2. Time Series Statistics

author: ECONOMIC FORECASTING date: Summer 2017 autosize: true

==

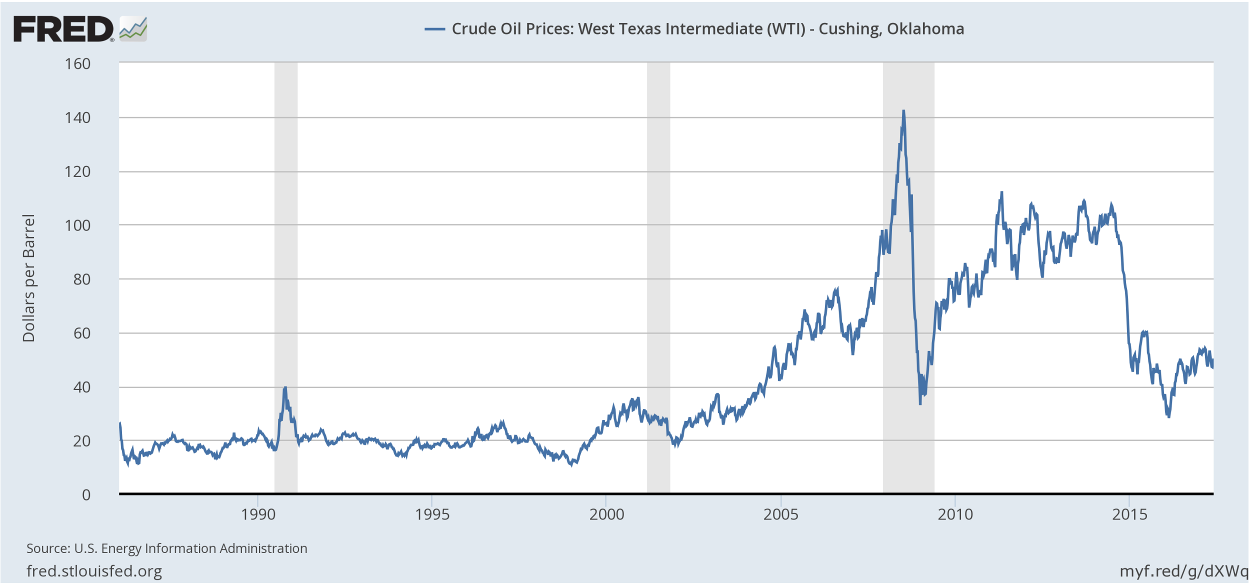

Figure 1: alt text

== There seems to be a relationship between this month's WTI price and last month's WTI price. - This is **serial correlation**.

Is this a good thing or a bad thing? - For everyday living, this is a good thing. - For those wanting to invoke the CLR, this is a *bad* thing.

Definition

Autocorrelation - Correlation between a lagged variable (one or more periods) and itself.

CLR Violation - Cross-sectional data : $\mathbf{C}(\epsilon_i, \epsilon_j) \neq 0$ - Time series data : $\mathbf{C}(\epsilon_t, \epsilon_{t-s}) \neq 0$ - Panel data: $\mathbf{C}(\epsilon_i, \epsilon_j) \neq 0$ and $\mathbf{C}(\epsilon_t, \epsilon_{t-s}) \neq 0$

Causes

Spatial - A random shock affecting local economic activity in one region may spill over into other regions.

Causes

Spatial - A random shock affecting local economic activity in one region may spill over into other regions.

Prolonged Influence - A random shock may have effects that persist more than one time period.

Causes

Spatial - A random shock affecting local economic activity in one region may spill over into other regions.

Prolonged Influence - A random shock may have effects that persist more than one time period.

Intertia - Due to psychological conditioning, past actions often have strong effects on current actions.

Causes Continued

Data Manipulation - Data may have gone through interpolation or a smoothing process.

