Problem 1: Invariant Transformations

The model is $y_{it} = x_{it}'\beta + a_i + u_{it}$. Assume that $E[u_{it}|x_i] = 0$, where $x_i = (x_{i1}, \ldots, x_{iT})$ is a $k \times T$ matrix.

Part A: Long-Differencing

Transforming. Let $T = 10$. Define the following:

$$\bar{y}_{i,5} = \frac{1}{5}\sum_{t=1}^{5} y_{it}, \quad \bar{y}_{i,10} = \frac{1}{10}\sum_{t=6}^{10} y_{it},$$

$$\bar{x}_{i,5} = \frac{1}{5}\sum_{t=1}^{5} x_{it}, \quad \bar{x}_{i,10} = \frac{1}{10}\sum_{t=6}^{10} x_{it},$$

$$\bar{u}_{i,5} = \frac{1}{5}\sum_{t=1}^{5} u_{it}, \quad \bar{u}_{i,10} = \frac{1}{10}\sum_{t=6}^{10} u_{it}.$$

In words, we're looking at the average over the first 5 periods, and the second five periods, for observation i. We want to write down a transformation that yields an equation for $\bar{y}_{i,10} - \bar{y}_{i,5}$. This means that

$$\frac{1}{5}\sum_{t=1}^{5} y_{it} = \frac{1}{5}\sum_{t=1}^{5} x_{it}'\beta + \frac{1}{5}\sum_{t=1}^{5} a_i + \frac{1}{5}\sum_{t=1}^{5} u_{it} \quad \Longrightarrow \quad \bar{y}_{i,5} = \bar{x}_{i,5}'\beta + a_i + \bar{u}_{i,5},$$

$$\frac{1}{5}\sum_{t=6}^{10} y_{it} = \frac{1}{5}\sum_{t=6}^{10} x_{it}'\beta + \frac{1}{5}\sum_{t=1}^{5} a_i + \frac{1}{5}\sum_{t=6}^{10} u_{it} \quad \Longrightarrow \quad \bar{y}_{i,10} = \bar{x}_{i,10}'\beta + a_i + \bar{u}_{i,10}.$$

So subtract and you'll get

$$\bar{y}_{i,10} - \bar{y}_{i,5} = (\bar{x}_{i,10} - \bar{x}_{i,5})'\beta + (\bar{u}_{i,10} - \bar{u}_{i,5}).$$

This transformation is invariant to a_i—we just added up a_i five times and then divided by five.

Consistency. Now let's come up with an estimator for β. For any variable w, define $\Delta\bar{w}_i = \bar{w}_{i,10} - \bar{w}_{i,5}$. Then we can write the transformation as

$$\Delta\bar{y}_i = \Delta\bar{x}_i{}'\beta + \Delta\bar{u}_i.$$

The typical OLS rigmarole will give

$$\hat{\beta} = \left(\frac{1}{n} \sum_{i=1}^{n} \Delta\bar{x}_i \Delta\bar{x}_i{}' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \Delta\bar{x}_i \Delta\bar{y} \quad \Longrightarrow \quad \hat{\beta} = \beta + \left(\frac{1}{n} \sum_{i=1}^{n} \Delta\bar{x}_i \Delta\bar{x}_i{}' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \Delta\bar{x}_i \Delta\bar{u}_i.$$

For consistency, we need the typical stuff, namely

- i.i.d. data,
- $E[\Delta\bar{x}_i \Delta\bar{x}_i{}']$ to be positive definite and finite,
- $E[\Delta\bar{x}_i \Delta\bar{u}_i] = 0$.

Then Khinchine's LLN implies convergence in probability of each sum, and thus Slutsky implies the convergence of the product to zero.

