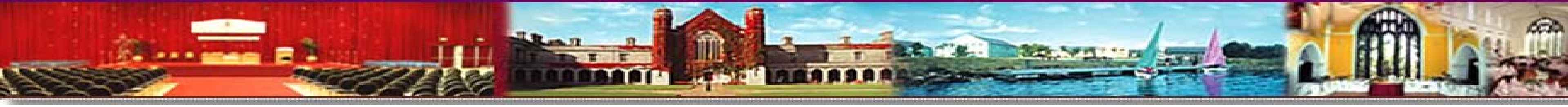
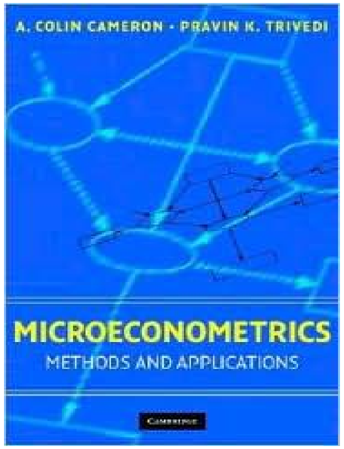
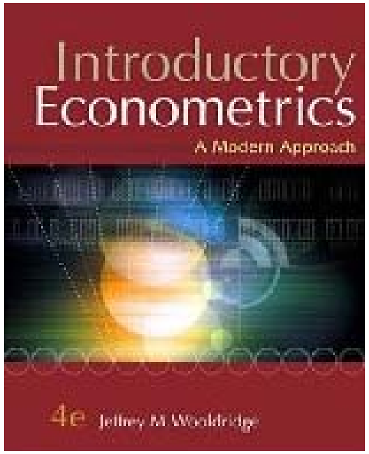
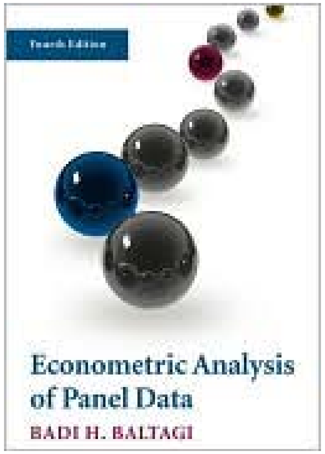
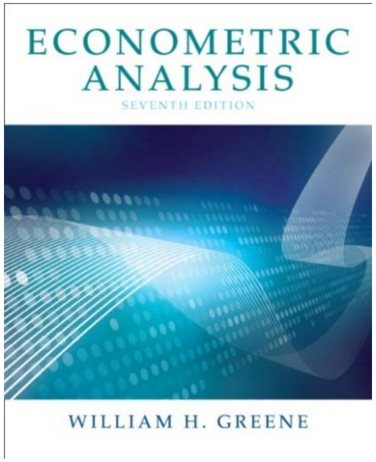


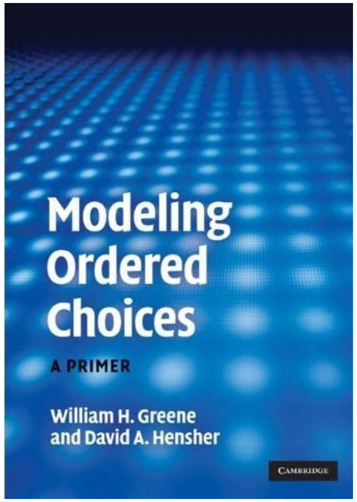


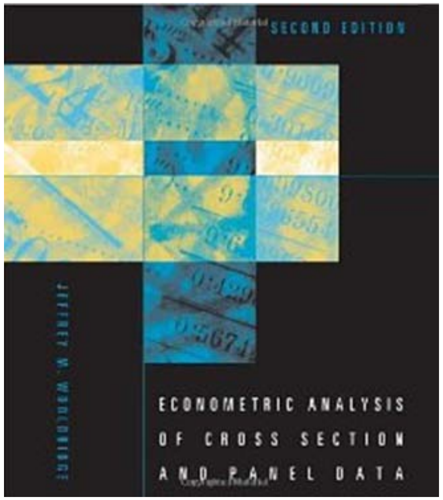


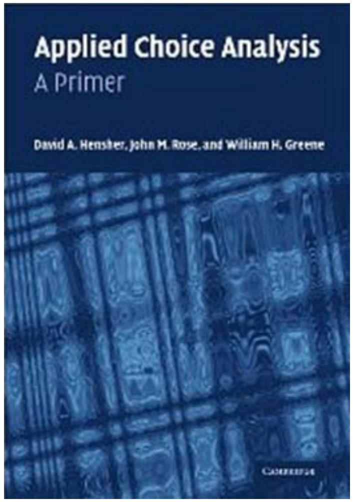


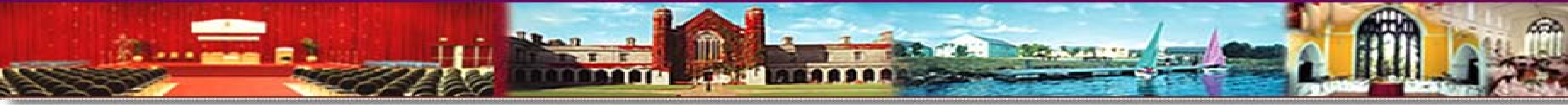


Part 3: Basic Linear Panel Data Models



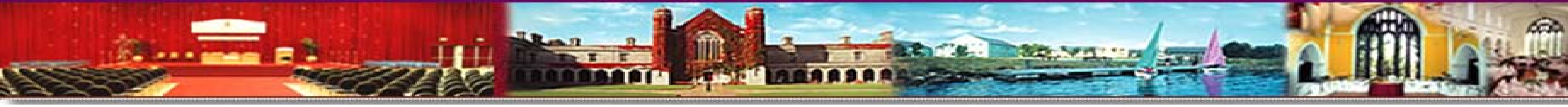


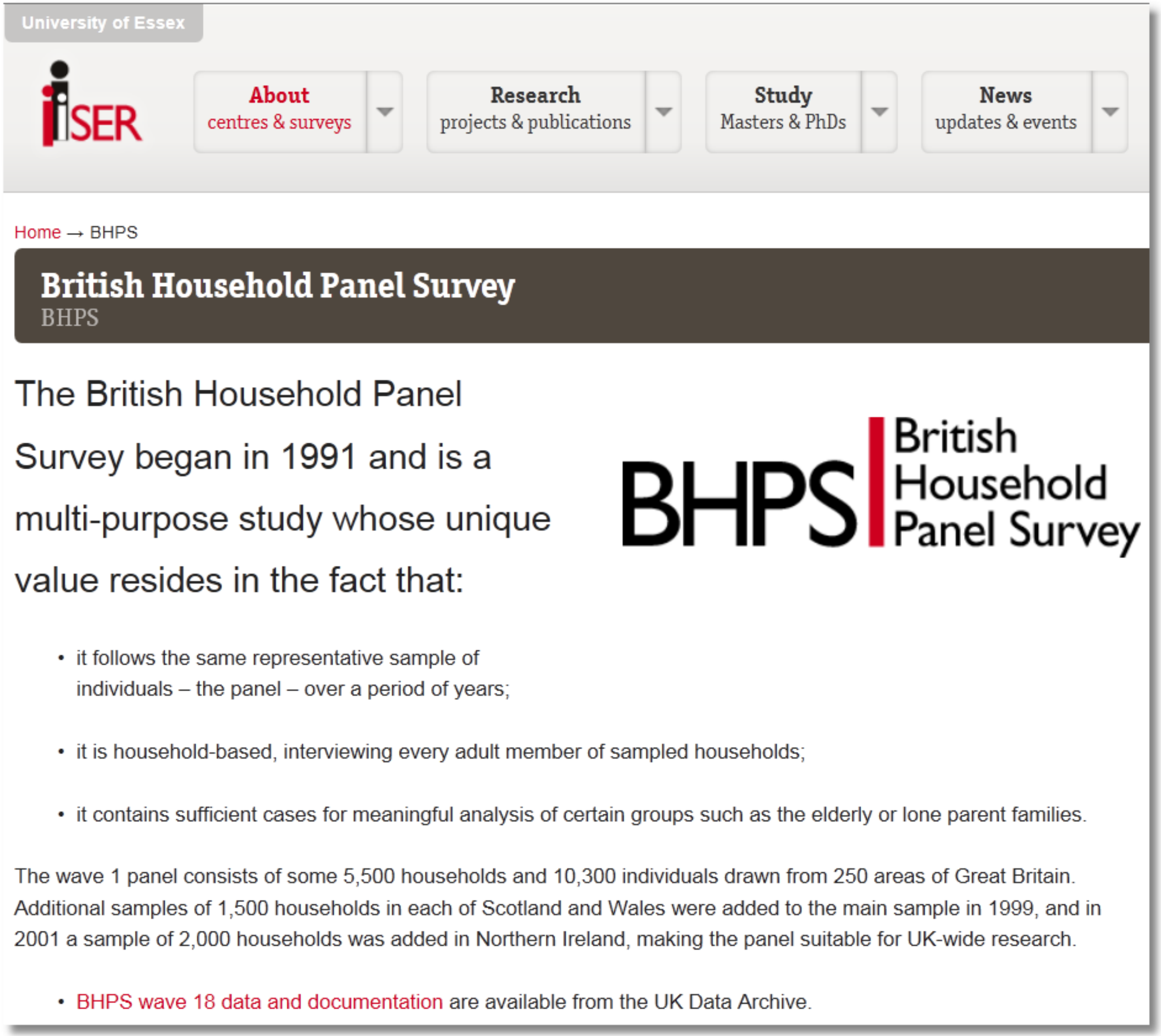




Benefits of Panel Data

- Time and individual variation in behavior unobservable in cross sections or aggregate time series
- Observable and unobservable individual heterogeneity
- Rich hierarchical structures
- More complicated models
- Features that cannot be modeled with only cross section or aggregate time series data alone
- Dynamics in economic behavior





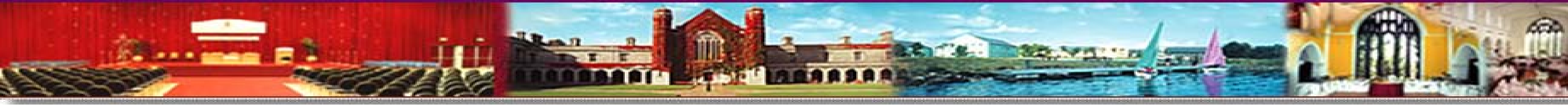
British Household Panel Survey
BHPS

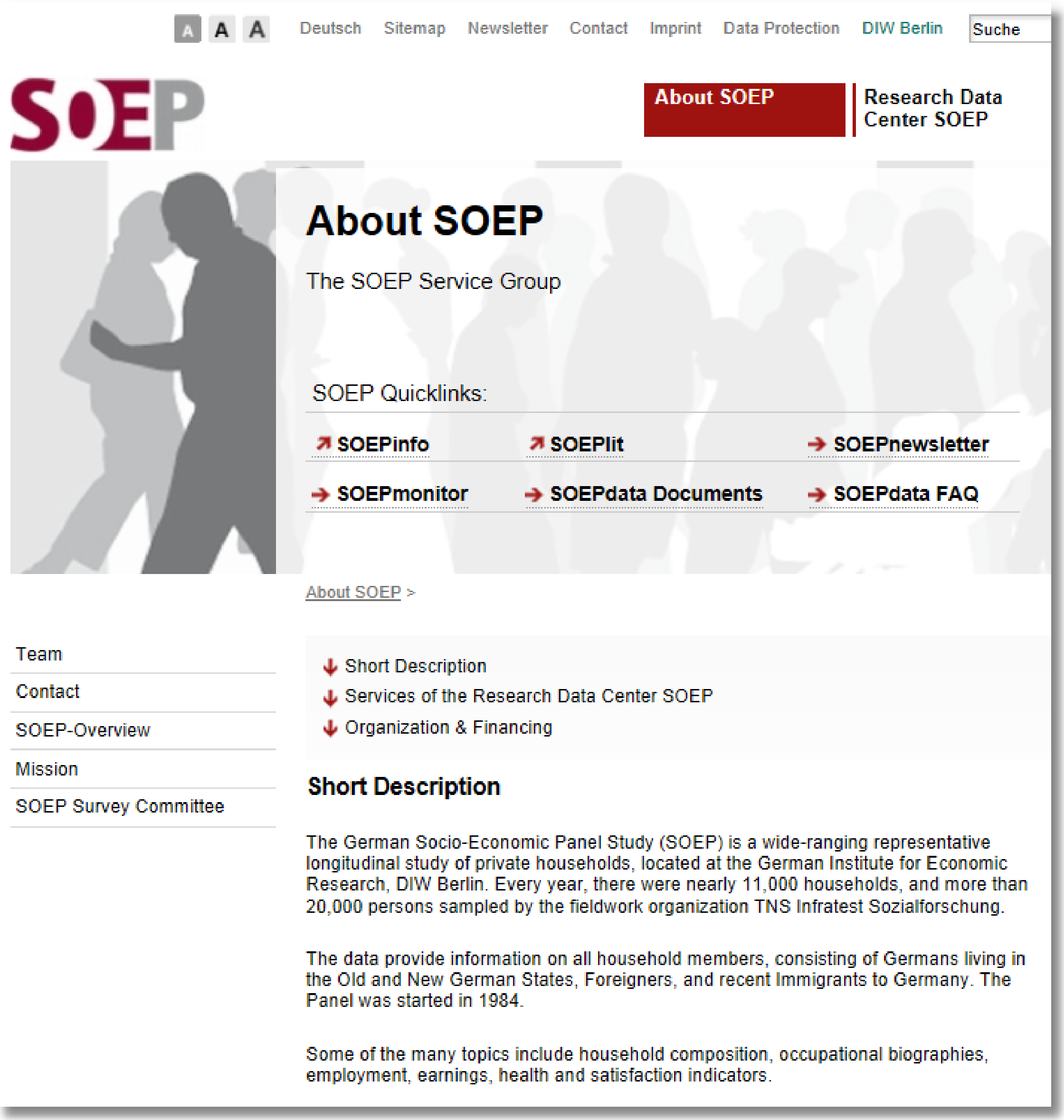
The British Household Panel Survey began in 1991 and is a multi-purpose study whose unique value resides in the fact that:

- it follows the same representative sample of individuals – the panel – over a period of years;
- it is household-based, interviewing every adult member of sampled households;
- it contains sufficient cases for meaningful analysis of certain groups such as the elderly or lone parent families.

The wave 1 panel consists of some 5,500 households and 10,300 individuals drawn from 250 areas of Great Britain. Additional samples of 1,500 households in each of Scotland and Wales were added to the main sample in 1999, and in 2001 a sample of 2,000 households was added in Northern Ireland, making the panel suitable for UK-wide research.

- BHPS wave 18 data and documentation are available from the UK Data Archive.





