t' \end{array}\right]'$ be the $n \times 1$ stationary multivariate time series. x_t is an $(n-1) \times 1$ vector of additional variables which can help explain the dynamics of the price change Δp_t. By the multivariate-version Wold theorem, (demeaned) X_t has a VMA representation: i.e.,

$$X_t = \varepsilon_t + \Theta_1 \varepsilon_{t-1} + \Theta_2 \varepsilon_{t-2} + \cdots, \tag{1.1}$$

where $\varepsilon_t = \left[\begin{array}{cc} \varepsilon_{\Delta p,t} & \varepsilon_{x,t}' \end{array}\right]'$ is the $n \times 1$ vector of white-noise process, $E[\varepsilon_t] = 0_{n\times 1}$, $Var[\varepsilon_t] = \Omega_{n\times n}$, $Cov[\varepsilon_t, \varepsilon_{t-k}] = 0_{n\times n}$ for all $k \neq 0$, and Θ_j are the $n \times n$ coefficient matrices. From (1.1), one obtains

$$\Delta p_t = \varepsilon_{\Delta p,t} + [\Theta_1]_1 \varepsilon_{t-1} + [\Theta_2]_1 \varepsilon_{t-2} + \cdots,$$

where $[\Theta_j]_1$ denotes the first row of matrix Θ_j.

- The limit of the price forecast $E[p_{t+k}|I_t]$, denoted by f_t^*, is given by

$$\begin{aligned}
f_t^* &= E[p_t + (p_{t+1} - p_t) + (p_{t+2} - p_{t+1}) + (p_{t+3} - p_{t+2}) + \cdots |I_t] \\
&= p_t + E[\Delta p_{t+1}|I_t] + E[\Delta p_{t+2}|I_t] + E[\Delta p_{t+3}|I_t] + \cdots \\
&= p_t + E[\varepsilon_{\Delta p,t+1} + [\Theta_1]_1\varepsilon_t + [\Theta_2]_1\varepsilon_{t-1} + [\Theta_3]_1\varepsilon_{t-2} + \cdots |I_t] \\
&\quad + E[\varepsilon_{\Delta p,t+2} + [\Theta_1]_1\varepsilon_{t+1} + [\Theta_2]_1\varepsilon_t + [\Theta_3]_1\varepsilon_{t-1} + \cdots |I_t] \\
&\quad + E[\varepsilon_{\Delta p,t+3} + [\Theta_1]_1\varepsilon_{t+2} + [\Theta_2]_1\varepsilon_{t+1} + [\Theta_3]_1\varepsilon_t + \cdots |I_t] + \cdots \\
&= p_t + ([\Theta_1]_1\varepsilon_t + [\Theta_2]_1\varepsilon_{t-1} + [\Theta_3]_1\varepsilon_{t-2} + \cdots) + ([\Theta_2]_1\varepsilon_t + [\Theta_3]_1\varepsilon_{t-1} \\
&\quad + [\Theta_4]_1\varepsilon_{t-2} + \cdots) + ([\Theta_3]_1\varepsilon_t + [\Theta_4]_1\varepsilon_{t-1} + [\Theta_5]_1\varepsilon_{t-2} + \cdots) + \cdots \\
&= p_t + \left(\sum_{j=1}^{\infty}[\Theta_j]_1\right)\varepsilon_t + \left(\sum_{j=2}^{\infty}[\Theta_j]_1\right)\varepsilon_{t-1} + \left(\sum_{j=3}^{\infty}[\Theta_j]_1\right)\varepsilon_{t-2} + \cdots.
\end{aligned} \tag{1.2}$$

Consider the random-walk decomposition of the form

$$p_t = m_t + s_t,$$

where $m_t = m_{t-1} + w_t$ and s_t is a zero-mean stationary process. With the restriction $f_t^* = m_t$, (1.2) implies

$$\begin{aligned}
s_t &= p_t - f_t^* \\
&= -\left(\sum_{j=1}^{\infty}[\Theta_j]_1\right)\varepsilon_t - \left(\sum_{j=2}^{\infty}[\Theta_j]_1\right)\varepsilon_{t-1} - \left(\sum_{j=3}^{\infty}[\Theta_j]_1\right)\varepsilon_{t-2} - \cdots,
\end{aligned}$$

which in turn suggests that the pricing error is

$$\sigma_s^2 = \sum_{k=0}^{\infty} C_k \Omega C_k',$$

where $C_k = \sum_{j=k+1}^{\infty} [\Theta_j]_1$.

