Panel Data Models

Simon Jackman
Department of Political Science
Stanford University

June 3, 2007

Sources of Variation in Panel Data

- n units/individuals; T time points.
- Total Variance = Between Variance plus Within Variance
- Between-Unit Variance: how much variation is cross-sectional (across units)?
- Within-Unit Variance: how much variation is longitudinal?
- Many political science panel data sets from comparative/IR are characterized by the total variance being largely cross-sectional variation; aggregate data from the U.S. states often looks this way too. The things that make units (countries/states) different from one another on y are more or less time-invariant; e.g., institutional or geographical characteristics of units.
- Graphical inspection a useful first step: boxplot the dependent variable by cross-sectional units (if they are not too many of them); similarly, by time.

Unit-Specific Omitted Variables?

Consider regression models for panel data, generically,

$$E(y_{it}|\mathbf{x}_{it}, c_i) = \beta_0 + \mathbf{x}_{it}\boldsymbol{\beta} + c_i$$

where $i = 1, \ldots, n$ indexes units; $t = 1, \ldots, T$ indexes time.

- c_i is an unobserved, time constant random quantity.
- a source of unmodelled, unit-specific heterogeneity in y_{it}
- c_i might be particularly relevant if the data display a good deal of cross-sectional or *between* variation, *and* the $\mathbf{x}_{it}$ available for analysis do a poor job of soaking up that cross-sectional variation.

Ignoring c_i?

$$E(y_{it}|\mathbf{x}_{it}, c_i) = \beta_0 + \mathbf{x}_{it}\boldsymbol{\beta} + c_i$$

- If $\mathbf{x}_{it}$ does a good job of capturing variation in y_{it}, then maybe we can ignore c_i?
- Specifically, if we re-write the model as

$$y_{it} = \beta_0 + \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$$

with the usual assumption about the disturbances,

$$E(u_{it}|\mathbf{x}_{it}, c_i) = 0 \iff E(\mathbf{x}_{it}u_{it}) = \mathbf{0}\,\forall\, i$$

If we make the further assumption

$$E(\mathbf{x}_{it}c_i) = \mathbf{0}$$

then we can ignore the c_i; roll them into a *compound error term*, $v_{it} = c_i + u_{it}$.

Ignoring c_i

- A situation in which we might ignore the c_i is when x_i is assigned randomly to cases, as in an experiment.

- In such a case the x_i are uncorrelated with the c_i (and anything else for that matter), and so estimation of $\boldsymbol{\beta}$ can ignore c_i.

- On the other hand, under these conditions, while $\hat{\boldsymbol{\beta}}$ is unbiased/consistent, there are *efficiency gains* to be had from extracting the c_i from the error term.

- This is what the random effects (RE) estimator does.

- When c_i is correlated with X_i then the fixed effects estimator is an attractive alternative.

Random Effects

- Model:

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it} \\ V(c_i|\mathbf{x}_i) &= \sigma_c^2 \end{aligned}$$

- Assumptions:
  1. **Strict exogeneity:** $E(u_{it}|\mathbf{x}_{it}, c_i) = 0$.
  2. **Orthogonality of unit-specific effects and errors:** $E(c_i|\mathbf{x}_{it}) = E(c_i) = 0$.
  3. **Homoskedasticity of idiosyncratic error:** $V(u_{it}) = \sigma_u^2$

Random Effects

- Together these assumptions imply that in the model

$$
\begin{aligned}
y_{it} &= \mathbf{x}_{it}\boldsymbol{\beta} + v_{it} \\
v_{it} &= c_i + u_{it}
\end{aligned}
$$

we have

$$E(v_{it}|\mathbf{x}_i) = 0, i = 1, \ldots, n; t = 1, \ldots T.$$

Random Effects

- Let $\mathbf{v}_i = (v_{i1}, \ldots, v_{it})'$. Then under the further assumption of no serial correlation

$$E(u_{it}u_{it}|\mathbf{x}_i, c_i) = 0 \,\forall\, t \neq s$$

we have

$$\begin{aligned} V(v_{it}) &= V(c_i) + 2\times C(c_i, u_{it}) + V(u_{it}) \\ &= \sigma_c^2 + \sigma_u^2 \\ C(v_{it}, v_{is}) &= C[(c_i + u_{it}), (c_i + u_{is})] \\ &= V(c_i) = \sigma_c^2 \end{aligned}$$

