Regression: Inference

Mark Hendricks

Autumn 2017

FINM Intro: Regression

Outline

Assumption: Full-rank

Assumption 1: $\mathbf{X}'\mathbf{X}$ is full rank.

Equivalently, assume that there is no exact linear relationship among any of the regressors.

- Clearly, the existence of OLS estimator requires that this assumption be satisfied.
- Multicollinearity refers to the case where this assumption fails.

The Classic Model Classic Inference Large Sample Properties

Assumption: Exogeneity

Assumption 2: ϵ is exogenous to the regressors, $\mathbf{x}$.

$$\mathbb{E}\left[\epsilon \mid \mathbf{x}\right] = 0$$

The exogeneity assumption,

- implies that ϵ is uncorrelated with $\mathbf{x}$.
- implies that ϵ is uncorrelated with any function of $\mathbf{x}$.
- does NOT imply that ϵ is independent of $\mathbf{x}$.

Statistics as variables

To judge the OLS forecast, remember that

- A statistic is a function of random variables.
- Thus, a statistic is itself a random variable with a mean, variation, and distribution.
- A good statistic/forecast will be centered tightly around the true population value.

Unbiased statistics

An estimate is **unbiased** if its expectation equals the population value.

- Consider the sample estimator, $\bar{x}$, for a sample of n.
- Suppose we have a variable x with population mean μ_x.;
- Verify that $\overline{x}$ is unbiased:

$$\begin{aligned}\mathbb{E}\left[\overline{x}\right] =& \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] \\ =& \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[x_i\right] \\ =& \mu_x\end{aligned}$$

Is OLS estimate unbiased?

Check if the OLS estimator is unbiased:

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{b}\right] =& \mathbb{E}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}\right] \\ =& \mathbb{E}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}\right)\right] \\ =& \boldsymbol{\beta}+\mathbb{E}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\epsilon}\right]\end{aligned}$$

But $\mathbb{E}\left[\boldsymbol{b}\right] = \boldsymbol{\beta}$ if, and only if,

$$\mathbb{E}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\epsilon}\right] = 0$$

which is guaranteed by the exogeneity assumption but not simply by orthogonality.

Variance of the mean estimator

Continue with the example of the sample mean estimator, $\bar{x}$ for sample size n.

- Suppose $\text{var}\left[x\right] = \sigma^2$.
- What is the variance of the sample mean estimator, $\bar{x}$?

$$\text{var}\left[\bar{x}\right] = \frac{\sigma^2}{n}$$

Variance of OLS estimator

The variance of the OLS estimator is

$$\text{var}\left[\boldsymbol{b} \mid \mathbf{x}\right] = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\Sigma\mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

where $\Sigma = \mathbb{E}\left[\boldsymbol{\epsilon\epsilon}' \mid \mathbf{x}\right]$.

- From above we have that $\mathbb{E}\left[\epsilon\right] = 0$.
- Thus, Σ is the variance of ϵ.
- So far we have used the first two moments of ϵ, along with its relation to $\mathbf{x}$, but no assumption on the distribution of ϵ has been made.

Heteroscedasticity and autocorrelation of residuals

- **Heteroscedasticity** refers to the case where the residuals but have distinct variances,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

- **Autocorrelation** refers to the case where residuals are correlated. In this case, Σ is not diagonal.

Assumption: Homoscedastic and orthogonal residuals

Assumption 3:
The residuals are uncorrelated across observations, with identical variances,

$$\Sigma = \mathbb{E}\left[\boldsymbol{\epsilon}\boldsymbol{\epsilon}' \mid \mathbf{x}\right] = \sigma^2 \mathcal{I}_n$$

Gauss-Markov Theorem

With these assumptions, the OLS estimator, $\boldsymbol{b}$, is the minimum variance linear unbiased estimator of $\boldsymbol{\beta}$.