Remark 1.1. Larger orders should have a greater information content, so it makes sense to think that signed volumes and/or signed concave transformations of volume can explain the dynamics of Δp_t. Letting V_t denote the volume of the trade (in shares or value), therefore, the candidates of x_t would include q_t, $q_t V_t$, and $q_t \sqrt{V_t}$.

2. Multiple Securities and Multiple Prices

2.1. Pairs Trading

There are many situations where prices are closely linked by "common" economic factors. For instance, bid and ask quotes on the same security reflect a single common value plus trade-related costs, so that they may vary in a stationary fashion. No arbitrage also suggests some comovements between prices without boundless divergence; e.g., forward and spot prices, the prices of options and their underlying assets, etc.

- *Pairs trading* is the strategy of identifying two securities with relative prices that are constant in the long run but exhibit considerable variation in the short run. When the relative prices diverge from the long-run value, one buys undervalued security (i.e., the security whose price will increase) and sells for the overvalued (i.e., the security whose price will decrease).

2.2. The Structural and Statistical Models

Consider a structural model of the form

$$m_t = m_{t-1} + u_t \tag{2.1}$$

$$p_{1,t} = m_t + c q_t \tag{2.2}$$

$$p_{2,t} = m_{t-1}, \tag{2.3}$$

where $u_t \sim WN(0, \sigma_u^2)$ and q_t is a trade direction indicator with $E[q_t] = 0, E[q_t^2] = 1, E[q_t q_{t-k}] = 0,$ and $E[u_t q_t] = 0$.

- There is one efficient price m_t that is common to both prices, $p_{1,t}$ and $p_{2,t}$. This model may describe trading in primary and crossing markets in that the second price $p_{2,t}$ is the quote midpoint in the primary market, lagged to reflect transmission and processing delays.

Starting at the fixed value m_0, one obtains from (2.1)

$$
\begin{aligned}
m_1 &= m_0 + u_1 \\
m_2 &= m_1 + u_2 = m_0 + u_2 + u_1 \\
m_3 &= m_2 + u_3 = m_0 + u_3 + u_2 + u_1 \\
&\vdots \\
m_t &= m_0 + u_t + u_{t-1} + \cdots + u_1 .
\end{aligned}
$$

So, the mean and variance of m_t are

$$
\begin{aligned}
E[m_t] &= m_0 \\
Var[m_t] &= \sigma_u^2 t.
\end{aligned}
$$

- From (2.2) and (2.3), it shows that

$$
\begin{aligned}
E[p_{1,t}] &= E[m_t + cq_t] = m_0 \\
Var[p_{1,t}] &= Var[m_t + cq_t] = \sigma_u^2 t + c^2 && (2.4) \\
E[p_{2,t}] &= E[m_{t-1}] = m_0 \\
Var[p_{2,t}] &= Var[m_{t-1}] = \sigma_u^2 t. && (2.5)
\end{aligned}
$$

  Since the variances are time dependent in (2.4) and (2.5), it confirms that $p_{1,t}$ and $p_{2,t}$ are nonstationary. However, the price changes $\Delta p_{1,t}$ and $\Delta p_{2,t}$ are stationary, because

$$
\begin{aligned}
E[\Delta p_{1,t}] &= E[u_t + c(q_t - q_{t-1})] = 0 \\
Var[\Delta p_{1,t}] &= Var[u_t + c(q_t - q_{t-1})] = \sigma_u^2 + 2c^2 \\
E[\Delta p_{2,t}] &= E[u_{t-1}] = 0 \\
Var[\Delta p_{1,t}] &= Var[u_{t-1}] = \sigma_u^2.
\end{aligned}
$$

