Semiparametric and Nonparametric Methods in Political Science

Lecture 1: Semiparametric Estimation

Michael Peress, University of Rochester and Yale University

Overview of Semi and Nonparametric Models

- Parametric Model: statistical model characterized by finite dimensional unknown parameter
  - $y_n \sim N(\mu, \sigma^2)$
  - $y_n \sim N(\beta' x_n, \sigma^2)$ (normal-linear model)

- Nonparametric Model: statistical model characterized by infinite dimensional unknown parameter
  - $y_n \sim f$ where f is unknown (density estimation)
  - $y_n = g(x_n) + \varepsilon_n,\ E[\varepsilon_n \mid x_n] = 0$ (nonparametric regression)

Overview of Semi and Nonparametric Models

- Semiparametric Model: statistical model characterized by finite dimensional parameter of interest and infinite dimensional "nuisance" parameter
    - $y_n = \beta' x_n + \varepsilon_n,\ \varepsilon_n \mid x_n \sim F(\varepsilon \mid x)$ with $E[\varepsilon_n \mid x_n] = 0$ (semiparametric linear model)
        - β is "parameter of interest"
        - F is a nuisance parameter
    - $y_t = \beta' x_t + \varepsilon_t,\ (x_t, \varepsilon_t)$ is stationary and ergodic and $E[\varepsilon_t \mid x_t] = 0$
        - β is "parameter of interest"
        - Stochastic process characterizing (x_t, ε_t) is a nuisance parameter

Overview of Semi and Nonparametric Models

- More Semiparametric/Nonparametric Models:
  - $y_n = g(\beta' x_n) + \varepsilon_n,\ \varepsilon_n \mid x_n \sim F(\varepsilon \mid x)$ with $E[\varepsilon_n \mid x_n] = 0$ (linear index model)
  - $y_n = g(x_n) + \beta' z_n + \varepsilon_n,\ \varepsilon_n \mid x_n \sim F(\varepsilon \mid x)$ with $E[\varepsilon_n \mid x_n] = 0$ (partially linear model)
  - $\Pr(y_n = 1) = G(\beta' x_n)$ (semiparametric binary choice)

Overview of Semi and Nonparametric Models

- Parametric Models:
  - MLE is efficient if parametric model is correct
  - MLE is often inconsistent if parametric model is incorrect
  - $\sqrt{N}$ -convergence rate

- Nonparametric Models:
  - More generality, but…
    - Theory more difficult
    - Implementation difficult
    - Slower convergence (slower than parametric rate of $\sqrt{N}$)
    - Efficiency loss (relative to MLE if parametric model is corr.)

Overview of Semi and Nonparametric Models

- Semiparametric Models:
  - More generality, and…
    - Often, $\sqrt{N}$ -convergence for parameter of interest
    - Often, easy to implement
    - Often, little efficiency loss
  - Theory can be very hard, but some important cases are sufficiently worked out so that we don't have to worry about it

- Lecture 1 will focus on easy but powerful semiparametric estimators
- Lecture 2 will focus on basics of nonparametric estimation
- Lecture 3 will focus on applications of nonparametric estimators and more advanced semiparametric estimators

Overview of Semi and Nonparametric Models

- Examples of "Easy" Semiparametric Estimators:
  - OLS w/ robust se's - semiparametric because OLS is consistent even if error terms are non-normal and heteroskedastic
  - Poisson regression w/ robust se's - semiparametric because estimator is consistent when dependent variable is not Poisson distributed
  - Linear-nonlinear models w/ Newey-West se's – semiparametric because OLS/MLE are consistent even when dependent variable exhibits time series dependence
  - Short panels with clustered standard errors – semiparametric because OLS/MLE are consistent even when dependent variable exhibits group correlation/time series dependence

Overview of Semi and Nonparametric Models

- Why Use These Semiparametric Estimators?
  - Easy to apply:
    - Some parametric alternatives are VERY computationally intensive
  - Skip the specification step (which is sometimes near impossible)
    - Modeling heteroskedasticity
    - Selecting ARMA structure (w/ time series and panel data)
    - Selecting between negative binomial, zero-inflated, zero-truncated, etc., in count models