About SOEP

The SOEP Service Group

SOEP Quicklinks:

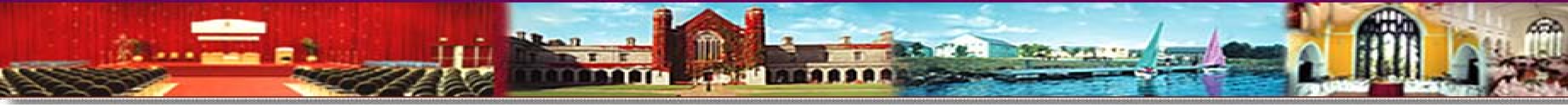
- Short Description
- Services of the Research Data Center SOEP
- Organization & Financing

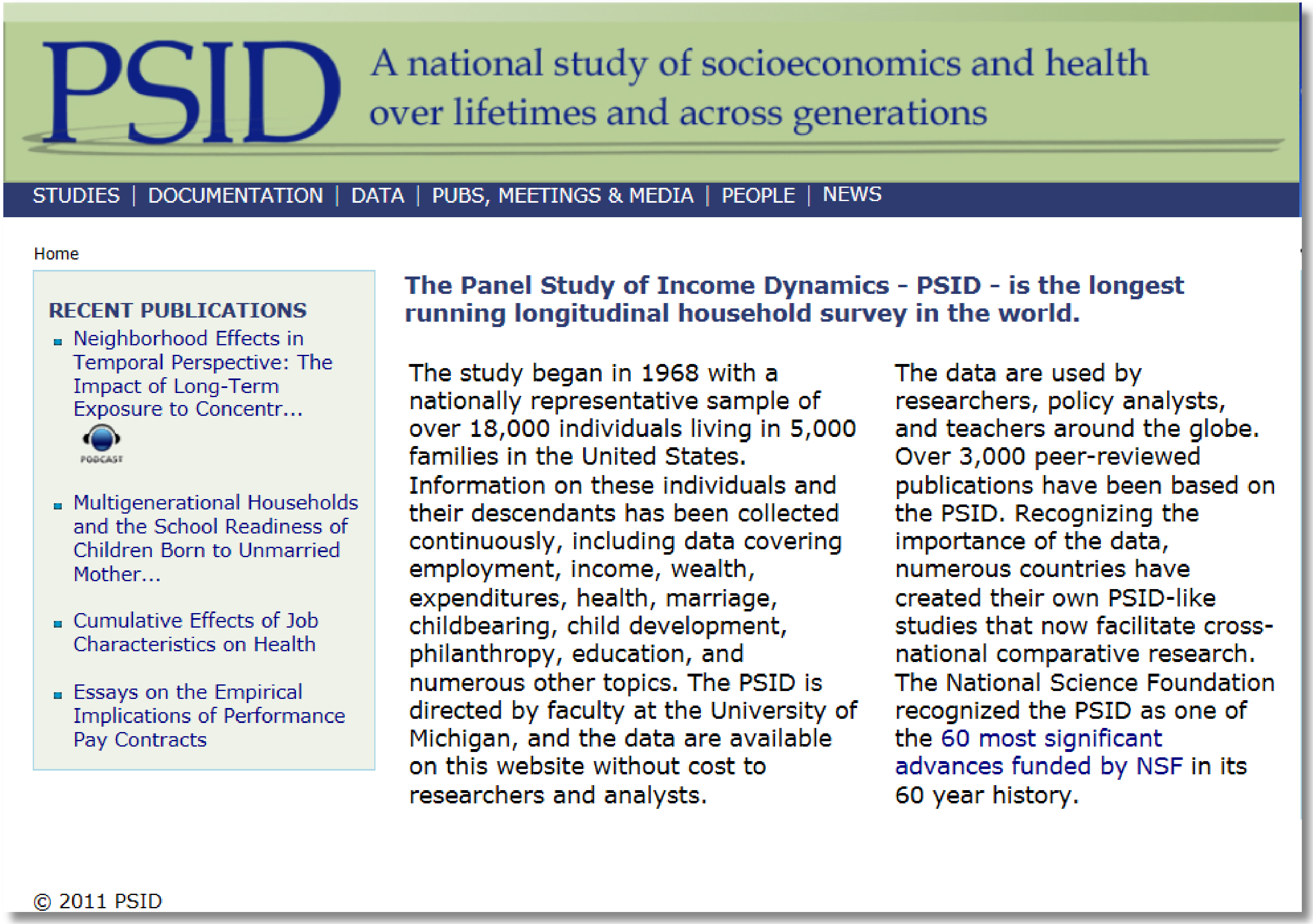
Short Description

The German Socio-Economic Panel Study (SOEP) is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin. Every year, there were nearly 11,000 households, and more than 20,000 persons sampled by the fieldwork organization TNS Infratest Sozialforschung.

The data provide information on all household members, consisting of Germans living in the Old and New German States, Foreigners, and recent Immigrants to Germany. The Panel was started in 1984.

Some of the many topics include household composition, occupational biographies, employment, earnings, health and satisfaction indicators.





PSID
STUDIES | DOCUMENTATION | DATA | PUBS, MEETINGS & MEDIA | PEOPLE | NEWS
Home
PODCAST

A national study of socioeconomics and health over lifetimes and across generations

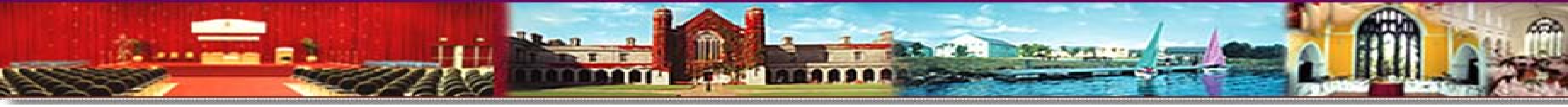
RECENT PUBLICATIONS

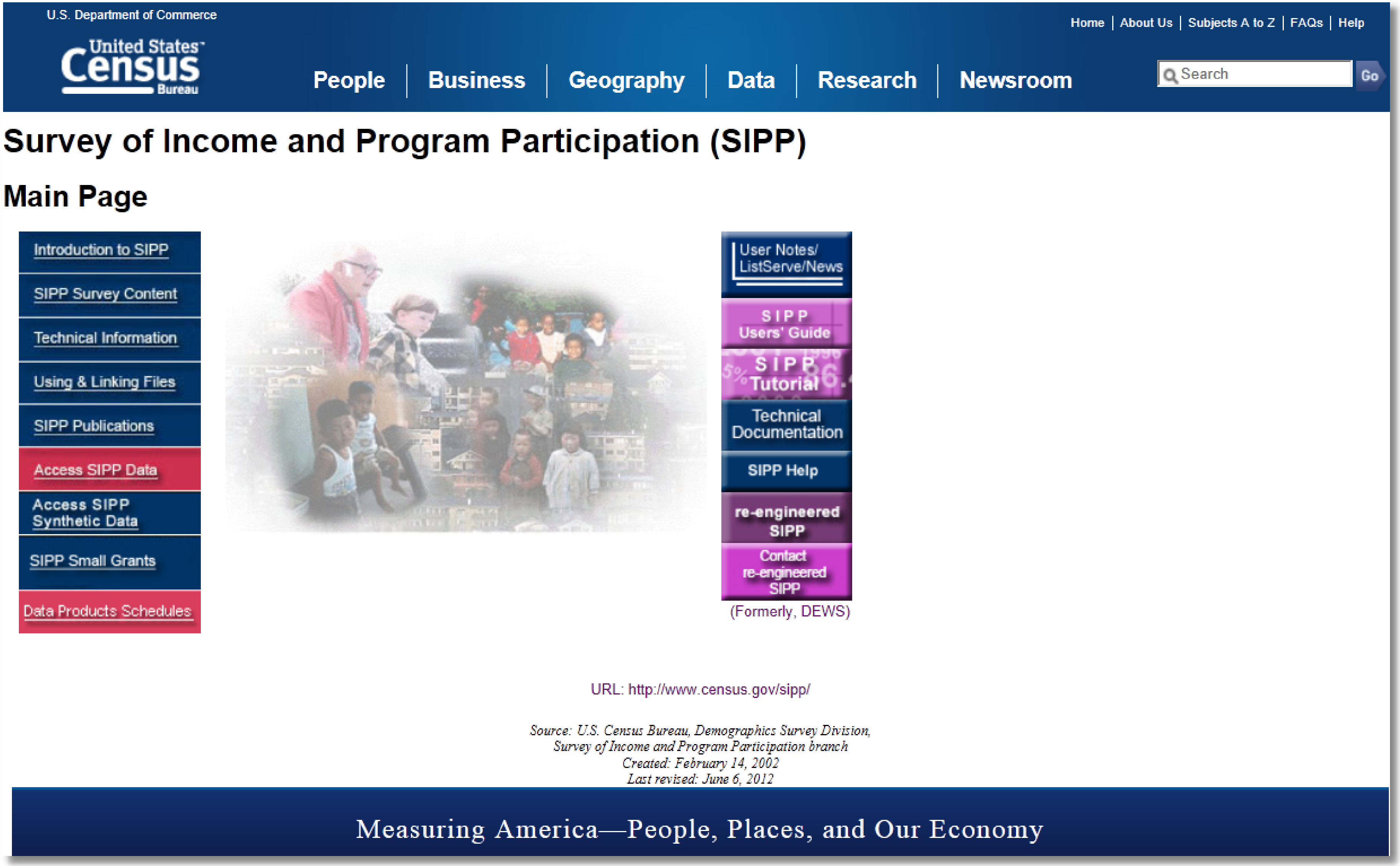
- Neighborhood Effects in Temporal Perspective: The Impact of Long-Term Exposure to Concentr...
- Multigenerational Households and the School Readiness of Children Born to Unmarried Mother...
- Cumulative Effects of Job Characteristics on Health
- Essays on the Empirical Implications of Performance Pay Contracts

The Panel Study of Income Dynamics - PSID - is the longest running longitudinal household survey in the world.

The study began in 1968 with a nationally representative sample of over 18,000 individuals living in 5,000 families in the United States. Information on these individuals and their descendants has been collected continuously, including data covering employment, income, wealth, expenditures, health, marriage, childbearing, child development, philanthropy, education, and numerous other topics. The PSID is directed by faculty at the University of Michigan, and the data are available on this website without cost to researchers and analysts.

The data are used by researchers, policy analysts, and teachers around the globe. Over 3,000 peer-reviewed publications have been based on the PSID. Recognizing the importance of the data, numerous countries have created their own PSID-like studies that now facilitate cross-national comparative research. The National Science Foundation recognized the PSID as one of the 60 most significant advances funded by NSF in its 60 year history.





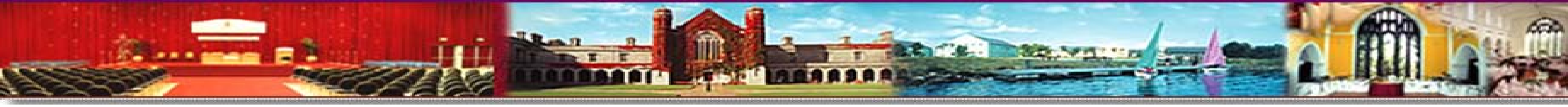
Survey of Income and Program Participation (SIPP)

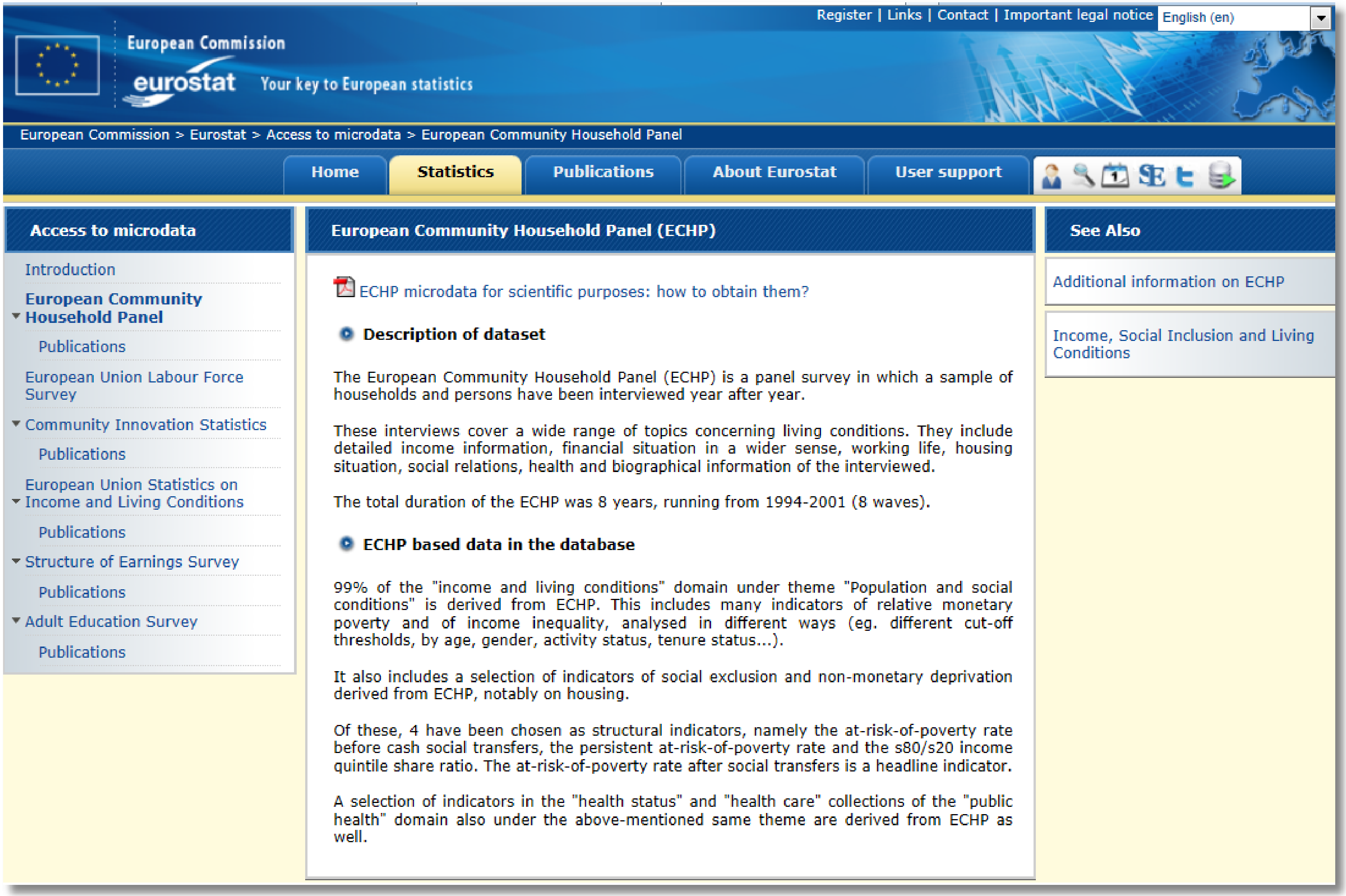
Main Page

URL: http://www.census.gov/sipp/

Source: U.S. Census Bureau, Demographics Survey Division,
Survey of Income and Program Participation branch
Created: February 14, 2002
Last revised: June 6, 2012

Measuring America—People, Places, and Our Economy




Register | Links | Contact | Important legal notice
English (en)
European Commission
eurostat
Your key to European statistics
European Commission > Eurostat > Access to microdata > European Community Household Panel
Home
Statistics
Publications
About Eurostat
User support

Access to microdata

- Introduction
- **European Community Household Panel**
  - Publications
- European Union Labour Force Survey
- Community Innovation Statistics
  - Publications
- European Union Statistics on Income and Living Conditions
  - Publications
- Structure of Earnings Survey
  - Publications
- Adult Education Survey
  - Publications

European Community Household Panel (ECHP)

ECHP microdata for scientific purposes: how to obtain them?

Description of dataset

The European Community Household Panel (ECHP) is a panel survey in which a sample of households and persons have been interviewed year after year.

These interviews cover a wide range of topics concerning living conditions. They include detailed income information, financial situation in a wider sense, working life, housing situation, social relations, health and biographical information of the interviewed.

The total duration of the ECHP was 8 years, running from 1994-2001 (8 waves).