Part B: Individual and Time Effects

Now the model is $y_{it} = x_{it}'\beta + a_i + \lambda_t + u_{it}$. The term λ_t is common to every i, but changes over time. For any variable w_{it}, define $\Delta w_{it} = w_{it} - w_{i,t-1}$. Consider

$$\begin{aligned} y_{it} &= x_{it}'\beta + a_i + \lambda_t + u_{it}, \\ y_{i,t-1} &= x_{i,t-1}'\beta + a_i + \lambda_{t-1} + u_{i,t-1} \\ \Longrightarrow \quad \Delta y_{it} &= \Delta x_{it}'\beta + \Delta\lambda_t + \Delta u_{it}. \end{aligned}$$

We can similarly define

$$\Delta y_{jt} = \Delta x_{jt}'\beta + \Delta\lambda_t + \Delta u_{jt}.$$

Subtract the two and you have

$$\Delta y_{it} - \Delta y_{jt} = (\Delta x_{it} - \Delta x_{jt})'\beta + (\Delta u_{it} - \Delta u_{jt}).$$

Part C: Individual and Cluster-Time Effects

Now the model is $y_{ict} = x_{ict}'\beta + a_{ic} + \lambda_{ct} + u_{ict}$. Do the same thing for Part B, except focus on a particular cluster c. For any variable w_{ict}, define $\Delta w_{ict} = w_{ict} - w_{ic,t-1}$. Consider

$$\begin{aligned} y_{ict} &= x_{ict}'\beta + a_{ic} + \lambda_{ct} + u_{ict}, \\ y_{ic,t-1} &= x_{ic,t-1}'\beta + a_{ic} + \lambda_{c,t-1} + u_{ic,t-1} \\ \Longrightarrow \quad \Delta y_{ict} &= \Delta x_{ict}'\beta + \Delta\lambda_{ct} + \Delta u_{ict}. \end{aligned}$$

We can similarly define

$$\Delta y_{jct} = \Delta x'_{jct}\beta + \Delta\lambda_{ct} + \Delta u_{jct}.$$

Subtract the two and you have

$$\Delta y_{ict} - \Delta y_{jct} = (\Delta x_{ict} - \Delta x_{jct})'\beta + (\Delta u_{ict} - \Delta u_{jct}).$$

Part D: Time Effects vs. Cluster-Time Effects

Some gobbledygook about omitted variables. Don't care.

Part E: Estimations

Time Effects. Notice that when we estimate these, our transformation implicitly includes $t-1$ terms. As such, we have to start our sums from $t=2$ since otherwise we would be summing with some $t=0$ term, which isn't a thing. Furthermore, since our transformation requires $j \neq i$, we always have $j = i-1$ in the summation. This however already requires that we begin the summation at $i=2$. Therefore

$$\hat{\beta} = \left[\sum_{i=2}^{n}\sum_{t=2}^{T}(\Delta x_{it} - \Delta x_{i-1,t})(\Delta x_{it} - \Delta x_{i-1,t})'\right]^{-1}\sum_{i=2}^{n}\sum_{t=2}^{T}(\Delta x_{it} - \Delta x_{i-1,t})(\Delta y_{it} - \Delta y_{i-1,t})'.$$

For consistency, we'll need

- $E[(\Delta x_{it} - \Delta x_{i-1,t})(\Delta x_{it} - \Delta x_{i-1,t})']$ to be full rank,
- $E[(\Delta x_{it} - \Delta x_{i-1,t})(\Delta u_{it} - \Delta u_{i-1,t})'] = 0$.

Strict exogeneity $E[u_{it}|x_i, a_i]$ and i.i.d. gives the latter result.

Cluster Effects. Things are a little bit messier with clusters. Suppose there are C clusters, and n_c individuals in cluster c. Then we'll have

$$\hat{\beta} = \left[\sum_{c=1}^{C}\sum_{i=2}^{n_c}\sum_{t=2}^{T}(\Delta x_{ict} - \Delta x_{i-1,ct})(\Delta x_{ict} - \Delta x_{i-1,ct})'\right]^{-1}\sum_{c=1}^{C}\sum_{i=2}^{n_c}\sum_{t=2}^{T}(\Delta x_{it} - \Delta x_{i-1,t})(\Delta y_{it} - \Delta y_{i-1,t})'.$$

For consistency, we'll need

- $E[(\Delta x_{ict} - \Delta x_{i-1,ct})(\Delta x_{ict} - \Delta x_{i-1,ct})']$ to be full rank,
- $E[(\Delta x_{ict} - \Delta x_{i-1,ct})(\Delta u_{ict} - \Delta u_{i-1,ct})'] = 0$.

