บทที่ 3

ระเบียบวิธีวิจัย

3.1 แบบจำลองที่ใช้ในการศึกษา เพื่อทำการพยากรณ์ราคาหลักทรัพย์

ในการศึกษาเพื่อหาความสัมพันธ์ของการเคลื่อนไหวของราคาหลักทรัพย์โดยใช้แบบจำลอง ARFIMA-FIGARCH ในการศึกษาความสัมพันธ์ถึงราคาปิดของหลักทรัพย์ในปัจจุบันและราคาปิดของหลักทรัพย์ในอดีต

3.2 ข้อมูลที่ใช้ในการศึกษา

ในการศึกษาเป็นข้อมูลทุติยภูมิ (secondary data) ประเภทอนุกรมเวลา (time series data) ใช้ข้อมูลตัวอย่าง (sample period) ของราคาหลักทรัพย์ปิดรายวันในตลาดหลักทรัพย์ของประเทศไทย (SET) ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2548 - 30 เมษายน พ.ศ.2553 เป็นจำนวน 1302 วัน แล้วทำการแปลงข้อมูลให้อยู่ในรูปอัตราผลตอบแทนของราคาหลักทรัพย์

โดยรวบรวมข้อมูลจากฐานข้อมูล Returns2007 จากศูนย์การเงินและการลงทุน (Pinancial Invesment Center: FIC) มหาวิทยาลัยเชียงใหม่ข้อมูลที่นำมาทำการศึกษาคือหลักทรัพย์ในกลุ่มพัฒนาอสังหาริมทรัพย์ จำนวน 3 บริษัท คือ

1. บริษัทแลนด์ แอนด์ เฮาส์ จำกัด (มหาชน) : LH
2. บริษัทควอลิตี้เฮาส์ จำกัด (มหาชน) : QH
3. บริษัทแสนสิริ จำกัด (มหาชน) : SIRI

3.3 วิธีการศึกษาวิเคราะห์

3.3.1 ปรับข้อมูลให้อยู่ในรูปอัตราผลตอบแทนของราคาหลักทรัพย์

ดังสูตรคำนวณต่อไปนี้

$$y_t = \log\left(\frac{P_t}{P_{t-1}}\right)$$

คือ อัตราผลตอบแทนของดัชนีราคาหลักทรัพย์ ณ เวลา t

y_t คือ อัตราผลตอบแทนของดัชนีราคาหลักทรัพย์ ณ เวลา t

P_t คือ ราคาปิดของหลักทรัพย์ที่สนใจในคาบเวลาปัจจุบัน

P_{t-1} คือ ราคาปิดของหลักทรัพย์ที่สนใจในคาบเวลาที่ผ่านมา

3.3.2 การทดสอบความนิ่ง หรือ Unit Root Test

ทำการทดสอบว่าข้อมูลที่จะนำมาศึกษามีความนิ่งหรือไม่ โดยการนำไปทดสอบ Unit Root ซึ่งทำการทดสอบด้วยวิธี Augmented Dickey-Fuller test (ADF) และ Phillips-Perron (PP) ดังสมการต่อไปนี้

1) วิธีของ Dickey-Fuller คือ DF (Dickey-Fuller Test), ADF (Augmented Dickey-Fuller Test) กำหนดในสมการ (3.1)

$$X_t = \rho X_{t-1} + \varepsilon_t \qquad (3.1)$$

โดยกำหนดสมมติฐานหลัก $H_0: \rho = 1$

และสมมติฐานรอง $H_1: |\rho| < 1$

ถ้ายอมรับ H_0 แสดงว่าข้อมูลมีลักษณะไม่นิ่ง แต่ถ้าปฏิเสธ H_0 แสดงว่าข้อมูลนั้นมีลักษณะนิ่ง และจากสมการ (3.1) สามารถแปลงเป็นสมการได้ดังต่อไปนี้คือ

กรณีไม่มีค่าคงที่และแนวโน้มเวลา $\Delta X_t = \theta X_{t-1} + \varepsilon_t$ (3.2)

