Panel Data

Li Gan

September 2011

Many panel data sets are publicly available. The most well-known ones for economists include Panel Study of Income Dynamics (PSID), Health and Retirement Study (HRS), NLSY (National Longitudinal Study of Youth), and COMPUSTAT. The first three are at household or individual level, and the last one is at firm level.

In China, at least one publicly available panel data sets at household level: CHNS (China Health and Nutrition Survey), administered by University of Carolina's population center.

Collecting panel data sets is typically a lot more expensive per observation than collecting cross section data sets. Compare with the cross section data sets, panel data sets have two distinguishing advantages.

(1) Dynamics. When studying dynamic behavior, using panel data sets is more appropriate than cross section data. Studying dynamics typically requires understanding the evolvement of some stock variables. For example, at household level, how wealth, human capital, and health capital evolve is typically very important in understanding many household behaviors. Studying these stock variables often requires following how these variables evolve over time. In this regard, observing the same households across time is very critical. At the firm level, it is even more important since almost all important factors such as capital stock, inventory stock, human capital, etc are stock variables.

(2) Endogeneity. Having observed the same household or firm multiple times can take care of one important source of endogeneity – the time-invarying unobserved heterogeneity. This can be seen more directly by the following example:

Consider a simple linear model:

$$E(BMI_i\,|x, c) = \alpha_0 + \alpha_0 * calorie\text{-}intake_i + X\beta + c_i. \qquad (1)$$

In (1), c_i represents unobserved factors such as genetics, for example.

If Cov(x,c) = 0, then OLS is consistent. If Cov(x,c) $\neq$ 0, then OLS is NOT consistent. We have two ways to solve this problem:

(i) using a proxy -- *BMI* of siblings, etc.

(ii) using the *IV* approach: $Cov(z, c) = 0$, and $Cov(z, x) \neq 0$, where z may include income, number of people in the family, # of kids in the family, etc.

However, if we observe the same person repeatedly, then more options arise. Suppose y and x are observed at two time periods.

$$E(y_t|X_t, c) = \beta_0 + X_t \beta + c, \qquad t = 1, 2 \tag{2}$$

Again, assume c is consistent (time invariant). Add an error term:

$$E(y_t \mid X_t, c) = \beta_0 + X_t \beta + c + u_t \tag{3}$$

By definition $E(X_t' u_t) = 0$. If we assume $E(X_t' c) = 0$, then OLS produces consistent estimates, otherwise no consistency.

Alternatively, if $E(X_t' c) \neq 0$, take the difference over time of (3):

$$\Delta y_t = \Delta X_t \beta + \Delta u_t .$$

We now can use OLS if: (a) $E(\Delta X_t' \Delta u_t) = 0$
(b) rank of $E(\Delta X_t' \Delta X_t)$ has full rank.

Condition (a) implies:

$$\begin{aligned} E\left(\Delta X_t' \Delta u_t\right) &= E\left(X_2' u_2\right) + E\left(X_1' u_1\right) - E\left(X_2' u_1\right) - E\left(X_1' u_2\right) \\ &= -E\left(X_2' u_1\right) - E\left(X_1' u_2\right) \end{aligned}$$

So $E(X_t' u_t) = 0$ cannot guarantee $E(\Delta X_t' \Delta u_t) = 0$. More conditions are necessary.

Discussions: consider the general panel model:

$$y_{it} = X_{it} \beta + c_i + u_{it}, \quad t = 1, 2, \dots T, \text{ and } i = 1, 2, \dots N. \tag{4}$$

(1) Random effect or fixed effect? (c_i)

Random effect : c_i is random. More precisely, $E(x_{it}' c_i) = 0$.
Fixed effect: c_i is an arbitrary constant. More precisely, $E(X_{it}' c_i) \neq 0$.

If random effect, c_i is drawn independent of x_{it}. Example: $N(0, \sigma_c^2)$.
If fixed effect, c_i can be drawn from a distribution of x_{it}. Example: $N(g(x_{it}), \sigma^2(x_{it}))$.

Note here in the fixed effect model, c_i could be random. Therefore, the difference between the so-called random effect model and the fixed effect model is NOT about randomness of the c_i. The difference is all about if c_i is correlated with x_{it}. More recently, the panel data model is often referred as *unobserved effects model*.

(2) **Strict exogeneity of x_{it}:**

It is possible for y_{it} to affect x_{it} for $s > t$. For example, in the example that studies the effect of debt level on inflation, inflation at t will affect debt level at later years.

Strict exogeneity:

$$E(y_{it}|x_{i1}, x_{i2}, \dots x_{iT}) = E(y_{it}|x_{it}) = x_{it}\beta + c_i$$

This assumption implies:

$$E(u_{it}|x_{i1}, x_{i2}, \dots x_{iT}, c_i) = 0, \text{ or } E(x_{it}' u_{is}) = 0, \text{ for any } i \text{ and } s.$$

This is a necessary condition for estimation.

Failures of the strict exogeneity condition:

Example 1: Cross country data. Study the effect of debt on inflation.

$$CPI_{it} = X_{it}\beta + \gamma Debt_{it} + c_i + u_{it}$$

Problem: past debt level may affect future inflation level.

Example 2: Program evaluation, consider a work-training program, denoted as pwg_{it}:

$$\log(wage_{it}) = \theta_t + z_{it}\gamma + \delta\, pwg_{it} + c_i + u_{it}$$

Problem: participation in the program may not be random. It is possible that:

$$Cov(u_{it}, pwg_{it}) > 0.$$

Those who are more active in the labor are more likely to seek work training program.

Example 3: Distribution lag model:

.

$$patents_{it} = \theta_t + z_{it}\gamma + \sum_{\tau=0}^{k} \delta_\tau RD_{it-\tau} + c_i + u_{it}$$

Is RD_{it-j} correlated with today's u_{it}? A shock in patents may affect the earning ability of the firm, and hence affect future spending in R&D. Cov(u_{it},x_{it+1}) ≠ 0, So strict exogeneity fails.

Example 4: Lagged dependent variable

$$\log(wage_{it}) = \theta \log(wage_{it-1}) + z_{it}\gamma + c_i + u_{it}.$$

In this case, since Cov(log($wage_{it}$),c_{it}) > 0, it must be the case that: Cov(x_{it+1},c_{it}) > 0, since x_{it+1} includes log($wage_{it}$). So the Strict Exogeneity condition fails.

Example 1 again: it may be useful to include lagged dependent variable in the model.

$$CPI_{it} = \sum_{\tau}^{K} \alpha_\tau CPI_{it-\tau} + X_{it}\beta + \gamma Debt_{it} + c_i + u_{it}$$

Estimation: Random Effect Model

The basic model is:

$$y_{it} = x_{it}\beta + c_i + u_{it}$$

*Assumptions RE*1:

(a) $E(u_{it}|x_{it}, c_i) = 0$, where $t=1,\ldots\ldots,T$
(b) $E(c_i|x_{it}) = 0$, where $t=1,\ldots\ldots,T$

Rewrite (1) into: $y_t = X_t \beta + c_i j_T + u_t$, where j_T is the $Tx1$ vector of ones:

$$j_T = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{Tx1} . \qquad \text{Let } v_i = \begin{pmatrix} c_i \\ \vdots \\ c_i \end{pmatrix} + \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix} = \begin{pmatrix} c_i + u_{i1} \\ \vdots \\ c_i + u_{iT} \end{pmatrix}$$

The model becomes: $y_i = x_i\beta + v_i$. Let $\Omega = E(v_i, v_i')$.

Assumption RE2: Rank $E(x_i'\Omega^{-1}x_i)=K$.
Assumption RE3: Homoskedasticity, which can be written as:

$$E(u_i u_i' | x_i, c_i) = \sigma_c^2 I, \text{ and } E(c_i | x_i) = \sigma_c^2.$$

With these two assumptions, we have:

$$E(v_i' v_i) = E\left[\begin{pmatrix} c_i + u_{i1} \\ \vdots \\ c_i + u_{iT} \end{pmatrix}\begin{pmatrix} c_i + u_{i1} & \cdots & c_i + u_{i1} \end{pmatrix}\right]$$

$$= \begin{pmatrix} \sigma_c^2 + \sigma_u^2 & \cdots & \sigma_c^2 \\ \vdots & \ddots & \vdots \\ \sigma_c^2 & \cdots & \sigma_c^2 + \sigma_u^2 \end{pmatrix} \equiv \Omega$$

Therefore, in the model $y_i = x_i\beta + v_i$, the error term v_i is no longer iid. An estimator that can do better than *OLS* is *GLS* (generalized least square).

