บทที่ 3

กรอบทฤษฎีและระเบียบวิธีวิจัย

3.1 ทฤษฎีที่เกี่ยวข้อง

วิธีการวิเคราะห์แบบ Box - Jenkins เป็นวิธีที่นิยมใช้ในการพยากรณ์ข้อมูลที่เป็นอนุกรมเวลา เนื่องจากมีความแม่นยำและเหมาะสมในการพยากรณ์ข้อมูลระยะสั้นในอนาคตซึ่งให้ค่าพยากรณ์ที่ใกล้เคียงกับความเป็นจริง เป็นที่ยอมรับอย่างกว้างขวางว่าเมื่อใดที่ผู้วิจัยต้องการศึกษาผลของข้อมูลที่จะเกิดขึ้นในอนาคต วิธีการวิเคราะห์แบบ Box - Jenkins มักถูกนำมาใช้ร่วมในการวิเคราะห์เสมอ ดังนั้นในการศึกษาการพยากรณ์ราคาน้ำมันดิบจึงนำวิธีการวิเคราะห์แบบ Box - Jenkins เป็นเครื่องมือในการศึกษา รวมทั้งใช้วิธีการพรรณนา วิธีการวิเคราะห์เชิงปริมาณร่วมในการอธิบายผลการศึกษาครั้งนี้ด้วย

3.1.1 แนวคิดการพยากรณ์อนุกรมเวลา

อนุกรมเวลาเป็นค่าสังเกตของกลุ่มหรือชุดข้อมูลที่มีขึ้น ณ ช่วงเวลาต่างๆ อนุกรมเวลานั้น สามรถแบ่งได้เป็น 2 ลักษณะ คือ อนุกรมเวลาต่อเนื่อง เป็นค่าสังเกตที่กระทำในเวลาที่ต่อเนื่องกันและอนุกรมเวลาที่ไม่ต่อเนื่องเป็นค่าสังเกตที่กระทำ ณ จุดเวลาใดเวลาหนึ่งที่ไม่มีความต่อเนื่องกัน การวิเคราะห์อนุกรมเวลาเป็นการวิเคราะห์หาค่าสังเกตที่มีการเปลี่ยนแปลงไปตามค่าเวลาซึ่งการเปลี่ยนแปลงมีทั้งการเปลี่ยนแปลงที่ไม่มีรูปแบบและการเปลี่ยนแปลงที่มีรูปแบบ ซึ่งการเปลี่ยนแปลงอย่างหลังนี้สามารถที่จะนำมาใช้ในการพยากรณ์รูปแบบอนุกรมเวลาในอนาคตได้ (ทรงศักดิ์ ศรีบุญจิตต์ และอารี วิบูลย์พงศ์)

การศึกษาอนุกรมเวลาของสินค้ารายเดือน ณ ตลาดต่างๆ ต้องกำหนดรูปแบบให้กับอนุกรมเวลา เนื่องจากมีความจำเป็นที่จะต้องพิจารณาความมีเสถียรภาพและลักษณะรูปแบบพฤติกรรมราคาว่ามีอิทธิพลของฤดูกาลและแนวโน้มเข้ามาเกี่ยวข้องกับอนุกรมเวลาที่ต้องการศึกษาหรือไม่ เมื่อได้รูปแบบที่อยู่ในรูปแบบของสมการมาทำการการวิเคราะห์อนุกรมเวลา เพื่อทำนายราคาสินค้าที่ต้องการศึกษาไปล่วงหน้า เพื่อนำค่าการพยากรณ์ที่ได้มาเป็นประโยชน์ตามวัตถุประสงค์ของการศึกษาที่ต้องการนำไปใช้ต่อไป

3.1.2 การทดสอบความนิ่งของข้อมูลอนุกรมเวลา (Unit Root Test)

ในการศึกษาเชิงประจักษ์ที่อาศัยข้อมูลอนุกรมเวลา นั้นมีข้อสมมุติว่าอนุกรมเวลานั้นจะต้องมีลักษณะนิ่ง ซึ่งคำว่านิ่งสามารถอธิบายได้ดังนี้

กระบวนการเฟ้นสุ่ม (stochastic process) จะถูกเรียกว่านิ่ง ถ้าค่าเฉลี่ย (mean) และความแปรปรวนของกระบวนการเฟ้นสุ่มดังกล่าวมีค่าคงที่เมื่อเวลาเปลี่ยนไปและค่าของความแปรปรวนระหว่างสองคาบเวลาจะขึ้นอยู่กับระยะทาง (distance) หรือความล่าหรือล้าหลัง (lag) ระหว่างคาบเวลาทั้งสองดังกล่าวเท่านั้น และไม่ขึ้นอยู่กับเวลาที่เกิดขึ้นจริงที่ความแปรปรวนร่วมได้ถูกคำนวณ (Gujarati, 1995)

คำนิยามของคำว่านิ่งของกระบวนการเฟ้นสุ่ม ตามที่นิยามนี้เป็นที่รู้จักกันว่าเป็น weakly stationary stochastic process ซึ่งใช้กันมากในทางปฏิบัติ (Spanos, 1986 and Gujarati, 1995) จากคำนิยามดังกล่าวเราสามารถเขียนคำนิยามนี้ในรูปของสัญลักษณ์ทางคณิตศาสตร์ได้ดังนี้

กระบวนการเฟ้นสุ่ม (x_t) จะถูกเรียกว่า “ นิ่ง ” ถ้า

$$\text{Mean} \quad : E(xt) = \text{constant} = \mu \tag{3.1}$$

$$\text{Variance} : V(x_t) = \text{constant} = \sigma^2 \tag{3.2}$$

$$\text{Covariance} : Cov(x_t, x_{t+k}) = E(x_t - \mu) = \sigma - \mu \tag{3.3}$$

แสดงให้เห็นได้ว่าค่าเฉลี่ยและความแปรปรวนมีค่าคงที่ เมื่อเวลาเปลี่ยนไป ในขณะที่ค่าความแปรปรวนร่วมเกี่ยว (covariance) ระหว่างสองคาบเวลาจะขึ้นอยู่กับช่องว่าง (gap) ระหว่างคาบเวลาเท่านั้น ไม่ได้ขึ้นอยู่กับเวลาที่เกิดขึ้นจริง และถ้าหากเงื่อนไขใดเงื่อนไขหนึ่งไม่เป็นไปตามที่กล่าวมานี้ กระบวนการเฟ้นสุ่มดังกล่าว จะถูกเรียกว่ามีลักษณะไม่นิ่ง(Charemza and Deadman, 1992)

