Distribution simulations

The presence of an unobserved, individual effect c_i in a panel data model will create a correlation over time in the outcome variable, even if the idiosyncratic error is completely random. Still, with the standard assumptions in place, ordinary least squares will yield a consistent estimator. Consistent but inefficient. The correlated structure of the composite error, which imclused the individual effect, must be made in order to correctly identify the variance-covariance structure.

The following illustrates the variance of the linear estimator, relative to the estimate that fully accounts for the fixed effects. The data generating process is defined by $\mathbf{y} = 10\mathbf{c} + \mathbf{x} + \epsilon$, where $\epsilon \sim N(0,1)$ and $c_i \in \{1,2,3,4,5\}$. By construction, $\mathbf{x}$ and $\mathbf{c}$ are orthogonal, so that the pooled linear estimator will be consistent. Figure 1 and 2 together show why the simple, pooled estimator will be inefficient. The following code generates a panel data set with $N = 5$ and $T = 20$, and then plots the full data set with an overall linear fit in Figure 1. It is apparent that the linear fit is subject to the spread of the covariate $\mathbf{x}$ from within a particular group. Generally, increased variation in the cofactors will yield a more precise estimator; but the variable intercept (read: fixed effect) that is relegated to the error term could potentially reverse this relationship.

```
library(ggplot2)
c <- rep(c(1,2,3,4,5), 20); x <- rnorm(100); eps <- rnorm(100)
y <- 10 * c + x + eps
p <- ggplot(data.frame(c, x, y), aes(x = x, y = y, color=c))
(p <- p + geom_point() + geom_smooth(method=lm))
```

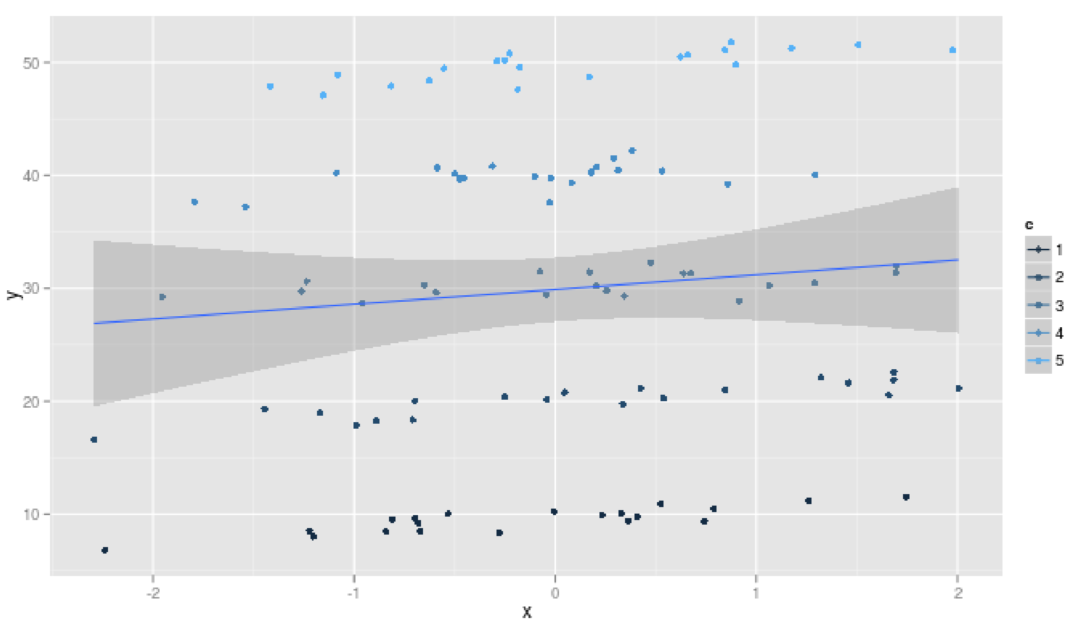

Figure 1: Fixed effect scatterplot with pooled, linear fit

We can directly calculate the variance of the coefficient on $\mathbf{x}$, but this is too much effort. Instead, we can bootstrap the distributions, since we really only care about the relative efficiency of the estimators. Figure

