An Introduction to Empirical Likelihood

Textbooks

- Greene, W.H., 2012, *Econometric Analysis* (7th ed.), Chapter 12.3.2.
- Cameron, A.C. and Trivedi, P.K., 2005, *Microeconometrics: Methods and Applications*, Chapter 6.8.
- Hansen, B.E., 2013, *Econometrics*, Chapter 14.
- Anatolyev, S. and Gospodinov, N., 2011, *Methods for Estimation and Inference in Modern Econometrics*, Chapter 2.

Section 1

Introduction

Introduction

- Method of Moments vs. Maximum Likelihood.
- Many economic models can be written in the form

$$E[g(Z_i, \beta)] = 0$$

where $\beta \in \mathcal{B}$ is a k-dimensional parameter of interest, and $g(\cdot, \cdot)$ is a $\mathbb{R}^m$ valued function. $k \leq m$.

Introduction

- Method of Moments vs. Maximum Likelihood.
- Many economic models can be written in the form

$$E[g(Z_i, \beta)] = 0$$

where $\beta \in \mathcal{B}$ is a k-dimensional parameter of interest, and $g(\cdot, \cdot)$ is a $\mathbb{R}^m$ valued function. $k \leq m$.

Examples I

Linear model with endogenous variables and instruments

- Structural equation $Y_i = X_i\beta_0 + \epsilon_i$
- X_i is an endogenous variable.
- W_i is a vector of instruments.
- We estimate β_0 via the moment condition

$$E\left[\epsilon_i W_i\right] = E\left[(y_i - X_i\beta)W_i\right] = 0.$$

Examples I

Linear model with endogenous variables and instruments

- Structural equation $Y_i = X_i\beta_0 + \epsilon_i$
- X_i is an endogenous variable.
- W_i is a vector of instruments.
- We estimate β_0 via the moment condition

$$E\left[\epsilon_i W_i\right] = E\left[(y_i - X_i\beta) W_i\right] = 0.$$

Examples II

Asset pricing in Dynamic Rational Expectation Model.

- Hansen and Singleton (1982).
- The rational expectation theory implies

$$E_t \left[b \frac{U'(C_{t+j}, \beta_0)}{U'(C_t, \beta_0)} X_{t+j} - 1 \right] = 0$$

where b is a discount factor, $U(\cdot)$ is a utility function known up to a parameter β, and $X_{t+j} = (P_{t+j} + D_{t+j})/P_{t+j}$ is the interest rate.

Outline

1. Review GMM and introduce Empirical Likelihood (EL)
2. Generalize EL and develop its asymptotic theory
3. Extend EL into high-dimensional moments

Outline

1. Review GMM and introduce Empirical Likelihood (EL)
2. Generalize EL and develop its asymptotic theory
3. Extend EL into high-dimensional moments

Outline

1. Review GMM and introduce Empirical Likelihood (EL)
2. Generalize EL and develop its asymptotic theory
3. Extend EL into high-dimensional moments

Section 2

Empirical Likelihood

Generalized Method of Moments

- Let $E_n[\cdot] = n^{-1} \sum_{i=1}^{n} \cdot$ be the empirical mean.
- The GMM estimator

$$\widehat{\beta}_{\text{GMM}} = \arg\min_{\beta \in \mathcal{B}} E_n[g(Z_i, \beta)]' W E_n[g(Z_i, \beta)],$$

  where W is a positive-definite weighting matrix.
- GMM gives a unified framework of estimation for many economic models.
- Empirical Likelihood (EL) is an alternative.

Generalized Method of Moments

- Let $E_n[\cdot] = n^{-1}\sum_{i=1}^{n} \cdot$ be the empirical mean.
- The GMM estimator

$$\widehat{\beta}_{\text{GMM}} = \arg\min_{\beta \in \mathcal{B}} E_n[g(Z_i, \beta)]' W E_n[g(Z_i, \beta)],$$

  where W is a positive-definite weighting matrix.
- GMM gives a unified framework of estimation for many economic models.
- Empirical Likelihood (EL) is an alternative.