Problem 2: Measurement Error in Panel Data

We're using the model $y_{it} = x_{it}'\beta + a_i + u_{it}$. x_{it} is a scalar that we can't observe directly. We instead have a mismeasured version of it,

$$x_{it}^* = x_{it} + \epsilon_{it}.$$

Let $x_i = (x_{i1}, \ldots, x_{iT})$ and $\epsilon_i = (\epsilon_{it}, \ldots, \epsilon_{iT})$. Assume the following:

- $E[a_i|x_i] = 0$,
- $E[\epsilon_{it} x_{it}] = 0$,
- $E[u_{it}\epsilon_{it}] = 0$,
- $E[a_i \epsilon_{it}] \neq 0$.
- $\{y_{it}, x_{it}, u_{it}\}$ are i.i.d.

Part A: Observation Implies Consistency

Let $v_{it} = a_i + u_{it}$. Therefore

$$E[v_{it}x_{it}] = E[a_i x_{it}] + E[u_{it}x_{it}] = 0,$$

because I guess we are also assuming that $E[u_{it}x_{it}] = 0$ even though it's not stated anywhere.[1] This means we have a rather ordinary looking model,

$$y_{it} = x_{it}'\beta + v_{it},$$

with the estimator

$$\hat{\beta} = \beta + \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}^2\right]^{-1} \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}v_{it}.$$

Consistency requires for $E[x_{it}^2]$ to be finite and for $E[x_{it}v_{it}] = 0$. Then the two terms will converge via Khinchine and their product will converge to zero via Slutsky.

[1] I hate this class.

Part B: Mismeasured OLS Estimator

Once we allow for mismeasurement, we the model becomes

$$
\begin{aligned}
y_{it} &= (x_{it}^* - \epsilon_{it})\beta + a_i + u_{it} \\
&= x_{it}^*\beta + a_i + u_{it} - \epsilon_{it}\beta \\
&= x_{it}^*\beta + \tilde{v}_{it}.
\end{aligned}
$$

Then the estimator is

$$\hat{\beta}^* = \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it}^*)^2\right]^{-1}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}x_{it}^* y_{it}.$$

Part C: Mismeasured OLS Probability Limit

Plug in $x_{it}^*\beta + \tilde{v}_{it}$ for y_{it}. You'll end up with

$$\hat{\beta}^* = \beta + \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it}^*)^2\right]^{-1}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}x_{it}^*\tilde{v}_{it}.$$

Supposing that $E[(x_{it}^*)^2]$ is finite, then the probability limit gives

$$\hat{\beta}^* \xrightarrow{p} \beta + E\left[(x_{it}^*)^2\right]^{-1}E[x_{it}^*\tilde{v}_{it}].$$

So let's break this apart a bit, in particular the latter term.

$$
\begin{aligned}
E[x_{it}^*\tilde{v}_{it}] &= E[(x_{it} + \epsilon_{it})(a_i + u_{it} - \epsilon_{it}\beta)] \\
&= E[x_{it}a_i] + E[x_{it}u_{it}] - \beta E[x_{it}\epsilon_{it}] + E[\epsilon_{it}a_i] + E[\epsilon_{it}u_{it}] - \beta E[\epsilon_{it}^2].
\end{aligned}
$$

From our assumptions, we're left with two nonzero terms. In particular, $E[\epsilon_{it}a_i] \neq 0$ and $\beta E[\epsilon_{it}^2] > 0$. The conclusion is that

$$\hat{\beta}^* - \beta = E\left[(x_{it}^*)^2\right]^{-1}\left(E[\epsilon_{it}a_i] - \beta E[\epsilon_{it}^2]\right).$$

It ain't consistent.

Part D: Sign of Asymptotic Bias

The inverted term is clearly positive. But we don't know the sign of $E[\epsilon_{it}a_i]$, nor do we know its magnitude relative to $\beta E[\epsilon_{it}^2]$. So we can only know for sure what the sign of the bias is if both terms end up being of the same sign.

