Panel Data

Definition

Panel data (longitudinal data) refers to data obtained from sampling schemes where the same set of units (individuals, firms, countries, etc.) are observed over time.

- If observations are made on units at a fixed point in time, we obtain cross-section data.
- If one unit is observed over time, we obtain time-series data.
- Panel data combine the two.
- Panel data are usually observed at regular time intervals (monthly, quarterly, yearly, etc.), and are balanced (all units are observed at all periods).
- Panel data can be
  - a short panel: many units and few time periods,
  - a long panel: many time periods and few units,
  - a large panel: many units and many time periods.

Panel data

- In economics, short panels are synonymous with micro panels.
- Panels with small to moderate number of units are called macro panels.

Some well-known Micro Panels:

1. Panel Study of Income Dynamics (PSID), by Institute of Social Research at University of Michigan, http://psidonline.isr.umich.edu.
2. National Longitudinal Survey (NLS), a set of surveys sponsored by the Bureau of Labor Statistics, http://www.bls.gov/nls/home.htm.
3. Current Population Survey (CPS), by Bureau of Census for the Bureau of Labor Statistics, http://www.census.gov/cps.
4. Living Standard Measurement Study (LSMS), by World Bank, http://www.worldbank.org/LSMS.
5. German Social-Economic Panel, http://www.diw.de/soep.
6. Canadian Survey of Labor Income Dynamics (SLID), collected by Statistics Canada, www.statcan.gc.ca.

Panel data

Some well-known Macro Panels:

1. Penn World Table, www.nber.org.
2. World Bank, http://data.worldbank.org.
3. International Monetary Fund, http://www.imf.org.
4. Bureau of Economic Analysis, https://www.bea.gov/data.
5. Bureau of Labor Statistics, https://www.bls.gov/data/.
6. FreddieMac, http://www.freddiemac.com/research/indices/house-price-index.page.

Panel data

The State Traffic Fatality Data Set

panelreps.R | fatality

Filter

	state	year	spircons	unrate	perinc	emppop	beertax	sobapt	mormon	mlda	dry	yngdrv	vmiles	breath	jaild	comserd
1	1	1982	1.37	14.4	10544.152	50.69204	1.53937948	30.35570	0.328290	19.00	25.006300	0.211572	7233.887	0	0	0
2	1	1983	1.36	13.7	10732.798	52.14703	1.78899074	30.33360	0.343410	19.00	22.994200	0.210768	7836.348	0	0	0
3	1	1984	1.32	11.1	11108.791	54.16809	1.71428561	30.31150	0.359240	19.00	24.042601	0.211484	8262.990	0	0	0
4	1	1985	1.28	8.9	11332.627	55.27114	1.65254235	30.28950	0.375790	19.67	23.633900	0.211140	8726.917	0	0	0
5	1	1986	1.23	9.8	11661.507	56.51450	1.60990703	30.26740	0.393110	21.00	23.464701	0.213400	8952.854	0	0	0
6	1	1987	1.18	7.8	11944.000	57.50988	1.55999994	30.24530	0.411230	21.00	23.792400	0.215527	9166.302	0	0	0
7	1	1988	1.17	7.2	12368.624	56.83453	1.50144362	30.22330	0.430180	21.00	23.792400	0.218328	9674.323	0	0	0
8	4	1982	1.97	9.9	12309.069	56.89330	0.21479714	3.95890	4.919100	19.00	0.000000	0.209012	6810.157	0	1	1
9	4	1983	1.90	9.1	12693.808	57.55363	0.20642203	3.89010	4.831070	19.00	0.000000	0.203855	6587.495	0	1	1
10	4	1984	2.14	5.0	13265.934	60.37902	0.29670331	3.82260	4.744610	19.00	0.000000	0.209127	6709.970	0	1	1
11	4	1985	1.86	6.5	13726.695	58.64853	0.38135594	3.75620	4.659710	21.00	0.000000	0.188428	6771.263	0	1	1
12	4	1986	1.78	6.9	14107.327	60.28018	0.37151703	3.69100	4.576320	21.00	0.000000	0.171539	8129.008	0	1	1
13	4	1987	1.72	6.2	14241.000	60.21506	0.36000001	3.62690	4.494420	21.00	0.000000	0.168724	9370.654	0	1	1
14	4	1988	1.68	6.3	14408.085	60.49767	0.34648702	3.56400	4.413990	21.00	0.000000	0.161005	9815.721	0	1	1
15	5	1982	1.19	9.8	10267.303	54.47586	0.65035802	22.96720	0.328290	21.00	36.712799	0.204903	7208.500	0	0	0
16	5	1983	1.20	10.1	10433.486	53.81479	0.67545873	23.00090	0.343410	21.00	36.430099	0.194169	7175.917	0	0	0
17	5	1984	1.22	8.9	10916.483	54.67128	0.59890109	23.03460	0.359240	21.00	36.104000	0.186380	7084.820	0	0	0
18	5	1985	1.12	8.7	11149.364	54.97712	0.57733053	23.06840	0.375790	21.00	35.904999	0.189292	7253.918	0	0	0
19	5	1986	0.92	8.7	11399.381	55.56186	0.56243551	23.10220	0.393110	21.00	39.569599	0.161957	7468.999	0	0	0