Overview of Semi and Nonparametric Models

- Drawbacks:
    - Efficiency loss relative to parametric model
- However:
    - Parametric model may be wrong!
    - Usually, semiparametric estimators achieve semiparametric efficiency bounds (they are efficient under maintained assumptions)
    - Often, not much efficiency loss
    - Often, these semiparametric estimators give robustness practically “for free” since we don’t have to estimate the nuisance parameters

Heteroskedasticity in the Linear Model

- Parametric Linear Model:
  1. $y_n = \beta_0' x_n + \varepsilon_n$ (linearity)
  2. (x_n, ε_n) are independent (independence)
  3. $E[x_n x_n']$ has full rank (identification)
  4. $E[\varepsilon_n \mid x_n] = 0$
  5. $\varepsilon_n \sim N(0, \sigma^2)$ (homoskedasticity and normality)
- Under 1-5, OLS is MLE; OLS is unbiased, normally distributed, consistent, and asymptotically normal; the information equality holds; and OLS is efficient

Heteroskedasticity in the Linear Model

- Semiparametric Linear Model:
  1. $y_n = \beta_0 ' x_n + \varepsilon_n$ (linearity)
  2. (x_n, ε_n) are independent (independence)
  3. $E[x_n x_n ']$ has full rank (identification)
  4. $E[\varepsilon_n | x_n] = 0$
  5. ~~$\varepsilon_n \sim N(0, \sigma^2)$ (homoskedasticity and normality)~~
- Consider OLS as semiparametric estimator
- Under 1-4, OLS is unbiased, ~~normally distributed~~, consistent, and asymptotically normal; ~~the information equality holds~~; and ~~OLS is efficient~~

Heteroskedasticity in the Linear Model

- Properties of OLS as semiparametric estimator:

$$\hat{\beta} = \left[\tfrac{1}{N} \sum_{n=1}^{N} x_n x_n' \right]^{-1} \left[\tfrac{1}{N} \sum_{n=1}^{N} x_n \underbrace{y_n}_{=(\beta_0' x_n + \varepsilon_n)} \right] = \beta_0 + \left[\tfrac{1}{N} \sum_{n=1}^{N} x_n x_n' \right]^{-1} \left[\tfrac{1}{N} \sum_{n=1}^{N} x_n \varepsilon_n \right]$$

- By law of iterated expectations,

$$E[x_n \varepsilon_n] = E[x_n \underbrace{E[\varepsilon_n \mid x_n]}_{=0}] = 0$$

- OLS is unbiased:

$$E[\hat{\beta} \mid x] = \beta_0 + \left[\tfrac{1}{N} \sum_{n=1}^{N} x_n x_n' \right]^{-1} \left[\tfrac{1}{N} \sum_{n=1}^{N} \underbrace{E[x_n \varepsilon_n]}_{=0} \right] = \beta_0$$

Heteroskedasticity in the Linear Model

- OLS is consistent:

$$\tfrac{1}{N}\sum_{n=1}^{N} x_n \varepsilon_n \xrightarrow{prob.} \underbrace{E[x_n \varepsilon_n]}_{=0} = 0$$

$$\hat{\beta} = \beta_0 + \left[\tfrac{1}{N}\sum_{n=1}^{N} x_n x_n{}'\right]^{-1} \underbrace{\left[\tfrac{1}{N}\sum_{n=1}^{N} x_n \varepsilon_n\right]}_{\xrightarrow{prob.} 0} \xrightarrow{prob.} \beta_0 + \left[\tfrac{1}{N}\sum_{n=1}^{N} x_n x_n{}'\right]^{-1} 0 = \beta_0$$