Problem 3: Difference-In-Difference Estimation

The model is $Y_{it} = \beta D_{it} + \alpha_i + \lambda_t + u_{it}$, where D_{it} is a binary treatment variable. Let $T = 2$, and assume that there are two groups: a treatment group where $D_{i1} = 0$ and $D_{i2} = 1$, and the control group where $D_{i1} = 0$ and $D_{i2} = 0$.

Part A: Expected Differences

We want to see what the expected difference in Y is over the two periods between the two groups. First of all,

$$Y_{i2} - Y_{i1} = \beta D_{i2} + \alpha_i + \lambda_2 + u_{i1} - \beta D_{i1} - \alpha_i - \lambda_2 + u_{i1} = \beta(D_{i2} - D_{i1}) + \lambda_1 - \lambda_2.$$

Now we can compare the the two groups by conditioning on the treatment variables. In particular,

$$E[Y_{i2} - Y_{i1} | D_{i1} = 0, D_{i2} = 1] = \beta + \lambda_2 - \lambda_1,$$

$$E[Y_{i2} - Y_{i1} | D_{i1} = 0, D_{i2} = 0] = \lambda_2 - \lambda_1.$$

Solve this for β, giving

$$\beta = E[Y_{i2} - Y_{i1} | D_{i2} = 0, D_{i1} = 1] - E[Y_{i2} - Y_{i1} | D_{i2} = 0, D_{i1} = 0].$$

For an estimator of β, we can use

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (Y_{i2} - Y_{i1}) 1\{D_{i1} = 0, D_{i2} = 1\}}{\sum_{i=1}^{n} 1\{D_{i1} = 0, D_{i2} = 1\}} - \frac{\sum_{i=1}^{n} (Y_{i2} - Y_{i1}) 1\{D_{i1} = 0, D_{i2} = 0\}}{\sum_{i=1}^{n} 1\{D_{i1} = 0, D_{i2} = 0\}},$$

where $1\{D_{i1} = 0, D_{i2} = 0\}$ indicates whether the bracketed event occurs. Since we're assuming that $D_{i1=0}$ for both groups, this really only indicates a difference in the second event. Therefore we can simplify the notation a little and write

$$\hat{\beta} = \frac{\sum_{i=1}^{n} \Delta Y_{i2} \Delta D_{i2}}{\sum_{i=1}^{n} \Delta D_{i2}} - \frac{\sum_{i=1}^{n} \Delta Y_{i2} (1 - \Delta D_{i2})}{\sum_{i=1}^{n} (1 - \Delta D_{i2})}.$$

Part B: First-Difference Transformation

This is a fucking mess and her solutions are as bad as everything else she does. Moving on.

Problem 4: Fixed-Effects Estimation

The model is $y_{it} = x_{it}'\beta + a_i + u_{it}$.

Part A: Estimator

Average the observation i over time to get $\bar{y}_i = \bar{x}_i'\beta + a_i + \bar{u}_i$. Then we can write

$$\begin{aligned} y_{it} - \bar{y}_i &= (x_{it} - \bar{x}_i)'\beta + (u_{it} - \bar{u}_i) \\ \implies \quad \tilde{y}_{it} &= \tilde{x}_{it}\beta + \tilde{u}_{it}. \end{aligned}$$

Now go a step further and "stack" the times into a single matrix. That is, write

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix} = \begin{bmatrix} x_{i1,1} & x_{i1,2} & \dots & x_{i1,k} \\ x_{i2,1} & x_{i2,2} & \dots & x_{i2,k} \\ \vdots & & & \\ x_{iT,1} & x_{iT,2} & \dots & x_{iT,k} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{bmatrix},$$

which we can sum up more concisely as

$$\tilde{y}_i = \tilde{X}_i'\beta + \tilde{u}_i.$$

Now the typical OLS rigmarole leads to

$$\hat{\beta} = \left(\frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i \tilde{y}_i.$$

Part B: Consistency and Asymptotic Normality

The usual manipulations gives

$$\hat{\beta} = \beta + \left(\frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i \tilde{u}_i.$$

Consistency follows if i.i.d., if $E[\tilde{X}_i \tilde{X}_i']$ has full rank, and if $E[\tilde{X}_i \tilde{u}_i] = 0$; we can invoke Khinchine and Slutsky as usual. $E[\tilde{X}_i \tilde{u}_i] = 0$ is actually implied by $E[u_{it}|X_i, a_i]$.