Random Effects: Estimation by FGLS

Thus

$$\mathbf{\Omega} = E(\mathbf{v}_i\mathbf{v}_i') = \begin{bmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \dots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & \dots & \vdots \\ \vdots & & \ddots & \sigma_c^2 \\ \sigma_c^2 & \dots & \dots & \sigma_c^2 + \sigma_u^2 \end{bmatrix}$$

and if we knew or could estimate σ_c^2 and σ_u^2 we could implement the FGLS *random effects estimator*

$$\hat{\boldsymbol{\beta}}_{RE} = \left(\sum_{i=1}^{n} \mathbf{X}_i' \hat{\mathbf{\Omega}}^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^{n} \mathbf{X}_i' \hat{\mathbf{\Omega}}^{-1} \mathbf{y}_i \right)$$

Random Effects: Estimation by FGLS

Hence, implement in two stages:

1. Run "pooled OLS": $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + v_{it}$, to yield residuals $\hat{v}_{it}^{(\text{OLS})}$.

2. Form consistent estimators

$$
\begin{aligned}
\hat{\sigma}_v^2 &= \frac{1}{NT - K}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{v}_{it}^{(\text{OLS})})^2 \\
\hat{\sigma}_c^2 &= \frac{1}{NT(T-1)/2 - k}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{v}_{it}^{(\text{OLS})}\hat{v}_{is}^{(\text{OLS})} \\
\hat{\sigma}_u^2 &= \hat{\sigma}_v^2 - \hat{\sigma}_c^2 \text{ (i.e., within variance = total variance minus between variance)}
\end{aligned}
$$

Note no guarantee that $\hat{\sigma}_v^2 > 0$; negative estimate of the random effects variance can arise if the u_{it} are negatively serially correlated (in which case we're violating the assumptions underlying the model); time-specific fixed effects can soak some of this up.

Random Effects: Estimation by FGLS

With estimates of σ_c^2 and σ_v^2, now compute

$$\hat{\boldsymbol{\beta}}_{RE} = \left(\sum_{i=1}^{n} \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^{n} \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{y}_i \right)$$

where

$$\hat{\boldsymbol{\Omega}} = \begin{bmatrix} \hat{\sigma}_c^2 + \hat{\sigma}_u^2 & \hat{\sigma}_c^2 & \dots & \hat{\sigma}_c^2 \\ \hat{\sigma}_c^2 & \hat{\sigma}_c^2 + \hat{\sigma}_u^2 & \dots & \vdots \\ \vdots & & \ddots & \hat{\sigma}_c^2 \\ \hat{\sigma}_c^2 & \dots & \dots & \hat{\sigma}_c^2 + \hat{\sigma}_u^2 \end{bmatrix}$$

Random Effects Estimation via ML

See Hsiao (2003, §3.3.3). Under joint normality of the random effects and the idiosycratic errors,

$$\begin{bmatrix} c_i \mid \mathbf{x}_{it} \\ u_{it} \mid \mathbf{x}_{it} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_c^2 & 0 \\ 0 & \sigma_u^2 \end{bmatrix} \right), \forall\, i, t$$

we have

$$\mathbf{y}_i | \mathbf{x}_i \sim N(\mathbf{x}_{it}\boldsymbol{\beta}, \boldsymbol{\Omega}), \;\; i = 1, \ldots, n$$

where $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$ and $\boldsymbol{\Omega}$ (a T-by-T variance-covariance matrix) is defined previously (a function of the unknown parameters σ_c^2 and σ_u^2).

Random Effects Estimation via ML

$$\mathbf{y}_i | \mathbf{x}_i \sim N(\mathbf{x}_{it}\boldsymbol{\beta}, \boldsymbol{\Omega}), \;\; i = 1, \ldots, n$$

This means we have the following likelihood function

$$\mathcal{L}(\boldsymbol{\beta}, \sigma_c^2, \sigma_u^2 | \mathbf{Y}, \mathbf{X}) = p(\mathbf{Y} | \mathbf{X}, \boldsymbol{\beta}, \sigma_c^2, \sigma_u^2) = \prod_{i=1}^{n} \phi_T(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\beta}, \sigma_c^2, \sigma_u^2)$$

where $\phi_T(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\beta}, \sigma_c^2, \sigma_u^2)$ is the T-variate normal density

$$(2\pi)^{-T/2} |\boldsymbol{\Omega}|^{-1/2} \exp\left[\frac{-1}{2}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta})' \boldsymbol{\Omega}^{-1} (\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta})\right]$$

Random Effects Estimation via ML

- Maximize the log of the likelihood function wrt $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_c^2, \sigma_u^2)'$ to yield $\hat{\boldsymbol{\theta}}_{\text{MLE}}$.

- Minus the inverse of the Hessian matrix is an estimate of the variance-covariance matrix of $\hat{\boldsymbol{\theta}}_{\text{MLE}}$.

