Estimation, inference and hypothesis testing

Daniele Bianchi[1]

whitesphd.com

[1]School of Economics and Finance
Queen Mary, University of London

Summary

This week provides an overview of estimation and inference. Testing a financial theory is often a three-step process: first, the model parameters should be estimated. Next, the distribution of the parameters should be defined. Finally, formal hypothesis tests must be conducted. This week we will focus on the first two steps. I keep the explanation intentionally "generic" by design and focus only on the case of identically and independently distributed RVs. All the concepts covered this week will be used in the next few weeks when we will cover linear regressions, time series analysis and univariate volatility modeling.

Contents

Estimation

M-Estimators

The use of M-estimators is pervasive in financial economics. Three common types of M-estimators:

- Method of moments
- Maximum likelihood
- Classical minimum distance

In this module we will focus only on the **Maximum Likelihood** estimator.

Maximum likelihood estimation (MLE)

Definition (Maximum likelihood)

Maximum likelihood uses the distribution of the data to *estimate any unknown parameters* by finding the values which make the data as likely as possible to have been observed – in other words, by maximizing the likelihood.

Maximum likelihood estimation (MLE) begins by specifying the *joint* distribution $f(\mathbf{y};\theta)$, of the observable data $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, as a function of a $k \times 1$ vector θ which contains all parameters.

Let $L(\theta, \mathbf{y}) = f(\mathbf{y};\theta)$, the MLE $\hat{\theta}$ is defined as the solution

$$\hat{\theta} = \arg\max_{\theta} L(\theta, \mathbf{y}),$$

Maximum likelihood estimation (MLE)

The $\arg\max$ is used in place of $\max$ to indicate that the maximum may not be unique and to indicate that the *global* maximum is required.

- Since $L(\theta, \mathbf{y})$ is strictly positive, the log of the likelihood can be used to estimate θ.
- The log-likelihood is defined as $l(\theta, \mathbf{y}) = \ln L(\theta, \mathbf{y})$.
- In most situation the MLE can be found by solving the k by 1 score vector

$$\frac{\partial l(\theta, \mathbf{y})}{\partial \theta} = 0,$$

In the next few slides we are going to review the ML estimation of a Poisson Model.

MLE of a Poisson model

Suppose $y_i \overset{i.i.d}{\sim} \text{Poisson}(\lambda)$. The PDF of a single observation is

$$f(y_i; \lambda) = \frac{\exp(-\lambda)\,\lambda^{y_i}}{y_i!},$$

Since the data are i.i.d, the joint likleihood is simple the product of the n individual likelihoods, i.e.,

$$f(\mathbf{y}; \lambda) = L(\lambda; \mathbf{y}) = \prod_{i=1}^{n} \frac{\exp(\lambda)\,\lambda^{y_i}}{y_i!},$$

The log-likelihood is

$$l(\lambda; \mathbf{y}) = \sum_{i=1}^{n} \left(-\lambda + y_i \ln(\lambda) - \ln(y_i!)\right),$$

which can be further simplified to

$$l(\lambda; \mathbf{y}) = -n\lambda + \ln(\lambda) \sum_{i=1}^{n} y_i - \sum_{j=1}^{n} \ln(y_j!),$$

MLE of a Poisson model

The first derivative is

$$\frac{\partial l\left(\lambda;\mathbf{y}\right)}{\partial\lambda} = -n + \lambda^{-1}\sum_{i=1}^{n} y_i,$$

The MLE is found by setting such derivative equal to 0 and solving, i.e.,

$$\begin{aligned}
-n + \hat{\lambda}^{-1}\sum_{i=1}^{n} y_i &= 0\\
\hat{\lambda}^{-1}\sum_{i=1}^{n} y_i &= n\\
\sum_{i=1}^{n} y_i &= n\hat{\lambda}\\
\hat{\lambda} &= n^{-1}\sum_{i=1}^{n} y_i,
\end{aligned}$$

Thus the MLE in a Poisson model is the sample mean.

Conditional MLE

Interest often lies in the distribution of a random variable Y conditional on one or more observed values $\mathbf{X}$.

