ECON 613: Applied Econometrics

Methods

January 21, 2019

Overview: Linear Models

- Study the relationship between an outcome variable y and a set of regressors x.
  - Conditional Prediction.
  - Causal inference.
  - Example: propensity to consume.
- Loss function approach

$$L(e) = L(y - \hat{y})$$

where $\hat{y} = E(y \mid x)$ is a predictor of y, and the error $e = y - \hat{y}$

Squared Loss Function

- Squared error loss: $L(e) = e^2$
- Optimization problem

$$\min_{\beta} \sum_{i}^{N} (y_i - f(x_i, \beta))^2$$

Linear Prediction

- $E[y \mid x] = x'\beta$
- OLS

$$y = x\beta + e$$

- Derivation

$$\begin{aligned} L(\beta) &= (y - x\beta)'(y - x\beta) \\ &= y'y - 2y'x\beta + \beta'X'X\beta \end{aligned}$$

Then

$$\frac{\partial L(\beta)}{\partial \beta} = -2x'y + 2x'x\beta = 0$$

- Formula

$$\hat{\beta} = (x'x)^{-1}x'y$$

Properties

see 4.4.4 and 4.4.5.

Properties of an estimator

- Unbiasedness: $E(\hat{\theta}) = \theta$.
- Consistency: $plim \hat{\theta}_n = \theta$.
- Efficiency: Reach Cramer-Rao lower bound asymptotically.

Codes

```
////R
fitR = lm(Y~X)

////Matlab
fitM = fitlm(X,Y,'linear')

////Stata
fitM = reg Y X
```

Codes

```
////R

####  define X matrix and y vector
X = as.matrix(cbind(1,X))
y = as.matrix(Y)

####  estimate the coeficients beta
####  beta = ((X'X)^(-1))X'y
beta  = solve(t(X)%*%X)%*%t(X)%*%y
```

Maximum Likelihood

GMM

Boostrap

Numerical Optimization

Introduction to MLE

Consider a parametric model in which the joint distribution of $Y = (Y_1, \ldots, Y_n)$ has a density $\ell(y, \theta)$ with respect to a measure μ. Then consider $P_\theta = \ell(y, \theta)\mu$ where $\theta \in \Theta \in \mathbb{R}^p$. Once $y = (y_1, \ldots, y_n)$ is observed, the maximum likelihood method consists of estimating the parameter θ a value $\hat{\theta}(y)$ that maximizes the likelihood function $\theta \to \ell(y, \theta)$. Formally, a maximum likelihood estimator of θ is a solution to the maximization problem

$$\max_\theta \ell(Y; \theta)$$

or

$$\max_\theta \log(\ell(Y; \theta))$$

Feasible examples: Poisson distribution

Consider a dependent variable that takes only non negative integer values $0, 1, 2, \ldots$, and one assumes that the dependent variable follows a Poisson distribution, and we wishes to estimate the Poisson parameter.

- Given $y_i \sim f(\lambda, y_i) = \frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}$
- Likelihood $\mathcal{L}(y; \lambda) = \prod_{i=1}^{N} \frac{\exp(-\lambda)\lambda^{y_i}}{y_i!} = \frac{\exp(-N\lambda)\lambda^{\sum_i^N y_i}}{\prod_{i=1}^{N} y_i!}$
- Log likelihood
  $\log \mathcal{L}(y; \lambda) = -N\lambda + \sum_i^N y_i \log(\lambda) - \sum_i^N \log(y_i!)$
- Estimate

$$\frac{\partial \log \mathcal{L}(y; \lambda)}{\partial \lambda} = 0 \Longrightarrow \widehat{\lambda} = \frac{\sum_i^N y_i}{N}$$

Feasible examples: Least Squares

- Normality assumption $e \sim \mathbb{N}(0, \sigma^2)$, then $y \sim \mathbb{N}(x\beta, \sigma^2)$.
- Likelihood $L(\beta) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp(-0.5\sigma^{-2}(y - x\beta)'(y - x\beta))$
- log likelihood $\log L(\beta) = -\frac{N}{2} \log \sigma^2 - \frac{1}{2\sigma^2}(y - x\beta)'(y - x\beta)$
- $\beta = (x'x)^{-1}x'y$