- The assumption on Σ simplifies the variance of the OLS estimator,

$$\begin{aligned} \text{var}\left[\boldsymbol{b} \mid \mathbf{x}\right] = & \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}' \sigma^2 \mathcal{I}_n \mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ = & \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} \end{aligned}$$

- Any other linear, unbiased estimator of $\boldsymbol{\beta}$ will have larger variance.
- This is known as the Gauss-Markov theorem. It depends on the above assumptions regarding linearity, exogeneity, full-rank, and residual covariance structure.

Outline

The Classic Model

Classic Inference

Large Sample Properties

OLS Inference

The distribution of the OLS estimates is required in order to assess statistical significance.

- Above the mean and variance of $\boldsymbol{b}$ were derived without making any distributional assumptions.

Assumption: Normality of residuals

Assumption 4: The residuals, ϵ are normally distributed.

$$\epsilon \,|\mathbf{x} \sim \mathcal{N}\left(\mathbf{0}, \Sigma\right)$$

Distribution of OLS estimator

Assumptions 1, 2, 3, 4 imply

$$\boldsymbol{b} \,|\mathbf{x} \sim \mathcal{N}\left(\boldsymbol{\beta}, \Omega\right)$$

where

$$\Omega = \sigma^2 \left(X'X\right)^{-1}$$

Often, these 4 assumptions are referred to as the classical regression model.

- Note that many results were derived without any distributional assumptions.

OLS Z-test

Testing the significance of an element of $\boldsymbol{b}$, would simply be a z-test:

$$\frac{b_j - \beta_j}{\sigma \ \omega_{jj}} \sim Z$$

where ω_{jj}^2 is the (j,j) element of $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$. Thus, $\sigma^2\omega_{ij}^2$ is the (i,j) element of Ω.

- However, this statistic depends on σ, which is unknown.
- Instead one must use the sample estimate of the variance,

$$s^2 = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-k-1}$$

OLS t-test

The t-test assesses statistical significance of an element of $\boldsymbol{b}$,

$$\frac{b_j - \beta_j}{s \ \omega_{jj}} \sim t(n-k-1)$$

which depends only on observable data as well as the hypothesized value, β_j.

R-squared

The **R-squared**, or coefficient of determination, in a regression is defined as

$$R^2_{y,x} = \frac{\text{regression sum of squares}}{\text{total sum of squares}}$$

$$= 1 - \frac{\text{error sum of squares}}{\text{total sum of squares}}$$

Algebraically, this is

$$R^2_{y,x} = \frac{\boldsymbol{b} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$
$$= 1 - \frac{\boldsymbol{e}'\boldsymbol{e}}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

R-squared versus correlation

Intuitively, the R-squared is the square of the correlation between y and the projection of y onto $\mathbf{x}$.

$$R^2_{y,\mathbf{x}} = [\text{corr}(\mathbf{Y}, \mathbf{PY})]^2$$

In a univariate regression of y on x,

$$R^2_{y,x} = [\text{corr}(y, x)]^2$$

Caveat: Regressing on a constant

The interpretation and formula for R-squared does not hold if there is no constant regressor.

- Without a constant, the R-squared will not necessarily be between 0 and 1.
- Without a constant, the R-squared will not necessarily be the square of the correlation between the sample **Y** and the projected Y values.
- Without a regressor, the fit can be improved simply by shifting the sample **Y** data by a constant.

OLS F-test

An F-test will determine the joint significance of the linear regression:

$$\frac{R^2_{y,\mathbf{x}}}{1-R^2_{y,\mathbf{x}}}\left(\frac{n-k-1}{k}\right) \sim F(k, n-k-1)$$

Namely, this tests whether all coefficients are jointly equal to zero. (We are assuming the regression includes a constant.)

- Note the use of the R-squared stat.
- It is simple to generalize this to test whether $\boldsymbol{b}$ is jointly equal to a non-zero hypothesis vector, $\boldsymbol{\beta}^*$.