- Curious fact: with n fixed and asymptotics in T, the estimate of σ_c^2 is inconsistent. There is never enough variation in the c_i; since n fixed (Hsiao 2003, 41).

- In `R`: see `lme` in package `nlme` (Pinheiro and Bates 2000); the FGLS approach is implemented in package `plm`.

Fixed Effects Estimator

- Treat the c_i as *parameters* to be estimated; but (for purposes of implementation) then apply transformations to the data that effectively make the c_i disappear.

- Suppose we have at least two time periods, and average the model

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$$

within a unit to yield

$$\bar{y}_i = \bar{\mathbf{x}}_i\boldsymbol{\beta} + c_i + \bar{u}_i$$

where $\bar{y}_i = T^{-1}\sum_{i=1}^{T} y_{it}$ and similarly for $\bar{\mathbf{x}}_i$. Note that the c_i appears in the averaged model, since the average of a constant is the constant. But if we then consider the *differenced* model, the c_i drops out:

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{x}_i)\boldsymbol{\beta} + u_{it} - \bar{u}_i$$

Fixed Effects Estimator

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{x}_i)\boldsymbol{\beta} + u_{it} - \bar{u}_i$$

- Running OLS on this differences model produces the *fixed effects estimator* of $\boldsymbol{\beta}$.

- Equivalent to running the regression and including a series of mutually exclusive and exhaustive dummy variables for the units/individuals (the unique values of the i subscript); but this is a computationally klunky way to do it for large n.

- Obvious that information about $\boldsymbol{\beta}$ comes from the within-unit covariation between $\mathbf{x}_i$ and $\mathbf{y}_i$; hence the fixed effects estimator is sometimes called the *within estimator*, and the differencing above is sometimes called the *within transformation*.

Fixed Effects Estimator

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{x}_i)\boldsymbol{\beta} + u_{it} - \bar{u}_i$$

- Any variable that has no "within" variation has to be dropped from the analysis (such a variable is co-linear with the fixed effects for the units).
- If a variable is time-invariant within one unit (but not all), then the fixed effect for that unit isn't identified; `agl` example (the U.S. is time-invariant on the `left` cabinet seats variable).

Fixed Effects Estimator

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{x}_i)\boldsymbol{\beta} + u_{it} - \bar{u}_i$$

- Curious fact: with T fixed and $n \to \infty$ (there are more units in the analysis), we are estimating an ever-increasing number of c_i (but with no more information about any of the c_i), giving rise to *incidental parameters* (e.g., Lancaster 2000).

- This turns out to be especially consequential in the case of non-linear/discrete data models.

- But for linear models, under strict exogeneity, inference for $\boldsymbol{\beta}$ is independent of inferences about c_i (e.g., the within-estimator proceeds after sweeping out the c_i) and we can obtain consistent estimators of $\boldsymbol{\beta}$ despite lacking consistent estimators of c_i as $n \to \infty$.

Fixed Versus Random Effects

- Fixed effects estimator: under strict exogeneity and conditional iid assumptions above, $\hat{\boldsymbol{\beta}}_{FE}$ is consistent.

- Random effects estimator: under same assumptions as above, plus $E(c_i|\mathbf{x}_i) = E(c_i)$ (orthogonality of unit-specific effects and predictors), $\hat{\boldsymbol{\beta}}_{RE}$ is also consistent, but more efficient than FE.

- Use *RE* if the orthogonality assumption holds; but often it doesn't.

Hausman Test for Fixed versus Random Effects

Proposition 1. [Hsiao (2003, Lemma 3.5.1)] *Based on a sample of n observations, consider two estimates $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$ that are both consistent and asymptotically normal, with $\hat{\boldsymbol{\beta}}_0$ attaining the Cramer-Rao bound. Let $\hat{\mathbf{q}} = \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0$. Then the limiting joint distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta})$ and $\sqrt{n}\hat{\mathbf{q}}$ has zero covariance between these two terms.*
Proof: *Rao (1973, 317).*

Corollaries:

1. $V(\hat{\mathbf{q}}) = V(\hat{\boldsymbol{\beta}}_1) - V(\hat{\boldsymbol{\beta}}_0)$

2. (Hausman 1978): Under $H_0 : E(c_i|\mathbf{x}_i) = 0$ (orthogonality), $m = \hat{\mathbf{q}}'V(\hat{\mathbf{q}})^{-1}\hat{\mathbf{q}} \sim \chi^2_k$, where k is the dimension of $\mathbf{q}$.

This suggests an easily-implemented, asymptotically valid test ($T \rightarrow \infty$): simply compare $\hat{\boldsymbol{\beta}}_{FE}$ and $\hat{\boldsymbol{\beta}}_{RE}$ using the testing framework above. If m exceeds a critical quantile of the χ^2_k density, then reject H_0 (orthogonality $\Rightarrow$ *RE*) in favor of H_A (*FE*).