Consider $y_i = x_i\beta + v_i$, (4.1) or:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}\beta + \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix} \qquad (4.1')$$

The covariance matrix of (4.1) is given by

$$\Lambda = \begin{pmatrix} \Omega & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Omega \end{pmatrix}$$

$$\begin{aligned}
\hat{\beta}_{GLS} &= \left(x'\Lambda^{-1}x\right)^{-1} x'\Lambda^{-1}y \\
&= \left(\left(x_1^{'} \cdots x_N^{'}\right) \begin{pmatrix} \Omega^{-1} & & 0 \\ & \ddots & \\ 0 & & \Omega^{-1} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \right)^{-1} \left(x_1^{'} \cdots x_N^{'}\right) \begin{pmatrix} \Omega^{-1} & & 0 \\ & \ddots & \\ 0 & & \Omega^{-1} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \\
&= \left(\begin{pmatrix} x_1^{'}\Omega^{-1} & \cdots & x_1^{'}\Omega^{-1} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \right)^{-1} \begin{pmatrix} x_1^{'}\Omega^{-1} & \cdots & x_1^{'}\Omega^{-1} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \\
&= \left(\sum_{i=1}^{N} x_i^{'}\Omega^{-1}x_i \right)^{-1} \sum_{i=1}^{N} x_i^{'}\Omega^{-1}y_i \\
&\equiv \hat{\beta}_{RE}
\end{aligned}$$

In this model, Ω is known. However, when Ω is unknown, we could use $\hat{\Omega} \to \Omega$, which requires us to obtain estimates for both σ_c^2 and σ_u^2. Note we have:

$$\sigma_v^{\ 2} = \sigma_c^{\ 2} + \sigma_u^{\ 2} = \frac{1}{T}\sum_{t=1}^{T} E(v_{it}^{\ 2}) \quad \text{for all } i.$$

Therefore, averaging v_{it}^2 across all i and t would give a consistent estimate of σ_v^2:

$$\hat{\sigma}_v^{\ 2} = \frac{1}{NT-K}\sum_{i=1}^{N}\sum_{i=1}^{T} \hat{\hat{v}}_{it}^{\ 2},$$

where $\hat{\hat{v}}_{it}$ is the residual from the OLS regression:

$$\hat{\hat{v}}_{it} = y_{it} - x_{it}\hat{\beta}_{OLS}.$$

Next we need to find a consistent estimator for σ_c^2. Recall that $\sigma_c^2 = E(v_{it}\, v_{is})$ all $t \neq s$. Therefore, for all i, there are $T/(T-1)/2$ non-redundant error products that can be used to estimate σ_c^2:

$$\hat{\sigma}_c^{\ 2} = \frac{1}{\left[NT(T-1)/2-K\right]}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T} \hat{\hat{v}}_{it}\hat{\hat{v}}_{is}$$

is a consistent estimator of σ_c^2. Given $\hat{\sigma}_c^2$ and $\hat{\sigma}_v^2$, then we can obtain $\hat{\sigma}_u^2$:

$$\hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_c^2 \qquad (4.2)$$

Note in practice (4.2) may not be positive. A negative value for $\hat{\sigma}_u^2$ is indicative of a substantial amount of negative serial correlation of u_{it}. If this occurs, we must use heteroscedastic-robust covariance matrix to construct $\hat{\beta}_{RE}$.

Robust variance matrix estimator:

If *Assumption 3* (homoscedastic) doesn't hold, then we must calculate the heteroscedastic-robust covariance matrix. Define the *Tx1* residual vector from the pooled OLS regressions: $\hat{\hat{v}}_i = y_i - x_i\hat{\beta}_{OLS}$, where i = 1, …, N. Define:

$$\hat{\Omega} = \frac{1}{N}\sum_{i=1}^{n}\hat{\hat{v}}_i\hat{\hat{v}}_i' \ .$$

The reason that $\hat{\Omega} \xrightarrow{p} \Omega$ because it is the average over N. The random effect estimator is given by:

$$\hat{\beta}_{RE} = \left(\sum_{i=1}^{N} x_i'\hat{\Omega}^{-1}x_i\right)^{-1}\left(\sum_{i=1}^{N} x_i'\hat{\Omega}^{-1}y_i\right)$$

The necessary condition for the estimator to work is iid across individuals.

Estimation: Fixed-Effect model

Consider the panel model again:

$$y_{it} = x_{it}\beta + c_i + u_{it}, \quad i = 1, 2, \ldots, N;\ t = 1, 2, \ldots T.$$

Assumption FE1: $E(u_{it}|c_i, x_i) = 0$, but $E(c_i|x_{it}) \neq 0$, where $t = 1, \ldots, T$.

One way to estimate the model is by running a regression with a dummy variable for each individual:

$$y_{it} = x_{it}\beta + \sum_{i=1}^{N} c_i D_{it} + v_{it},$$

where $D_{it} = 1$ if ith person; and $D_{it} = 0$ otherwise.

In this case, it is important to note that $\hat{c}_i \xrightarrow{p} c_i$ iff $T \to \infty$. Note this is not the typical assumption of $N \to \infty$. In fact, we often assume T is fixed but $N \to \infty$ for the purpose of robust covariance estimation.

Therefore, typically $\hat{c}_i$ is not a consistent estimator of c_i. However, $\hat{c}_i$ is an unbiased estimator of c_i. In many applications, c_i is a nuisance parameter, so we don't really need to know them.

There are many ways to get rid of c_i before we estimate our model.

Method 1: **Time-Demeaning:**

Take away c_i by taking way the average across i:

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

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)\beta + (u_{it} - \bar{u}_i), \text{ or } \ddot{y}_{it} = \ddot{x}_{it}\beta + \ddot{u}_{it} \qquad (5)$$

Given the strict exogeneity condition, u_{it} is uncorrelated with x_{it} for all t=1,…..,T:

$$E(\ddot{x}_{it}'\ddot{u}_{it}) = E((x_{it} - \bar{x}_i)'(u_{it} - \bar{u}_i)) = 0$$

The OLS estimator of (5) is consistent:

$$\hat{\beta}_{FE} = \left(\sum_{i}^{N} \ddot{x}_i{}' \ddot{x}_i\right)^{-1} \left(\sum_{i}^{N} \ddot{x}_i{}' \ddot{y}_i\right)^{-1}$$

$$= \left(\sum_{i}^{N}\sum_{t}^{T} \ddot{x}_{it}\ddot{x}_{it}\right)^{-1} \left(\sum_{i}^{N}\sum_{t}^{T} \ddot{x}_{it}\ddot{y}_{it}\right)^{-1}$$

Is $\hat{\beta}_{FE}$ efficient asymptotically? Or the question: is $\ddot{u}_{it}$ iid?

For the same individual,

$$E\left(\ddot{u}_{it}{}^2\right) = E\left(u_{it} - \overline{u}_i\right)^2$$

$$= \sigma_u^2\left(1 - \frac{1}{T}\right)$$

$\Rightarrow$ homoscedasticity across individual t. For $t \neq s$, the covariance is:

$$E\left(\ddot{u}_{it}\ddot{u}_{is}\right) = -\frac{\sigma_u^2}{T}$$

Which shows the time demeaned error $\ddot{u}_{it}$ serially correlated. As T gets large, this correlation becomes smaller.

As it can be seen later, however, this serial correlation does not cause any problems. Because asymptotically, it is as if the covariance structure were iid with some minor correlation.

This set of de-meaned equation can be obtained by premultiplying a time demeaning matrix, Q_T, defined as:

$$Q_T \equiv I_T - j_T\left(j_T{}' j_T\right)^{-1} j_T{}'$$

$$= I_T - \frac{1}{T} j_T j_T{}'$$

$$= \begin{pmatrix} 1-\frac{1}{T} & & -\frac{1}{T} \\ & \ddots & \\ -\frac{1}{T} & & 1-\frac{1}{T} \end{pmatrix}$$

Q_T is symmetric, idempotent with rank T-1. We have:

$$Q_T j_T \equiv I_T j_T - j_T\left(j_T{}' j_T\right)^{-1} j_T{}' j_T = 0$$

$$Q_T y_i \equiv I_T y_i - j_T (j_T' j_T)^{-1} j_T' y_i$$
$$= y_i - \frac{1}{T}\begin{pmatrix} 1 & & 1 \\ & \ddots & \\ 1 & & 1 \end{pmatrix} y_i$$
$$= y_i - \bar{y}_i$$

Therefore, we have:

$Q_T y_i = y_i - \bar{y}_i = \ddot{y}_i$. Similarly, $Q_T x_i = x_i - \bar{x}_i = \ddot{x}_i$. $Q_T u_i = u_i - \bar{u}_i = \ddot{u}_i$. Therefore, (5) can be written as:

$$Q_T y_i = Q_T x_i \beta + Q_T u_i \qquad (6)$$

The correlation between the transformed x_i and transformed u_i is given by:

$$\ddot{x}_{it}' \ddot{u}_{it} = (Q_T x_i)' Q_T u_i$$
$$= x_i' Q_T' Q_T u_i$$
$$= x_i' Q_T' u_i = \ddot{x}_{it}' u_i$$

Therefore,

$$\sqrt{N}\left(\hat{\beta}_{FE} - \beta\right) = \left(\frac{1}{N}\sum_i^N \ddot{x}_i' \ddot{x}_i\right)^{-1} \left(\frac{1}{\sqrt{N}}\sum_i^N \ddot{x}_i' \ddot{u}_i\right)$$
$$= \left(\frac{1}{N}\sum_i^N \ddot{x}_i' \ddot{x}_i\right)^{-1} \left(\frac{1}{\sqrt{N}}\sum_i^N \ddot{x}_i' u_i\right)$$

Note:

$E(\ddot{x}_i' u_i) = 0$, and

$$Var(\ddot{x}_i' u_i) = E(\ddot{x}_i' u_i u_i' \ddot{x}_i) = \sigma_u^2 E(\ddot{x}_i' \ddot{x}_i)$$

By *CLT*,

$$\frac{1}{\sqrt{N}}\sum_i^N \ddot{x}_i' u_i \xrightarrow{d} N\left(0, \sigma_u^2 E(\ddot{x}_i' \ddot{x}_i)\right).$$

Then we must have:

$$\sqrt{N}\left(\hat{\beta}_{FE} - \beta\right) \xrightarrow{d} N\left(0, \sigma_u^2 \left(\frac{1}{N}\sum_i^N \ddot{x}_i' \ddot{x}_i\right)^{-1}\right)$$

Therefore, the asymptotic covariance of $\hat{\beta}_{FE}$ is the same as if $\ddot{u}_{it}$ were iid, with one difference:

If *OLS*, $Var\left(\hat{\beta}_{FE}\right) = Var\left(\ddot{u}_{it}\right)\left(\sum_{i}^{N} \ddot{x}_i^{'}\ddot{x}_i\right)^{-1}$.

