Lecture 1: Time series regression

Geert Mesters (Universitat Pompeu Fabra, Barcelona GSE and VU Amsterdam)

Introduction

Introduction

- In this set of slides we will revisit the most widely used and abused statistic/econometric methodology employed in applied work: the linear regression model
- Focus: Implications of running a regression with time series data
- Specifically, here we study the following predictive regression

$$Y_{t+h} = \beta_0 + \beta_1 X_{1\,t} + ... + \beta_p X_{p\,t} + \epsilon_t$$

  where $h \geq 0$ and Y_t, $X_{1\,t}$, ..., $X_{p\,t}$ and ϵ_t are stationary time series
- Unless otherwise stated, the term stationary should be interpreted loosely: strict stationary, covariance stationary, strong mixing, etc.

Introduction

In particular we will focus on:

1. Estimation and inference
2. Measuring marginal effects
3. Goodness-of-fit

Regression and Stationarity

Regression and Stationarity

- It is important to devote some extra thoughts to the theme of regression and stationarity
- Special care needs to be devoted to understanding the stationarity properties of the data when carrying out a regression
- Here we focus on the case in which the data are stationary
- A number of approaches exist for handling nonstationary data. Important to mention cointegration techniques, which allow to handle multiple nonstationary time series that share common trends. We will not deal with these methods.

Spurious Regression

- Spurious Regression is a classic example of why it is important to understand the stationarity properties of the data before running a regression.
- We will illustrate this concept using a real dataset example as well as a simulation study

Current Themes In Economics

One of the key questions currently debated in economic circles is:

What is the impact of hipsters on economic growth?

Hipsters and Economic Growth

To investigate this question empirically, I carry out the following exercise:

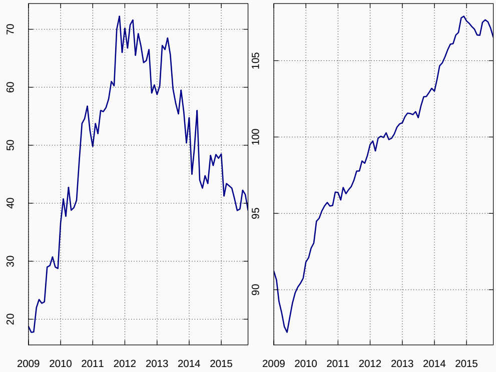
- Measure the popularity of hipsters X_t using the monthly number of searches for the word hipster on google in the US
- Measure economic growth Y_t using the monthly level of industrial production in the US
- Run the regression

$$Y_{t+1} = \beta_0 + \beta_1 X_t + \epsilon_t$$

- We can carry out inference on the impact of hipsters on economic growth on the basis of the estimate of the β_1 coefficient

Hipsters and Economic Growth

Hipsters and Industrial Production



Hipsters and Economic Growth

Estimation Results

	Estimate	Std. Error	t-ratio	p-value
Intercept	90.685	2.143	42.316	< 0.001
Hipsters	0.176	0.041	4.238	< 0.001
R^2	0.18			

Remarks

- The results would seem to suggest that the popularity of hipsters is a significant predictor of future industrial production
- The R^2 of the regression is about 20%! (large by econ standards)
- Hipsters predict economic growth. That is so mainstream!
- The result seems fishy however. The plot of the series shows that the series are clearly not stationary.
- Could the correlation between the two series be artificial and due to nonstationarity?

Monte Carlo Study

- In order to shed some light into this phenomenon we carry out a Monte Carlo exercise
- Let Y_t and X_t be random walks generated as

$$Y_t = Y_{t-1} + \epsilon_{Y\,t} \text{ and } X_t = X_{t-1} + \epsilon_{X\,t}$$

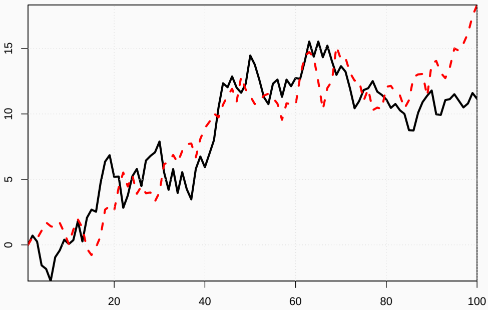
  where $\epsilon_{Y\,t}$ and $\epsilon_{X\,t}$ are independent normals with mean zero and variance 0.01
- Consider estimating the model

