Some Recent Developments in Microeconometrics

A. Colin Cameron
Univ of Calif. - Davis

July 21, 2006

1. INTRODUCTION

- *Presented to the 23rd Annual Summer Meeting of the Society for Political Methodology, July 20-22, 2006, University of California - Davis. These slides are available at cameron.econ.ucdavis.edu. A completed paper will be available end September 2006. This work draws considerably on Cameron and Trivedi (2005).*
- By late 1970's well-established theory for
  - LS, ML and IV in
  - nonlinear cross-section and linear panel models.
- This survey considers more recent microeconometrics methods.
- Microeconometrics emphasizes
  - causative inference
  - controlling for heterogeneity
  - potentially nonlinear models.

Example: Earnings and Schooling

- Interested in the causative effect of a one year increase in education

$$\left.\frac{\partial y_i}{\partial s_i}\right|_{\mathbf{x}_{2i}=\mathbf{x}_2^*}.$$

- Simple linear model is

$$y_i = \alpha s_i + \mathbf{x}_{2i}'\boldsymbol{\beta}_2 + u_i.$$

- **Complications** include:
  - **causation:** s_i is selected by the individual and likely endogenous
  - **heterogeneity:** the marginal effect may differ across individuals
  - **nonlinearity:** the relationship may be nonlinear

Outline of Talk

1. INTRODUCTION
2. STATISTICAL INFERENCE
3. ESTIMATION METHODS
4. CAUSATION
5. DATA ISSUES

2. STATISTICAL INFERENCE

1. Robust Inference
2. Bootstrap
3. Weak Instruments

Notation

- Typical observation
$$\mathbf{w}_i = (y_i, \mathbf{x}_i, \mathbf{z}_i),$$
where y_i and $\mathbf{x}_i$ are as usual and $\mathbf{z}_i$ is (optional) **instrument**.
- $\boldsymbol{\theta}$ is a generic $q \times 1$ parameter vector
- Assume independence over i (or sometimes clustering)
- Linear regression model has $k \times 1$ parameter vector $\boldsymbol{\beta}$ and
$$\begin{aligned} y_i &= \mathbf{x}_i'\boldsymbol{\beta} + u_i \\ \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u}. \end{aligned}$$

2.1 Robust Inference

- Consider **m-estimator** $\widehat{\boldsymbol{\theta}}$ such as ML or OLS that solves

$$\sum_i \mathbf{h}_i(\mathbf{w}_i, \boldsymbol{\theta}) = \mathbf{0}.$$

- Then $\widehat{\boldsymbol{\theta}}$ is asymptotically normal with **sandwich variance matrix**

$$\mathrm{V}[\widehat{\boldsymbol{\theta}}] = \left[\sum_i \left. \frac{\partial \mathbf{h}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\widehat{\boldsymbol{\theta}}} \right]^{-1} \sum_i \mathbf{h}_i(\widehat{\boldsymbol{\theta}}) \mathbf{h}_i(\widehat{\boldsymbol{\theta}})' \left[\sum_i \left. \frac{\partial \mathbf{h}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\widehat{\boldsymbol{\theta}}} \right]^{-1}.$$

- This leads to **robust or sandwich standard errors**.
  - This is the robust option in STATA.

Robust Inference: Examples

- White (1980) showed this for **OLS with heteroskedastic errors.**
  - Big impact
  - Do inefficient OLS rather than efficient feasible GLS
  - But get standard errors that are correct.
- White (1982) and Huber (1967) did this for **quasi-MLE**.
  - Do inefficient quasi-MLE
  - But get standard errors that are correct
  - e.g. For counts do Poisson not negative binomial.
- Hansen (1982) did this for **GMM**.
- Amemiya (1985) and Newey and McFadden (1994) give quite **general treatments** of estimation and inference.

Robust Inference: Clustering

- White (1984), Arellano (1987) and Liang and Zeger (1986) adapted this to **clustering**.
  - This is the cluster option in STATA.
  - Do inefficient estimator assuming independence.
  - But get standard errors that are correct.
- In practice the **number of clusters may be small** e.g. 10. For inference
  - Cameron, Gelbach and Miller (2006a) propose cluster version of the Wild bootstrap
  - Donald and Lang (2004) propose alternative two-step grouping estimator and use of $t(G-2)$ distribution where $G =$ number of clusters. See also Angrist and Lavy (2002).
- Cameron, Gelbach and Miller (2006b) extend one-way cluster robust to **multi-way clustering**.

2.2 Bootstrap Methods

- **Bootstrap** due to Efron (1979)
  - provides an alternative asymptotic approximation for the distribution of a statistic
  - does so by viewing the sample as the population and obtaining B resamples leading to B realizations of the statistic
  - there are many, many ways to bootstrap.
- A **bootstrap without asymptotic refinement**
  - is no better than regular asymptotic theory
  - though is popular as it may be simpler to implement
  - leading example is bootstrap estimate of standard error.
- A **bootstrap with asymptotic refinement**
  - is asymptotically better than regular asymptotic theory
  - hopefully then does better in finite samples
  - leading example is the bootstrap-t method.
- **Microeconometricians** rarely do bootstrap with asymptotic refinement.

Bootstrap standard errors (no asymptotic refinement)

1. For data $\mathbf{w}_1, ..., \mathbf{w}_N$ do the following B times
   - Draw a bootstrap resample $\mathbf{w}_1^*, ..., \mathbf{w}_N^*$ by sampling with replacement from the original data (bootstrap pairs)
   - Obtain estimate $\widehat{\theta}^*$ of θ, where for simplicity θ is scalar.
2. The **bootstrap estimate of standard error** is simply the standard error of the B estimates $\widehat{\theta}_1^*, ..., \widehat{\theta}_B^*$:

$$s_{\widehat{\theta},\text{Boot}} = \sqrt{\frac{1}{B-1}\sum\nolimits_{b=1}^{B}(\widehat{\theta}_b^* - \overline{\widehat{\theta}}^*)^2}, \text{ where } \overline{\widehat{\theta}}^* = B^{-1}\sum\nolimits_{b=1}^{B}\widehat{\theta}_b^*.$$

3. To test $H_0 : \theta = \theta_0$ use $t = (\widehat{\theta} - \theta_0)/s_{\widehat{\theta},\text{Boot}}$.
4. This is **asymptotically no better than a regular Wald test**.