Note $Var\left(\ddot{u}_{it}\right) \neq \sigma_u^2 = Var\left(u_{it}\right)$. We need to make an adjustment of the variance estimate directly from *OLS*.

Now how to estimate σ_u^2?

Note the error term in *FE* model is $\ddot{u}_{it}$, so we cannot directly use the *SSR*. $\sum_{t=1}^{T} E\left(\ddot{u}_{it}^2\right) = (T-1)\sigma_u^2$ which means $E\left(\ddot{u}_{it}^2\right) \leq \sigma_u^2$. Therefore, directly using OLS variance estimate would yield a smaller variance estimate. In another words, OLS still problematic in calculating variance, but since the difference is only a constant, we can correct it – in this sense that OLS of the demeaned equation does not cause problems.

$$\frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} E\left(u_{it}^2\right) = \sigma_u^2$$

Now define fixed effect residual: $\hat{u}_{it} = \ddot{y}_{it} - \ddot{x}_{it}\hat{\beta}_{FE}$.

A consistent and unbiased estimator of $E\left(\ddot{u}_{it}^2\right)$ is:

$$E\left(\ddot{u}_{it}^2\right) = \frac{1}{NT-K} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2$$

Then a consistent estimator of σ_u^2 is:

$$\hat{\sigma}_u^2 = \frac{1}{N(T-1)-K} \sum_{i=1}^{N} \sum_{t=1}^{T} E\left(\hat{u}_{it}^2\right)$$

.

In the denominator is *N(T-1)-K* instead of *NT-K*. This difference could be substantial if *T* is small.

More complications:

1. Serial correlation: $u_{it} = \rho u_{it-1} + \varepsilon_{it}$

Typically, we simply use the residual to test if there is a serial correlation.

However, it is not simple in this case because the error term after the transformation is $\ddot{u}_{it}$.

Since $E(\ddot{u}_{it}\ddot{u}_{is}) = -\frac{\sigma_u^2}{T}$, for the equation $\ddot{u}_{it} = \delta\ddot{u}_{it-1} + \varepsilon_{it}$, the H_0 of no-serial correlation is: H_0: $\delta = -1/(T-1)$.

To find if there a serial correlation, let $\hat{u}_{it} = \ddot{y}_{it} - \ddot{x}_{it}\hat{\beta}_{FE}$, run a regression of $\hat{u}_{it}$ on $\hat{u}_{it-1}$, $t = 2,,\ldots\ldots T$, and $i=1,\ldots.N$. Test if the coefficient is -1/(T-1). Note the standard error for the coefficient used in the test should be robust standard error.

If we find serial correlation, then we have to use robust variance estimator in the original FE estimator.

$$A\,\mathrm{var}(\hat{\beta}_{FE}) = \left(\sum_{i=1}^{N}\ddot{x}_i'\ddot{x}_i\right)^{-1}\left(\sum_{i=1}^{N}\ddot{x}_i'\hat{u}_i\hat{u}_i'\ddot{x}_i\right)\left(\sum_{i=1}^{N}\ddot{x}_i'\ddot{x}_i\right)^{-1}$$

This robust variance matrix estimator is valid in the presence of heteroskedasticity or serial correlation provided that T is small relation to N.

Method 2: First Differencing

Again, consider the linear panel data model,

$$y_{it} = x_{it}\beta + c_i + u_{it},\ \ i = 1, 2, \ldots, N;\ t = 1, 2, \ldots T. \qquad (6.0)$$

To get rid of the unobserved c_i, take the first differencing, we have:

$$\Delta y_{it} = \Delta x_{it}\beta + \Delta u_{it},. \qquad (6.1)$$

A pooled regression of Δy_{it} on Δx_{it}, denoted as $\hat{\beta}_{FD}$, can yield consistent estimator of β, given the assumptions:

Assumption *FD1*: $E(\Delta x_{it}'\Delta u_{it}) = 0$, $t = 2, 3,\ldots,T$.
Assumption *FD2*: rank $\sum_{t=2}^{T} E(\Delta x_{it}'\Delta x_{it})$ is K. $t = 2, 3,\ldots,T$.

With the two assumptions, the OLS estimate of (6.1) is consistent.

If u_{it} in (6.0) follows random walk, then the OLS estimates of (6.1), $\hat{\beta}_{FD}$, is efficient. If u_{it} in (6.0) is iid, then Δu_{it} in (6.1) will be serially correlated:

$Var(\Delta u_{it}) = 2\sigma_u^2$, and $E(\Delta u_{it}\Delta u_{it-1}) = -\sigma_u^2$.

Therefore, we would adopt robust variance matrix:

$$A\operatorname{var}\left(\hat{\beta}_{FD}\right)=\left(\sum_{i=1}^{N}\Delta x_i^{'}\Delta x_i\right)^{-1}\left(\sum_{i=1}^{N}\Delta x_i^{'}\Delta\hat{u}_i\Delta\hat{u}_i{}'\,\Delta x_i\right)\left(\sum_{i=1}^{N}\Delta x_i^{'}\Delta x_i\right)^{-1}$$

where $\Delta\hat{u}_{it} = \Delta y_{it} - \Delta x_{it}\hat{\beta}_{FD}$.

Discussions: serial correlation of u_{it}.

Again, when u_{it} are iid, then a regression of the following model should yield:

$$\Delta u_{iT} = \rho\Delta u_{iT-1} + \varepsilon_{it}.$$

Testing if there is a serial correlation is equivalent to testing if H_0: $\rho=0.5$.
Note here that we only use the observation at T and T-1 for Δu_{iT} and T-1 and T-2 for Δu_{iT-1}. We are not use information from other years to avoid the unnecessary panel data issue.

It is possible that Δu_{it} is actually iid. This occurs when u_{it} is random walk.

Comparing $\hat{\beta}_{FE}$ **and** $\hat{\beta}_{FD}$***:***

Depends on error structure: If u_{it} is serially uncorrelated, the *FE* is more efficient. However, when u_{it} is a random walk, then *FD* is more efficient.

It is important to note that when there is a strong exogeneity, the difference between *FD* and *FE* is only the sampling error. Therefore, one can use Hausman-type test (to compare the difference between the *FE* estimates and *FD* estimates).

Prefer FD: 1. easier to estimate

Prefer FE: 1. more efficient (*FD* losses *N* observations)

Test of strict exogeneity: when T > 2, we can use the FE method.

$$y_{it} = x_{it}\beta + w_{it+1}\delta + c_i + u_{it},$$

where w_{it+1} is a subset of x_{it+1}. Testing $\delta=0$ is a test for strict exogeneity.

Test of strict exogeneity when T = 2. In this case, we may have to use FD method only.

$$\Delta y_{it} = \Delta x_{it}\beta + w_t\delta + \Delta u_{it},$$

where w_t is a subset of x_t. Again, testing $\delta=0$ is a test for strict exogeneity.

Comparing $\hat{\beta}_{FE}$ **and** $\hat{\beta}_{RE}$

In some situations, it is often only possible to apply random effect model. The key variable we are interested in have no variation or very little variation over time.

Example: we are interested in how the size of the local labor market affect local the unemployment rate (Gan and Zhang, *Journal of Econometrics*, 2006, page 127-152.). We have panel data of 295 monthly city unemployment rates.

$$\textit{unemployment rate}_{ct} = \alpha_c + X_{ct}\beta + \gamma \log(\textit{size}_c) + u_{ct}$$

where X_{ct} includes: unemployment benefit, measurement of industry composition, percentage of people who are young (youth share), net migration rates, log of square miles of area (not changing over time), and finally the log of the size of the market (average employment), which does not change overtime.

There is a model developed in the paper that argues that a larger size of the market would yield a lower unemployment.