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

- Intuitively, β_1 should be equal to zero since Y_t and X_t are unrelated
- We can investigate whether this is the case by simulating this model and then carrying out classic least squares inference

Monte Carlo Results: One Replication

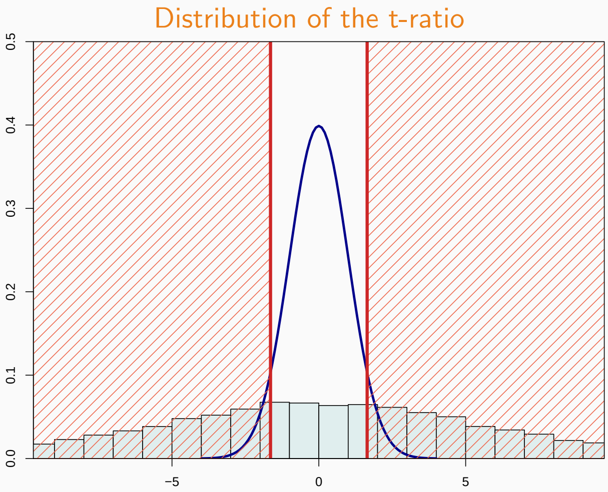
Y_t and X_t



Monte Carlo Results: One Replication

	Estimate	Std. Error	t value	p-value
(Intercept)	3.7581	0.4855	7.741	< 0.01
X	0.4719	0.1491	3.165	< 0.01

Monte Carlo Results



Monte Carlo Results

Prob($H_0 : \beta_1 = 0$ rejected at 5% level)	0.80
Average R^2	0.24
$q_{0.95}$ R^2	0.69

Remarks

- The results shows that even if Y_t and X_t are unrelated, the t-test signals significance in 80% of the cases!
- The R^2 is on average 25%. It can be larger then 69%
- The correlation is the consequence of the fact that both variables are "trending". Thus, when regressing one variable on the other, it is likely that the correlation will be substantial even if the two processes are unrelated.
- This issue was forcefully shown in a famous paper by Granger and Newbold (1974) who named this phenomenon spurious regression

Regression with Time Series

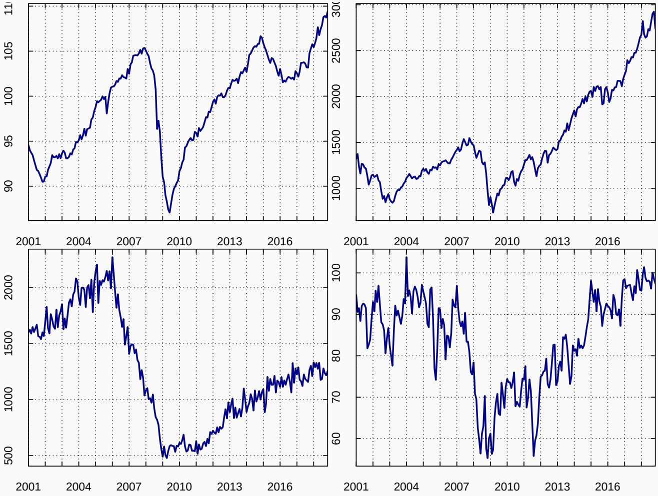
Problem: Predicting Industrial Production

- We will illustrate the concepts introduced in this set of slides with an empirical application
- We are interested in predicting h-period ahead (e.g. $h = 1$) Industrial Production using current Industrial Production, S&P500, New Housing Starts, and Consumer Sentiment
- The dataset contains monthly measurement of these four variables from 2000 to 2018
- The regression model uses the growth rates of the variables (which appear to be stationary)