2 clearly illustrates that the pooled estimate is inefficient relative to directly modelling the true fixed effect. We use $B = 100$ iterations, regenerating the data each time. For this simple example, the value of the individual identifier is an argument in the data generating process. Even if this were not the case, however, the distributions in Figure 2 suggest that there must be a better way to weight the estimator to achieve a greater efficiency. And in fact, there is — clustered variance, or the `robust` option in Stata. The procedure to directly calculate the clustered variance is detailed in the lecture notes, and will not be presented again here.

```
B <- 100
fe.res <- rep(NA, B); ols.res <- rep(NA, B)

for (i in 1:B) {
  c <- rep(c(1,2,3,4,5), 20); x <- rnorm(100); eps <- rnorm(100)
  y <- 10 * c + x + eps

  ols <- lm(y ~ x)
  ols.res[i] <-  ols$coefficients[["x"]]

  fe <- lm(y ~ x + factor(c))
  fe.res[i] <- fe$coefficients[["x"]]
}
```

The geometric argument for the spread of the OLS estimates is straightforward: variation in the covariates will have differential impacts on the slope of the linear fit, depending on which strata is represented. If by chance, for example, $\mathbf{x}$ observations were disproportionately selected from the upper tail of the normal distribution when $\mathbf{c} == 1$, then the pooled linear fit will slope downward. If $\mathbf{c} == 5$, however, the linear fit will slope upwards. This alone should give pause in assessing the efficiency of the pooled estimator. Conditional on the covariates, all observations should be given equal weight. The clustered variances help to mitigate this effect by appropriately reweighting the observations.

```
labels <- c(rep("FE", B), rep("pooled", B))
sim <- data.frame(coefficient=c(fe.res, ols.res), method=labels)
ggplot(sim, aes(x = coefficient, fill=method)) + geom_density(alpha=0.2)
```

Suppose that the model was not linear, but rather characterized by a limited dependent variable. What will happen to the consistency and efficiency of the pooled estimate, without taking into account the correlated structure of the error? Figure 3 indicates that the additional variation in the composite error is not averaged away. Instead, the pooled OLS estimator is centered around the wrong estimate, suggesting that the impact of $\mathbf{x}$ on $\mathbf{y}$ is smaller than it is in truth. The value of directly modelling the error is complicated by the nonlinear Probit model. The following code first generates a binary dependent variable from the random covariates, and then esimates the generalized linear model using the Probit function as the binomial link.

```
fe.probit <- rep(NA, B); pooled.probit <- rep(NA, B)

for (i in 1:B) {
  c <- rep(c(1,2,3,4,5), 20); x <- rnorm(100); eps <- rnorm(100)
  y <- ifelse(c + x + eps > 5, 1, 0)

  pool <- glm(y ~ x, family = binomial(link = "probit"))
  pooled.probit[i] <- pool$coefficients[["x"]]

  fe <- glm(y ~ x + factor(c), family = binomial(link = "probit"))
  fe.probit[i] <- fe$coefficients[["x"]]
}
```