Misspecification - Omission of a relevant variable - Improper variable form

Omission of a Relevant Independent Variable

True Model: $y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \epsilon_t$

You Regress: $y_t = \alpha + \beta_1 x_{1,t} + \epsilon_t$

This leads to: $y_t = \alpha + \beta_1 x_{1,t} + (\beta_2 + \epsilon_t) \implies$ **y** s are correlated by $\beta_2 + \epsilon_t$

Consequences of OVB

- **Coefficients are biased**
- **Alters the error structure**
  - SEs are smaller than they should be $\implies$ t-stats are inflated $\implies$ inferences are incorrect
  - Reject the null hypothesis that $\beta_1 = 0$
  - Spurious regression/results

Consequences of "pure" Autocorrelation

Coefficients are unbiased

Coefficients are inefficient

SEs of the coefficients are biased (typically downward)

What to Do?

1. Test to Identify
2. Take action
   - Remove autocorrelation
   - Model autocorrelation
3. Action should be determined by what caused the autocorrelation
   - If you caused autocorrelation, $\implies$ remove
   - If autocorrelation is pure, $\implies$ model

Identification

First order autocorrelation can be visually obvious - Long positive/negative error "runs" (positive autocorrelation) - Alternating positive/negative error "runs"" (negative autocorrelation) - Visual inspection can be inconclusive

Enter the **Durbin-Watson** Statistic - Only approriate for first-order autocorrelation

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

Measuring Autocorrelation

Assumptions

1. $\{x_t\}$ is stationary time series.
    - **stationarity** requires $\mu(t) = \mu$ and $\sigma^2(t) = \sigma^2$
        - where μ and σ^2 are constants
2. **Autocovariance** (λ_k):
$$\lambda_k = \mathbf{E}[(x_t - \mu)(x_{t+k} - \mu)]$$
3. **Autocorrelation** (ρ_k):
$$\rho_k = \frac{\lambda_k}{\sigma^2}$$

Sample Statistics

Sample λ_k

$$\mathbf{c_k} = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})$$

Sample ρ_k

$$\mathbf{r_k} = \frac{\mathbf{c_k}}{c_0}$$

In R, use the `acf()` function - For example, the lag 1 autocorrelation for variable x could be: **`acf(x)$acf[2]`**

Correlogram

Visual representation of the autocorrelation function

Confidence intervals for testing $\rho_k = 0$ is a function of n, the number of observations

$$-\frac{1}{n} \pm \frac{2}{\sqrt{n}}$$

since the distribution of r_k is:

$$r_k \sim^a N\left(-\frac{1}{n}, \frac{1}{n}\right)$$

Visual Identification 1a

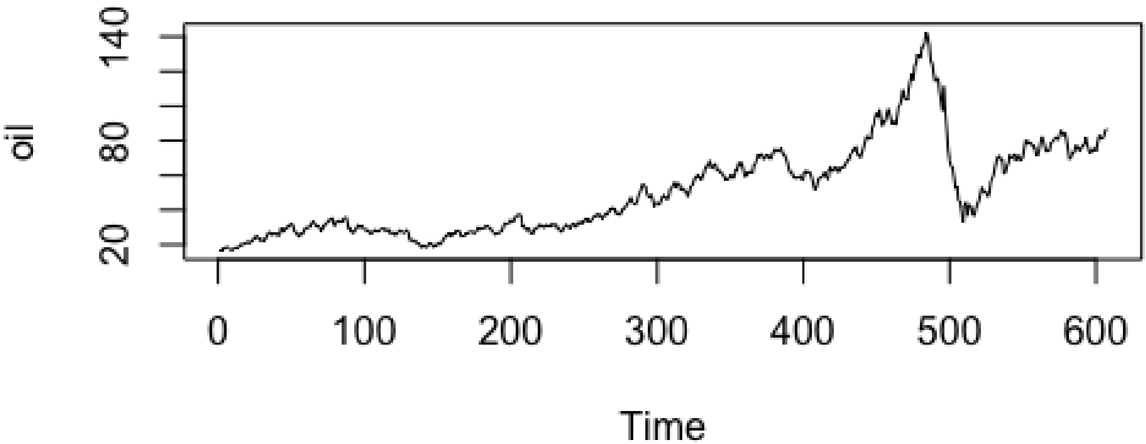

Figure 2: alt text

Visual Identification 1b

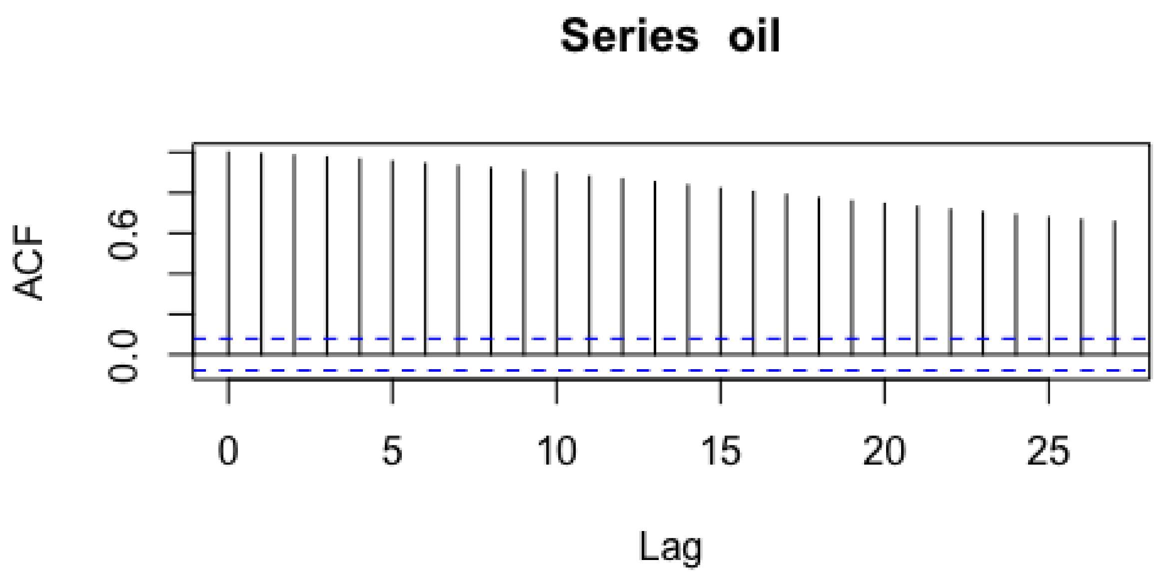

Figure 3: alt text

Visual Identification 2a

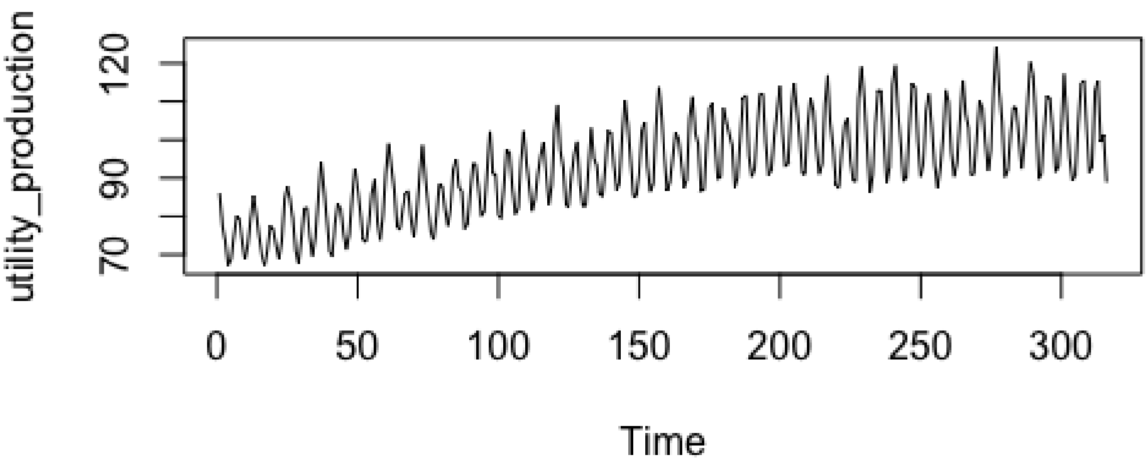

Figure 4: alt text

Visual Identification 2b

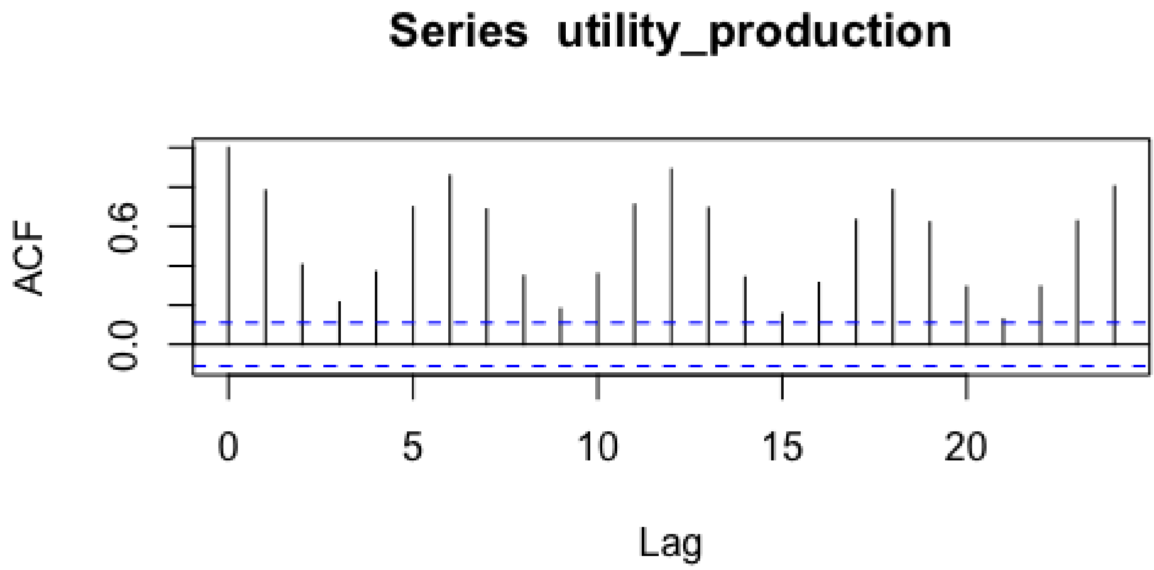

Figure 5: alt text

Durbin-Watson

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

After some math... $DW = 2(1 - \hat{\rho})$

DW	4	2	0
ρ	-1	0	1

	Positive	Negative
H_0	$\rho = 0$	$\rho = 0$
H_1	$\rho > 0$	$\rho < 0$

Need some critical values to make decisions...

Durbin-Watson Critical Values

Positive Correlation - $DW > dU \rightarrow \rho = 0$ - $DW < dL \rightarrow \rho > 0$ - $dL < DW < dU \rightarrow$ inconclusive

Negative Correlation - $DW < 4 - dU \rightarrow \rho = 0$ - $DW > 4 - dL \rightarrow \rho < 0$ - $4 - dL > DW > 4 - dU \rightarrow$ inconclusive

Hypothesis Testing

Case #1 ($\alpha = 0.05$) - $n = 50$ - 4 independent variables - $DW = 0.70$ - Conclusion: Positive Autocorrelation

Case #2 ($\alpha = 0.05$) - $n = 100$ - 2 independent variables - $DW = 1.70$ - Conclusion: Negative Autocorrelation

Notes on Durbin-Watson

1. Only appropriate to test for first-order autocorrelation
2. "Weak" test due to inconclusive ranges
3. Biased towards failing to reject the null hypothesis of no autocorrelation