- The likelihood for a single observation is $f\left(y_i|\mathbf{x}_i\right)$, and when Y_i are conditionally i.i.d., then

$$L\left(\theta;\mathbf{y}|\mathbf{X}\right)=\prod_{i=1}^{n} f\left(y_i|\mathbf{x}_i\right),$$

- The log-likelihood is defined as

$$l\left(\theta;\mathbf{y}|\mathbf{X}\right)=\sum_{i=1}^{n}\ln f\left(y_i|\mathbf{x}_i\right),$$

- The key aspect of the conditional MLE is that the relationship between Y and $\mathbf{X}$ has to be specified.
- For e.g., $y|\mathbf{x}\sim N\left(\mu,\sigma^2\right)$ where $\mu=\boldsymbol{\beta}'\mathbf{x}$.

Conditional MLE of a Poisson model

Suppose y_i is conditionally on $\mathbf{x}_i$ i.i.d. distributed $\text{Poisson}\left(\lambda_i\right)$ where $\lambda_i = \exp\left(\theta x_i\right)$. The likelihood is defined as

$$L\left(\theta; \mathbf{y}|\mathbf{x}\right) = \prod_{i=1}^{n} \frac{\exp\left(\lambda_i\right) \lambda_i^{y_i}}{y_i!},$$

and the log-likelihood is[1]

$$l\left(\theta; \mathbf{y}|\mathbf{x}\right) = \sum_{i=1}^{n} \left(-\exp\left(\theta x_i\right) + y_i\left(\theta x_i\right) - \ln\left(y_i!\right)\right),$$

The score of the likelihood is

$$\frac{\partial l\left(\theta; \mathbf{y}|\mathbf{x}\right)}{\partial \theta} = \sum_{i=1}^{n} -x_i \exp\left(\hat{\theta} x_i\right) + x_i y_i = 0,$$

[1] N.B., we simply replace λ in the simple MLE with $\lambda_i = \exp\left(\theta x_i\right)$.

Conditional MLE of a Poisson model

This score cannot be analytically solved. It is possible, however, to show the score has conditional expectation 0 since $E\left[y_i|x_i\right] = \lambda_i$. Therefore,

$$
\begin{aligned}
E\left[\frac{\partial l\left(\theta;\mathbf{y}|\mathbf{x}\right)}{\partial\theta}|\mathbf{X}\right] &= E\left[\sum_{i=1}^{n} -x_i \exp\left(\hat{\theta}x_i\right) + x_i y_i|\mathbf{X}\right] \\
&= \sum_{i=1}^{n} E\left[-x_i \exp\left(\theta x_i\right)|\mathbf{X}\right] + E\left[x_i y_i|\mathbf{X}\right] \\
&= \sum_{i=1}^{n} -x_i\lambda_i + x_i E\left[y_i|\mathbf{X}\right] \\
&= \sum_{i=1}^{n} -x_i\lambda_i + x_i\lambda_i = 0,
\end{aligned}
$$

This opens up the possibility to use numerical procedures, such as the Expectations-Maximization (EM) algorithm to estimate the model.

Properties of estimators

Properties of estimators

The first step in assessing the performance of an economic model is the estimation of the parameters. There are a number of desirable properties estimators may possess. These are:

- Bias and consistency
- Asymptotic normality
- Efficiency

Bias

Any discrepancy between the expected value of an estimator and the population parameter is known as bias.

Definition (Bias)

The Bias of an estimator $\hat{\theta}$ is defined

$$B\left[\hat{\theta}\right] = E\left[\hat{\theta}\right] - \theta_0,$$

where θ_0 is used to denote the population (or "true") value of the parameter.

- When an estimator has a bias of 0 it is said to be **unbiased**.
- Unfortunately, many estimators are not unbiased.

Consistency

Consistency is a closely related concept that measures whether a parameter will be far from the population value in *large samples*.

Definition (Consistency)

An estimator $\hat{\theta}_n$ is said to be consistent if $\text{plim}\hat{\theta}_n = \theta_0$. The explicit dependence of the estimator on the sample size is used to clarify that these form a sequence $\left\{\hat{\theta}_n\right\}_{n=1}^{\infty}$.