ทรงศักดิ์ ศรีบุญจิตต์ และอารี วิบูลย์พงศ์ (2542) ได้อธิบายว่าการประมาณค่าทางเศรษฐมิติโดยใช้ข้อมูลอนุกรมเวลานั้นมีข้อสมมติเกี่ยวกับความนิ่ง (stationary) ของข้อมูล สมมติว่าแบบจำลองมีลักษณะดังนี้

$$Y_t = \alpha + \beta X_t + u_{1t} \tag{3.4}$$

$$X_t = X_{t-1} + u_{2t} \quad ; \; u_{2t} \sim iid(0, \sigma^2_{u2}) \tag{3.5}$$

โดยที่ u_{2t} เป็นอนุกรมเวลาของตัวแปรสุ่ม (random variables)ที่มีการแจกแจงปกติที่เหมือนกันและเป็นอิสระต่อกัน โดยค่าเฉลี่ย (mean) เท่ากับศูนย์และค่าความแปรปรวน (variance) คงที่ ซึ่งตัวแปร X เป็นแนวเดินเชิงสุ่ม (random walk) และเป็น integrated of order one,I(1) เพราะฉะนั้นตัวแปร Y ก็จะเป็น I(1) ด้วย ตามทฤษฎีเศรษฐมิติแล้วการถดถอยด้วยตัวแปรที่มีลักษณะไม่นิ่ง (non-stationary) ค่าสถิติ t ที่ใช้โดยปกติจะมีการแจกแจงไม่มาตรฐาน (nonstandard distribution)เพราะฉะนั้นการใช้ตารางมาตรฐานสำหรับการทดสอบค่าสถิติแบบที่ใช้โดยทั่วไปอาจทำให้เกิดข้อสรุปที่ผิดพลาดและเกิดการถดถอยที่ไม่ถูกต้อง (spurious regression) (Johnston and Dinardo,1997)

การทดสอบ unit root สามารถทดสอบได้โดยการทดสอบ ADF (Augmented Divkey-Fuller (ADF) test) (Said and Dickey , 1984) และ DF (Dick-Fuller (DF) test) (Dicky and Fuller, 1981) โดยมีสมมติฐานว่าง (null hypothesis) ของการทดสอบ DF (Dick-Fuller (DF) test) คือ $H_0 : \rho = 1$ จากสมการ (3.6)

$$X_t = \rho X_{t-1} + \varepsilon_t \tag{3.6}$$

โดยถ้าการทดสอบพบว่า $|\rho| < 1$ แล้วแสดงว่า X_t จะมีลักษณะนิ่ง แต่ถ้า $\rho = 1$ แล้ว X_t จะมีลักษณะไม่นิ่ง นอกจากวิธีนี้สามารถทดสอบได้อีกทางหนึ่งซึ่งเหมือนกับสมการ (3.6) คือ

$$\Delta X_t = \theta X_{t-1} + \varepsilon_t \tag{3.7}$$

ซึ่งก็คือ $X_t = (1+\theta) X_{t-1} + \varepsilon_t$ ซึ่งก็คือสมการที่ (3.6) โดยที่ $\rho = (1+\theta)$ ถ้าค่า θ ในสมการ (3.7) มีค่าเป็นลบ จะได้ว่า ρ ในสมการ (3.6) จะมีค่าน้อยกว่า 1 ดังนั้นสามารถสรุปได้ว่าการปฏิเสธ $H_0 : \theta = 0$ ซึ่งเป็นการยอมรับ $H_a : \theta < 0$ หมายความว่า $\rho<1$ และ X_t มี integration of order zero (Charemza and Deadman, 1992) นั่นคือ X_t มีลักษณะนิ่งและถ้าไม่สามารถปฏิเสธ $H_0 : \theta = 0$ ได้จะหมายความว่า X_t มีลักษณะไม่นิ่งโดยถ้า X_t มีแนวเดินเชิงสุ่มซึ่งมีความโน้มเอียงทั่วไปรวมอยู่ด้วย(random walk with drift) สามรถเขียนแบบจำลองได้ดังนี้

$$\Delta X_t = \alpha + \theta X_{t-1} + \varepsilon_t \tag{3.8}$$

และมีแนวโน้มตามเวลาเชิงเส้นเขียนแบบจำลองได้ดังนี้

$$\Delta X_t = \alpha + \beta_t + \theta X_{t-1} + \varepsilon_t \quad \textbf{(3.9)}$$

โดยตัวพารามิเตอร์ที่อยู่ในความสนใจในทุกสมการ คือ θ นั่นคือ ถ้า θ=0; X_t จะมี unit root โดยการเปรียบเทียบค่าสถิติ t–statistic ที่คำนวณได้กับค่าที่เหมาะสมที่อยู่ในตาราง Dickey-Fuller (Dickey–Fuller Tables) (Enders, 1995) หรือกับ ค่าวิกฤติ MacKinnon (MacKinnon Critical Values) (Gujarati, 2003)

อย่างไรก็ตามค่าวิกฤติ (critical values) จะไม่เปลี่ยนแปลง ถ้าสมการ (3.7), (3.8), (3.9) ถูกแทนที่โดยกระบวนการเชิงอัตถดถอย (autoregressive processes) ดังรูปสมการต่อไปนี้

$$\Delta X_t = \theta X_{t-1} + \sum_{i=1}^{\rho} \varphi_i \Delta X_{t-i} + \varepsilon_t \quad \textbf{(3.10)}$$

$$\Delta X_t = \alpha + \theta X_{t-1} + \sum_{i=1}^{\rho} \varphi_i \Delta X_{t-i} + \varepsilon_t \quad \textbf{(3.11)}$$

$$\Delta X_t = \alpha + \beta_t + \theta X_{t-1} + \sum_{i=1}^{\rho} \varphi_i \Delta X_{t-i} + \varepsilon_t \quad \textbf{(3.12)}$$

