

Chương 7. Tự tương quan

Autocorrelation

Các giả thiết của mô hình CLRM (nhắc lại)

1. Mô hình là tuyến tính $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$
2. Kỳ vọng U_i bằng 0: $E(u_i | X_{2i}, X_{3i}) = 0$
3. Các U_i thuần nhất: $\text{var}(u_i) = \sigma^2$
4. Không có sự tương quan giữa các U_i : $\text{cov}(u_i, u_j) = 0, i \neq j$
5. Không có quan hệ tuyến tính giữa các biến giải thích.
 $\lambda_1 \mathbf{1} + \lambda_2 X_{2i} + \lambda_3 X_{3i} \neq \mathbf{0},$
 $\forall \lambda_1, \lambda_2, \lambda_3 \neq (0, 0, 0)$

Uncorrelated versus correlated disturbances

Assumption:

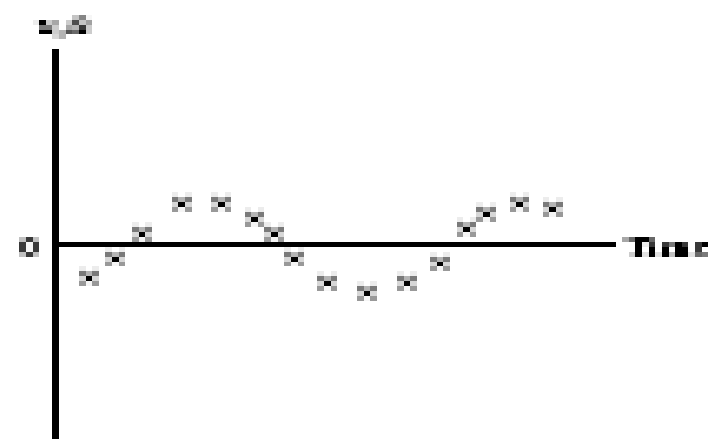
The errors are uncorrelated

$$\text{cov}(u_i, u_j) = E(u_i u_j) = 0$$

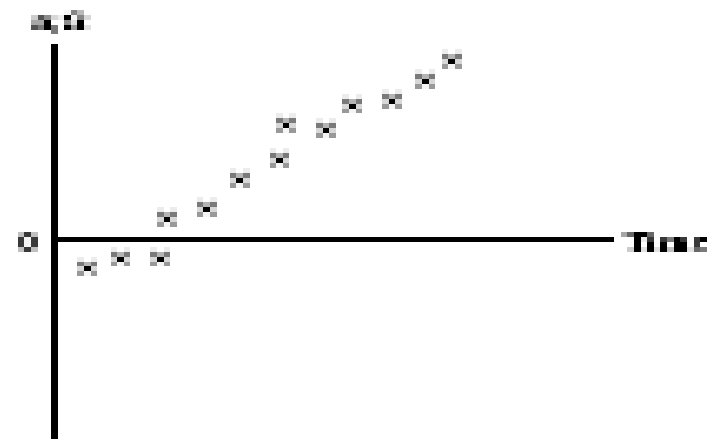
Assumption:

The errors are correlated

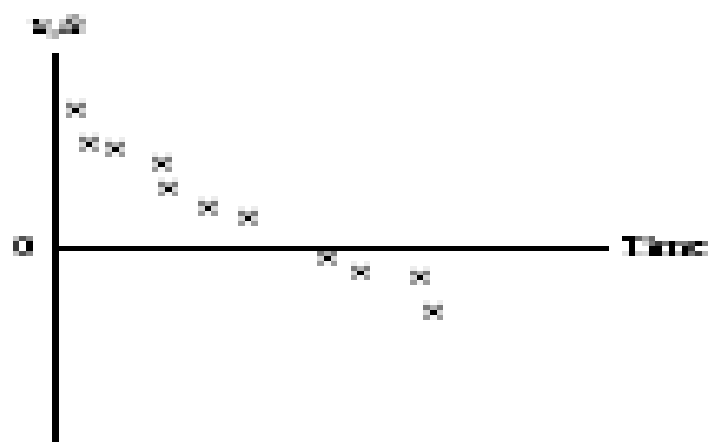
$$\text{cov}(u_i, u_j) = E(u_i u_j) = \sigma_{ij} \neq 0 \text{ for some } i \neq j$$