ECHP based data in the database

99% of the "income and living conditions" domain under theme "Population and social conditions" is derived from ECHP. This includes many indicators of relative monetary poverty and of income inequality, analysed in different ways (eg. different cut-off thresholds, by age, gender, activity status, tenure status...).

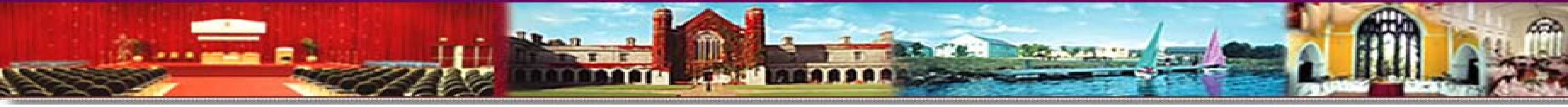
It also includes a selection of indicators of social exclusion and non-monetary deprivation derived from ECHP, notably on housing.

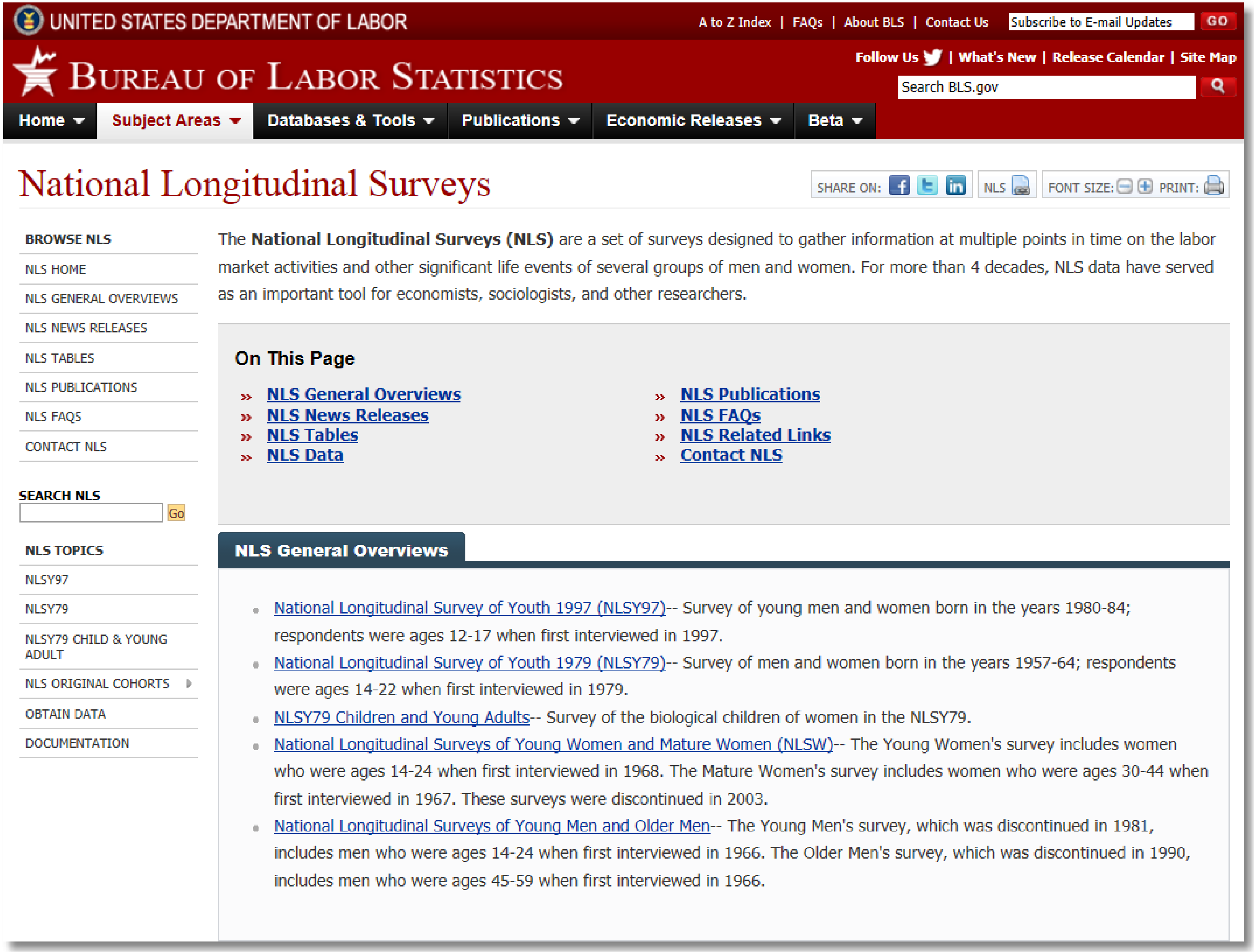
Of these, 4 have been chosen as structural indicators, namely the at-risk-of-poverty rate before cash social transfers, the persistent at-risk-of-poverty rate and the s80/s20 income quintile share ratio. The at-risk-of-poverty rate after social transfers is a headline indicator.

A selection of indicators in the "health status" and "health care" collections of the "public health" domain also under the above-mentioned same theme are derived from ECHP as well.

See Also

- Additional information on ECHP
- Income, Social Inclusion and Living Conditions





National Longitudinal Surveys

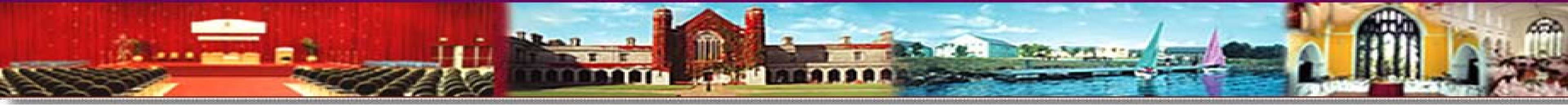
The **National Longitudinal Surveys (NLS)** are a set of surveys designed to gather information at multiple points in time on the labor market activities and other significant life events of several groups of men and women. For more than 4 decades, NLS data have served as an important tool for economists, sociologists, and other researchers.

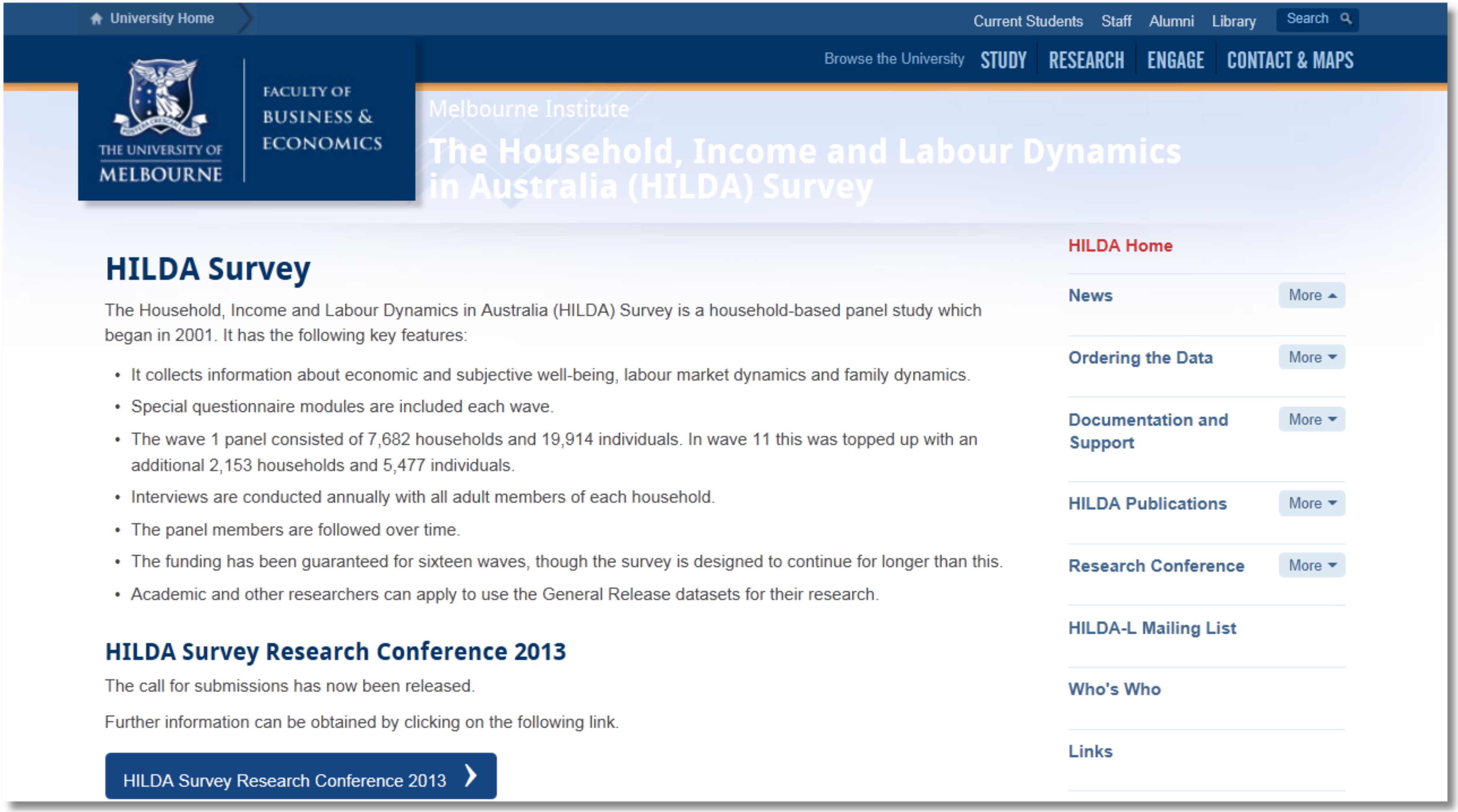
On This Page

» NLS General Overviews
» NLS News Releases
» NLS Tables
» NLS Data
» NLS Publications
» NLS FAQs
» NLS Related Links
» Contact NLS

NLS General Overviews

- National Longitudinal Survey of Youth 1997 (NLSY97)-- Survey of young men and women born in the years 1980-84; respondents were ages 12-17 when first interviewed in 1997.
- National Longitudinal Survey of Youth 1979 (NLSY79)-- Survey of men and women born in the years 1957-64; respondents were ages 14-22 when first interviewed in 1979.
- NLSY79 Children and Young Adults-- Survey of the biological children of women in the NLSY79.
- National Longitudinal Surveys of Young Women and Mature Women (NLSW)-- The Young Women's survey includes women who were ages 14-24 when first interviewed in 1968. The Mature Women's survey includes women who were ages 30-44 when first interviewed in 1967. These surveys were discontinued in 2003.
- National Longitudinal Surveys of Young Men and Older Men-- The Young Men's survey, which was discontinued in 1981, includes men who were ages 14-24 when first interviewed in 1966. The Older Men's survey, which was discontinued in 1990, includes men who were ages 45-59 when first interviewed in 1966.





The Household, Income and Labour Dynamics in Australia (HILDA) Survey

HILDA Survey

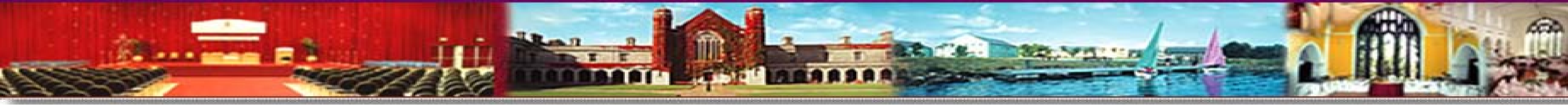
The Household, Income and Labour Dynamics in Australia (HILDA) Survey is a household-based panel study which began in 2001. It has the following key features:

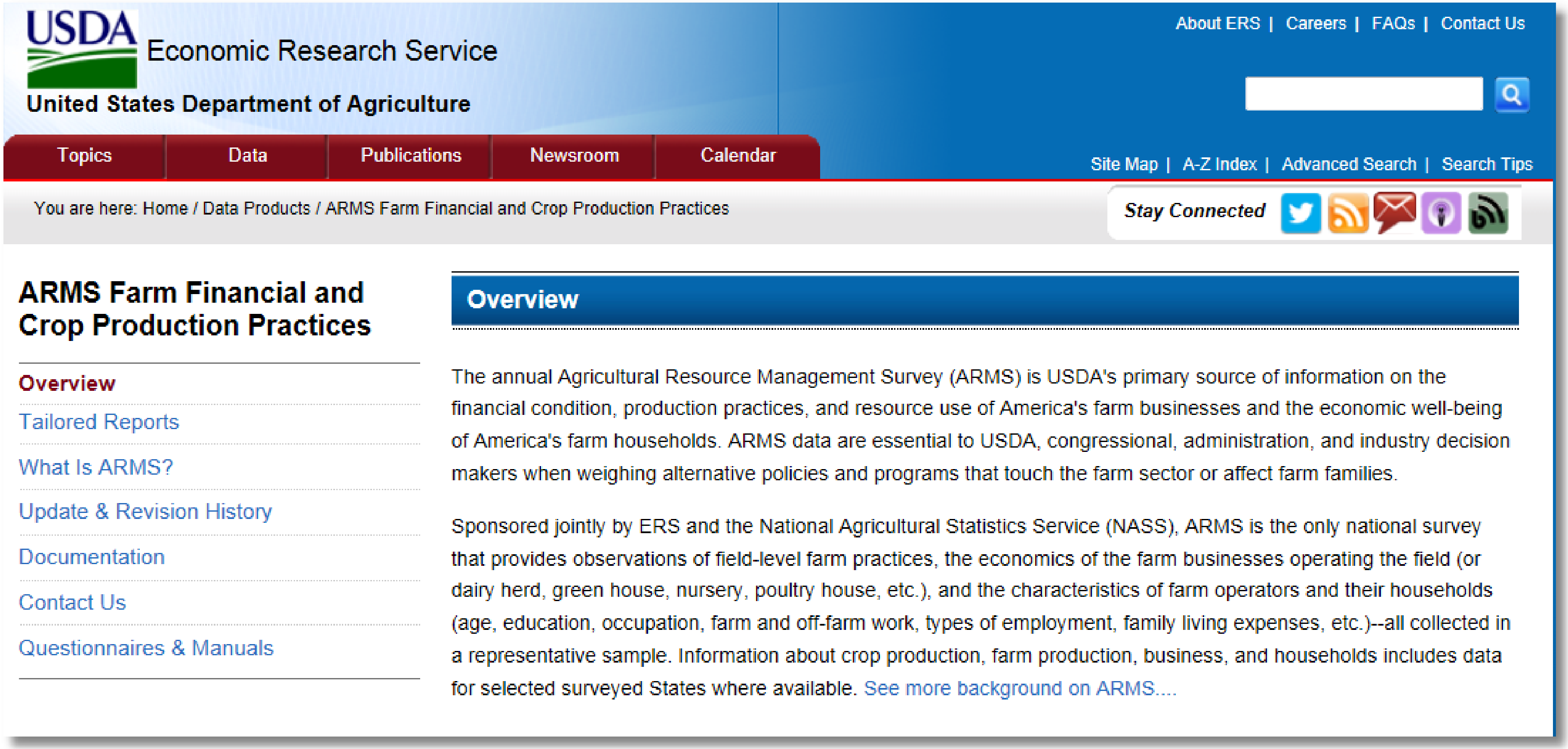
- It collects information about economic and subjective well-being, labour market dynamics and family dynamics.
- Special questionnaire modules are included each wave.
- The wave 1 panel consisted of 7,682 households and 19,914 individuals. In wave 11 this was topped up with an additional 2,153 households and 5,477 individuals.
- Interviews are conducted annually with all adult members of each household.
- The panel members are followed over time.
- The funding has been guaranteed for sixteen waves, though the survey is designed to continue for longer than this.
- Academic and other researchers can apply to use the General Release datasets for their research.