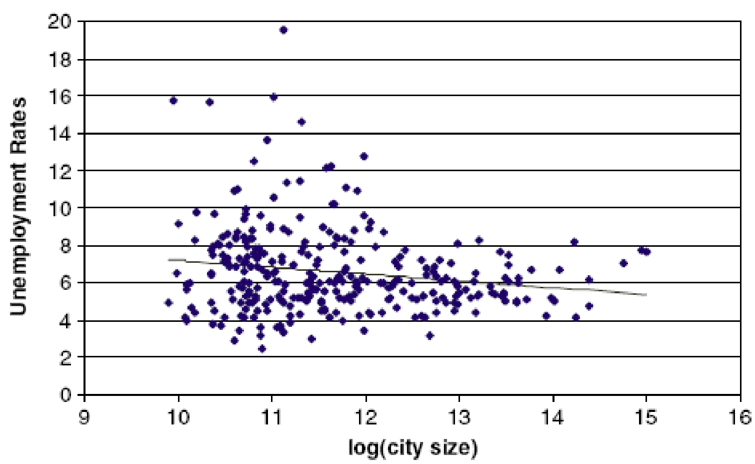

Fig. 2. Logarithm of city size and mean unemployment rates.

The paper finds that an increase in two standard deviation of the city size would result a decrease of 0.15 percentage points of unemployment rate.

Table 6
Unemployment rate mean regression results

Variables	(1)	(2)
time fixed effect	Yes	Yes
city random effect	Yes	Yes
constant	0.485	1.423
	(0.520)	(0.539)
Lagged unemployment rate	0.874	0.873
	(0.0020)	(0.0021)
INDCOM	−18.64	−18.66
	(1.033)	(1.033)
RISK	10.61	1.467
	(5.69)	(5.88)
INDCOM × RISK	360.8	424.9
	(163.8)	(163.4)
unemployment benefit	1.058	1.019
	(0.128)	(0.127)
youth share	−1.765	−2.147
	(0.528)	(0.531)
mean net migration rate	−0.0051	0.0080
	(0.011)	(0.011)
log(miles^2)	−0.141	−0.203
	(0.140)	(0.139)
[log(*miles*^{2})]^2	0.0105	0.0179
	(0.0095)	(0.0095)
log(size)		−0.0752
		(0.0129)
R^2	0.915	0.916
No. of obs.	50439	50439

Standard errors are in parentheses.

In this example, T is relatively large. Given that, we can show next that $\hat{\beta}_{FE}$ and $\hat{\beta}_{RE}$ are close to each other.

In the case of varying x_{it}, we can compare $\hat{\beta}_{FE}$ and $\hat{\beta}_{RE}$. These two estimators do exhibit some interesting relationships.

Again, define:

$$Q_T \equiv I_T - j_T\left(j_T' j_T\right)^{-1} j_T' = I_T - P_T$$

$$= I_T - \frac{1}{T} j_T j_T'$$

$$= \begin{pmatrix} 1-\frac{1}{T} & & -1 \\ & \ddots & \\ -1 & & 1-\frac{1}{T} \end{pmatrix}$$

Then:

$$Q_T j_T \equiv I_T j_T - j_T\left(j_T' j_T\right)^{-1} j_T' j_T = 0$$

For *FE* model, the transformed equation can be written as:

$$Q_T y_i = Q_T x_i \beta + Q_T u_i$$

For *RE* model:

$$\begin{aligned}\Omega &= \sigma_u^2 I_T + \sigma_c^2 j_T j_T' \\ &= \sigma_u^2 I_T + T\sigma_c^2 j_T\left(j_T j_T'\right)^{-1} j_T' \\ &= \sigma_u^2\left(P_T + Q_T\right) + T\sigma_c^2 P_T \\ &= \left(\sigma_u^2 + T\sigma_c^2\right)P_T + \sigma_u^2 Q_T \\ &= \left(\sigma_u^2 + T\sigma_c^2\right)\left(P_T + \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_c^2} Q_T\right) \\ &= \left(\sigma_u^2 + T\sigma_c^2\right)\left(P_T + \eta Q_T\right)\end{aligned}$$

where $\eta = \dfrac{\sigma_u^2}{\sigma_u^2 + T\sigma_c^2}$.

Define $S_T(\eta) \equiv P_T + \eta Q_T$. Note we have: $P_T P_T = I$, and $P_T Q_T = 0$, $P_T = I_T - Q_T$.
It is easy to show that $S_T^{-1} = P_T + \dfrac{1}{\eta} Q_T$ by showing $S_T S_T^{-1} = I$.
Again, one can also show that $S_T^{-1/2} = P_T + \dfrac{1}{\sqrt{\eta}} Q_T$, since $S_T^{-1/2} S_T^{-1/2} = S_T^{-1}$.

Given that, $\Omega = \left(\sigma_u^2 + T\sigma_c^2\right)\left(P_T + \eta Q_T\right)$ and the previous equation, we have:

$$\Omega^{-1/2} = \left(\sigma_u^2 + T\sigma_c^2\right)^{-1/2}\left(P_T + \frac{1}{\sqrt{\eta}} Q_T\right).$$

Again, define $\lambda = 1 - \sqrt{\eta} = 1 - \dfrac{\sigma_u}{\left(\sigma_u^2 + T\sigma_c^2\right)^{1/2}}$. Then,

$$
\begin{aligned}
\Omega^{-1/2} &= \frac{1-\lambda}{\sigma_u}\left(P_T + \frac{1}{1-\lambda}Q_T\right)\\
&= \frac{1}{\sigma_u}\left(P_T - \lambda P_T + Q_T\right)\\
&= \frac{1}{\sigma_u}\left(I_T - \lambda P_T\right)
\end{aligned}
$$

Further, define $G_T = I_T - \lambda P_T$. Consider a transformation of

$$G_T y_i = G_T x_i \beta + G_T v_i \qquad (7)$$

It is easy to verify that the error term in (6) is iid:

$$E\left(G_T v_i v_i^{'} G_T\right) = \frac{1}{\sigma_u^2}\Omega^{-1}E\left(v_i v_i^{'}\right) = \frac{1}{\sigma_u^2}$$

Therefore, OLS of (6) yields an estimate:

$$
\begin{aligned}
\hat{\beta}_{OLS} &= \left(\sum_{i=1}^{N} x_i^{'} G_T G_T x_i\right)^{-1} \sum_{i=1}^{N} x_i^{'} G_T G_T y_i\\
&= \left(\frac{1}{\sigma_u^2}\sum_{i=1}^{N} x_i \Omega^{-1} x_i^{'}\right)^{-1} \frac{1}{\sigma_u^2}\sum_{i=1}^{N} x_i^{'} \Omega^{-1} y_i\\
&= \left(\sum_{i=1}^{N} x_i \Omega^{-1} x_i^{'}\right)^{-1} \sum_{i=1}^{N} x_i^{'} \Omega^{-1} y_i\\
&= \hat{\beta}_{RE}
\end{aligned}
$$

Rewrite equations (6) and (7) here:

$$Q_T y_i = Q_T x_i \beta + Q_T v_i \qquad (6)$$

$$G_T y_i = G_T x_i \beta + G_T v_i \qquad (7)$$

where $G_T = I_T - \lambda P_T$ and $Q_T = I_T - P_T$. The difference between the two transformation is:

$$\lambda = 1 - \frac{\sigma_u}{\left(\sigma_u^2 + T\sigma_c^2\right)^{1/2}}.$$

As T is large or σ_c/σ_u is large, $\lambda \to 1$, random effect model and the fixed effect is close to each other. Note whether T is large is not related to N. It has nothing to do with the N.

STATA Deviation

Simple command for panel data model:

(1) xtreg

Example:

xtreg ln_w grade age* ttl_exp ttl_exp2 tenure tenure2 black not_smsa south, fe i(idcode)

fe – fixed effect model
"re" -- random effect model GLS
"be" – between-effects model
"mle" – random effect MLE

i(idcode), specifies the variable name that contains the unit to which the observation belongs

(2) Given the model:

$$y_{it} = x_{it}\beta + c_i + u_{it}$$

We can allow u_{it} to be serially correlated: $u_{it} = \rho u_{it-1} + \varepsilon_{it}$

Since the model has a time series aspect, we need to tsset the data.

tsset
xtregar ln_w grade age* ttl_exp ttl_exp2 tenure tenure2 black not_smsa south, fe

Dynamic Panel Data Models

Or unobserved Effect models without the strict exogeneity assumption

Strict exogeneity:

$$E(y_{it}|x_{i1}, x_{i2}, \dots x_{iT}) = E(y_{it}|x_{it}) = x_{it}\beta + c_i$$

Once x_{it} and c_i are controlled for, x_{is} has no partial effect on y_{it} for $s \neq t$. In this case, x_{it} is strictly exogenous, conditional on the unobserved effect c_i.

Another way to state the strict exogeneity is: $E(u_{it}|x_{i1}, x_{i2}, \dots x_{iT}) = 0 \Rightarrow E(x_{it}'u_{is})=0$.

This assumption is much stronger than $E(x_{it}'u_{it})=0$.

How to take care this type of problem, allowing u_{it} to be correlated with future values of x_{it}, i.e., $x_{it+1}, x_{it+2}, \dots, x_{iT}$, but not with past x_{it}, i.e., $x_{it-1}, x_{it-2}, \dots x_{i1}$.

