Toulouse School of Economics

M1 INTERMEDIATE ECONOMETRICS

Inference in the classical linear model

Koen Jochmans   François Poinas

2025 — 2026

Data on milk production in 247 dairy farms.

The variables available are

- milk production (in liters/year)
- number of milking cows,
- number of man-equivalent units,
- the number of hectares devoted to pasture and crops,
- the kilograms of feed fed to the dairy cows.

We postulate a Cobb-Douglas production function for output O, that is

$$O = A\, I_1^{\beta_1}\, I_2^{\beta_2}\, \cdots I_k^{\beta_k}.$$

Here I_k are the various inputs and β_k the associated output elasticities.

A is total factor productivity,

Taking logs gives

$$\log(O) = \log(I_1)\beta_1 + \cdots \log(I_k)\beta_k + \log(A).$$

This gives the regression model

$$Y = X'\beta + e$$

for $Y = \log(O)$, $X_k = \log(I_k)$, and $e = \log(A)$ which we estimate by least squares.

```
Linear regression                               Number of obs     =        247
                                                F(4, 242)         =     945.03
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9405
                                                Root MSE          =     .14126
```

log_milk	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
log_cows	.6712744	.046485	14.44	0.000	.5797074	.7628413
log_labor	.0251551	.033468	0.75	0.453	-.0407708	.091081
log_land	-.020511	.0239257	-0.86	0.392	-.0676402	.0266182
log_feed	.3888347	.0296418	13.12	0.000	.3304458	.4472235
_cons	5.42975	.2148372	25.27	0.000	5.00656	5.852939

The table by default reports quantities that can serve to test certain aspects of the model.

A more interesting question could be whether the dairy farms exhibit constant returns to scale in production.

This hypothesis translates to the restriction that

$$\beta_1 + \cdots + \beta_k = 1.$$

Is the data supportive of this hypothesis?

How do we approach such a question?

Hypotheses, statistics, and decision rules

First consider a linear contrast $\theta = r'\beta$ for some chosen (non-random) vector r.

Want to see if there is evidence in the data against the null hypothesis $\mathbb{H}_0 : \theta = \theta_0$ in favor of the alternative hypothesis $\mathbb{H}_1 : \theta \neq \theta_0$, where θ_0 is some chosen value.

Based on a decision rule involving a statistic (some function of the data):

Evaluate whether the distance $|\hat{\theta} - \theta_0|$ is 'large'.

Would like a decision rule to have good properties.

Mostly, size control and power.

We first look at this problem in the classical linear regression model.

When

$$Y = X'\beta + e, \qquad e|X \sim N(0, \sigma^2),$$

the estimator $\hat{\theta} = r'\hat{\beta}$ satisfies

$$\hat{\theta}|\boldsymbol{X} \sim N(\theta, \sigma^2 r'(\boldsymbol{X}'\boldsymbol{X})^{-1} r).$$

We will suppose that σ^2 is known, so only the mean is unknown here.

Under the null, $\theta = \theta_0$ is known.

So we know the distribution of the distance $|\hat{\theta} - \theta_0|$ under the null.

The distribution of $(\hat{\theta} - \theta_0)|\boldsymbol{X}$ depends on $\boldsymbol{X}$ through its variance.

It is more convenient to standardize.

Under the null,

$$\frac{\hat{\theta}-\theta_0}{\sigma\sqrt{r'(\boldsymbol{X}'\boldsymbol{X})^{-1}r}} \sim N(0,1).$$

Hence,

$$\mathbb{P}_{\theta_0}\left(\frac{\hat{\theta}-\theta_0}{\sigma\sqrt{r'(\boldsymbol{X}'\boldsymbol{X})^{-1}r}} \leq \varepsilon\right) = \Phi(\varepsilon),$$

and so

$$\mathbb{P}_{\theta_0}\left(\left|\frac{\hat{\theta}-\theta_0}{\sigma\sqrt{r'(\boldsymbol{X}'\boldsymbol{X})^{-1}r}}\right| > \varepsilon\right) = \Phi(-\varepsilon) + (1-\Phi(\varepsilon)) = 2(1-\Phi(\varepsilon)).$$

Size/significance level

Say we reject the null in favor of the alternative if

$$T = \left| \frac{\hat{\theta} - \theta_0}{\sigma \sqrt{r'(\boldsymbol{X}'\boldsymbol{X})^{-1} r}} \right| > c$$

for a chosen critical value c.

Then the probability of incorrectly rejecting the null is

$$\mathbb{P}_{\theta_0}(T > c) = 2(1 - \Phi(c))$$

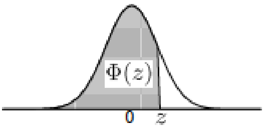
so if we want this probability to be equal to some $\alpha \in (0, 1)$ we choose

$$c = c_\alpha = \Phi^{-1}(1 - \alpha/2).$$

We call α the size (or significance level) of the test.