HILDA Survey Research Conference 2013

The call for submissions has now been released.

Further information can be obtained by clicking on the following link.




USDA
Economic Research Service
United States Department of Agriculture
About ERS | Careers | FAQs | Contact Us
Topics
Data
Publications
Newsroom
Calendar
Site Map | A-Z Index | Advanced Search | Search Tips
Stay Connected

You are here: Home / Data Products / ARMS Farm Financial and Crop Production Practices

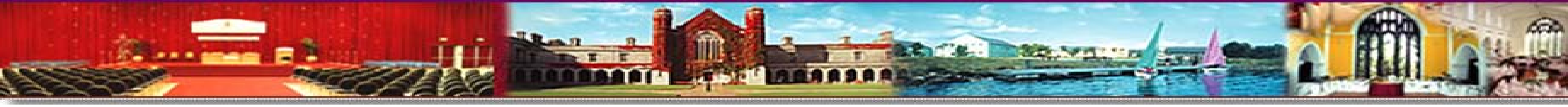
ARMS Farm Financial and Crop Production Practices

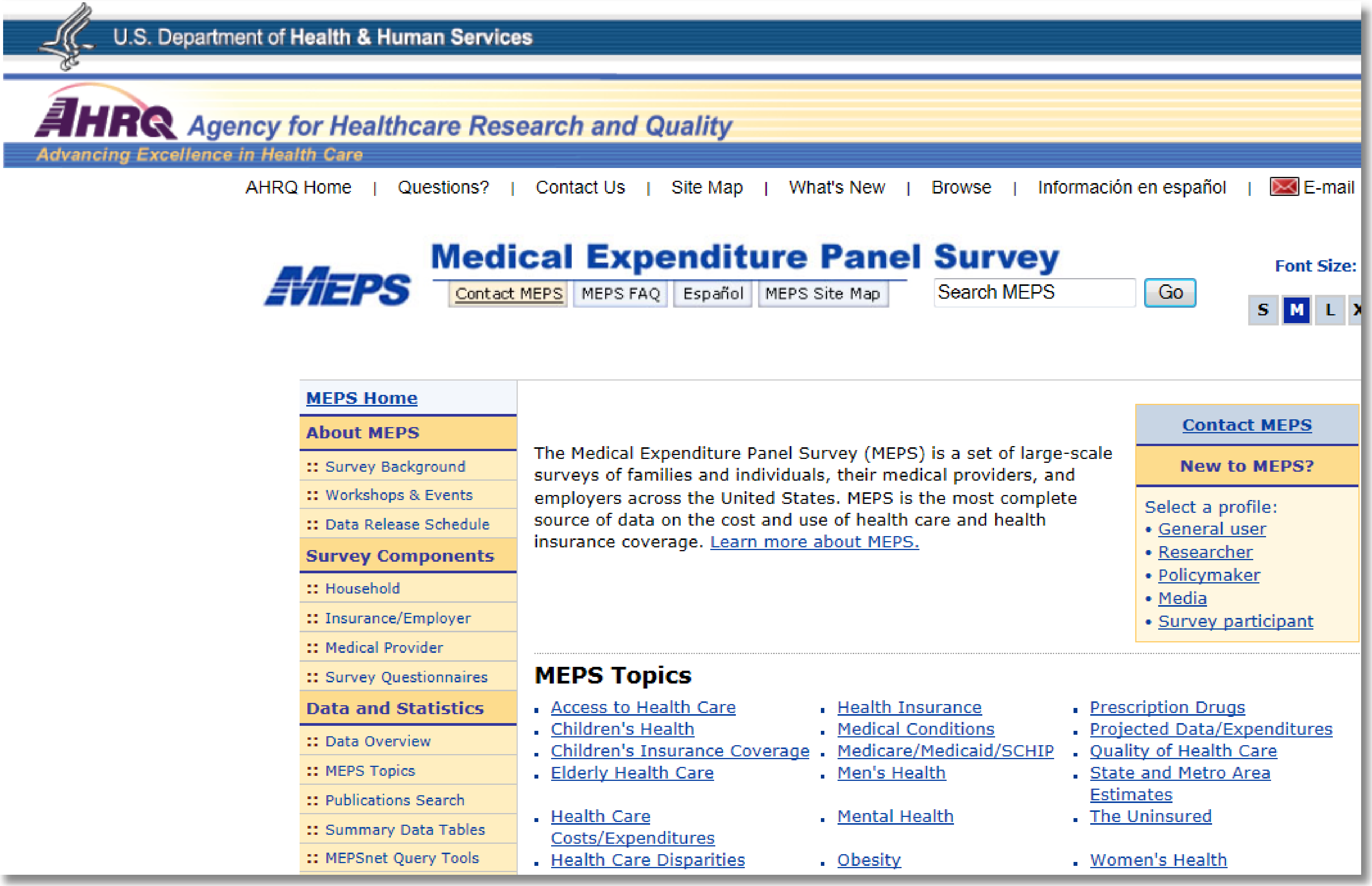
Overview
Tailored Reports
What Is ARMS?
Update & Revision History
Documentation
Contact Us
Questionnaires & Manuals

Overview

The annual Agricultural Resource Management Survey (ARMS) is USDA's primary source of information on the financial condition, production practices, and resource use of America's farm businesses and the economic well-being of America's farm households. ARMS data are essential to USDA, congressional, administration, and industry decision makers when weighing alternative policies and programs that touch the farm sector or affect farm families.

Sponsored jointly by ERS and the National Agricultural Statistics Service (NASS), ARMS is the only national survey that provides observations of field-level farm practices, the economics of the farm businesses operating the field (or dairy herd, green house, nursery, poultry house, etc.), and the characteristics of farm operators and their households (age, education, occupation, farm and off-farm work, types of employment, family living expenses, etc.)--all collected in a representative sample. Information about crop production, farm production, business, and households includes data for selected surveyed States where available. See more background on ARMS....





U.S. Department of Health & Human Services

AHRQ Agency for Healthcare Research and Quality

Advancing Excellence in Health Care

AHRQ Home | Questions? | Contact Us | Site Map | What's New | Browse | Información en español | E-mail

MEPS **Medical Expenditure Panel Survey**

Contact MEPS | MEPS FAQ | Español | MEPS Site Map

Search MEPS Go

Font Size: S M L X

MEPS Home

About MEPS

:: Survey Background
:: Workshops & Events
:: Data Release Schedule

Survey Components

:: Household
:: Insurance/Employer
:: Medical Provider
:: Survey Questionnaires

Data and Statistics

:: Data Overview
:: MEPS Topics
:: Publications Search
:: Summary Data Tables
:: MEPSnet Query Tools

The Medical Expenditure Panel Survey (MEPS) is a set of large-scale surveys of families and individuals, their medical providers, and employers across the United States. MEPS is the most complete source of data on the cost and use of health care and health insurance coverage. Learn more about MEPS.

Contact MEPS

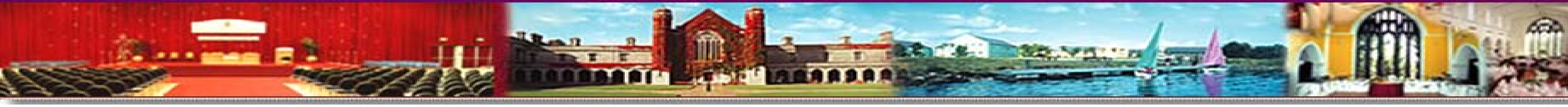
New to MEPS?

Select a profile:

- General user
- Researcher
- Policymaker
- Media
- Survey participant

MEPS Topics

- Access to Health Care
- Children's Health
- Children's Insurance Coverage
- Elderly Health Care
- Health Care Costs/Expenditures
- Health Care Disparities
- Health Insurance
- Medical Conditions
- Medicare/Medicaid/SCHIP
- Men's Health
- Mental Health
- Obesity
- Prescription Drugs
- Projected Data/Expenditures
- Quality of Health Care
- State and Metro Area Estimates
- The Uninsured
- Women's Health

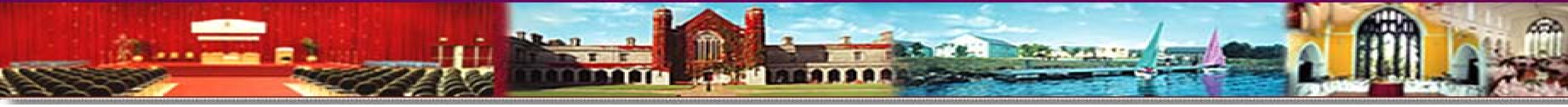


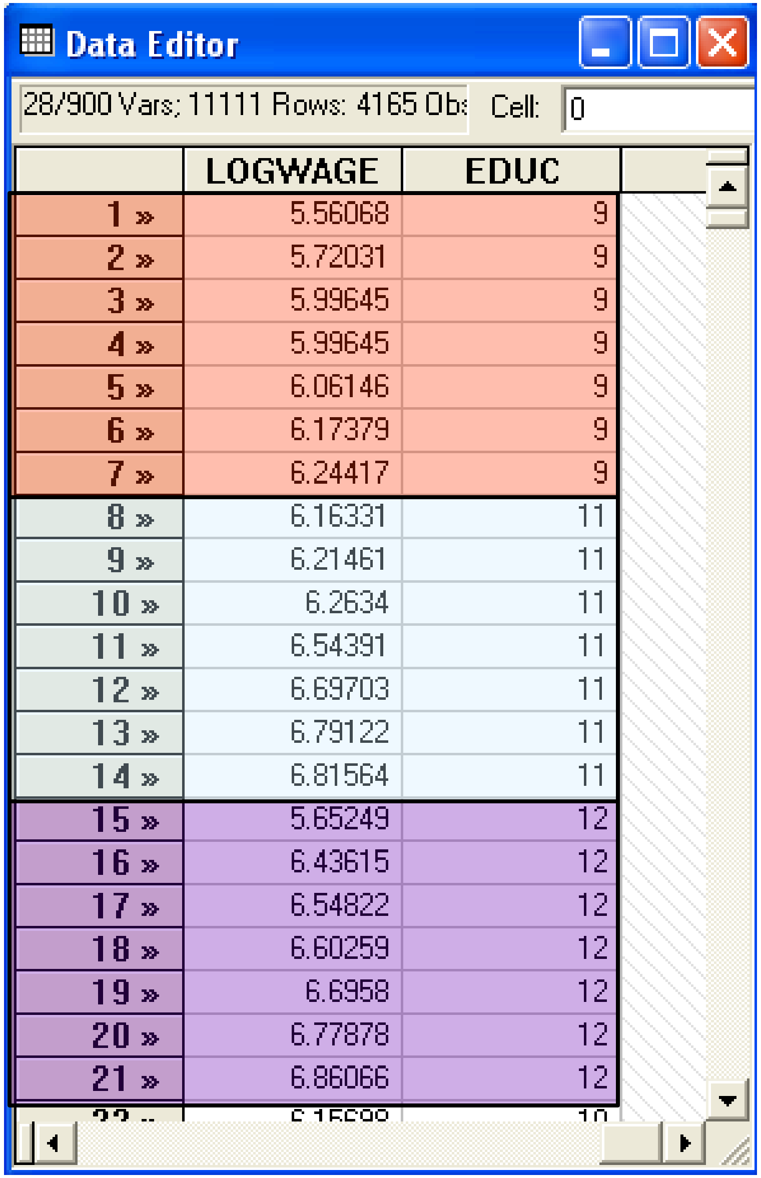
Cornwell and Rupert Data

Cornwell and Rupert Returns to Schooling Data, 595 Individuals, 7 Years
Variables in the file are

EXP	= work experience
WKS	= weeks worked
OCC	= occupation, 1 if blue collar,
IND	= 1 if manufacturing industry
SOUTH	= 1 if resides in south
SMSA	= 1 if resides in a city (SMSA)
MS	= 1 if married
FEM	= 1 if female
UNION	= 1 if wage set by union contract
ED	= years of education
BLK	= 1 if individual is black
LWAGE	= log of wage = dependent variable in regressions

These data were analyzed in Cornwell, C. and Rupert, P., "Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variable Estimators," Journal of Applied Econometrics, 3, 1988, pp. 149-155. See Baltagi, page 122 for further analysis. The data were downloaded from the website for Baltagi's text.

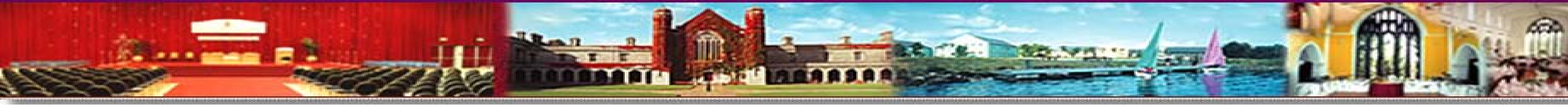




28/900 Vars; 11111 Rows: 4165 Ob

Cell: 0

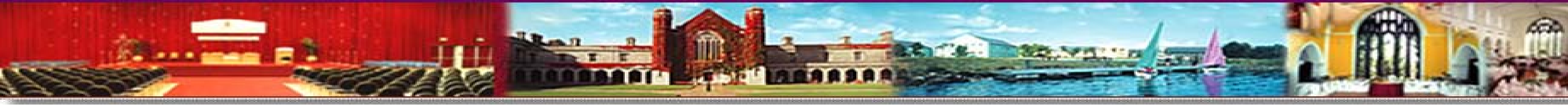
	LOGWAGE	EDUC
1 »	5.56068	9
2 »	5.72031	9
3 »	5.99645	9
4 »	5.99645	9
5 »	6.06146	9
6 »	6.17379	9
7 »	6.24417	9
8 »	6.16331	11
9 »	6.21461	11
10 »	6.2634	11
11 »	6.54391	11
12 »	6.69703	11
13 »	6.79122	11
14 »	6.81564	11
15 »	5.65249	12
16 »	6.43615	12
17 »	6.54822	12
18 »	6.60259	12
19 »	6.6958	12
20 »	6.77878	12
21 »	6.86066	12



Balanced and Unbalanced Panels

- Distinction: Balanced vs. Unbalanced Panels
- A notation to help with mechanics

  $z_{i,t}$, $i = 1,\ldots,N$; $\boxed{t = 1,\ldots,T_i}$
- The role of the assumption
  - Mathematical and notational convenience:
    - □ Balanced, n=NT
    - □ Unbalanced: $\boxed{n = \sum_{i=1}^{N} T_i}$
  - Is the fixed T_i assumption ever necessary? Almost never.
- Is unbalancedness due to **nonrandom** attrition from an otherwise balanced panel? This would require special considerations.



Application: Health Care Usage

German Health Care Usage Data, 7,293 Individuals, Varying Numbers of Periods

This is an unbalanced panel with 7,293 individuals. There are altogether 27,326 observations. The number of observations ranges from 1 to 7.