Example 11.2: (lagged dependent variable) Static model with feedback:

We are interested in the how the spread of *HIV* affect condoms sales.

$$y_{it} = z_{it}\gamma + \rho y_{it-1} + w_{it}\beta + c_i + u_{it}$$

y_{it} is per capita condom sales, w_{it} measures the HIV spread (HIV infection rate, for example); $\rho \neq 0$, then state dependence; z_{it} is strictly exogenous and finally w_{it} is sequentially exogenous.

$$E(u_{it}|z_{it}, w_{it}, \dots w_{i1}, c_i) = 0.$$

However, w_{it} (*HIV* spread) is influenced by past y_{it} (condom sales).

$$w_{it} = z_{it}\delta + \rho_1 y_{it-1} + c_i + r_{it}$$

Example: patents and R&D spending. Consider a model,

$$patents_{it} = \theta_t + z_{it}\gamma + \sum_{\tau=0}^{k} \delta_\tau RD_{it-\tau} + c_i + u_{it}$$

Is RD_{it-j} correlated with today's u_{it}? A shock in patents may affect the earning ability of the firm, and hence affect future spending in R&D. $Cov(u_{it}, x_{it+1}) \neq 0$, So strict exogeneity fails.

However, it is still possible to estimate such a model without outside instrument variables, as long as we have the so-called sequential exogeneity, i.e, Cov(x_{it}, u_{is}) = 0 for all $t \leq s$. All previous models can be considered as one version of the sequential exogeneity.

It is necessary to recognize here that u_{it} cannot be affect by any previous z_{it} or RD_{it}. It has to be random in the sense that it is the "act of God." If we allow u_{it} to be affected by previous z_{it} or RD_{it}, the basic causality that the model implies is wrong.

What happens if the strict exogeneity fails? Both FE and FD estimators are inconsistent and biased.

For the *FE* estimator:

$$p\lim\left(\hat{\beta}_{FE}\right) = \beta + \left[\frac{1}{T}\sum_t E\left(\ddot{x}_{it}'\ddot{x}_{it}\right)\right]^{-1}\left[\frac{1}{T}\sum_t E\left(\ddot{x}_{it}'u_{it}\right)\right]$$

Under sequential exogeneity, it is possible to bound the bias:

$$E\left(\ddot{x}_{it}'u_{it}\right) = E\left[\left(x_{it} - \bar{x}_{it}\right)'u_{it}\right] = -E\left(\bar{x}_{it}'u_{it}\right) \qquad \text{(since } E\left(x_{it}'u_{it}\right) = 0\text{)}$$

Therefore,

$$\frac{1}{T}\sum_{t=1}^{T} E\left(\ddot{x}_{it}'u_{it}\right) = -\frac{1}{T}\sum_{t=1}^{T} E\left(\bar{x}_i'u_{it}\right) = -E\left(\bar{x}_i'\bar{u}_i\right)$$

To bound the inconsistency, $\text{var}\left(\bar{x}_i\right)$ and $\text{var}\left(\bar{u}_i\right)$ are of order $\frac{1}{T}$.

By Cauchy-Schwartz inequality

$$E\left(\bar{x}_i'\bar{u}_i\right) \leq \left[\text{var}\left(\bar{x}_i\right)\text{var}\left(\bar{u}_i\right)\right]^{1/2} = O\left(T^{-1}\right)$$

So when T is large, the bias could be small. The key condition for this to hold is that $|\rho_1|<1$.

For the *FD* estimator: When $T > 2$:

1st differencing: $\Delta y_{it} = \Delta x_{it}\beta + \Delta u_{it}$,. $t = 2, 3, \ldots, T$

$$p\lim\hat{\beta}_{FD} = \beta + \left[\frac{1}{T}\sum_{t=1}^{T} E\left(\Delta x_{it}'\Delta x_{it}\right)\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T} E\left(\Delta x_{it}'\Delta u_{it}\right)\right]$$

in which E($\Delta x_{it}'\Delta u_{it}$)= - E($x_{it}'u_{it-1}$) $\neq$ 0. Therefore, the *FD* estimate is biased and inconsistent. Note this bias does not depend on *T*.

How to estimate this type of models? *GMM*.

Again, 1st differencing to remove c_i:

$$\Delta y_{it} = \Delta x_{it}\beta + \Delta u_{it},. \quad t = 2, 3, \ldots, T$$

Now under the sequential exogeneity assumption ($s \leq t$):

$$E(x_{is}' u_{it}) = 0, \ s = 1,2,\ldots, t$$
$$\Rightarrow E(x_{is}' \Delta u_{it}) = 0, \ s = 1, 2, \ldots, t-1$$

So at time t we can use x_{it-1}^o as potential instruments for Δx_{it}, where

$$x_{it}^o \equiv (x_{i1}, x_{i2}, \ldots, x_{it})$$

Obviously, for Δx_{it}, it is likely that x_{it-1} would be correlated with Δx_{it}. However, when the sequential exogeneity condition implies $E(x_{is}' u_{it}) = 0$ when $s < t$, only the set of variables x_{it-2}^o can be instrumental variables for Δx_{it}.

Various types of models with the sequential exogeneity would imply different sets of instruments. These sets of instruments differ on the number of lags necessary to ensure that the instruments and errors are uncorrelated. When the number lags increases, it is less likely that the instrument variables are correlated with Δx_{it}.

To estimate such a model, we have to apply GMM.

General Method of Moments (GMM)

$$y_i = x_t\beta + u_i, \quad \text{with } E(x_i' u_i) \neq 0$$

Assumption 1: $E(z_i' u_i) = 0$
Assumption 2: $E(z_i' x_i) \neq 0$

So the moment conditions are: $E(z_i' u_i) = E(z_i' (y_i - x_t\beta)) = 0$

Find β, such that:

$$\min_{\beta} \left(\sum_{i=1}^{N} z_i' (y_i - x_i\beta) \right)' W \left(\sum_{i=1}^{N} z_i' (y_i - x_i\beta) \right)$$

The solution to this problem:

$$\hat{\beta}_{MM} = (X'ZWZ'X)^{-1} X'ZWZ'Y$$
$$= \beta + (X'ZWZ'X)^{-1} X'ZWZ'u,$$

and the covariance of the estimator:

$$Var(\hat{\beta}_{MM}) = (X'ZWZ'X)^{-1} X'ZW\Lambda WZ'X(X'ZWZ'X)^{-1}$$

where Λ = E($z_i' u_i u_i'$). This is a very long but intuitive covariance matrix.

The next step is to find the optimal weighting matrix *W*, which is the *GMM*.

When $W = \Lambda^{-1}$, then the optimal covariance matrix is reached. In this case,

$$\hat{\beta}_{GMM} = (X'Z\Lambda^{-1}Z'X)^{-1} X'Z\Lambda^{-1}Z'Y \text{, and}$$

$$Var(\hat{\beta}_{GMM}) = (X'Z\Lambda^{-1}Z'X)^{-1} X'Z\Lambda^{-1}\Lambda\Lambda^{-1}Z'X(X'Z\Lambda^{-1}Z'X)^{-1}$$
$$= (X'Z\Lambda^{-1}Z'X)^{-1} X'Z\Lambda^{-1}Z'X(X'Z\Lambda^{-1}Z'X)^{-1}$$
$$= (X'Z\Lambda^{-1}Z'X)^{-1}$$

How to do it? 2SLS

(1) A regression of Δx_{it} on x_{it-1}^o
(2) Use $\Delta\hat{x}_{it}$ in previous equation.

The potential problem of this is that Δx_{it} on x_{it-1}^o could be poorly identified.

The residual of 2SLS could be used to construct optimal weighting matrix – *GMM*. Suppose using all x_{it-1}^o as *IV*s. The first-differenced model is:

$$\Delta y_i = \Delta x_i\beta + \Delta u_i.$$

Define $z_i = \begin{pmatrix} x_{i1}^o & & 0 \\ & \ddots & \\ 0 & & x_{iT-1}^o \end{pmatrix}$

Efficient GMM estimator:

(1) Consistent estimator (2SLS):

$$\Delta\hat{\hat{u}}_i = \Delta y_i - \Delta x_i \hat{\hat{\beta}}_{2SLS}$$

(2) Optimal weighting matrix:

$$\hat{W} = \left(\frac{1}{N} \sum_{i=1}^{N} x_i^o{}' \Delta\hat{\hat{u}}_i \Delta\hat{\hat{u}}_i{}' x_i^o \right)^{-1}$$

Finally, the GMM estimate:

$$(3)\ \hat{\beta}_{GMM} = \left(\frac{1}{N} \sum_{i}^{N} x_i^o{}' x_i \hat{W} x_i{}' x_i^o \right)^{-1} \left(\frac{1}{N} \sum_{i}^{N} x_i^o{}' x_i \hat{W}_i x_i^o y_i \right)$$

$$A\text{var}\sqrt{N}\hat{\beta}_{GMM} = \left(X' Z \hat{W} Z' X \right)^{-1}$$

If $E(z_i' u_i u_i' z_i) = E(z_i' \Omega z_i)$, the GMM is equivalent to 3SLS.