Advantages of Panel Data

- Panel data enable us to control for individual heterogeneity.
- Panels give more sources of variation, less collinearity among the variables, and yield more efficient estimators.
- With panel data, it is easier to study dynamic models.
- They are more suitable for identifying and measuring effects that are not detectable in pure cross-section and pure time-series data.
- Micro panel data gathered on individuals, firms, and households can be measured more accurately than similar variables measured at the macro level. Biases resulted from aggregation over time or individuals may be reduced or eliminated.

Limitations of Panel Data

- Design and/or data collection problems.
- Measurement errors.
- Selectivity problems: self-selection, non-response, attrition.
- Short-time dimension.
- Potential dependence over cross-sections:
  - macro panels with long time dimension that do not account for cross-sectional dependence may lead to misleading inference.

Static Panel Data Models

Definition

A static panel data model (or a panel data regression model) differs from a cross-sectional or time-series model in that it involves double subscripts, i.e.,

$$y_{it} = \alpha + \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + \varepsilon_{it}, \quad i = 1, 2, \ldots, N; \quad t = 1, 2, \ldots, T;$$

where i denotes cross-sections, t denotes time periods, y_{it} is the dependent variable, $\boldsymbol{x}_{it}$ is a $K \times 1$ vector of explanatory variables, ε_{it} is the disturbance term, α is a scalar parameter, $\boldsymbol{\beta}$ is a $K \times 1$ vector of parameters.

- If $\{\varepsilon_{it}\}_{i,t}^{N,T}$ are independent and identically distributed (iid), then the panel data regression model is no different than a multiple linear regression model.
- If $\varepsilon_{it} = \mu_i + \nu_{it}$, where μ_i denotes the unobserved individual specific effects and $\{\nu_{it}\}$ are the so-called idiosyncratic errors which are usually i.i.d. across i and t. Then, the model above called the one-way error component model.

Static Panel Data Models

- Moreover, if $\varepsilon_{it} = \mu_i + \lambda_t + \nu_{it}$, where λ_t denotes the unobserved time-specific effects, then the model is called the two-way error component model.

- For example, μ_i could stand for unobserved (time-invariant) innate ability of individual i, whereas λ_t could stand for unobserved (individual invariant) macro economic shock at period t.

- μ_i and λ_t could be correlated with the time-varying explanatory variables $\boldsymbol{x}_{it}$ in an arbitrary manner. Then, $\{\mu_i\}$ and $\{\lambda_t\}$ need be treated as unknown parameters. In this case, the model is called the fixed effects model.

- If μ_i and λ_t are uncorrelated with $\boldsymbol{x}_{it}$, $\{\mu_i\}$ and $\{\lambda_t\}$ can be treated as iid random variables. In this case, the model is called the random effects model.

- Hence, we can have

  1. one-way fixed effects model,
  2. one-way random effects model,
  3. two-way fixed effects model,
  4. two-way random effects model.

Static Panel Data Models

The one-way error component model is given by

$$y_{it} = \alpha + \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + \varepsilon_{it}, \quad \varepsilon_{it} = \mu_i + \nu_{it}$$

for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$;

- i denotes cross-sections such as individuals, firms, countries, etc.,
- t denotes time periods,
- α is the intercept of the model,
- $\boldsymbol{x}_{it}$ denotes the explanatory variables,
- μ_i denotes the unobserved individual specific effects,
- ν_{it} denotes the idiosyncratic errors.

Fixed effects and Random effects

- Recall that the way in which μ_i relates to $\boldsymbol{x}_{it}$ determines fixed effects model vs. random effects model.
- If μ_i relates to $\boldsymbol{x}_{it}$ in an arbitrary manner (i.e., dependent), $\{\mu_i\}$ need be treated as fixed parameters and we end up with the fixed effects model.
    - For example, in a panel of firms, μ_i could stand for the unobserved managerial skills.
- If μ_i is uncorrelated with $\boldsymbol{x}_{it}$, $\{\mu_i\}$ can be treated as a set of iid random variables and we end up with the random effects model.
    - For example, in (large) micro panels, individuals are randomly drawn from a large population, it is typical to assume that μ_i is uncorrelated with $\boldsymbol{x}_{it}$.
- Note that there is a model called correlated random effects model which can be considered as the middle ground between the fixed effects model and the random effects model. In this model, μ_i is a linear function of $\boldsymbol{x}_{it}$.