Likelihood Approach

- Consider a random sample from population distribution $F(x)$ with density $f(x)$.
- The joint likelihood function is

$$\prod_{i=1}^{n} f(x_i).$$

- The parametric method assumes that $f(\cdot)$ is known up to a finite-dimensional parameter.

Nonparametric Likelihood

- A nonparametric method assigns probability p_i to each observation x_i for $i = 1, \ldots, n$.
- It can be viewed as a sampling experiment from the multinomial population concentrated at the sample points.
- The Nonparametric Maximum Likelihood Estimator (NPMLE) solves

$$\max_{\mathbf{p}} \prod_{i=1}^{n} p_i \text{ s.t. } p_i \geq 0, \sum_{i=1}^{n} p_i = 1,$$

or equivalently

$$\max_{\mathbf{p}} \sum_{i=1}^{n} \log p_i \text{ s.t. } \sum_{i=1}^{n} p_i = 1.$$

Nonparametric Likelihood

- A nonparametric method assigns probability p_i to each observation x_i for $i = 1, \ldots, n$.
- It can be viewed as a sampling experiment from the multinomial population concentrated at the sample points.
- The Nonparametric Maximum Likelihood Estimator (NPMLE) solves

$$\max_{\mathbf{p}} \prod_{i=1}^{n} p_i \ \text{ s.t. } p_i \geq 0, \ \sum_{i=1}^{n} p_i = 1,$$

or equivalently

$$\max_{\mathbf{p}} \sum_{i=1}^{n} \log p_i \ \text{ s.t. } \sum_{i=1}^{n} p_i = 1.$$

Solution

- To solve the constrained optimization problem,

$$\mathcal{L}(\mathbf{p}, \mu) = \frac{1}{n}\sum_{i=1}^{n} \log p_i - \mu \left(\sum_{i=1}^{n} p_i - 1\right).$$

- The first-order condition gives

$$\partial\mathcal{L}/\partial p_i = (np_i)^{-1} - \mu = 0,$$

  for all $i = 1, \ldots, n$.
- As $\sum_{i=1}^{n} p_i = 1$, we have $\mu = 1$, so that

$$\widehat{p}_i = 1/n.$$

- This is exactly the Empirical Distribution Function.

Solution

- To solve the constrained optimization problem,

$$\mathcal{L}(\mathbf{p}, \mu) = \frac{1}{n}\sum_{i=1}^{n} \log p_i - \mu \left(\sum_{i=1}^{n} p_i - 1\right).$$

- The first-order condition gives

$$\partial\mathcal{L}/\partial p_i = (np_i)^{-1} - \mu = 0,$$

for all $i = 1, \ldots, n$.

- As $\sum_{i=1}^{n} p_i = 1$, we have $\mu = 1$, so that

$$\widehat{p}_i = 1/n.$$

- This is exactly the Empirical Distribution Function.

Solution

- To solve the constrained optimization problem,

$$\mathcal{L}(\mathbf{p}, \mu) = \frac{1}{n}\sum_{i=1}^{n} \log p_i - \mu \left(\sum_{i=1}^{n} p_i - 1\right).$$

- The first-order condition gives

$$\partial\mathcal{L}/\partial p_i = (np_i)^{-1} - \mu = 0,$$

  for all $i = 1, \dots, n$.
- As $\sum_{i=1}^{n} p_i = 1$, we have $\mu = 1$, so that

$$\widehat{p}_i = 1/n.$$

- This is exactly the Empirical Distribution Function.

EL with Moment Constraints

- Again, the model is $E[g(Z_i, \beta)] = 0$.
- The EL problem is

$$\max_{\beta, \mathbf{p}} \sum_{i=1}^{n} \log p_i$$

$$\text{s.t. } \sum_{i=1}^{n} p_i g(Z_i, \beta) = 0 \text{ and } \sum_{i=1}^{n} p_i = 1.$$

- If $k = m$, then

$$\widehat{\beta} = \arg_{\beta} \left(\frac{1}{n} \sum_{i=1}^{n} g(Z_i, \beta) = 0 \right),$$

which is the method of moments.