Bootstrap-t procedure (asymptotic refinement)

1. For data $\mathbf{w}_1, ..., \mathbf{w}_N$ do the following B times
   - Draw a bootstrap resample $\mathbf{w}_1^*, ..., \mathbf{w}_N^*$ by sampling with replacement from the original data (bootstrap pairs)
   - Obtain estimate $\widehat{\theta}^*$, standard error $s_{\widehat{\theta}^*}$ and t-statistic $t^* = (\widehat{\theta}^* - \theta_0)/s_{\widehat{\theta}^*}$.
2. The empirical distribution of the B t-statistics $t_1^*, ..., t_B^*$, is used to estimate the distribution of $t = (\widehat{\theta} - \theta_0)/s_{\widehat{\theta}}$ computed from the original sample.
3. In particular, on a nonsymmetric two-sided test at 5 percent reject $H_0 : \theta = \theta_0$ if t is less than the 2.5 percentile or more than the 97.5 percentile of $t_1^*, ..., t_B^*$.
4. **Asymptotic refinement**
   with test size $= 0.05 + O(N^{-1})$ rather than $0.05 + O(N^{-0.5})$.
5. Reason: It bootstraps t (not $\widehat{\theta}$), and t is **asymptotically pivotal** (meaning no unknown parameters as $\mathcal{N}[0,1]$ asymptotically).

Bootstrap Extensions

- Theory for asymptotic refinement based on **Edgeworth expansions**: Beran (1982), Hall (1992).
- **Microeconometrics literature:**
    - Bootstrap for over-identified GMM model recenters: Hall and Horowitz (1996), Brown and Newey (2002).
    - Number of bootstraps: Andrews and Buchinsky (2000), and Davidson and MacKinnon (2000).
    - Horowitz (2001) surveys bootstrap theory and MacKinnon (2000) practice.
    - For OLS with clustered data Cameron, Gelbach and Miller (2006a) apply a cluster version of the Wild bootstrap.

Bootstrap Extensions (continued)

- **Bootstrap in nonstandard settings** - nonsmooth estimators and less than $\sqrt{N}$-consistent estimator - is focus of current theory work.
  - Politis and Romano (1994) propose subsampling
  - Abrevaya and Huang (2005) for maximum score estimator
  - Abadie and Imbens (2006a) for matching treatment effects estimators
  - Moreira, Porter, and Suarez (2004) for IV with weak instruments

2.3 Weak Instruments

- **OLS is inconsistent** in model $y_i = \mathbf{x}_i'\boldsymbol{\beta} + u_i$ if $\text{Cor}[\mathbf{x}_i, u_i] \neq \mathbf{0}$.
- Assume there exists **instrument** $\mathbf{z}_i$ such that $\text{Cor}[\mathbf{z}_i, u_i] = \mathbf{0}$.
  The **IV estimator** for a just-identified model is $\widehat{\boldsymbol{\beta}}_{\text{IV}} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$.
  $\widehat{\boldsymbol{\beta}}_{\text{IV}}$ is asymptotically normal with

$$\text{V}[\widehat{\boldsymbol{\beta}}_{\text{IV}}] = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\Sigma\mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1}, \text{ where } \Sigma = \text{E}[\mathbf{uu}'|\mathbf{Z}].$$

- A **weak instrument** is one for which $\text{Cor}[\mathbf{z}_i, \mathbf{x}_i]$ is small. Then
  - $\widehat{\boldsymbol{\beta}}_{\text{IV}}$ is imprecise (this is well-known).
  - $\widehat{\boldsymbol{\beta}}_{\text{IV}}$ can be more inconsistent than OLS if $\text{Cor}[\mathbf{z}_i, u_i]$ departs slightly from zero (this is pointed out by Bound et al. (1995) but ignored).
  - $\widehat{\boldsymbol{\beta}}_{\text{IV}}$ can be biased and quite nonnormal even in large samples.

The Problem of Weak Instruments

- The last of these is called the **problem of weak instruments**.
  - Regular asymptotic theory performs poorly in finite samples.
  - Theoreticians established key results early, e.g. Nagar (1959).
  - Applied researchers to highlight the problem were Nelson and Startz (1990) and Bound, Jaeger and Baker (1995).
  - Staiger and Stock (1997) provided influential theory.
- **Big impact on microeconometrics**
  - Applied researchers need to show there is no weak instruments problem, typically by first-stage F-test exceeding 10.
  - There is a big theoretical literature, including new testing procedures. Andrews and Stock (2005) provide recent survey.
  - There is movement away from using IV to measure causation.

3. ESTIMATION METHODS

1. Generalized Method of Moments
2. Simulation-Based Estimation
3. Markov Chain Monte Carlo for Bayesian Analysis
4. Empirical Likelihood
5. Quantile Regression
6. Semiparametric Methods

3.1 Generalized Method of Moments

- Starting point is **moment condition** $\mathsf{E}[\mathbf{h}_i(\mathbf{w}_i, \boldsymbol{\theta})] = \mathbf{0}$.
- In **just-identified case** method of moments solves $\sum_i \mathbf{h}_i(\boldsymbol{\theta}) = \mathbf{0}$.
- In **over-identified case** this is not feasible as more equations ($\dim[\mathbf{h}_i]$) than unknowns ($\dim[\boldsymbol{\theta}]$).
- The **generalized method of moments (GMM) estimator** $\widehat{\boldsymbol{\theta}}_{\text{GMM}}$ maximizes the quadratic form

$$Q(\boldsymbol{\theta}) = \left[\sum\nolimits_i \mathbf{h}_i(\boldsymbol{\theta})\right] \mathbf{W} \left[\sum\nolimits_i \mathbf{h}_i(\boldsymbol{\theta})\right],$$

  where $\mathbf{W}$ is a $\dim[\mathbf{h}] \times \dim[\mathbf{h}]$ symmetric weighting matrix
- Example is **two-stage least squares (2SLS)**
  - $\mathsf{E}[(y - \mathbf{x}'\beta)\mathbf{z}] = 0$ where $\dim[\mathbf{z}] > \dim[\mathbf{x}]$.
  - $Q(\beta) = (\mathbf{y} - \mathbf{X}\beta)'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\beta)$ so here $\mathbf{W} = (\mathbf{Z}'\mathbf{Z})^{-1}$.