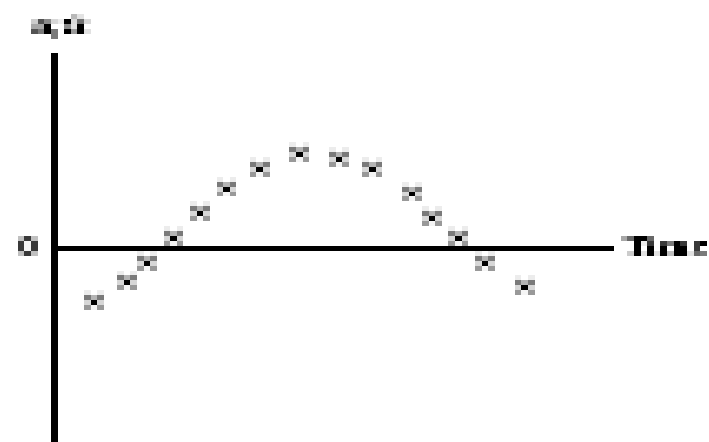
(a)



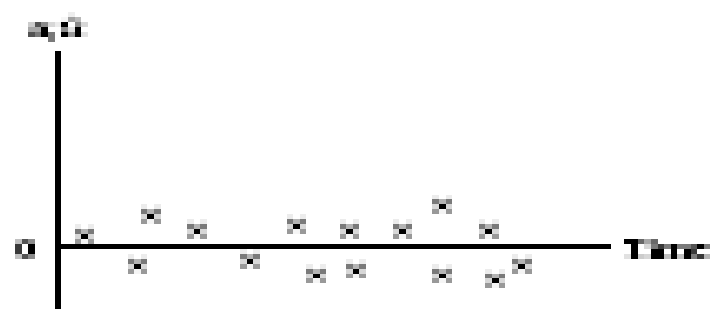
(b)



(c)



(d)



7.1. Nguyên nhân của tự tương quan (TTQ)

- Mô hình chuỗi thời gian thương có tính quán tính
- Hiện tượng mạng nhện
- Trễ
- TTQ có thể xuất hiện vì các vấn đề của mô hình
 - Bỏ sót biến
 - Xử lý số liệu
 - Mô hình định dạng sai

7.2. Ước lượng OLS khi có TTQ

- Xét $Y_t = \beta_1 + \beta_2 X_t + u_t$ với giả thiết $E(u_t, u_{t+s}) \neq 0$ với $s \neq 0$. Như là điểm xuất phát, ta giả thiết nhiễu sinh ra theo cách sau:

$$U_t = \rho u_{t-1} + \varepsilon_t \quad (-1 < \rho < 1) \quad (*)$$

- ρ : Hệ số tự hiệp phương sai
- ε_t : nhiễu ngẫu nhiên thỏa mãn:

$$E(\varepsilon_t) = 0$$

$$\text{var}(\varepsilon_t) = \sigma_\varepsilon^2$$

$$\text{COV}(\varepsilon_t, \varepsilon_{t+s}) = 0 \quad s \neq 0$$

(ε_t thường được gọi là nhiễu trắng)

- Lược đồ (*) gọi là lược đồ tự hồi quy bậc nhất AR(1).
- Lược đồ tự hồi quy bậc hai:

$$U_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t \quad (-1 < \rho_1, \rho_2 < 1)$$

- Khi $|\rho| < 1$ thì AR(1) dừng.
- Chỉ ra được:
 - + $\hat{\beta}_2$ là ước lượng tuyến tính không chệch.
 - + $\hat{\beta}_2$ không còn hiệu quả.

Vậy $\hat{\beta}_2$ không còn là ước lượng tuyến tính không chệch tốt nhất nữa. Ta có thể tìm được BLUE không?

7.3. Ước lượng tuyến tính không chệch tốt nhất khi có TTQ

- Dùng GLS: $Y_t = \beta_1 + \beta_2 X_t + u_t$ với U_t là AR(1)
- Biến đổi:

$$Y_t - \rho Y_{t-1} = \beta_1(1 - \rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t$$

với ε_t là nhiễu trắng.