EL with Moment Constraints

- Again, the model is $E[g(Z_i, \beta)] = 0$.
- The EL problem is

$$\max_{\beta, \mathbf{p}} \sum_{i=1}^{n} \log p_i$$

$$\text{s.t. } \sum_{i=1}^{n} p_i g(Z_i, \beta) = 0 \text{ and } \sum_{i=1}^{n} p_i = 1.$$

- If $k = m$, then

$$\widehat{\beta} = \arg_{\beta} \left(\frac{1}{n} \sum_{i=1}^{n} g(Z_i, \beta) = 0 \right),$$

which is the method of moments.

EL with Moment Constraints

- Again, the model is $E[g(Z_i, \beta)] = 0$.
- The EL problem is

$$\max_{\beta,\mathbf{p}} \sum_{i=1}^{n} \log p_i$$

$$\text{s.t. } \sum_{i=1}^{n} p_i g(Z_i, \beta) = 0 \text{ and } \sum_{i=1}^{n} p_i = 1.$$

- If $k = m$, then

$$\widehat{\beta} = \arg_\beta \left(\frac{1}{n} \sum_{i=1}^{n} g(Z_i, \beta) = 0 \right),$$

which is the method of moments.

Solution when $k < m$

- The Lagrangian

$$\mathcal{L}(\beta, \mathbf{p}, \lambda, \mu) = \frac{1}{n}\sum_{i=1}^{n} \log p_i - \lambda' \sum_{i=1}^{n} p_i g(Z_i, \beta) - \mu \left(\sum_{i=1}^{n} p_i - 1 \right).$$

- The first order condition with respect to p_i,

$$(np_i)^{-1} - \lambda' g(Z_i, \beta) - \mu = 0. \tag{1}$$

- Multiple both sides by p_i and sum over i,

$$1 - \lambda' \sum_{i=1}^{n} p_i g(Z_i, \beta) - \mu \sum_{i=1}^{n} p_i = 1 - \mu = 0.$$

Solution when $k < m$

- The Lagrangian

$$\mathcal{L}(\beta, \mathbf{p}, \lambda, \mu) = \frac{1}{n}\sum_{i=1}^{n} \log p_i - \lambda' \sum_{i=1}^{n} p_i g(Z_i, \beta) - \mu \left(\sum_{i=1}^{n} p_i - 1 \right).$$

- The first order condition with respect to p_i,

$$(np_i)^{-1} - \lambda' g(Z_i, \beta) - \mu = 0. \tag{1}$$

- Multiple both sides by p_i and sum over i,

$$1 - \lambda' \sum_{i=1}^{n} p_i g(Z_i, \beta) - \mu \sum_{i=1}^{n} p_i = 1 - \mu = 0.$$

Solution when $k < m$

- The Lagrangian

$$\mathcal{L}(\beta, \mathbf{p}, \lambda, \mu) = \frac{1}{n}\sum_{i=1}^{n} \log p_i - \lambda' \sum_{i=1}^{n} p_i g(Z_i, \beta) - \mu \left(\sum_{i=1}^{n} p_i - 1 \right).$$

- The first order condition with respect to p_i,

$$(np_i)^{-1} - \lambda' g(Z_i, \beta) - \mu = 0. \tag{1}$$

- Multiple both sides by p_i and sum over i,

$$1 - \lambda' \sum_{i=1}^{n} p_i g(Z_i, \beta) - \mu \sum_{i=1}^{n} p_i = 1 - \mu = 0.$$

Solution when $k < m$ (Continue)

- Plug in $\mu = 1$ into (1), $p_i = [n(1 + \lambda' g(Z_i, \beta))]^{-1}$, which is slightly different from $1/n$.
- Substitute the expression of p_i into the criterion function, we obtain the EL estimator $\widehat{\beta}$ and multiplier $\widehat{\lambda}$ via the saddlepoint problem

$$\max_{\beta \in \mathcal{B}} \min_{\lambda} - \sum_{i=1}^{n} \log\left(1 + \lambda' g(Z_i, \beta)\right).$$

Computation

- The inner loop and the outer loop

$$\max_{\beta}\left\{\min_{\lambda(\beta)}\left[-\sum_{i=1}^{n}\log\left(1+\lambda(\beta)'g(Z_i,\beta)\right)\right]\right\}.$$

- The inner loop is globally convex in λ.
- The outer loop is neither convex nor concave in general in β.

