Phd Program in Transportation

Transport Demand Modeling

João Zeferino - 2018

zeferino@dec.uc.pt

(*adapted from Anabela Ribeiro's*)

Session 6

PANEL MODELS

General Framework

Class Structure

	Dia	Description
1	2 November 2018 (6 hours)	A - Panel data models: Main issues One way component error models Two way component error models Fixed effects Random effects *Exercises and notes on the Home Assignment* *Panel Data Models*

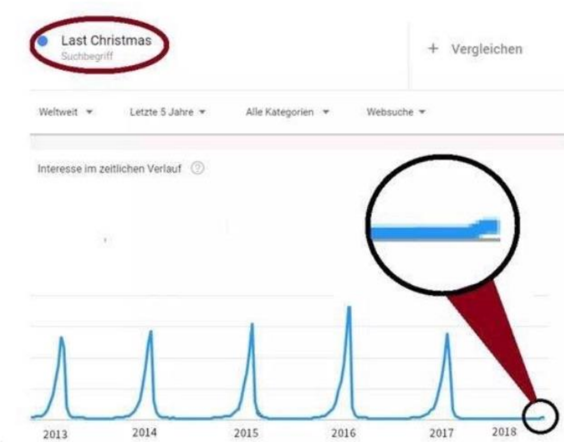
General Framework

Correlation does not imply causation!

Logical fallacies

– ***Cum hoc ergo propter hoc:*** **“***with this, therefore because of this”*

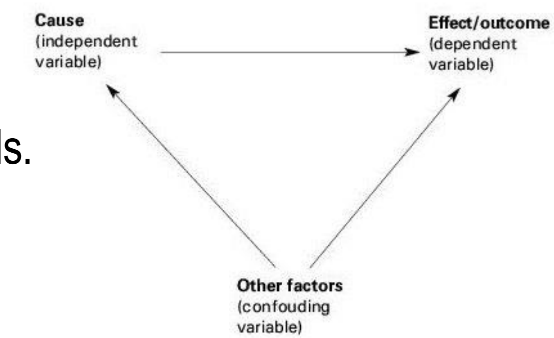
– ***Post hoc ergo propter hoc:*** **“***after this, therefore because of this*”



Example: When ice cream sales increase, swimming pool deaths also seem to increase.

Hence, eating ice cream leads to drownings in swimming pools.

(…)



Correlation is a valuable type of scientific evidence, but after confirmed as real, then every possible causative relationship must be systematically explored.

General Framework

Regression types

Time Series - Regression between a set of variables across time in a certain individual/unit - cross-sectional data in time

Cross section – Regression between a set of variables in a certain moment in time, for a set of individuals/units

Panel Data – Has both a cross-sectional and a time series dimension, where all cross section units are observed during the whole time period

Other names are pooled data, micropanel data, longitudinal data, event history analysis and cohort analysis

Panel Data Models: General Presentation

Panel Data

- Balanced
- Unbalanced (missing values)

Country	Year	Y	X1	X2	X3
1	2000	6,0	7,8	5,8	1,3
1	2001	4,6	0,6	7,9	7,8
1	2002	9,4	2,1	5,4	1,1
2	2000	9,1	1,3	6,7	4,1
2	2001	8,3	0,9	6,6	5,0
2	2002	0,6	9,8	0,4	7,2
3….	2000…	9,1…	0,2…	2,6…	6,4…

Advantages

Longitudinal or cross-sectional time-series data

Panel data are better able to study the dynamics of adjustment. Cross-sectional distributions that look relatively stable hide a multitude of changes

Spells of unemployment, job turnover, residential and income mobility are better studied with panels.

Panel data are also well suited to study the duration of economic states like unemployment and poverty, and if these panels are long enough, they can shed light on the speed of adjustments to economic policy changes.

Micropanel data gathered on individuals, firms and households may be more accurately measured than similar variables measured at the macro level

Biases resulting from aggregation over firms or individuals may be reduced or eliminated

Panel models allow controlling for aggregate effects and individual **heterogeneity**

Advantages - Example

For example, suppose that a cross-section of public transit agencies reveals that, on average, public transit subsidies are associated with 20% increased ridership.

If a homogeneous population of public transit firms is considered, this might lead to think that a firm's ridership will increase by 20% given transit subsidies

However, an alternative explanation in a sample of heterogeneous public transit firms is that the subsidies have no effect (0% increase) on four fifths of the firms, and raise ridership by 100% on one fifth of the firms. Although these competing hypotheses cannot be tested using a cross-sectional sample (**in the absence of a cross-sectional variable that "explains" this difference)**, it is possible to test between them by identifying the effect of subsidies on a **cross section of time series for the different transit systems.**

Thus, **testing for cross-sectional homogeneity is equivalent to testing the null hypothesis that these additional parameters are equal to zero**

Panel Data Models: General Presentation

Main issues – specification issues to be accounted for

Explanatory variables - Multicollinearity

Cross-sectional - Heteroscedasticity

Time-series - Correlation in the disturbance terms

Panel Data - Heterogeneity

Main issues

Multicollinearity is a state of very high intercorrelations or inter-associations among the independent variables. It is a type of disturbance in the data, and if present the statistical inferences made about the data **may not** be reliable.

Can be detected through correlation matrix, scatterplot or **variance inflation factor**

- Variance inflation factor

Step 1 - run an ordinary least square regression that has Xi as a function of all the other explanatory variables

$$X_i = \alpha_0 + \alpha_2 X_2 + \alpha_3 X_3 + \cdots + \alpha_k X_k + e$$

Step 2 - calculate the VIF factor as

$$\mathrm{VIF}_i = \frac{1}{1 - R_i^2}$$

Step 3 - Analyze the magnitude of multicollinearity by considering the size of VIF.
A rule of thumb is that if VIF > 10 then multicollinearity is high (a cutoff of 5 is also used).

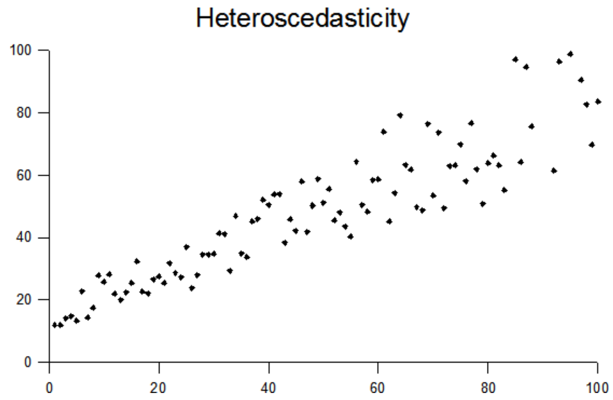
Good planning in the study design is the best remedy for **multicollinearity**.

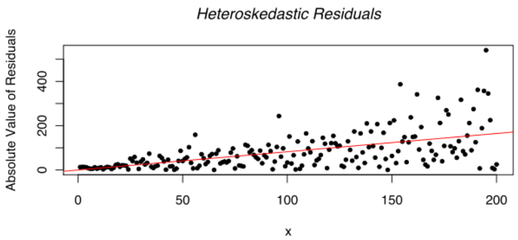
Panel Data Models: General Presentation

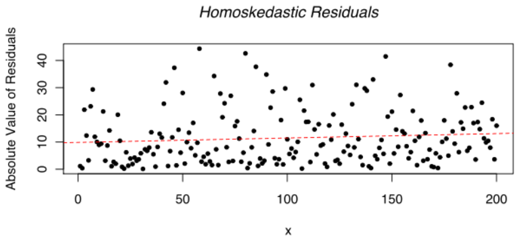
Main issues – (typically Cross-sectional)

Heteroscedasticity, which refers to the variance of the disturbances not being constant across observations.

For example, when analyzing household mobility patterns, there is often greater variation in mobility among high-income families than low-income families, possibly due to the greater flexibility in traveling allowed by higher incomes.





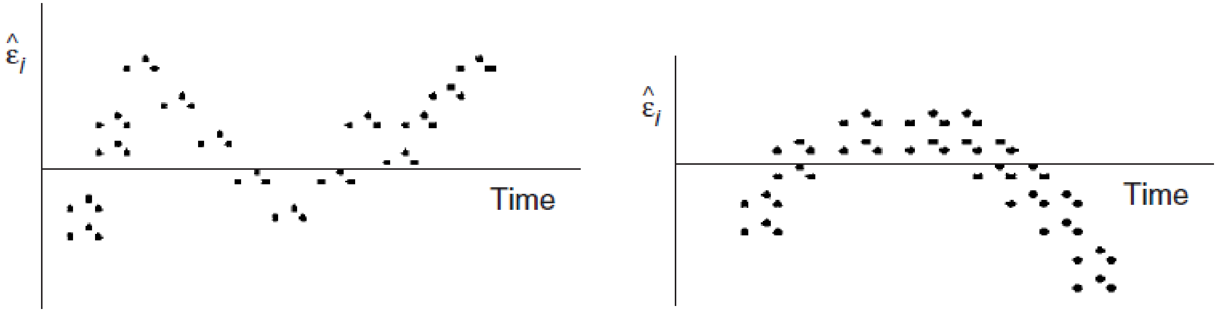


It does not affect consistency of regression but it affects its efficiency.

Main issues – (typically Time-series)

Correlation in the disturbance terms - A second issue is **serial correlation** of the disturbance terms, which occurs in time-series studies when the disturbances associated with observations in one time period are dependent on disturbances from prior time periods.

Positive correlation – the estimates of the standard errors are smaller than the true
Negative correlation – the estimates of the standard errors are bigger than the true



The regression estimates are unbiased but its errors are biased.

It does not affect consistency of regression but it affects its efficiency.

Main issues – (Panel Data)

Heterogeneity - Compared with cross-sectional or time-series data, panel data raise new specification issues that need to be considered during analysis

The most important of these is ***heterogeneity*** bias

Heterogeneity refers to the differences across cross-sectional units that may not be appropriately reflected in the available data (explanatory variable/s).

Heterogeneity - if not accounted for explicitly, may lead to model parameters that are inconsistent and/or meaningless. With panel data, it is possible to account for cross-sectional ***heterogeneity*** by introducing additional parameters into the model.

Heterogeneity in panel data can be tested using several test of hypothesis – Z Test (if the distributions follow a normal distribution).

Main issues – new specification issues

?