จำนวนของ lagged difference terms ที่จะนำเข้ามารวมในสมการนั้นจะต้องมีมากพอที่จะทำให้พจน์ค่าความคลาดเคลื่อน (error terms) มีลักษณะเป็น serially independent และเมื่อนำเอาการทดสอบ DF (Dickey – Fuller (DF) test) มาใช้กับสมการ (3.10) ถึง (3.12) เรียกว่าการทดสอบ ADF (augmented Dickey–Fuller (ADF) test) ค่าสถิติทดสอบ ADF (ADF test statistic) มีการแจกแจงเชิงเส้นกำกับ (asymptotic distribution) เหมือนกับสถิติ DF (DF statistic) ดังนั้นสามารถใช้ค่าวิกฤติ (critical values) แบบเดียวกัน (Gujarati, 2003)

3.1.3 การทดสอบความนิ่งของอนุกรมเวลาแบบฤดูกาล (Seasonal Unit Root Test)

ความสำคัญของการนำข้อมูลอนุกรมเวลามาทำการทดสอบ seasonal unit root เนื่องจากข้อมูลอนุกรมเวลาบางชุดมีความไม่นิ่งของฤดูกาลเข้ามาเกี่ยวข้อง ซึ่งถ้าหากนำข้อมูลที่มีความไม่นิ่งของฤดูกาลมาทำการประมาณค่าแล้วอาจทำให้ผลลัพธ์มีความคลาดเคลื่อนเกิดขึ้นได้ ดังนั้นจึงต้องทำการทดสอบแบบฤดูกาลก่อนโดยความนิ่งที่ทดสอบนั้นจะมีด้วยกัน 3 แบบ คือ ความนิ่งแบบมาตรฐาน (seasonal unit root at the zero frequency) ความนิ่งแบบเป็นรายครึ่งปี (unit root at the biannual frequency) และความนิ่งแบบรายไตรมาส (unit root with an quarterly frequency) การทดสอบความไม่นิ่ง มีรูปสมการดังนี้

$$X_{1t-1} = (1 + L + L^2 + L^3)\,X_{t-1} = X_{t-1} + X_{t-2} + X_{t-3} + X_{t-4} \qquad \textbf{(3.13)}$$

$$X_{2t-1} = (1 - L + L^2 - L^3)\,X_{t-1} = X_{t-1} - X_{t-2} + X_{t-3} - X_{t-4} \qquad \textbf{(3.14)}$$

$$X_{3t-1} = (1 - L^2)\,X_{t-1} = X_{t-1} - X_{t-3} \qquad \textbf{(3.15)}$$

$$X_{3t-2} = (L - L^3)\,X_{t-2} = X_{t-2} - X_{t-4} \qquad \textbf{(3.16)}$$

จากรูปแบบของตัวแปรที่สมมติขึ้นทั้ง 4 ตัวแปร คือ X_{1t-1} , X_{2t-1} ,X_{3t-1} และ X_{3t-2} ซึ่งค่าของทั้ง 4 ตัวแปรนี้สามารถหาค่าได้โดยอาศัยสมการ Hylleberge

$$(1 - L^4)X_t = \gamma_1 X_{1t-1} - \gamma_2 X_{2t-1} + \gamma_3 X_{3t-1} - \gamma_4 X_{3t-2} + \varepsilon_t \qquad \textbf{(3.17)}$$

โดยสมมุติฐานว่าง (null hypothesis) ของการทดสอบว่ามีความนิ่งแบบมาตรฐาน $H_0 : \gamma_1=0$ เมื่อทำการทดสอบค่า t-test แล้ว $\gamma_1 =0$ (ยอมรับสมมติฐานว่าง) แสดงว่า $\Delta_4 X_t$ มีลักษณะไม่นิ่งแบบมาตรฐาน สำหรับการทดสอบความนิ่งแบบเป็นรายครึ่งปีกำหนดให้สมมุติฐานว่าง คือ $H_0 : \gamma_2=0$ เมื่อทำการทดสอบค่า t-test แล้วพบว่า $\gamma_2 = 0$ (ยอมรับสมมติฐานว่าง) แสดงว่า $\Delta_4\, X_t$ มีลักษณะไม่นิ่งแบบรายครึ่งปี และสุดท้ายคือการทดสอบความนิ่งแบบรายไตรมาส โดยใช้การทดสอบ F-test สมมุติฐานว่างคือ $\gamma_3 = \gamma_4 = 0$ เมื่อทำการทดสอบแล้วค่า F-test ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติ (ยอมรับสมมติฐานว่าง) แสดงว่า $\Delta_4\, X_t$ มีลักษณะไม่นิ่งแบบรายไตรมาส ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 5% ใช้ค่าทดสอบที่ได้จากตาราง HEGY-test มีดังนี้

ตาราง 3.1 5% critical values for HEGY test

Deterministic terms	Null: Alternative: T	$\gamma_1 = 0$ $\gamma_1 < 0$	$\gamma_2 = 0$ $\gamma_2 < 0$	$\gamma_3 = 0$ $\gamma_3 < 0$	$\gamma_4 = 0$ $\gamma_4 \neq 0$		$\gamma_3 \cap \gamma_4 = 0$ $\gamma_3 \neq 0$ and/or $\gamma_4 \neq 0$
					$\gamma_4 < 0$	$\gamma_4 > 0$	
No intercept	48	-1.95	-1.95	-1.93	-2.11	2.05	3.26
No seasonal dummies	100	-1.97	-1.92	-1.90	-2.01	2.00	3.12
No trend	136	-1.93	-1.94	-1.92	-1.99	1.99	3.14
	200	-1.94	-1.95	-1.92	-1.98	1.97	3.16
Intercept	48	-2.96	-1.95	-1.90	-2.06	2.04	3.04
No seasonal dummies	100	-2.88	-1.95	-1.90	-1.99	1.97	3.08
No trend	136	-2.89	-1.91	-1.88	-1.98	1.97	3.00
	200	-2.87	-1.92	-1.90	-1.98	1.96	3.12
Intercept	48	-3.08	-3.04	-3.61	-2.37	2.35	7.68
Seasonal dummies	100	-2.95	-2.94	-3.44	-2.32	2.29	7.72
Trend	136	-2.94	-2.90	-3.44	-2.31	2.28	7.66
	200	-2.91	-2.89	-3.38	-2.33	2.32	7.53
Intercept	48	-3.56	-1.91	-1.92	-2.05	1.96	2.95
No seasonal dummies	100	-3.47	-1.94	-1.89	-1.97	1.98	2.98
Trend	136	-3.46	-1.96	-1.90	-1.97	1.92	3.04
	200	-3.44	-1.95	-1.92	-1.97	1.96	3.07
Intercept	48	-3.71	-3.08	-3.66	-2.26	2.34	6.55
Seasonal dummies	100	-3.53	-2.94	-3.48	-2.32	2.28	6.60
Trend	136	-3.52	-2.93	-3.44	-2.78	2.31	6.62
	200	-3.49	-2.91	-3.41	-2.27	2.31	6.57