Problem: Predicting Industrial Production

Industrial Production, S&P 500,

New Housing Starts, Consumer Sentiment

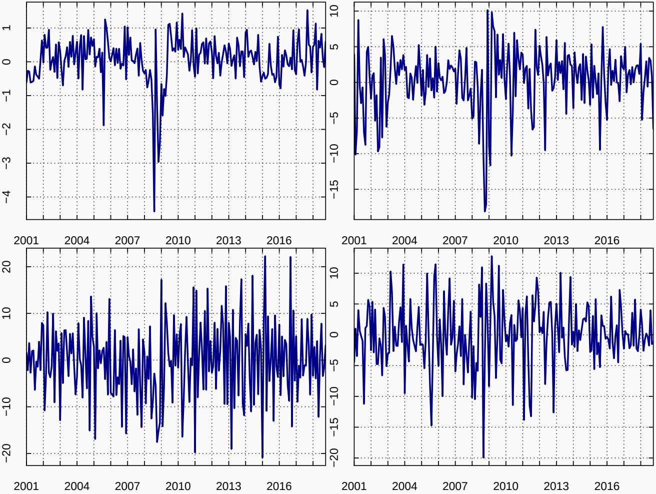


Levels

Problem: Predicting Industrial Production

Industrial Production, S&P 500,

New Housing Starts, Consumer Sentiment

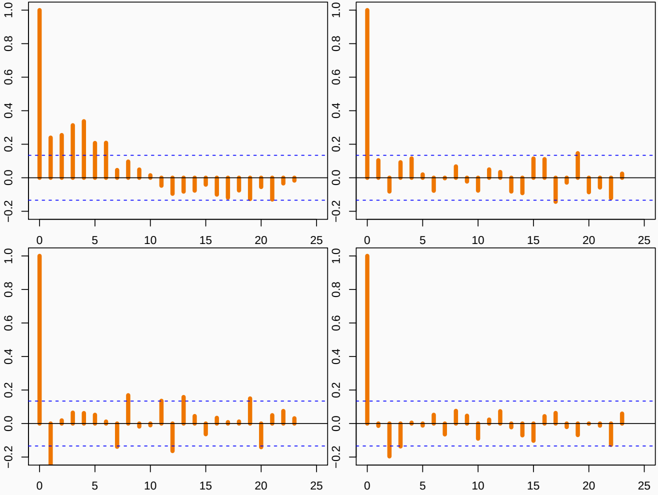


Growth Rates

Problem: Predicting Industrial Production

Industrial Production, S&P 500,

New Housing Starts, Consumer Sentiment



ACF Growth Rates

Estimation and Inference

Linear Regression with Time Series Data

- We shall assume that the data is generated as

$$Y_t = X_t'\beta_0 + \epsilon_t$$

where $X_t = (1, X_{1\,t}, ..., X_{p\,t})'$

- It is convenient to express the model in matrix form as

$$\underbrace{Y}_{T\times 1} = \underbrace{X}_{T\times(p+1)} \underbrace{\beta_0}_{(p+1)\times 1} + \underbrace{\epsilon}_{T\times 1}$$

- Note that β_0 denotes the true model parameter that we wish to learn from the data

Least Squares Estimator

- Inference on the linear regression model is carried out by least squares

- As it is well known, the least squares estimator is defined as

$$\hat{\beta} = \arg\min_{\beta} \sum_{t=1}^{T} (Y_t - X_t'\beta)^2$$

- We will see that several classic properties of the least squares estimator are retained in the time series setting.

- However, the distribution of the estimator can differ depending on the dependence properties of the data
  (implying that testing procedures might have to be appropriately modified)

Least Squares Estimator

- As it is well known, the least squares estimator is available in closed form and can be written as

$$\hat{\beta} = \left(\sum_{t=1}^{T} X_t X_t'\right)^{-1} \sum_{t=1}^{T} X_t Y_t = (X'X)^{-1} X' Y$$

- It is straightforward to show that the estimator can be conveniently represented as

$$\hat{\beta} = \beta_0 + \left(\sum_{t=1}^{T} X_t X_t'\right)^{-1} \sum_{t=1}^{T} X_t \epsilon_t = \beta_0 + (X'X)^{-1} X' \epsilon$$