- Heteroscedasticity
- Correlation in the disturbance terms
- Heterogeneity

One Way and Two Way component models

Overcoming specification problems in panel data:

One-Way error component models:

variable-intercept models across individuals **or** time

Two-Way error component models:

variable-intercept models across individuals **and** time

Introduction of dummy variables in the model

Modeling specifications:

With fixed effects: effects that are in the sample

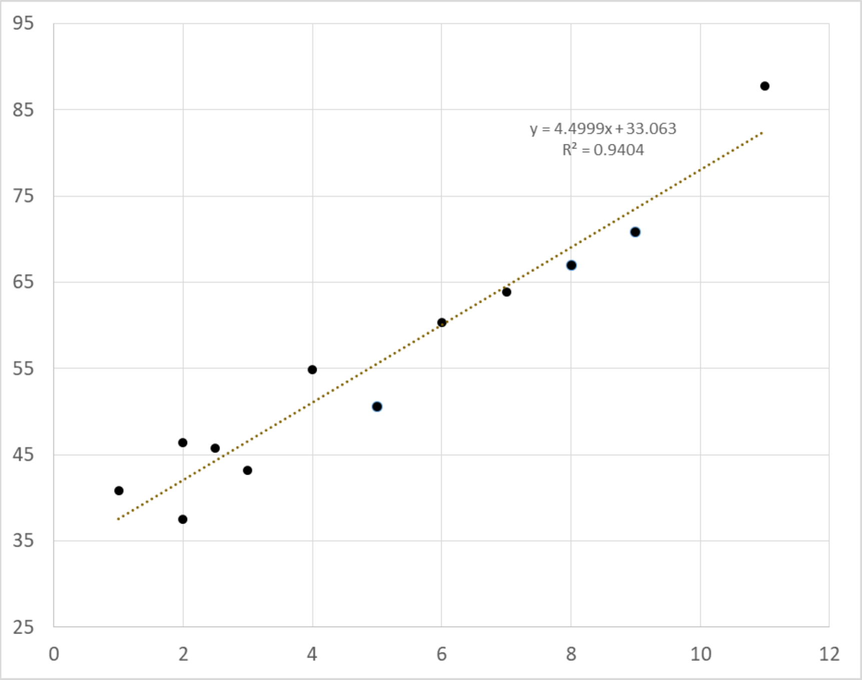
With random effect: effect randomly drawn from a population

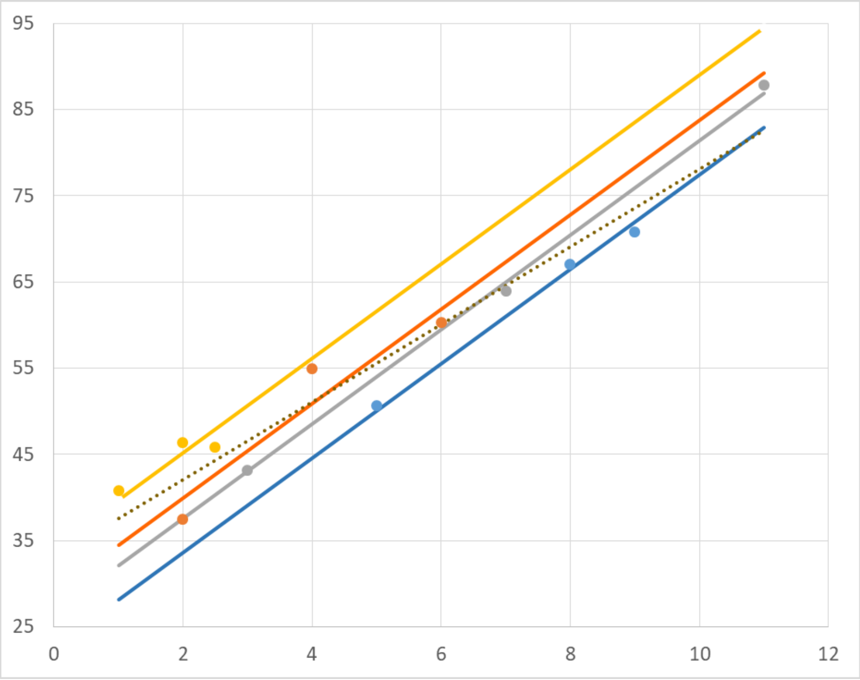
Panel Data Models: Variable-intercept Error Component Models

Variable-intercept models across individuals or/and time (**one/two-way models**) are the simplest and most straightforward models for accounting for cross-sectional ***heterogeneity*** in panel data, which arise when the null hypothesis of overall ***homogeneity*** is rejected

- The variable-intercept model assumes that the effects of omitted variables may be individually unimportant but are collectively significant, and thus is considered to be a random variable that is independent of included independent variables.
- Because heterogeneity effects are assumed to be constant for given cross-sectional units or for different cross-sectional units during one time period, they are absorbed by the intercept term as a means to account for individual or time heterogeneity

Panel Data Models:
Variable-intercept Error Component Models
Example 1





Panel Data Models:
One-Way Error Component Models

A **Panel Data regression** is generally written as:

$$Y_{it} = \alpha + X_{it}^{'}\beta + u_{it}, i = 1,\dots,n; t = 1,\dots,T \qquad (1)$$

where *i* refers to the cross-sectional units, *t* refers to the time periods, α is a scalar, β is a *Px1* vector, and X_{it} is the *ith* observation on the *Pth* explanatory variable, and u_{it} is the error term

A **One-Way Error component model** for the disturbances, which is the most commonly utilized in panel data formulation, is specified as

$$u_{it} = \mu_i + v_{ij} \qquad (2)$$

where $\boldsymbol{\mu_i}$ is the unobserved cross-sectional specific effect and $\boldsymbol{v_{it}}$ are random disturbances

Panel Data Models: One-Way Error Component Models

Fixed Effects (FE)

When the $\boldsymbol{\mu_i}$ are assumed to be fixed parameters to be estimated, and the $\boldsymbol{\upsilon_{it}}$ are random disturbances that follow the usual regression assumptions, then combining both yields the following model, where inference is conditional on the particular *n* cross-sectional units that are observed, and is thus called a ***Fixed Effects model***

$$Y_{it} = \alpha + X_{it}^{'}\beta + \mu_i + \upsilon_{it}, i = 1,\ldots,n; t = 1,\ldots,T \qquad (3)$$

on which ordinary least squares (OLS), which provide best linear unbiased estimators (BLUE), are used to obtain α, β and μ_i

Panel Data Models:
One-Way Error Component Models

Fixed Effects (FE)

FE explore the relationship between predictor and outcome variables **within** an entity (country, person, company, etc.).

Each entity has its own individual characteristics that may or may not influence the predictor variables (for example, being a male or female)

When using FE we assume that something within the individual may impact or bias the predictor or outcome variables and **we need to control for this**.
This is the rationale behind the assumption of the correlation between entity's error term and predictor variables.

FE remove the effect of those time-invariant characteristics so we can assess the net effect of the predictors on the outcome variable.

Panel Data Models: One-Way Error Component Models

Fixed Effects (FE)

Another important assumption of the FE model is that those time invariant characteristics are unique to the individual and should not be correlated with other individual characteristics.

Each entity is different therefore the entity's error term and the constant (which captures individual characteristics) should not be correlated with the others.

If the error terms are correlated, then FE is no suitable since inferences may not be correct and you need to model that relationship (**probably using random-effects**), this is the main rationale for the **Hausman test** .

Fixed-effects models are designed to study the causes of changes within a person or entity.

Panel Data Models: One-Way Error Component Models

Fixed Effects (FE)

Large Samples: when *n* is large, many indicator variables are included, and a least squares dummy variable (LSDV) estimator is obtained.

Testing for the joint significance of the included fixed effects parameters (the dummy variables) is straightforwardly conducted using the **Chow *F* test**

$$F_0 = \frac{(RRSS - URSS)/(n-1)}{URSS/(nT-n-P)} \overset{H_0}{\sim} F_{n-1,n(T-1)-P} \qquad (4)$$

where RRSS are the restricted residual sums of squares from OLS on the pooled model and URSS are the unrestricted residual sums of squares from the LSDV regression

- **If the null is true (no fixed effects) then the correct procedure is to estimate a single regression from all the data.**
- **If the null is not true (a significant value for F) than we have to account for fixed effects**

Panel Data Models: One-Way Error Component Models

Random Effects (RE)

The fixed effects specification suffers from an obvious shortcoming in that it requires the estimation of are too many parameters and the associated loss of degrees of freedom

This can be avoided if the μ_i can be assumed random. Unlike the fixed effects model where inferences are conditional on the particular cross-sectional units sampled, an alternative formulation, called the ***Random Effects model***.

$$\mu_i \sim IID(0,\sigma_\mu^2), \quad \upsilon_{it} \sim IID(0,\sigma_\upsilon^2) \qquad (5)$$

- The μ_i and υ_{it} are independent, and X_{it} are independent of the μ_i and υ_{it} for all i, t

Panel Data Models: One-Way Error Component Models

Random Effects (RE)

The random effects model is an appropriate specification if we are drawing *n* individuals randomly from a large population

> This is usually the case for household panel studies. Care is taken in the design of the panel to make it "representative" of the population we are trying to make inferences about

The individual effect is characterised as random and inference pertains to the population from which this sample was randomly drawn.

Furthermore, it can be shown that a random effects specification implies a **homoscedastic** disturbances variance, $VAR(u_{it}) = \sigma_{\mu}^{2} + \sigma_{\upsilon}^{2}$ for all *i, t, and* serial correlation only for disturbances of the same cross-sectional unit.

Panel Data Models: One-Way Error Component Models

Fixed vs. Random Effects

This is not as easy a choice as it might seem. In fact, the fixed versus random effects issue has generated a hot debate in the biometrics and statistics literature which has spilled over into the panel data econometrics literature….

Most commonly accepted:

The most important issue when considering these alternative specifications is the context of the analysis.

The essential difference between these two modeling specifications is:

- whether the inferences are conditional on the effects that are in the sample, i.e. inferences from the estimated model are confined to the effects in the model (the fixed effects model is appropriate);
- whether the inferences are made unconditionally with respect to the population of effects, i.e. inferences are made about a population of effects, from which the effects in the model are a random sample (suited for the random effects model).

Panel Data Models: One-Way Error Component Models

Fixed vs. Random Effects (Lagrange and Hausman Tests)

The FE model has a considerable virtue in that it does not assume that the individual effects μ_i are uncorrelated with the regressors X_{it} as is assumed by the random-effects model.

In fact, the RE model may be biased and inconsistent due to omitted variables (Hausman and Taylor, 1981; Chamberlain, 1978).