Different Types of Models with the Sequential Exogeneity

Model 1: $y_{it} = \rho y_{it-1} + c_i + u_{it}$.

$E(u_{it}| y_{it-1}, \ldots, y_{i1}, c_i) = 0$

Differencing: $\Delta y_{it} = \rho\Delta y_{it-1} + \Delta u_{it}$, $t \geq 2$.

Obviously the strict exogeneity fails: $\text{Cov}(\Delta y_{it-1}, \Delta u_{it}) = \text{Cov}(y_{it-1}, -u_{it-1}) \neq 0$.

Note that $y_{it-2}, y_{it-3}, \ldots, y_{i1}$ are uncorrelated with Δu_{it}. Therefore, in this case, the set of instruments for Δy_{it} is: $\{y_{it-2}, y_{it-3}, \ldots, y_{i1}\}$.

Example: Testing for persistency in county crime rates

$$\log(\textit{crime rate}_{it}) = \alpha + \rho\log(\textit{crime rate}_{it-1}) + c_i + u_{it.}$$

First differencing:

$$\Delta\log(\textit{crime rate}_{it}) = \rho\Delta\log(\textit{crime rate}_{it-1}) + \Delta u_{it.}$$

IVs are: $\log(\textit{crime rate}_{it-2})$, $\log(\textit{crime rate}_{it-3})$. The first stage regression is:

$\Delta\log(\textit{crime rate}_{it-1})$ on $\log(\textit{crime rate}_{it-2})$, $\log(\textit{crime rate}_{it-3})$.

This regression yields a p-value of .023 (which means F is large).

However, the estimated ρ turns out to be not significant from zero.

Model 2: $y_{it} = \rho y_{it-1} + c_i + u_{it}$, and $u_{it} = a u_{it-1} + \varepsilon_{it}$.

Error term u_{it} is AR(1).

Plug the error equation into the original equation:

$$y_{it} = \rho y_{it-1} + c_i + a u_{it-1} + \varepsilon_{it}.$$

Note this implies that Cov(y_{it}, u_{it-1}) ≠ 0. Take the first difference of the original equation (we need to work with the original equation first):

$$\Delta y_{it} = \rho \Delta y_{it-1} + \Delta u_{it}$$

Note Cov($y_{it-2}, \Delta u_{it}$) = Cov($y_{it-2}, -u_{it-1}$) ≠ 0. Therefore, the instruments for Δy_{it-1} have to be $y_{it-3}, y_{it-4}, \ldots, y_{i1}$.

Model 3: When both exogenous and endogenous variables are present

Consider a model:

$$y_{it} = z_{it}\gamma + w_{it}\delta + c_i + u_{it}.$$

E($z_{it}'u_{is}$) = 0, for any t and s, strict exogeneity.

E($w_{it}'u_{is}$) = 0, if $t \leq s$, sequential exogeneity.

First differencing: $\Delta y_{it} = \Delta z_{it}\gamma + \Delta w_{it}\delta + \Delta u_{it}$.

Note that Δu_{it} are uncorrelated with ($w_{it-1}, w_{it-1}, \ldots, w_{i1}$) and z_{is} for any s.

IVs available at time t are ($z_{is}, w_{it-1}, w_{it-2}, \ldots, w_{i1}$) for any s.

Model 4: contemporaneous correlation between the error and some explanatory variables

$$y_{it} = z_{it}\gamma + w_{it}\delta + c_i + u_{it}.$$

E($z_{it}'u_{is}$) = 0, for any t and s, strict exogeneity.

E($w_{it}'u_{is}$) = 0, if $t < s$, sequential exogeneity.

E($w_{it}'u_{it}$) ≠ 0, contemporaneous correlation.

Again, first differencing: $\Delta y_{it} = \Delta z_{it}\gamma + \Delta w_{it}\delta + \Delta u_{it}$.

Note that Δu_{it} are uncorrelated with ($w_{it-2}, w_{it-3}, \ldots, w_{i1}$) and z_{is} for any s.

IVs available at time t are ($z_{is}, w_{it-2}, w_{it-3}, \ldots, w_{i1}$) for any s.

Example: the effect of cigarette smoking on wage

$$\log(wage_{it}) = z_{it}\gamma + \delta_1 cigs_{it} + c_i + u_{it}.$$

If $cigs_{it}$ and $income_{it}$ are correlated, and $income_{it}$ and $wage_{it}$ are correlated, then $cigs_{it}$ and u_{it} are correlated.

IVs in this case would be ($cigs_{it-2}, cigs_{it-3}, \ldots, cigs_{i1}, z_{is}$)

Example: effect of prison population on crime rates.

$$\log(crime_{it}) = \theta_t + \beta_1 \log(prison_{it}) + x_{it}\gamma + c_i + u_{it}$$

Differencing:

$$\Delta\log(crime_{it}) = \xi_t + \beta_1 \Delta\log(prison_{it}) + \Delta x_{it}\gamma + \Delta u_{it}$$

Simultaneity between $\Delta\log(crime_{it})$ and $\Delta\log(prison_{it})$ makes estimation inconsistent. *IV*s for $\Delta\log(prison_{it})$

(1) whether a final decision was reached on overcrowding litigation in the current year.
(2) whether a final decision was reached in the previous two years.

Litigations have several stages
pre-filing
filing
preliminary decision by the court
final decision
further action
release from these restrictions

States have overcrowding litigations in any given year, states where such litigation occurred might be in any on of these stages.

OLS estimate: β_1 -0.181(SE = 0.048)
2SLS estimate: β_1 -1.032(SE = 0.37)

Example: we are interested in how per student spending would affect average scores of a student.

$$avgscore_{it} = \theta_t + \delta_1 spending_{it} + z_{it}\gamma + c_i + u_{it}$$

Note in this model, z_{it} should contain average family income for school *i* at time *t*. However, we are unable to collect such data. Therefore, $income_{it}$ are missing but they are correlated with spending.

1^{st} differencing :

$$\Delta avgscore_{it} = \theta_t + \delta_1 \Delta spending_{it} + \Delta z_{it}\gamma + \Delta u_{it}$$

We need *IV*s for $\Delta spending_{it}$. In this model, using lagged spending may not be a good idea since spending may affect *avgscore* with a lag – past spending may affect current *avgscore*. Therefore, one cannot use the lagged variables as IVs. Additional IVs are necessary.

Possibility for additional IVs: use exogenous changes in property tax that arise because of an unexpected change in tax laws (Such changes occur in California in 1978 and in Michigan 1994)

Example (Gan and Zhang, 2007, NBER Working Paper. The results are not reported in the paper): the effect of market size on housing market transactions, using Texas city-year data. We are interested in how the market size, characterized by the log of population size, $lpop_{ct}$, affect the log of the average housing price $lprice_{ct}$, and log of the month in the market to sell ($inventory_{ct}$). We are also interested in how the exogenous shocks, such as unemployment rate, $urate_{ct}$, affect $lprice_{ct}$ and $inventory_{ct}$.

We try three different specifications. The key paramaeter is the coefficient on $lpop_{ct}$. The prediction of the model is that coefficient is positive for lpricect.

$lprice_{ct} = lprice_{ct-1} + interest\text{-}rate_t + points_t + urate_{ct} + lpop_{ct}$

$lprice_{ct-1}$	$interest\text{-}rate_t$	$points_t$	$urate_{ct}$	$lpop_{ct}$	
	-.035 (.0055)	-.257 (.0127)	-.0106 (.0022)	.335 (.0426)	FE
	.0151 (.0072)	-.157 (.018)	-.0103 (.0038)	.942 (.0115)	FE & AR(1) (rho = .705)
.431 (.0061)	-.0021 (.00020)	-.00562 (.0014)	-.0011 (.0002)	.0209 (.0089)	Dynamic panel (rho .0209 (.0089))

$inventory_{ct} = inventory_{ct-1} + interest\text{-}rate_t + points_t + urate_{ct} + lpop_{ct}$

$inventory_{ct-1}$	$interest\text{-}rate_t$	$points_t$	$urate_{ct}$	$lpop_{ct}$	
	1.309 (.194)	-.0377 (.453)	.633 (.090)	-2.424 (1.509)	FE
	.0341 (.146)	1.173 (.423)	.223 (.086)	.363 (.269)	FE & AR(1) (rho = .803)
.0651 (.0212)	.608 (.029)	1.748 (.101)	.447 (.032)	.509 (1.172)	Dynamic

The three different specifications yield very different estimates.

Model 5: Measurement error induced endogeneity

Consider the classic measurement error model:

$y_{it} = x_{it}^{*}\beta + c_i + u_{it}, \quad t = 1, 2, \ldots T,$ and $i = 1, 2, \ldots N.$

where E(u_{it} |x_t^{*}, x_i, c_i), strict exogeniety. Let

$x_{it} = x_{it}^{*} + r_{it},$ where x_{it}^{*} and r_{it} are uncorrelated.