LS Estimator - i.i.d. Setting

Recall that in the classic i.i.d. setting, under appropriate assumptions, the least squares estimator has the following properties:

1. Consistency

$$\hat{\beta} \xrightarrow{p} \beta_0$$

2. Asymptotically normality

$$\hat{\beta} \overset{a}{\sim} \mathcal{N}\left(\beta_0, \sigma^2 M^{-1} / T\right)$$

where σ^2 is the variance of ϵ_t and $M = \mathbb{E} X_t X_t'$

LS Estimator - Time Series Setting

- In the next slides we will look into the properties of the least squares estimator with time series data
- We shall assume that our data is generated by an α-mixing process, meaning that we are allowing for both time-series dependence and heterogeneity across observations

Strong Mixing

Definition

We say that the time series $\{Z_t\}$ is strong mixing of size $-a$ if the mixing coefficients of the $\{Z_t\}$ process satisfy

$$\alpha(m) = O(m^{-a-\epsilon}) \ ,$$

for some $\epsilon > 0$, where

$$\alpha(m) = \sup_n \alpha(\mathcal{B}^n_{-\infty}, \mathcal{B}^\infty_{n+m})$$

with

$$\alpha(\mathcal{B}_1, \mathcal{B}_2) = \sup_{b_1 \in B_1, b_2 \in B_2} |P(b_1 \cap b_2) - P(b_1)P(b_2)|$$

Also known as polynomial mixing, see White (2001, page 46)

Strong Mixing

An appealing property of α-mixing processes is that they have "hereditary" properties in the sense that functions of α-mixing processes are still α-mixing with the same size.

Theorem

Let $\{X_t\}$ and $\{\epsilon_t\}$ be an sequences of random variables with α mixing coefficients of size $-a$ and let $g : \mathbb{R}^{p+1} \times \mathbb{R} \to \mathbb{R}$ be a measurable function.

Then, $\{g(X_t, \epsilon_t)\}$ is α mixing with coefficients of size $-a$.

In particular, this theorem can be used to show that $\{X_t \epsilon_t\}$, $\{X_t X_t'\}$ are α mixing with coefficients of size $-a$.

Strong Mixing

Theorem

Law of Large Numbers for α-mixing Process Let $\{Z_t\}$ be an sequence of random variables with $\mathbb{E}Z_t = \mu$ for all t and with α mixing coefficients of size $-r/(r-1)$ for $r > 1$ such that $\mathbb{E}|Z_t|^{r+\delta} < C < \infty$ for all t for some $\delta > 0$.

Then,

$$\overline{Z}_T \xrightarrow{p} \mu \ .$$

Note that in comparison to standard LLN for i.i.d. data here we have a trade-off between moments and dependence: if few moments exist (i.e. the data is heavy tailed) then we require strong dependence decay
(and vice verse)

LS Estimator - Consistency

Theorem

Assume that

1. $\{X_t, \epsilon_t\}$ *is an* α*-mixing sequence of size* $-r/(r-1)$ *for* $r > 1$;
2. $\mathbb{E}|X_{i\,t}\epsilon_t|^{r+\delta} < \Delta < \infty$ *and* $\mathbb{E}|X_{i\,t}^2|^{r+\delta} < \Delta < \infty$ *for some* $\delta > 0$ *for all* t *and* i;
3. $\mathbb{E}X_t\epsilon_t = 0$ *for each* t *and* $\lim_{T\to\infty} \mathbb{E}X'X/T \to M$ *positive definite.*

Then

$$\hat{\beta} \xrightarrow{p} \beta_0 \ .$$

LS Estimator - Consistency

- (Sketch of proof.) It is straightforward to check that under these assumptions we have

$$
\begin{aligned}
\hat{\beta} &= \beta_0 + \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1} \sum_{t=1}^{T} X_t \epsilon_t \\
&= \beta_0 + \left(\frac{1}{T} \sum_{t=1}^{T} X_t X_t' \right)^{-1} \frac{1}{T} \sum_{t=1}^{T} X_t \epsilon_t \\
&\overset{p}{\to} \beta_0 + M^{-1} \cdot 0 = \beta_0
\end{aligned}
$$