With the intent of identifying potential correlation between the individual effects and the regressors, the **Hausman's Test** (Hausman, 1978) examines the null hypothesis of no correlation between the individual effects and X_{it}

A rejection of the null hypothesis of no correlation suggests the possible inconsistency of the RE model and the possible preference for a FE specification (test value significant).

Panel Data Models:
One-Way Error Component Models

Fixed vs. Random Effects (Lagrange and Hausman Tests)

Breusch and Pagan's **Lagrange multiplier** statistic, is used to test the null hypothesis that there are no group effects in the RE model. Arguably, a rejection of the null hypothesis is as likely to be due to the presence of fixed effects. The statistic is computed from the ordinary least squares residuals from a pooled regression.

- **Large values of LM favor the FE and RE models over the classical model with no common effects.**

A second statistic is **Hausman's Chi Squared** statistic for testing whether the GLS estimator is an appropriate alternative to the LSDV estimator. Computation of the Hausman statistic requires estimates of both the random and fixed effects models.

- **Large values of H weigh in favor of the FE model.**

Panel Data Models: Two-Way Error Component Models

Fixed Effects

The disturbances presented in (2) are further generalized to include time-specific effects. This generalization is called a **Two-Way Error components model**, whose disturbances are written as

$$u_{it} = \mu_i + \lambda_t + v_{ij}, i = 1,\ldots,n; t = 1,\ldots,T \qquad (10)$$

where μ_i is the unobserved cross-sectional specific effect, $\boldsymbol{\lambda_t}$ denotes the unobservable time effects, and υ_{it} are random disturbances. Here $\boldsymbol{\lambda_t}$ is individual invariant and accounts for any time-specific effect that is not included in the regression

When the μ_i and λ_t are assumed to be fixed parameters to be estimated and are random disturbances that follow the usual regression assumptions, combining (1) and (10) yields a model where inferences are conditional on the particular *n cross-sectional units and are* to be made over the specific time period of observation

$$u_{it} = \mu_i + v_{ij}$$

where μ_i is the unobserved cross-sectional specific effect and υ_{it} are random disturbances

Panel Data Models: Two-Way Error Component Models

Fixed Effects

This model is called a **two-way fixed effects error component model** and is given as

$$Y_{it} = \alpha + X_{it}^{'}\beta + \mu_i + \lambda_t + \upsilon_{it}, i = 1,\ldots,n; t = 1,\ldots,T \qquad (11)$$

where X_{it} are assumed independent of the υ_{it} for all *i, t.*

Inference for this ***two-way fixed-effects model*** is conditional on the particular *n individuals* and over the *T time periods of observation.* Similar to the one-way fixed-effects model, the computational difficulties involved with obtaining the OLS estimates for *β*.

Panel Data Models: Two-Way Error Component Models

Fixed Effects

Testing for the joint significance of the included cross-sectional and time period fixed effects parameters (the dummy variables) is straightforwardly computed using an *F test*

$$F_0 = \frac{(RRSS - URSS)/(n+T-2)}{URSS/(n-1)(T-1)-P} \overset{H_0}{\sim} F_{(n+T-2),(n-1)(T-1)-P} \quad (12)$$

where RRSS are the restricted residual sums of squares from OLS on the pooled model and URSS are the unrestricted residual sums of squares from the regression using the within transformation of Wallace and Hussain (1969)

Panel Data Models: Two-Way Error Component Models

Random Effects

Similar to the one-way error component model case, if both the μ_i and λ_t are random with

$$\mu_i \sim IDD(0,\sigma_\mu^2),\ \lambda_t \sim (0,\sigma_\lambda^2),\ \upsilon_{it} \sim IDD(0,\sigma_\upsilon^2) \qquad (13)$$

The μ_i, λ_t and υ_{it} are independent, and X_{it} are independent of the μ_i, λ_t and υ_{it} for all i, t.

This formulation is called the ***Random-Effects model***

Furthermore, it can be shown that a random effects specification implies a homoscedastic disturbances variance, $VAR(u_{it}) = \sigma_\mu^2 + \sigma_\lambda^2 + \sigma_\upsilon^2$ for all *i, t,* and serial correlation only *for* disturbances of the same cross-sectional unit.

Panel Data Models: Two-Way Error Component Models

Random Effects

Estimation of the two-way random-effects model is typically accomplished using the GLS estimators of Wallace and Hussain (1969) or by using maximum likelihood estimation (Baltagi and Li, 1992)

For this model specification, Breusch and Pagan (1980) derived a **Lagrange-Multiplier** test for the null hypothesis;

$H_0 = \sigma_\mu^2 = \sigma_\lambda^2 = 0$ this test is based on the normality of the disturbances

If the **Two-Way Error component model** specification is **significant**, the quality of its estimates should be always **better** than a **One-way Error Component model** in Fixed Effects or Random Effects model.

Panel Data Models: Example 2

Nlogit with Grunfeld Investment Equation

$$I_{it} = \alpha + \beta_1 F_{it} + \beta_2 C_{it} + v_{it} \qquad (16)$$

where I_{it} t denotes real gross investment for firm *i* in year *t*, F_{it} is the real value of the firm (shares outstanding) and C_{it} is the real value of the capital stock. These panel data consist of 10 large US manufacturing firms over 20 years, 1935–54

Table 1

Variable Abbreviation	Variable Description
invest	Gross investment, defined as additions to plant and equipment plus maintenance and repairs in millions of dollars deflated by the implicit price deflator of producers' durable equipment (base 1947)
value	Market value of the firm, defined as the price of common shares at December 31 (base 1947)
capital	Stock of plant and equipment, defined as the accumulated sum of net additions to plant and equipment deflated by the implicit price deflator for producers' durable equipment (base 1947)
firm	General Motors (GM), US Steel (US), General Electric (GE), Chrysler (CH), Atlantic Rening (AR), IBM, Union Oil (UO), Westinghouse (WH), Goodyear (GY), Diamond Match (DM), American Steel (AS)
year	Year of data
firmcod	Numeric code that identifies each firm

Panel Data Models: Example 2

Ordinary Least Squares Model Estimates
Gross Investment

```
+----------------------------------------------------+
| Ordinary    least squares regression               |
| Model was estimated Oct 15, 2010 at 09:10:35AM     |
| LHS=INVEST   Mean                 =   133.3119     |
|              Standard deviation   =   210.5872     |
| WTS=none     Number of observs.   =        220     |
| Model size   Parameters           =          3     |
|              Degrees of freedom   =        217     |
| Residuals    Sum of squares       =   1768678.     |
|              Standard error of e  =   90.28063     |
| Fit          R-squared            =   .8178870     |
|              Adjusted R-squared   =   .8162086     |
| Model test   F[  2,   217] (prob) = 487.28 (.0000) |
| Diagnostic   Log likelihood       =  -1301.299     |
|              Restricted(b=0)      =  -1488.643     |
|              Chi-sq [   2]  (prob) = 374.69 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =   9.019390     |
|              Akaike Info. Criter. =   9.019388     |
| Autocorrel   Durbin-Watson Stat.  =   .3566636     |
|              Rho = cor[e,e(-1)]   =   .8216682     |
+----------------------------------------------------+

+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -38.4100540      8.41337092     -4.565   .0000
 VALUE   |     .11453436       .00551883     20.753   .0000   988.577805
 CAPITAL |     .22751413       .02422825      9.390   .0000   257.108541
```

Panel Data Models: Example 2

Fixed Effects Panel Data Model Estimates (One-way Error) Gross Investment

```
+----------------------------------------------------+
| Least Squares with Group Dummy Variables           |
| Ordinary    least squares regression               |
| Model was estimated Nov 05, 2012 at 06:09:19PM     |
| LHS=INVEST   Mean                 =   133.3119     |
|              Standard deviation   =   210.5872     |
| WTS=none     Number of observs.   =        220     |
| Model size   Parameters           =         13     |
|              Degrees of freedom   =        207     |
| Residuals    Sum of squares       =   523718.7     |
|              Standard error of e  =   50.29952     |
| Fit          R-squared            =   .9460750     |
|              Adjusted R-squared   =   .9429489     |
| Model test   F[ 12,   207] (prob) = 302.64 (.0000) |
| Diagnostic   Log likelihood       =  -1167.426     |
|              Restricted(b=0)      =  -1488.643     |
|              Chi-sq [ 12]  (prob) = 642.44 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =   7.893402     |
|              Akaike Info. Criter. =   7.893264     |
| Estd. Autocorrelation of e(i,t)      .549274       |
+----------------------------------------------------+
```

Estimated Fixed Effects

Group	Coefficient	Standard Error	t-ratio
1	-70.29907	47.37535	-1.48387
2	101.90474	23.76871	4.28735
3	-235.56939	23.28607	-10.11632
4	-27.80911	13.41858	-2.07243
5	-114.60252	13.50246	-8.48753
6	-23.16020	12.07589	-1.91789
7	-66.54422	12.24204	-5.43572
8	-57.54649	13.33791	-4.31451
9	-87.21454	12.28873	-7.09712
10	-6.56803	11.27363	-.58260
11	-20.57820	11.29779	-1.82144

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
VALUE	.11012912	.01129984	9.746	.0000	988.577805
CAPITAL	.31003344	.01654048	18.744	.0000	257.108541

Significant F value – account for effects

Panel Data Models: Example 2

Fixed Effects Panel Data Model Estimates (One-way Error) Gross Investment

```
|             Test Statistics for the Classical Model                 |
+---------------------------------------------------------------------+
|       Model             Log-Likelihood  Sum of Squares    R-squared |
|(1)  Constant term only    -1488.64328  .9711984910D+07    .0000000 |
|(2)  Group effects only    -1327.50901  .2244546885D+07    .7688890 |
|(3)  X - variables only    -1301.29920  .1768678402D+07    .8178870 |
|(4)  X and group effects   -1167.42554  .5237186622D+06    .9460750 |
+---------------------------------------------------------------------+
|                           Hypothesis Tests                          |
|         Likelihood Ratio Test             F Tests                   |
|         Chi-squared   d.f.  Prob.        F     num. denom.   P value |
|(2) vs (1)   322.269     10  .00000   69.533    10     209    .00000 |
|(3) vs (1)   374.688      2  .00000  487.284     2     217    .00000 |
|(4) vs (1)   642.435     12  .00000  302.639    12     207    .00000 |
|(4) vs (2)   320.167      2  .00000  340.079     2     207    .00000 |
|(4) vs (3)   267.747     10  .00000   49.207    10     207    .00000 |
+---------------------------------------------------------------------+
```