- Normality/homoskedasticity not needed for these results

Heteroskedasticity in the Linear Model

- OLS is asymptotically normal:

$$\sqrt{N}(\hat{\beta}-\beta_0) = \left[\tfrac{1}{N}\sum_{n=1}^{N} x_n x_n'\right]^{-1} \left[\tfrac{1}{\sqrt{N}}\sum_{n=1}^{N} x_n \varepsilon_n\right]$$

$$\downarrow \; LLN \qquad\qquad \downarrow \; CLT$$

$$E[x_n x_n']^{-1} \qquad N(0, Var(x_n \varepsilon_n))$$

$$\sqrt{N}(\hat{\beta}-\beta_0) \xrightarrow{dist.} N(0, \underbrace{E[x_n x_n']^{-1}}_{bread} \underbrace{Var(x_n \varepsilon_n)}_{meat} \underbrace{E[x_n x_n']^{-1}}_{bread})$$

- Estimate asymptotic distribution using:

$$E[x_n x_n']^{-1} \approx \tfrac{1}{N}\sum_{n=1}^{N} x_n x_n'$$

$$Var(x_n \varepsilon_n) = E[\varepsilon_n{}^2 x_n x_n'] \approx \tfrac{1}{N}\sum_{n=1}^{N} \varepsilon_n{}^2 x_n x_n'$$

Heteroskedasticity in the Linear Model

- Implementation:
  - In stata, “regress y x1 x2, robust”
  - In r, use *sandwich* package:
    - “lm1 <- lm(Y ~ X1 + X2)”
    - “sw1 <- sandwich(lm1)”

Heteroskedasticity in the Linear Model

- Overview:
  - Apply OLS when homoskedasticity/normality do not hold
  - Benefit: robustness
  - Drawback: less efficiency

Heteroskedasticity in the Linear Model

- Example:
  - OLS is MLE when errors are normal and homoskedastic
  - LAD is MLE when errors are double exponential and homoskedastic
  - OLS will be more efficient than LAD when errors are normal and homoskedastic, robust se's will be correct for both estimators
  - LAD will be more efficient than OLS when errors are double exponential and homoskedastic, robust se's will be correct for both estimators

Heteroskedasticity in the Linear Model

- Example Continued:
  - Generate 1000 Monte Carlo data sets with N=500, X1~N(0,1), X2~N(0,1), Beta=(-.5,1.5,-1.0), and errors either N(0,1) or DExp(0,1)

	DGP = Normal-Linear			**DGP = DExp-Linear**		
	Beta1	*Beta2*	*Beta3*	*Beta1*	*Beta2*	*Beta3*
Rel. Eff. OLS/LAD	0.81	0.80	0.82	1.31	1.37	1.28
OLS Overconfidence	1.06	1.04	1.05	0.97	1.02	1.02
LAD Overconfidence	1.04	1.04	1.02	0.98	0.98	1.04
OLS se / LAD se	0.89	0.89	0.89	1.15	1.15	1.15

Nonproportional Dispersion in Count Models

- Parametric Poisson Model:
  1. $y_n \sim Poisson(\lambda_n)$, $\lambda_n = e^{\beta_0' x_n}$
  2. (y_n, x_n) are iid
- Notice that $E[y_n \mid x_n] = Var(y_n \mid x_n) = \lambda_n = e^{\beta_0' x_n}$
- We can derive the log-likelihood function:

$$l(y,x;\beta) = \sum_{n=1}^{N} y_n \beta' x_n - e^{\beta' x_n} - \log y_n!$$

Nonproportional Dispersion in Count Models

- Semiparametric Poisson Model:
  1. ~~$y_n \sim Poisson(\lambda_n),\ \lambda_n = e^{\beta_0' x_n}$~~ $E[y_n \mid x_n] = e^{\beta_0' x_n}$
  2. (y_n, x_n) are iid
- Semiparametric estimator defined by:

$$\hat{\beta} = \arg\max_{\beta} \tfrac{1}{N} \sum_{n=1}^{N} y_n \beta' x_n - e^{\beta' x_n} - \log y_n!$$

Nonproportional Dispersion in Count Models

- Consistency of semiparametric Poisson regression:

$$\hat{\beta} = \arg\max_{\beta} \tfrac{1}{N}\sum_{n=1}^{N} y_n \beta' x_n - e^{\beta' x_n} - \log y_n!$$