As $\alpha \downarrow 0$ we have that $c_\alpha \uparrow +\infty$. Cannot fully eliminate type-I errors.

The c.d.f. of the standard normal distribution



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891

The test in terms of a χ^2_1 variable

We have

$$\mathbb{P}_{\theta_0}(T > c_\alpha) = \mathbb{P}_{\theta_0}(T^2 > c_\alpha^2)$$

so we can equally look at the squared deviation from the null.

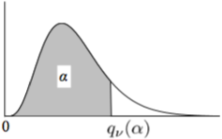
Under the null,

$$W = T^2 \sim \chi^2_1.$$

Here this leads to the same decision rule, but it generalizes easily to the multivariate case.

The χ^2-distribution

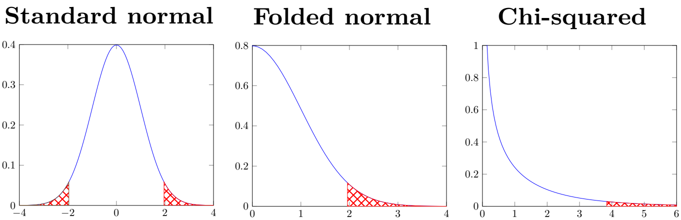
The quantile function of the χ^2 distribution



ν	0.500	0.600	0.700	0.800	0.850	0.900	α 0.925	0.950	0.975	0.990	0.995	0.999	0.995
1	0.455	0.708	1.074	1.642	2.072	2.706	3.170	3.841	5.024	6.635	7. 879	10.83	12.12
2	1.386	1.833	2.408	3.219	3.794	4.605	5.181	5.991	7.378	9.210	10 .60	13.82	15.20
3	2.366	2.946	3.665	4.642	5.317	6.251	6.905	7.815	9.348	11.34	12.84	16.27	17.73
4	3.357	4.045	4.878	5.989	6.745	7.779	8.496	9.488	11.14	13.28	14.86	18.47	20.00
5	4.351	5.132	6.064	7.289	8.115	9.236	10.01	11.07	12.83	15.09	16.75	20.52	22.11
6	5.348	6.211	7.231	8.558	9.446	10.64	11.47	12.59	14.45	16.81	18.55	22.46	24.10
7	6.346	7.283	8.383	9.803	10.75	12.02	12.88	14.07	16.01	18.4 8	20.28	24.32	26.02
8	7.344	8.351	9.524	11.03	12.03	13.36	14.27	15.51	17.53	20. 09	21.95	26.12	27.87
9	8.343	9.414	10.66	12.24	13.29	14.68	15.63	16.92	19.02	21 .67	23.59	27.88	29.67
10	9.342	10.47	11.78	13.44	14.53	15.99	16.97	18.31	20.48	23.21	25.19	29.59	31.42

Null distribution

Critical region for our test (with size .05, and so $c_{.05}. = 1.96$) depicted in three different manners.



p-value

The p-value is the probability, under the null, that we would observe a statistic at least as large as the one calculated from the data. It equals

$$2(1 - \Phi(T)).$$

Seeing a small p-value roughly means that our statistic is 'unusually' large.

Because critical values monotonically increase as the significance level goes down the p-value also represents a cut-off value of α where a decision of accepting the null turns into a decision of rejecting the null.

The p-value is the smallest significance level at which our test would lead to a rejection.

An example

In the dairy-farm data we have

$$\theta = r'\beta = (1, 1, \ldots, 1)\,\beta$$

and we test $\mathbb{H}_0 : \theta = 1$.

Here,

$$T = 2.083, \qquad W = 4.34.$$

At the 5%-level the relevant critical values are, respectively 1.96 (for T) and 3.84 (for W).

We thus reject the null of constant returns to scale at this significance level.

The p-value is 0.0384. So we would not be able to reject the null at the 1% level, for example.

Power

The power is the probability of rejecting the null when it is false.

It equals

$$\mathbb{P}_\theta(T > c_\alpha)$$

and depends on θ.

Because

$$\frac{\hat{\theta} - \theta_0}{\sigma\sqrt{r'(\boldsymbol{X}'\boldsymbol{X})^{-1}r}} = \frac{\hat{\theta} - \theta}{\sigma\sqrt{r'(\boldsymbol{X}'\boldsymbol{X})^{-1}r}} + \frac{\theta - \theta_0}{\sigma\sqrt{r'(\boldsymbol{X}'\boldsymbol{X})^{-1}r}}$$

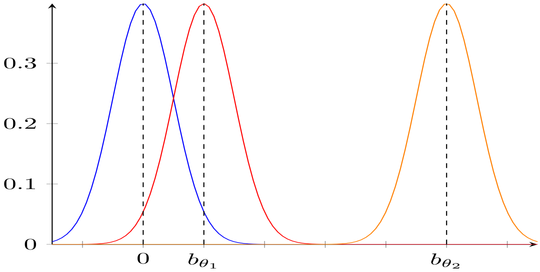
and the first right-hand side term is standard normal, we have that

$$\mathbb{P}_\theta(T > c_\alpha|\boldsymbol{X}) = \Phi\left(-c_\alpha - b_\theta\right) + \left(1 - \Phi(c_\alpha - b_\theta)\right)$$

for

$$b_\theta = \frac{\theta - \theta_0}{\sigma\sqrt{r'(\boldsymbol{X}'\boldsymbol{X})^{-1}r}}.$$

b_θ shows how a violation of the null is reflected in the distribution of our statistic.