For normality, write

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' \right)^{-1} \frac{\sqrt{n}}{n} \sum_{i=1}^{n} \tilde{X}_i \tilde{u}_i.$$

Evidently we don't need positive definiteness anymore for reasons that are obscure because this class is poorly taught. But as long as we have full rank of the expectation of the first term, and as long as

$$E[\tilde{X}_i \tilde{u}_i \tilde{u}_i' \tilde{X}_i'] < \infty,$$

then we can invoke the Lindeberg-Levy central limit theorem on the other term. Specifically, that term will converge in distribution to

$$\mathcal{N}\left(0, E[\tilde{X}_i \tilde{u}_i \tilde{u}_i' \tilde{X}_i']\right).$$

Part C: Asymptotic Variance

From the above result, we have

$$\text{Avar}\left(\sqrt{n}(\hat{\beta} - \beta)\right) = E\left[\tilde{X}_i \tilde{X}_i'\right]^{-1} E[\tilde{X}_i \tilde{u}_i \tilde{u}_i' \tilde{X}_i'] E\left[\tilde{X}_i \tilde{X}_i'\right]^{-1}.$$

Now let's have some homoskedasticity, i.e. $E[u_i u_i' | X_i, a_i] = \sigma_u^2 I_t$. Hopefully this will have some implications as far as the asymptotic variance goes.

Take it for granted that $E[\tilde{u}_i \tilde{u}_i'] = \sigma_u^2 Q_t$, that Q_t is symmetric and idempotent, and that $\tilde{X}_i = X_i Q_t$. Then we can rewrite the asymptotic variance as

$$\text{Avar}\left(\sqrt{n}(\hat{\beta} - \beta)\right) = E\left[X_i Q_t X_i'\right]^{-1} E[X_i Q_t \tilde{u}_i \tilde{u}_i' Q_t X_i'] E\left[X_i Q_t X_i'\right]^{-1}.$$

For reasons I am quite confused about, we can replace $\tilde{u}_i \tilde{u}_i'$ with $\sigma_u^2 Q_t$. Noting again that $Q_t Q_t = Q_t$, we get

$$\text{Avar}\left(\sqrt{n}(\hat{\beta} - \beta)\right) = E\left[X_i Q_t X_i'\right]^{-1} E[X_i Q_t \sigma_u^2 Q_t X_i'] E\left[X_i Q_t X_i'\right]^{-1}.$$

So suck that sigma out and the asymptotic variance is just

$$\text{Avar}\left(\sqrt{n}(\hat{\beta} - \beta)\right) = \sigma_u^2 E\left[X_i Q_t X_i'\right]^{-1}.$$

Okay then, so what the hell is Q_t? The derivation in the notes is complete and utter garbage—big surprise. Take my word for it that

$$\begin{aligned} Q_t &= \left(I_t - \frac{1}{T}\mathcal{I}_t\right) \\ &= \begin{bmatrix} 1 - \frac{1}{T} & -\frac{1}{T} & \dots & -\frac{1}{T} \\ \vdots & \ddots & \vdots & -\frac{1}{T} \\ -\frac{1}{T} & -\frac{1}{T} & \dots & 1 - \frac{1}{T} \end{bmatrix}. \end{aligned}$$

Part D: Time-Invariant Regressor

A time-invariant regressor means that $x_{it} = x_{is} = \lambda$ for all t, s. The implication is that $\bar{x}_i = \lambda$. And therefore $x_{it} - \lambda = 0$. So for i, we'd have a row of zeroes in $\tilde{X}_i$. Then we can't invert things. No good.

Problem 5: Fixed-Effects and First-Difference

The model is $y_{it} = x_{it}'\beta + a_i + u_{it}$.