- Importantly, $p_t = \begin{bmatrix} p_{1,t} & p_{2,t} \end{bmatrix}'$ is cointegrated with $\beta = \begin{bmatrix} 1 & -1 \end{bmatrix}'$; i.e., the linear combination $\beta' p_t = p_{1,t} - p_{2,t} = (m_t + cq_t) - m_{t-1} = u_t + cq_t$ is stationary, since

$$
\begin{aligned}
E[\beta' p_t] &= E[u_t + cq_t] = 0 \\
Var[\beta' p_t] &= Var[u_t + cq_t] = \sigma_u^2 + c^2.
\end{aligned}
$$

Let Δp_t be $\begin{bmatrix} \Delta p_{1,t} & \Delta p_{2,t} \end{bmatrix}'$. Since Δp_t is stationary, the Wold theorem implies that the VMA representation for Δp_t exists.

Proposition 2.1. *The price change vector Δp_t, implied by (2.1), (2.2), and (2.3), follows a*

VMA(1) process. That is, the corresponding statistical representation for Δp_t is given by

$$\Delta p_t = \varepsilon_t + \Theta\varepsilon_{t-1} \tag{2.6}$$

with $\Omega = Var[\varepsilon_t]$.

Proof. The autocovariances of Δp_t are

$$\begin{aligned}
\Gamma_0 &= \begin{bmatrix} Var[\Delta p_{1,t}] & Cov[\Delta p_{1,t}, \Delta p_{2,t}] \\ Cov[\Delta p_{2,t}, \Delta p_{1,t}] & Var[\Delta p_{2,t}] \end{bmatrix} \\
&= \begin{bmatrix} Var[u_t + c(q_t - q_{t-1})] & Cov[u_t + c(q_t - q_{t-1}), u_{t-1}] \\ Cov[u_{t-1}, u_t + c(q_t - q_{t-1})] & Var[u_{t-1}] \end{bmatrix} \\
&= \begin{bmatrix} \sigma_u^2 + 2c^2 & 0 \\ 0 & \sigma_u^2 \end{bmatrix},
\end{aligned} \tag{2.7}$$

$$\begin{aligned}
\Gamma_1 &= \begin{bmatrix} Cov[\Delta p_{1,t}, \Delta p_{1,t-1}] & Cov[\Delta p_{1,t}, \Delta p_{2,t-1}] \\ Cov[\Delta p_{2,t}, \Delta p_{1,t-1}] & Cov[\Delta p_{2,t}, \Delta p_{2,t-1}] \end{bmatrix} \\
&= \begin{bmatrix} Cov[u_t + c(q_t - q_{t-1}), u_{t-1} + c(q_{t-1} - q_{t-2})] & Cov[u_t + c(q_t - q_{t-1}), u_{t-2}] \\ Cov[u_{t-1}, u_{t-1} + c(q_{t-1} - q_{t-2})] & Cov[u_{t-1}, u_{t-2}] \end{bmatrix} \\
&= \begin{bmatrix} -c^2 & 0 \\ \sigma_u^2 & 0 \end{bmatrix},
\end{aligned} \tag{2.8}$$

and $\Gamma_k = 0$ for $|k| > 1$, which suggests that Δp_t follows a VMA(1) process. □

- From (2.6), one computes that

$$\begin{aligned}
\Gamma_0 &= E[(\Delta p_t - E[\Delta p_t])(\Delta p_t - E[\Delta p_t])'] \\
&= E[(\varepsilon_t + \Theta\varepsilon_{t-1})(\varepsilon_t + \Theta\varepsilon_{t-1})'] \\
&= E[\varepsilon_t \varepsilon_t'] + \Theta E[\varepsilon_{t-1}\varepsilon_{t-1}']\Theta' \\
&= \Omega + \Theta\Omega\Theta',
\end{aligned} \tag{2.9}$$