Computation

- The inner loop and the outer loop

$$\max_{\beta}\left\{\min_{\lambda(\beta)}\left[-\sum_{i=1}^{n}\log\left(1+\lambda(\beta)'g(Z_i,\beta)\right)\right]\right\}.$$

- The inner loop is globally convex in λ.
- The outer loop is neither convex nor concave in general in β.

Differences from GMM

- GMM: Two-step, iterative, continuous-updating (Hansen, Heaton and Yaron, 1996). Basic idea: stick to $p_i = 1/n$.
- Altonji and Segal (1996): large bias in two-step GMM
- EL: more flexible **p**. One-step estimator. Self-normalize.
- High-order improvement
  - Kitamura(2001): Asymptotic optimality under the generalized Neyman-Pearson criterion
  - Newey and Smith (2004): High-order bias correction

Example (Imbens, 1997)

- The logarithm of hourly wage of 827 men in 1971–1978.
- The model is borrowed from Card (1994)

$$\begin{aligned} \ln y_{it} &= \mu_t + \omega_i + u_{it} + \varepsilon_{it} \\ u_{it} &= \alpha u_{it-1} + \eta_{it} \end{aligned}$$

  where μ_t is a common time-varying component, ω_i is the individual fixed effect, η_{it} is the shock to the autoregressive component.
- Unknown parameters: $\mu_1, \ldots, \mu_T$; $\sigma^2_{\eta,1}, \ldots, \sigma^2_{\eta,T}$; σ^2_{ε}; σ^2_{ω} and α.
- Imbens (1997) constructs 44 moments to estimate the 19 parameters. He compares two-step GMM, iterated GMM and EL.

Example (Imbens, 1997)

- Real data estimates

	2S-GMM	GMMi	EL	s.e.
σ_ω^2	0.110	0.111	0.123	0.053
σ_ε^2	0.040	0.039	0.041	0.003
α	0.913	0.912	0.899	0.057

- Simulation: true value and bias divided by the average of s.e.

	True value	2S-GMM	GMMi	EL
σ_ω^2	0.10	-0.28	-0.28	-0.16
σ_ε^2	0.05	-0.31	-0.30	-0.22
α	0.50	0.05	0.06	0.02

Example (Imbens, 1997)

- Real data estimates

	2S-GMM	GMMi	EL	s.e.
σ_ω^2	0.110	0.111	0.123	0.053
σ_ε^2	0.040	0.039	0.041	0.003
α	0.913	0.912	0.899	0.057

- Simulation: true value and bias divided by the average of s.e.

	True value	2S-GMM	GMMi	EL
σ_ω^2	0.10	-0.28	-0.28	-0.16
σ_ε^2	0.05	-0.31	-0.30	-0.22
α	0.50	0.05	0.06	0.02

Literature

- Owen (1988): Original paper on EL
- Qin and Lawless (1993): EL with estimating equations
- Imbens (1997): Introduce EL into econometrics
- Kitamura (1997): EL in time series
- Newey and Smith (2004): Generalized EL

Section 3

Generalized Empirical Likelihood

Generalized Empirical Likelihood

- Let $\rho(\cdot)$ be a smooth scalar function that satisfies

$$\rho(0) = 0 \text{ and } \partial\rho(0)/\partial v = \partial^2\rho(0)/\partial v^2 = -1.$$

  GEL estimator solves

$$\min_{\beta} \sup_{\lambda} \sum_{i=1}^{n} \rho\left(\lambda' g(Z_i, \beta)\right).$$

- The Generalized Empirical Likelihood (GEL) contains several important estimators as special cases.
  - EL: $\rho(v) = \log(1 - v)$,
  - continuous-updating: $\rho(v) = -\frac{1}{2}v^2 - v$,
  - Exponential Tilting (Kitamura and Stutzer, 1997): $\rho(v) = 1 - \exp(v)$.