- Tính được

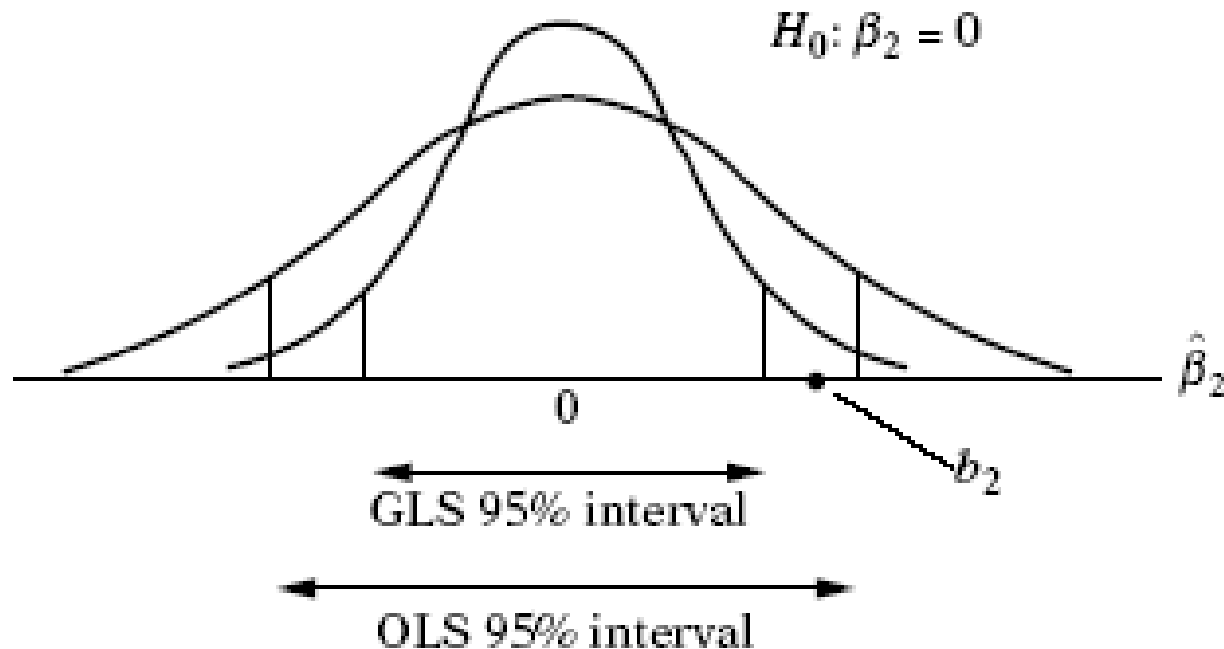
$$\hat{\beta}_2^{\text{GLS}} = \frac{\sum_{t=2}^n (x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{\sum_{t=2}^n (x_t - \rho x_{t-1})^2} + C$$

$$\text{var } \hat{\beta}_2^{\text{GLS}} = \frac{\sigma^2}{\sum_{t=2}^n (x_t - \rho x_{t-1})^2} + D$$

C, D là hằng số điều chỉnh trong thực hành, có thể bỏ qua.

7.4. Hậu quả của việc sử dụng OLS khi có TTQ

- Các ước lượng OLS là LUE, nhưng không hiệu quả nữa.
- Phương sai OLS thường chệch.



- Kđ T và F không đáng tin cậy.
- Ước lượng chênh σ^2 thực, dùng như ước lượng thấp σ^2 .
- R^2 có thể là độ đo không đáng tin cậy.
- Các phương sai và sai số tiêu chuẩn đã tính cũng có thể không hiệu quả.

7.5. Phát hiện có TTQ

- V/đ chính: chúng ta không quan sát được yếu tố ngẫu nhiên (chỉ qs được e_t thu được từ OLS).
- Chúng ta có thể sử dụng phương pháp đồ thị
 - Vẽ đồ thị phần dư theo thời gian
 - Vẽ lược đồ tương quan (và tương quan riêng)
- Kđ TTQ sử dụng:
 - Kđ Durbin-Watson
 - Kđ Breuch-Godfrey
 - Một số kđ khác

A simple idea

- We use OLS to estimate unbiased parameters
- Then we compute the residuals
- Why not look at the regression
$$\hat{u}_t = \rho \hat{u}_{t-1} + v_t$$
- and test if the parameter is significantly different from 0
- For some reason this is not our first choice!