กรณีมีค่าคงที่ $\Delta X_t = \alpha + \theta X_{t-1} + \varepsilon_t$

(3.3)

กรณีมีทั้งค่าคงที่และแนวโน้มเวลา $\Delta X_t = \alpha + \beta t + \theta X_{t-1} + \varepsilon_t$ (3.4)

โดยกำหนดสมมติฐานหลัก $H_0: \theta = 0$

และสมมติฐานรอง $H_1: \theta < 0$

การยอมรับ H_0 แสดงว่าข้อมูลมีลักษณะไม่นิ่ง แต่ถ้าปฏิเสธ H_0 แสดงว่าข้อมูลนั้นมีลักษณะนิ่ง นอกจากนี้ ถ้าสมการ (3.2) (3.3) และ (3.4) เข้าสู่ Autoregressive Process จะได้สมการดังนี้

กรณีไม่มีค่าคงที่และแนวโน้มเวลา $$\Delta X_t = \theta X_{t-1} + \sum_{i=1}^{p} \phi_i \Delta X_{t-i} + e_t \qquad (3.5)$$

กรณีมีเฉพาะค่าคงที่ $$\Delta X_t = \alpha + \theta X_{t-1} + \sum_{i=1}^{p} \phi_i \Delta X_{t-i} + e_t \qquad (3.6)$$

กรณีมีทั้งค่าคงที่และแนวโน้มเวลา $$\Delta X_t = \alpha + \beta t + \theta X_{t-1} + \sum_{i=1}^{p} \phi_i \Delta X_{t-i} + e_t \qquad (3.7)$$

โดย X_t คือ ราคาหลักทรัพย์ ณ เวลา t

X_{t-1} คือ ราคาหลักทรัพย์ ณ เวลา t-1

$\alpha, \beta, \theta, \phi$ คือ ค่าพารามิเตอร์

t คือ ค่าแนวโน้ม

ε_t คือ ค่าความคลาดเคลื่อนเชิงสุ่มของราคาหลักหลักทรัพย์

โดยมีสมมติฐานคือ

ถ้ายอมรับสมมติฐานหลักแสดงว่า x_t มีลักษณะไม่นิ่ง (Non Stationary) ให้ทำการทดสอบข้อมูลระดับผลต่างลำดับที่ 1 (1^{st} difference) พิจารณาว่ายอมรับสมมติฐานหลักหรือไม่ปฏิเสธสมมติฐานหลักแสดงว่าข้อมูลนั้นมีลักษณะนิ่งแล้ว

2) วิธี Phillips-Perron หรือ PP test พัฒนาวิธีการศึกษา Unit Root จาก Dicey-Fuller โดยมีสมมติฐานเกี่ยวกับการกระจายค่าความคลาดเคลื่อน (Distribution of the errors) ซึ่งทฤษฎีนี้สนับสนุนการทดสอบของ Dickey-Fuller มีสมมติฐานค่าความคลาดเคลื่อนไม่ขึ้นกับค่าสถิติ (Statistically independent) และมีค่าความแปรปรวนคงที่ (Constant Variance) ซึ่งในการใช้วิธีการศึกษานี้ต้องมั่นใจว่า error terms ไม่มีสหสัมพันธ์ (uncorrelated) และมีค่าความแปรปรวนคงที่

โดยพิจารณาสมการ Regression ดังนี้

$$y_t = a_0^* + a_1^* y_{t-1} + \mu_t \qquad (3.8)$$

และ

$$y_t = \tilde{a}_0 + \tilde{a}_1 y_{t-1} + \tilde{a}_2 (t - T/2) + u_t \qquad (3.9)$$

เมื่อ T คือ จำนวนข้อมูล

μ_t คือ distributed term

$E\mu_t = 0$ สมมติฐานของDickey-Fuller นั้น disturbance term ต้องเป็นIndependent และ Homogeneous แต่การทดสอบของ Phillips-Perron กำหนดให้ disturbance เป็น weakly dependent และ Heterogeneous