One-way Fixed effects model

$$y_{it} = \alpha + \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + \varepsilon_{it}, \quad \varepsilon_{it} = \mu_i + \nu_{it}$$

- μ_i in ε_{it} is permitted to be correlated with $\boldsymbol{x}_{it}$ in an arbitrary manner.
- Hence, they need be treated as fixed parameters to be estimated together with other model parameters.
- Since both α and μ_i are fixed parameters, we cannot separately identify and estimate them.
- Think of μ_i as the intercept for the ith unit, and there is also intercept α that apply to all units. If we include all unit specific intercepts and α together, and try to estimate them we run into the perfect multicollinearity problem.
- Hence, we need to impose a constraint either on μ_is or on α, i.e., either $\sum_{i=1}^{N} \mu_i = 0$ or $\alpha = 0$.

One-way Fixed effects model

- If we choose to impose $\sum_{i=1}^{N} \mu_i = 0$, then μ_i represents the deviation of the ith unit from the common mean α.
- If we choose to impose $\alpha = 0$, then μ_is are free parameters. We will adopt this assumption.

Assumption (A)

(i) The idiosyncratic errors $\nu_{it} \sim$ iid$(0, \sigma_\nu^2)$ for all i and t; and (ii) $(\boldsymbol{x}_{it}^{'}, \mu_i)^{'}$ are uncorrelated with ν_{it}.

- Assumption A implies that $\mathbb{E}(y_{it}|\boldsymbol{x}_{it}, \mu_i) = \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + \mu_i$, because $\mathbb{E}(\nu_{it}|\boldsymbol{x}_{it}, \mu_i) = 0$.
- Assumption A allows for several estimation methods for the model parameters.

Least-squares dummy variable estimation

- The one-way fixed effects model can be estimated using the so-called least-squares dummy variable (LSDV) method.
- The model can be written as

$$y_{it} = \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + \mu_i + \nu_{it}.$$

- The LSDV approach minimizes the sum of the squared residuals to estimate the model parameters.
- Let

$$\mathcal{S}(\boldsymbol{\beta}, \boldsymbol{\mu}) = \sum_{i=1}^{N}\sum_{t=1}^{T}\nu_{it}^2 = \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} - \mu_i)^2.$$

- Then, the LSDV estimator is the solution from

$$\underset{\{\boldsymbol{\beta}, \boldsymbol{\mu}\}}{\arg\min}\ \mathcal{S}(\boldsymbol{\beta}, \boldsymbol{\mu}).$$

Least-squares dummy variable estimation

- Taking the partial derivatives of $\mathcal{S}(\boldsymbol{\beta}, \boldsymbol{\mu})$ with respect to (wrt) μ_is and setting it to zero, we get, for $i = 1, 2, \ldots, N$,

$$\frac{\partial \mathcal{S}(\boldsymbol{\beta}, \boldsymbol{\mu})}{\partial \mu_i} = \frac{\partial \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} - \mu_i)^2}{\partial \mu_i} \quad \text{(for fixed } i\text{)}$$

$$= -2 \sum_{t=1}^{T} (y_{it} - \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} - \mu_i) \overset{set}{=} 0$$

$$\Rightarrow \widehat{\mu}_i = \frac{1}{T} \sum_{t=1}^{T} y_{it} - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} = \bar{y}_{i\cdot} - \bar{\boldsymbol{x}}_{i\cdot}^{'}\boldsymbol{\beta}$$

- Substituting $\widehat{\mu}_i$s back into $\mathcal{S}(\boldsymbol{\beta}, \boldsymbol{\mu})$ yields the concentrated sum of squares:

$$\mathcal{S}(\boldsymbol{\beta}, \widehat{\boldsymbol{\mu}}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \bar{y}_{i\cdot} - [\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_{i\cdot}]^{'} \boldsymbol{\beta} \right)^2$$

$$= \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it}^{*} - \boldsymbol{x}_{it}^{*'}\boldsymbol{\beta})^2.$$

Least-squares dummy variable estimation

- Taking the partial derivatives of $\mathcal{S}(\boldsymbol{\beta}, \widehat{\boldsymbol{\mu}})$ wrt $\boldsymbol{\beta}$ and setting it to zero, we get

$$\frac{\partial \mathcal{S}(\boldsymbol{\beta}, \widehat{\boldsymbol{\mu}})}{\partial \boldsymbol{\beta}} = -2 \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{x}_{it}^{*} (y_{it}^{*} - \boldsymbol{x}_{it}^{*'} \boldsymbol{\beta}) \stackrel{set}{=} \mathbf{0}$$

$$\Rightarrow \widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{x}_{it}^{*} \boldsymbol{x}_{it}^{*'} \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{x}_{it}^{*} y_{it}^{*}.$$

- Recall that $y_{it}^{*} = y_{it} - \bar{y}_{i\cdot}$ and $\boldsymbol{x}_{it}^{*} = \boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_{i\cdot}$.
- Hence, the LSDV estimator utilizes the within variation in each i.
- There is indeed an estimator called the within estimator to estimate the one-way fixed effects model.
- Note that the LSDV estimator is equivalent to the within estimator.