- First order condition:

$$0 = \tfrac{1}{N}\sum_{n=1}^{N} x_{n,k}(y_n - e^{\hat{\beta}' x_n})$$

- In large samples:

$$0 = E[x_{n,k}(y_n - e^{\hat{\beta}' x_n})] = E[x_{n,k}E[(y_n - e^{\hat{\beta}' x_n}) \mid x_n]] = E[x_{n,k}(e^{\beta_0' x_n} - e^{\hat{\beta}' x_n}) \mid x_n]]$$

- Hence, $\hat{\beta} \xrightarrow{prob.} \beta_0$ as long as conditional mean is correctly specified (even if Poisson assumption does not hold)

Nonproportional Dispersion in Count Models

- What about standard errors?
- Define $\psi(y_n, x_n; \beta) = y_n \beta' x_n - e^{\beta' x_n} - \log y_n!$
- Taylor expansion argument:

$$\sqrt{N}(\hat{\beta} - \beta_0) \xrightarrow{dist.} N(0, Q^{-1} V Q^{-1})$$

where,

$$Q = E[\psi_{\beta\beta}(y_n, x_n; \beta_0)], \qquad V = Var(\psi_{\beta}(y_n, x_n; \beta_0))$$

- If MLE assumptions hold, $Q = -V$
- If not, must use “sandwich” estimator, i.e. robust se’s

Nonproportional Dispersion in Count Models

- Implementation:
  - In stata: “poisson y x1 x2, robust”
  - In r, use *sandwich* package:
    - “pm1 <- glm(Y ~ X1 + X2,family=”poisson”)”
    - “sw1 <- sandwich(pm1)”

Nonproportional Dispersion in Count Models

- What we get:
  - Robustness to overdispersion
  - Robustness to zero-inflation, zero-truncation, one inflation, etc.
  - All we need is correctly specified conditional mean
- What we don't get
  - Efficiency (MLE is more efficient if parametric model is correct)
  - Predicted values easily generated, but not predicted distribution (since distribution is not Poisson)
  - If we want predicted values, we can use procedures discussed in lecture 3

Nonproportional Dispersion in Count Models

- Same principal extends to negative binomial model
- Same principal extends to other exponential family models (i.e. consistency holds as long as conditional mean is correctly specified)
- “, robust” does not provide any benefit for logit, probit, ordered logit, multinomial logit, etc.:
  - These models are only correct if parametric model is correct
  - If parametric model is correct, se’s = robust se’s in large samples

Time Series Dependence in the Linear Model

- Semiparametric Time Series Linear Model:
  1. $y_t = \beta_0 ' x_t + \varepsilon_t$ (linearity)
  2. ~~(x_t, ε_t) are independent~~ (x_t, ε_t) are stationary and ergodic
  3. $E[x_t x_t ']$ has full rank (identification)
  4. $E[\varepsilon_t \mid x_t] = 0$
  5. ~~$\varepsilon_t \sim N(0, \sigma^2)$ (homoskedasticity and normality)~~
- Consider OLS as semiparametric estimator
- Under 1-4, OLS is unbiased, ~~normally distributed~~, consistent, and asymptotically normal; ~~the information equality holds~~; ~~OLS is efficient~~

Time Series Dependence in the Linear Model

- OLS estimator: $\hat{\beta} = \beta_0 + \left[\frac{1}{T} \sum_{t=1}^{T} x_t x_t' \right]^{-1} \left[\frac{1}{T} \sum_{t=1}^{T} x_t \varepsilon_t \right]$
- OLS is unbiased: $E[\hat{\beta} \mid x] = \beta_0 + \left[\frac{1}{T} \sum_{t=1}^{T} x_t x_t' \right]^{-1} \left[\frac{1}{T} \sum_{t=1}^{T} \underbrace{E[x_t \varepsilon_t]}_{=0} \right] = 0$
- Under stationarity and ergodicity, $\frac{1}{T} \sum_{t=1}^{T} x_t \varepsilon_t \xrightarrow{prob.} E[x_t \varepsilon_t] = 0$
- Hence, OLS is consistent