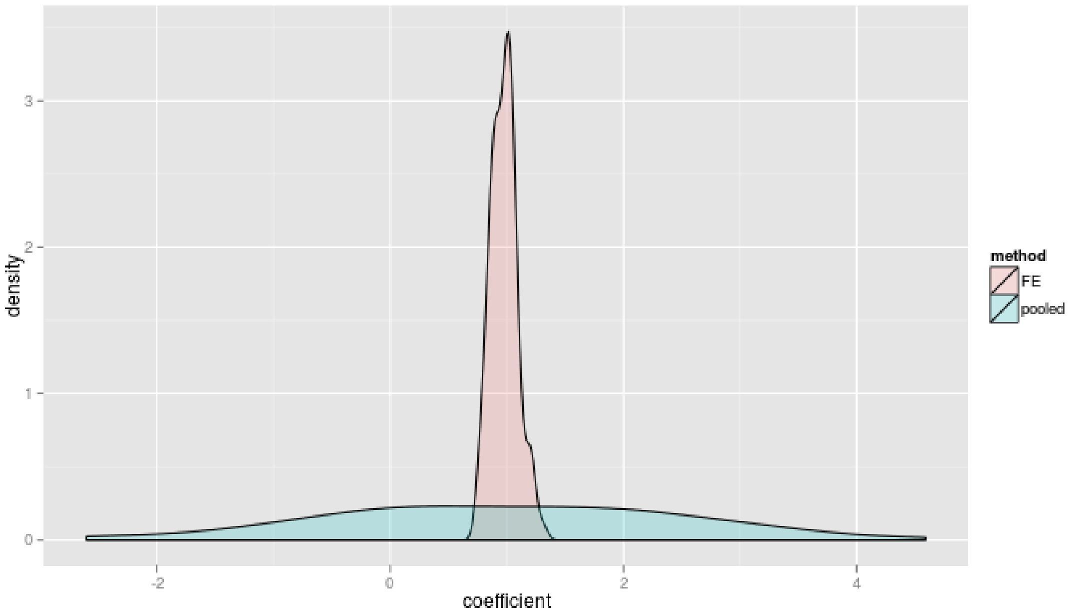

Figure 2: Pooled OLS versus fixed effects, simulated distributions

```
labels <- c(rep("FE", B), rep("pooled", B))
sim.probit <- data.frame(coefficient=c(fe.probit, pooled.probit), method=labels)
ggplot(sim.probit, aes(x = coefficient, fill=method)) + geom_density(alpha=0.2)
```

Two-way, fixed effects panel model

Consider the fixed effects, two-way component panel data model:

$$y_{it} = \alpha + x_{it}\beta + \mu_i + \lambda_t + \epsilon_{it}$$

The fixed effects estimator of β can be obtained by regressing $\mathbf{y}$ on $\mathbf{X}$, $\mathbf{Z}_1$, and $\mathbf{Z}_2$, where $\mathbf{Z}_1 = \mathbb{I}_N \otimes \kappa$ is a matrix of unit indicators and $\mathbf{Z}_2 = \iota \otimes \mathbb{I}_T$ is a matrix of time period indicators, with κ a vector of ones of dimension T and ι a vector of ones of dimension N. Note that $\dim(\mathbf{Z}_1) = TN \times N$ and $\dim(\mathbf{Z}_2) = TN \times T$, assuming a balanced panel.

The computation for this regression is daunting, however, since it requires the inversion of a $(k + N + T - 1) \times (k + N + T - 1)$ matrix. The Frisch-Waugh (FW) theorem suggests that instead of a direct regression, we can demean the variables across time and units. The FW theorem proves that a one-way within transformation will yield the same estimator as a fixed effects regression; and the theorem can be extended for both individual and time effects. The error component structure has the form $u_{it} = \mu_i + \lambda_t + \epsilon_{it}$, which is can be translated into matrix form: $u = (\mathbb{I}_N \otimes \kappa)\alpha + (\iota \otimes \mathbb{I}_T)\lambda + \epsilon$, with $\alpha = [\alpha_1, \alpha_2, \ldots \alpha_N]'$ and $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_T]'$. The error structure suggests a candidate *purging* matrix, which removes the individual- and time-specific effects, along with the overall mean. Call this matrix $\mathbf{Q} = \mathbb{I}_N \otimes \mathbb{I}_T - \mathbb{I}_N \otimes \kappa\kappa'/T - \mathbb{I}_T \otimes \iota\iota'/N + \iota\iota'/N \otimes \kappa\kappa'/T$, which will remove, in turn, the fixed time, unit, and total (through space and time) effect. If both $\mathbf{y}$ and $\mathbf{X}$ are sorted by unit and time, then a regression of $\mathbf{Qy}$ on $\mathbf{QX}$ should yield an unbiased estimate of β with a properly identified error structure.