Ljung-Box Q Statistic

- Useful for testing higher autocorrelation
- NO lagged dependent variables asa regressors

$$H_0\colon\ \rho_1 = 0, \rho_2 = 0, ..., \rho_p = 0$$
$$H_1\colon\ \text{at least one } \rho_\tau \neq 0$$

$$\mathbf{Q} = n(n+2)\sum_{\tau=1}^{p}\hat{\rho}_\tau^2/(n-\tau) \rightarrow^d \chi^2_{(\rho)}$$

$$\hat{\rho_\tau} = \frac{\sum_{t=\tau+1}^{n} e_t - e_{t-\tau}}{\sum_{t=1}^{n} e_t^2} \;\; \tau = 1, 2, ..., p$$

Breusch-Godfrey LM Test

- Useful for autocorrelation of *any* order in models with, or without lagged dependent variables / endogenous regressors

$$H_0\text{: } \rho_1 = 0, \rho_2 = 0, ..., \rho_p = 0$$
$$H_1\text{: some } \rho_\tau \neq 0$$

Step 1: Estimate $y = X\beta + \epsilon$ by OLS and save e_t's

Step 2: Regress e_t on X_t and $e_{t-1}, e_{t-2}, ..., e_{t-p}$

Step 3: $n \cdot R^2 \rightarrow^d \chi^2_{(p)}$

Specification Error

Removing Autocorrelation - Adding an omitted variable + What independent variable do all time series have in common? **Time**

$$y_t = \alpha + \delta t + X\beta + \epsilon_t$$

...extremely crude, but useful.

- Changing Functional Form
  - Example: $ln(y), ln(X)$, more/different X's
    - can be applied to t as well

Removal through Generalized Differences

Transforming the Model

$$\text{Model}: \;\; y_t = \alpha + \beta_1 X_{1,t} + \epsilon_t \tag{1}$$

$$where: \epsilon_t = \rho\epsilon_{t-1} + v_t$$
$$\text{and}: \;\; y_{t-1} = \alpha + \beta_1 X_{1,t-1} + \epsilon_{t-1} \tag{2}$$

Generalized Differences Continued...

Multiplying Equation (2) by ρ

$$\rho y_{t-1} = \rho\alpha + \rho\beta_1 X_{1,t-1} + \rho\epsilon_{t-1} \tag{3}$$

Subtracting Equation (3) from Equation (1)

$$y_t - \rho y_{t-1} = \alpha - \rho\alpha + (\beta_1 X_{1,t} - \rho\beta_1 X_{1,t-1}) + (\epsilon_t - \epsilon_{t-1}) \tag{4}$$

Simplifying...

$$y_t' = \alpha(1-\rho) + \beta_1 X_t' + v_t \tag{5}$$

where

$$y_t' = y_t - \rho y_{t-1} X_t' = X_t - \rho X_{t-1}$$

Notice that we can estimate eqution (5) by OLS and that β_1 is β_1 from Equation (1)

Applying Generalized Differences

Multuplie methods to choose from... - **Cochrane-Orcutt**

- **Hildreth-Lu**
- **First Differences**
  - Case where $\rho = 1$

Cochrane-Orcutt Steps

1. Estimate original equation $y_t = X_t\beta + \epsilon_t$ by OLS and save e_i's. Here, $\epsilon_t = \rho\epsilon_{t-1} + v_t$
2. Estimate ρ using

$$\hat{\rho} = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=2}^{n} e_{t-1}^2}$$

Use $\hat{\rho}$ to quasi-difference the variables (drop first observation) and estimate the quasi-differenced equation by OLS - This is the "Cohcrane-Orcutt" two-step estimation of the β vector

C-O Iteratively

Use $\hat{\beta}$ vector from step 3 to compute a new set of residuals: $e^* = y - X_t\hat{\beta}$ - no estiamtion, just computation - If the model includes an intercept: $\hat{\alpha} = \bar{y} - \bar{X}_t\hat{\beta}$

Use residuals to calculate a new $\hat{\rho}$

Repeat until desired convergence - Usually based on the sum of squared residuals

Hildreth-Lu

Same set-up as Cochrane-Orcutt - a trial and error method to minimize the sum of squared residuals (SSR)

Choose the minimum and maximum ρ values - Ex: $0.3 < \rho < 0.4$

Choose step values - Ex: 0.01

Select value of ρ that minimizes SSR