ที่มา: Pattherson (2000: 276)

3.1.4 แนวคิดการวิเคราะห์ข้อมูลอนุกรมเวลา Box-Jenkins

วิธีการวิเคราะห์อนุกรมเวลาของ Box-Jenkins เป็นวิธีการพยากรณ์ข้อมูลอนุกรมเวลาโดยการหารูปแบบที่เหมาะสมให้กับข้อมูลเพื่อใช้กับการพยากรณ์ในระยะช่วงเวลาสั้นๆ การกำหนดรูปแบบที่เหมาะสมให้กับอนุกรมเวลานั้นจะใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบออโต (ACF) และค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนแบบออโต (PACF) ร่วมใช้ในการพิจารณา รูปแบบที่จะกำหนดให้กับอนุกรมเวลาจะเป็นรูปแบบในกลุ่มของ ARIMA(p,d,q) (integrated autoregressive-moving average order p and q) โดยเป็นการรวมของรูปแบบ AR(p) และรูปแบบ MA(q) เข้าด้วยกัน ส่วน d คือ จำนวนครั้งของการหาผลต่าง

รูปแบบ AR(p) หมายถึง รูปแบบที่แสดงว่าค่าสังเกต X_t จะขึ้นอยู่กับค่า X_{t-1} $X_{t-2}, \ldots$ X_{t-p} หรือค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่า

รูปแบบ MA(q) หมายถึง รูปแบบที่แสดงว่าค่าสังเกต X_t จะขึ้นอยู่กับค่าของความคลาดเคลื่อน $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots \varepsilon_{t-p}$

รูปแบบ AR(p) MA(q) ARMA(p,q) และ ARIMA(p,d,q) มีการกำหนดรูปแบบดังนี้

AR(p) คือ $X_t = \theta_0 + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t$

MA(q) คือ $X_t = \theta_0 + \varepsilon_t - \theta_0 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$

ARMA(p,q) คือ $X_t = \theta_0 + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$

ARIMA(p,d,q) คือ $\Delta^d X_t = \theta_0 + \phi_1 \Delta^d X_{t-1} + \ldots + \phi_1 \Delta^d X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - .. - \theta_q \varepsilon_{t-q}$

3.1.5 แบบจำลองการพยากรณ์ โดยวิธี Box-Jenkins

อนุกรมเวลาที่นำมาศึกษาเพื่อการพยากรณ์โดยวิธีของ Box-Jenkins ด้วยวิธี ARIMA ก่อนกำหนดรูปแบบให้กับอนุกรมเวลาที่ไม่นิ่งจะต้องแปลงอนุกรมเวลาดังกล่าวให้นิ่งก่อน ดังนั้นการพิจารณาว่าอนุกรมเวลามีความนิ่งหรือไม่ จะพิจารณาได้จากลักษณะดังต่อไปนี้

1) ค่าเฉลี่ย $E(X_t)$ คงที่สำหรับทุกค่าของ t หรือไม่ ทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆแล้วหาค่าเฉลี่ยของอนุกรมเวลาแต่ละส่วน ถ้าค่าเฉลี่ยแต่ละส่วนย่อยไม่แตกต่างกันมาก สรุปได้ว่า $E(X_t)$ คงที่

2) ค่าความแปรปรวน $V(X_t)$ คงที่ สำหรับทุกๆค่าของ t หรือไม่ทำได้โดยการหาค่าเบี่ยงเบนมาตรฐานของอนุกรมเวลาแต่ละส่วน ถ้าค่าเบี่ยงเบนมาตรฐานในแต่ละส่วนไม่แตกต่างกันมาก สรุปว่า $V(X_t)$ คงที่

3) พิจารณาจากแนวโน้มและ/หรือปัจจัยฤดูกาล ด้วยการพลือตอนุกรมเวลา ในกรณีที่มีแนวโน้มและ/หรือปัจจัยฤดูกาล มักจะเห็นชัดเจนได้จากรูปที่เรียกว่า คอเรลโลแกรม (correlogram)

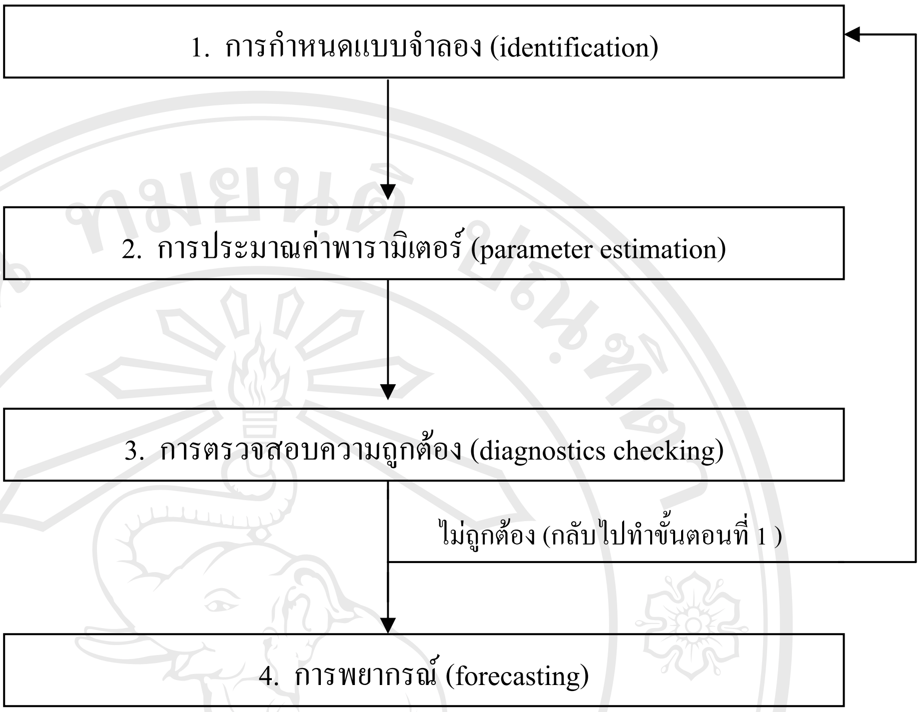
4) พิจารณาจาก correlogram ของค่าสัมประสิทธิ์สหสัมพันธ์แบบออโตของตัวอย่าง (r_k) กรณีที่อนุกรมเวลาเป็นแบบ stationary ค่า correlogram ของสัมประสิทธิ์สหสัมพันธ์ (r_k) จะมีค่าลดลงค่อนข้างเร็ว เมื่อ k มีค่าเพิ่มขึ้นมาก ดังนั้นค่า autocorrelation (r_k) มีค่าลดลงค่อนข้างช้าจะเป็นข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้ม แต่ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ (r_k) มีค่าลดลงค่อนข้างช้า และมีค่าค่อนข้างคงที่สูงที่ k=L ,2L ,3L จะเป็นข้อสังเกตว่าอนุกรมชุดนี้มีแนวโน้ม (trend) และอิทธิพลของฤดูกาล (seasonal) และการเคลื่อนไหวของค่า correlogram ของ autocorrelation (r_k) มีลักษณะคล้ายลูกคลื่น โดยเฉพาะคลื่นจะครบรอบใน 2 ช่วงเวลา แสดงว่าอนุกรมเวลามีอิทธิพลของฤดูกาลเข้ามาเกี่ยวข้อง