จากสมการ ที่ (3.9) แทน $\tilde{a}_0$ ด้วย μ แทน $\tilde{a}_0$ ด้วย α และ $\tilde{a}_2$ ด้วย β จะได้

$$y_t = \mu + \beta(t - T/2) + \alpha y_{t-1} + u_t \qquad (3.10)$$

เมื่อ μ, α และ β เป็นค่าสัมประสิทธิ์ของ least-square regression

การทดสอบค่าสัมประสิทธิ์ทางสถิติของสมการถดถอยภายใต้วิธีของ Phillips-Perron นั้นข้อมูลจะถูกสร้างจากสมการ

$$y_t = y_{t-1} + \mu_t$$

ให้ $\tau_\mu, \tau_\alpha, \tau_\beta$ เป็นการทดสอบค่าสถิติของการทดสอบแบบ t(t-statistics) สำหรับสมมติฐานหลัก คือ $\mu = 0, \alpha = 1, \beta = 0$ ซึ่งก็คือ $\tilde{a}_0 = 0, \tilde{a}_1 = 1, \tilde{a}_2 = 0$ ดังนั้นค่าสถิติของ Phillips-Perron คือ

$$Z(t_\alpha) = \left(\tilde{S}/\tilde{\sigma}_{T\omega}\right)t_\alpha - \left(T^3/4\sqrt{3}D_x^{1/2}\tilde{\sigma}_{T\omega}\right)\left(\tilde{\sigma}^2_{T\omega} - \tilde{S}^2\right) \qquad (3.11)$$

$$Z(t_\mu) = \left(\tilde{S}/\tilde{\sigma}_{T\omega}\right)t_\mu - \left(T^3/24D_x^{1/2}E_x\tilde{\sigma}_{T\omega}\right)\left(\tilde{\sigma}^2_{T\omega} - \tilde{S}^2\right)\left(T^{-3/2}\sum y_{t-1}\right) \qquad (3.12)$$

$$Z(t_\beta) = \left(\tilde{S}/\tilde{\sigma}_{T\omega}\right)t_\beta - \left(T^3/2D_x^{1/2}\tilde{\sigma}_{T\omega}\right)^{-1/2}\left[T^{-2}\sum\left(y_{t-1} - \bar{y}_{-1}\right)^2\right]^{-1/2}$$
$$\left(\tilde{\sigma}^2_{T\omega} - \tilde{S}^2\right)\left[(1/2)T^{-3/2}\sum y_{t-1} - T^{-5/2}\sum ty_{t-1}\right] \qquad (3.13)$$

เมื่อ $D_x = \det(X'X)$ คือ determinant ของเมตริกซ์ X

$\tilde{S}$ คือ Standard error ของสมการถดถอย และ ω คือจำนวนการประมาณค่าของ Autocorrelations

$$\tilde{\sigma}^2_{T\varpi} = T^{-1}\sum_1^T u_1^2 + 2T\sum_{s=1}^1 \sum_{t=s+1}^T u_t u_{t=s} \tag{3.14}$$

ทั้ง $\tilde{s}^2$ และ $\tilde{\sigma}^2_{T\varpi}$ เป็นการประมาณค่าของ $\sigma_\mu^2 = \lim_{T\to\alpha} E\left(u_T^2\right)$ และ $\sigma^2 = \lim_{T\to\alpha} E\left(T^{-1}s_T^2\right)$

เมื่อ $s_T = \sum u_t$ และผลรวมทั้งหมดมากกว่า 1
สำหรับ Joint hypothesis $\beta = 0$ และ $\alpha = 1$ ใช้ค่าสถิติ $Z(\Phi_3)$ ดังนี้

$$Z(\Phi_3) = \left(S^2/\sigma_{T\varpi}^2\right)\Phi_3 - \left(1/2\sigma_{T\varpi}^2\right)\left(\sigma_{T\varpi}^2 - S^2\right)\left[T(\alpha-1) - \left(T^6/48D_x\right)\left(\sigma_{T\varpi}^2 - S^2\right)\right] \tag{3.15}$$