Strong Mixing

Theorem

Central Limit Theorem for α-mixing Process
Let $\{Z_t\}$ be an sequence of random variables with $\mathbb{E}Z_t = \mu$ for all t and with α mixing coefficients of size $-r/(r-2)$ for $r > 2$ such that $\mathbb{E}|Z_t|^{r+\delta} < \Delta < \infty$ for all t for some $\delta > 0$ and

$$\sigma^2_{LR} = \lim_{T\to\infty} T\mathbb{V}ar(\overline{Z}_T)$$

Then,

$$\frac{\overline{Z}_T - \mu}{\sqrt{\sigma^2_{LR}/T}} \xrightarrow{d} \mathcal{N}(0,1) \ .$$

Note that in comparison to the LLN we require a stronger moment assumption as well as the long run variance of the process to converge

LS Estimator - Asymptotic Normality

Theorem

Assume that

1. $\{X_t, \epsilon_t\}$ *is an* α*-mixing sequence of size* $-r/(r-2)$ *for* $r > 2$;
2. $\mathbb{E}|X_{i\,t}\epsilon_t|^{r+\delta} < \Delta < \infty$ *and* $\mathbb{E}|X_{i\,t}^2|^{r+\delta} < \Delta < \infty$ *for some* $\delta > 0$ *for all* t *and* i;
3. $\mathbb{E}X_t\epsilon_t = 0$ *for each* t, $\lim_{T\to\infty} \mathbb{V}ar T^{-1/2} X'\epsilon \to V$ *positive definite and* $\lim_{T\to\infty} \mathbb{E}X'X/T \to M$ *positive definite.*

Then

$$\sqrt{T} D^{-1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, I_{p+1}) \ ,$$

where

$$D = M^{-1} V M^{-1}$$

LS Estimator - Asymptotic Normality

- (Sketch of proof.) It turns out that under the assumption the least square estimator satisfies the requirements to apply the CLT for α-mixing data.
- We have to work out the variance of the estimator:

$$\begin{aligned}
\mathbb{V}ar(\hat{\beta}) &= \mathbb{V}ar\left[\frac{T}{T}\left(\sum_{t=1}^{T} X_t X_t'\right)^{-1}\sum_{t=1}^{T} X_t \epsilon_t\right] \\
&= M^{-1}\mathbb{V}ar\left(\frac{1}{T}\sum_{t=1}^{T} X_t \epsilon_t\right)M^{-1} \\
&\stackrel{p}{\to} T^{-1}M^{-1}VM^{-1}
\end{aligned}$$

LS Estimator

- It is interesting to study the properties of estimator under two special cases
- We will consider two cases:
  1. ϵ_t is white noise and $\{\epsilon_t\}$ is independent of $\{X_t\}$
  2. ϵ_t is covariance stationary and $\{\epsilon_t\}$ is independent of $\{X_t\}$
- Note that "$\{\epsilon_t\}$ independent of $\{X_t\}$" means that X_t and ϵ_s are independent for each t and s. This is sometimes called "strict exogeneity" in econometric textbooks.

LS Estimator - Case 1

- In the first case the variance of the least squares estimator is

$$\begin{aligned}\mathbb{V}ar\left(T^{-1/2}\sum_{t=1}^{T}X_t\epsilon_t\right) &= \mathbb{E}\left[\mathbb{V}ar\left(T^{-1/2}\sum_{t=1}^{T}X_t\epsilon_t\middle|\{X_t\}\right)\right] \\ &= \mathbb{E}\left[T^{-1}\sum_{t=1}^{T}X_t\mathbb{V}ar\left(\epsilon_t|\{X_t\}\right)X_t'\right] \\ &= \sigma^2\mathbb{E}\left[T^{-1}\sum_{t=1}^{T}X_tX_t'\right]\to\sigma^2 M\end{aligned}$$

- This implies that the asymptotic variance of the least squares estimator is the same as in the i.i.d case