Consistency requires an estimator to exhibit two features as the sample size becomes large.

- First, any bias must be shrinking.
- Second, the distribution of $\hat{\theta}$ around θ_0 must be shrinking in such a way that virtually all of the probability mass is arbitrarily close to θ_0.

Asymptotic normality

While unbiasedness and consistency are highly desirable properties of any estimator, alone these do not provide a method to perform inference.

The primary tool in econometrics for inference is the central limit theorem (CLT). There are a variety CLTs, the Lindberg-Lévy CLT is the simplest

Theorem (Lindberg-Lévy)

Let $\{y_i\}$ be a sequence of i.i.d. random variables with $\mu \equiv E\left[Y_i\right]$ and $\sigma^2 \equiv V\left[Y_i\right] < \infty$. If $\sigma^2 > 0$, then

$$\frac{\overline{y}_n - \mu}{\overline{\sigma}_n} = \sqrt{n}\frac{\overline{y}_n - \mu}{\sigma} \rightarrow N\left(0, 1\right),$$

where $\overline{y}_n = n^{-1}\sum_{i=1}^{n} y_i$ and $\overline{\sigma}_n = \sqrt{\frac{\sigma^2}{n}}$.

Lindberg-Lévy states that as long as i.i.d. data have 2 moments – the mean and the variance – the sample mean will be asymptotically normal.

Efficiency

A final concept, efficiency, is useful for ranking consistent asymptotically normal estimators.

Definition (Relative efficiency)

Let $\hat{\theta}_n$ and $\tilde{\theta}_n$ be two $\sqrt{n}$-consistent asymptotically normal estimators for θ_0. If the asymptotic variance of $\hat{\theta}_n$, i.e., $\mathrm{avar}\left(\hat{\theta}_n\right)$ is

$$\mathrm{avar}\left(\hat{\theta}_n\right) < \mathrm{avar}\left(\tilde{\theta}_n\right),$$

then $\hat{\theta}_n$ is said to be relatively efficient to $\tilde{\theta}_n$.

From the concept of relative efficiency it is quite easy to derive what is an "asymptotically efficient estimator"; that is, if $\mathrm{avar}\left(\hat{\theta}_n\right) < \mathrm{avar}\left(\tilde{\theta}_n\right)$, for any choice of $\tilde{\theta}_n$ then $\hat{\theta}_n$ is said to be the **efficient** estimator.

Distribution theory and inference

Distribution of the MLE

Recall that MLE is defined as the maximum of the log-likelihood of the data with respect to the parameters, i.e.,

$$\hat{\theta} = \arg\max_{\theta} l\left(\theta; \mathbf{y}\right),$$

Definition (Asymptotic distribution of the MLE)

The asymptotic distribution of the MLE $\hat{\theta}$ is

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, \mathcal{I}^{-1}\right),$$

where

$$\mathcal{I} = -E\left[\left.\frac{\partial^2 l\left(\theta; y_i\right)}{\partial\theta\partial\theta'}\right|_{\theta=\theta_0}\right],$$

A related, perhaps more useful, concept to the information matrix $\mathcal{I}$ is the Cramér-Rao lower bound.

Theorem (Cramér-Rao inequality)

Let $f(\mathbf{y};\theta)$ be the joint density of $\mathbf{y}$ where θ is a k dimesional parameter vector. Let $\hat{\theta}$ be a consistent estimator of θ with finite covariance. Under some regularity condition on $f(\cdot)$

$$avar\left(\hat{\theta}\right) \geq \mathcal{I}^{-1}(\theta),$$

where

$$\mathcal{I} = -E\left[\left.\frac{\partial^2 \ln f(Y_i;\theta)}{\partial\theta\partial\theta'}\right|_{\theta=\theta_0}\right],$$

Distribution of the MLE

Two important remarks:

- The important implication of the Cramér-Rao theorem is that MLE, which are generally consistent, are asymptotically efficient.
- This guarantee makes a strong case for using ML when available.