Time Series Dependence in the Linear Model

- Large sample distribution:

$$\sqrt{T}(\hat{\beta}-\beta_0) = \left[\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right]^{-1} \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t \varepsilon_t\right]$$

$$\downarrow \; LLN \qquad\qquad \downarrow \; CLT$$

$$\underbrace{E[x_t x_t']}_{=Q}{}^{-1} \qquad N\left(0, \underbrace{\lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t \varepsilon_t\right)}_{=V}\right)$$

$$\sqrt{T}(\hat{\beta}-\beta_0) \xrightarrow{dist.} N(Q^{-1}VQ^{-1})$$

Time Series Dependence in the Linear Model

- Tricky part is estimating $V = \lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t \varepsilon_t \right)$
- If $x_t \varepsilon_t$ are independent, then
  $$Var\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t \varepsilon_t \right) = \frac{1}{T} \sum_{t=1}^{T} E[\varepsilon_t^2 x_t x_t'] \approx \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 x_t x_t'$$
- If $x_t \varepsilon_t$ not independent, then covariance terms make it hard
  $$Var\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t \varepsilon_t \right) = \frac{1}{T} \sum_{t=1}^{T} Var(x_t \varepsilon_t) + \frac{2}{T} \sum_{s<t} Cov(x_t \varepsilon_t, x_s \varepsilon_s)$$
- Newey-West (and related procedures) provide a way to estimate
  $$\lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t \varepsilon_t \right)$$

Time Series Dependence in the Linear Model

- Define,

$$\hat{\gamma}(s) = \tfrac{1}{T-s-1} \sum_{t=s+1}^{T} \varepsilon_t \varepsilon_{t-s} x_t x_{t-s}'$$

and estimate,

$$\hat{V} = \hat{\gamma}(0) + \sum_{s=1}^{m_T} (1 - \tfrac{s}{m_T+1})[\hat{\gamma}(s) + \hat{\gamma}(s)']$$

- Select m_T such that $m_T \to \infty$ as $T \to \infty$
- Automatic procedures for choosing m_T efficiently are available
- Rule of thumb is $m_T = \left\lfloor 4\left(\tfrac{T}{100}\right)^{2/9} \right\rfloor$
- Newey-West is special case of spectral density approach to covariance matrix estimation (w/ a Bartlett Kernel)

Time Series Dependence in the Linear Model

- What we get:
  - Robustness to heteroskedasticity and autocorrelation
  - No need to select appropriate ARMA model (there is some bandwidth selection going on in the background, but this part is largely automated)
- What we lose:
  - Efficiency: If correct ARMA model is selected, then MLE will be more efficient

Time Series Dependence in Nonlinear Models

- Newey-West standard errors can be used to correct for time series dependence in nearly any nonlinear model
- For many nonlinear models, incorporating time series dependence is extremely difficult
- Parametric time-series versions of standard estimators cannot be estimated in most (or even all?) statistical packages
  - Time series logit, time series probit, time series count, etc.

Time Series Dependence in Nonlinear Models

- Binomial Probit with AR1 errors (parametric model)
  1. $y_t^* = \beta' x_t + \varepsilon_t,\ y_t = 1\{y_t^* \geq 0\}$ (probit model)
  2. $\varepsilon_t = \rho\varepsilon_{t-1} + u_t,\ u_t \sim N(0,1),\ u_t$ are iid (AR1 errors)
- MLE involves complicated T-dimensional integral
  - Define $A(y_1,...,y_T) = \{x : x_t \in [0,\infty)\ if\ y_t = 1,\ x_t \in (-\infty,0]\ if\ y_t = 0\}$
  - $\mu_t(x;\beta) = \beta' x_t,\ \Omega_{s,t}(\rho) = \dfrac{\rho^{|s-t|}}{1-\rho}$
  - $\Pr(y_1,...,y_T;\beta,\rho) = \int_{\vartheta \in A(y_1,...,y_T)} \phi(\vartheta;\mu(x;\beta),\Omega(\rho))d\vartheta$
- Really hard to compute! (stata/r don't do it right now)