ถ้าไม่มีการนำ Trend มาคำนวณในสมการถดถอยด้วย สมมติฐานของ $\alpha = 1$ จะเป็น

$$Z\left(ta_1^{\bullet}\right) = Z\left(t\alpha^*\right) = \left(S/\sigma_{T\varpi}\right)t\alpha^{\bullet} - \left(1/2\sigma_{T\varpi}\right)\left(\sigma_{T\varpi}^2 - S^2\right)\left[T^{-2}\sum\left(y_{t-1} - Y_{-1}\right)^2\right]^{-1/2} \tag{3.16}$$

และ

$$Y_{-1} = T^{-1}\sum_{t=1}^{T} y_{t-1} \tag{3.17}$$

ค่าสถิติที่ใช้ในการทดสอบสมมติฐานคือ

$Z\left(ta_1^*\right)$ ใช้ในการทดสอบสมมติฐานว่า $a_1^* = 1$

$Z\left(t\tilde{a}_1\right) = Z(t\alpha)$ ใช้ในการทดสอบสมมติฐานว่า $\tilde{a}_1 = \alpha = 1$

$Z\left(t\tilde{a}_2\right) = Z(t\beta)$ ใช้ในการทดสอบสมมติฐานว่า $\tilde{a}_2 = \beta = 0$

$Z(\phi_3)$ ใช้ในการทดสอบสมมติฐานว่า $\tilde{a}_1 = 1 = \alpha$ และ $\tilde{a}_2 = 0 = \beta$

ค่าวิกฤติสำหรับค่าสถิติของ Phillips-Perron เปรียบเทียบกับการทดสอบของ Dickey-Fuller ได้ดังนี้ ค่าวิกฤติสำหรับ $Z\left(ta_1^*\right)$ และ $Z\left(t\tilde{a}_1\right)$ คือ τ_μ และ τ_τ ของ Dickey-Fuller และค่าวิกฤติสำหรับ $Z(\phi_{\_})$ คือค่าสถิติ ϕ_3 ของ Dickey-Fuller

3.3.3 ทดสอบ Long memory

ซึ่งก็คือ กระบวนการ Fractionally integrated เป็นการหาค่ากลางระหว่าง I(0) และ I (1) โดยที่ d จะอยู่ในช่วง $-\frac{1}{2}$<d< $\frac{1}{2}$

3.3.4 ประมาณค่าพารามิเตอร์ด้วยแบบจำลอง ARFIMA-FIGARCH

สมการ ARFIMA (P,d_1,Q)-FIGARCH (p,d_2,q)

$$\varphi(L)(1-L)^d(y_t-\mu_t)=\vartheta(L)\varepsilon_t\ ,\ \varepsilon_t|\psi_{t-1}\sim D(0,h_t)\ , \tag{3.18}$$

$$\mu_t=\sum_{j=0}^{r}\gamma_j t^j+\sum_{i=1}^{m/2}(\theta_i\cos\omega_i t+\lambda_i\sin\omega_i t)\ , \tag{3.19}$$

$$\phi(L)(1-L)^{d_2}\varepsilon_t^2=\alpha_0+\sum_{i=1}^{m/2}(K_i\cos\omega_i t+\tau_i\sin\omega_i t)+[1-\beta(L)]\nu_t\ , \tag{3.20}$$

โดยที่ L = the backshift operator $(L^s\varepsilon_t=\varepsilon_{t-s})$, $\varphi(L)=1-\sum_{j=1}^{P}\varphi_j L^j$, $\vartheta(L)=1+\sum_{j=1}^{Q}\vartheta_j L^j$, $\phi(L)=1-\sum_{j=1}^{q}\phi_j L^j$,$\beta(L)=\sum_{j=1}^{p}\beta_j L^j$, $\nu_t=\varepsilon_t^2-h_t$, รากของพหุนาม $\varphi(L)=0$ และ $\phi(L)=0$ อยู่นอกวงกลมหน่วย (unit circle) , $-1<d_1<0.5$,$0<d_2<1$, $\omega_i=\frac{2\pi i}{m}$,$m=365$