In the following we are going to see as an example how to derive the asymptotic distribution of the MLE for the Poisson model.

Inference in a Poisson MLE

Recall that the log-likelihood in a Poisson MLE is

$$l\left(\lambda;\mathbf{y}\right)=-n\lambda+\ln\left(\lambda\right)\sum_{i=1}^{n}y_i-\sum_{i=1}^{n}\ln\left(i\right),$$

and that the first-order condition is

$$\frac{\partial l\left(\lambda;\mathbf{y}\right)}{\partial\lambda}=-n+\lambda^{-1}\sum_{i=1}^{n}y_i,$$

The MLE was previously shown to be $\hat{\lambda}=n^{-1}\sum_{i=1}^{n}y_i$. The second derivative can be defined as,

$$\frac{\partial^2 l\left(\lambda;y_i\right)}{\partial\lambda^2}=-\lambda^{-2}y_i,$$

Inference in a Poisson MLE

Now take the expectation of the negative of the second derivative

$$
\begin{aligned}
\mathcal{I} = -E\left[\frac{\partial^2 l\left(\lambda; y_i\right)}{\partial \lambda^2}\right] &= -E\left[-\lambda^{-2} y_i\right], \\
&= \left[\lambda^{-2} E\left[y_i\right]\right], \\
&= \left[\lambda^{-2} \lambda\right] = \lambda^{-1},
\end{aligned}
$$

and so

$$
\sqrt{n}\left(\hat{\lambda} - \lambda_0\right) \xrightarrow{d} N\left(0, \lambda\right),
$$

since $\mathcal{I}^{-1} = \lambda$.

Hypothesis testing

Null and alternative hypotheses

Econometrics models are estimated in order to test hypothesis, for e.g., whether a financial theory is supported by data.

Formal hypothesis testing begins by specifying the null and the alternative hypotheses

Definition (Null and alternative hypothesis)

The null hypothesis, denoted by H_0, is a statement about the population values of some parameters to be tested.

The alternative hypothesis, denoted by H_1 is the complementary hypothesis to the null.

Notice that:

- The null hypothesis cannot be accepted; the data can either lead to *rejection of the null* or a **failure to reject the null**.

Null and alternative hypotheses

Possible alternative tests

- **Two-sided alternative**

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_1 : \theta \neq \theta_0$$

- **One-sided alternative**

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_1 : \theta < \theta_0$$

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_1 : \theta > \theta_0$$

Test, critical value and errors

Three important concepts are: (1) the testing rule, (2) the critical value and (3) the Type I vs Type II errors.

Definition (Hypothesis test)

A hypothesis test is a rule that specifies which values to reject H_0 in favor of H_1. Hypothesis testing require, for e.g., a test statistic. The null is rejected when the t-statistics is larger than the critical value.

Definition (Critical value)

The critical value for a α-sized test, denoted by C_α, is the value where a test statistic, T, indicates rejection of the null hypothesis when the null is true.

The region where the test statistic is outside of the critical value, i.e., $T > C_\alpha$ is known as **rejection region**.

Test, critical value and Type I vs Type II errors

Since the sample is random, the test statistics T, however defined, is also random. As a result, the same test procedure can lead to different conclusions in different samples.

As such, there are always two ways such a procedure can be in error:

1. Reject the null hypothesis when it is in fact true - **Type I error**
2. Not reject the null hypothesis when it is false - **Type II error**

		Decision	
		Accept H_0	Reject H_0
Truth	H_0	Correct	Type I Error
	H_1	Type II Error	Correct

Test, critical value and Type I vs Type II errors

The **size** or level of a test, denoted by α, is the probability of rejecting the null when the null is true, i.e., probability of a Type I error.

The **power** of the test is the probability of rejecting the null when the alternative is true. The power is equivalently defined as 1 minus the probability of a Type II error β.

The significance level of the test is under the control of the analyst, and commonly fixed at $\alpha = (0.01,\ 0.05,\ 0.10)$

		Decision	
		Accept H_0	Reject H_0
Truth	H_0 True	Correct $(1-\alpha)$	Type I Error α
	H_0 False	Type II Error β	Correct $(1-\beta)$

P-value and confidence interval

Two other key concepts in hypothesis testing are the **p-value** and the **confidence interval**.