Dealing With Violations of IID Assumptions

- Plausible that there is/are:

  1. unit-wise heteroskedasticity: $V(u_{it}|\mathbf{x}_{it}) \neq V(u_{jt}|\mathbf{x}_{jt})$ for some $i \neq j$. Put differently, do we need to let σ_u^2 pick up an i subscript (varying across units).
  2. within-unit serial correlation of the disturbances: i.e., $C(u_{it}, u_{is}) \neq 0$ for some $t \neq s$.
  3. contemporaneous shocks: $C(u_{it}, u_{jt}) \neq 0$ for some $i \neq j$.

- Modern/frequentist treatment of these problems usually (a) ignores at estimation stage; (b) use a "robust" estimator of the variance-covariance matrix of the estimator (robust in the sense that we have a consistent estimator of the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ as $n \to \infty$).

Asymptotically Robust Variance-Covariance Estimators with Panel Data

1. `white1` in `plm`: asymptotically robust to general patterns of heteroskedasticity

2. `white2` in `plm`: asymptotically robust to unit-wise heteroskedasticity

3. `arellano` in `plm` (Arellano 1987): asymptotically robust to arbitrary forms of any heteroskedasticity or serial correlation.

4. `pcse` package implements the Beck and Katz (1995) "panel-corrected" standard errors: asymptotically robust to unit-wise heteroskedasticity and contemporaneous correlations between units (asymptotics in T, not n).

Discrete Panel Data

- Restrict attention here to binary data

- Two cases to consider:

  1. Static, with unit-wise heterogeneity
  2. Dynamic (see earlier lectures re transition models)

Binary Panel Data, Static Model, with Unit-Specific Effects

$$\Pr(y_{it} = 1 | \mathbf{x}_{it}, \boldsymbol{\beta}, \alpha_i) = F(\mathbf{x}_{it}\boldsymbol{\beta} + \alpha_i)$$

- F is usually the logistic or normal CDF (logit or probit, respectively).
- Classical approach: estimation via maximum likelihood, properties of estimators are asymptotic (almost no finite sample results).
- Note immediately that if there is no time variation in y_{it} then the fixed effects estimator of α_i will be (or ought to be)

$$\hat{\alpha}_i^{(\text{MLE})} = \begin{cases} +\infty & \iff \sum_{t=1}^{T} y_{it} = T \text{ (all ones)} \\ -\infty & \iff \sum_{t=1}^{T} y_{it} = 0 \text{ (all zeros)} \end{cases}$$

i.e., units like these have to be dropped from the analysis if we deploy a fixed effects estimator.

Binary Panel Data, Static Model, with Unit-Specific Effects: Incidental Parameters Problem

- No consistency possible as $n \rightarrow \infty$, because of the incidental parameters problem (Neyman and Scott 1948; Lancaster 2000).

- That is, the issue is not just the inconsistent estimation of α_i as $n \rightarrow \infty$ with T fixed. Because the logit/probit model is nonlinear, the inconsistency in $\hat{\alpha}_i$ afflicts $\hat{\boldsymbol{\beta}}$. Well known that in this case (e.g., Hsiao 2003, §7.3.1)

$$\text{plim}_{n \rightarrow \infty} \hat{\boldsymbol{\beta}} = 2\boldsymbol{\beta}.$$

Binary Panel Data, Static Model, with Unit-Specific Effects: Incidental Parameters Problem

- Solution (Neyman and Scott 1948): can we find functions that do not depend on the incidental parameters (in this case, the α_i), but are functions of the data and the unknown parameters, i.e.,

$$\Psi(\mathbf{y}_1, \ldots, \mathbf{y}_n | \boldsymbol{\beta}, \mathbf{X})$$

  that converge to zero in probability as $n \rightarrow \infty$? An estimator of $\boldsymbol{\beta}$ based on setting Ψ to zero and solving for $\boldsymbol{\beta}$ will be consistent (subject to some regularity conditions).

- How to find such a Ψ? Trick: is there a sufficient statistic τ_i for α_i that does not depend on $\boldsymbol{\beta}$? If so, then condition on τ_i and maximize the *conditional likelihood*.

Binary Panel Data, Static Model, with Unit-Specific Effects: Incidental Parameters Problem

- For logit we can use this trick as follows:

$$p(\mathbf{y}_i) = \frac{\exp\left[\alpha_i \sum_{t=1}^{T} y_{it} + \boldsymbol{\beta}' \sum_{t=1}^{T} \mathbf{x}_{it} y_{it}\right]}{\prod_{t=1}^{T}\left[1 + \exp(\mathbf{x}_{it}\boldsymbol{\beta} + \alpha_i)\right]}$$

since $\tau_i = \sum_{t=1}^{T} y_{it}$ is a sufficient statistic for α_i.