เพื่อให้ ได้ค่า positive variance ของ h_t ที่ปราศจากการยับยั้งการเพิ่มขึ้นของพารามิเตอร์ในสมการความแปรปรวนโดย สูตร ลอการิทึมของความแปรปรวนสามารถ $(\nu_t=\varepsilon_t^2-Inh_t)$ องค์ประกอบเชิงฤดูกาลสามารถเขียนมาในอีกรูปหนึ่งได้คือ

$$\mu_t=\sum_{j=0}^{r}\gamma_t t^j+\sum_{k=1}^{12}c_k m_{kt}\ , \tag{3.21}$$

ที่ m_{kt} คือ ตัวแปร dummy ที่เป็นค่าเฉลี่ยต่อเดือนที่ต่อเนื่อง ในทางเดียวกันเราสามารถอธิบายตามส่วนประกอบเชิงฤดูกาลสำหรับความแปรปรวน ได้ดังนี้ (Piotr Fiszede,2008)

$$\phi(L)(1-L)^{d}\,\varepsilon_t^2 = \alpha_0 + \sum_{k=1}^{12}\delta_k m_{kt} + [1-\beta(L)]v_t \qquad (3.22)$$

3.3.5 หาค่า log-likelihood, Akaike Information criteria (AIC), Schwartz Information criteria (SIC)

เพื่อหาแบบจำลองที่ดีที่สุด โดยพิจารณาจากค่า Akaike Information Criterion (AIC) และ Schwartz Criterion (SC) รูปแบบของแบบจำลองที่ให้ค่า AIC และ SC น้อยที่สุดจะเป็นรูปแบบที่ดีที่สุด โดย Akaike Information Criterion (AIC) สามารถคำนวณได้ดังนี้

Akaike Information Criterion (AIC) $\quad -2\iota/\eta + 2k/\eta$

Schwartz Criterion (SC) $\quad -2\iota/\eta + k\log\eta/\eta$

โดยที่ k เป็นจำนวนของพารามิเตอร์ที่ทำการประมาณค่า

η เป็นจำนวนของค่าสังเกต

ι เป็นค่าของ log likelihood function ที่ใช้พารามิเตอร์ที่ถูกประมาณค่า k ตัว

3.3.6 ทดสอบหาค่าความคลาดเคลื่อนร้อยละเฉลี่ย (The Mean Absolute Percentage Error: MAPE)

สูตรของ MAPE คือ

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\mathrm{A}_t - F_t}{\mathrm{A}_t}\right|$$

เมื่อให้ A_t คือค่าที่แท้จริง และ F_t คือค่าที่คาดการณ์ไว้

ค่าความแตกต่างระหว่าง A_t และ F_t ได้แสดงให้เห็นโดย ค่าที่แท้จริง A_t อีกครั้ง ค่าสัมบูรณ์ที่ได้จากการคำนวณนี้ คือผลบวกของทุกๆ จุดที่เหมาะสมหรือทุกๆ จุดที่คาดการณ์ ไว้ในห้วงเวลา และได้แสดงให้เห็นอีกครั้งด้วยจำนวนบนจุดที่เหมาะสม n ทำให้เกิดข้อผิดพลาดด้านเปอร์เซ็นต์ ดังนั้นจึงสามารถเปรียบเทียบข้อผิดพลาดของ time series ที่เหมาะสมซึ่งมีความแตกต่างกัน

ถ้าค่า MAPE น้อยกว่า 10% แสดงว่า มีความแม่นยำสูงมากในการคาดการณ์ ถ้าค่า MAPE อยู่ระหว่าง 10-20% แสดงว่าการคาดการณ์อยู่ในระดับดี ถ้าค่า MAPE อยู่ระหว่าง 20-50%แสดงว่าการคาดการณ์อยู่ใน ระดับพอสมควร ถ้าค่า MAPE มากกว่า 50% ขึ้นไป แสดงว่าเกิดความคลาดเคลื่อน (ผิดพลาด)ในการคาดการณ์