Some difficulties

- Non-uniqueness of the Likelihood Function
- Non-existence of a solution to the Maximization Problem
- Multiple Solutions to the Maximization Problem

Asymptotic Properties (1): Convergence

Definition

Under a set of regularity conditions, there exists a sequence of maximum likelihood estimators converging almost surely to the true parameter value θ_0

- The variables $Y_i, i = 1, 2, \ldots$ are independent and identically distributed with density $f(y; \theta), \theta \in \Theta \in \mathbb{R}^p$
- The parameter space Θ is compact.
- The log likelihood function $\mathcal{L}(y, \theta)$ is continuous in θ and is a measurable function of y.
- The log-likelihood function is such that $(1/n)\mathcal{L}_n(y, \theta)$ converges surely to $E_{\theta_0} log(f(Y_i; \theta))$ uniformly in $\theta \in \Theta$. $E_{\theta_0} log(f(Y_i; \theta))$ exists.

Asymptotic Properties (2): Asymptotic Normality

- The log likelihood function $\mathcal{L}_n(\theta)$ is twice continuously differentiable in an open neighborhood of θ_0
- The matrix (Fisher Information Matrix)

$$\mathcal{I}_1(\theta_0) = E_{\theta_0}\left(-\frac{\partial^2 \log f(Y_1;\theta_0)}{\partial\theta\partial\theta'}\right)$$

Definition

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to \mathbb{N}(0, \mathcal{I}_1(\theta_0)^{-1}).$$

Concentrated Likelihood Function

Definition

Let the parameter set $\theta = (\alpha, \beta)$. The solutions $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ to the mazimization problem $\max_{\alpha,\beta} \log \mathcal{L}(y; \alpha, \beta)$ can be obtained via the following two-step procedure:

a) Maximize the log-likelihood function with respect α given β. The maximum value is attained for values of α in a set $A(\beta)$ depending on the parameter β. Thus, if $\alpha \in A(\beta)$, the log-likelihood value is

$$\log \mathcal{L}_c(y; \beta) = \max_{\alpha} \log \mathcal{L}(y; \alpha, \beta)$$

The mapping $\log \mathcal{L}_c$ is called the concentrated (in α) log likelihood function.

b) In a second step, maximize the concentrated log-likelihood function with respect to β.

Application

Consider the likelihood

$$\mathcal{L}(y,\beta,\sigma) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y - x\beta)'(y - x\beta)$$

- First step

$$\frac{\partial\mathcal{L}(y;\beta,\sigma)}{\partial\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}(y - x\beta)'(y - x\beta) = 0$$

  Then

$$\sigma^2(\beta) = \frac{1}{n}(y - x\beta)'(y - x\beta)$$

- Substituting $\sigma^2(\beta)$ into the likelihood

$$\mathcal{L}_c(y,\beta,\sigma) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\frac{1}{n}(y - x\beta)'(y - x\beta) - \frac{n}{2}$$

Hypothesis Testing

Three procedures to do tests

Likelihood Ratio

- The likelihood ratio statistic is

$$LR = 2(\ell(\theta) - \ell(\tilde{\theta}))$$

where $\hat{\theta}$ and $\tilde{\theta}$ are the restricted and unrestricted maximum likelihood estimates of θ.
- Wilk's theorem shows that

$$LR \sim \chi^2(r)$$

where r is the number of restrictions.

Additional Tests

- Wald Test
- LM test

We will see in GMM.

In practice

- The regularity conditions are strong.
- What happens if we weaken them?

Problem 1

Number of parameters increases with the number of observations

- Convergence holds
- Estimates may be biased

Problem 2

True parameter value θ_0 does not belong to Θ: The model is misspecified

- Convergence holds to a parameter that is not the true parameter.

Problem 3

Correlated Observations

- Convergence does not hold.

Problem 3

Discontinuity of the likelihood function

- Numerical problems.