Pooled OLS:

$$\hat{\beta}_{POLS} = \beta + \frac{Cov(x_{it}, c_i + u_{it} - \beta r_{it})}{Var(x_{it})}$$
$$= \beta + \frac{Cov(x_{it}, c_i) - \beta\sigma_r^2}{Var(x_{it})}$$

where $\sigma_r^2 = \text{var}(r_{it}) = \text{cov}(x_{it}, r_{it})$. From previous equation, there are two sources of asymptotic bias: correlation between c_i and x_{it}, and the measurement bias. If x_{it} and c_i are positively correlated, then two source of bias tend to cancel each other out.

Now consider the first differencing: $\Delta y_{it} = \Delta x_{it}^{*}\beta + \Delta u_{it} = \Delta x_{it}\beta - \Delta r_{it}^{*}\beta + \Delta u_{it}$

$$\hat{\beta}_{FD} = \beta + \frac{Cov(\Delta x_{it}, \Delta u_{it} - \beta\Delta r_{it})}{Var(x_{it})}$$
$$= \beta - \beta\frac{Cov(\Delta x_{it}, \Delta r_{it})}{Var(\Delta x_{it})}$$
$$= \beta - 2\beta\frac{Cov(\Delta x_{it}, \Delta r_{it})}{Var(\Delta x_{it})}$$
$$= \beta - 2\beta\frac{\left[\sigma_r^2 - 2Cov(\Delta r_{it}, \Delta r_{it-1})\right]}{Var(\Delta x_{it})}$$
$$= \beta\left(1 - \frac{\sigma_r^2(1-\rho_r)}{\sigma_x^2(1-\rho_x) + \sigma_r^2(1-\rho_r)}\right)$$

where $\rho_x = \text{corr}(x_{it}^{*}, x_{it-1}^{*})$, $\rho_r = \text{corr}(r_{it}, r_{it-1})$.

$Var(\Delta x_{it}) = 2\left[\sigma_x^2(1-\rho_x) + \sigma_r^2(1-\rho_r)\right]$. As $\rho_x \to 1$, bias $\to -\beta$.

How to estimate such a model?

Consistent a more general form:

$$y_{it} = z_{it}\beta + w_{it}^{*}\beta + c_i + u_{it},$$

Let $r_{it} = w_{it} - w_{it}^{*}$.

Assume strict exogeneity of w_{it}^{*} and z_{it}. Replacing w_{it}^{*} with w_{it} and 1st differencing:

$$\Delta y_{it} = \Delta z_{it}\beta + w_{it}\beta - r_{it}\beta + \Delta u_{it}$$

The standard classic error in variable assumption:

$$E(r_{it}| z_i, w_i^{*}, c_i) = 0, \quad t = 1, 2, \ldots, T,$$

which implies r_{it} is uncorrelated with z_{is} and w_{is}^{*} for all t and s.

However, Δr_{it} and Δw_{it} are correlated , so we need *IV* for Δw_{it}.

Two approaches:

(i) Measure w_{it}^{*} twice.

(ii) If $E(r_{it}r_{is}) = 0$ for any $t \neq s$, no serial correlation in measurement error.
In this case, we have instruments readily available.
Therefore, the set of variables $(w_{it-2}, w_{it-3, \ldots,} w_{i1})$ and $(w_{it+1}\ w_{it+2, \ldots,} w_{iT})$ as instruments for Δw_{it}.

Models with Individual Specific Slopes

$$y_{it} = x_{it}\beta + c_i + g_i t + u_{it}, \qquad (9)$$

For example, each city, individual, firm etc. is allowed to have its own time trend. "random trend model." If y_{it} is natural log of some variables (such as GDP), then (9) is sometimes referred as "random growth model," in which g_i is roughly the average growth rate over a period (holding the explanatory variables fixed).

We want to allow (c_i, g_i) to be arbitrarily correlated with x_{it}. Our primary interest is to consistently estimate β.

Strict exogeneity, $E(u_{it}|x_{i1}, \ldots, x_{it}, c_i, g_i) = 0$

One approach: (1) first differencing.

$$\Delta y_{it} = \Delta x_{it}\beta + g_i + \Delta u_{it}, \qquad (10)$$

(2) fixed effect model of (*). Obviously it requires $T \geq 3$

Example: Friedberg, L (1998): "Did Unilateral Divorce Raise Divorce Rates? Evidence from Panel Data," *American Economic Review* 88, 608-627.

Issue: do "unilateral and no-fault divorce" laws encourage more divorces?

A substantial increase in divorce rates: 2.2 per thousand people in 1960 to 5.0 in 1985. At the same time, states substantially liberalized and simplified their divorce laws: (i) most states used to require both spouses had to consent in the absence of fault; no longer so. (ii) Most states also adopted some form of no-fault divorce, eliminating the need for one spouse to prove a transgression by the other.

A recent trend to abandon this – family values, etc.

$$divorce\ rate_{st} = c_0 + c_1{*}unilateral_{st} + c_{2s}{*}state_s + c_{3t}{*}year_t + c_{4s}{*}state_s{*}time_t + c_{5s}{*}state_s{*}time_t^2 + u_{st}$$

Data and estimates:

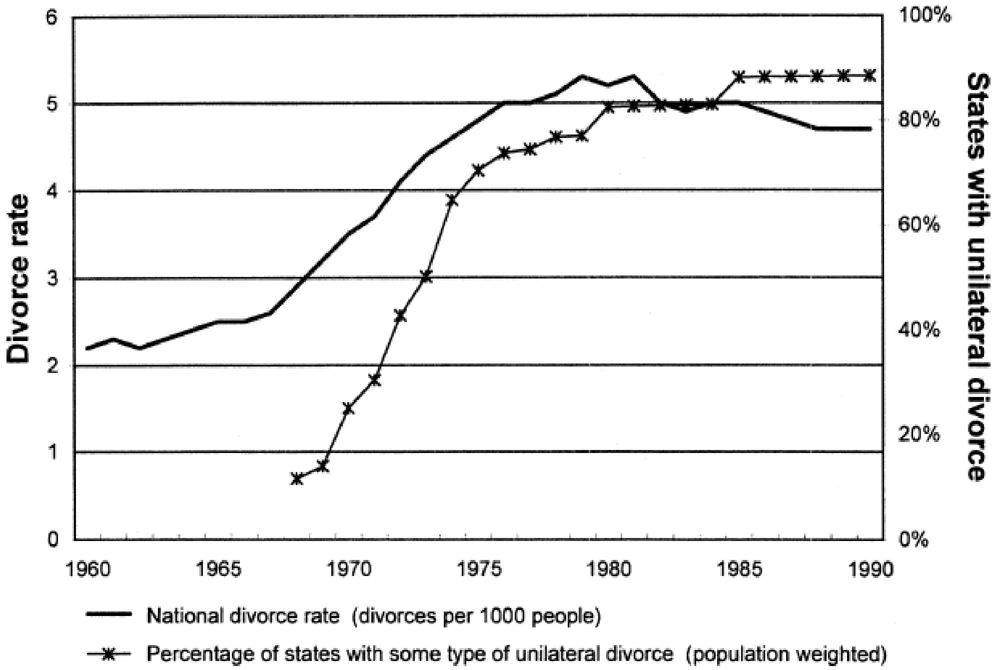

FIGURE 1. THE DIVORCE RATE AND DIVORCE LAWS

From the graph, it looks like a quadratic growth (Data is listed in the paper).

TABLE 3—REGRESSION RESULTS

	Dependent variable: Divorce rate (divorces per 1,000 people)				
Independent variables:	3.1	3.2	3.3	3.4[a]	3.5
Unilateral	1.802 (0.087)	1.509 (0.090)	0.004 (0.056)	0.447 (0.050)	0.441 (0.055)
Adjusted R^2	0.314	0.362	0.946	0.976	0.982
Year effects[b]	No	Yes, $F = 4.9$	Yes, $F = 89.0$	Yes, $F = 95.3$	Yes, $F = 8.9$
State effects[b]	No	No	Yes, $F = 217.3$	Yes, $F = 196.2$	Yes, $F = 131.1$
State trend, linear[b]	No	No	No	Yes, $F = 24.7$ State trends and effects, $F = 256.0$	Yes, $F = 9.3$
State trend, quadratic[b]	No	No	No	No	Yes, $F = 6.5$ State trends and effects, $F = 224.8$

Notes: Regressions of state divorce rates, 1968–88, on whether a state has a unilateral-divorce law. Unilateral-divorce law is defined according to column (1) of Table 1. Standard errors are in parentheses. Regressions are weighted by state population. $N = 1{,}043$. Data for some states in certain years is missing. For others it is incomplete, which is accounted for by a set of dummies described in the Appendix.

[a] The coefficients on the state effects and trends are shown in Tables 6 and 7.

[b] All reported F-statistics have corresponding p-values that are smaller than 0.00005.

Therefore, having quadratic term in "time" is important. Note if no year effect and state effect (column 3.1), the estimate is big and significant. However, with year and state effect (columns 3.2 and 3.3), the estimate are no longer significant; finally, allowing trend, the estimates (columns 3.4 and 3.5) become significant again and estimates are reasonable.