LS Estimator - Case 2

- In the second case the variance of the least squares estimator is

$$
\begin{aligned}
\mathbb{V}ar\left(T^{-1/2}\sum_{t=1}^{T} X_t\epsilon_t\right) &= T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left(X_t\epsilon_t\epsilon_s X_s'\right) \\
&= T^{-1}\left[\sum_{t=1}^{T}\mathbb{E}\left(\epsilon_t^2 X_t X_t'\right) + \sum_{t\neq s}\mathbb{E}\left(X_t\epsilon_t\epsilon_s X_s'\right)\right] \\
&= T^{-1}\left[\sum_{t=1}^{T}\mathbb{E}\left[\mathbb{E}\left(\epsilon_t^2 X_t X_t' | \{X_t\}\right)\right] + \sum_{t\neq s}\mathbb{E}\left[\mathbb{E}\left(X_t\epsilon_t\epsilon_s X_s' | \{X_t\right.\right.\right. \\
&= T^{-1}\left[\sigma^2\sum_{t=1}^{T}\mathbb{E}\left(X_t X_t'\right) + \sum_{t\neq s}\gamma_{t-s}\mathbb{E}\left(X_t X_s'\right)\right]
\end{aligned}
$$

LS Estimator - Case 2

- Notice that the variance of the least squares estimator is "larger" than what we have in the white noise/iid case
- In this case, the least squares estimator is still consistent but is no longer optimal.

Estimating the Covariance Matrix of the LS Estimator

- In the time-series setting, the least squares estimator is still normally distributed. However, the difference with respect to the classic case is that the covariance matrix of the estimator might be different depending on the properties of the data

- The asymptotic normality of the least squares estimator is key to carry out inference

- The covariance matrix of the least squares estimator is in general not known so we need to estimate this from data.

Estimating the Covariance Matrix of the Estimator

- If $\{\epsilon_t\}$ is white noise and independent of $\{X_t\}$ we can estimate the asymptotic covariance matrix of the least squares estimator using the standard formula used in the i.i.d. setting

- The asymptotic covariance matrix of the least squares estimator is

$$\widehat{S} = s^2 \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1}$$

where

$$s^2 = \frac{1}{T - p - 1} \sum_{t=1}^{T} \hat{\epsilon}_t^2$$

with $\hat{\epsilon}_t = Y_t - X_t' \hat{\beta}$

Estimating the Covariance Matrix of the Estimator

- The estimation of the covariance of the estimator is more challenging in when we allow for generic dependence

- It is possible to estimate the covariance matrix of the estimator using the Newey-West estimator. Let

$$\widehat{V}_{NW} = \sum_{t=1}^{T} \hat{\epsilon}_t^2 X_t X_t' + \sum_{l=1}^{L} \sum_{t=1}^{T} w_l \, \hat{\epsilon}_t \hat{\epsilon}_{t-l} \left(X_t X_{t-l}' + X_{t-l} X_t' \right)$$

where

$$w_l = 1 - \frac{l}{L+1}$$

Then, a consistent estimator of the covariance of $\hat{\beta}$ is

$$\widehat{S}_{NW} = \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1} \widehat{V}_{NW} \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1}$$

Remarks

- In general, we do not know whether the innovation terms of the regression are white noise or not. Thus, it is unclear a priori which estimator should be used.
- In practice, after estimating a linear regression model we can check the model residuals to assess if there is evidence of autocorrelation. If the residuals are significantly autocorrelated, it may be more appropriate to use the Newey-West estimator
- Notice the that Newey-West estimator depends on the choice of a "bandwidth" parameter L. In practice, there are plug-in formulas that take care of this.