Consideration of group effects improve regression

Panel Data Models: Example 2

Fixed Effects Panel Data Model Estimates (Two-way Error) Gross Investment

```
+----------------------------------------------------+
| Least Squares with Group and Period Effects        |
| Ordinary    least squares regression               |
| Model was estimated Nov 07, 2012 at 01:14:13AM     |
| LHS=INVEST   Mean                 =   133.3119     |
|              Standard deviation   =   210.5872     |
| WTS=none     Number of observs.   =        220     |
| Model size   Parameters           =         32     |
|              Degrees of freedom   =        188     |
| Residuals    Sum of squares       =   459399.9     |
|              Standard error of e  =   49.43295     |
| Fit          R-squared            =   .9526976     |
|              Adjusted R-squared   =   .9448978     |
| Model test   F[ 31,   188] (prob) = 122.14 (.0000) |
| Diagnostic   Log likelihood       =  -1153.012     |
|              Restricted(b=0)      =  -1488.643     |
|              Chi-sq [ 31]  (prob) = 671.26 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =   7.937036     |
|              Akaike Info. Criter. =   7.934958     |
| Estd. Autocorrelation of e(i,t)     .569917        |
+----------------------------------------------------+
```

μ_i

Estimated Fixed Effects - Full sets of effects, normalized t

Group	Coefficient	Standard Error	t-ratio
1	-53.14755	42.84415	-1.24049
2	149.17125	16.34968	9.12380
3	-192.46291	15.79790	-12.18282
4	35.02275	11.28051	3.10471
5	-63.87868	16.09003	-3.97008
6	42.16480	12.69795	3.32060
7	-8.17149	15.43398	-.52945
8	6.90560	11.49035	.60099
9	-29.34076	13.70930	-2.14021
10	65.11490	15.55637	4.18574
11	48.62208	15.58761	3.11928

Estimated Fixed Effects - Full sets of effects, normalized t

Period	Coefficient	Standard Error	t-ratio
1	41.85916	15.38287	2.72115
2	24.89993	15.03606	1.65601
3	5.48352	15.46655	.35454
4	6.23543	14.87055	.41931
5	-21.24024	14.71160	-1.44377
6	2.03439	14.74346	.13799
7	25.37139	14.65503	1.73124
8	23.85983	14.71735	1.62120
9	4.08671	14.58562	.28019
10	3.53909	14.57617	.24280
11	-7.68033	14.59933	-.52607
12	14.10477	14.64576	.96306
13	6.98162	14.67759	.47567
14	3.52843	14.82120	.23807
15	-23.34160	14.88759	-1.56786
16	-25.52857	14.85283	-1.71877
17	-12.97548	14.79992	-.87673
18	-14.62988	15.04933	-.97213
19	-16.65342	15.74591	-1.05763
20	-39.93475	16.15229	-2.47239

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
VALUE	.11668113	.01293303	9.022	.0000	988.577805
CAPITAL	.35143569	.02104860	16.696	.0000	257.108541
Constant	-72.3935959	12.7315764	-5.686	.0000	

λ_t

Fixed Effects Panel Data Model Estimates (Two-way Error)
Gross Investment

Test Statistics for the Classical Model

	Model	Log-Likelihood	Sum of Squares	R-squared
(1)	Constant term only	-1488.64328	.9711984910D+07	.0000000
(2)	Group effects only	-1327.50901	.2244546885D+07	.7688890
(3)	X - variables only	-1301.29920	.1768678402D+07	.8178870
(4)	X and group effects	-1167.42554	.5237186622D+06	.9460750
(5)	X ind.&time effects	-1153.01185	.4593999310D+06	.9526976

Hypothesis Tests

	Likelihood Ratio Test			F Tests			
	Chi-squared	d.f.	Prob.	F	num.	denom.	P value
(2) vs (1)	322.269	10	.00000	69.533	10	209	.00000
(3) vs (1)	374.688	2	.00000	487.284	2	217	.00000
(4) vs (1)	642.435	12	.00000	302.639	12	207	.00000
(4) vs (2)	320.167	2	.00000	340.079	2	207	.00000
(4) vs (3)	267.747	10	.00000	49.207	10	207	.00000
(5) vs (4)	28.827	19	.06875	1.385	19	188	.13801
(5) vs (3)	296.575	30	.00000	17.860	30	188	.00000

Consideration of group effects and time effects improve regression

Panel Data Models: Example 2

Random Effects Panel Data Model Estimates (One-way Error) Gross Investment

```
+----------------------------------------------------+
| Least Squares with Group Dummy Variables           |
| Ordinary    least squares regression               |
| Model was estimated Oct 15, 2010 at 09:35:46AM     |
| LHS=INVEST   Mean                 =   133.3119     |
|              Standard deviation   =   210.5872     |
| WTS=none     Number of observs.   =        220     |
| Model size   Parameters           =         13     |
|              Degrees of freedom   =        207     |
| Residuals    Sum of squares       =   523718.7     |
|              Standard error of e  =   50.29952     |
| Fit          R-squared            =   .9460750     |
|              Adjusted R-squared   =   .9429489     |
| Model test   F[ 12,   207] (prob) = 302.64 (.0000) |
| Diagnostic   Log likelihood       =  -1167.426     |
|              Restricted(b=0)      =  -1488.643     |
|              Chi-sq [ 12]  (prob) = 642.44 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =   7.893402     |
|              Akaike Info. Criter. =   7.893264     |
| Estd. Autocorrelation of e(i,t)     .549274        |
+----------------------------------------------------+
```

```
+-----------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)        |
| Estimates:  Var[e]              =   .253004D+04     |
|             Var[u]              =   .562055D+04     |
|             Corr[v(i,t),v(i,s)] =   .689588         |
| Lagrange Multiplier Test vs. Model (3) =  874.75    |
| ( 1 df, prob value =  .000000)                      |
| (High values of LM favor FEM/REM over CR model.)    |
| Baltagi-Li form of LM Statistic =         874.75    |
| Fixed vs. Random Effects (Hausman)    =    2.89     |
| ( 2 df, prob value =  .235736)                      |
| (High (low) values of H favor FEM (REM).)           |
|             Sum of Squares              .186431D+07 |
|             R-squared                   .808040D+00 |
+-----------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
VALUE	.10924931	.00978586	11.164	.0000	988.577805
CAPITAL	.30782652	.01634860	18.829	.0000	257.108541
Constant	-53.8343750	24.5716850	-2.191	.0285	

Random Effects Panel Data Model Estimates (Two-way Error)
Gross Investment

```
+----------------------------------------------------------+
| Least Squares with Group and Period Effects              |
| Ordinary    least squares regression                     |
| Model was estimated Oct 15, 2010 at 10:10:47AM           |
| LHS=INVEST   Mean                   =    133.3119        |
|              Standard deviation     =    210.5872        |
| WTS=none     Number of observs.     =         220        |
| Model size   Parameters             =          32        |
|              Degrees of freedom     =         188        |
| Residuals    Sum of squares         =    459399.9        |
|              Standard error of e    =    49.43295        |
| Fit          R-squared              =    .9526976        |
|              Adjusted R-squared     =    .9448978        |
| Model test   F[ 31,   188] (prob)   = 122.14 (.0000)     |
| Diagnostic   Log likelihood         =   -1153.012        |
|              Restricted(b=0)        =   -1488.643        |
|              Chi-sq [ 31]  (prob)   = 671.26 (.0000)     |
| Info criter. LogAmemiya Prd. Crt.   =    7.937036        |
|              Akaike Info. Criter.   =    7.934958        |
| Estd. Autocorrelation of e(i,t)        .569917           |
+----------------------------------------------------------+
```

```
+----------------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i) + w(t)      |
| Estimates:  Var[e]              =    .244362D+04         |
|             Var[u]              =    .447819D+04         |
|             Corr[v(i,t),v(i,s)] =    .646968             |
|             Var[w]              =    .122879D+04         |
|             Corr[v(i,t),v(j,t)] =    .334600             |
| Lagrange Multiplier Test vs. Model (3) =  881.07         |
| ( 2 df, prob value =  .000000)                           |
| (High values of LM favor FEM/REM over CR model.)         |
| Fixed vs. Random Effects (Hausman)     =    5.72         |
| ( 2 df, prob value =  .057275)                           |
| (High (low) values of H favor FEM (REM).)                |
|             Sum of Squares            .186431D+07        |
|             R-squared                 .808040D+00        |
+----------------------------------------------------------+
```

```
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 VALUE   |     .11107050       .01021747    10.871   .0000   988.577805
 CAPITAL |     .33700305       .01975302    17.061   .0000   257.108541
 Constant|  -63.1362933      23.9608695     -2.635   .0084
```

Panel Data Models: Example 2