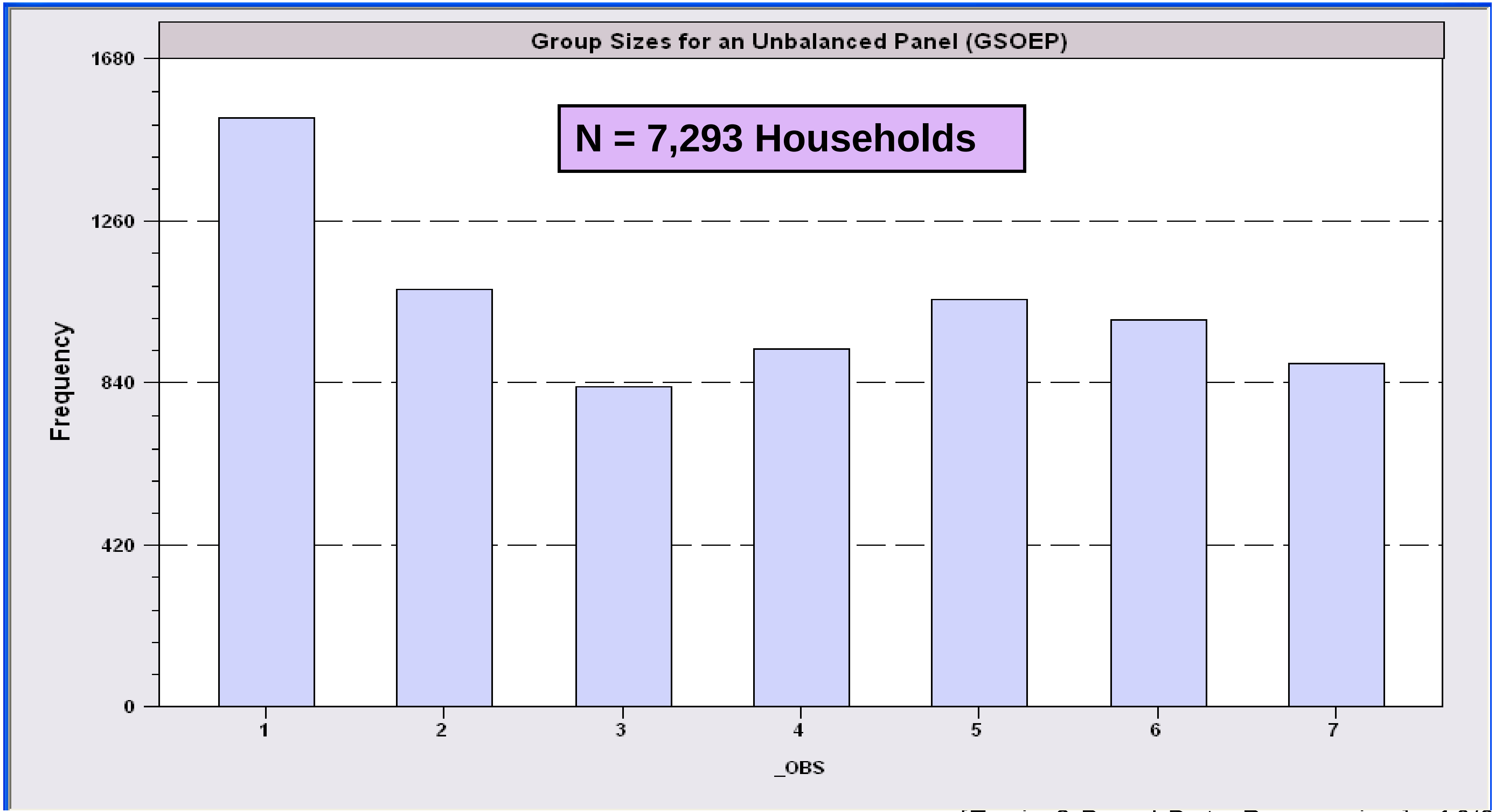
(Frequencies are: 1=1525, 2=2158, 3=825, 4=926, 5=1051, 6=1000, 7=987).

(Downloaded from the JAE Archive)

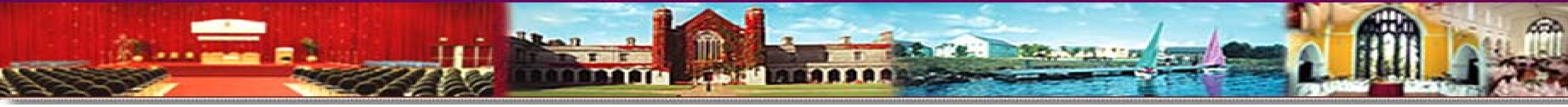
Variables in the file are

DOCTOR = 1(Number of doctor visits > 0)
HOSPITAL = 1(Number of hospital visits > 0)
HSAT = health satisfaction, coded 0 (low) - 10 (high)
DOCVIS = number of doctor visits in last three months
HOSPVIS = number of hospital visits in last calendar year
PUBLIC = insured in public health insurance = 1; otherwise = 0
ADDON = insured by add-on insurance = 1; otherswise = 0
HHNINC = household nominal monthly net income in German marks / 10000.
(4 observations with income=0 were dropped)
HHKIDS = children under age 16 in the household = 1; otherwise = 0
EDUC = years of schooling
AGE = age in years
MARRIED = marital status

An Unbalanced Panel: RWM's GSOEP Data on Health Care



N = 7,293 Households



Fixed and Random Effects

- Unobserved individual effects in regression: E[y_{it} | $\mathbf{x}_{it}$, c_i]

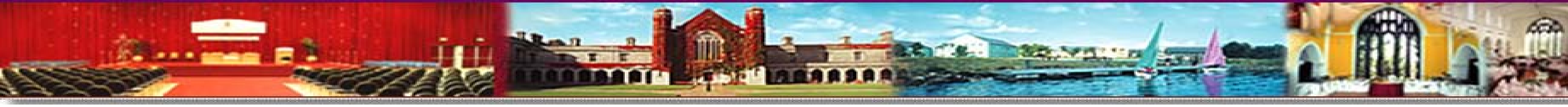
  Notation: $\boxed{y_{it} = \mathbf{x}_{it}'\beta + c_i + \varepsilon_{it}}$

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{x}_{i1}' \\ \mathbf{x}_{i2}' \\ \vdots \\ \mathbf{x}_{iT_i}' \end{bmatrix} \quad T_i \text{ rows, K columns}$$

- Linear specification:

  Fixed Effects: E[c_i | $\mathbf{X}_i$] = g($\mathbf{X}_i$). Cov[$\mathbf{x}_{it}$,c_i] $\neq \mathbf{0}$
  effects are correlated with included variables.

  Random Effects: E[c_i | $\mathbf{X}_i$] = 0. Cov[$\mathbf{x}_{it}$,c_i] = $\mathbf{0}$



Convenient Notation

- Fixed Effects – the ‘dummy variable model’

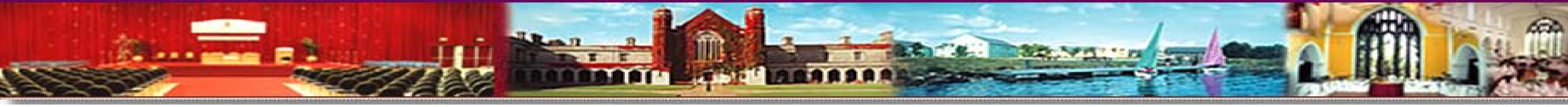
$$\boxed{y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}}$$

Individual specific constant terms.

- Random Effects – the ‘error components model’

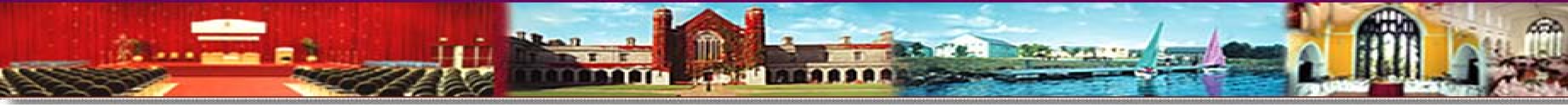
$$\boxed{y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} + u_i}$$

Compound (“composed”) disturbance



Estimating **β**

- **β** is the partial effect of interest
- Can it be estimated (consistently) in the presence of (unmeasured) c_i?
  - Does pooled least squares “work?”
  - Strategies for “controlling for c_i” using the sample data



The Pooled Regression

- Presence of omitted effects

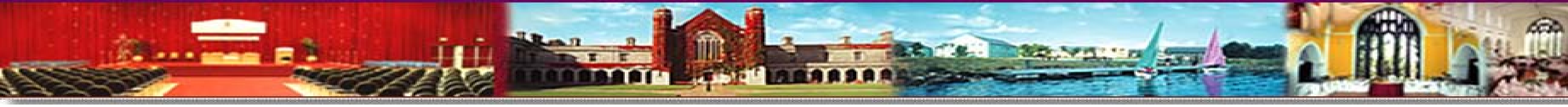
$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + \varepsilon_{it}, \text{ observation for person i at time t}$$

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{i}c_i + \boldsymbol{\varepsilon}_i, \; T_i \text{ observations in group i}$$

$$= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{c}_i + \boldsymbol{\varepsilon}_i, \text{ note } \mathbf{c}_i = (c_i, c_i, \ldots, c_i)'$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{c} + \boldsymbol{\varepsilon}, \; \Sigma_{i=1}^{N} T_i \text{ observations in the sample}$$

- Potential bias/inconsistency of OLS – depends on ‘fixed’ or ‘random’



OLS with Individual Effects

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$= \boldsymbol{\beta} + \left[(1/N)\Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right]^{-1}\left[(1/N)\Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{c}_i\mathbf{i}\right] \text{ (part due to the omitted } c_i\text{)}$$

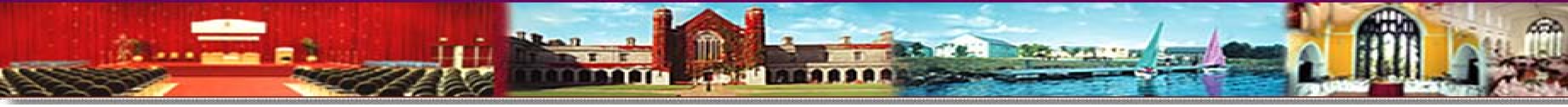
$$+ \left[(1/N)\Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right]^{-1}\left[(1/N)\Sigma_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\varepsilon}_i\right] \text{ (covariance of } \mathbf{X} \text{ and } \boldsymbol{\varepsilon} \text{ will = 0)}$$

The third term vanishes asymptotically by assumption

$$\text{plim } \mathbf{b} = \boldsymbol{\beta} + \text{plim}\left[\frac{1}{N}\Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right]^{-1}\left[\Sigma_{i=1}^{N}\frac{T_i}{N}\overline{\mathbf{x}}_i c_i\right] \text{ (left out variable formula)}$$

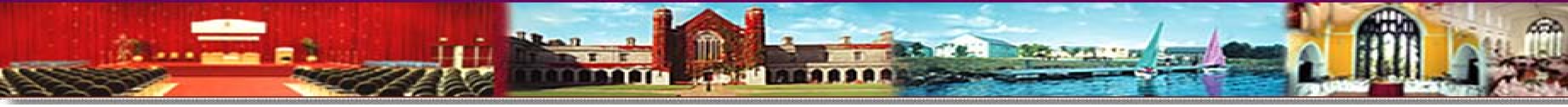
So, what becomes of $\left[\Sigma_{i=1}^{N} w_i \overline{\mathbf{x}}_i c_i\right]$?

plim $\mathbf{b} = \boldsymbol{\beta}$ if the covariance of $\overline{\mathbf{x}}_i$ and c_i converges to zero.



Estimating the Sampling Variance of b

- $s^2(\mathbf{X}'\mathbf{X})^{-1}$? Inappropriate because
  - Correlation across observations
  - (Possibly) Heteroscedasticity
- A ‘robust’ covariance matrix
  - Robust estimation (in general)
  - The White estimator
  - A Robust estimator for OLS.



Cluster Estimator

Robust variance estimator for Var[**b**]

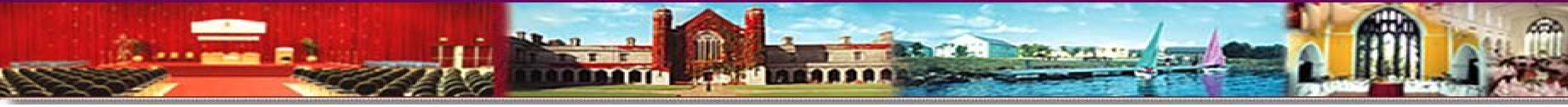
Est.Var[**b**]

$$= \boxed{(\mathbf{X'X})^{-1}\left[\Sigma_{i=1}^{N}(\Sigma_{t=1}^{T_i}\mathbf{x}_{it}\hat{v}_{it})(\Sigma_{t=1}^{T_i}\mathbf{x}_{it}'\hat{v}_{it})\right](\mathbf{X'X})^{-1}}$$

$$= (\mathbf{X'X})^{-1}\left[\Sigma_{i=1}^{N}\left(\Sigma_{t=1}^{T_i}\Sigma_{s=1}^{T_i}\hat{v}_{it}\hat{v}_{is}\mathbf{x}_{it}\mathbf{x}_{is}'\right)\right](\mathbf{X'X})^{-1}$$

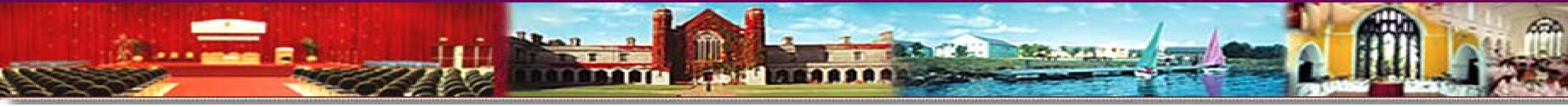
$$\hat{v}_{it} = \text{ a least squares residual } = \widehat{\varepsilon_{it} + c_i}$$

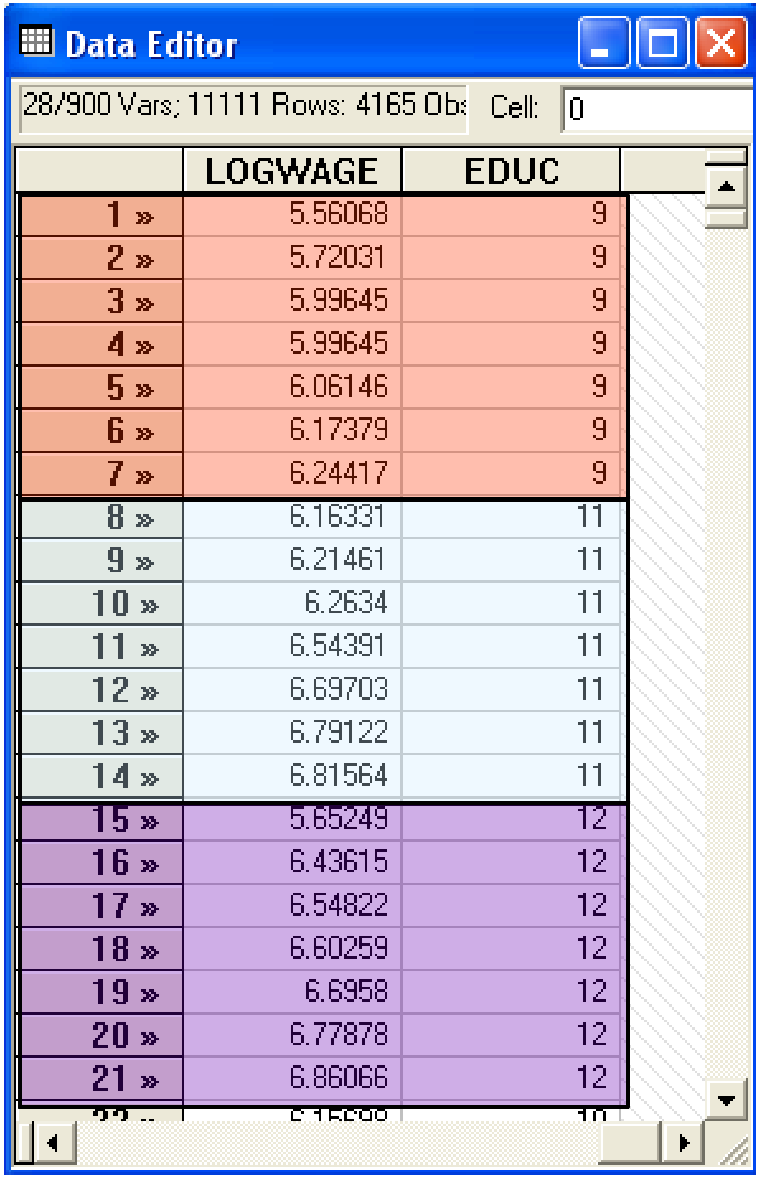
(If $T_i = 1$, this is the White estimator.)