Illustration: Predicting Industrial Production

- Consider our original regression model for Industrial Production
- Estimate the model over the full sample by least squares
- Carry out inference using both classic and Newey-West estimators

Illustration: Predicting Industrial Production

Classic Inference

	Estimate	Std. Error	t-value	p-value
Intercept	0.055	0.0503	1.094	0.275
IP	0.237	0.0728	3.257	0.001***
S&P500	0.021	0.0119	1.822	0.070*
House	0.010	0.0064	1.684	0.093*
Sentiment	0.109	0.2189	0.500	0.617
R^2	0.09			

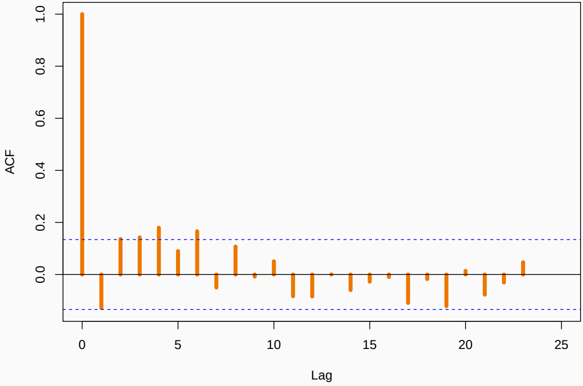
Illustration: Predicting Industrial Production

Newey-West Inference

	Estimate	Std. Error	t-value	p-value
Intercept	0.055	0.0548	1.003	0.316
IP	0.237	0.0824	2.878	0.004***
S&P500	0.021	0.0096	2.261	0.024**
House	0.010	0.0100	1.081	0.281
Sentiment	0.109	0.1807	0.606	0.545
R^2	0.09			

Illustration: Predicting Industrial Production

Residuals



$$Q(12) = 27.0047 \text{ (p-val} < 0.01)$$

Illustration: Predicting Industrial Production

- Residuals of the model exhibit weak albeit significant correlation. Newey-West more appropriate estimator to use in this setting.
- Inference using Classic and Newey-West approach is comparable. However, significance of housing disappears using Newey-West.
- Notice that in this application, most of the forecasting power comes from the lagged values of IP

Measuring Marginal Effects

Measuring Marginal Effects

- When carrying out regression, often we are concerned in measuring the marginal effect of a predictor on the dependent variable
- In a time series setting, the concept of marginal effect is richer
- Typically, marginal effects change across horizons
- It is often of interest to measure the marginal effect of a predictor and how it varies with the horizon

Measuring Marginal Effects

- Let's make the regression coefficients (explicitly) depend of the predictive regression on the forecast horizons h

$$Y_{t+h} = \beta_0^{(h)} + \beta_1^{(h)} X_{1\,t} + ... + \beta_p^{(h)} X_{p\,t} + \epsilon_t^{(h)}$$

- Then, the collection of coefficients

$$\{\beta_i^{(h)}\}_{h=1}^{H}$$

synthesize the marginal effect of variable x_i on y from $h = 1$ to H conditional on the information available at time t

- The estimation of marginal effects by sequence of regression is quite popular these days in macro under the name of local projections.

Illustration: Predicting Industrial Production

- Consider our original regression model for Industrial Production

- We are concerned with estimating the marginal effect of the S&P 500 growth rates on Industrial Production from $h = 1$ to 6

- We carry out inference on the marginal effects of the S&P 500 growth rates using NW covariance estimator

Illustration: Predicting Industrial Production

Marginal Effect of S&P 500 on Industrial Production

(Point Estimate + 90% NW Conf Interval)

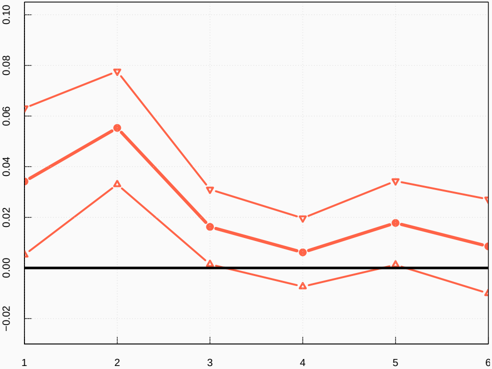


Illustration: Predicting Industrial Production

- The effects of S&P500 are significant up to 3 months ahead.
- Interestingly, effect of S&P 500 are larger and with stronger significance at an horizon of $h = 2$
- Does it make sense to find that the effects of the S&P 500 are stronger at $h = 3$? Yes! The stock market is considered as a leading indicator of economic activity, that is the stock market (which is forward looking) typically anticipates movements of real activity.