Time Series Dependence in Nonlinear Models

- Alternative approach: semiparametric estimation
- Claim (Poirier and Ruud, 1986): Probit is still consistent when observations are dependent
- Why?
  - MLE is $\hat{\theta} = \arg\max_{\theta} \frac{1}{T} \sum_{t=1}^{T} \log f(y_t;\theta)$
  - MSE is consistent because $\frac{1}{T} \sum_{t=1}^{T} \log f(y_t;\theta) \to E[\log f(y_t;\theta)]$
  - Information inequality implies $E[\log f(y_t;\theta)]$ is minimized at θ_0, the "true" parameter value
  - Information inequality will continue to hold for all models that have the same marginals

Time Series Dependence in Nonlinear Models

- Hence, we can apply Newey-West standard errors to probit to obtain consistent estimates with corrects standard errors
- Same result holds for other models:
  - Parametric Poisson models w/ time series dependence are difficult to obtain
  - In Poisson case, using Newey West standard errors give estimator that is robust to over/under dispersion, zero-inflation, and time series dependence

Time Series Dependence in Nonlinear Models

- Implementation (linear model):
  - In stata, “newey y x1 x2, lag(#)”
  - In r, *sandwich* package:
    - “lm1 <- lm(Y ~ X1 + X2)”
    - “sw1 <- NeweyWest(lm1)”
- Implementation (nonlinear models):
  - In stata, using nwest package
  - In r, *sandwich* package:
    - “glm1 <- glm(Y ~ X1 + X2,family=”poisson”)”
    - “sw1 <- NeweyWest (glm1)”

Time/Group Dependence in Panel Data

- Semiparametric Linear Panel Data Model (short panels with many individuals):
    1. $y_{n,t} = \beta' x_{n,t} + \varepsilon_{n,t}$ (linear model)
    2. $E[\varepsilon_{n,t} \mid x_{n,t}] = 0$
    3. $E[x_{n,t} x_{n,t}']$ has full rank (identification)
    4. $(\varepsilon_{n,1}, \ldots, \varepsilon_{n,T})$ are independent over n (independence)
- OLS estimator:

$$\hat{\beta} = \beta_0 + \left[\tfrac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} x_{n,t} x_{n,t}' \right]^{-1} \left[\tfrac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} x_{n,t} \varepsilon_{n,t} \right]$$

$$= \beta_0 + \left[\tfrac{1}{N} \sum_{n=1}^{N} \underbrace{\left(\tfrac{1}{T} \sum_{t=1}^{T} x_{n,t} x_{n,t}' \right)}_{= z_n} \right]^{-1} \left[\tfrac{1}{N} \sum_{n=1}^{N} \underbrace{\left(\tfrac{1}{T} \sum_{t=1}^{T} x_{n,t} \varepsilon_{n,t} \right)}_{= \omega_n} \right] = \beta_0 + \left[\tfrac{1}{N} \sum_{n=1}^{N} z_n \right]^{-1} \left[\tfrac{1}{N} \sum_{n=1}^{N} \omega_n \right]$$

Time/Group Dependence in Panel Data

- OLS is unbiased: $E[\hat{\beta} \mid x] = \beta_0$
- OLS is consistent since, $\frac{1}{N}\sum_{n=1}^{N}\omega_n \xrightarrow{prob.} 0$
- If $x_{n,t}$ are independent over n and t, then sandwich estimator provides correct standard errors
- Otherwise,

$$Var\left(\frac{1}{\sqrt{NT}}\sum_{n=1}^{N}\sum_{t=1}^{T}x_{n,t}\varepsilon_{n,t}\right) = \frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}Var(x_{n,t}\varepsilon_{n,t}) + \frac{2}{NT}\sum_{n=1}^{N}\sum_{s<t}Cov(x_{n,t}\varepsilon_{n,t}, x_{n,s}\varepsilon_{n,s})$$

Time/Group Dependence in Panel Data

- Alternatively,

$$\sqrt{N}(\hat{\beta}-\beta_0) = \left[\tfrac{1}{N}\sum_{n=1}^{N} z_n\right]^{-1} \left[\tfrac{1}{\sqrt{N}}\sum_{n=1}^{N} \omega_n\right]$$