Problems with multicollinearity

If regressor j is highly correlated with the other regressors, then the variance of the coefficient estimate can be written as

$$\text{var}\left[b_j \mid \mathbf{x}\right] = \left(\frac{1}{1 - R^2_{x_j,\mathbf{x}_{[-j]}}}\right) \frac{\sigma^2}{\sum_{i=1}^{n} \left(\mathbf{x}_{i,j} - \bar{\mathbf{x}}_j\right)^2}$$

where

$$R^2_{x_j,\mathbf{x}_{[-j]}}$$

denotes the R-squared from regressing x_j on the remaining regressors, (all columns except column j of $\mathbf{X}$.)

Variance inflation factor

The term

$$\frac{1}{1 - R^2_{x_j,\mathbf{x}_{[-j]}}}$$

is known as the **variance inflation factor**.

- The VIF is closely related to the condition number of $\mathbf{X}'\mathbf{X}$.
- The condition number in linear algebra measures the sensitivity of inverting a matrix.
- It compares the largest and smallest eigenvalues.
- Most software packages will warn the user if the condition number (VIF) is too big.

Example of multicollinearity

Consider a regression examining how the unemployment rate responds to the short and long ends of the U.S. Treasury yield curve.

- Such a regression would be at least a crude attempt to think about the dual mandate of the Fed.
- They are supposed to balance a stable money supply against stimulating full employment.
- Monetary stimulation could effect the yield curve and unemployment.

Regressing with multicollinearity

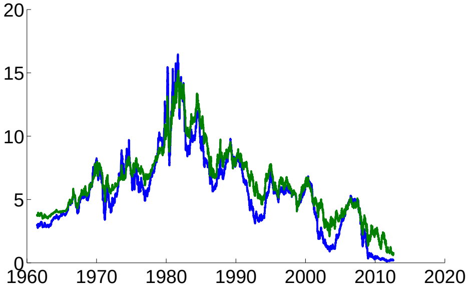


Figure: VIF=7.6. Condition of $X'X = 13.7$

Warning: Correlation of interest rates.

This is an unsophisticated model, but it serves as a basic warning.

- Many applications use interest rates as explanatory variables.
- However, many different rates are highly correlated.
- Recall that multicollinearity decreases confidence in the OLS estimate.

Variation in regressors

The sample variance of x_j reduces the variance of the OLS estimate b_j.

- Again write

$$\text{var}\left[b_j \mid \mathbf{x}\right] = \left(\frac{1}{1 - R^2_{x_j, \mathbf{x}_{[-k]}}}\right) \frac{\sigma^2}{\sum_{i=1}^{n} \left(\mathbf{x}_{i,j} - \bar{\mathbf{x}}_j\right)^2}$$

- The denominator of the second term is the scaled sample variance.
- If there is not enough variation in the regressor data, then the OLS estimation can not precisely estimate β_j.

Example: Considering variation in the regressor

The standard error of the OLS estimator depends on the variation in the regressors.

- The standard error of $\boldsymbol{b}$ decreases as the variation in $\mathbf{X}$ increases, holding other things equal.
- Recall that net interest margin refers to the spread in the lending and borrowing of banks.
- Consider using this as a regressor, for some financial or economic data.

Little variation in market risk

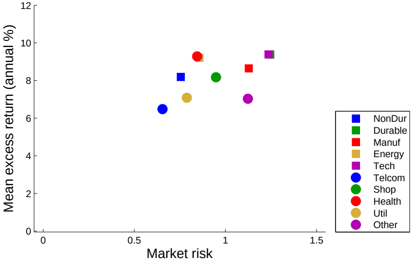


Figure: Data Source: Ken French. Monthly 1926-2011.

Little variation in market risk

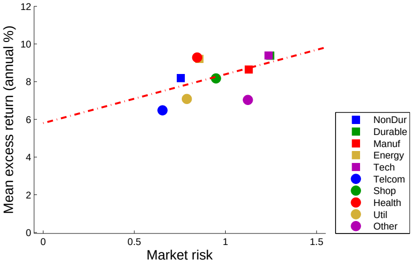


Figure: Data Source: Ken French. Monthly 1926-2011.