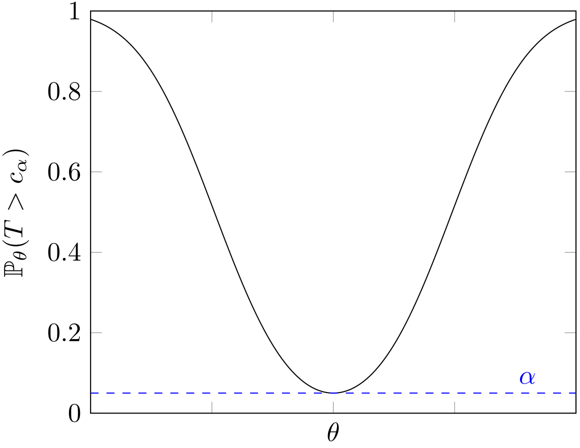


In terms of a χ_1^2 random variable

The square of a normal variable with mean zero and variance one follows a Chi-squared distribution.

The square of a normal variable with mean b_θ and variance follows a non-central Chi-squared distribution with non-centrality parameter $b'_\theta b_\theta$.

Power function



Confidence intervals

Let

$$T_{\theta_*} = \left| \frac{\hat{\theta} - \theta_*}{\sigma \sqrt{r'(\boldsymbol{X}'\boldsymbol{X})^{-1} r}} \right| .$$

The set

$$C = \{\theta_* : T_{\theta_*} \leq c_\alpha\}$$

collects all values θ_* for which we would not reject the null hypothesis $\mathbb{H}_0 : \theta = \theta_*$ with a test of size α.

The probability that the set C covers/contains the true value θ is equal to

$$\mathbb{P}_\theta(\theta \in C) = \mathbb{P}_\theta(T_\theta \leq c_\alpha) = \Phi(c_\alpha) - \Phi(-c_\alpha) = 1 - \alpha.$$

The set C is called a confidence interval with coverage probability $1-\alpha$.

Easy to see that, here,

$$C = \left[\hat{\theta} - c_\alpha \, \sigma \sqrt{r'(\boldsymbol{X}'\boldsymbol{X})^{-1} r} \, , \, \hat{\theta} + c_\alpha \, \sigma \sqrt{r'(\boldsymbol{X}'\boldsymbol{X})^{-1} r} \right] .$$

Multiple linear restrictions

Now we wish to test a collection of m linear contrasts jointly.

Can write this as testing $\mathbb{H}_0 : \theta = \theta_0$ for

$$\theta = R'\beta = \begin{pmatrix} r_1'\beta \\ r_2'\beta \\ \vdots \\ r_m'\beta \end{pmatrix},$$

with $R = (r_1, r_2, \ldots, r_m)$.

Testing m restrictions jointly is not the same as testing m restrictions separately.

Size control is difficult in multiple testing.

Like before we have

$$\hat{\theta}|\boldsymbol{X} \sim N(\theta, \sigma^2 R' (\boldsymbol{X}'\boldsymbol{X})^{-1} R),$$

or, equivalently,

$$\frac{(R' (\boldsymbol{X}'\boldsymbol{X})^{-1} R)^{-1/2}}{\sigma} (\hat{\theta} - \theta) \sim N(0, I_m),$$

which is now multivariate.

The squared distance of the standardized estimator from the null is thus

$$W = (\hat{\theta} - \theta_0)' \frac{(R' (\boldsymbol{X}'\boldsymbol{X})^{-1} R)^{-1}}{\sigma^2} (\hat{\theta} - \theta_0)$$

and follow a χ^2_m distribution under the null.

The decision rule remains to reject the null when $W > c_\alpha$, where the critical value c_α is chosen to control size at α.

We do this, as before, by taking c_α to be the $(1-\alpha)$th quantile of the χ^2_m distribution.

The power analysis is as before, only now with the distribution under the alternative becoming non-central Chi-squared with m degrees of freedom.

Like before, we can get a confidence set by 'inverting' a test statistic.

Let

$$W_{\theta_*} = (\hat{\theta} - \theta_*)' \frac{(R' \, (\boldsymbol{X}'\boldsymbol{X})^{-1} R)^{-1}}{\sigma} (\hat{\theta} - \theta_*)$$

Then

$$C = \{\theta_* : W_{\theta_*} \leq c_\alpha\}$$

is a confidence ellipsoid with coverage probability $1 - \alpha$.