Binary Panel Data, Static Model, with Unit-Specific Effects: Incidental Parameters Problem

- We rule out the cases where $\tau_i = \sum_{t=1} y_{it} = T$ or $\tau_i = 0$.

- There are $\binom{T}{\tau_i}$ binary sequences $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})$, $y_{it} \in \{0, 1\}$ that could generate $\sum_{t=1}^{T} y_{it} = \tau_i$; we have to sum over these possible sequences to form $p(\mathbf{y}_i | \tau_i)$.

- Notation: let $\mathcal{B}_i$ be the set of $\binom{T}{\tau_i}$ sequences $\mathbf{d}_{ij} = (d_{ij1}, \ldots, d_{ijT})'$ where

  1. $d_{ijt} \in \{0, 1\}$
  2. $\sum_{t=1}^{T} d_{ijt} = \tau_i = \sum_{t=1}^{T} y_{it}$
  3. j indexes the $\binom{T}{\tau_i}$ sequences in $\mathcal{B}_i$

- Then the conditional probability of $\mathbf{y}_i$ given $\tau_i = \sum_{t=1}^{T} y_{it}$ is

$$p(\mathbf{y}_i | \tau_i) = \frac{\exp\left[\boldsymbol{\beta}' \sum_{t=1}^{T} \mathbf{x}_{it} y_{it}\right]}{\sum_{\mathbf{d}_{ij} \in \mathcal{B}_i} \exp\left[\boldsymbol{\beta}' \sum_{t=1}^{T} \mathbf{x}_{it} d_{ijt}\right]}$$

Binary Panel Data, Static Model, with Unit-Specific Effects: Incidental Parameters Problem

- The resulting conditional MLEs of $\boldsymbol{\beta}$ have good properties as $n \to \infty$.
- This estimator often referred to as the Chamberlain (1980) conditional MLE estimator for fixed effects binary panel data models.
- Easy case with $T = 2$. We ignore $\tau_i = 0$ or $\tau_i = 2$, leaving just $\tau_i = 1$. Let

$$\omega_i = \begin{cases} 0 & \Longleftrightarrow \quad \mathbf{y}_i = (1, 0) \\ 1 & \Longleftrightarrow \quad \mathbf{y}_i = (0, 1) \end{cases}$$

- $\Pr(\omega_i = 1 | \tau_i = 1) = F[(\mathbf{x}_{i2} - \mathbf{x}_{i1})\boldsymbol{\beta}] = F_i$
- Log-likelihood:

$$\log \mathcal{L} = \sum_{i:\tau_i=1} \left\{ \omega_i \log F_i + (1 - \omega_i) \log(1 - F_i) \right\}.$$

Binary Panel Data, Static Model, with Unit-Specific Effects: Incidental Parameters Problem

- Note that we don't estimate the unit-specific effects α_i; indeed, the only thing that makes this model work is that we estimate $\boldsymbol{\beta}$ conditioning on a sufficient statistic for them.

- Roughly analagous to subtracting of the unit-specific means of $\mathbf{y}_i$ and $\mathbf{x}_i$ to get the within-estimator/fixed-effects estimator in the linear/continuous case (in the sense that the unit-specific effects then drop out of the estimation).

- We can't implement this solution with a probit model; the probit likelihood doesn't factor in the nice way that the logit likelihood does, so as to let us get a sufficient statistic for the unit-specific α_i.

- That said, some limited Monte Carlo evidence suggests fixed effects probit not bad (Hsiao 2003, §7.3.1.c).

Binary Panel Data, Static Model, Random Unit-Specific Effects

- We can gain efficiency if the α_i are treated as random effects.

- Generally, suppose α_i come from some distribution $G(\alpha_i|\boldsymbol{\delta})$.

- Let $F_{it} = F(\mathbf{x}_{it}\boldsymbol{\beta} + \alpha_i)$; $F \equiv$ logistic or normal CDF for random effects panel logit/probit, respectively.

- Maximize the *marginal* log-likelihood

$$\log \mathcal{L} = \sum_{i=1}^{N} \log \int \prod_{t=1}^{T} F_{it}^{y_{it}}(1 - F_{it})^{1-y_{it}} dG(\alpha_i|\boldsymbol{\delta})$$

- Elaborations by Chamberlain (1980) to deal with the case where the α_i may be correlated with the regressors $\mathbf{x}_{it}$; gives rise to a nasty estimation problem.