$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$

$$N(0, E[z_n]^{-1} Var(\omega_n) E[z_n]^{-1}) \qquad E[z_n] \qquad N(0, Var(\omega_n))$$

$$Var(\omega_n) \approx \tfrac{1}{N}\sum_{n=1}^{N} \omega_n \omega_n' = \tfrac{1}{N}\sum_{n=1}^{N}\left[\tfrac{1}{T}\sum_{t=1}^{T} x_{n,t}\varepsilon_{n,t}\right]\left[\tfrac{1}{T}\sum_{t=1}^{T} x_{n,t}\varepsilon_{n,t}\right]'$$

- Notice that this will not work with small-long panels since LLN in N will not kick it
- As long as ω_n are independent, variance estimator is accurate (does not require any assumption about time-series dependence)
- Clustering will not control for a common time effect

Time/Group Dependence in Panel Data

- Same principal holds if two-way structure is not individuals/time, but individuals/groups (e.g. countries, states)

$$\hat{\beta} = \beta_0 + \left[\frac{1}{G} \sum_{g=1}^{G} \left\{ \frac{1}{I_g} \sum_{i=1}^{I_g} x_{g,i} x_{g,i}' \right\} \right]^{-1} \left[\frac{1}{G} \sum_{g=1}^{G} \left\{ \frac{1}{I_g} \sum_{i=1}^{I_g} x_{g,i} \varepsilon_{g,i} \right\} \right]$$

$$\sqrt{G}(\hat{\beta} - \beta_0) = \left[\frac{1}{G} \sum_{g=1}^{G} \left\{ \frac{1}{I_g} \sum_{i=1}^{I_g} x_{g,i} x_{g,i}' \right\} \right]^{-1} \left[\frac{1}{\sqrt{G}} \sum_{g=1}^{G} \left\{ \frac{1}{I_g} \sum_{i=1}^{I_g} x_{g,i} \varepsilon_{g,i} \right\} \right]$$

Time/Group Dependence in Panel Data

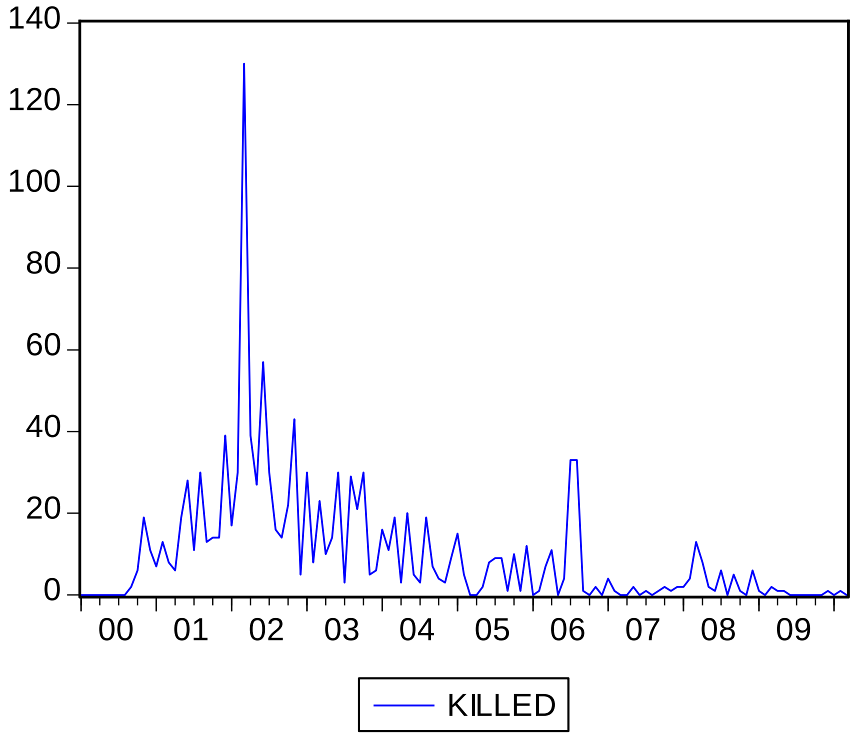
- Suppose there is a group effect, $\varepsilon_{g,i} = u_g + \xi_{g,i}$, u_g and $\xi_{g,i}$ are iid and independent of each other
  - (i.e. individuals in different countries, u_g representing a country effect, possibly due to omitted country variables)
  - If group (e.g. country) fixed effects are excluded, must cluster
  - If group fixed effects are included, no need to cluster
- If country fixed effects are omitted, then clustering deals with within country correlation (as long as $x_{g,i}$ and u_g are not dependent, in which case OLS w/out fixed effects is inconsistent)