เมื่อพิจารณาจากการตรวจสอบแล้วว่าอนุกรมเวลาที่ศึกษามีความไม่นิ่ง จะต้องทำการกำหนดรูปแบบให้กับอนุกรมเวลาที่ไม่นิ่งเสียก่อน โดยการหาผลต่างสำหรับอนุกรมเวลาที่มีแนวโน้ม ถ้าอนุกรมเวลาที่มีอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลจนได้อนุกรมเวลาที่มีความนิ่งถ้าอนุกรมเวลามีทั้งแนวโน้มและอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลจนได้อนุกรมเวลาที่มีความนิ่งแต่ถ้าอนุกรมเวลามีความแปรปรวนไม่คงที่ ให้แปลงอนุกรมเวลาเดิมโดยการหา logarithm $Z_t = \text{In}(X_t)$ จนกว่าจะได้อนุกรมเวลาใหม่ที่มีความแปรปรวนคงที่ จากอนุกรมเวลาใหม่เป็น stationary series แล้วจะทำตามขั้นตอนของ Box-Jenkins ดังนี้

ขั้นตอนการพยากรณ์โดยวิธีของ Box-Jenkins มี 4 ขั้นตอน ดังต่อไปนี้

1. การกำหนดแบบจำลอง (identification)
2. การประมาณค่าพารามิเตอร์ (parameter estimation)
3. การตรวจสอบความถูกต้อง (diagnostic checking)
4. การพยากรณ์ (forecasting)

ขั้นตอนต่างๆ สามารถสรุปเป็นแผนภูมิการวิจัยตามวิธี Box-Jenkins ได้ดังนี้



รูป 3.1 การแสดงขั้นตอนของ Box-Jenkins
ที่มา: Gujarati (2003)

1) การกำหนดแบบจำลอง (identification)

การกำหนดแบบจำลองให้กับอนุกรมเวลาที่เป็น stationary series เป็นการหารูปแบบ ARMA(p,q) ที่คาดว่าจะเหมาะสมให้กับอนุกรมเวลาโดยที่ autocorrelation : P_k คือการวัดความสัมพันธ์ของแต่ละช่วงเวลา โดยมีช่วงที่ย้อนหลังไป k หน่วยเวลา โดยที่ P_k มีค่าอยู่ในช่วง [-1,1] โดยพิจารณาเปรียบเทียบค่า autocorrelation (r_k) ของอนกรมเวลาตัวอย่างกับค่า autocorrelation (P_k) ของอนุกรมเวลาของประชากรที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

$$r_k = \frac{\sum_{t=a}^{n-k}(X_{t-q})(X_{t+k-q})}{\sum_{t=a}^{n}(X_{t-q})^2} \qquad \textbf{(3.18)}$$

โดยที่ $X_t = \sum_{t=a}^{n}(X_t)$

q = จำนวนเวลาสุดท้ายที่ย้อนหลัง

Partial Autocorrelation: (r_{kk}) คือการวัดความสัมพันธ์ของแต่ละช่วงเวลาโดยมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลา โดยพิจารณาเปรียบเทียบค่า partial autocorrelation (r_{kk}) ของอนุกรมเวลาตัวอย่างกับค่า partial autocorrelation (r_{kk}) ของอนุกรมเวลาของประชากรที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

$$r_{kk} = \frac{\sum_{j=1}^{k-1} (r_{k-1})(r_{k-j})}{\sum_{j=1}^{k-1} (r_{k-1,j})(r_j)} \tag{3.19}$$

การพิจารณาเปรียบเทียบแต่ละรูปแบบต้องพิจารณา r_k , r_{kk} กับ P_k และP_{kk} พร้อมกันหลายๆค่า จึงมักจะพิจารณาจากคอเรลโลแกรมที่ได้จากการพล็อต r_k , r_{kk} , P_kและ P_{kk} ในช่วงเวลา k ดังนั้นการพิจารณาเปรียบเทียบจะเป็นการเปรียบเทียบคอเรลโลแกรมของค่า autocorrelation ของอนุกรมเวลาตัวอย่าง (r_k) กับค่า autocorrelation ของอนุกรมเวลาของประชากร (P_k) และคอลเรลโลแกรมของค่า partial autocorrelation ของอนุกรมเวลาตัวอย่าง (P_{kk}) สำหรับแต่ละรูปแบบจะมีคอเรลโลแกรมของ P_k และ P_{kk} ต่างกัน อนุกรมเวลาที่จะนำมากำหนดรูปแบบจะต้องเป็นอนุกรมเวลาที่ stationary เท่านั้น หากไม่เป็น stationary จะต้องแปลงให้เป็น stationary เสียก่อน