Part A: Identical Estimators for Two Periods

The within-group transformation for period 1 yields

$$
\begin{aligned}
& y_{i1} - \frac{1}{2}(y_{i1} + y_{i2}) = \left[x_{i1} - \frac{1}{2}(x_{i1} + x_{i2})\right]' \beta + u_{i1} - \left[\frac{1}{2}(u_{i1} + u_{i2})\right] \\
\implies \quad & \frac{1}{2}(y_{i1} - y_{i2}) = \left[\frac{1}{2}(x_{i1} - x_{i2})\right]' \beta + \left[\frac{1}{2}(u_{i1} - u_{i2})\right] \\
\implies \quad & (y_{i1} - y_{i2}) = \left[(x_{i1} - x_{i2})\right]' \beta + \left[(u_{i1} - u_{i2})\right] \\
\implies \quad & (y_{i2} - y_{i1}) = (x_{i2} - x_{i1})'\beta + (u_{i2} - u_{i1}).
\end{aligned}
$$

The last line is precisely the first-different transformation.

Part B: More than Two Periods

Fixed-Effects. The fixed-effects transformation is

$$
\begin{aligned}
& y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)'\beta + u_{it} - \bar{u}_i \\
\implies \quad & \tilde{y}_i = \tilde{x}_i'\beta + \tilde{u}_i.
\end{aligned}
$$

Then we can do the typical OLS shindig, giving

$$
\hat{\beta}_{FE} = \beta + \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i \tilde{u}_i'
$$

With strict exogeneity, i.e. $E[u_{it}|X_i, a_i] = 0$, the law of iterated expectations gives

$$\begin{aligned}
E[\tilde{X}_i \tilde{u}_i] &= E\left[\sum_{t=1}^{T} \tilde{x}_{it}\tilde{u}_{it}\right] \\
&= E\left[\sum_{t=1}^{T} E[\tilde{x}_{it}\tilde{u}_{it}|X_i, a_i]\right] \\
&= E\left[\sum_{t=1}^{T} \tilde{x}_{it} E[\tilde{u}_{it}|X_i, a_i]\right] \\
&= 0.
\end{aligned}$$

Then the probability limit of the estimator is $\hat{\beta}_{FE} \xrightarrow{p} \beta$.

First-Difference. The first-difference transformation is

$$\begin{aligned}
& y_{it} - y_{i,t-1} = (x_{it} - x_{i,t-1})'\beta + u_{it} - u_{i,t-1} \\
\implies \quad & \Delta y_i = \Delta x_t' \beta + \Delta u_i.
\end{aligned}$$

Throw down the usual OLS stuff again and we get

$$\hat{\beta}_{FD} = \beta + \left(\frac{1}{n}\sum_{i=1}^{n} \Delta X_i \Delta X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \Delta X_i \Delta u_i'.$$

With strict exogeneity, i.e. $E[u_{it}|X_i, a_i] = 0$, the law of iterated expectations gives

$$\begin{aligned}
E[\Delta X_i \Delta u_i] &= E\left[\sum_{t=2}^{T} \Delta x_{it} \Delta u_{it}\right] \\
&= E\left[\sum_{t=2}^{T} E[\Delta x_{it} \Delta u_{it}|X_i, a_i]\right] \\
&= E\left[\sum_{t=2}^{T} \Delta x_{it} E[\Delta u_{it}|X_i, a_i]\right] \\
&= 0.
\end{aligned}$$

Then the probability limit of the estimator is $\hat{\beta}_{FD} \xrightarrow{p} \beta$. Keep in mind that we need to start at $t = 2$ because there is a $t - 1$ implicit in the definition of Δ, and we can't have $t = 0$ entering the calculation because $t = 0$ isn't a thing.

Part C: Failure of Strict Exogeneity

If strict exogeneity fails, then we cannot say that $E[\tilde{X}_i\tilde{u}_i']$ or $E[\Delta X_i \Delta u_i'] = 0$, and therefore we have biased estimators.