Define $\mathbb{P}_1 = \mathbb{I}_{NT} - \mathbf{Z}_1(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'$ and $\mathbb{P}_2 = \mathbb{I}_{NT} - \mathbf{Z}_2(\mathbf{Z}_2'\mathbf{Z}_2)^{-1}\mathbf{Z}_2'$ to be the projection matrices for individual and time fixed effects, respectively. It is sufficient to prove that $\mathbb{P}_1\mathbb{P}_2 = \mathbf{Q}$ to create a composition projection

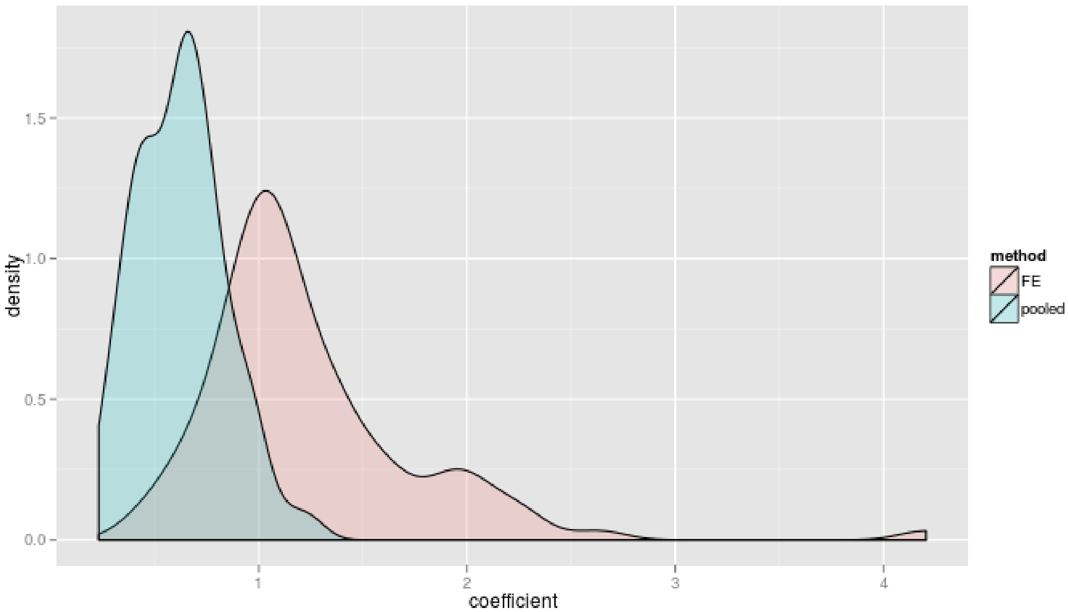

Figure 3: Limited dependent variable: inconsistency of pooled estimator

matrix: first a within transformation ignoring the time effects followed by a within transformation ignoring the individual effects. First, note that:

$$\begin{aligned}
\mathbb{P}_2 = \mathbb{I}_{NT} - \mathbf{Z}_2(\mathbf{Z}_2'\mathbf{Z}_2)^{-1}\mathbf{Z}_2' &= \mathbb{I}_{NT} - \mathbf{Z}_2((\iota \otimes \mathbb{I}_T)'(\iota \otimes \mathbb{I}_T))^{-1}\mathbf{Z}_2' \\
&= \mathbb{I}_{NT} - \mathbf{Z}_2((\iota' \otimes \mathbb{I}_T')(\iota \otimes \mathbb{I}_T))^{-1}\mathbf{Z}_2' \\
&= \mathbb{I}_{NT} - \mathbf{Z}_2(\iota'\iota \otimes \mathbb{I}_T)^{-1}\mathbf{Z}_2' \\
&= \mathbb{I}_{NT} - \mathbf{Z}_2(N \cdot \mathbb{I}_T)^{-1}\mathbf{Z}_2' \\
&= \mathbb{I}_{NT} - N^{-1}\mathbf{Z}_2\mathbf{Z}_2' \\
&= \mathbb{I}_{NT} - N^{-1}(\iota \otimes \mathbb{I}_T)(\iota \otimes \mathbb{I}_T)' \\
&= \mathbb{I}_{NT} - N^{-1}(\iota\iota' \otimes \mathbb{I}_T) \\
&= \mathbb{I}_{NT} - (\iota\iota'/N \otimes \mathbb{I}_T)
\end{aligned}$$