Kđ d-Durbin-Watson

- Kđ nổi tiếng nhất cho TTQ là kđ Durbin-Watson.
- Các giả thiết:
 1. Mô hình hồi quy chứa hệ số chặn
 2. X cố định trong phép lấy mẫu lặp
 3. U_t phân bố chuẩn
 4. U_t là AR(1)
 5. Mô hình không chứa giá trị trễ của biến phụ thuộc. Kđ không áp dụng cho MH sau
$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_k X_{kt} + \gamma Y_{t-1} + u_t$$
 6. Không có các quan sát bị mất trong dữ liệu

- Thống kê Durbin-Watson:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

- Ta có

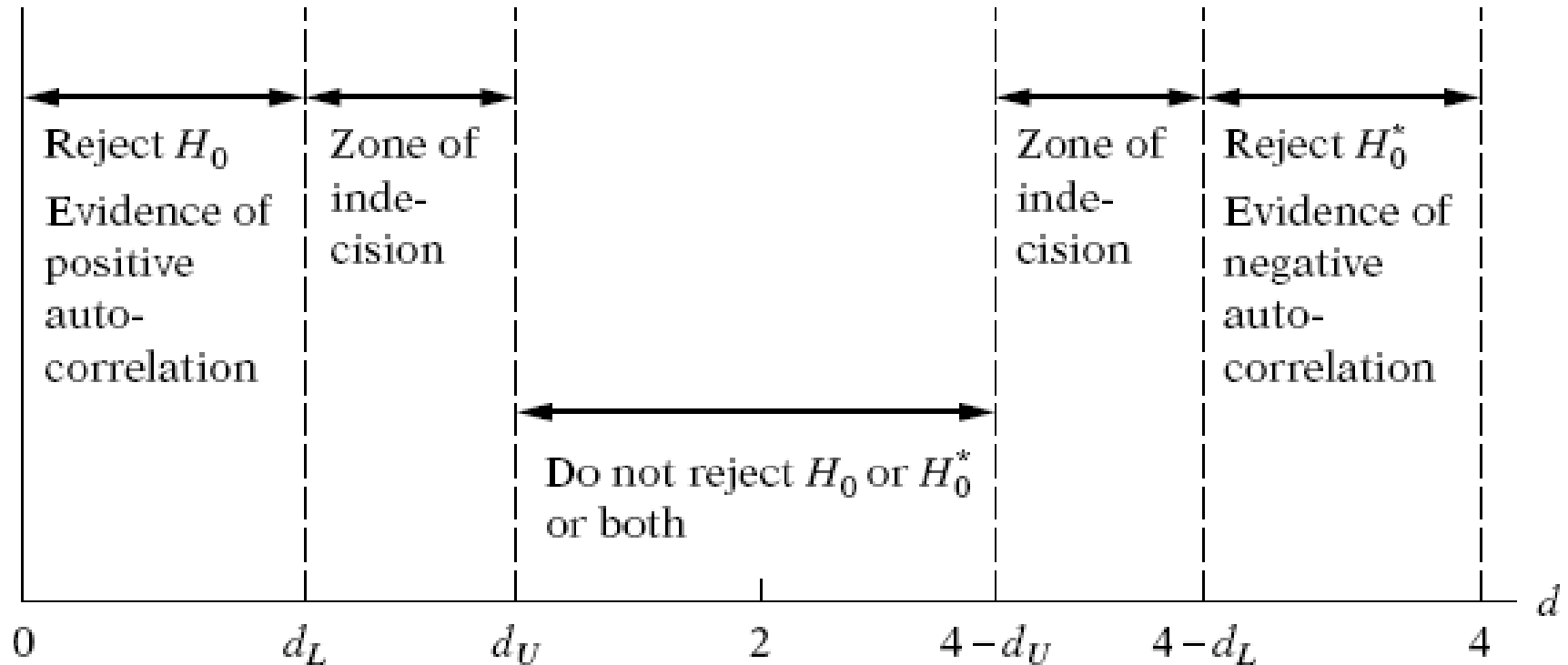
$$d = \frac{\sum_{t=2}^n e_t^2 + \sum_{t=2}^n e_{t-1}^2 - 2 \sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2}$$

$$\approx \frac{2 \sum_{t=2}^n e_t^2 - 2 \sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} \approx 2 \left(1 - \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} \right) \approx 2(1 - \hat{\rho})$$

- $\hat{\rho} = \frac{\sum e_t e_{t-1}}{\sum e_t^2}$ là hệ số tự tương quan bậc nhất của mẫu, đó là ƯL của ρ . Vì $-1 \leq \rho \leq 1$ nên $0 \leq d \leq 4$.