```
+--------------------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i) + w(t)          |
| Estimates:  Var[e]              =    .244362D+04             |
|             Var[u]              =    .447819D+04             |
|             Corr[v(i,t),v(i,s)] =    .646968                 |
|             Var[w]              =    .122879D+04             |
|             Corr[v(i,t),v(j,t)] =    .334600                 |
| Lagrange Multiplier Test vs. Model (3) =  881.07             |
| ( 2 df, prob value =   .000000)                              |
| (High values of LM favor FEM/REM over CR model.)             |
| Fixed vs. Random Effects (Hausman)     =    5.72             |
| ( 2 df, prob value =   .057275)                              |
| (High (low) values of H favor FEM (REM).)                    |
|             Sum of Squares           .186431D+07             |
|             R-squared                .808040D+00             |
+--------------------------------------------------------------+
```

Large values of LM favor the effects model over the classical model with no common effects.

Large values of H weigh in favor of the fixed effects model.

A large value of the LM statistic in the presence of a small H statistic (as in our application) argues in favor of the random effects model.

***"LIMDEP*, Version 9, Student , Reference Guide"**

Panel Data Models: Example 3

NLOGIT with Impact of safety belt use on road accident fatalities

The effectiveness of safety belt use in reducing motor vehicle–related fatalities has been the subject of much research interest in the past few years. To investigate the hypothesised relationship between various exogenous factors including seat belt usage rates and traffic fatalities

Derrig et al. (2002) compiled a panel data set of demographic, socioeconomic, political, insurance, and roadway variables for all 50 US states over a 14-year period (1983 through 1996) for a total of 700 observations (state-years)

This data set was subsequently enriched with additional information by R. Noland of Imperial College, U.K., and is used in this analysis

Panel Data Models: Example 3

US States Map

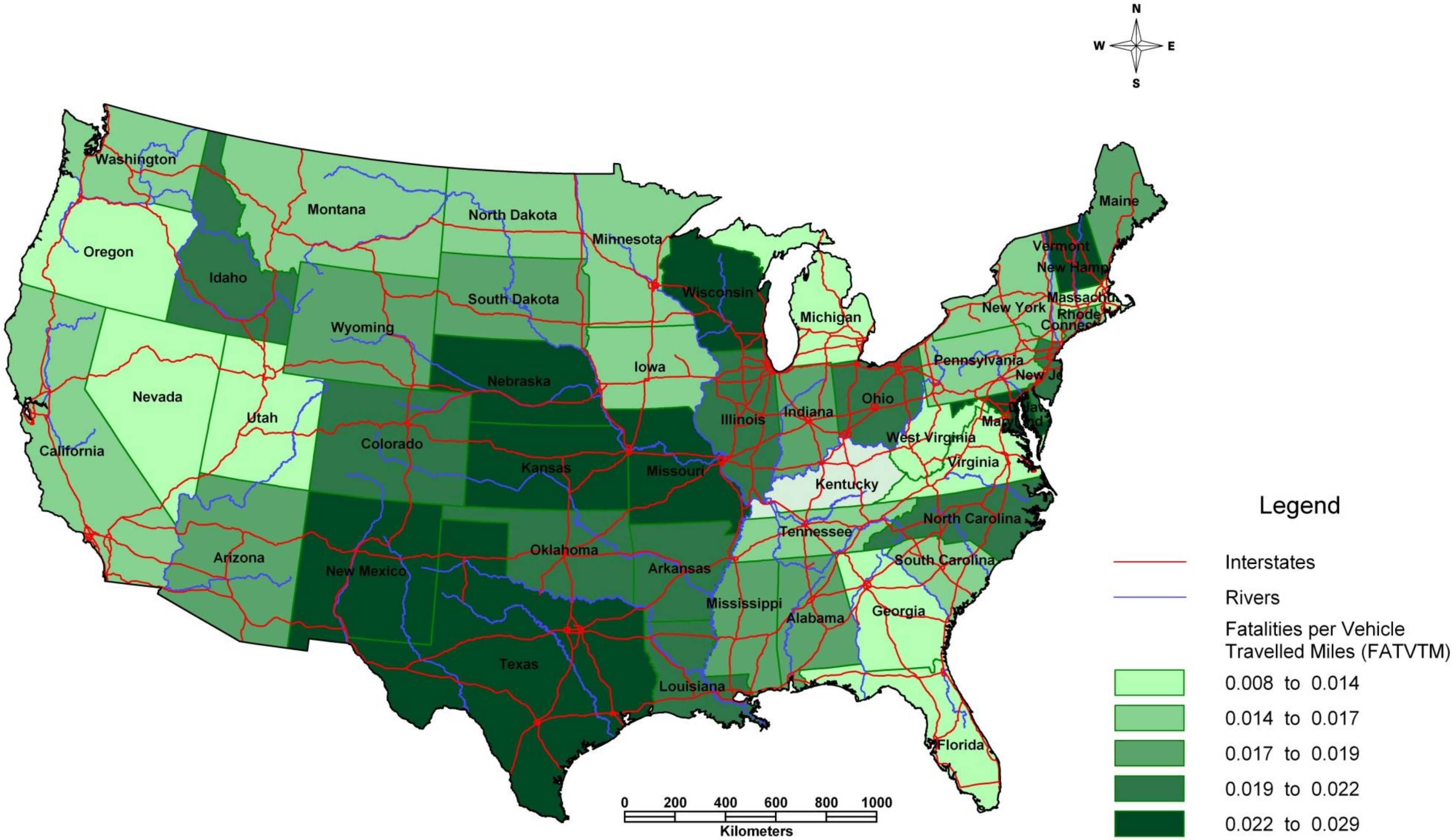
Panel Data Models: Example 3

Table 1

Variable Abbreviation	Variable Description
STATE	State
YEAR	Year
STATENUM	State ID number
DEATHS	Number of traffic-related deaths
INJURED	Number of traffic-related injuries
PRIMLAW	Primary seat belt law
SECLAW	Secondary seat belt law
TOTVMT	Total VMT
PI92	Per capita income 1992 $
POP	Total population
HOSPAREA	Hospitals per square mile
ETHANOL	Alcohol consumption total ethanol by volume
ETHPC	Per capita ethanol consumption
SEATBELT	Percent seat belt use
PERC1524	Percent population 15–24
PERC2544	Percent population 25–44
PERC4564	Percent population 45–64
PERC65P	Percent population 65 plus
PERC75P	Percent population 75 plus
INMILES	Total lane miles (excluding local roads)
PRECIP	Annual precipitation in inches
EDUC	Percent bachelors degrees
PERCRINT	Percentage of vehicle miles driven in rural interstate highway miles

Panel Data Models: Example 3

Sample - Fatalities per vehicle travelled mile (FATVTM) - 1995



Panel Data Models: Example 3

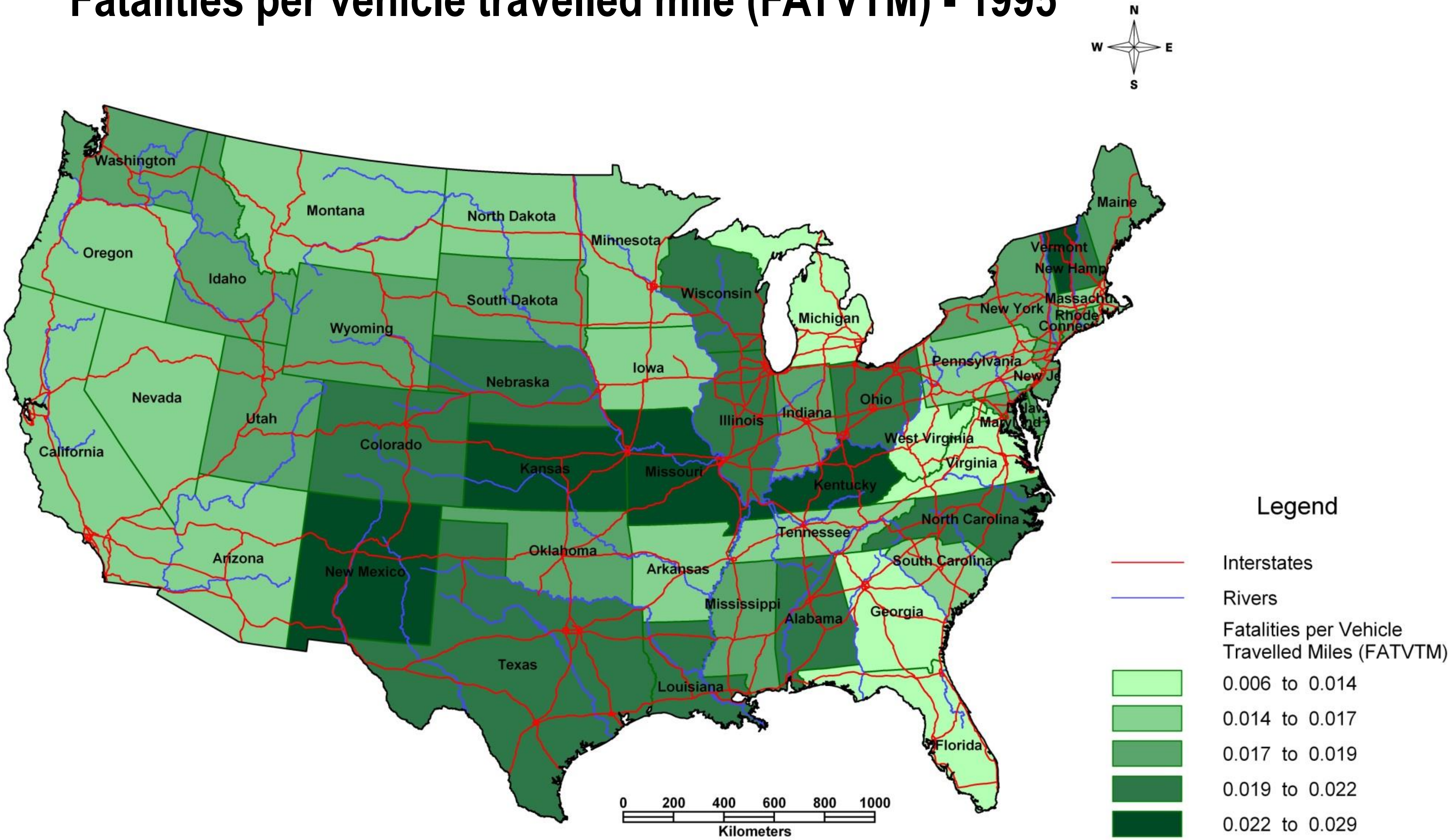
Ordinary Least Squares Model Estimates
Fatalities per vehicle travelled mile (FATVTM)

```
+----------------------------------------------------+
| Ordinary    least squares regression               |
| Model was estimated Oct 14, 2010 at 08:40:34AM     |
| LHS=FATVMT    Mean                =   .1866178E-01 |
|               Standard deviation  =   .4588015E-02 |
| WTS=none      Number of observs.  =        350     |
| Model size    Parameters          =         13     |
|               Degrees of freedom  =        337     |
| Residuals     Sum of squares      =   .2366157E-02 |
|               Standard error of e =   .2649762E-02 |
| Fit           R-squared           =   .6779165     |
|               Adjusted R-squared  =   .6664477     |
| Model test    F[ 12,   337] (prob) =  59.11 (.0000) |
| Diagnostic    Log likelihood      =   1586.645     |
|               Restricted(b=0)     =   1388.380     |
|               Chi-sq [ 12]  (prob) = 396.53 (.0000) |
| Info criter.  LogAmemiya Prd. Crt. =  -11.83010    |
|               Akaike Info. Criter. =  -11.83014    |
| Autocorrel    Durbin-Watson Stat.  =   .7254229    |
|               Rho = cor[e,e(-1)]   =   .6372885    |
+----------------------------------------------------+
```

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
Constant	1.99567406	.19989676	9.984	.0000	
YEAR	-.00093815	.999589D-04	-9.385	.0000	1993.00000
PI92	-.461127D-06	.103842D-06	-4.441	.0000	20460.8038
POP	.919607D-10	.449527D-10	2.046	.0416	.513867D+07
HOSPAREA	-.36752679	.10279119	-3.575	.0004	.00258178
ETHPC	1.97828609	.46514010	4.253	.0000	.00184147
SEATBELT	-.00142761	.00144334	-.989	.3233	.58400000
PERC1524	-.16843079	.02807833	-5.999	.0000	.14278826
PERC2544	-.22129246	.01923444	-11.505	.0000	.31660983
PERC4564	.05561139	.01714655	3.243	.0013	.19345771
PERC75P	-.34256374	.02808903	-12.196	.0000	.05521571
LNMILES	-.212927D-08	.230052D-08	-.926	.3553	162439.608
PRECIP	.653803D-04	.129058D-04	5.066	.0000	37.1971429

Panel Data Models: Example 3

Ordinary Least Squares Model Estimates
Fatalities per vehicle travelled mile (FATVTM) - 1995



Panel Data Models: Example 3

Fixed Effects Panel Data Model Estimates (One-way Error)
Fatalities per vehicle travelled mile (FATVTM)