Asymptotic Properties

- Let $J \equiv E\left[\partial g(Z, \beta_0)/\partial \beta'\right]$ and $V \equiv E\left[g(Z, \beta_0) g(Z, \beta_0)'\right]$.
- With random sampling,

$$\sqrt{n}\left(\widehat{\beta} - \beta_0\right) \Rightarrow \mathrm{N}\left(0, \left(J' V^{-1} J\right)^{-1}\right).$$

- First-order asymptotically equivalent to the efficient GMM.

Asymptotic Tests

- Wald test, LM test and overidentification test.
- Moreover, it inherits a *likelihood ratio*-type test. Let $(\widehat{\beta}, \widehat{\lambda})$ and $(\tilde{\beta}, \tilde{\lambda})$ be the estimates without and with constraints, respectively. Then

$$GELR = 2\sum_{i=1}^{n}\left[\rho\left(\tilde{\lambda}'g(Z_i, \tilde{\beta})\right) - \rho\left(\widehat{\lambda}'g(Z_i, \widehat{\beta})\right)\right].$$

- Under the null,

$$GELR \Rightarrow \chi_q^2,$$

  where q is the number of restrictions.
- Convenience of LR test: no need to estimate the variance.

Asymptotic Tests

- Wald test, LM test and overidentification test.
- Moreover, it inherits a *likelihood ratio*-type test. Let $(\widehat{\beta}, \widehat{\lambda})$ and $(\tilde{\beta}, \tilde{\lambda})$ be the estimates without and with constraints, respectively. Then

$$GELR = 2\sum_{i=1}^{n}\left[\rho\left(\tilde{\lambda}'g(Z_i, \tilde{\beta})\right) - \rho\left(\widehat{\lambda}'g(Z_i, \widehat{\beta})\right)\right].$$

- Under the null,

$$GELR \Rightarrow \chi_q^2,$$

  where q is the number of restrictions.
- Convenience of LR test: no need to estimate the variance.

Example

- Overidentification Test for EL. $\rho(v) = \log(1 - v)$.
- H_0: Correct model specification. $E[g(Z, \beta_0)] = 0$.
  - Likelihood without restriction: $-n \log n$
  - Likelihood with restriction: $-n \log n - \sum_{i=1}^{n} \log\left(1 + \widehat{\lambda}' g\left(Z_i, \widehat{\beta}\right)\right)$.
- Test statistic

$$ELR = 2 \sum_{i=1}^{n} \log\left(1 + \widehat{\lambda}' g\left(Z_i, \widehat{\beta}\right)\right) \Rightarrow \chi_m^2.$$

Example

- Overidentification Test for EL. $\rho(v) = \log(1 - v)$.
- H_0: Correct model specification. $E[g(Z, \beta_0)] = 0$.
  - Likelihood without restriction: $-n \log n$
  - Likelihood with restriction: $-n \log n - \sum_{i=1}^{n} \log\left(1 + \widehat{\lambda}' g\left(Z_i, \widehat{\beta}\right)\right)$.
- Test statistic

$$ELR = 2 \sum_{i=1}^{n} \log\left(1 + \widehat{\lambda}' g\left(Z_i, \widehat{\beta}\right)\right) \Rightarrow \chi^2_m.$$

Section 4

Relaxed Empirical Likelihood

High-Dimensional Moments

- Large data, large model vs. parsimonious modeling.
  - Altonji, Smith and Vidangos (2013)
  - Eaton, Kortum and Kramarz (2011)
  - Han, Orea, Schmidt (2005)
  - Angrist (1990)