The within estimator

- To derive the within estimator, we need to first stack the observations in a consistent way so that we can write the model for the entire sample.

$$\underbrace{\begin{pmatrix} y_{11} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT} \end{pmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{pmatrix} \boldsymbol{x}'_{11} \\ \vdots \\ \boldsymbol{x}'_{1T} \\ \vdots \\ \boldsymbol{x}'_{N1} \\ \vdots \\ \boldsymbol{x}'_{NT} \end{pmatrix}}_{\boldsymbol{X}} \boldsymbol{\beta} + \underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}}_{\mathbf{I}_N \otimes \boldsymbol{l}_T} \underbrace{\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix}}_{\boldsymbol{\mu}} + \underbrace{\begin{pmatrix} \nu_{11} \\ \vdots \\ \nu_{1T} \\ \vdots \\ \nu_{N1} \\ \vdots \\ \nu_{NT} \end{pmatrix}}_{\boldsymbol{\nu}}$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}_\mu \boldsymbol{\mu} + \boldsymbol{\nu}$$

where $\boldsymbol{Z}_\mu = \mathbf{I}_N \otimes \boldsymbol{l}_T$.

The within estimator

- Let $\mathbb{X} = (\boldsymbol{X}, \boldsymbol{Z}_\mu)$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}^{'}, \boldsymbol{\mu}^{'})^{'}$. Then, the one-way fixed effects model is given by

$$\boldsymbol{y} = \mathbb{X}\boldsymbol{\theta} + \boldsymbol{\nu}.$$

- We can now utilize the least-squares methodology under Assumption A:

$$\widehat{\boldsymbol{\theta}} = (\mathbb{X}^{'}\mathbb{X})^{-1}\mathbb{X}^{'}\boldsymbol{y}.$$

- Let $\mathbb{P}_\mu = \boldsymbol{Z}_\mu(\boldsymbol{Z}_\mu^{'}\boldsymbol{Z}_\mu)^{-1}\boldsymbol{Z}_\mu^{'}$ and $\mathbb{Q}_\mu = \mathbf{I}_{NT} - \mathbb{P}_\mu$. These are the projection and orthogonal projection matrices we are familiar with.

- Recall that $\mathbb{P}_\mu \boldsymbol{Z}_\mu = \boldsymbol{Z}_\mu$ and $\mathbb{Q}_\mu \boldsymbol{Z}_\mu = \mathbf{0}_{NT\times N}$.

- Using the inverse of a partitioned matrix formula, one can find (the within estimator)

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^{'}\mathbb{Q}_\mu \boldsymbol{X})^{-1}\boldsymbol{X}^{'}\mathbb{Q}_\mu \boldsymbol{y}.$$

The within estimator

- Notice that

$$\mathbb{Q}_\mu = \mathbf{I}_N \otimes \left(\mathbf{I}_T - \boldsymbol{l}_T(\boldsymbol{l}_T'\boldsymbol{l}_T)^{-1}\boldsymbol{l}_T'\right) = \mathbf{I}_N \otimes \mathbb{Q}_{\boldsymbol{l}_T}$$
$$= \mathbf{I}_N \otimes \begin{pmatrix} 1-\frac{1}{T} & -\frac{1}{T} & \cdots & -\frac{1}{T} \\ -\frac{1}{T} & 1-\frac{1}{T} & \cdots & -\frac{1}{T} \\ \vdots & \vdots & \cdots & \vdots \\ -\frac{1}{T} & -\frac{1}{T} & \cdots & 1-\frac{1}{T} \end{pmatrix}$$

.

- Hence, for example,

$$\begin{pmatrix} \mathbb{Q}_{\boldsymbol{l}_T} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbb{Q}_{\boldsymbol{l}_T} & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbb{Q}_{\boldsymbol{l}_T} \end{pmatrix} \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_N \end{pmatrix} = \begin{pmatrix} \boldsymbol{y}_1 - \bar{y}_{1\cdot} \\ \boldsymbol{y}_2 - \bar{y}_{2\cdot} \\ \vdots \\ \boldsymbol{y}_N - \bar{y}_{N\cdot} \end{pmatrix} = \begin{pmatrix} \boldsymbol{y}_1^* \\ \boldsymbol{y}_2^* \\ \vdots \\ \boldsymbol{y}_N^* \end{pmatrix}.$$

- The same argument applies to each column in $\boldsymbol{X}$. Hence, the LSDV is equivalent to the within estimator.