```
+-----------------------------------------------------------+
| Least Squares with Group Dummy Variables                  |
| Ordinary    least squares regression                      |
| Model was estimated Oct 14, 2010 at 09:01:11AM            |
| LHS=FATVMT   Mean                 =   .1866178E-01        |
|              Standard deviation   =   .4588015E-02        |
| WTS=none     Number of observs.   =        350            |
| Model size   Parameters           =         61            |
|              Degrees of freedom   =        289            |
| Residuals    Sum of squares       =   .5945459E-03        |
|              Standard error of e  =   .1434312E-02        |
| Fit          R-squared            =   .9190699             |
|              Adjusted R-squared   =   .9022677             |
| Model test   F[ 60,   289] (prob) =  54.70 (.0000)        |
| Diagnostic   Log likelihood       =   1828.360            |
|              Restricted(b=0)      =   1388.380            |
|              Chi-sq [ 60]  (prob) = 879.96 (.0000)        |
| Info criter. LogAmemiya Prd. Crt. =  -12.93348            |
|              Akaike Info. Criter. =  -12.93707            |
| Estd. Autocorrelation of e(i,t)      .069813              |
| White/Hetero. corrected covariance matrix used.           |
+-----------------------------------------------------------+
```

Estimated Fixed Effects

Group	Coefficient	Standard Error	t-ratio
1	1.25732	.36450	3.44944
2	1.25310	.36657	3.41845
3	1.25095	.36598	3.41805
4	1.25870	.36472	3.45116
5	1.22224	.36246	3.37209
6	1.24874	.36523	3.41908
7	1.23164	.36304	3.39261
8	1.24125	.36592	3.39216
9	1.24002	.36399	3.40673
10	1.25091	.36461	3.43082
11	1.24188	.36529	3.39970
12	1.25548	.36539	3.43597
13	1.24269	.36327	3.42080
14	1.24707	.36417	3.42438
15	1.25450	.36357	3.45049
16	1.25665	.36317	3.46022
17	1.25351	.36477	3.43647
18	1.25396	.36604	3.42577
19	1.24160	.36598	3.39257
20	1.23887	.36435	3.40027

μ_i

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
YEAR	-.00063812	.00018494	-3.450	.0006	1993.00000
PI92	.567135D-06	.317186D-06	1.788	.0747	20460.8038
POP	.143018D-08	.547381D-09	2.613	.0094	.513867D+07
HOSPAREA	-.29725499	.50333761	-.591	.5552	.00258178
ETHPC	2.24840288	3.37243604	.667	.5054	.00184147
SEATBELT	-.00498158	.00148673	-3.351	.0009	.58400000
PERC1524	.07483185	.02716235	2.755	.0062	.14278826
PERC2544	.03984851	.02650378	1.504	.1336	.31660983
PERC75P	.20873172	.12617346	1.654	.0990	.05521571
LNMILES	-.643958D-07	.293653D-07	-2.193	.0290	162439.608
PRECIP	-.188007D-04	.118190D-04	-1.591	.1126	37.1971429

Fixed Effects Panel Data Model Estimates (One-way Error)
Fatalities per vehicle travelled mile (FATVTM)

Test Statistics for the Classical Model

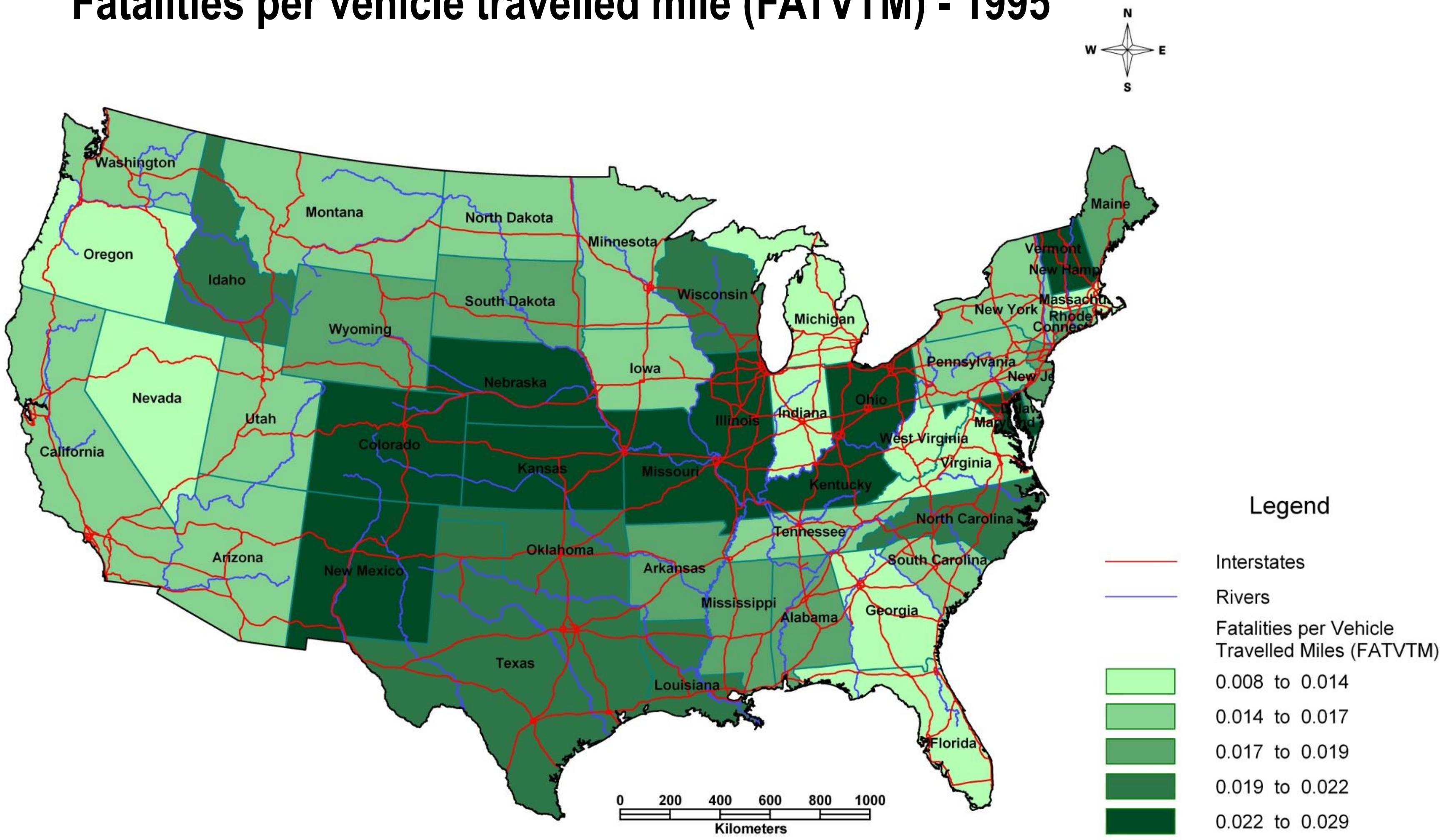
	Model	Log-Likelihood	Sum of Squares	R-squared
(1)	Constant term only	1388.37998	.7346408134D-02	.0000000
(2)	Group effects only	1710.41141	.1166528091D-02	.8412111
(3)	X - variables only	1581.26641	.2440012876D-02	.6678631
(4)	X and group effects	1828.35953	.5945458916D-03	.9190699

Hypothesis Tests

	Likelihood Ratio Test			F Tests			
	Chi-squared	d.f.	Prob.	F	num.	denom.	P value
(2) vs (1)	644.063	49	.00000	32.435	49	300	.00000
(3) vs (1)	385.773	11	.00000	61.787	11	338	.00000
(4) vs (1)	879.959	60	.00000	54.700	60	289	.00000
(4) vs (2)	235.896	11	.00000	25.276	11	289	.00000
(4) vs (3)	494.186	49	.00000	18.307	49	289	.00000

Panel Data Models: Example 3

Fixed Effects Panel Data Model Estimates (One-way Error)
Fatalities per vehicle travelled mile (FATVTM) - 1995



Fixed Effects Panel Data Model Estimates (Two-way Error)
Fatalities per vehicle travelled mile (FATVTM)