Application: Cornwell and Rupert

```
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant       5.66098218      .04685914    120.808    .0000
 OCC           -.11220205       .01464317     -7.662    .0000     .51116447
 SMSA           .15504405       .01233744     12.567    .0000     .65378151
 MS             .09569050       .02133490      4.485    .0000     .81440576
 FEM           -.39478212       .02603413    -15.164    .0000     .11260504
 ED             .05688005       .00267743     21.244    .0000    12.8453782
 EXP            .01043785       .00054206     19.256    .0000    19.8537815
+-----------------------------------------------------------------------+
| Covariance matrix for the model is adjusted for data clustering.      |
| Sample of   4165 observations contained    595 clusters defined by    |
|      7 observations (fixed number) in each cluster.                   |
| Sample of   4165 observations contained      1 strata defined by      |
|   4165 observations (fixed number) in each stratum.                   |
+-----------------------------------------------------------------------+
 Constant       5.66098218      .10026368     56.461    .0000
 OCC           -.11220205       .02653437     -4.229    .0000     .51116447
 SMSA           .15504405       .02540156      6.104    .0000     .65378151
 MS             .09569050       .04656766      2.055    .0399     .81440576
 FEM           -.39478212       .05319458     -7.421    .0000     .11260504
 ED             .05688005       .00568214     10.010    .0000    12.8453782
 EXP            .01043785       .00131647      7.929    .0000    19.8537815
```



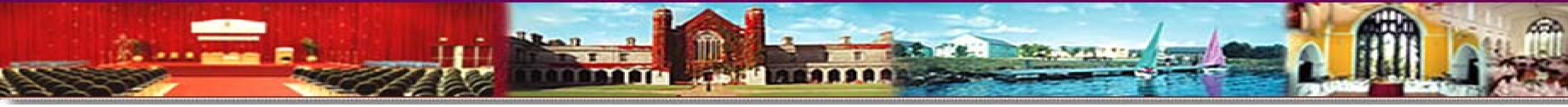


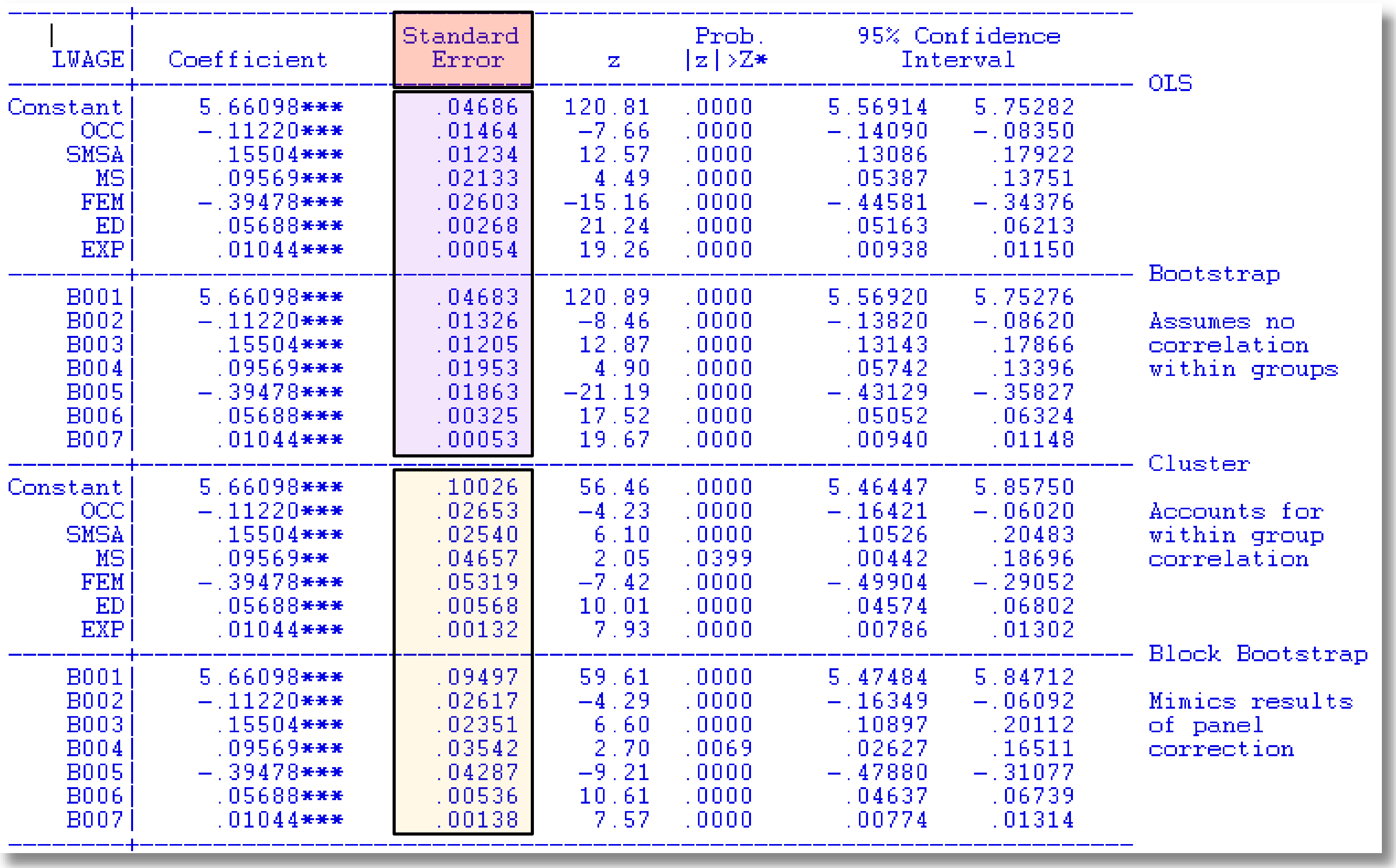
28/900 Vars; 11111 Rows: 4165 Obs Cell: 0

	LOGWAGE	EDUC
1 »	5.56068	9
2 »	5.72031	9
3 »	5.99645	9
4 »	5.99645	9
5 »	6.06146	9
6 »	6.17379	9
7 »	6.24417	9
8 »	6.16331	11
9 »	6.21461	11
10 »	6.2634	11
11 »	6.54391	11
12 »	6.69703	11
13 »	6.79122	11
14 »	6.81564	11
15 »	5.65249	12
16 »	6.43615	12
17 »	6.54822	12
18 »	6.60259	12
19 »	6.6958	12
20 »	6.77878	12
21 »	6.86066	12

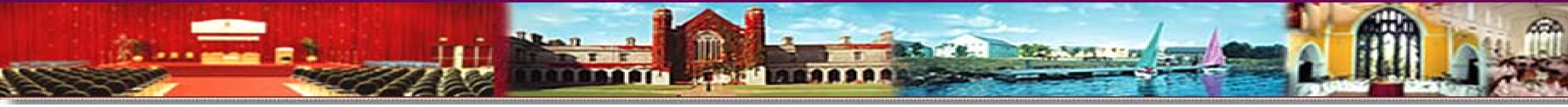
Bootstrap variance for a panel data estimator

- **Panel Bootstrap = Block Bootstrap**
- **Data set is N groups of size T_i**
- **Bootstrap sample is N groups of size T_i drawn with replacement.**





```
--------+------------------------------------------------------------------
       |                       Standard              Prob.       95% Confidence
  LWAGE|    Coefficient          Error          z    |z|>Z*          Interval
--------+------------------------------------------------------------------ OLS
Constant|     5.66098***        .04686      120.81   .0000      5.56914    5.75282
     OCC|     -.11220***        .01464       -7.66   .0000      -.14090    -.08350
    SMSA|      .15504***        .01234       12.57   .0000       .13086     .17922
      MS|      .09569***        .02133        4.49   .0000       .05387     .13751
     FEM|     -.39478***        .02603      -15.16   .0000      -.44581    -.34376
      ED|      .05688***        .00268       21.24   .0000       .05163     .06213
     EXP|      .01044***        .00054       19.26   .0000       .00938     .01150
--------+------------------------------------------------------------------ Bootstrap
    B001|     5.66098***        .04683      120.89   .0000      5.56920    5.75276
    B002|     -.11220***        .01326       -8.46   .0000      -.13820    -.08620   Assumes no
    B003|      .15504***        .01205       12.87   .0000       .13143     .17866   correlation
    B004|      .09569***        .01953        4.90   .0000       .05742     .13396   within groups
    B005|     -.39478***        .01863      -21.19   .0000      -.43129    -.35827
    B006|      .05688***        .00325       17.52   .0000       .05052     .06324
    B007|      .01044***        .00053       19.67   .0000       .00940     .01148
--------+------------------------------------------------------------------ Cluster
Constant|     5.66098***        .10026       56.46   .0000      5.46447    5.85750
     OCC|     -.11220***        .02653       -4.23   .0000      -.16421    -.06020   Accounts for
    SMSA|      .15504***        .02540        6.10   .0000       .10526     .20483   within group
      MS|      .09569**         .04657        2.05   .0399       .00442     .18696   correlation
     FEM|     -.39478***        .05319       -7.42   .0000      -.49904    -.29052
      ED|      .05688***        .00568       10.01   .0000       .04574     .06802
     EXP|      .01044***        .00132        7.93   .0000       .00786     .01302
--------+------------------------------------------------------------------ Block Bootstrap
    B001|     5.66098***        .09497       59.61   .0000      5.47484    5.84712
    B002|     -.11220***        .02617       -4.29   .0000      -.16349    -.06092   Mimics results
    B003|      .15504***        .02351        6.60   .0000       .10897     .20112   of panel
    B004|      .09569***        .03542        2.70   .0069       .02627     .16511   correction
    B005|     -.39478***        .04287       -9.21   .0000      -.47880    -.31077
    B006|      .05688***        .00536       10.61   .0000       .04637     .06739
    B007|      .01044***        .00138        7.57   .0000       .00774     .01314
--------+------------------------------------------------------------------
```



The Fixed Effects Model

$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{d}_i\alpha_i + \boldsymbol{\varepsilon}_i$, for each individual

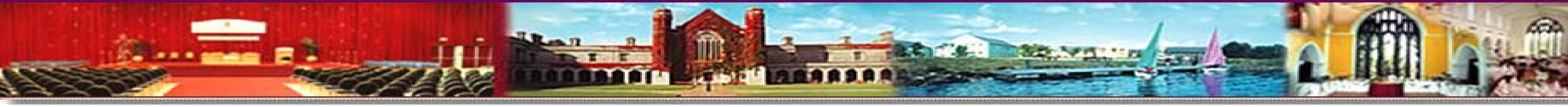
$$\begin{pmatrix}\mathbf{y}_1\\ \mathbf{y}_2\\ \vdots\\ \mathbf{y}_N\end{pmatrix} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{d}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{X}_2 & \mathbf{0} & \mathbf{d}_2 & \mathbf{0} & \mathbf{0}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \mathbf{X}_N & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{d}_N\end{bmatrix}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\alpha}\end{pmatrix} + \boldsymbol{\varepsilon}$$

$$= [\mathbf{X},\mathbf{D}\boldsymbol{\varepsilon}]\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\alpha}\end{pmatrix} +$$

$$= \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon}$$

$E[c_i \mid \mathbf{X}_i] = g(\mathbf{X}_i)$; Effects are correlated with included variables.

$\text{Cov}[\mathbf{x}_{it}, c_i] \neq \mathbf{0}$



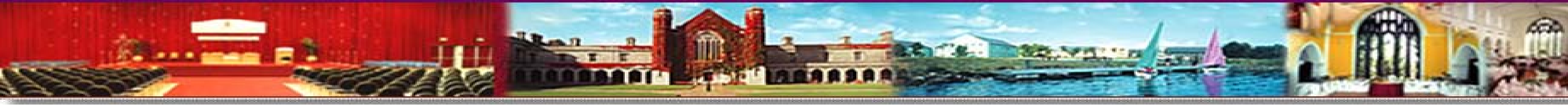
Estimating the Fixed Effects Model

- The FEM is a plain vanilla regression model but with many independent variables
- Least squares is unbiased, consistent, efficient, but inconvenient if N is large.

$$\begin{pmatrix}\mathbf{b}\\ \mathbf{a}\end{pmatrix} = \begin{bmatrix}\mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{D}\\ \mathbf{D}'\mathbf{X} & \mathbf{D}'\mathbf{D}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}'\mathbf{y}\\ \mathbf{D}'\mathbf{y}\end{bmatrix}$$

Using the Frisch-Waugh theorem

$$\mathbf{b} \quad = [\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{X}]^{-1}\left[\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{y}\right]$$



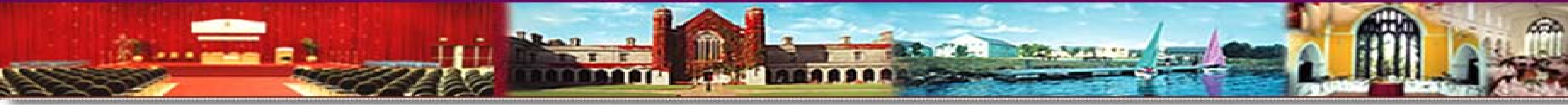
The Within Groups Transformation Removes the Effects

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + \varepsilon_{it}$$

$$\bar{y}_i = \bar{\mathbf{x}}_i\boldsymbol{\beta} + c_i + \bar{\varepsilon}_i$$

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

Use least squares to estimate **β**.