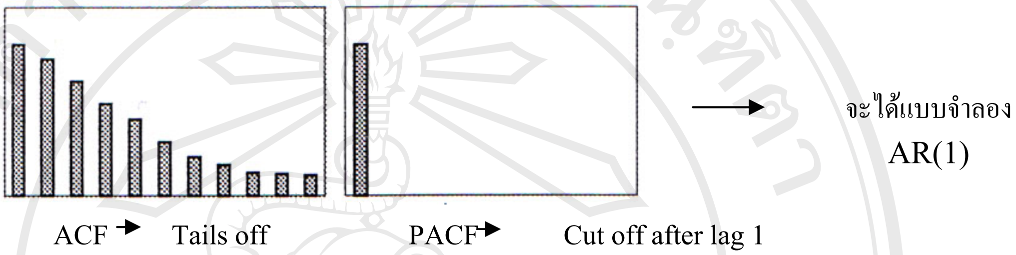
การกำหนดแบบจำลองว่าควรจะมี autoregressive, p เท่าใด differencing, d ที่ลำดับเท่าใดและ moving average, q เท่าใด โดยพิจารณาจาก ACF และ PACF โดยใช้ตาราง 3.1 พิจารณาร่วมในแบบจำลอง

ตาราง 3.2 การพิจารณา ACF และ PACF

ชนิดของแบบจำลอง	รูปแบบของ ACF	รูปแบบของ PACF
AR(p)	ลู่โค้งเข้าหาแกน	เกิดค่าที่ชัดเจนเพียง p ค่าแล้วหายไป
MA(q)	เกิดค่าที่ชัดเจนเพียง q ค่าแล้วหายไป	ลู่โค้งเข้าหาแกน
ARMA(p,q)	ลู่โค้งเข้าหาแกน	ลู่โค้งเข้าหาแกน

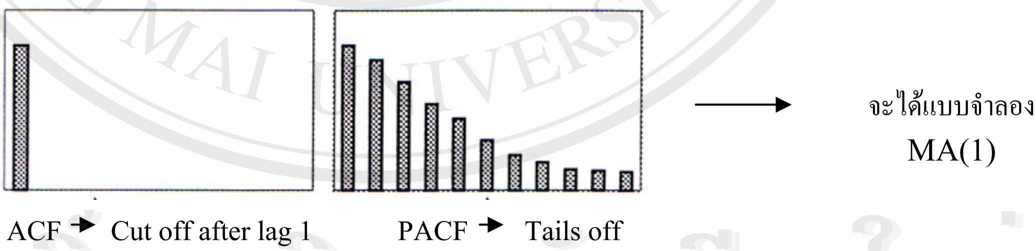
ที่มา : Gujarati (2003)

จากตาราง 3.1 สามารถกำหนดรูปแบบของแบบจำลองได้ดังนี้ หากคอเรลโลแกรมของ ACF มีลักษณะโค้งลู่เข้าหาแกนในระนาบ ขณะที่คอเรลโลแกรม PACF เกิดค่าขึ้นมาไม่กี่ค่าแล้วก็หายไป จำนวนแท่งของค่าที่เกิดขึ้นมาให้นับเป็นค่าที่ p ของ AR(p) เช่น คอเรลโลแกรมของ ACF ที่โค้งลู่เข้าแกนระนาบและ PACF ที่มีแท่งคอเรลโลแกรมเกิดขึ้น 1 แท่ง นั่นคือ แบบจำลองจะมีลักษณะเป็น AR(1) ดังรูป 3.2



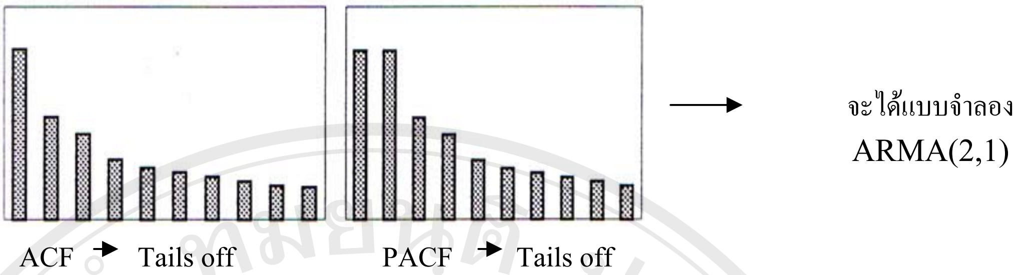
รูป 3.2 ตัวอย่างการเกิดแบบจำลอง AR(1)

สำหรับ ACF เกิดขึ้นมาไม่กี่ค่าแล้วหายไป ในขณะที่ PACF จะลู่โค้งเข้าหาแกนระนาบนั้น เช่น ถ้าค่า ACF เกิดแท่งคอเรลโลแกรมขึ้นเพียง 1 แท่งและหลังจากนั้นก็หายไป ในขณะที่ PACF โค้งลู่เข้าหาแกนระนาบ สรุปได้ว่าแบบจำลองจะมีลักษณะเป็น MA(1)



รูป 3.3 ตัวอย่างการเกิดแบบจำลอง MA(1)

ถ้า ACF และ PACF โค้งเข้าหาแกนระนาบทั้งคู่แบบจำลองจะเป็น ARMA(p,q) และเมื่อรวมกับการทดสอบความนิ่งในขั้นตอนที่ 1 แล้วจะหาค่าของผลต่างได้ ซึ่งผลจากค่าของผลต่างจำนวน d ครั้งนั้นก็จะได้แบบจำลอง ARIMA(p,d,q)



รูป 3.4 ตัวอย่างการเกิดแบบจำลอง ARMA (2,1)

นอกจากการพิจารณา ACF และ PACF ยังมีค่าทางสถิติอื่นที่สำคัญต้องพิจารณาร่วม ได้แก่

ก. ค่า Root Mean Square Error (RMSE) เป็นค่าที่แสดงความคลาดเคลื่อนระหว่างค่าที่ประมาณได้จากแบบจำลองกับค่าข้อมูลจริง ซึ่งถ้าค่า RMSE มีค่าเข้าใกล้ศูนย์แสดงว่าแบบจำลองนี้มีความคลาดเคลื่อนน้อยสามารถใช้เป็นตัวแทนข้อมูลจริงได้อย่างเหมาะสม ซึ่งมีสมการดังนี้

$$\mathrm{RMSE}=\sqrt{\frac{1}{\mathrm{T}}\sum_{t=1}^{t}(\mathrm{X}_t^s-\mathrm{X}_t^a)} \qquad \textbf{(3.20)}$$