A similar, nearly symmetric argument can be made to show that $\mathbb{P}_1 = \mathbb{I}_{NT} - (\mathbb{I}_N \otimes \kappa\kappa'/T)$. It follows that the sequential projection using $\mathbb{P}_1$ and $\mathbb{P}_2$ is equivalent to the two-way demeaning matrix $\mathbf{Q}$:

$$\begin{aligned}
\mathbb{P}_2\mathbb{P}_1 &= (\mathbb{I}_{NT} - (\iota\iota'/N \otimes \mathbb{I}_T))\,(\mathbb{I}_{NT} - (\mathbb{I}_N \otimes \kappa\kappa'/T)) \\
&= \mathbb{I}_{NT}^2 - (\iota\iota'/N \otimes \mathbb{I}_T) - (\mathbb{I}_N \otimes \kappa\kappa'/T) + (\iota\iota'/N \otimes \mathbb{I}_T)(\mathbb{I}_N \otimes \kappa\kappa'/T) \\
&= \mathbb{I}_{NT} - (\iota\iota'/N \otimes \mathbb{I}_T) - (\mathbb{I}_N \otimes \kappa\kappa'/T) + (\iota\iota'/N \otimes \kappa\kappa'/T) \\
&= \mathbb{I}_N \otimes \mathbb{I}_T - (\iota\iota'/N \otimes \mathbb{I}_T) - (\mathbb{I}_N \otimes \kappa\kappa'/T) + (\iota\iota'/N \otimes \kappa\kappa'/T) = \mathbf{Q}
\end{aligned}$$

The sequential projection onto the fixed effect matrices is numerically equivalent to a two-way within transformation. The bilinear and associative properties of the Kronecker product in the steps above ensure that $\mathbb{P}_1\mathbb{P}_2 = \mathbb{P}_2\mathbb{P}_1 = \mathbf{Q}$, so that the ordering of the within transformations make no difference. Note that $\mathbf{Q}$ is itself a projection matrix, such that $\mathbf{Q}$ is idempotent and symmetric. The estimator can be simplified:

$$\beta = ((\mathbf{QX})'\mathbf{QX})^{-1}(\mathbf{QX})'\mathbf{Qy} = (\mathbf{X}'\mathbf{Q}'\mathbf{QX})^{-1}\mathbf{X}'\mathbf{Q}'\mathbf{Qy} = (\mathbf{X}'\mathbf{QX})^{-1}\mathbf{X}'\mathbf{Qy}$$

The results depend crucially on the panel being balanced. Otherwise, the within transformations become much, much more complicated. The non-uniform structure requires individual and special treatment for each unit in the data set. The block diagonal matrices are of various sizes, and the dummy variable matrices must be tailored to suit the various time intervals. This is not to say that it cannot be done, however, but the demeaning process becomes complicated, circumstantial.

Extra: Some additional insight can be obtained by denoting $\mathbf{A}_k$ as the demeaning matrix of dimension k. Premultiplying $\mathbf{X}$ by $\mathbf{A}_n$ will return a matrix with deviations from column means; this is the standard case that was presented in class. If we apply the appropriately dimensioned $\mathbf{A}$ matrix to the transpose of $\mathbf{X}$, however, we can achieve row means. Note that $\mathbf{A}_k$ is symmetric and idempotent, so that $\mathbf{A}_k\mathbf{A}_k = \mathbf{A}_k$ and $\mathbf{A}_k' = \mathbf{A}_k$. If we wanted to transform the matrix toward deviations from row means, we would post-multiply by $\mathbf{A}_k$: $(\mathbf{A}_k\mathbf{X}')' = \mathbf{X}\mathbf{A}'_k = \mathbf{X}\mathbf{A}_k$. This suggests the form of $\mathbb{P}_1$ versus $\mathbb{P}_2$ above.