The within estimator

- Under Assumption A,

$$\begin{aligned}
\mathbb{E}(\widehat{\boldsymbol{\beta}}|\mathbb{X}) &= (\boldsymbol{X}^{'}\mathbb{Q}_{\mu}\boldsymbol{X})^{-1}\boldsymbol{X}^{'}\mathbb{Q}_{\mu}\mathbb{E}(\boldsymbol{y}|\boldsymbol{X}) = \boldsymbol{\beta} \\
\mathsf{Var}(\widehat{\boldsymbol{\beta}}|\mathbb{X}) &= (\boldsymbol{X}^{'}\mathbb{Q}_{\mu}\boldsymbol{X})^{-1}\boldsymbol{X}^{'}\mathbb{Q}_{\mu}\sigma_{\nu}^{2}\mathbf{I}_{NT}\mathbb{Q}_{\mu}\boldsymbol{X}(\boldsymbol{X}^{'}\mathbb{Q}_{\mu}\boldsymbol{X})^{-1} \\
&= \sigma_{\nu}^{2}(\boldsymbol{X}^{'}\mathbb{Q}_{\mu}\boldsymbol{X})^{-1}.
\end{aligned}$$

- For inference on $\boldsymbol{\beta}$, an estimator of σ_{ν}^{2} is needed so that we can obtain $\widehat{\mathsf{Var}(\widehat{\boldsymbol{\beta}}}|\mathbb{X}) = \widehat{\sigma}_{\nu}^{2}(\boldsymbol{X}^{'}\mathbb{Q}_{\mu}\boldsymbol{X})^{-1}$. To this end, use the variation of the residuals, i.e.,

$$\widehat{\sigma}_{\nu}^{2} = \widehat{\boldsymbol{\nu}}^{'}\widehat{\boldsymbol{\nu}}/(NT - N - K).$$

where $\widehat{\boldsymbol{\nu}} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_{\mu}\widehat{\boldsymbol{\mu}}$ and $(\widehat{\boldsymbol{\beta}}^{'}, \widehat{\boldsymbol{\mu}}^{'})^{'}$ are the LSDV estimators.

Testing for the fixed effects

- It is often the case that the researcher needs to test for the joint significance of the individual dummies, i.e., test for

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_N = \alpha.$$

- Under the null hypothesis, the model is the pooled model (pooled OLS)

$$y_{it} = \alpha + \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + \nu_{it}.$$

- Under the alternative hypothesis, the model is the LSDV specification.
- Let RSS_r be the restricted residual sum of squares from fitting the null model, with $NT - K - 1$ degrees of freedom.
- Let RSS_u be the restricted residual sum of squares from fitting the alternative model, with $NT - N - K$ degrees of freedom.
- Under Assumption A and normality of ν_{it}, the F-test for testing H_0 is

$$F = \frac{(RSS_r - RSS_u)/(N-1)}{RSS_u/(NT-N-K)} \sim F_{N-1,\, NT-N-K}$$

Robust estimates of standard errors

- Recall that $\text{Var}(\widehat{\boldsymbol{\beta}}|\mathbb{X}) = \sigma_\nu^2(\boldsymbol{X}'\mathbb{Q}_\mu\boldsymbol{X})^{-1}$, where σ_ν^2 is estimated by $\widehat{\sigma}_\nu^2 = \widehat{\boldsymbol{\nu}}'\widehat{\boldsymbol{\nu}}/(NT - N - K)$.

- Note that this is correct under Assumption A. However, if the iid assumption on ν_{it} is violated, it is not correct.

- Recall that $\mathbb{Q}_{\boldsymbol{l}_T} = \mathbf{I}_T - \boldsymbol{l}_T(\boldsymbol{l}_T'\boldsymbol{l}_T)^{-1}\boldsymbol{l}_T' = \mathbf{I}_T - \frac{1}{T}\boldsymbol{l}_T\boldsymbol{l}_T'$.

- Recall also that $\mathbb{Q}_{\boldsymbol{l}_T}\boldsymbol{y}_i = \boldsymbol{y}_i - \bar{y}_{i\cdot} = \boldsymbol{y}_i^*$ for $i = 1, 2, \ldots, N$.