โดยที่ X_t^s คือ ค่าประมาณจากแบบจำลอง

X_t^a คือ ค่าที่แท้จริง/ข้อมูลจริง

T คือ จำนวนคาบเวลาที่ใช้

ข. ค่า Theil's Ineqvality Coefficient (U) หลักในการพิจารณาค่าความคลาดเคลื่อนนั้นเหมือนกับค่า RMSE แต่ต่างกันตรงที่ค่า Theil's Ineqvality Coefficient จะมีค่าระหว่าง 0 ถึง 1 ซึ่งถ้าเข้าใกล้ 0 แสดงว่าแบบจำลองนี้สามารถใช้เป็นตัวแทนข้อมูลจริงได้อย่างเหมาะสม ซึ่งมีสมการดังนี้

$$\mathrm{U}=\frac{\sqrt{\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}(\mathrm{X}_t^s-\mathrm{X}_t^a)}}{\sqrt{\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}(\mathrm{X}_t^s)}+\sqrt{\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}(\mathrm{X}_t^a)^2}} \qquad \textbf{(3.21)}$$

โดยที่ X_t^s คือ ค่าประมาณจากแบบจำลอง

X_t^a คือ ค่าที่แท้จริง/ข้อมูลจริง

T คือ จำนวนคาบเวลาที่ใช้

ค. ค่า Adjust R^2 คือ ค่า R^2 ที่ผ่านการขจัดอิทธิพลของตัวแปรอิสระต่างๆออกไปทั้งหมดแล้ว ถ้าหากค่า R^2 ยิ่งมีค่าเข้าใกล้ 100% มากเท่าใดก็อธิบายได้ว่าตัวแปรอิสระสามารถเป็นตัวอธิบายตัวแปรตามได้มากเท่านั้น ซึ่งมีสมการดังนี้

$$\bar{R}^2 = 1 - \frac{\sum \hat{u}_i^2 / (n-k)}{\sum y_i^2 / (n-1)} \quad \textbf{(3.22)}$$

ง. ค่า Akaike's Information Criterion (AIC) มักจะนิยมใช้ในแบบจำลองที่ไม่เป็นเชิงเส้น เป็นค่าสถิติที่อยู่ในรูป natural logarithm ค่าสถิตินี้สามารถนำไปใช้ในการหาค่าย้อนหลัง (lag length) ที่เหมาะสมได้อีกด้วย การพิจารณาค่า AIC นี้ถ้าหากค่า AIC มีค่าน้อยเพียงใดแล้วแสดงว่าแบบจำลองนี้สามารถใช้เป็นตัวแทนข้อมูลจริงได้อย่างเหมาะสม ซึ่งมีรูปแบบสมการดังนี้

$$AIC = \left(\frac{2k}{n}\right) + \log\left(\frac{\sum \hat{u}_i^2}{n}\right) \quad \textbf{(3.23)}$$

โดยที่ $\sum \hat{u}_i^2$ คือ ผลรวมของค่าความคลาดเคลื่อน

n คือ ค่าสังเกตทั้งหมด

จ. ค่า Schwarz's Bayesian Information Criterion (BIC) คือ วิธีวัดปรับได้อย่างดี (goodness of fit) เป็นวิธีประยุกต์ที่คล้ายกับวิธี Akaike's Information Criterion (AIC) การพิจารณาค่า BIC นั้น ถ้าหากว่าค่า BIC ยิ่งมีค่าน้อยมากเท่าใดแล้ว แสดงว่าแบบจำลองนั้นสามารถใช้เป็นตัวแทนของข้อมูลจริงได้อย่างเหมาะสม ซึ่งมีรูปแบบสมการดังนี้

$$\mathrm{BIC}=\log\left(\frac{\sum \hat{u}_i^2}{n}\right)+\frac{2k\log n}{n} \qquad \textbf{(3.24)}$$

2) การประมาณค่าพารามิเตอร์ในรูปแบบ (parameter estimation) โดยทั่วไปจะทำการประมาณ parameter โดยวิธีกำลังสองน้อยที่สุดโดยใช้การวิเคราะห์ตัวเลข (Numerical Analysis) สำหรับค่าประมาณแบบง่ายจะทำโดยการสร้างสมการที่มาจากความสัมพันธ์ระหว่าง P_k และพารามิเตอร์ สมการที่สร้างขึ้นจะมีจำนวนเท่ากับจำนวนพารามิเตอร์ที่ต้องการประมาณ ส่วนค่าประมาณที่ได้จากการวิเคราะห์ตัวเลขจะได้จากการแก้สมการที่สร้างขึ้นจากวิธีกำลังสองน้อยที่สุด ขั้นตอนของการวิเคราะห์ตัวเลขจะต้องมีการกำหนดค่าประมาณเริ่มต้น ซึ่งส่วนใหญ่จะใช้การประมาณแบบง่ายเป็นค่าประมาณเริ่มต้น เมื่อการวิเคราะห์สิ้นสุดจะได้ค่าประมาณสุดท้ายที่จะนำมาใช้ประโยชน์ในการสร้างสมการพยากรณ์ เมื่อพิจารณา ACF , PACF, ค่าสัมประสิทธิ์ของ (autoregressive; AR: p) และ (moving average; MA: q) ซึ่งสามารถใช้วิธีการถดถอยเชิงเส้นอย่างง่าย (simple least square) และวิธีการถดถอยแบบไม่เป็นเชิงเส้นอย่างง่าย (nonlinear) แล้วนำค่าดังกล่าวมาสร้างสมการความสัมพันธ์ที่จะนำไปใช้พยากรณ์และมีความเหมาะสมที่สุด