$$\begin{aligned}
\Gamma_1 &= E[(\Delta p_t - E[\Delta p_t])(\Delta p_{t-1} - E[\Delta p_{t-2}])'] \\
&= E[(\varepsilon_t + \Theta\varepsilon_{t-1})(\varepsilon_{t-1} + \Theta\varepsilon_{t-2})'] \\
&= \Theta E[\varepsilon_{t-1}\varepsilon_{t-1}'] \\
&= \Theta\Omega
\end{aligned} \tag{2.10}$$

and $\Gamma_k = 0$ for $|k| > 1$. Notice that equating (2.9) and (2.10) to (2.7) and (2.8) leads to

$$\Theta = (c^2 + \sigma_u^2)^{-1} \begin{bmatrix} -c^2 & c^2 \\ \sigma_u^2 & -\sigma_u^2 \end{bmatrix}. \tag{2.11}$$

Proposition 2.2. *A convergent VAR representation does not exist for the price change vector Δp_t.*

Proof. Corresponding to $\Delta p_t = (I + \Theta L)\varepsilon_t$ in (2.6), one writes a VAR representation as

$$\Phi(L)\Delta p_t = \varepsilon_t$$

with $\Phi(L) = (I + \Theta L)^{-1}$. Notice that $\Theta^2 = -\Theta$ from (2.11). So, the VAR coefficients, $\Phi(L)$, are determined as

$$\begin{aligned} \Phi(L) &= (I + \Theta L)^{-1} \\ &= I - \Theta L + \Theta^2 L^2 - \Theta^3 L^3 + \cdots \\ &= I - \Theta L - \Theta L^2 - \Theta L^3 + \cdots \\ &= I - \Theta(L + L^2 + L^3 + \cdots) \end{aligned}$$

which does not converge. □

Remark 2.3. While no convergent VAR representation is available for Δp_t, the *vector error correction model* (VECM), which is a variant of the VAR model, exists for Δp_t.

2.3. VECM

Suppose that a VAR model does exist for p_t, i.e.,

$$\Psi`(L)p_t = \varepsilon_t. \tag{2.12}$$

With $\Delta p_t = (I - L)p_t$, the VMA representation for $\Delta p_t = (I + \Theta L)\varepsilon_t$ is rewritten as

$$(I - L)p_t = (I + \Theta L)\varepsilon_t. \tag{2.13}$$

- From (2.12) and (2.13), then one obtains

$$\begin{aligned} \Psi(L) &= (I + \Theta L)^{-1}(I - L) \\ &= (I - \Theta L + \Theta^2 L^2 - \Theta^3 L^3 + \cdots)(I - L) \\ &= (I - \Theta L - \Theta L^2 - \Theta L^3 + \cdots)(I - L) \\ &= (I - \Theta L - \Theta L^2 - \Theta L^3 + \cdots) - (L - \Theta L^2 - \Theta L^3 - \Theta L^4 + \cdots) \\ &= I - (\Theta + I)L. \end{aligned}$$

Let $\Psi = \Theta + I$. Since $\Psi(L) = I - \Psi L$, one writes (2.12) as

$$(I - \Psi L)p_t = \varepsilon_t$$

or

$$p_t = \Psi p_{t-1} + \varepsilon_t. \tag{2.14}$$

Using (2.14), one obtains

$$\begin{aligned}
\Delta p_t &= p_t - p_{t-1} \\
&= (\Psi p_{t-1} + \varepsilon_t) - p_{t-1} \\
&= (\Psi - I)p_{t-1} + \varepsilon_t \\
&= \Theta p_{t-1} + \varepsilon_t,
\end{aligned}$$

which in turn indicates that the VECM for Δp_t is given by

$$\begin{aligned}
\Delta p_t &= (c^2 + \sigma_u^2)^{-1} \begin{bmatrix} -c^2 & c^2 \\ \sigma_u^2 & -\sigma_u^2 \end{bmatrix} p_{t-1} + \varepsilon_t \\
&= (c^2 + \sigma_u^2)^{-1} \begin{bmatrix} -c^2 \\ \sigma_u^2 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} p_{t-1} + \varepsilon_t \\
&= \begin{bmatrix} -\frac{c^2}{c^2+\sigma_u^2} \\ \frac{\sigma_u^2}{c^2+\sigma_u^2} \end{bmatrix} \times (p_{1,t-1} - p_{2,t-1}) + \varepsilon_t \\
&= \beta u_{t-1} + \varepsilon_t,
\end{aligned}$$

where $u_{t-1} = p_{1,t-1} - p_{2,t-1}$.