Outline

Is OLS robust?

How good is OLS if the assumptions do not hold?

- Financial data is usually non-normal—violating Assumption 4.
- Time-series models will almost always violate exogeneity—Assumption 2.
- Macro-economic data typically has correlated residuals, while asset prices show time-varying volatility—violations of Assumption 3.

OLS corrections

Two main ways to address these problems:

- Large sample properties. (Relax assumptions 2, 4.)
- Robust standard errors. (Relax assumption 3.)

Instrumental Variable Regression (IV) is also very important in dealing with assumption 2, but will not be discussed here.

Non-normality

Applications often do not satisfy Assumption 4, upon which the inference results relied.

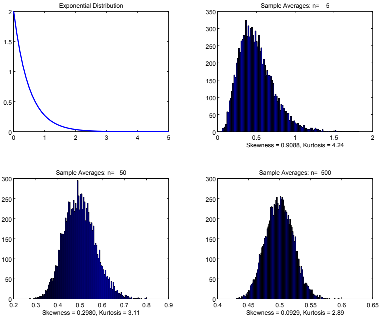
- However, the asymptotic distribution of the OLS estimate is an application of the Central Limit Theorem.
- In practice, inference often relies on having large data sets and appealing to the asymptotic results.

Central Limit Theorem

Covered in other September Review Modules.

- Basically, it says that as sample size increases, the sample average statistic converge to a normal distribution.
- Slightly more complicated for non-iid data, but weaker versions hold.
- Note that the OLS estimator can be rewritten as a sample average of ϵ!

Example - Central Limit Theorem



Assumption: Orthogonality of population residuals

Assumption 5: The population residuals are uncorrelated with the regressors.

$$\mathbb{E}\left[\mathbf{x}'\epsilon\right] = \mathbf{0}$$

- This assumption is much weaker than Assumption 2.
- This is a restriction on the population variables, not the fitted estimates, which have zero correlation by construction.

Consistency

A sample statistic is **consistent** if it converges to the true population value in probability.

- Suppose that Assumptions 1, 5 hold.
- Then the OLS estimator, $\boldsymbol{b}$ is consistent,

$$\text{plim} \quad \boldsymbol{b} = \boldsymbol{\beta}$$

- In practice, more attention is paid to having a consistent estimator than an unbiased estimator, due to the weaker assumption.

Asymptotic distribution of OLS

Under Assumptions 1,3, 5, the OLS estimate is asymptotically normal,

$$\boldsymbol{b} \mid \mathbf{x} \sim^{\text{asym}} \mathcal{N}\left(\boldsymbol{\beta}, \Omega\right)$$

where

$$\Omega = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

Heteroscedastic and autocorrelated inference

For many applications, particularly in time-series, Assumption 3 is clearly false.

- For practical purposes, this is not a big problem for inference.

Under Assumptions 1, 5, the OLS estimate is asymptotically normal,

$$\boldsymbol{b} \mid \mathbf{x} \sim^{\text{asym}} \mathcal{N}(\boldsymbol{\beta}, \Omega)$$

where

$$\Omega = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\Sigma\mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

OLS without iid errors

With non-iid errors, OLS is still unbiased (or consistent).

- Thus, it is appropriate to estimate with OLS, but one must use the larger variance given by

$$\text{var}\left[\boldsymbol{b} \mid \mathbf{x}\right] = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\Sigma\mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

- Non-OLS estimators, such as GLS, may have lower variances which allow for more confident inference.

References

- Cochrane, John. *Asset Pricing*. 2001.
- Greene, William. *Econometric Analysis.* 2011.
- Hamilton, James. *Time Series Analysis.* 1994.
- Wooldridge, Jeffrey. *Econometric Analysis of Cross Section and Panel Data.* 2011.