Binary Panel Data: `R` packages

- `glmmML`: generalized linear models with random intercepts. Does fixed effects as well. Can also handle Poisson data. Implements the marginal likelihood model described above for random effects, with choices of a normal/logistic/Cauchy distribution for the α_i.

Bayesian Approach

- All parameters are random (fixed versus random distinction meaningless in the Bayesian approach).

- Use Bayes Rule to update beliefs about random quantities (parameters, unit-specific effects) conditional on the data:

$$\underbrace{p(\boldsymbol{\theta}|\text{data})}_{\text{posterior}} \propto \underbrace{p(\boldsymbol{\theta})}_{\text{prior}} \underbrace{p(\text{data}|\boldsymbol{\theta})}_{\text{likelihood}}$$

- In the Bayesian approach we don't care about the distribution of estimates of θ under repeated sampling. In fact, repeated sampling and/or asymptotics a meaningless fiction in many social-science circumstances. Rather, Bayesian approach conditions on the data at hand to update about $\boldsymbol{\theta}$?

- The posterior density $p(\boldsymbol{\theta}|\mathbf{y})$ is all you need to make probability statements about $\boldsymbol{\theta}$; contrast this with the classical/frequentist focus on asymptotically-valid approximations to the sampling distribution of $\hat{\boldsymbol{\theta}}$.

Bayesian Example: Unit-Specific Heterogeneity ("Random Effects")

- Model/Likelihood:

$$\begin{aligned} y_{it} &\sim N(\mathbf{x}_{it}\boldsymbol{\beta} + c_i, \sigma^2) \\ c_i &\sim N(0, \omega^2) \end{aligned}$$

- Priors:

$$\begin{aligned} \boldsymbol{\beta} &\sim N(\mathbf{b}_0, \mathbf{B}_0) \\ \sigma^2 &\sim \text{inverse-Gamma}(\nu_0/2, \nu_0\sigma_0^2/2) \\ \omega^2 &\sim \text{inverse-Gamma}(\kappa_0/2, \kappa_0\omega_0^2/2) \end{aligned}$$

where $\mathbf{b}_0$, $\mathbf{B}_0$, ν_0, σ_0^2, κ_0 and ω_0^2 are user-supplied *hyperparameters* expressing the researcher's prior knowledge (or ignorance) about $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \omega^2)$.

Bayesian Example: Unit-Specific Heterogeneity ("Random Effects")

Applying Bayes Rule:

$$
\begin{aligned}
p(\boldsymbol{\theta}|\mathbf{y},\mathbf{X}) &\propto p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta},\mathbf{X}) \\
&= \underbrace{p(\boldsymbol{\beta})\, p(\sigma^2)\, p(\omega^2)}_{\text{priors}} \underbrace{\prod_{i=1}^{n} \phi_T(\mathbf{y}_i|\mathbf{x}_i,\sigma^2,\omega^2)}_{\text{likelihood}}
\end{aligned}
$$

where

$$
\begin{aligned}
p(\boldsymbol{\beta}) &\propto \exp\left[\frac{-1}{2}(\boldsymbol{\beta} - \mathbf{b}_0)'\mathbf{B}_0^{-1}(\boldsymbol{\beta} - \mathbf{b}_0)\right] \\
p(\sigma^2) &\propto (\sigma^2)^{-\frac{\nu_0}{2}-1} \exp\left(\frac{-\nu_0\sigma_0^2}{2\sigma^2}\right) \\
p(\omega^2) &\propto (\omega^2)^{-\frac{\kappa_0}{2}-1} \exp\left(\frac{-\kappa_0\omega_0^2}{2\omega^2}\right)
\end{aligned}
$$

Bayesian Example: Unit-Specific Heterogeneity ("Random Effects")

Applying Bayes Rule:

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y},\mathbf{X}) &\propto p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta},\mathbf{X}) \\ &= \underbrace{p(\boldsymbol{\beta})\, p(\sigma^2)\, p(\omega^2)}_{\text{priors}} \underbrace{\prod_{i=1}^{n} \phi_T(\mathbf{y}_i|\mathbf{x}_i,\sigma^2,\omega^2)}_{\text{likelihood}} \end{aligned}$$

and where

$$\begin{aligned} \phi_T(\mathbf{y}_i|\mathbf{x}_i,\sigma^2,\omega^2) &\propto |\boldsymbol{\Omega}|^{-1/2} \exp\left[\frac{-1}{2}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta})\right] \\ \boldsymbol{\Omega} &= \begin{bmatrix} \sigma_c^2+\sigma_u^2 & \sigma_c^2 & \dots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2+\sigma_u^2 & \dots & \vdots \\ \vdots & & \ddots & \sigma_c^2 \\ \sigma_c^2 & \dots & \dots & \sigma_c^2+\sigma_u^2 \end{bmatrix} \end{aligned}$$

Bayesian Example: Unit-Specific Heterogeneity ("Random Effects")

- Posterior density looks extremely difficult to characterize analytically (lots of cumbersome/tedious mathematics)

- Modern Bayesian approach is to use Monte Carlo methods to sample an arbitrarily large number of times from the posterior density

- Modern algorithms lets us sample from the posterior density even though we can't characterize it analytically.