Definition (P-value)

The p-value is the probability of observing a value as large as the observed t-statistic given the null is true. The p-value is also:

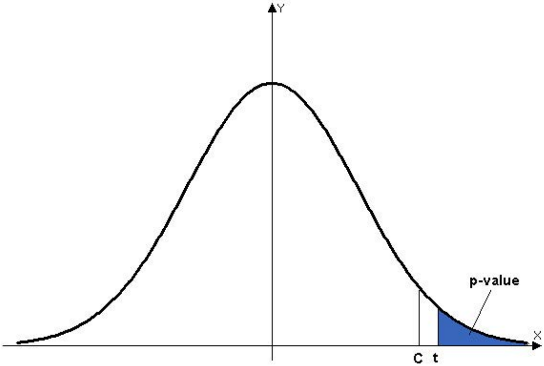
- The largest size α where the null hypothesis cannot be rejected
- The smallest size where the null hypothesis can be rejected

Two main advantages of using the p-value over the t-statistics:

- A p-value is that it immediately demonstrates which test sizes would lead to rejection: anything above the p-value.
- P-values can be interpreted without knowledge of the distribution of the test statistics.

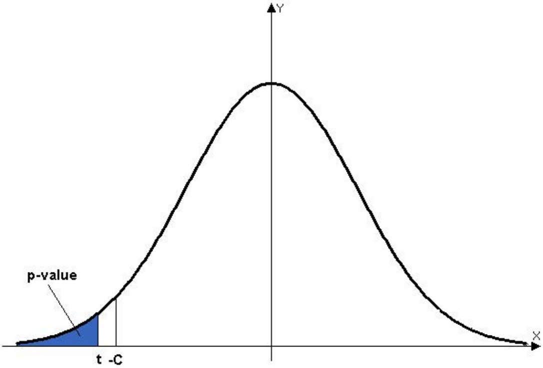
P-value and confidence interval

The p-value of a test is the probability of obtaining a test statistic more extreme than the observed sample value t, given that H_0 is true: p-value $= P(T > t)$



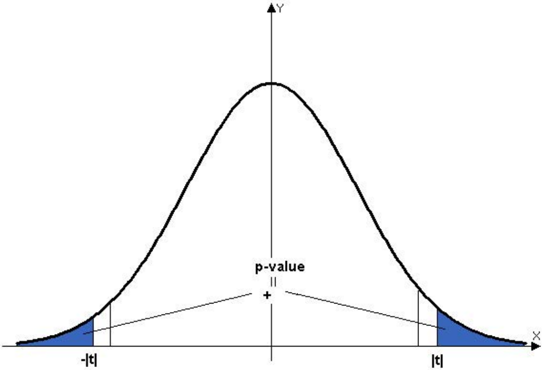
P-value and confidence interval

The p-value of a test is the probability of obtaining a test statistic more extreme than the observed sample value t, given that H_0 is true: p-value $= P(T < t)$



P-value and confidence interval

The p-value of a test is the probability of obtaining a test statistic more extreme than the observed sample value t, given that H_0 is true: p-value $= P(T > |t|) + P(T < -|t|)$



P-value and confidence interval

Definition (Confidence interval)

A confidence interval for a scalar parameter is the range values $\theta_0 \in \left(\underline{C}_\alpha, \overline{C}_\alpha\right)$ where the null $H_0 : \theta = \theta_0$ cannot be rejected for a size of α.

One comment is in order:

- The formal definition of a confidence interval is not usually sufficient to uniquely identify the conditional interval.
- Suppose that a $\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, \sigma^2\right)$.
- The common confidence interval is

$$\left(\hat{\theta} - 1.96\sigma, \hat{\theta} + 1.96\sigma\right),$$

- In general symmetric confidence intervals should be used, especially for asymptotically normal parameters.