- $\rho=1$ thì $d=0$, $\rho=-1$ thì $d=4$, $\rho=0$ thì $d=2$.

The Durbin-Watson test, III



The critical values d_L and d_U are given in Table D.5A in Gujarati

Legend

H_0 : No positive autocorrelation

H_0^* : No negative autocorrelation

Một số giá trị cận trên và cận dưới của thống kê Durbin-Watson

j	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
65	1.567	1.629	1.536	1.662	1.503	1.696	1.471	1.731	1.438	1.767
70	1.583	1.641	1.554	1.672	1.525	1.703	1.494	1.735	1.464	1.768
75	1.598	1.652	1.571	1.680	1.543	1.709	1.515	1.739	1.487	1.770
80	1.611	1.662	1.586	1.688	1.560	1.715	1.534	1.743	1.507	1.772
85	1.624	1.671	1.600	1.696	1.575	1.721	1.550	1.747	1.525	1.774
90	1.635	1.679	1.612	1.703	1.589	1.726	1.566	1.751	1.542	1.776
95	1.645	1.687	1.623	1.709	1.602	1.732	1.579	1.755	1.557	1.778
100	1.654	1.694	1.634	1.715	1.613	1.736	1.592	1.758	1.571	1.780
150	1.720	1.746	1.706	1.760	1.693	1.774	1.679	1.788	1.665	1.802
200	1.758	1.778	1.748	1.789	1.738	1.799	1.728	1.810	1.718	1.820

Note: n = number of observations; k = number of explanatory variables excluding the constant term.

Adapted from N. E. Savin and K. J. White, "The Durbin-Watson Test for Serial Correlation, with Extreme Small Samples or Many Regressors," *Econometrica*, 45, (Nov 1977): 1989–1996.

The Durbin-Watson Decision Rule once more

<i>One-sided Test Against $\rho > 0$</i>		
$d < d_L$	$d_L < d < d_U$	$d > d_U$
Reject H_0	Inconclusive	Fail to Reject H_0
<i>One-sided Test Against $\rho < 0$</i>		
$d > (4 - d_L)$	$(4 - d_U) < d < (4 - d_L)$	$d < (4 - d_U)$
Reject H_0	Inconclusive	Fail to Reject H_0
<i>Two-sided Test Against $\rho \neq 0$</i>		
$d < d_L$	$d_L < d < d_U$	$d > d_U$
or	or	or
$d > (4 - d_L)$	$(4 - d_U) < d < (4 - d_L)$	$d < (4 - d_U)$
Reject H_0	Inconclusive	Fail to Reject H_0

Kđ Breuch-Godfrey (BG)

- Kiểm định Breuch-Godfrey có tính “tự nhiên” hơn theo nghĩa mà nó sử dụng hồi quy.
- Xét mô hình:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t$$

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + v_t$$

Kiểm định Breuch-Godfrey

1. Sử dụng OLS ước lượng phần dư

$$\hat{u}_t = Y_t - \hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \cdots + \hat{\beta}_k X_{kt}$$

2. Hồi quy phần dư lên các giá trị trễ VÀ biến giải thích gốc thu được R^2

$$\hat{u}_t = \alpha_1 + \alpha_2 X_{2t} + \cdots + \alpha_k X_{kt} + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \cdots + \rho_p \hat{u}_{t-p} + v_t$$

3. Với n đủ lớn, BG chỉ ra: $(n-p)R^2 \sim \chi^2(p)$

Nếu $(n-p)R^2 > \chi^2(p)$ thì H_0 bị bác bỏ.

(Chú ý: Kiểm định F không đúng nữa)

7.6. Các biện pháp khắc phục:

1. Khi cấu trúc TTQ đã biết

Xét mô hình

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad (1)$$

$u_t = \rho u_{t-1} + v_t$ với v_t thỏa mãn các giả thiết OLS

Viết lại mô hình theo $t-1$

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1}$$

Biến đổi mô hình

$$Y_t - \rho Y_{t-1} = \beta_1 + \beta_2 X_t + u_t - \rho \beta_1 - \rho \beta_2 X_{t-1} + \rho u_{t-1}$$