GMM: Discussion

- $\widehat{\theta}_{\text{GMM}}$ is asymptotically normal with variance matrix

$$\mathsf{V}[\widehat{\theta}_{\text{GMM}}] = (\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{W}\Sigma\mathbf{W}\mathbf{G}(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1},$$

  and $\mathbf{G} = \sum_i \partial \mathbf{h}_i(\theta)/\partial\theta$ and $\Sigma = \mathsf{V}[\sum_i \mathbf{h}_i(\theta)]$.
- Hansen (1982) proposed GMM.
- Optimal GMM (given choice of $\mathbf{h}_i(\theta)$) uses $\mathbf{W} = \widehat{\Sigma}^{-1}$.
  But this is found to work poorly in finite samples.
- GMM
  - Can be viewed as a generalization of 2SLS
  - Nests many other estimation procedures including ML and LS
  - Peculiar to econometrics and widely used as a framework
  - Used in econometrics GMM when others would use the more specialized generalized linear models.

3.2 Simulation-Based Estimation

- Suppose **conditional density** of y given regressors **x,** unobservables **u**, and parameters $\boldsymbol{\theta} = [\boldsymbol{\theta}_1' \ \boldsymbol{\theta}_2']'$ **is an integral**

$$f(y|\mathbf{x}, \boldsymbol{\theta}) = \int f(y|\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}_1) g(\mathbf{u}|\boldsymbol{\theta}_2) d\mathbf{u}.$$

- Problems if $f(y|\mathbf{x}, \boldsymbol{\theta})$ is **not of closed form**.
- For low dimension **u** can use **Gaussian quadrature**.
- For high dimension **u** use **Monte Carlo methods**.

Maximum Simulated Likelihood

- The **MSL estimator** maximizes the **simulated log-likelihood function**

$$\widehat{L}_N(\boldsymbol{\theta}) = \sum_{i=1}^{N} \ln \widehat{f}(y_i|\mathbf{x}_i, \mathbf{u}_i^{(S)}, \boldsymbol{\theta}),$$

- Here $\widehat{f}(\cdot)$ is a Monte Carlo estimate or **simulator**

$$\widehat{f}(y_i|\mathbf{x}_i, \mathbf{u}_i^{(S)}, \boldsymbol{\theta}) = \frac{1}{S}\sum_{s=1}^{S} f(y_i|\mathbf{x}_i, \mathbf{u}_i^{s}, \boldsymbol{\theta}),$$

where $\mathbf{u}_i^{(S)} = (\mathbf{u}_i^1, ..., \mathbf{u}_i^S)$ are S draws $\mathbf{u}_i^s$ with marginal density $g(\mathbf{u}_i|\boldsymbol{\theta}_2)$.

- Many possible simulators exist - require $\widehat{f}_i \xrightarrow{p} f_i$ as $S \rightarrow \infty$.
- MSLE $\overset{LD}{=}$ MLE if $N \rightarrow \infty$ and additionally $S \rightarrow \infty$ (need $N/S \rightarrow \infty$).

Simulation-Based Estimation: Discussion

- Potential to estimate **rich parametric models**.
- Leading applications are to **flexible multinomial models**
  - Multinomial probit with more than four choices
  - Random parameters logit.
- **Computationally expensive** plus many tricks including Halton sequences and antithetic sampling.
- **Method of Simulated Moments (MSM)** is less computational
  - Suppose an unbiased simulator exists.
  - Then need as little as $S = 1$ draws for each observation (though there is an efficiency loss).
  - Not applicable to MLE as there is no unbiased simulator for $\ln f_i$.
  - Due to McFadden (1989) and Pakes and Pollard (1989).
- **Books** by Gouriéroux and Monfort (1996) and Train (2003).

3.3 Markov Chain Monte Carlo for Bayesian Analysis

- **Standard Bayesian setup** with **posterior density**

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{L(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{f(\mathbf{y})},$$

  with likelihood $L(\mathbf{y}|\boldsymbol{\theta})$, prior $\pi(\boldsymbol{\theta})$ and normalizing constant $f(\mathbf{y}) = \int L(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ (conditioning on $\mathbf{X}$ is suppressed).
- **Closed form** for $p(\boldsymbol{\theta}|\mathbf{y})$ exists only in special cases.
  e.g. normal likelihood plus normal prior yields normal prior.
- Can use **numerical integration** to approximate e.g. posterior mean.
  e.g. Importance sampling - see Geweke (1989).
- **Modern methods** instead use **Monte Carlo integration**, yielding draws $\widehat{\boldsymbol{\theta}}_1, ..., \widehat{\boldsymbol{\theta}}_S$ from the posterior.
- **Additional advantage** is that given $\widehat{\boldsymbol{\theta}}_1, ..., \widehat{\boldsymbol{\theta}}_S$ can summarize many features of the posterior, not just the posterior mean.

MCMC for Bayesian Analysis: Methods

- The **Gibbs sampler** provides one way to make draws
  - Suppose $\boldsymbol{\theta} = [\boldsymbol{\theta}_1' \ \boldsymbol{\theta}_2']'$ and it is possible to draw from the conditional posteriors $p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{y})$ and $p(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1, \mathbf{y})$.
  - Begin with initial $\boldsymbol{\theta}_1^{(0)}$, draw $\boldsymbol{\theta}_2^{(1)}$ from $p(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1^{(0)}, \mathbf{y})$, then draw $\boldsymbol{\theta}_1^{(1)}$ from $p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2^{(1)}, \mathbf{y})$, etc.
  - By Markov chain theory can show that eventually get draws $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ from the unconditional posterior $p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y})$.
- The **Metropolis-Hastings algorithm** can be used when the conditional posteriors.

MCMC for Bayesian Analysis: Discussion

- Many **subtleties**
  - Often a mix of closed-form, Gibbs and MH
  - Convergence can be slow and hard to establish
  - Can do Bayesian inference or choose weak prior and do classical inference.
- In **microeconometrics**
  - Especially useful for limited dependent variables models where can use data augmentation (e.g. Chib (1992) for Tobit model)
  - Chib (2001) and books by Koop (2003) and Lancaster (2004)
  - Perhaps used more in other fields.