Specifying hypothesis

Formalised in terms of θ, a null hypothesis is

$$H_0 : \mathbf{R}(\theta) = 0,$$

where $\mathbf{R}(\cdot)$ is a function from $\mathbb{R}^k$ to $\mathbb{R}^m$, $m \leq k$, where m represents the number of hypothesis in a composite null.

While this specification is very flexible, testing non-linear hypotheses on the parameters require some technical advancement which will not be covered in this module.

In this module, we will focus on linear equality restrictions of the form

$$H_0 : \mathbf{R}\theta - \mathbf{r} = 0,$$

where $\mathbf{R}$ is a $m \times k$ matrix and $\mathbf{r}$ is a $m \times 1$ vector.

Classical tests

Three classes of statistics will described to test hypotheses:

- Wald
- Lagrange Multiplier (LM)
- Likelihood Ratio (LR)

In this module we are going to focus only on the Wald and the Likelihood Ratio tests.

Wald tests

Wald tests are possibly the most natural method to test a hypothesis and are often the simplest to compute since only the unrestricted model must be estimated.

In this respect, Wald tests directly exploits the asymptotic normality of the estimated parameters to form test statistics with asymptotic χ^2_ν distributions.[2]

Given $\sqrt{n}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} N\left(0,\Sigma\right)$, and the linear form of the null hypothesis for the m restrictions $H_0 : \mathbf{R}\theta = \mathbf{r}$, the Wald statistics takes the form,

$$W = n\left(\mathbf{R}\hat{\theta}-\mathbf{r}\right)'\left(\mathbf{R}\Sigma\mathbf{R}'\right)^{-1}\left(\mathbf{R}\hat{\theta}-\mathbf{r}\right) \xrightarrow{d} \chi^2_m$$

[2]Recall that a χ^2_ν RV is defined as the sum of ν independent standard normals squared, i.e., $\sum_{i=1}^{\nu} z_i^2$ where $z_i \overset{iid}{\sim} N(0,1)$.

A t-test is a special case of a Wald and is applicable to tests involving a single hypothesis. Suppose the null is

$$H_0 : \mathbf{R}\theta - r = 0,$$

where $\mathbf{R}$ is a 1 by k, and so

$$\sqrt{n}\left(\mathbf{R}\hat{\theta} - r\right) \xrightarrow{d} N\left(0, \mathbf{R}\Sigma\mathbf{R}'\right),$$

The *studentized* version can be formed by subtracting the mean and dividing by the standard deviation,

$$t = \frac{\sqrt{n}\left(\mathbf{R}\hat{\theta} - r\right)}{\sqrt{\mathbf{R}\Sigma\mathbf{R}'}} \xrightarrow{d} N(0, 1),$$

and the test statistics can be compared to the critical values from a standard normal to conduct hypothesis test.

t-tests

t-tests are used in commonly encountered test statistic, the t-stat, a test of the null that a parameter is 0 against the alternative that it is not.

The t-stat is popular because most models are written in such a way that if a parameter $\theta = 0$ then it will have no impact.

Definition (t-stat)

The t-stat of a parameter θ_j is the t-test value of the null $H_0 : \theta_j = 0$ against a two-sided alternative $H_1 : \theta_j \neq 0$. It is defined as

$$t - stat \equiv \frac{\hat{\theta}_j}{\sigma_\theta}, \qquad \text{with} \qquad \sigma_\theta = \sqrt{\frac{\mathbf{e}_j \Sigma \mathbf{e}_j'}{n}},$$

where $\mathbf{e}_j$ is a vector of 0s with 1 in the jth position.

Likelihood ratio test

Likelihood ratio (LR) tests examine how "likely" the data are under the null and the alternative.

If the hypothesis is valid then the data should be (approximately) equally likely under each.

The LR test statistic is defined as

$$LR = -2\left(l\left(\tilde{\theta};\mathbf{y}\right) - l\left(\hat{\theta};\mathbf{y}\right)\right),$$

where $\tilde{\theta}$ is defined as

$$\tilde{\theta} = \arg\max_{\theta} l\left(\theta;\mathbf{y}\right) \qquad \text{s.t.} \qquad \mathbf{R}\theta - \mathbf{r} = 0,$$

whereas $\hat{\theta}$ is the unconstrained estimator.