$$Y_t - \rho Y_{t-1} = \beta_1(1 - \rho) + \beta_2(X_t - \rho X_{t-1}) + v_t$$

$$Y_t^* = \beta_1^* + \beta_2 X_t^* + v_t \quad (2)$$

Ước lượng bình phương tổng quát

- Nếu biết ρ thì chúng ta biến đổi dữ liệu về dạng (2) và ước lượng bằng OLS
- Cách làm đó gọi là GLS
- Chú ý:
 - Hồi quy Y^* đối với X^* có hay không có hệ số chặn phụ thuộc vào phương trình gốc có hệ số chặn hay không.
 - Để tránh mất 1 quan sát trong (2), ta lấy qs đầu:
$$Y_1^* = Y_1 \sqrt{1 - \rho^2}, X_1^* = X_1 \sqrt{1 - \rho^2}$$
- Khi ρ chưa biết, chúng ta có thể bắt đầu bằng ước lượng nó và sau đó sử dụng dạng biến đổi (2).

2. Khi ρ chưa biết

a) Phương pháp sai phân cấp 1:

$$\boxed{P=1} \quad Y_t = \beta_1 + \beta_2 X_t + \beta_3 t + u_t \text{ với } u_t \text{ là AR}(1) \quad (1)$$

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + \beta_3 (t-1) + u_{t-1}$$

$$\Delta Y_t = \beta_2 \Delta X_t + \beta_3 + \varepsilon_t \text{ thỏa mãn CLRM} \quad (2)$$

Hệ số chặn của (2) là hệ số biến xu thế của MH gốc.

(2) là phương trình sai phân cấp 1.

$$\boxed{P=-1} \quad \frac{Y_t + Y_{t-1}}{2} = \beta_1 + \beta_2 \frac{X_t + X_{t-1}}{2} + \frac{\varepsilon_t}{2}$$

Là mô hình hồi quy trung bình trượt (2 thời kì)

b) Ước lượng ρ dựa trên thống kê d-Durbin-Watson

- Ta có: $d \approx 2(1 - \hat{\rho})$ hay $\hat{\rho} \approx 1 - \frac{d}{2}$
- Ta ước lượng ρ , sau đó ước lượng dạng sai phân tổng quát (2) như trong mục 1.
- Chú ý:
 - Quan hệ $\hat{\rho} = 1 - \frac{d}{2}$ không đúng trong mẫu nhỏ. Theil và Nagar đã giới thiệu một công thức cho mẫu nhỏ (Bài tập 12.6 Gujarati).
 - Ước lượng này áp dụng cho mẫu lớn.

C) Thủ tục lặp Cochrane-Orcutt để ước lượng ρ

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad (1)$$

Giả sử $u_t = \rho u_{t-1} + \varepsilon_t \quad (2)$

B1: ƯL (1) thu e_i

B2: ƯL (2): $e_t = \rho e_{t-1} + v_t$ thu được $\hat{\rho}$

B3: ƯL $Y_t - \hat{\rho} Y_{t-1} = \beta_1^* + \beta_2^* (X_t - \hat{\rho} X_{t-1}) + e_t^*$
thu được $\hat{\beta}_1^*, \hat{\beta}_2^*$

B4: Tính phần dư mới $e_t^{**} = Y_t - \hat{\beta}_1^* - \hat{\beta}_2^* X_t$

B5: Trở lại bước 2: ƯL $e_t^{**} = \rho e_{t-1}^{**} + w_t$ thu ƯL vòng 2 của ρ là $\hat{\rho}$

Quá trình lặp đến khi các ước lượng kế tiếp nhau của ρ khác nhau một lượng rất nhỏ.

d) Phương pháp Durbin-Watson 2 bước để ước lượng ρ

Xét mô hình $Y_t = \beta_1 + \beta_2 X_t + u_t$ (1)

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1}$$

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2 X_t + \rho \beta_2 X_{t-1} + u_t - \rho u_{t-1}$$

$$Y_t = \beta_1(1-\rho) + \beta_2 X_t - \rho \beta_2 X_{t-1} + \rho Y_{t-1} + \varepsilon_t \quad (2)$$

B1: Xem (2) mô hình hồi quy 4 biến. Hồi quy Y_t theo X_t, X_{t-1}, Y_{t-1} thu $\hat{\rho}$

B2: Biến đổi mô hình về phương trình sai phân tổng quát và ước lượng bằng OLS.

- Chú ý: ước lượng hệ số Y_{t-1} ($= \hat{\rho}$), tuy là ước lượng chệch của ρ nhưng là ước lượng vững của ρ .