Nonlinear Regression

Algorithmic Trading and Quantitative Strategies[1]

Petter Kolm

January 31, 2023

 If you find any typos, I would much appreciate you email them to me.

Readings

Later in this course, we will be using *nonlinear least squares* (NLS) to estimate market impact models. Then you will see an example of how nonlinear regression can be used for estimation and hypothesis testing. In this lecture we cover the basic aspects of NLS.

The following is optional reading:

- Kuan (2004) provides an asymptotic analysis of NLS.
- NLS is an example of so-called *extremum estimators*. Hayashi (2000) offers a general asymptotic analysis of a large class of extremum estimators.

Nonlinear Least Squares

We assume as before that we have collected data

$$\mathcal{D} := (\mathbf{X}, \mathbf{y}) \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{n \times p}$ and $\mathbf{y} \in \mathbb{R}^{n}$.

We want to estimate a nonlinear regression function of the form

$$y = f(\mathbf{x}, \boldsymbol{\beta}) + \varepsilon\,,\ \boldsymbol{\beta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^{p} \tag{2}$$

where $\boldsymbol{\Theta}$ is the "permissible" parameter space.

Here we will use the l^2-loss function. Nonlinear regression with this loss is referred to as *nonlinear least squares* (NLS).

Assumptions:

1. $\mathbb{E}[y \mid \mathbf{x}] = f(\mathbf{x}, \boldsymbol{\beta}_0)$ for some $\boldsymbol{\beta}_0 \in \boldsymbol{\Theta}$
2. $\mathbb{E}[(f(\mathbf{x}, \boldsymbol{\beta}_0) - f(\mathbf{x}, \boldsymbol{\beta}))^2] > 0$ for all $\boldsymbol{\beta} \in \boldsymbol{\Theta}, \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$

Solving NLS I

The nonlinear least squares estimator $\widehat{\boldsymbol{\beta}}$ is found by solving

$$\min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}} \mathrm{SSR}(\boldsymbol{\beta}) := \min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}} \frac{1}{2}\sum_{i=1}^{n}\left(y_i - f\left(\mathbf{x}_i, \boldsymbol{\beta}\right)\right)^2 \ .$$

Differentiating with respect to $\boldsymbol{\beta}$ we obtain the FOCs

$$\frac{\partial \mathrm{SSR}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\sum_{i=1}^{n} \nabla_{\boldsymbol{\beta}} f\left(\mathbf{x}_i, \boldsymbol{\beta}\right)\left(y_i - f\left(\mathbf{x}_i, \boldsymbol{\beta}\right)\right) = \mathbf{0}\,,$$

which represent a nonlinear system in $\boldsymbol{\beta}$ that can be solved numerically with a gradient descent or Newton method.

Solving NLS II

We can write the FOCs compactly as

$$\nabla_{\boldsymbol{\beta}} f(\mathbf{X}, \boldsymbol{\beta})' f(\mathbf{X}, \boldsymbol{\beta}) = \nabla_{\boldsymbol{\beta}} f(\mathbf{X}, \boldsymbol{\beta})' \mathbf{y} \,. \tag{3}$$

Of course, equation (3) reduces to the normal equations when $f(\mathbf{X}, \boldsymbol{\beta}) \equiv \mathbf{X}\boldsymbol{\beta}$.

Summary of Key Results for NLS I

Under the assumptions above, we have the following results:

1. NLS is consistent, i.e. $\widehat{\boldsymbol{\beta}} \to_p \boldsymbol{\beta}_0$.
2. The (asymptotic) variance of the NLS estimator is

$$\mathrm{var}[\widehat{\boldsymbol{\beta}}] = \left(\frac{1}{n}\sum_{i=1}^{n} \nabla_{\boldsymbol{\beta}} \widehat{f}_i' \nabla_{\boldsymbol{\beta}} \widehat{f}_i\right)^{-1} \cdot \left(\frac{1}{n}\sum_{i=1}^{n} \widehat{\varepsilon}_i^2 \nabla_{\boldsymbol{\beta}} \widehat{f}_i' \nabla_{\boldsymbol{\beta}} \widehat{f}_i\right) \cdot \left(\frac{1}{n}\sum_{i=1}^{n} \nabla_{\boldsymbol{\beta}} \widehat{f}_i' \nabla_{\boldsymbol{\beta}} \widehat{f}_i\right)^{-1},$$

   where $\widehat{\varepsilon}_i := y_i - f(\mathbf{x}_i, \widehat{\boldsymbol{\beta}})$.[2]
3. As always, the asymptotic standard error of each element of $\widehat{\boldsymbol{\beta}}$ is the square root of the appropriate diagonal element of the $\mathrm{var}[\widehat{\boldsymbol{\beta}}]$ matrix. This is a *heteroskedasticity consistent standard error* $\mathrm{SE}^*(\widehat{\beta}_i)$ for $\widehat{\beta}_i$.

Summary of Key Results for NLS II

4. For hypothesis testing we use the same t-, F- and Wald-ratios as for OLS.[3]

Remark: As for OLS, there is no significant additional computational cost associated with using the heteroskedasticity-robust formula above. Therefore, these should be used in practice.

[2] Just like in the linear case, this expression is sometimes multiplied by $n/(n-p)$ as a degrees of freedom adjustment, where p is the dimension of $\widehat{\beta}$. Of course, for n large, this adjustment has not impact.

[3] For you: Why is this the case?

References

Hayashi, Fumio (2000). *Econometrics*. Princeton University Press.

Kuan, Chung-Ming (2004). *Nonlinear Least Squares Theory*. URL: https://homepage.ntu.edu.tw/~ckuan/pdf/et01/ch8.pdf.