Least Squares Dummy Variable Estimator

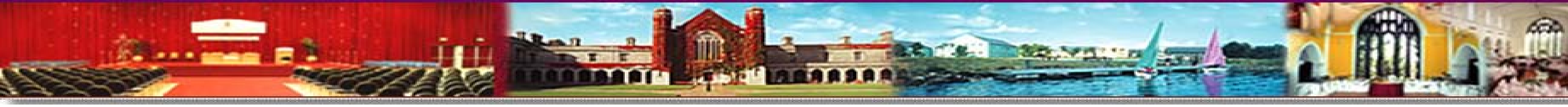
- **b** is obtained by **'within' groups least squares** (group mean deviations)

- **a** is estimated using the normal equations:

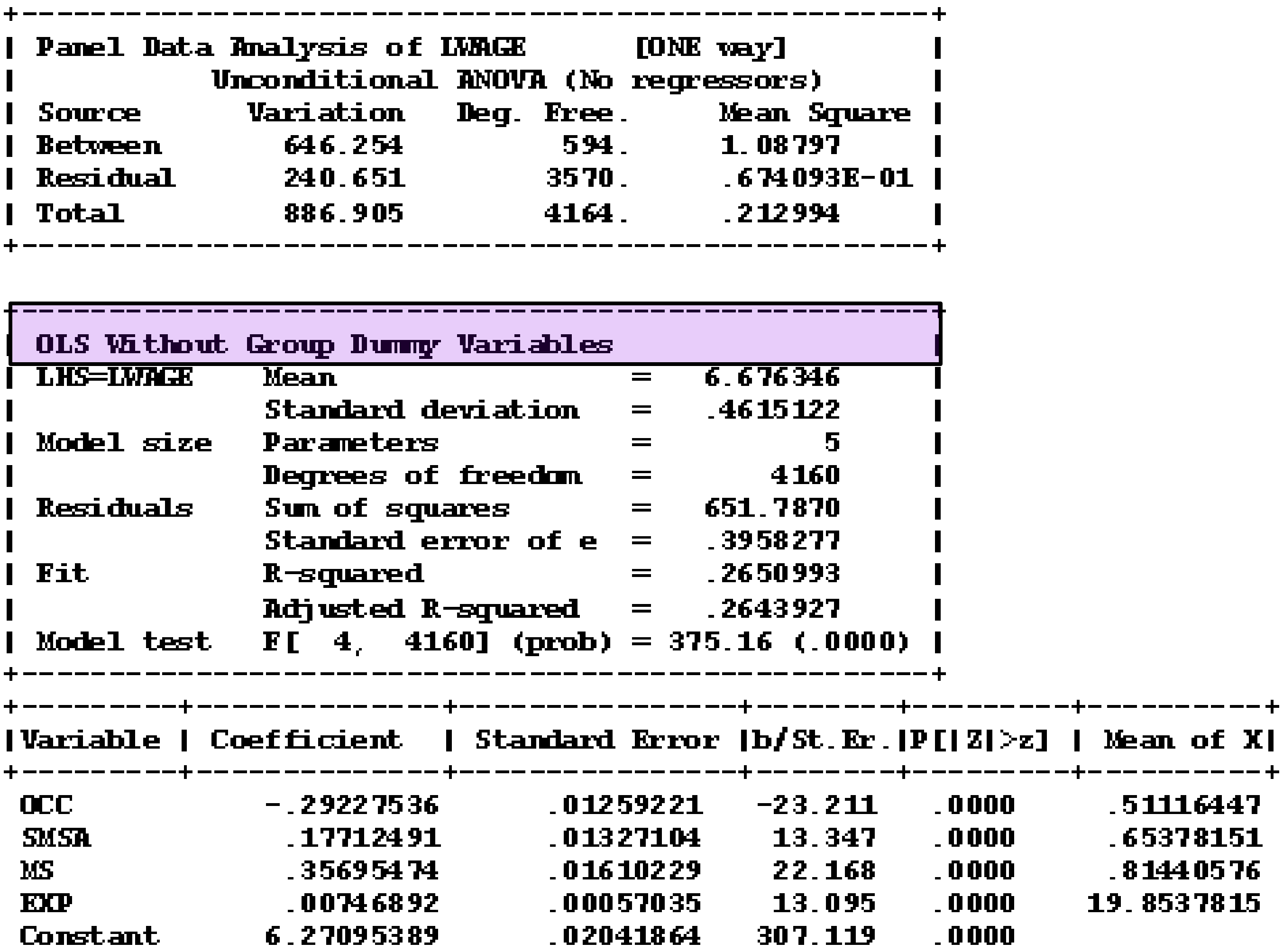
$$\mathbf{D'Xb+D'Da=D'y}$$

$$\mathbf{a = (D'D)^{-1}D'(y - Xb)}$$

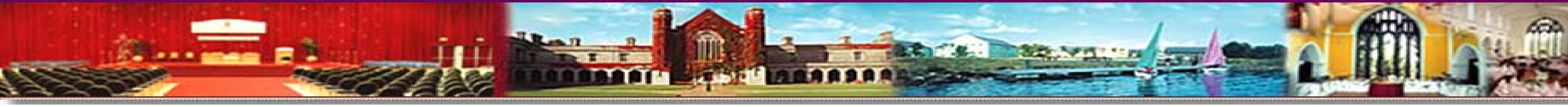
$$a_i = (1/T_i)\Sigma_{t=1}^{T_i}(y_{it} - \mathbf{x}_{it}'\mathbf{b}) = \bar{e}_i$$



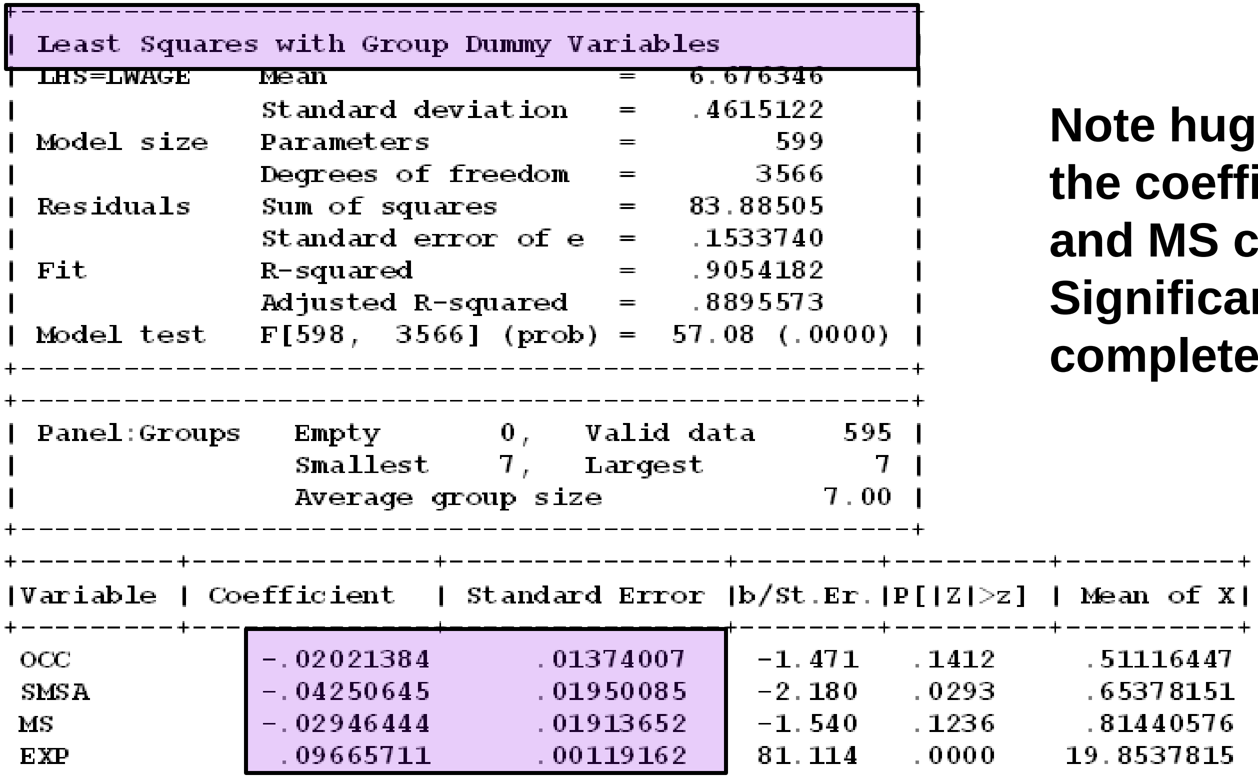
Application Cornwell and Rupert



```
+--------------------------------------------------+
| Panel Data Analysis of LWAGE      [ONE way]      |
|          Unconditional ANOVA (No regressors)     |
| Source       Variation    Deg. Free.    Mean Square |
| Between       646.254         594.      1.08797   |
| Residual      240.651        3570.      .674093E-01 |
| Total         886.905        4164.      .212994   |
+--------------------------------------------------+

+--------------------------------------------------+
| OLS Without Group Dummy Variables                |
| LHS=LWAGE     Mean                 =    6.676346 |
|               Standard deviation   =    .4615122 |
| Model size    Parameters           =           5 |
|               Degrees of freedom   =        4160 |
| Residuals     Sum of squares       =    651.7870 |
|               Standard error of e  =    .3958277 |
| Fit           R-squared            =    .2650993 |
|               Adjusted R-squared   =    .2643927 |
| Model test    F[  4,  4160] (prob) = 375.16 (.0000) |
+--------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 OCC          -.29227536       .01259221   -23.211   .0000      .51116447
 SMSA          .17712491       .01327104    13.347   .0000      .65378151
 MS            .35695474       .01610229    22.168   .0000      .81440576
 EXP           .00746892       .00057035    13.095   .0000    19.8537815
 Constant     6.27095389       .02041864   307.119   .0000
```



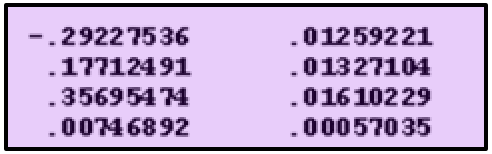
LSDV Results



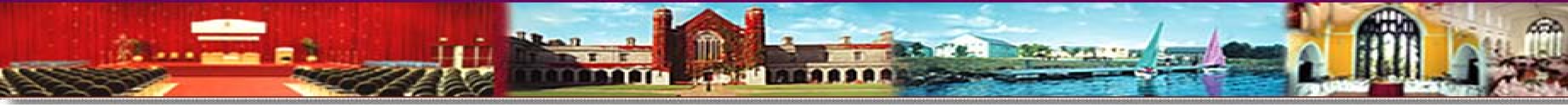
```
+----------------------------------------------------+
| Least Squares with Group Dummy Variables           |
| LHS=LWAGE     Mean                 =   6.676346    |
|               Standard deviation   =   .4615122    |
| Model size    Parameters           =        599    |
|               Degrees of freedom   =       3566    |
| Residuals     Sum of squares       =   83.88505    |
|               Standard error of e  =   .1533740    |
| Fit           R-squared            =   .9054182    |
|               Adjusted R-squared   =   .8895573    |
| Model test    F[598,  3566] (prob) =  57.08 (.0000) |
+----------------------------------------------------+
+----------------------------------------------------+
| Panel:Groups    Empty        0,   Valid data    595 |
|                 Smallest     7,   Largest         7 |
|                 Average group size             7.00 |
+----------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 OCC        -.02021384       .01374007     -1.471    .1412     .51116447
 SMSA       -.04250645       .01950085     -2.180    .0293     .65378151
 MS         -.02946444       .01913652     -1.540    .1236     .81440576
 EXP         .09665711       .00119162     81.114    .0000   19.8537815
```

Note huge changes in the coefficients. SMSA and MS change signs. Significance changes completely!

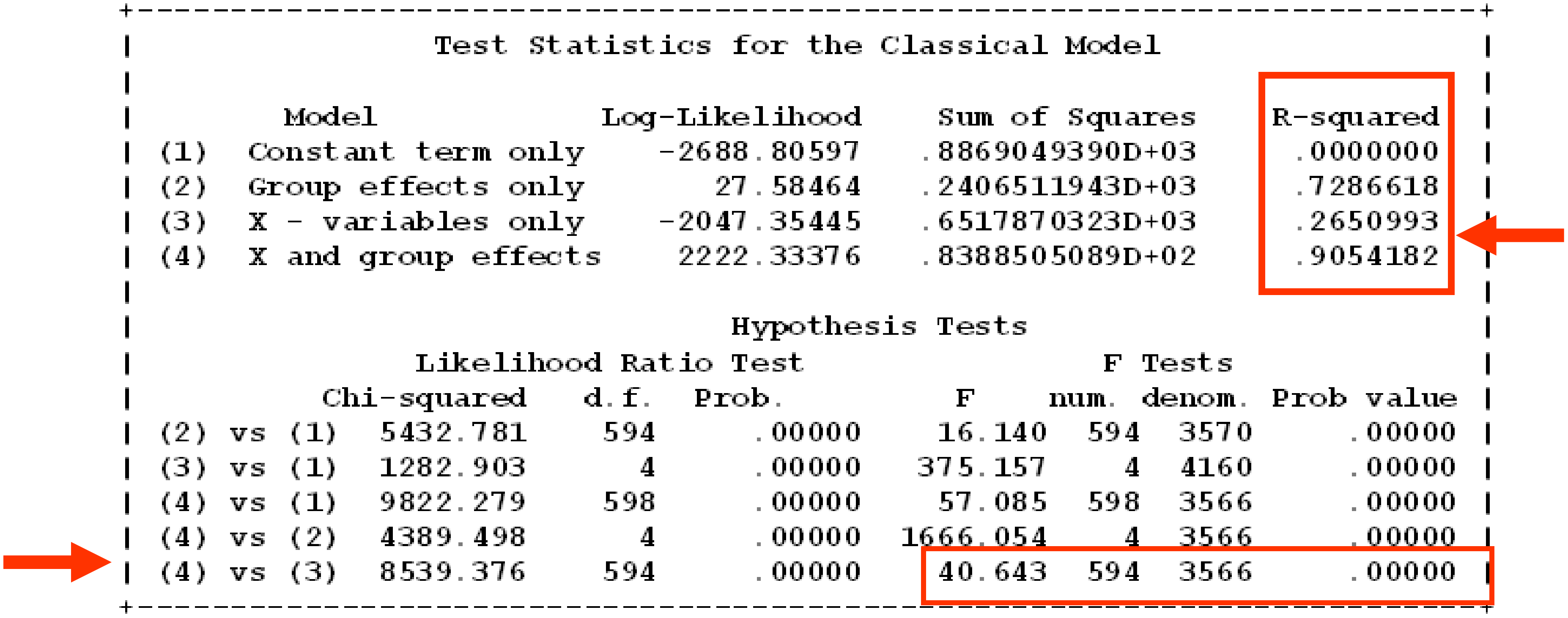
Pooled OLS



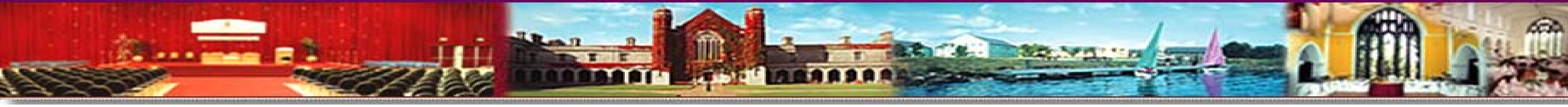
```
-.29227536      .01259221
 .17712491      .01327104
 .35695474      .01610229
 .00746892      .00057035
```



The Effect of the Effects



```
+----------------------------------------------------------------------+
|               Test Statistics for the Classical Model                |
|                                                                      |
|       Model                Log-Likelihood    Sum of Squares    R-squared |
| (1)  Constant term only     -2688.80597    .8869049390D+03    .0000000 |
| (2)  Group effects only        27.58464    .2406511943D+03    .7286618 |
| (3)  X - variables only     -2047.35445    .6517870323D+03    .2650993 |
| (4)  X and group effects     2222.33376    .8388505089D+02    .9054182 |
|                                                                      |
|                          Hypothesis Tests                            |
|              Likelihood Ratio Test               F Tests             |
|            Chi-squared   d.f.   Prob.          F     num.  denom.  Prob value |
| (2) vs (1)  5432.781      594    .00000     16.140   594   3570    .00000 |
| (3) vs (1)  1282.903        4    .00000    375.157     4   4160    .00000 |
| (4) vs (1)  9822.279      598    .00000     57.085   598   3566    .00000 |
| (4) vs (2)  4389.498        4    .00000   1666.054     4   3566    .00000 |
| (4) vs (3)  8539.376      594    .00000     40.643   594   3566    .00000 |
+----------------------------------------------------------------------+
```



A Caution About Stata and R^2

$$\text{R squared} = 1 - \frac{\text{Residual Sum of Squares}}{\text{Total Sum of Squares}}$$

Or is it? What is the total sum of squares?