- u_t is the disequilibrium error in the sense of short-term deviation from the long-run equilibrium relationship between $p_{1,t}$ and $p_{2,t}$, and βu_{t-1} represents the error correction term. If $u_{t-1} = 0$ (i.e., $p_{1,t-1} = p_{2,t-1}$), then $E[p_{1,t}] = E[p_{2,t}] = 0$. If $u_{t-1} > 0$ (i.e., $p_{1,t-1} > p_{2,t-1}$), for instance, the error correction term suggests that $E[\Delta p_{1,t}] < 0$ and $E[\Delta p_{2,t}] > 0$; consequently, $p_{1,t-1}$ will decrease to $p_{1,t}$ but $p_{2,t-1}$ will increase to $p_{2,t}$ in order to restore the long-run equilibrium level.

2.4. Cointegration in Practice

In identifying the pair of securities for pairs trading, it is important to estimate and test the cointegrating vector of prices. For estimation one uses Stock and Watson (1993)'s dynamic OLS method. For testing one applies the ADF test to estimated cointegrating residuals. See Reading 7 for details.

- There are two potential problems. First, data-snooping bias may occur, in that the size of a cointegration test (i.e., the probability of incorrectly rejecting the null of no cointegration) is likely to be understated if the price pair is selected from a large number of candidates due to a history of "interesting" comovement. Two, structural breaks in the long-run means of disequilibrium errors can play havoc with the forecasts of cointegrated models. For instance, apparent price divergence on the basis of historical data may actually represent new equilibrium.

3. Discussions

3.1. Timing and Signing

In limit order markets, one sets $q_t = 1$ for a trade priced at the prevailing ask and $q_t = -1$ for a trade priced at the prevailing bid, as all executions occur at posted quotes. The same is not true for hybrid markets in which a significant number of transactions occur somewhere between ask and bid quotes (i.e., price improvement occurs). In this case, the algorithm of Lee and Ready (1990) is useful to classify trades into buy and sells.

- According to Lee and Ready (1990), one sets $q_t = Sign(p_t - m_t^*)$ where m_t^* is the midpoint of the ask and bid quotes, as a trade priced above (or below) the quote midpoint is more likely to be buyer (or seller) initiated.

Remark 3.1. The Lee and Ready's scheme classifies all trades except those that occur at the midpoint of the bid and ask. These trades are classified using the "tick test" in which trades executed at a price higher than the previous trade are called buys and those executed at a lower price are called sells.

3.2. Event Time or Wall-Clock Time

The analysis can be set either in event time (wherein t indexes trades) or in wall-clock time (wherein t indexes fixed time intervals, e.g., minutes). With wall-clock time index, aggregation or transformation is necessary to handle multiple events occurring within the interval.

- The choice depends on the goal of the analysis. If the analysis involves a single security, an event index is better, since the process is more likely to be stationary in event time than in wall-clock time. If the microstructure-based estimates are to be used in a cross-sectional analysis, comparability across securities is important, so that wall-clock time is more appropriate; e.g., volatility per minute is more comparable across firms than volatility per trade.

3.3. Transaction Prices or Quote Midpoints

Quote midpoints have certain advantages over transaction prices. First, m_t^* is active between revisions, while p_t is stale by the next revision. Second, m_t^* reflects current information more reliably than the last-sale price, since a quote setter has a strong motivation to keep bid and ask up to date. Finally, the short-run volatility of m_t^* is lower than that of p_t, since transaction prices are subject to bid-ask bounce.

- If the analysis characterizes execution-related phenomena, like the pricing error or trade execution costs, the actual transaction price p_t is a correct choice. If the study estimates long-run price impact or the contribution of trades to the efficient price variance, then the quote midpoint m_t^* is a sensible alternative choice.