- In particular, the Gibbs sampler breaks the cumbersome, high-dimensional joint posterior density into its constituent conditional densities.

Computing Posterior Densities with the Gibbs Sampler

Consider a parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)'$. The posterior density $\pi(\boldsymbol{\theta}|\text{data})$ can be computed via the following scheme (the Gibbs sampler): we **initialize** the sampler with start values $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2^{(0)})'$. Then for $t = 1, \ldots, T, \ldots$,

1. sample $\boldsymbol{\theta}_1^{(t)}$ from $g_1(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2^{(t\text{-}1)}, \text{data})$

2. sample $\boldsymbol{\theta}_2^{(t)}$ from $g_2(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1^{(t)}, \text{data})$

This yields $\boldsymbol{\theta}^{(t)} = (\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)})$ $t = 1, \ldots, T, \ldots$.

The Gibbs Sampler

- Initialize the sampler with starting values $\boldsymbol{\theta}^{(0)}$.
- Let the sampler run, generating $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots$
- Under a very wide set of conditions, as $t \rightarrow \infty$, each Gibbs sample $\boldsymbol{\theta}^{(t)}$ can be regarded as samples from the posterior density $\pi(\boldsymbol{\theta}|\text{data})$.
- More formally: The sampler moves away from the initial values, providing a "random tour" of the (high-dimensional) parameter space, visiting locations in the parameter space with frequencies proportional to the posterior density.
- Even more formally: The output of the Gibbs sampler constitutes a **Markov chain** on the parameter space for $\boldsymbol{\theta}$, with transition probabilities such that the "equilibrium", "limiting", or "stationary" distribution of the Gibbs sampler is the posterior density $\pi(\boldsymbol{\theta}|\text{data})$.

Inference with the Gibbs Sampler

Extremely simple.

- Store output of the Gibbs sampler
- Compute any summary statistic you like (mean, median, confidence intervals). In particular, compute any function of the Gibbs sampler output
- These estimates get better with more Gibbs samples
- Modern computing power makes this approach to estimation and inference possible. First anticipated in the first half of the 20th century (Metropolis and Ulam 1949)
- Now implemented in freeware statistical programs: e.g., `R` packages such as `MCMCpack` (Martin, and Quinn 2007); my own `pscl` (Jackman 2007b); `OpenBUGS` (Spiegelhalter et al. 2003), or `JAGS`.

Inference with the Gibbs Sampler

See Geweke (1989) and Tanner (1996) for limit theorems for Monte Carlo based inference: e.g., **"Simulation Consistency"**. Let $\{\theta_t\}$ be a sequence of sampled values from the posterior density, $\pi(\theta|\mathbf{y})$. Then

$$\lim_{T\to\infty} \frac{1}{T - n} \sum_{t=n}^{T} g(\theta_t) \to E[g(\theta)|\mathbf{y}] = \int g(\theta)\pi(\theta|\mathbf{y})d\theta$$

This is an example of Monte Carlo integration. For surveys of Bayesian inference via Markov chain Monte Carlo, see any of my review articles (Jackman 2000a; 2000b; 2004), my in-progress book manuscript (Jackman 2007a), or the books by Gelman et al. (2004) or Gill (2002).

Bayesian Inference for Panel Data Regression Model with a Hierarchical Model for Unobserved Unit Level Heterogeneity ("Random Effects")

The model is $y_{it} \sim N(\mathbf{x}_{it}\boldsymbol{\beta} + \alpha_i, \sigma^2)$, with $\alpha_i \sim N(0, \omega^2)$. As shown earlier, the posterior density looks formidable, but a Gibbs sampler for this problem is trivial to implement in `OpenBUGS`/`JAGS`:

```
model{
  for(i in 1:N){                          ## loop over units
    for(j in 1:T){                        ## loop over time periods
      mu[i,j] <- x[i,j]*beta + alpha[i]
      y[i,j] ~ dnorm(mu[i,j],tau)         ## likelihood
    }
  }

  for(i in 1:N){
    alpha[i] ~ dnorm(0,tau.alpha)         ## priors for unit-specific terms
  }
  beta ~ dnorm(0,.0001)            ## prior for beta
  tau.alpha ~ dgamma(.01,.01)             ## prior for tau.alpha (precision)
  tau ~ dgamma(.01,.01)                   ## prior for tau (precision)

  sigma <- 1/sqrt(tau)                    ## std dev of idiosyncratic errors
  sigma.alpha <- 1/sqrt(tau.alpha)        ## std dev of unit-specific terms
}
```