- Similarly, $\mathbb{Q}_{\boldsymbol{l}_T}\boldsymbol{x}_{i,j} = \boldsymbol{x}_{i,j}^*$ for $j = 1, 2, \ldots, k$ and $i = 1, 2, \ldots, N$. Let $\boldsymbol{X}_i^*$ denote $(\boldsymbol{x}_{i,1}^*, \boldsymbol{x}_{i,2}^*, \ldots, \boldsymbol{x}_{i,k}^*)$.

- Also, let $\mathbb{Q}_{\boldsymbol{l}_T}\boldsymbol{\nu}_i = \boldsymbol{\nu}_i^*$ for $i = 1, 2, \ldots, N$.

- Then, pre-multiplying the model $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{l}_T\mu_i + \boldsymbol{\nu}_i$ (for $i = 1, 2, \ldots, N$) is transformed into

$$\boldsymbol{y}_i^* = \boldsymbol{X}_i^*\boldsymbol{\beta} + \boldsymbol{\nu}_i^*$$

because $\mathbb{Q}_{\boldsymbol{l}_T}\boldsymbol{l}_T = \mathbf{0}_{T\times 1}$.

Robust estimates of standard errors

Assumption (A$'$)

(i) The idiosyncratic errors ν_{it}s are independent over the i's, but not necessarily over the t's, i.e., $\boldsymbol{\nu}_i^* \sim (\mathbf{0}, \Omega_i)$ for $i = 1, 2, \ldots, N$, where Ω_i a $T \times T$ positive definite matrix (including the case of serial correlation); and (ii) $\boldsymbol{X}_i^*$ are uncorrelated with $\boldsymbol{\nu}_i^*$.

- Then, a robust estimator for the variance of $\widehat{\boldsymbol{\beta}}$ is (the so-called sandwich formula)

$$\widehat{\text{Var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}^*)} = (\boldsymbol{X}^{*'}\boldsymbol{X}^*)^{-1}\left(\sum_{i=1}^{N}\boldsymbol{X}_i^{*'}\widehat{\boldsymbol{\nu}}_i^*\widehat{\boldsymbol{\nu}}_i^{*'}\boldsymbol{X}_i^*\right)(\boldsymbol{X}^{*'}\boldsymbol{X}^*)^{-1}.$$

One-way Random effects model

$$y_{it} = \alpha + \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + \varepsilon_{it}, \quad \varepsilon_{it} = \mu_i + \nu_{it}$$

- If μ_i is uncorrelated with $\boldsymbol{x}_{it}$, they can be considered as a set of random variable iid $(0, \sigma_\mu^2)$, independent of the idiosyncratic errors ν_{it}.
- Also, assume that $\nu_{it} \sim \text{iid}(0, \sigma_\nu^2)$.
- Hence, we can extend the specification to include the time invariant variables, say $\boldsymbol{w}_i$ and the individual invariant variables $\boldsymbol{h}_t$:

$$\boldsymbol{y} = \mathbb{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} = \boldsymbol{Z}_\mu \boldsymbol{\mu} + \boldsymbol{\nu}$$

  where $\mathbb{X}$ contains all types of regressors, constant, time-invariant, time and individual varying.

- Let $\boldsymbol{\Omega}$ denote Var$(\boldsymbol{\varepsilon})$. Then, one can show that

$$\begin{aligned}\boldsymbol{\Omega} &= \mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{'}|\mathbb{X}) = \boldsymbol{Z}_\mu \mathbb{E}(\boldsymbol{\mu}\boldsymbol{\mu}^{'}|\mathbb{X})\boldsymbol{Z}_\mu^{'} + \mathbb{E}(\boldsymbol{\nu}\boldsymbol{\nu}^{'}|\mathbb{X}) \\ &= \sigma_\mu^2(\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma_\nu^2(\mathbf{I}_N \otimes \mathbf{I}_T)\end{aligned}$$

  where $\mathbf{J}_T = \boldsymbol{l}_T \boldsymbol{l}_T^{'}$, a $T \times T$ matrix of ones.

One-way Random effects model

- Estimation of this model involves $\mathbf{\Omega}^{-1}$, which can be computationally challenging because the size of $\mathbf{\Omega}$ is $NT \times NT$.
- Luckily, this problem has been considered by the literature and an analytical expression for $\mathbf{\Omega}^{-1}$ using the spectral decomposition was proposed.
- Let $\bar{\mathbf{J}}_T = \frac{1}{T}\mathbf{J}_T$ and $\mathbb{Q}_T = \mathbf{I}_T - \bar{\mathbf{J}}_T$.
- It has been shown that

$$\begin{aligned}\mathbf{\Omega} &= (T\sigma_\mu^2 + \sigma_\nu^2)(\mathbf{I}_T \otimes \bar{\mathbf{J}}_T) + \sigma_\nu^2(\mathbf{I}_N \otimes \mathbb{Q}_T) \\ &= \sigma_1^2 \mathbb{P}_\mu + \sigma_\nu^2 \mathbb{Q}_\mu\end{aligned}$$

  where $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\nu^2$.
- Then,

$$\mathbf{\Omega}^{-1} = \sigma_1^{-2}\mathbb{P}_\mu + \sigma_\nu^{-2}\mathbb{Q}_\mu$$

  and

$$\mathbf{\Omega}^{-1/2} = \sigma_1^{-1}\mathbb{P}_\mu + \sigma_\nu^{-1}\mathbb{Q}_\mu.$$