3.4 Empirical Likelihood

- $\pi_i = f(y_i|\mathbf{x}_i)$ denotes the **probability** that the i^{th} observation on y has realized value y_i.
- So maximize the **empirical log-likelihood function** $N^{-1}\sum_i \ln \pi_i$ w.r.t. $\pi_1, ..., \pi_N$, subject to any model constraints.
- The moment condition $\mathrm{E}[\mathbf{h}(\mathbf{w}_i, \boldsymbol{\theta})] = \mathbf{0}$ imposes the constraint that

$$\sum_{i=1}^{N} \pi_i \mathbf{h}(\mathbf{w}_i, \boldsymbol{\theta}) = \mathbf{0}.$$

- So maximize wrt to $\boldsymbol{\pi} = [\pi_1 ... \pi_N]'$, η, $\boldsymbol{\lambda}$, and $\boldsymbol{\theta}$ the Lagrangian

$$\mathcal{L}_{\mathrm{EL}}(\boldsymbol{\pi}, \eta, \boldsymbol{\lambda}, \boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \ln \pi_i - \eta\left(\sum_{i=1}^{N} \pi_i - 1\right) - \boldsymbol{\lambda}' \sum_{i=1}^{N} \pi_i \mathbf{h}(\mathbf{w}_i, \boldsymbol{\theta}).$$

Empirical Likelihood: Discussion

- Then the **EL estimator** $\widehat{\boldsymbol{\theta}}_{\text{EL}}$ is asymptotically normal with

$$\widehat{\text{V}}[\widehat{\boldsymbol{\theta}}_{\text{EL}}] = \left[\sum\nolimits_i \widehat{\pi}_i \left.\frac{\partial \mathbf{h}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\widehat{\boldsymbol{\theta}}}\right]^{-1} \sum\nolimits_i \widehat{\pi}_i \mathbf{h}_i(\widehat{\boldsymbol{\theta}}) \mathbf{h}_i(\widehat{\boldsymbol{\theta}})' \left[\sum\nolimits_i \widehat{\pi}_i \left.\frac{\partial \mathbf{h}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\widehat{\boldsymbol{\theta}}}\right]^{-1}.$$

- **Advantage:** Asymptotically equivalent to MM and GMM, but adding weights $\widehat{\pi}_i$ improves finite sample performance.
  Newey and Smith (2004) show that GEL has better second-order properties than GMM.
- **Disadvantage:** Difficult to compute $\widehat{\boldsymbol{\theta}}_{\text{EL}}$.
- **Literature**:
  - Due to Qin and Lawless (1994), building on Owen (1988).
  - Imbens (2002) provides a recent survey of empirical likelihood that contrasts EL with GMM.
  - Objective functions other than $N^{-1}\sum_i \ln \pi_i$ may be used, such as $N^{-1}\sum_i \pi_i \ln \pi_i$.

3.5 Quantile Regression

- **Least absolute deviations (LAD) estimator** minimizes $\sum_{i=1}^{N} |y_i - \mathbf{x}_i'\boldsymbol{\beta}|$.
- In the iid case, with $\mathbf{x}_i'\boldsymbol{\beta} = \beta$, $\widehat{\beta}_{\text{LAD}}$ is the sample median. More generally estimate quantiles other than the median.
- The q^{th} **quantile regression estimator** $\widehat{\boldsymbol{\beta}}_q$ minimizes over $\boldsymbol{\beta}_q$

$$Q_N(\boldsymbol{\beta}_q) = \sum_{i:y_i \geq \mathbf{x}_i'\boldsymbol{\beta}}^{N} q|y_i - \mathbf{x}_i'\boldsymbol{\beta}_q| + \sum_{i:y_i < \mathbf{x}_i'\boldsymbol{\beta}}^{N} (1-q)|y_i - \mathbf{x}_i'\boldsymbol{\beta}_q|.$$

  where we use $\boldsymbol{\beta}_q$ rather than $\boldsymbol{\beta}$ to make clear that different choices of q estimate different values of $\boldsymbol{\beta}$ (LAD estimator is $q = 0.5$).
- **Implementation:**
  - $\widehat{\boldsymbol{\beta}}_q$ is obtained by linear programming (STATA does this)
  - Standard errors often computed by bootstrap.

Quantile Regression: Examples

- Koenker and Bassett (1982) used quantile regression to test **heteroskedasticity**: nonconstant $\widehat{\beta}_q$ as q varies $\Rightarrow$ heteroskedasticity.
- Powell (1984, 1986) used as way to get censored LAD and related estimators in **Tobit models** without assuming normal errors.
- Buchinsky (1994) used quantile regression **in its own right**, studying the response of earnings to education at different quantiles of income.
- Koenker and Hallock (2001) and Koenker (2005) provide summaries.
- Chernozhukov and Hansen (2005) propose an **IV estimator**.
- Angrist et al. (2006) provide interpretation of quantile regression when the quantile function is **misspecified** (i.e. nonlinear in **x)**.

3.6 Semiparametric Regression

- Consider model $y_i = m(\mathbf{x}_i) + u_i$ where $m(\cdot)$ is **unspecified**.
- **Nonparametric regression** obtains $\widehat{m}(\mathbf{x})$ at different values of $\mathbf{x}$.
- Then
  - There are many methods including kernel regression and lowess
  - Because local average is taken rate of convergence is less than $N^{0.5}$
  - For multivariate $\mathbf{x}_i$ nonparametric regression is very noisy.
- **Semiparametric models** impose some structure on $m(\mathbf{x})$.
- Then
  - Part parametric and part nonparametric
  - Ideally find $N^{0.5}$ estimate for the parametric part
  - Ideally no efficiency loss compared to if nonparametric part was specified
  - Not all parameters may be identified (e.g. just up to scale).