```
+----------------------------------------------------+
| Least Squares with Group and Period Effects        |
| Ordinary    least squares regression               |
| Model was estimated Oct 14, 2010 at 09:16:28AM     |
| LHS=FATVMT   Mean                 =   .1866178E-01 |
|              Standard deviation   =   .4588015E-02 |
| WTS=none     Number of observs.   =        350     |
| Model size   Parameters           =         67     |
|              Degrees of freedom   =        283     |
| Residuals    Sum of squares       =   .5273813E-03 |
|              Standard error of e  =   .1365115E-02 |
| Fit          R-squared            =   .9282124     |
|              Adjusted R-squared   =   .9114704     |
| Model test   F[ 66,   283] (prob) =  55.44 (.0000) |
| Diagnostic   Log likelihood       =   1849.330     |
|              Restricted(b=0)      =   1388.380     |
|              Chi-sq [ 66]  (prob) = 921.92 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =  -13.01788     |
|              Akaike Info. Criter. =  -13.02266     |
| Estd. Autocorrelation of e(i,t)      .074958       |
+----------------------------------------------------+
```

Estimated Fixed Effects - Full sets of effects, normalized to

Group	Coefficient	Standard Error	t-ratio
1	.00857	.00161	5.33434
2	.00467	.00734	.63678
3	.00229	.00260	.87981
4	.01006	.00209	4.80648
5	-.02607	.01455	-1.79161
6	-.00014	.00251	-.05498
7	-.00922	.00529	-1.74424
8	-.00295	.00473	-.62258
9	-.00461	.00658	-.70154
10	.00222	.00229	.97268
11	-.00353	.00479	-.73756
12	.00596	.00363	1.64151
13	-.00419	.00478	-.87677
14	-.00075	.00118	-.63862
15	.00582	.00344	1.69190
16	.00678	.00427	1.58723
17	.00514	.00143	3.59424
18	.00665	.00239	2.77882
19	-.00534	.00362	-1.47510
20	-.00606	.00351	-1.72759

μ_i

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
PI92	.226658D-06	.280961D-06	.807	.4204	20460.8038
POP	.134814D-08	.587353D-09	2.295	.0223	.513867D+07
HOSPAREA	-.83312526	.86160797	-.967	.3343	.00258178
ETHPC	.83897209	1.59819810	.525	.6000	.00184147
SEATBELT	-.00281535	.00142595	-1.974	.0492	.58400000
PERC1524	.06577722	.02957799	2.224	.0268	.14278826
PERC2544	.04788101	.04200016	1.140	.2551	.31660983
PERC4564	.00321838	.04736540	.068	.9459	.19345771
PERC75P	.15764805	.12826417	1.229	.2199	.05521571
LNMILES	-.516750D-07	.295699D-07	-1.748	.0814	162439.608
PRECIP	-.363727D-04	.160696D-04	-2.263	.0242	37.1971429
Constant	-.01478536	.03059653	-.483	.6292	

Panel Data Models: Example 3

Fixed Effects Panel Data Model Estimates (Two-way Error)
Fatalities per vehicle travelled mile (FATVTM)

Estimated Fixed Effects - Full sets of effects, normalized to sum to 0

Period	Coefficient	Standard Error	t-ratio
1	.00254	.00064	3.99512
2	.00113	.00053	2.12076
3	-.00021	.00028	-.73363
4	-.00034	.00019	-1.77758
5	-.00092	.00027	-3.43895
6	-.00077	.00047	-1.61718
7	-.00144	.00076	-1.88969

λ_t

Test Statistics for the Classical Model

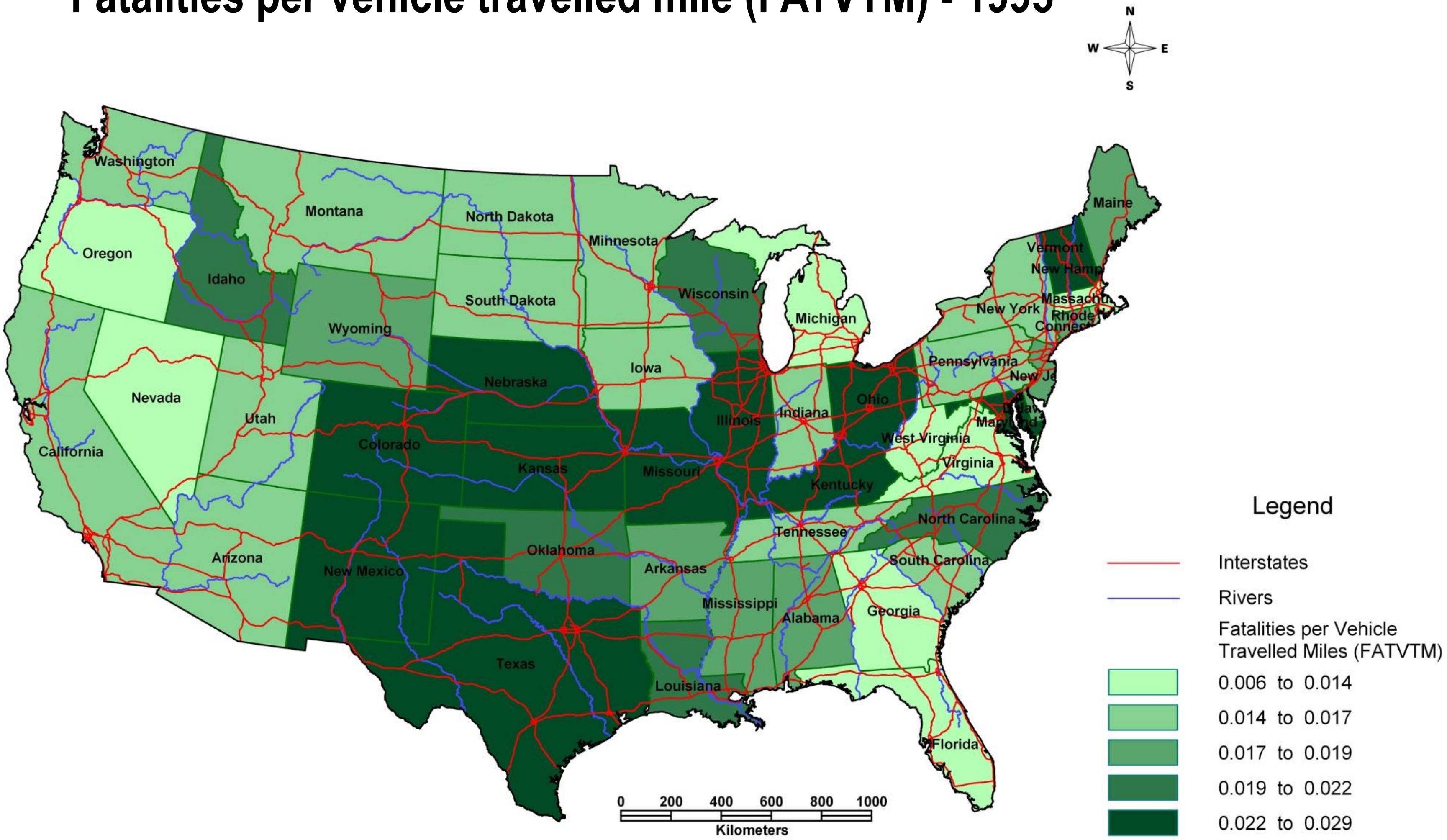
	Model	Log-Likelihood	Sum of Squares	R-squared
(1)	Constant term only	1388.37998	.7346408134D-02	.0000000
(2)	Group effects only	1710.41141	.1166528091D-02	.8412111
(3)	X - variables only	1546.00949	.2984616903D-02	.5937311
(4)	X and group effects	1820.82267	.6207110400D-03	.9155082
(5)	X ind.&time effects	1849.33751	.5273812609D-03	.9282124

Hypothesis Tests

	Likelihood Ratio Test			F Tests			
	Chi-squared	d.f.	Prob.	F	num.	denom.	P value
(2) vs (1)	644.063	49	.00000	32.435	49	300	.00000
(3) vs (1)	315.259	11	.00000	44.906	11	338	.00000
(4) vs (1)	864.885	60	.00000	52.191	60	289	.00000
(4) vs (2)	220.823	11	.00000	23.103	11	289	.00000
(4) vs (3)	549.626	49	.00000	22.462	49	289	.00000
(5) vs (4)	57.030	6	.00000	8.347	6	283	.00000
(5) vs (3)	606.656	56	.00000	23.546	56	283	.00000

Panel Data Models: Example 3

Fixed Effects Panel Data Model Estimates (Two-way Error)
Fatalities per vehicle travelled mile (FATVTM) - 1995



Panel Data Models: Example 3

Random Effects Panel Data Model Estimates (One-way Error)
Fatalities per vehicle travelled mile (FATVTM)

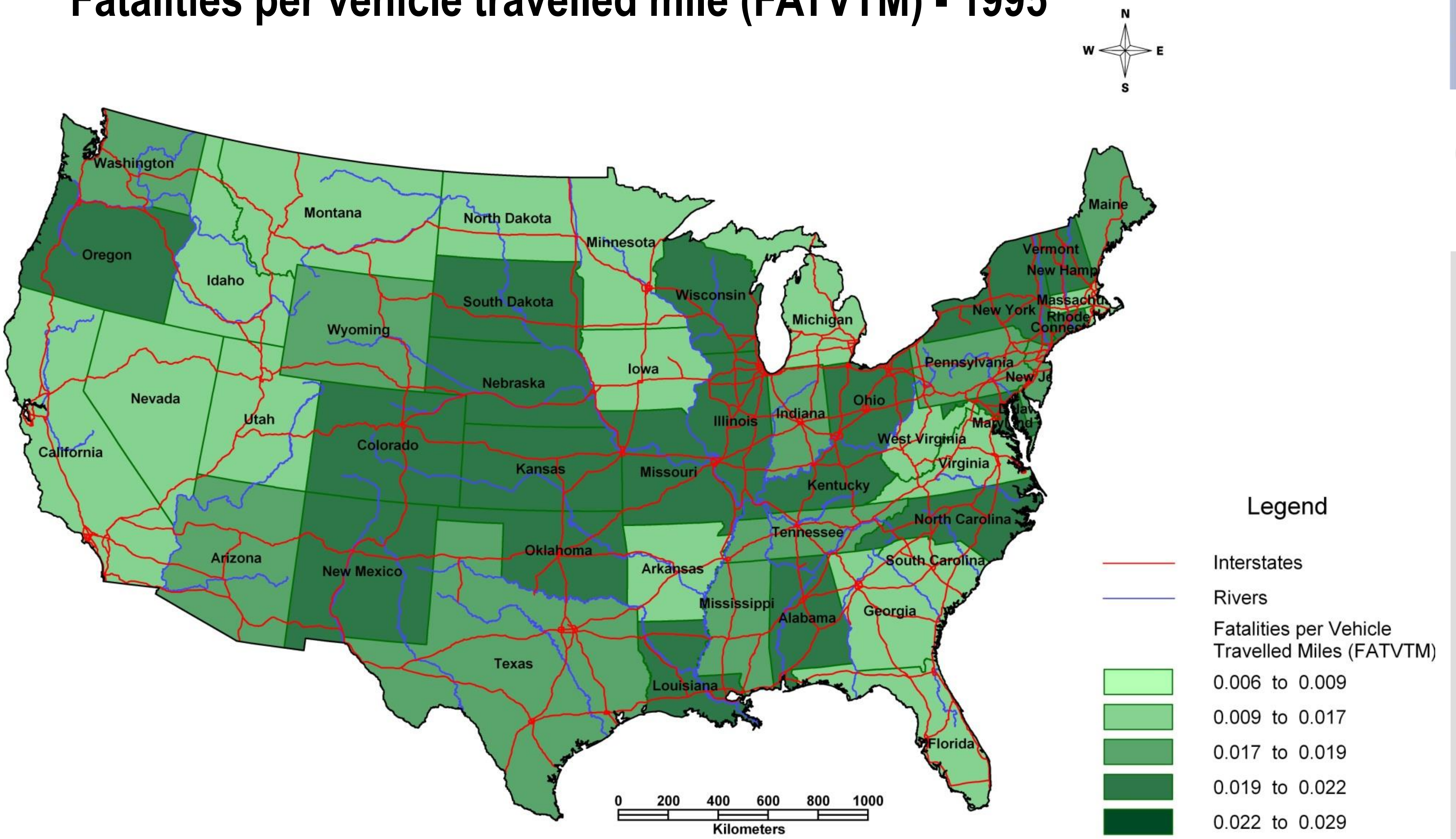
```
+----------------------------------------------------------------------+
| OLS Without Group Dummy Variables                                     |
| Ordinary    least squares regression                                  |
| Model was estimated Oct 14, 2010 at 09:28:52AM                        |
| LHS=FATVMT    Mean                     =   .1866178E-01               |
|               Standard deviation       =   .4588015E-02               |
| WTS=none      Number of observs.       =          350                 |
| Model size    Parameters               =           13                 |
|               Degrees of freedom       =          337                 |
| Residuals     Sum of squares           =   .2366157E-02               |
|               Standard error of e      =   .2649761E-02               |
| Fit           R-squared                =   .6779165                   |
|               Adjusted R-squared       =   .6664477                   |
| Model test    F[ 12,    337] (prob)    =  59.11 (.0000)               |
| Diagnostic    Log likelihood           =   1586.645                   |
|               Restricted(b=0)          =   1388.380                   |
|               Chi-sq [ 12]  (prob)     = 396.53 (.0000)               |
| Info criter.  LogAmemiya Prd. Crt.     =  -11.83010                   |
|               Akaike Info. Criter.     =  -11.83014                   |
+----------------------------------------------------------------------+
```