Model: A general model with individual-specific slopes:

$$y_{it} = x_{it}\beta + z_{it}\alpha_i + u_{it}. \qquad (10.1)$$

This model allows some time-invariant parameters α_i. Note there is no longer a need to have unobserved heterogeneity term c_i.

Example: (i) Polachak and Kim 1994 return to education may be different for different people. (ii) Lemieux (1998) unobserved heterogeneity is awarded differently in the union and nonunion sectors.

Assumption 1 (strict exogeneity): E($u_{it}|z_i, x_i, a_i$) = 0

Rewrite equation (10.1):

$$y_i = x_i\beta + z_i\alpha_i + u_i. \qquad (11)$$

Define $M_i = I_T - z_i(z_i'z_i)^{-1}z_i'$. It is interesting to note that $M_iz_i = 0$. So premultiplying M_i to (11) would eliminate the $z_{it}\alpha_i$ term.

Note that:

$$M_iy_i = y_i - z_i\hat{\gamma}, \qquad (12)$$

which is the residual from the regression of y_i on z_i.

$$M_ix_i = x_i - z_i\hat{\gamma}_2$$

which is the is OLS estimate of the regression of x_i on z_i. Therefore, the M_i premultiplied (11) is:

$$M_iy_i = M_ix_i\beta + M_iu_i.$$

If we let $M_iy_i = y_i - z_i\hat{\gamma} = \ddot{y}_t$, and $M_ix_i = x_i - x_i\hat{\gamma}_2 = \ddot{x}_t$. Then (12) becomes:

$$\ddot{y}_i = \ddot{x}_i\beta + \ddot{u}_i, \qquad (13)$$

The OLS of (13) is:

$$\hat{\beta}_{FE} = \left(\sum_i \ddot{x}_i{}'\ddot{x}_i\right)^{-1}\left(\sum_i \ddot{x}_i{}'\ddot{y}_i\right)$$
$$= \beta + \left(\sum_i \ddot{x}_i{}'\ddot{x}_i\right)^{-1}\left(\sum_i \ddot{x}_i{}'\ddot{u}_i\right)$$

Given the strict exogeneity assumption, $E(\ddot{x}_i{}'\ddot{u}_i) = 0$. So the estimate is consistent.

Further, if we assume that $E(u_i u_i'|z_i, x_i, a_i) = \sigma_u^2 I$ –
Or we could use robust standard error.

To obtain a consistent estimator of α_i in equations (10.1) or (11):

Premultiply (11) by $(z_i{}'z_i)^{-1}z_i{}'$ and rearrange to get:

$$(z_i{}'z_i)^{-1}z_i{}'y_i = (z_i{}'z_i)^{-1}z_i{}'z_i\alpha + (z_i{}'z_i)^{-1}z_i{}'x_i\beta + (z_i{}'z_i)^{-1}z_i{}'u_i$$

So we get:

$$\alpha_i = (z_i{}'z_i)^{-1}z_i{}'y_i - (z_i{}'z_i)^{-1}z_i{}'x_i\beta - (z_i{}'z_i)^{-1}z_i{}'u_i$$

Under Assumption 1, we have

$$E(\alpha_i) = \alpha = E\left[(z_i{}'z_i)^{-1}z_i{}'(y_i - x_i\beta)\right]$$

So, a consistent estimator of $E(\alpha_i)$ is:

$$\hat{\alpha} = \frac{1}{N}\sum_{i=1}^{N}(z_i{}'z_i)^{-1}z_i{}'\left(y_i - x_i\hat{\beta}_{FE}\right) \qquad (14)$$

Note $\hat{\alpha}$ in (14) averages over all α_i.

With fixed T, we cannot consistently estimate each α_i, as in the case of linear panel data model. However, for each i, the term $\hat{\alpha}_i$ could be unbiased. To get asymptotic covariance matrix we define:

$$\hat{\alpha}_i = (z_i{}'z_i)^{-1}z_i{}'\left(y_i - x_i\hat{\beta}_{FE}\right)$$

It is easy to see that

$$E(\hat{\alpha}_i \mid z, x) = (z_i{}'z_i)^{-1}z_i{}'\left(E(y_i) - x_iE\left(\hat{\beta}_{FE}\right)\right)$$
$$= (z_i{}'z_i)^{-1}z_i{}'(z_i\alpha_i + x_i\beta - x_i\beta)$$
$$= \alpha_i$$

Hausman and Taylor-Type Models

Consider a model:

$$y_{it} = x_{it}\beta + z_i\gamma + c_i + u_{it}.$$

We allow Cov(x_{it},c_i) ≠ 0, but assume Cov(z_i,c_i) = 0, and the strict exogeneity on u_{it}. The key coefficient here is γ. Since z_i do not vary with time, so the fixed-effect model cannot be used. A typical way to estimate this is to apply the random-effects model, where it is assumed that Cov(x_{it},c_i) = 0. The Hausman-Taylor type model has less strict assumption than the random-effects model by allowing x_{it} and c_i being correlated.

To estimate this model, first we estimate the fixed-effect model – note the both the $z_i\gamma$ term and c_i term are eliminated at this stage. From the estimates of the fixed-effects model, we can get residual. Note the residual includes three terms: $z_i\gamma$ term, c_i term, and the u_{it} term.

Second, we take the average (over time, for each i) to minimize the effect of the u_{it} term. The averaged residual now mostly consists of the z_i term and the c_i term. Because these two terms (z_i term and the c_i term) are uncorrelated, and z_i is observed, we can run regression of the averaged residual on z_i to obtain the consistent estimate of γ.

Example: we are interested in how the size of the local labor market affect local the unemployment rate (Gan and Zhang, *Journal of Econometrics*, 2006, page 127-152.). We have panel data of 295 monthly city unemployment rates.

$$\textit{unemployment rate}_{ct} = \alpha_c + X_{ct}\beta + \gamma \log(\textit{size}_c) + u_{ct}$$

It is possible to allow α_c and X_{ct} to be arbitrarily correlated, but assume that α_c and log($size_c$) to be uncorrelated.

How to estimate such a model:

Since E(z_i'c_i) = 0, we have an extra moment condition that can be used to identify γ.

Intuition: we first estimate β by fixed effect model. Then we use the residual to estimate γ, given

$$E(z_i' z_i)\gamma = E(z_i'(\bar{y}_i - \bar{x}_i\beta))$$

The estimate is given by:

$$\hat{\gamma} = \left(\frac{1}{N} \sum_{i=1}^{N} z_i ' z_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} z_i ' \left(\bar{y}_i - \bar{x}_i \hat{\beta}_{FE} \right) \right)$$

STATA Deviation

Consider the model: $y_{it} = x_{it}\beta + z_i\gamma + c_i + u_{it}$.

The STATA command on this type of models is `xthtaylor`. This command defines endogenous and exogenous variables. The endogenous variables are variables that are allowed to be correlated with c_i, while exogenous variables are variables that are not allowed to be correlated with c_i. In order to be able to identify γ, it is necessary to have at least one exogenous variable z_i to ensure the moment condition $E(z_i' c_i) = 0$. If all z_i are endogenous, then the model becomes the typical panel data fixed-effects model where γ cannot be identified.

The advantage of Hausman-Taylor model over the typical random-effects model is that it allows endogenous variables in x_{it} (allowing c_i and x_{it}) to be correlated. It even allows some of the z_i to be correlated.

Example 1: In Gan and Zhang, *Journal of Econometrics* 2006, page 127-152.)

In this example, the endogenous variables (that are allowed to be correlated with c_i) are X_{ct}, and the exogenous variables (that are uncorrelated with c_i) are log($size_c$).

`xthtaylor` *unemployment_rate* X_{ct} log($size_c$), `endog`(X_{ct})

Example 2: wage is a function of

How long this person has worked for the firm, wks
Binary variables if lives in a metropolitan area or in the south, smsa, and south
Marital status, *ms*;
Years of education, *ed*
A quadratic of work experience
In manufacturing or not,
Black or not, *blk*
Female or not, *fem*

It is expected that time-varying variables exp, exp2, wks, ms, and union are all correlated with the unobserved individual effect.

Assume exogenous variables *occ, south, smsa, ind, fem*, and *blk* are instruments for the endogenous, time-invariant variable *ed*.

```
xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog(exp exp2
    wks ms union ed)
```

Matched Pairs and Cluster Samples

Example, husband and wife, or siblings,

$$y_{i1} = x_{i1}\beta + f_i + u_{i1}$$
$$y_{i2} = x_{i2}\beta + f_i + u_{i2}$$

We just have to recognize that this is equivalent to a two-period panel data model. Or more generally, a cluster model, where number of people in a cluster may vary. This will cause problems of homoscedasticity – therefore, it is natural to use robust standard error.

Peer effect model:

$$y_{is} = x_{is}\beta + \overline{w}_{i(s)}\delta + v_{is}$$

There are many examples for this type of models. The issue here is if the cluster or groups are endogenously formed. If they are, this is similar to Hausman-Taylor model. However, if they are endogenously formed, then we have to find outside instruments.

A recent paper using Air Force Academy – in which case the group is exogenously formed is quite interesting.

`