Maximum likelihood estimation

Definition (Multivariate normal distribution)

A random n-vector $\boldsymbol{y}$ is said to have multivariate normal distribution with mean vector $\boldsymbol{\mu}_y$ and variance covariance matrix $\boldsymbol{\Sigma}$, denoted by $\boldsymbol{y} \sim N(\boldsymbol{\mu}_y, \boldsymbol{\Sigma})$, if its joint probability density function (pdf) is given by

$$f(\boldsymbol{y}; \boldsymbol{\mu}_y, \boldsymbol{\Sigma}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2} (\boldsymbol{y} - \boldsymbol{\mu}_y)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_y) \right).$$

- Note that in our model $\boldsymbol{y}$ has the mean $\mathbb{X}\boldsymbol{\theta}$, and the variance covariance matrix $\boldsymbol{\Omega}$.
- The likelihood of the model is given by

$$f(\boldsymbol{y}; \mathbb{X}\boldsymbol{\theta}, \boldsymbol{\Omega}) = (2\pi)^{-NT/2} |\boldsymbol{\Omega}|^{-1/2} \exp\left(-\frac{1}{2} (\boldsymbol{y} - \mathbb{X}\boldsymbol{\theta})' \boldsymbol{\Omega}^{-1} (\boldsymbol{y} - \mathbb{X}\boldsymbol{\theta}) \right).$$

- The log-likelihood of the model is given by

$$\ln f(\boldsymbol{y}; \mathbb{X}\boldsymbol{\theta}, \boldsymbol{\Omega}) = -\frac{NT}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Omega}| - \frac{1}{2} (\boldsymbol{y} - \mathbb{X}\boldsymbol{\theta})' \boldsymbol{\Omega}^{-1} (\boldsymbol{y} - \mathbb{X}\boldsymbol{\theta}).$$

Maximum likelihood estimation

- One can show that

$$|\mathbf{\Omega}| = (\sigma_\nu^2)^{N(T-1)}(\sigma_1^2)^N$$

and we already saw that $\mathbf{\Omega}^{-1} = \sigma_1^{-2}\mathbb{P}_\mu + \sigma_\nu^{-2}\mathbb{Q}_\mu$.

- After plugging-in these expressions to the log-likelihood, we can maximize the log-likelihood to find the maximum likelihood estimators of $(\boldsymbol{\theta}^{'}, \sigma_\nu^2, \sigma_\mu^2)^{'}$,

$$(\widehat{\boldsymbol{\theta}}, \widehat{\sigma}_\nu^2, \widehat{\sigma}_\mu^2)^{'} = \underset{\{\boldsymbol{\theta}, \sigma_\nu^2, \sigma_\mu^2\}}{\arg\max} \ln f(\boldsymbol{y}; \mathbb{X}\boldsymbol{\theta}, \mathbf{\Omega}).$$

Fixed effects vs Random effects

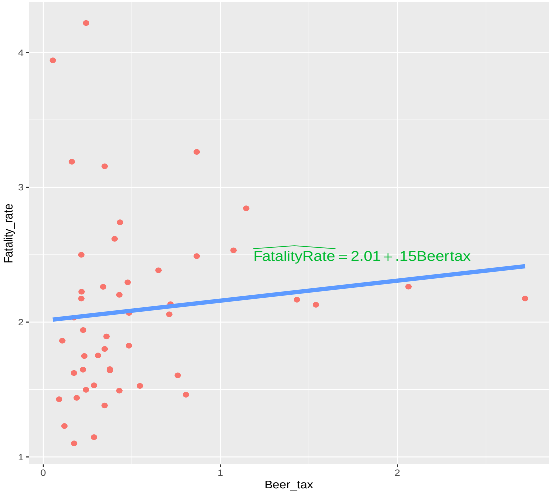
- How should we decide?
- Fixed effects control for the unobserved heterogeneity across i, but we cannot estimate the effects of time-invariant variables such as sex, race, etc.
- With fixed effects, prediction of the conditional mean is impossible, instead only changes in conditional mean caused by the changes or time-varying regressors can be predicted.
- Random effects overcome these difficulties, but the causal interpretation may then be unwarranted.
- Correlated random effects model may be more suitable as it allows the fixed effects to correlate with (some) time-varying regressors linearly.