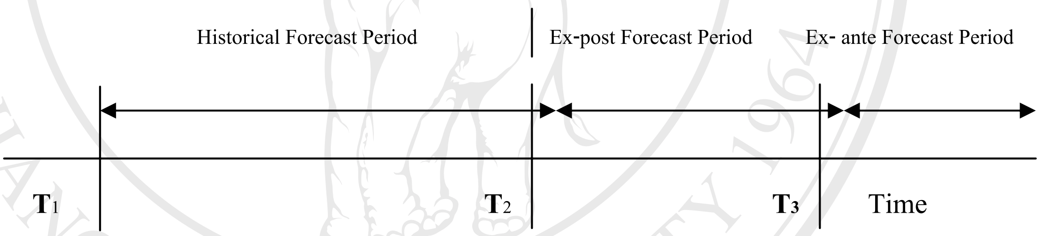
3) การตรวจสอบแบบจำลอง (Diagnostic Checking) เมื่อกำหนดรูปแบบและประมาณค่าพารามิเตอร์ในแบบจำลองจะต้องตรวจสอบทุกครั้งว่ารูปแบบที่กำหนดนั้นมีความเหมาะสมจริงหรือไม่ การตรวจสอบสามารถทำได้หลายวิธีได้แก่ การพิจารณาคอเรลโลแกรมของ r_k หรือของค่าความเคลื่อน การทดสอบค่าพารามิเตอร์ในแบบจำลองด้วยการทดสอบแบบ t และการทดสอบความเหมาะสมของแบบจำลองโดยการทดสอบของ Box และ Pierce หรือการทดสอบของ Box and Ljung หากตรวจสอบพบว่าแบบจำลองนั้นมีความเหมาะสมแล้วจะใช้รูปแบบนั้นในการพยากรณ์ต่อไป แต่หากแบบจำลองนั้นไม่เหมาะสมต้องทำตามขั้นตอนที่ 1 เพื่อกำหนดรูปแบบแบบจำลองใหม่ คือ การตรวจสอบความถูกต้องโดยการนำแบบจำลองที่เลือกไว้มาวาดกราฟเพื่อดูความเหมาะสมกับข้อมูลเบื้องต้นว่า สมการแบบจำลองใดสามารถเป็นตัวแทนข้อมูลเบื้องต้นได้ดีที่สุด นอกจากการพล็อตคอเรลโลแกรมยังสามารถใช้ Box-Pierce ซึ่งใช้ Q-statistic วิเคราะห์ความเหมาะสมของแบบจำลอง ซึ่งมีรูปสมการดังนี้

$$Q=n\sum_{k=1}^{m}\rho_k^2 \qquad \textbf{(3.25)}$$

โดยที่ n คือ จำนวนข้อมูล

M คือ ค่า Lag Lengh

4) การพยากรณ์ (forecasting) การพยากรณ์สามารถทำได้ทั้งแบบจุด (point forecast) และแบบช่วง (interval forecast) ซึ่งควรเป็นพยากรณ์ในช่วงเวลาสั้นๆ เพื่อความเหมาะสมกับเครื่องมือทางเศรษฐมิติ ซึ่งในที่นี้ก็คือ วิธี ARIMA การพยากรณ์โดยวิธี ARIMA นี้ควรเป็นการพยากรณ์ในช่วงระยะเวลาสั้นและการพยากรณ์ข้อมูลไปข้างหน้านั้นจำเป็นต้องใช้แบบจำลองที่สามารถให้ค่าพยากรณ์ที่ถูกต้องแม่นยำที่สุด ดังนั้นการพยากรณ์จึงต้องมีการทดสอบแบบจำลอง โดยแบ่งการพยากรณ์ออกเป็น 3 ช่วง คือ ช่วง historical forecast ใช้ข้อมูลตั้งแต่ T_1 ถึง T_2 ช่วง ex-post forecast ใช้ข้อมูลตั้งแต่ T_2 ถึง T_3 และช่วง ex-ante forecast เป็นการพยากรณ์ในช่วงเวลาอนาคตตั้งแต่ T_3 เป็นต้นไป



รูป 3.5 ช่วงเวลาที่ใช้ในการพยากรณ์

ที่มา: Pindyck and Rubinfeld (1997)

3.2 ระเบียบวิธีวิจัย

วิธีการวิเคราะห์การพยากรณ์ราคาน้ำมันดิบรายเดือนของประเทศไทยด้วยแบบจำลอง ARIMA โดยวิธี Box-Jenkins และทดสอบความนิ่งของข้อมูลโดยใช้วีการทดสอบ unit root ใช้วิธีการพรรณนาและวิธีการวิเคราะห์เชิงปริมาณ ซึ่งมีขั้นตอน ดังต่อไปนี้

3.2.1 การพิจารณาข้อมูลราคานำเข้าน้ำมันดิบของประเทศไทยว่าแนวโน้มของข้อมูลนั้นมีเสถียรภาพหรือไม่

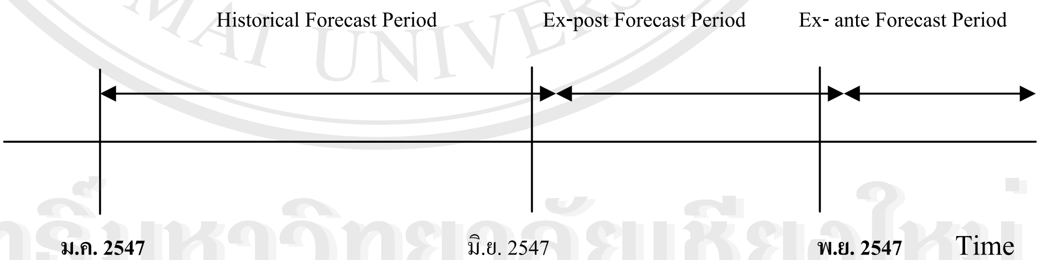
3.2.2 เตรียมทำการทดสอบว่าข้อมูลมีความนิ่งหรือไม่โดยทำการทดสอบ unit root โดยการเปลี่ยนข้อมูลอนุกรมเวลาราคานำเข้าน้ำมันดิบให้อยู่ในรูปลอการิทึมธรรมชาติก่อน

3.2.3 การทดสอบ unit root โดยวิธี Augmented Dickey-Fuller โดยใช้ราคานำเข้าน้ำมันดิบรายเดือนจาก 4 แหล่ง คือ ประเทศโอมาน ประเทศดูไบ แหล่ง Brent และ Forcados

3.2.4 ทำการกำหนดแบบจำลอง ARIMA(p,d,q) โดยการพิจารณาจากคอเรลโลแกรม ค่า ACF และ PACF

3.2.5 การตรวจสอบความถูกต้องซึ่งใช้ค่า Q-statistic วิเคราะห์ความเหมาะสมของแบบจำลอง ว่ามีความเหมาะสมหรือไม่

3.2.6 การพยากรณ์ทำการพยากรณ์ราคาน้ำมันดิบจากทั้ง 4 แหล่ง โดยแบ่งการพยากรณ์ออกเป็น 3 ช่วง คือ historical forecast , ex-post forecast และ ex-ante forecast



รูป 3.6 ช่วงเวลาที่ใช้ในการพยากรณ์ราคาจริง