High-Dimensional Moments

- Large data, large model vs. parsimonious modeling.
  - Altonji, Smith and Vidangos (2013)
  - Eaton, Kortum and Kramarz (2011)
  - Han, Orea, Schmidt (2005)
  - Angrist (1990)

High-Dimensional Moments

- Large data, large model vs. parsimonious modeling.
  - Altonji, Smith and Vidangos (2013)
  - Eaton, Kortum and Kramarz (2011)
  - Han, Orea, Schmidt (2005)
  - Angrist (1990)

What Happens when $m > n$

- Problem of optimally-weighted GMM.

$$\mathbb{E}_n\left[g_i(\beta)\right]' \; W_{\text{GMM}} \; \mathbb{E}_n\left[g_i(\beta)\right]$$

- Problem of EL. For any $\beta \in \mathcal{B}$, the constraints

$$\sum_{i=1}^{n} p_i = 1 \text{ and } \sum_{i=1}^{n} p_i g_i(\beta) = \; 0_m$$

What Happens when $m > n$

- Problem of optimally-weighted GMM.

$W_{\text{GMM}} = \widehat{V}^{-1}$
$\widehat{V}$: $m \times m$, rank n

$$\mathbb{E}_n [g_i(\beta)]' \; W_{\text{GMM}} \; \mathbb{E}_n [g_i(\beta)]$$

- Problem of EL. For any $\beta \in \mathcal{B}$, the constraints

$$\sum_{i=1}^{n} p_i = 1 \text{ and } \sum_{i=1}^{n} p_i g_i(\beta) = 0_m$$

What Happens when $m > n$

- Problem of optimally-weighted GMM.

$$\mathbb{E}_n \left[g_i\left(\beta\right)\right]' \; W_{\text{GMM}} \; \mathbb{E}_n \left[g_i\left(\beta\right)\right]$$

- Problem of EL. For any $\beta \in \mathcal{B}$, the constraints

$$\sum_{i=1}^{n} p_i = 1 \text{ and } \sum_{i=1}^{n} p_i g_i\left(\beta\right) = 0_m$$

What Happens when $m > n$

- Problem of optimally-weighted GMM.

$$\mathbb{E}_n \left[g_i \left(\beta \right) \right]' \; W_{\mathrm{GMM}} \; \mathbb{E}_n \left[g_i \left(\beta \right) \right]$$

- Problem of EL. For any $\beta \in \mathcal{B}$, the constraints

$$\sum_{i=1}^{n} p_i = 1 \text{ and } \sum_{i=1}^{n} p_i g_i \left(\beta \right) = 0_m$$

m equations
n free parameters

Extension: Relaxed EL

- Shi (2013) proposes the first asymptotically normal estimator that allows $m > n$.
- Relax equality constraints of the standard EL problem by inequality constraints

$$\max_{\beta,\mathbf{p}} \sum_{i=1}^{n} \log p_i$$

$$\text{s.t.} \quad \sum_{i=1}^{n} p_i = 1,$$

$$\left| \sum_{i=1}^{n} p_i h_j(Z_i, \beta) \right| \leq \tau, \forall\, j = 1, \ldots, n.$$

where τ is a tuning parameter, and $h_j = g_j / \widehat{\sigma}_j$ is the scale-standardized function.

Extension: Relaxed EL

- Shi (2013) proposes the first asymptotically normal estimator that allows $m > n$.
- Relax equality constraints of the standard EL problem by inequality constraints

$$\max_{\beta,\mathbf{p}} \sum_{i=1}^{n} \log p_i$$

$$\text{s.t.} \quad \sum_{i=1}^{n} p_i = 1,$$

$$\left| \sum_{i=1}^{n} p_i h_j(Z_i, \beta) \right| \leq \tau, \forall\, j = 1, \ldots, n.$$

where τ is a tuning parameter, and $h_j = g_j / \widehat{\sigma}_j$ is the scale-standardized function.

Summary

1. What is EL and why
2. GEL and the asymptotic theory
3. REL for high-dimensional moments