Semiparametric Regression: Examples

- **Partial linear model**
  - $y_i = \mathbf{x}_i'\boldsymbol{\beta} + g(z_i) + u_i$ where $g(\cdot)$ is unspecified
  - Estimators include Robinson (1988) differencing estimator
  - Example is sample selection where $g(\cdot)$ is multiple of inverse Mills ratio.
- **Single-index model**
  - $y_i = g(\mathbf{x}_i'\boldsymbol{\beta}) + u_i$ where $g(\cdot)$ is unspecified
  - Estimators include Stoker (1986) average derivative estimator and Ichimura (1993) weighted semiparametric least squares estimator
  - Example is binary choice with $\Pr[y_i = 1] = g(\mathbf{x}_i'\boldsymbol{\beta})$.
- **Many other examples**, especially for microeconometrics in limited dependent variable models
  - Manski (1975) proposed early example
  - Pagan and Ullah (1999) provide survey.

4. CAUSATION

1. Potential Outcomes Model
2. Differences in Differences
3. Regression Discontinuity
4. Instrumental Variables
5. Panel Data
6. Structural Models

- Angrist and Krueger (1999) survey many methods.

4.1 Potential Outcomes Model

- Focus on causal effect of binary variable d called a **treatment indicator**.
- The **outcome** y is a continuous variable that takes value

$$y_i = \begin{cases} y_i(1) & \text{if treated } (d_i = 1) \\ y_i(0) & \text{if control } (d_i = 0) \end{cases}$$

- The problem is that we observe only one of $y_i(0)$ and $y_i(1)$.
  i.e. for observed y_i we are missing data on the **counterfactual**.
- **Pure randomization** of treatment permits computation of the average treatment effect.
  $(\bar{y}_1 - \bar{y}_0)$ provides as estimate of $\mathrm{E}[y(1)] - \mathrm{E}[y(0)]$.

- The challenge is to extend this to cases where **individuals choose treatment**.
- Do this by assuming that **treatment assignment is random, once one controls using regressors**.
- **Formally** it is assumed (Rubin (1978)) that

$$(y(0), y(1)) \perp d \mid \mathbf{x}.$$

The assumption is given several names, including conditional independence, unconfoundedness, ignorability, and selection on observables only.

Potential Outcomes Model: Propensity Score matching

- Suppose the treatment effect is **constant across individuals**. Then a **control function approach** estimates treatment effect by $\widehat{\tau}$ from OLS of
$$y_i = \alpha + \mathbf{x}_i'\boldsymbol{\beta} + \tau d_i + u_i.$$
- If instead the treatment effect differ across individuals, **matching methods** compare $y_i(0)$ and $y_j(1)$ for similar individuals.
  - Match on $\mathbf{x}$ is obvious, but has problems for high-dimension $\mathbf{x}$
  - Instead match on predicted **propensity score** $p(\mathbf{x}) = \Pr[d = 1|\mathbf{x}]$. Rosenbaum and Rubin (1983) show that $(y(0), y(1)) \perp d \mid \mathbf{x} \Rightarrow (y(0), y(1)) \perp d \mid p(\mathbf{x})$
  - Use a flexible model for the propensity score e.g. semiparametric binary choice.
  - Various matching methods are used - nearest neighbors, kernel, stratification, ...
  - Abadie and Imbens (2006) present results for statistical inference.
  - References include Heckman, Ichimura, and Todd (1997) and Dehejia and Wahba (1999).

4.2 Differences in Differences

- Suppose groups are defined so that one group receives treatment and the other group does not.
  - e.g. a policy is applied in one state but not another state
  - A simple group differences (**treatment-control comparison**) in means fails to control for state-specific effects.
- Now suppose people in one group move over time from no treatment to treatment.
  - e.g. a policy change occurs over time in one state
  - A simple time differences (**before-after comparison**) in means fails to control for time-specific effects.
- Now suppose initially no group receives treatment but **over time some groups are treated while others are not**
  - This is setup for **differences-in-differences (DID)**
  - Use $\widehat{\text{ATE}} = (\Delta y \text{ for treated}) - (\Delta y \text{ for not treated})$
  - Assumes $y_{it} = \phi_i + \delta_t + \alpha d_{it} + \varepsilon_{it}$, $t = 0, 1$, $d_{it} = 1$ if treated
  - Ashenfelter (1978) early example
  - Imbens and Athey (2006) consider nonlinear (DID) models.

4.3 Instrumental Variables

- Now allow for **treatment selection on unobservables**.
- IV provides a general solution, provided there is an **instrument** that is correlated with being treated but does not directly cause y.
  - There are many creative examples proposed.
  - The interest in IV methods has been reduced given the weak instruments problem.
- The treatment literature emphasizes **binary treatment**, in which case the variable being instrumented is binary.
  - Then the IV estimator can be interpreted as providing measuring a **local average treatment effect (LATE)** that depends on the instrument chosen and its particular values. Imbens and Angrist (1994).
  - A more general treatment effect is the **marginal treatment effect (MTE)** that includes LATE, ATE and ATET as special cases.

4.4 Panel Data

- **Panel data** permit identification despite selection on unobservables, provided the **unobservables are time-invariant**.
- Consider the linear model $y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}$, where α_i and ε_{it} are unobserved.
  [For binary treatment d_{it} is a component of $\mathbf{x}_{it}$].
- If α_i is correlated with $\mathbf{x}_{it}$ (and ε_{it} is uncorrelated with $\mathbf{x}_{it}$) then
  - OLS of y_{it} on $\mathbf{x}_{it}$ is inconsistent
  - OLS of Δy_{it} on $\Delta\mathbf{x}_{it}$ is consistent (**first-differences estimator**)
  - OLS of $(y_{it} - \bar{y}_i)$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$ is consistent (**fixed effects estimator**)
- **Key assumption** is that only the time-invariant component of the unobservable is correlated with regressors such as the treatment indicator.

Panel Data: Extensions

- Note that random effects estimators will be inconsistent.
  For this reason microeconometricians shy away from random effects models.
- Microeconometrics focuses on **extending fixed effects models to a wider range of models with short panels**
  - Arellano-Bond (1991) estimator for linear models with lagged dependent variables
  - Logit model
  - Logit model with lagged dependent variables
  - Models such as Poisson model with multiplicative unobservable:
    $\mathrm{E}[y_{it}] = \alpha_i \exp(\mathbf{x}_{it}'\boldsymbol{\beta})$.
  - Current literature on biased estimators for fixed T but small bias.
  - Cameron and Trivedi (2005, chapters 22-23) has survey.