N.B., Under the null $H_0 : \mathbf{R}\theta - \mathbf{r} = 0$, the $LR \xrightarrow{d} \chi^2_m$.

Choosing the tests

Choosing between different tests is far from trivial:

- Both tests have the same limiting distribution.
- They are both *asymptotically equivalent* in the sense they all have an identical asymptotic distribution and if one test reject, the other should reject as well.

As a result, there is no asymptotic argument that one should be favored over the other.

The simplest justifications for choosing one over the others are practical:

- If both the unrestricted and the restricted models are easy to estimate the LR is the natural choice (no parameters covariance needs to be estimated).
- If estimation under the alternative hypothesis is simpler, then Wald tests are reasonable.

Example: Testing the market risk premium

While most of the interest lies in predicting the market risk premium, testing whether such premium is different from zero is a natural application of what we learned in this chapter.

Let λ denote the **market premium** and σ^2 the variance of the return.

Since the market is a traded portfolio, it must be the case that the premium for holding market risk is the same as the mean of the market return.

We consider monthly data for the value-weighted market returns of the NYSE/AMEX and the risk free rate for the period 1927 to 2020.

Excess returns on the market are defined as the return to holding the market portfolio MKT minus the R_f risk-free rate.

Example: Testing the market risk premium

Figure 1: Market risk premium

These two charts show the time series of the market excess returns from July 1926 to October 2020. The left panel shows the time series of monthly excess returns while the right panel shows the distribution of the returns.

Example: Testing the market risk premium

The mean and the variance of the market excess returns can be computed using the MLE method as outlined above. The estimates are calculated according to

$$\begin{bmatrix} \hat{\lambda} \\ \hat{\sigma}^2 \end{bmatrix} = \begin{bmatrix} n^{-1} \sum_{i=1}^{n} MKT_i^e \\ n^{-1} \sum_{i=1}^{n} \left(MKT_i^e - \hat{\lambda} \right)^2 \end{bmatrix},$$

and defining $\hat{\epsilon}_i = MKT_i^e - \hat{\lambda}$, the covariance of the moment conditions is estimated by

$$\hat{\Sigma} = \begin{bmatrix} n^{-1} \sum_{i=1}^{n} \hat{\epsilon}_i^2 & 0 \\ 0 & 2 \left(n^{-1} \sum_{i=1}^{n} \hat{\epsilon}_i^2 \right)^2 \end{bmatrix},$$

such that the t-test on the market risk premium takes the form

$$t - \text{test} = \frac{\hat{\lambda}}{\sqrt{n^{-1} \sum_{i=1}^{n} \hat{\epsilon}_i^2}} = 4.2 > 1.96 = C_{\alpha=0.05},$$

Exercises

Exercises

Problem 2.1: Let $Y_i \sim N\left(\mu, \sigma^2\right)$. Derive the MLE estimators for μ, σ^2 where there are n observations.

Problem 2.2: Derive the asymptotic distribution of the MLE estimators calculated above.

Problem 2.3: When performing a hypothesis test, what are Type I and Type II errors?

Problem 2.4: Suppose $\{y_i\}$ are a set of transaction counts (trade counts) over 5-minute intervals which are believed to be i.i.d. distributed from a Poisson with parameter λ. Recall the PDF of a poisson

$$f(y_i;\lambda) = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!},$$

What is the log-likelihood for this problem? What is the MLE of λ? What is the variance of the MLE? Suppose that $\hat{\lambda} = 202.4$ and the sample size is 200; construct a 95% confidence interval for λ. Use a t-test to test the null $H_0 : \lambda = 200$ against $H_1 : \lambda \neq 200$ with a size of 5%.

Problem 2.5: Suppose $Y_i|X_i = x_i \overset{iid}{\sim} N\left(\beta_0 + \beta_1 x_i, \sigma^2\right)$. Write down the log-likelihood. Find the MLE for the unknown parameters. What is the asymptotic distribution of the parameters? Describe the t-test for $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$.