Now take my word for it that

$$\sum_{s=1}^{T} x_{is}u_{it} = x_{it}u_{it} \quad \text{and} \quad x_{it}\sum_{s=1}^{T} u_{is} = x_{it}u_{it}.$$

Then in the fixed effects case, we'll have

$$\begin{aligned}
E[\tilde{X}_i\tilde{u}_i] &= E\left[\sum_{t=1}^{T} \tilde{x}_{it}\tilde{u}_{it}\right] \\
&= E\left[\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(u_{it}-\bar{u}_i)\right] \\
&= E\left[\sum_{t=1}^{T} x_{it}u_{it}\right] - E\left[\sum_{t=1}^{T} x_{it}\bar{u}_i\right] - E\left[\sum_{t=1}^{T} \bar{x}_i u_{it}\right] + E\left[\sum_{t=1}^{T} \bar{x}_i\bar{u}_i\right] \\
&= \sum_{t=1}^{T} E[x_{it}u_{it}] - \sum_{t=1}^{T} E\left[x_{it}\frac{1}{T}\sum_{s=1}^{T} u_{is}\right] - \sum_{t=1}^{T} E\left[\frac{1}{T}\sum_{s=1}^{T} x_{is}u_{it}\right] + E\left[\frac{1}{T}\sum_{t=1}^{T} x_{it}\frac{1}{T}\sum_{s=1}^{T} u_{is}\right] \\
&= \sum_{t=1}^{T} E[x_{it}u_{it}] - \frac{1}{T}\sum_{t=1}^{T} E\left[x_{it}u_{it}\right] - \frac{1}{T}\sum_{t=1}^{T} E\left[x_{it}u_{it}\right] + \frac{1}{T}E\left[x_{it}u_{it}\right] \\
&= \sum_{t=1}^{T} E[x_i u_i]\left(1-\frac{1}{T}\right).
\end{aligned}$$

So the estimator converges in probability to

$$\hat{\beta}_{FE} = \beta + E\left[\tilde{X}_i\tilde{X}_i'\right]^{-1}\sum_{t=1}^{T} E[x_i u_i]\left(1-\frac{1}{T}\right) \neq \beta.$$

It ain't consistent.

For the first-difference estimator, we have

$$\begin{aligned}
E[\Delta X_t \Delta u_i] &= E\left[\sum_{t=2}^{T}(x_{it} - x_{i,t-1})(u_{it} - u_{i,t-1})\right] \\
&= E\left[\sum_{t=2}^{T}(x_{it} - x_{i,t-1})(u_{it} - u_{i,t-1})\right] \\
&= E\left[\sum_{t=2}^{T} x_{it}u_{it}\right] - E\left[\sum_{t=2}^{T} x_{it}u_{i,t-1}\right] - E\left[\sum_{t=2}^{T} x_{i,t-1}u_{i,t}\right] + E\left[\sum_{t=2}^{T} x_{i,t-1}u_{i,t-1}\right].
\end{aligned}$$

Evidently the middle two expectations cancel out. Then we can shift the index on the last expectation to write

$$E[\Delta X_t \Delta u_i] = E\left[\sum_{t=2}^{T} x_{it}u_{it}\right] - E\left[\sum_{t=2}^{T} x_{it}u_{i,t-1}\right] - E\left[\sum_{t=2}^{T} x_{i,t-1}u_{i,t}\right] + E\left[\sum_{t=1}^{T-1} x_{i,t}u_{i,t}\right].$$

We'll end up with

$$E[\Delta X_t \Delta u_i] = E[x_{i1}u_{i1}] + 2\sum_{t=2}^{T-1} E[x_{it}u_{it}] + E[x_{iT}u_{iT}].$$

So we get a bunch of junk in there when taking the probability limit. Importantly,

$$\hat{\beta}_{FD} = \beta + \left(\frac{1}{n}\sum_{i=1}^{n}\Delta X_i \Delta X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n}\Delta X_i \Delta u_i' \neq \beta.$$

Problem 6: AutoCorrelation

LOL

Problem 7: Time Series

Part A: Ordinary Least Squares

Consider the model $y_t = x_t'\beta + u_t$. Using OLS, we get the same ol' thing for the estimator, namely,

$$\hat{\beta} = \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} x_t y_t.$$

Substituting in the model for y_t, we get

$$\hat{\beta} = \beta + \left(\frac{1}{T} \sum_{t=1}^{T} x_t x_t' \right)^{-1} \frac{1}{T} \sum_{t=1}^{T} x_t u_t.$$

Consistency. For consistency, we'll need the obvious stuff,

- $E[x_t x_t']$ has full rank,
- $E[x_t u_t] = 0$.