Goodness-of-Fit

Goodness-of-Fit

- A natural question that arises when working with a model is: How good is the fit of the model?
- In regression analysis, typically the in-sample R^2 is used as a natural measure of goodness of fit
- However, in a time series setting is more interesting to validate the model using an out-of-sample prediction criterion
- Notice that the R^2 uses the same sample of data to estimate and to measure the goodness-of-fit. It provides over optimistic measures of goodness of fit.

Model Validation

A simple recipe for out-of-sample model validation:

- Divide the sample into two sub samples labelled as in-sample and out-of-sample
- Estimate the model using the in-sample data
- Predict the dependent variables using the out-of-sample data and the estimator of β obtained in-sample
- Measures goodness of fit by comparing the out-of-sample predictions with the out-of-sample realized values

Model Validation

- Classic index used to synthesize out-of-sample predictive ability is the Mean Square Error of the predictions
- Let $\hat{Y}_t$ denote the forecast of Y_t

$$MSE = \frac{1}{T - t_{\text{start}}} \sum_{t=t_{\text{start}}}^{T} (Y_t - \hat{Y}_t)^2$$

where t_{start} denotes the first out-of-sample observation

- It is also interesting to consider a R^2 type criterion:

$$\widetilde{R}^2 = 1 - \frac{MSE}{s^2_{oos}}$$

where s^2_{oos} is the out-of-sample variance of the dependent variable

Illustration: Predicting Industrial Production

- Consider our original regression model for Industrial Production
- Estimate the model by least squares from January 2000 to January 2010
- Use the least square estimator to predict future Industrial Production (growth rates) from January 2010 to the end of the sample
- Benchmark the forecasts again (a simple) constant growth model (i.e. predict future observations with the in-sample average of the Industrial Production growth rates)

Illustration: Predicting Industrial Production

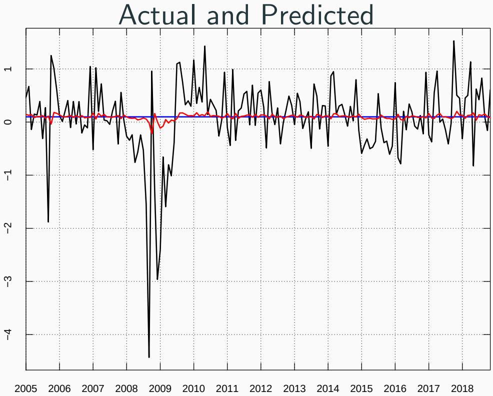


Illustration: Predicting Industrial Production

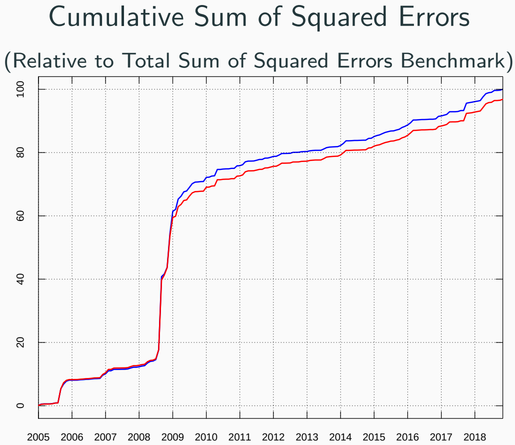


Illustration: Predicting Industrial Production

Out-of-Sample Validation

	Regression	Constant Growth
MSE	0.476	0.445
$\widetilde{R}^2$	5%	-0.01

References & Material

Reference & Material

- References:
  J. M. Wooldridge, "Introductory Econometrics: A Modern Approach", Chapter 10 and 11
  H. White, "Asymptotic Theory for Econometricians", Chapters 1 to 6

- Code:
  `regression-example.Rmd`
  `spurious-regression-example.Rmd`
  `spurious-regression-sim.Rmd`