4.5 Regression Discontinuity

- Suppose **treatment occurs** when a variable s **crosses a threshold** $\bar{s}$. So $d = 1(s > \bar{s})$.
- **Complication** is to suppose that outcome y also depends on s.

- Then can compare y for those with s just less $<$ than y to those with with s just less $>$ than y.
- Simplest approach is to assume s has a linear effect. Use $\widehat{\alpha}_{\text{OLS}}$ in

$$y_i = \beta + \alpha d_i + \gamma s_i + u_i.$$

- More flexible is to use $y_i = \beta + \alpha d_i + \gamma k(s_i) + u_i$, where $k(\cdot)$ is not specified and nonparametric methods are used.
- And adapt to fuzzy design where the treatment threshold is not exact.
- Hahn, Todd and Van der Klaauw (2001) provide theory. Ludwig and Miller (2006) have recent application.

4.6 Structural Models

- Classic way to secure identification was linear simultaneous equations model.
- This has fallen by the wayside.
  First, IV allows one to just focus on the equation of interest.
  Second, other methods developed to measure causative effects that require weaker assumptions.
- Main area with structural modelling is industrial organization.
  See Reiss and Wolak (2005).

5. DATA ISSUES

1. Sampling Schemes
2. Measurement Error
3. Multiple Imputation for Missing Data

- **Endogenous stratified sampling**
  - Leads to inconsistent parameter estimates
  - Can use weighted ML (Manski and Lerman (1977)), GMM methods (Imbens(1992)), inverse-probability weighted estimators (Wooldridge (2002))
  - Imbens and Lancaster (1996) give general treatment in likelihood framework
  - Sample selection is also a leading example (Heckman (1979))

Sampling Schemes: Exogenous Stratified Sampling

- **Exogenous stratified sampling**
  - Parameter estimates remain consistent
  - For OLS (and analogously for other estimates)
  - No need to use sample weights if maintain that $\mathrm{E}[y_i|\mathbf{x}_i] = \mathbf{x}_i'\boldsymbol{\beta}$
  - Should use sample weights if do not assume $\mathrm{E}[y_i|\mathbf{x}_i] = \mathbf{x}_i'\boldsymbol{\beta}$ but want to recover census coefficients (DuMouchel and Duncan, 1983).
  - Wooldridge (2001) gives a general treatment of weighted m-estimation.

- **Clustered Sampling**
  - Survey methods often induce dependence for subgroups of observations
  - e.g. several households on the same block may be interviewed.
  - Standard procedure is to use cluster-robust standard errors.
  - Could use sample design information to improve efficiency of estimation, but this is rarely done.
  - Many other social science disciplines use hierarchical linear models or multilevel models. These are not used in microeconometrics.
  - If errors are correlated with regressors then use cluster fixed effects estimators

5.2 Measurement Error

- For **linear regression**
  - Focus is on classical measurement error in regressor
  - $\text{plim}\,\widehat{\beta}_{\text{OLS}} = \lambda\beta$ where λ is the reliability ratio of x as a measure of x^*
  - Angrist and Krueger (1999, p.1346) and Bound, Brown, and Mathiewetz (2001, pp.3749-3830) summarize many validation studies for labor-related data. Measurement error is large enough to matter.
  - $\boldsymbol{\beta}$ can be identified by IV methods, replicated data or validation sample data, additional distributional assumptions. And bounds on $\boldsymbol{\beta}$ can be obtained by reverse regression. Wansbeek and Meijer (2000) review many identification methods.
- For **nonlinear regression**
  - No clear theory, just special results
  - Surveys by Carroll, Ruppert and Stefanski (1995) and Hausman (2001).

Measurement Error: Nonlinear Regression

- Nonlinear regression with **additive error**
  - IV methods do not easily extend (Y. Amemiya (1985) for polynomial regression)
  - Can use repeated measures (Hausman, Newey and Powell (1995), Li (2002), and Schennach (2004).
  - Schennach (2006) proposes an instrumental variables estimator.
- Nonlinear models with **nonadditive error** e.g. discrete outcome, counts
  - Measurement error in dependent variable also cause problems
  - Hausman, Abrevaya and Scott-Morton (1998) consider mismeasurement in the dependent variable in binary outcome models.
  - Guo and Li (2002) consider mismeasurement in a regressor in a Poisson model.
  - These papers take a parametric approach with strong assumptions.
- Some work relaxes assumption of iid measurement error in regressor
  - Kim and Solon (2005) consider standard linear panel estimators.
  - Mahajan (2006) considers binary regressor in nonparametric models.

5.3 Multiple Imputation for Missing Data

- Let $\mathbf{W} = (\mathbf{W}_{obs}, \mathbf{W}_{miss})$ denote data partitioned into observed and missing observations.
- Assume $\mathbf{W}$ has density $f(\mathbf{W}|\boldsymbol{\theta})$. Then given imputed value $\mathbf{W}_{miss}^{(l)}$ we can obtain the MLE based on $f(\mathbf{W}_{obs}, \mathbf{W}_{miss}^{(l)}|\boldsymbol{\theta})$.
- Do multiple imputations to account for imprecision in imputing $\mathbf{W}_{miss}^{(l)}$.
- Given m different imputed values for $\mathbf{W}_{miss}$ get m estimates $\widehat{\boldsymbol{\theta}}_r$, $r = 1, ..., m$ with associated variance matrices $\widehat{\mathbf{V}}_r = \widehat{\mathbf{V}}[\widehat{\boldsymbol{\theta}}_r]$. Then

$$\begin{aligned} \widehat{\boldsymbol{\theta}} &= \frac{1}{m}\sum_{r=1}^{m} \widehat{\boldsymbol{\theta}}_r \\ \widehat{\mathrm{V}}[\widehat{\boldsymbol{\theta}}] &= \frac{1}{m}\sum_{r=1}^{m} \widehat{\mathbf{V}}_r + \frac{1+\frac{1}{m}}{m-1}\sum_{r=1}^{m}(\widehat{\boldsymbol{\theta}}_r - \widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}}_r - \widehat{\boldsymbol{\theta}})'. \end{aligned}$$

- References include Rubin (1976, 1987).