Bayesian Inference for Panel Data Regression Model with a Non-Hierarchical model for Unobserved Unit Level Heterogeneity ("Fixed Effects")

$y_{it} \sim N(\mathbf{x}_{it}\boldsymbol{\beta} + \alpha_i, \sigma^2)$; no hierarchical structure linking the α_i:

```
model{
  for(i in 1:N){                              ## loop over units
    for(j in 1:T){                            ## loop over time periods
      mu[i,j] <- x[i,j]*beta + alpha[i]
      y[i,j] ~ dnorm(mu[i,j],tau)             ## likelihood
    }
  }

  for(i in 1:N){
    alpha[i] ~ dnorm(0,.0001)                 ## priors for unit-specific terms
  }
  beta ~ dnorm(0,.0001)                       ## prior for beta
  tau ~ dgamma(.01,.01)                       ## prior for tau (precision)
  sigma <- 1/sqrt(tau)                        ## std dev of idiosyncratic errors
}
```

Extensions to Varying Coefficient Models

Western (1998); see `corporatism` example on my web pages.

References

Beck, Nathaniel and Jonathan Katz. 1995. "What to Do (and Not to Do) with Time-Series Cross-Section Data in Comparative Politics." *American Political Science Review* 89:539--556.

Chamberlain, Gary. 1980. "Analysis of Covariance with Qualitative Data." *Review of Economic Studies* 47:225--238.

Gelman, Andrew, John B. Carlin, Hal S. Stern and Donald B. Rubin. 2004. *Bayesian Data Analysis*. Second ed. Boca Raton, Florida: Chapman and Hall.

Geweke, J. 1989. "Bayesian inference in econometric models using Monte Carlo integration." *Econometrica* 57:1317--1339.

Gill, Jeff. 2002. *Bayesian Methods: A social and behavioral sciences approach*. New York: Chapman and Hall.

Hausman, J. 1978. "Specification Tests in Econometrics." *Econometrica* 46:1251--1271.

Hsiao, Cheng. 2003. *Analysis of Panel Data*. Second ed. New York: Cambridge University Press.

Jackman, Simon. 2000a. "Estimation and Inference Are Missing Data Problems: Unifying Social Science Statistics via Bayesian Simulation." *Political Analysis* 8:307--332.

Jackman, Simon. 2000b. "Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo." *American Journal of Political Science* 44:375--404.

Jackman, Simon. 2004. "Bayesian Analysis for Political Research." In *Annual Reviews of Political Science*. Vol. 7 Palo Alto, California: Annual Reviews pp. 483--505.

Jackman, Simon. 2007a. *Bayesian Analysis for the Social Sciences*. Hoboken, New Jersey: Wiley.

Jackman, Simon. 2007b. *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University*. Stanford, California: Department of Political Science, Stanford University. R package version 0.93.
URL: *http://pscl.stanford.edu*

Lancaster, Tony. 2000. “The Incidental Parameter Problem since 1948.” *Journal of Econometrics* 95:391--413.

Martin, Andrew D., and Kevin M. Quinn. 2007. *MCMCpack: Markov chain Monte Carlo (MCMC) Package*. R package version 0.8-1.
URL: *http://mcmcpack.wustl.edu*

Metropolis, N. and S. Ulam. 1949. “The Monte Carlo Method.” *Journal of the American Statistical Association* 44:335--341.

Neyman, J. and E. L. Scott. 1948. “Consistent Estimates Based on Partially Consistent Observations.” *Econometrica* 16:1--32.

Pinheiro, José C. and Douglas M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. New York: Springer.

Rao, C. Radhakrishna. 1973. *Linear Statistical Inference and Its Applications*. Second ed. New York: Wiley.

Spiegelhalter, David J., Andrew Thomas, Nicky Best and Dave Lunn. 2003. *WinBUGS User Manual Version 1.4*. Cambridge, UK: MRC Biostatistics Unit.

Tanner, Martin A. 1996. *Tools for Statistical Inference: Methods for the Exploration of Postrior Distributions and Likelihood Functions*. Third ed. New York: Springer-Verlag.

Western, Bruce. 1998. "Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modelling Approach." *American Journal of Political Science* 42:1233--1259.