Maximum Likelihood

GMM

Boostrap

Numerical Optimization

Method of Moments

- Orthogonality condition in Linear Models

$$E(x(y' - x)) = 0 \tag{1}$$

- Moment Condition

$$\frac{1}{N}\sum_i x_i(y_i - x'\beta) \tag{2}$$

- Moment Estimator

$$\hat{\beta}_{\text{MM}} = (\sum_i x_i x_i')^{-1}(\sum_i x_i y_i) \tag{3}$$

Nonlinear Model

- Consider

$$Y_i = g(X_i, b_0) + u_i$$

- Orthogonality Condition

$$E[X'(y - g(X, b_0))] = 0$$

- Moments condition

$$E_0 h(Y, X, a_0) = 0$$

- The function h is H-dimensional and the parameter a is of size K.

Formal Idea

Definition

The basic idea of generalized method of moments is to choose a value for *a* such that the sample mean is closest to zero.

$$\frac{1}{n}\sum_{i=1}^{n} h(Y_i, X_i, a)$$

Formal Definition

Definition

Let $\mathbb{S}_n$ be an $(H \times H)$ symmetric positive definite matrix that may depend on the observations. The generalized method of moments (GMM) estimator associated with $\mathbb{S}_n$ is a solution $\tilde{a}_n(\mathbb{S}_n)$ to the problem

$$min_a \left[\sum_{i=1}^{n} h(Y_i, X_i, a)\right]' \mathbb{S}_n \left[\sum_{i=1}^{n} h(Y_i, X_i, a)\right]$$

Assumptions

H1 The variables (Y_i, X_i) are independent and identically distributed.

H2 The expectation $E_0 h(Y, X, a)$ exists and is zero when a is equal to the true value a_0 of the parameter of interest.

H3 The matrix $\mathbb{S}_n$ converges almost surely to a nonrandom matrix $\mathbb{S}_0$

H4 The parameter a_0 is identified from the equality constraints, i.e. $E_0 h(Y, X, a)' \mathbb{S}_0 E_0 h(Y, X, a) = 0$

H5 The parameter value a_0 is known to belong to a compact set $\mathcal{A}$

H6 The quantity $(1/n) \sum_{i=1}^{n} h(Y_i, X_i, a)$ converges almost surely and uniformly in a to $E_0 h(Y, X, a)$

H7 The function $h(Y, X, a)$ is continuous in a

H8 The matrix $\left[E_0 \frac{h(Y, X, a)}{\partial a} \right]' \mathbb{S}_0 \left[E_0 \frac{h(Y, X, a)}{\partial a'} \right]$ is nonsingular, which implies $H \geq K$.

Asymptotic Normality

Under the assumptions, we have

$$\sqrt{n}(\tilde{a}_n(\mathbb{S}_n) - a_0) \sim \mathbb{N}(0, \Sigma(\mathbb{S}_0))$$

where

$$\Sigma(\mathbb{S}_0) = \left(\left[E_0 \frac{h(Y,X,a)}{\partial a} \right]' \mathbb{S}_0 \left[E_0 \frac{h(Y,X,a)}{\partial a'} \right] \right)^{-1}$$
$$\left(\left[E_0 \frac{h(Y,X,a)}{\partial a} \right]' \mathbb{S}_0 V_0(h(Y,X,a_0)) \mathbb{S}_0 \left[E_0 \frac{h(Y,X,a)}{\partial a'} \right] \right)^{-1}$$
$$\left(\left[E_0 \frac{h(Y,X,a)}{\partial a} \right]' \mathbb{S}_0 \left[E_0 \frac{h(Y,X,a)}{\partial a'} \right] \right)^{-1}$$

Optimal GMM

- $\mathbb{S}_0$ is not known.
- Two-step procedure
  - Estimate

$$min_a \left[\sum_{i=1}^{n} h(Y_i, X_i, a)\right]' I \left[\sum_{i=1}^{n} h(Y_i, X_i, a)\right]$$

    where I is the identity matrix, and recover $\hat{a}$.
  - Matrix of variance/covariance

$$\hat{\mathbb{S}} = \frac{1}{N} \sum_{i=1}^{n} h(Y_i, X_i, \hat{a}) h(Y_i, X_i, \hat{a})'$$

Relationship to IV.

- Nonlinear 2SLS is a very good application of GMM.

Inference

- Over identification test see iv section.

Applications

- Matrix of variance/covariance in practice
- Indirect Inference
- Simulated method of moments