Conclusion

- Microeconometricians are very ambitious in their desire to obtain marginal effects that
  - can be given a causative interpretation
  - permit individual heterogeneity
  - are obtained under minimal assumptions
  - with statistical inference also under minimal assumptions.
- This has led to a literature and toolkit that goes way beyond extending linear structural equation models to a nonlinear setting.

Some References

Abadie, A., and G.W. Imbens (2006a), "Large Sample Properties of Matching Estimators for Average Treatment Effects," Econometrica, 235-267.

Abadie, A., and G.W. Imbens (2006b), "On the Failure of the Bootstrap for Matching Estimators," unpublished manuscript.

Abrevaya, J., and J. Huang (2005), "On the Bootstrap of the Maximum Score Estimator," Econometrica, 1175-1204.

Amemiya, T. (1985), Advanced Econometrics, Cambridge, MA, Harvard University Press.

Amemiya, Y. (1985), "Instrumental Variable Estimator for the Nonlinear Error in Variables Model," Journal of Econometrics, 28, 273-289.

Andrews, D.W.K., and M. Buchinsky (2000), "A Three-Step Method for choosing the Number of Bootstrap Replications," Econometrica, 68, 23-51.

Andrews, D.W.K., and J. Stock (2005), "Inference with Weak instruments," invited paper, 2005 World Congress of the Econometric Society.

Angrist, J., V. Chernozhukov, and I. Ferandez-Val (2006), "Quantile Regression Under Misspecification, with an Application to the U.S. Wage Structure," Econometrica, 539-563.

Angrist, J.D., and A.B. Krueger (1999), "Empirical Strategies in Labor Economics," in Handbook of Labor Economics, O.C. Ashenfelter and D.E. Card (Eds.), Volume 3A, 1277-1397, Amsterdam, North-Holland.

Angrist, J., and V. Lavy (2002), "The Effect of High School Matriculation Awards: Evidence from Randomized Trials," NBER Working Paper 9389.

Arellano, M. (1987), "Computing Robust Standard Errors for Within-Group Estimators," Oxford Bulletin of Economics and Statistics, 49, 431-434.

Arellano, M., and S. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," Review of Economic Studies, 58, 277-298.

Ashenfelter, 0. (1978), "Estimating the Effect of Training Programs on Earnings," Review of Economics and Statistics, 60, 47-57.

Athey, S. and G.W. Imbens (2006), "Identification and Inference in Nonlinear Difference-in-Difference Models," Econometrica, 431-497.

Beran, R. (1982), "Estimating Sampling Distributions: The Bootstrap and Competitors," Annals of Statistics, 10. 212-225.

Bertrand, M., E. Duflo and S. Mullainathan (2004), "How Much Should We Trust Differences-in-Differences Estimates?" Quarterly Journal of Economics, 119, 249-275

Bound, J., C. Brown, and N. Mathiowetz (2001), "Measurement Error in Survey Data" in Handbook of Econometrics, J.J. Heckman and E.E. Leamer (Eds.), Volume 5, Amsterdam, North-Holland.

Bound, J., D.A. Jaeger, and R.M. Baker (1995), "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak," Journal of the American Statistical Association, 90, 443-450.

Buchinsky, M. (1994), "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression," Econometrica, 62, 405-458.

Cameron, A.C., Gelbach, J., and D.L. Miller (2006a), "Bootstrap-Based Improvements for Inference with Clustered Errors," Working Paper No. 06-??, Department of Economics, University of California - Davis.

Cameron, A.C., Gelbach, J., and D.L. Miller (2006b), "Robust Inference with Multi-way Clustering," Working Paper No. 06-??, Department of Economics, University of California - Davis.

Cameron, A.C., and P.K. Trivedi (2005), Microeconometrics: Methods and Applications, Cambridge, Cambridge Universtiy Press.

Carroll, R.J., D. Ruppert, and L.A. Stefanski (1995), Measurement Error in Nonlinear Models, London, Chapman and Hall.

Chernozhukov, V., and C. Hansen (2005), "An IV Model of Quantile Treatment Effects," Econometrica, 245-262.

Chib, S. (1992), "Bayes Regression for the Tobit Censored Regression Model," Journal of Econometrics, 58, 79-99.

Chib, S. (2001), "Markov Chain Monte Carlo Methods: Computation and Inference," in J.J. Heckman and E.E. Leamer (Eds.), Handbook of Econometrics, Volume 5, 3570-3649, Amsterdam, North-Holland.

Dehejia, R.H., and S. Wahba (1999), "Reevaluating the Evaluation of Training Programs," Journal of the American Statistical Association, 94, 1053-1062.

Donald, S. G., and Lang, K. (2004), "Inference with Differences in Differences and Other Panel Data", unpublished manuscript.

DuMouchel, W.K., and G.J. Duncan (1983), "Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples," Journal of the American Statistical Association, 78, 535-43.

Efron, B. (1979), "Bootstrapping Methods: Another Look at the Jackknife," Annals of Statistics, 7, 1-26.

Geweke, J. (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," Econometrica, 57, 1317-1339.

Gouriéroux, C., and A. Monfort (1996), Simulation Based Econometrics Methods, New York, Oxford University Press.

Hahn, J., P. Todd and W. Van der Klaauw (2001), "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," Econometrica, 69, 201-209.

Hall, P. (1992), The Bootstrap and Edgeworth Expansion, New York: Springer-Verlag.

Hall, P., and J.L. Horowitz (1996), "Bootstrap Critical Values for Tests Based on Generalized Method of Moments Estimators," Econometrica, 64, 891-916.

Hansen, L.P. (1982), "Large Sample Properties of Generalized Methods of Moments Estimators," Econometrica, 1029-1054.

Hausman, J.A. (2001), "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left," Journal of Economic Perspectives, 15, 57-68.

Hausman, J.A., J. Abrevaya, and F.M. Scott-Morton (1998), "Misclassification of the Dependent Variable in a Discrete Response Setting," Journal of Econometrics, 87, 239-269.

Hausman, J.A., W.K. Newey, and J.L. Powell (1995), "Nonlinear Errors in Variables: Estimation of Some Engel Curves," Journal of Econometrics, 65, 205-233.