Traffic deaths and alcohol taxes

- Variables:
  - Traffic fatality rate (#traffic deaths in state i in year t, per $1e4$ state residents
  - Tax on a case of beer
  - Other: legal driving age, legal drinking age, drunk driving laws, vehicle miles per driver, state socio-economic data, etc.
- Observational unit: a year in a US State: 48 States and 7 years, 336 observations.

Panel data

Figure: Fatality vs Beer tax, for 1982



Traffic deaths and alcohol taxes

- Why might there be higher more traffic deaths in states that have higher alcohol taxes?
  - Quality (age) of automobiles
  - Quality of roads
  - Culture around drinking and driving
  - Density of cars on the road

Traffic deaths and alcohol taxes

- These omitted factors could cause omitted variable bias.
- For example (1), take traffic density.
    - High traffic density means more traffic deaths.
    - (Western) states with lower traffic density have lower alcohol taxes.
- Then, the two conditions for omitted variable bias are satisfied.
- Specifically, high taxes could reflect high traffic density (so the LS coefficient would be biased positively – high taxes, more deaths).
- Panel data lets us eliminate omitted variable bias when the omitted variables are constant over time within a given state.

Traffic deaths and alcohol taxes

- For example (2), take cultural attitudes towards drinking and driving.
  - arguably are a determinant of traffic deaths.
  - potentially are correlated with the beer tax.
- Then, the two conditions for omitted variable bias are satisfied.
- Specifically, high taxes could pick up the effect of cultural attitudes towards drinking (so the LS coefficient would be biased).
- Panel data lets us eliminate omitted variable bias when the omitted variables are constant over time within a given state.

Traffic deaths and alcohol taxes

- Potential omitted variable bias from variables that vary across states but are constant over time, $\{\mu_i\}$:
  - culture of drinking and driving
  - quality of roads
  - vintage of autos on the road
- Potential omitted variable bias from variables that vary over time but are constant across states , $\{\lambda_t\}$:
  - improvements in auto safety over time
  - changing national attitudes towards drunk driving

Traffic deaths and alcohol taxes

TABLE 10.1 Regression Analysis of the Effect of Drunk Driving Laws on Traffic Deaths

Dependent variable: traffic fatality rate (deaths per 10,000).

Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Beer tax	0.36** (0.05)	−0.66* (0.29)	−0.64+ (0.36)	−0.45 (0.30)	−0.69* (0.35)	−0.46 (0.31)	−0.93** (0.34)
Drinking age 18				0.028 (0.070)	−0.010 (0.083)		0.037 (0.102)
Drinking age 19				−0.018 (0.050)	−0.076 (0.068)		−0.065 (0.099)
Drinking age 20				0.032 (0.051)	−0.100+ (0.056)		−0.113 (0.125)
Drinking age						−0.002 (0.021)	
Mandatory jail or community service?				0.038 (0.103)	0.085 (0.112)	0.039 (0.103)	0.089 (0.164)
Average vehicle miles per driver				0.008 (0.007)	0.017 (0.011)	0.009 (0.007)	0.124 (0.049)
Unemployment rate				−0.063** (0.013)		−0.063** (0.013)	−0.091** (0.021)
Real income per capita (logarithm)				1.82** (0.64)		1.79** (0.64)	1.00 (0.68)
Years	1982–88	1982–88	1982–88	1982–88	1982–88	1982–88	1982 & 1988 only
State effects?	no	yes	yes	yes	yes	yes	yes
Time effects?	no	no	yes	yes	yes	yes	yes
Clustered standard errors?	no	yes	yes	yes	yes	yes	yes

Traffic deaths and alcohol taxes

***F*-Statistics and *p*-Values Testing Exclusion of Groups of Variables**

Time effects = 0			4.22 (0.002)	10.12 (< 0.001)	3.48 (0.006)	10.28 (< 0.001)	37.49 (< 0.001)
Drinking age coefficients = 0				0.35 (0.786)	1.41 (0.253)		0.42 (0.738)
Unemployment rate, income per capita = 0				29.62 (< 0.001)		31.96 (< 0.001)	25.20 (< 0.001)
$\bar{R}^2$	0.091	0.889	0.891	0.926	0.893	0.926	0.899

These regressions were estimated using panel data for 48 U.S. states. Regressions (1) through (6) use data for all years 1982 to 1988, and regression (7) uses data from 1982 and 1988 only. The data set is described in Appendix 10.1. Standard errors are given in parentheses under the coefficients, and *p*-values are given in parentheses under the *F*-statistics. The individual coefficient is statistically significant at the $^{+}$10%, *5%, or **1% significance level.