Time/Group Dependence in Panel Data

- Implementation:
  - In stata, “regress y x1 x2, cluster(ind)”

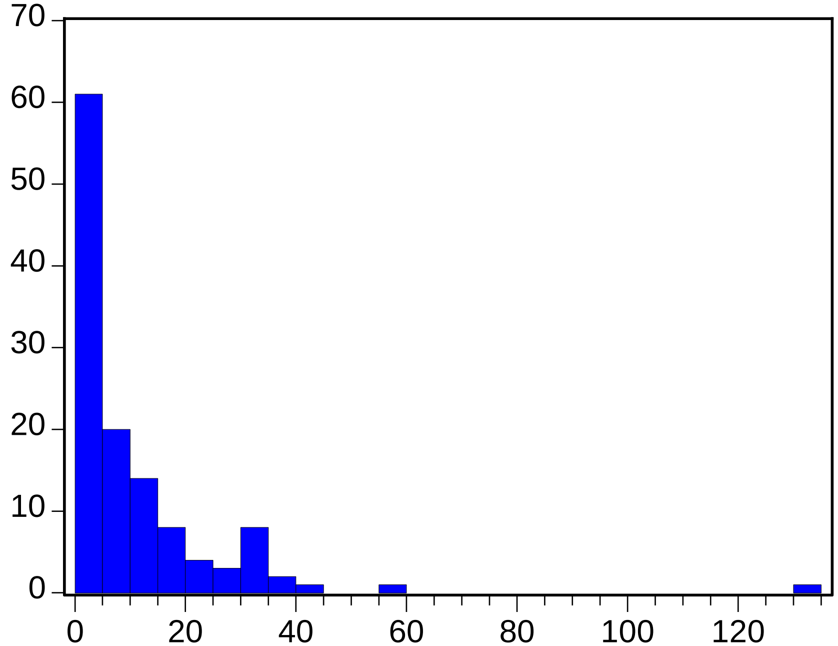
Example: Monthly Terror Attacks in Israel

- Number of Israelis Killed:



Example: Monthly Terror Attacks in Israel

- Number of Israelis Killed:



Series: KILLED
Sample 2000M01 2010M03
Observations 123

Mean	10.00813
Median	5.000000
Maximum	130.0000
Minimum	0.000000
Std. Dev.	15.73265
Skewness	4.160091
Kurtosis	29.29346
Jarque-Bera	3897.928
Probability	0.000000

Example: Monthly Terror Attacks in Israel

- Control for:
  - Election period (3 months leading up to Israeli election)
  - Post peace summit (6 months following peace summit)
  - Right-wing Israeli prime-minister

Example: Monthly Terror Attacks in Israel

- Linear Model in *stata*:
  - Calculate by hand, $m_T = 4$
  - Naïve standard errors, “regress killed killed_m1 elec postsummit rightpm”
  - Robust standard errors, “regress killed killed_m1 elec postsummit rightpm, robust”
  - Newey-West standard errors “newey killed killed_m1 elec postsummit rightpm, lag(4)”

Example: Monthly Terror Attacks in Israel

- Poisson Model in *r*:
  - m_T calculated automatically
  - “mod1 <- glm(killed ~ killed_m1 + elec + postsummit + rightpm,family="poisson",data=xls1)”
  - “coef <- summary(mod1)$coefficients[1:5,1]”
  - “se1 <- summary(mod1)$coefficients[1:5,2]”
  - “se2 <- sqrt(diag(sandwich(mod1)))”
  - “se3 <- sqrt(diag(NeweyWest(mod1)))”