But we'll need two more things to deal with the awkwardness of time series data.

- $\{y_t, x_t\}$ is an ergodic stationary sequence across time.

  Strict stationarity means that the joint distribution of $(z_t, z_{t-1}, \ldots)$ depends only on the lags, not on t. For example, (z_{10}, z_9, z_8) has the same joint distribution as (z_4, z_3, z_2).

  Ergodicity means that as two events are separated by an infinite amount of time, then they are independent. More accurately, for two functions $f(\cdot)$ and $g(\cdot)$,

$$\lim_{n\to\infty} \left| E\left[f(z_t, \ldots, z_{t+k}) g(z_{t+n}, \ldots, z_{t+n+\ell}) \right] \right| = \lim_{n\to\infty} \left| E\left[f(z_t, \ldots, z_{t+k}) \right] \right| \left| E\left[g(z_{t+n}, \ldots, z_{t+n+\ell}) \right] \right|.$$

This gives us consistency via the **ergodic theorem**, which states that if $\{z_t\}$ is an ergodic stationary process with $E[z_t] = \mu$, then

$$\frac{1}{T} \sum_{t=1}^{T} z_t \xrightarrow{p} \mu.$$

Normality. For normality, we need one more condition: $\{x_t u_t\}$ must be a martingale difference sequence with $E[u^2 x_t x_t'] < \infty$.

- z_t is a **martingale** with respect to $(w_{t-1}, w_{t-2}, \ldots, w_1)$ if

$$E[z_t | w_{t-1}, w_{t-2}, \ldots, w_1] = z_{t-1}.$$

- If z_t is a martingale, then

$$E[z_t - z_{t-1} | w_{t-1}, w_{t-2}, \ldots, w_1] = 0.$$

  This is a **martingale difference sequence**. The equation follows because z_{t-1} is included in w_{t-1}, and therefore it evaluates to $z_{t-1} - z_{t-1} = 0$.

Then we can invoke the **central limit theorem for ergodic stationary martingale difference sequences**, which states that if $\{z_t\}$ is a martingale difference sequence that is stationary and ergodic with $E[z_t z_t'] = \Sigma$, then

$$\frac{1}{\sqrt{T}} = \sum_{t=1}^{T} z_t \xrightarrow{d} \mathcal{N}(0, \Sigma).$$

In our case, we have convergence to the (thankfully) unremarkable

$$\mathcal{N}\left(0, E[x_t x_t']^{-1} E[u_t^2 x_t x_t'] E[x_t x_t']^{-1}\right).$$

Part B: Ordinary Least Squares for AR(1) Model

The model is $y_t = \rho y_{t-1} + u_t$, where y_{t-1} is the lagged dependent variable. As long as $|\rho| < 1$, then the series is covariance-stationary and is an ergodic sequence. Then the estimator is

$$\hat{\rho} = \rho + \left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1} y_{t-1}'\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} y_{t-1} u_t.$$

The consistency requirements are exactly what you'd think. The martingale difference sequence is

$$E[u_t | y_{t-1}, \ldots, y_1, u_{t-1}, \ldots, u_t],$$

which via the law of iterated expectations somehow gives

$$E[y_{t-1} u_t | y_{t-2} u_{t-1}, \ldots, y_1 u_2].$$

Then we can get our normal distribution on.

Part C: Nonlinear Least Squares

The model is $y_t = g(x_t; \theta) + u_t$. Adopting the results from the cross-sectional nonlinear last squares, we'd have to make the following changes.

- Instead of $E[u_i x_i]$ as we'd have in a cross section, we'd need $E[u_t | x_t, x_{t-1}, \ldots, x_1]$.
- Instead of i.i.d., we'd need ergodic stationarity. These first two points (along with others that are the same as in the cross-sectional case) give us consistency.
- We need the score,
$$\frac{\partial g(x_t; \theta_0)}{\partial \theta} u_t,$$

to be a matingale difference sequence. This allows us to use the CLT for ergodic stationary martingale difference sequences.