```
+--------------------------------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)                              |
| Estimates:  Var[e]                       =   .205299D-05                  |
|             Var[u]                       =   .496825D-05                  |
|             Corr[v(i,t),v(i,s)]          =   .707603                      |
| Lagrange Multiplier Test vs. Model (3) =    291.00                        |
| ( 1 df, prob value =  .000000)                                            |
| (High values of LM favor FEM/REM over CR model.)                          |
| Baltagi-Li form of LM Statistic =           291.00                        |
| Fixed vs. Random Effects (Hausman)        =     .00                       |
| (12 df, prob value = 1.000000)                                            |
| (High (low) values of H favor FEM (REM).)                                 |
|             Sum of Squares                   .358714D-02                  |
|             R-squared                        .511715D+00                  |
+--------------------------------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
YEAR	-.00049436	.00011501	-4.298	.0000	1993.00000
PI92	-.291797D-06	.155039D-06	-1.882	.0598	20460.8038
POP	.115470D-09	.975382D-10	1.184	.2365	.513867D+07
HOSPAREA	-.72287233	.19328036	-3.740	.0002	.00258178
ETHPC	1.37233196	.82313631	1.667	.0955	.00184147
SEATBELT	-.00464050	.00128596	-3.609	.0003	.58400000
PERC1524	.00364361	.02335547	.156	.8760	.14278826
PERC2544	-.04071682	.02683827	-1.517	.1292	.31660983
PERC4564	.03763409	.02527058	1.489	.1364	.19345771
PERC75P	-.08847938	.04345419	-2.036	.0417	.05521571
LNMILES	-.439628D-08	.505405D-08	-.870	.3844	162439.608
PRECIP	.886500D-05	.137572D-04	.644	.5193	37.1971429
Constant	1.02171594	.22911037	4.459	.0000	

Panel Data Models: Example 3

Random Effects Panel Data Model Estimates (One-way Error)
Fatalities per vehicle travelled mile (FATVTM) - 1995



Panel Data Models: Example 3

Alternative Modeling Specifications for Fatalities per VMT (t statistics in parentheses)

Independent Variable	Model 1 OLS: No Correction	Model 2 Fixed Effects	Model 3 Two-Way Fixed Effects	Model 4 Random Effects	Model 5 Fixed Effects w/ AR(1)	Model 6 Random Effects w/ AR(1)
	Parameter Estimates					
Constant	1.69 (9.22)	—	—	0.83 (3.83)	—	0.28 (1.2)
YEAR	–7.90E-04 (–8.62)	–6.71E-04 (–3.75)	—	–4.10E-04 (–3.87)	–4.50E-04 (–2.21)	–1.30E-04 (–1.15)
PI92	–6.50E-07 (–7.12)	3.30E-07 (1.61)	5.5E-08 (0.23)	–1.90E-07 (–1.33)	1.30E-07 (0.62)	–3.70E-07 (–2.57)
POP	1E-10 (2.4)	1.1E-09 (2.21)	9.3E-10 (1.91)	1.30E-10 (1.35)	1.50E-09 (2.81)	1.20E-10 (1.26)
HOSPAREA	–0.18 (–1.88)	–0.35 (–0.61)	–0.55 (–0.95)	–0.64 (–3.22)	–0.19 (–0.35)	–0.54 (–2.79)
ETHPC	1.62 (3.65)	2.6 (1.72)	1.81 (1.19)	1.43 (1.73)	–1.97 (–1.11)	–3.70E-03 (–0.004)
SEATBELT	–2.20E-03 (–1.67)	–4.00E-03 (–3.3)	2.37E-03 (–1.87)	–4.10E-03 (–3.590)	–2.50E-03 (–1.95)	–2.30E-03 (–1.91)
PERC1524	–0.148 (–5.63)	0.084 (3.22)	0.074 (2.85)	0.036 (1.713)	0.13 (4.41)	0.032 (1.33)
PERC2544	–0.184 (–10.61)	0.097 (2.77)	0.081 (2.26)	1.22E-02 (0.51)	0.16 (4.3)	0.031 (1.24)
PERC4564	0.081 (5.48)	0.063 (1.66)	0.022 (0.56)	0.037 (1.6)	0.052 (1.25)	0.011 (0.51)
PERC75P	–0.298 (–11.29)	0.226 (2.29)	0.15 (1.48)	–6.20E-03 (–0.15)	0.31 (2.980)	3.80E-03 (0.090)
INMILES	–2.4E-09 (–1.11)	–3.6E-08 (–1.47)	–3.6E-08 (–1.49)	–2.8E-09 (–0.55)	–3.3E-08 (–1.35)	–4.4E-09 (–0.88)
PRECIP	–3.10E-05 (–0.8)	3.30E-04 (2.210)	2.30E-04 (1.61)	2.10E-04 (2.77)	2.40E-04 (1.48)	1.80E-04 (2.46)
Model Statistics						
N	400	400	400	400	350	350
R^2	0.650	0.916	0.923	0.650	0.926	0.651

Panel Data Models: Example 3

Conclusions

The results and the significance of the parameter estimates in particular show ample variation between the different specifications

- The parameter for percent seat belt use (*seatbelt) is* significant at the 99% level for both the fixed- and random-effects specifications but loses much of this significance when incorporating a two-way fixed-effects model or a correction for serial correlation
  - Indicating that without correction for serial correlation, the standard error of the parameters was downward biased; that is, the models without correction underestimated the standard error
  - Autocorrrelation effects may bias the estimators efficiency, but there is no prove to systematically sustain a bias of the estimates
- On the other hand, the parameters for the hospitals per square miles (*HOSPAREA) variable are significant* for both random-effects specifications but are not significant for any of the fixed-effects formulations
  - This fact may indicate that the hospitals density is clearly linked to state characteristics and occupation densities, which are intrinsic to the state panel that was calibrated (μ_i)

Panel Data Models: Example 3

Conclusions

This brings up the interesting and at times highly debated issue of model selection

As previously discussed, when inferences are confined to the effects in the model, the effects are more appropriately considered to be fixed

When inferences are made about a population of effects from which those in the data are considered to be a random sample, then the effects should be considered random

In these cases, a fixed-effects model is defended on grounds that inferences are confined to the sample. In favor of selecting the fixed-effects rather than the random-effects formulation based on the LM Test and the Hausman tests.

```
+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]              =  .205299D-05   |
|             Var[u]              =  .496825D-05   |
|             Corr[v(i,t),v(i,s)] =  .707603       |
| Lagrange Multiplier Test vs. Model (3) =  291.00 |
| ( 1 df, prob value =  .000000)                   |
| (High values of LM favor FEM/REM over CR model.) |
| Baltagi-Li form of LM Statistic =         291.00 |
| Fixed vs. Random Effects (Hausman)    =      .00 |
| (12 df, prob value = 1.000000)                   |
| (High (low) values of H favor FEM (REM).)        |
|             Sum of Squares         .358714D-02   |
|             R-squared              .511715D+00   |
+--------------------------------------------------+
```

```
+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]              =  .253004D+04   |
|             Var[u]              =  .562055D+04   |
|             Corr[v(i,t),v(i,s)] =  .689588       |
| Lagrange Multiplier Test vs. Model (3) =  874.75 |
| ( 1 df, prob value =  .000000)                   |
| (High values of LM favor FEM/REM over CR model.) |
| Baltagi-Li form of LM Statistic =         874.75 |
| Fixed vs. Random Effects (Hausman)    =     2.89 |
| ( 2 df, prob value =  .235736)                   |
| (High (low) values of H favor FEM (REM).)        |
|             Sum of Squares         .186431D+07   |
|             R-squared              .808040D+00   |
+--------------------------------------------------+
```

Panel Data Models: Bibliography

Washington, S., Karlaftis, M., Mannering, F. (2003), Statistical and Econometric Methods for Transportation Data Analysis, Chapman & Hall/CRC.

Wooldridge, J. (2001), Econometric Analysis of Cross Section and Panel Data, MIT Press, Cambridge, Massachusetts.

Baltagi, B, (2005), Econometric Analysis of Panel Data, Third Edition, John Willey & Sons, Ltd., Chichester.

___(1996), A Handbook of the Theory with Applications, Advanced Studies in Theoretical and Applied Econometrics , Vol. 33, Mátyás, László; Sevestre, Patrick (Eds.), Springer.

Hsiao, C. (2003), Analysis of Panel Data, Second Edition, Cambridge University Press, Cambridge.

Frees, E. (2004), Longitudinal and Panel Data, Analysis and Applications in the Social Sciences, Cambridge University Press, Cambridge.