Heckman, J.J. (1978), "Dummy Endogenous Variables in a Simultaneous Equations System," Econometrica, 46, 931-960.

Heckman, J.J. (1979), "Sample Selection as a Specification Error," Econometrica, 47, 153-161.

Heckman, J.J., H. Ichimura, and P. Todd (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," Review of Economic Studies, 64, 605-654.

Horowitz, J. L. (1997), "Bootstrap Methods in Econometrics: Theory and Numerical Performance," in Kreps and Wallis eds., Advances in Econometrics, Vol. 7.

Huber, P.J. (1967), "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions," in Proceedings of the Fifth Berkeley Symposium, J. Neyman (Ed.), 1, 221-233, Berkeley, CA, University of California Press.

Imbens, G.W. (1992), "An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling," Econometrica, 60, 1187-1214.

Imbens, G.W. (2002), "Generalized Method of Moments and Empirical Likelihood," Journal of Business and Economic Statistics, 20, 493-506.

Imbens, G.W., and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effect," Econometrica, 62, 467-475.

Imbens, G.W., and T. Lancaster (1996), "Efficient Estimation and Stratified Sampling," Journal of Econometrics, 74, 289-318.

Kim, B, and G. Solon (2005), “Implications of Mean-reverting Measurement Error for Longitudinal Studies of Wages and Employment,” Review of Economics and Statistics, 87, 193-196.

Koenker, R. (2005), “Quantile Regression,” Econometric Society Monograph, Cambridge, UK, Cambridge University Press.

Koenker, R., and G. Bassett (1978), “Regression Quantiles,” Econometrica, 46, 33-50.

Koenker, R., and K.F. Hallock (2001), “Quantile Regression,” Journal of Economic Perspectives, 15, 143-156.

Koop, G. (2003), Bayesian Econometrics, New York, Wiley.

Lancaster, T. (2004), An Introduction to Modern Bayesian Econometrics, Oxford, Blackwell.

Li, T. (2002), “Robust and Consistent Estimation of Non-linear Errors-in-Variables Models,” Journal of Econometrics, 110, 1-26.

Liang, K.-Y., and S.L. Zeger (1986), “Longitudinal Data Analysis Using Generalized Linear Models,” Biometrika, 73, 13-22.

Ludwig, J., and D.L. Miller (2006),“Does Head Start Improve Children’s Life Chances?: Evidence from a Regression Discontinuity Design,” Quarterly Journal of Economics, forthcoming.

Mahajan, A. (2006), "Identification and Estimation of regression Models with Misclassification," Economtrica, 631-665.

Manski, C.F. (1975), "The Maximum Score Estimator of the Stochastic Utility Model of Choice," Journal of Econometrics, 3, 205-228.

Manski, C.F., and S.R. Lerman (1977), "The Estimation of Choice Probabilities from Choice-Based Samples," Econometrica, 45, 1977-1988.

McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," Econometrica, 57, 995-1026.

Moulton, B.R. (1986), "Random Group Effects and the Precision of Regression Estimates," Journal of Econometrics, 32, 385-397.

Moulton, B.R. (1990), "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," Review of Economics and Statistics, 72, 334-38.

Moreira, M.J., J.R. Porter, and G.A. Suarez (2004), "Bootstrap and Higher-Order Expansion Validity When Instruments May Be Weak," NBER Technical Working Paper 302.

Nelson, C.R., and R. Startz (1990), "The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument Is a Poor One," Journal of Business, 63, S125-140.

Newey, W.K., and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," in Handbook of Econometrics, R.F. Engle and D. McFadden (Eds.), Volume 4, 2111-2245, Amsterdam, North-Holland.

Newey, W.K., and R.J. Smith (2004), "Higher Order Properties of GMM and Genralized Empirical Likelihood Estimators," Econometrica, 219-255.

Owen, A.B. (1988), "Empirical Likelihood Ratios Confidence Intervals for a Single Functional," Biometrika, 75, 237-249.

Pagan, A.R., and A. Ullah (1999), Nonparametric Econometrics, Cambridge, UK, Cambridge University Press.

Pakes, A.S., and D. Pollard (1989), "Simulation and the Asymptotics of Optimization Estimators," Econometrica, 57, 1027-1057.

Politis, D.N., and J.P. Romano (1994), "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions," Annals of Statistics, 22, 2031-2050.

Powell, J.L. (1984), "Least Squares Absolute Deviations Estimation for the Censored Regression Model," Journal of Econometrics, 25, 303-325.

Powell, J.L. (1986), "Censored Regression Quantiles," Journal of Econometrics, 32, 143-155.

Qin, J., and J. Lawless (1994), "Empirical Likelihood and General Estimating Equations," Annals of Statistics, 22, 300-325.

Reiss, P.C., and F.A. Wolak (2005), Structural Econometric Modeling: Rationales and Examples from Industrial Organization, Handbook of Econometrics: volume 6, forthcoming.

Rosenbaum, P. and D.B. Rubin (1983), "The Central Role of Propensity Score in Observational Studies for Causal Effects," Biometrika, 70, 41-55.

Rubin, D.B. (1976), "Inference and Missing Data," Biometrika, 63, 581-592.

Rubin, D.B. (1978), "Bayesian Inference for Causal Effects," Annals of Statistics, 6, 34-58.

Rubin, D.B. (1987), Multiple Imputation for Nonresponse in Surveys, New York, John Wiley.

Staiger, D., and J. Stock (1997), "Instrumental Variables Regression with Weak Instruments," Econometrica, 65, 557-586.

Train, K.E. (2003), Discrete Choice Methods with Simulation, Cambridge, UK, Cambridge University Press.

Wansbeek, T., and E. Meijer (2000), Measurement Error and Latent Variables in Econometrics, Amsterdam, North-Holland.

White, H. (1980a), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-838.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-25.

White, H. (1984), Asymptotic Theory for Econometricians, San Diego, Academic Press.

Wooldridge, J.M. (2001), "Asymptotic Properties of Weighted M-Estimators for Standard Stratified Samples," Econometric Theory, 17, 451-470.

Wooldridge, J.M. (2003), "Cluster-Sample Methods in Applied Econometrics," American Economic Review, 93, 133-138.