$$\text{Conventional: Total Sum of Squares} = \sum_{i=1}^{N} \sum_{t=1}^{T_i} (y_{it} - \bar{y})^2$$

$$\text{"Within Sum of Squares"} = \sum_{i=1}^{N} \sum_{t=1}^{T_i} (y_{it} - \bar{y}_i)^2$$

For the FE model above,

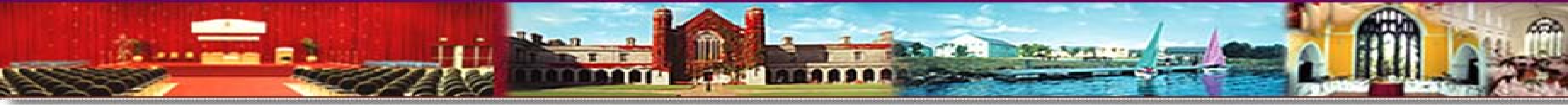
$R^2 = 0.90542$

$R^2 = 0.65142$

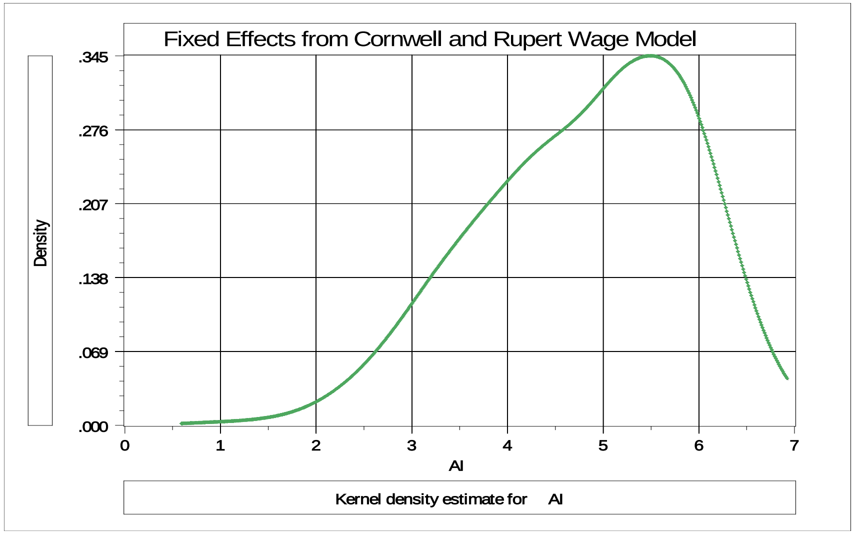
Which should appear in the denominator of R^2

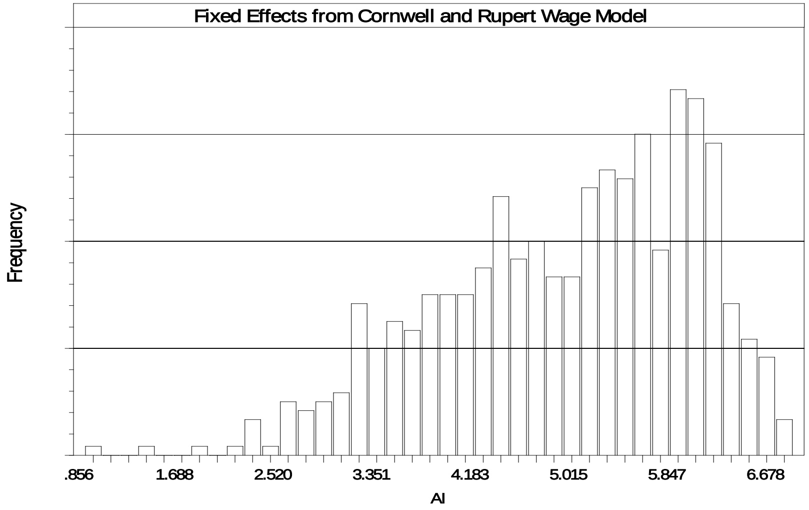
The coefficient estimates and standard errors are the same. The calculation of the R^2 is different. In the **areg** procedure, you are estimating coefficients for each of your covariates plus each dummy variable for your groups. In the **xtreg, fe** procedure the R^2 reported is obtained by only fitting a mean deviated model where the effects of the groups (all of the dummy variables) are assumed to be fixed quantities. So, all of the effects for the groups are simply subtracted out of the model and no attempt is made to quantify their overall effect on the fit of the model.

Since the SSE is the same, the R^2=1−SSE/SST is very different. The difference is real in that we are making different assumptions with the two approaches. In the **xtreg, fe** approach, the effects of the groups are fixed and **unestimated quantities are subtracted out of the model** before the fit is performed. In the **areg** approach, the group effects are estimated and affect the total sum of squares of the model under consideration.

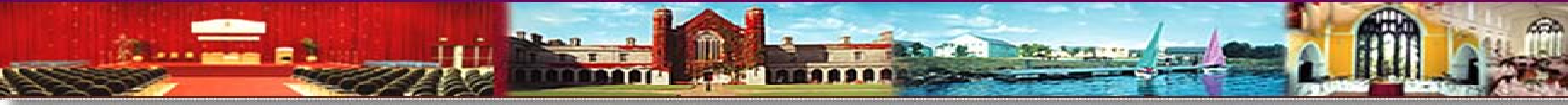


Examining the Effects with a KDE



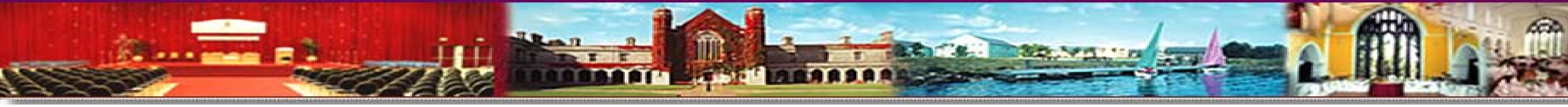


Mean = 4.819,
Standard deviation = 1.054.



Robust Covariance Matrix for LSDV Cluster Estimator for Within Estimator

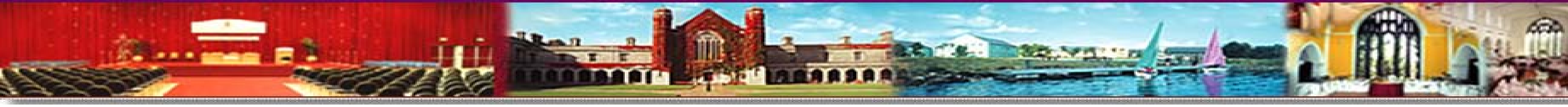
```
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
|OCC     |    -.02021           .01374007    -1.471   .1412    .5111645|
|SMSA    |    -.04251**         .01950085    -2.180   .0293    .6537815|
|MS      |    -.02946           .01913652    -1.540   .1236    .8144058|
|EXP     |     .09666***        .00119162    81.114   .0000   19.853782|
+--------+------------------------------------------------------------+
+---------------------------------------------------------------------+
| Covariance matrix for the model is adjusted for data clustering.    |
| Sample of   4165 observations contained    595 clusters defined by  |
|      7 observations (fixed number) in each cluster.                 |
+---------------------------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
|DOCC    |    -.02021           .01982162    -1.020   .3078      .00000|
|DSMSA   |    -.04251           .03091685    -1.375   .1692      .00000|
|DMS     |    -.02946           .02635035    -1.118   .2635      .00000|
|DEXP    |     .09666***        .00176599    54.732   .0000      .00000|
+--------+------------------------------------------------------------+
```



Time Invariant Regressors

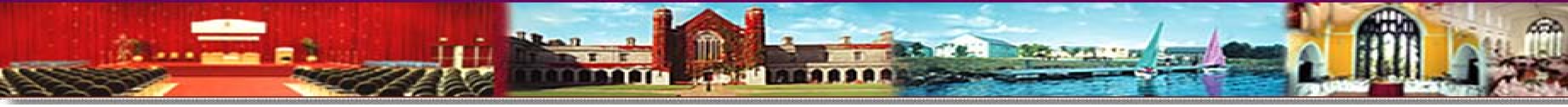
- Time invariant $\mathbf{x}_{it}$ is defined as invariant for all i. E.g., sex dummy variable, FEM and ED (education in the Cornwell/Rupert data).

- If $\mathbf{x}_{it,k}$ is invariant for all t, then the group mean deviations are all 0.



FE With Time Invariant Variables

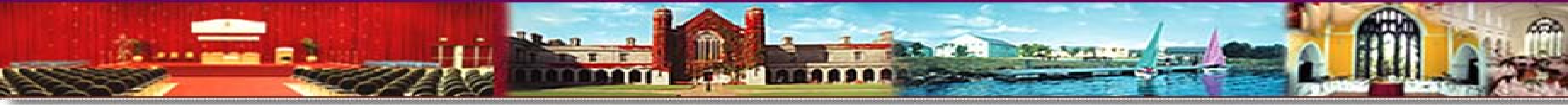
```
+----------------------------------------------------+
| There are  3 vars. with no within group variation. |
| FEM      ED        BLK                             |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 EXP     |      .09671227       .00119137    81.177   .0000   19.8537815
 WKS     |      .00118483       .00060357     1.963   .0496   46.8115246
 OCC     |     -.02145609       .01375327    -1.560   .1187     .51116447
 SMSA    |     -.04454343       .01946544    -2.288   .0221     .65378151
 FEM     |        .000000     ......(Fixed Parameter).......
 ED      |        .000000     ......(Fixed Parameter).......
 BLK     |        .000000     ......(Fixed Parameter).......
+--------------------------------------------------------------------+
|             Test Statistics for the Classical Model                |
+--------------------------------------------------------------------+
|       Model               Log-Likelihood    Sum of Squares  R-squared |
|(1)  Constant term only     -2688.80597         886.90494     .00000 |
|(2)  Group effects only        27.58464         240.65119     .72866 |
|(3)  X - variables only     -1688.12010         548.51596     .38154 |
|(4)  X and group effects     2223.20087          83.85013     .90546 |
+--------------------------------------------------------------------+
```



Drop The Time Invariant Variables Same Results

```
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 EXP     |    .09671227       .00119087    81.211   .0000   19.8537815
 WKS     |    .00118483       .00060332     1.964   .0495   46.8115246
 OCC     |   -.02145609       .01374749    -1.561   .1186     .51116447
 SMSA    |   -.04454343       .01945725    -2.289   .0221     .65378151

+--------------------------------------------------------------------+
|             Test Statistics for the Classical Model                |
+--------------------------------------------------------------------+
|       Model              Log-Likelihood    Sum of Squares  R-squared |
|(1)  Constant term only    -2688.80597       886.90494      .00000 |
|(2)  Group effects only      27.58464        240.65119      .72866 |
|(3)  X - variables only    -1688.12010       548.51596      .38154 |
|(4)  X and group effects    2223.20087        83.85013      .90546 |
+--------------------------------------------------------------------+
```

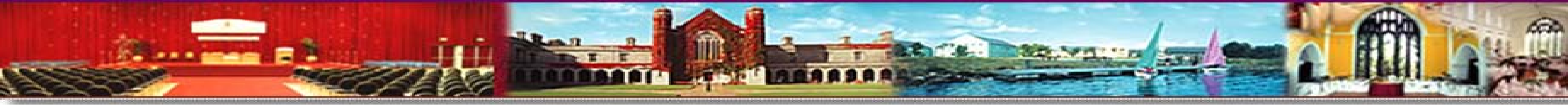
No change in the sum of squared residuals



Fixed Effects Vector Decomposition

Efficient Estimation of Time Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects

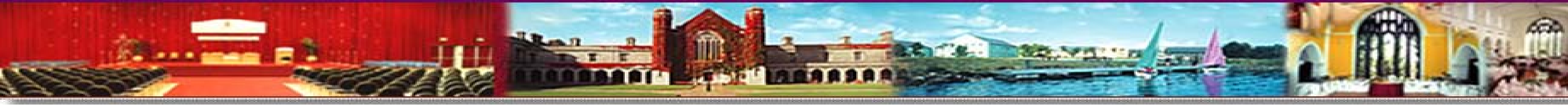
Thomas Plümper and Vera Troeger
Political Analysis, 2007



Introduction

[T]he FE model ... does not allow the estimation of time invariant variables. A second drawback of the FE model ... results from its **inefficiency** in estimating the effect of variables that have very little within variance.

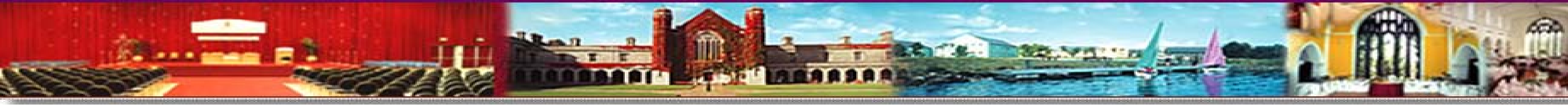
This article discusses a remedy to the related problems of estimating time invariant and rarely changing variables in FE models with unit effects



The Model

$$\boxed{y_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k x_{kit} + \sum_{m=1}^{M} \gamma_m z_{mi} + \varepsilon_{it}}$$

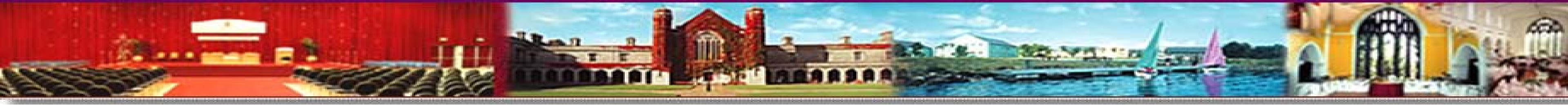
where α_i denote the N unit effects.



Fixed Effects Vector Decomposition

Step 1: Compute the fixed effects regression to get the “estimated unit effects.” “We run this FE model with the sole intention to obtain estimates of the unit effects, α_i.”

$$\boxed{\hat{\alpha}_i = \bar{y}_i - \sum_{k=1}^{K} b_k^{FE} \bar{x}_{ki}}$$



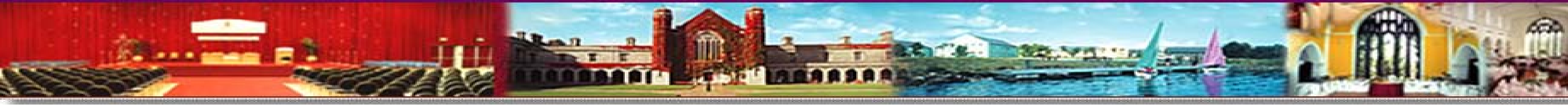
Step 2

Regress a_i on $\mathbf{z}_i$ and compute residuals

$$\boxed{a_i = \sum_{m=1}^{M} \gamma_m z_{im} + h_i}$$

h_i is orthogonal to $\mathbf{z}_i$ (since it is a residual)

Vector $\mathbf{h}_i$ is expanded so each element h_i is replicated T_i times - $\mathbf{h}$ is the length of the full sample.

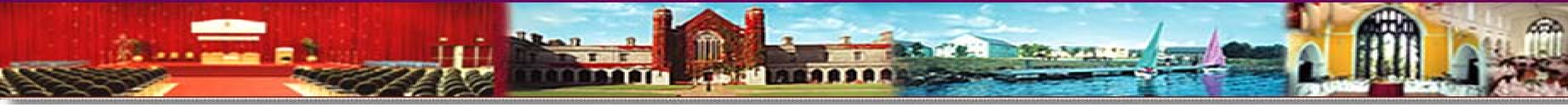


Step 3

Regress y_{it} on a constant, **X**, **Z** and **h** using ordinary least squares to estimate α, **β**, **γ**, δ.

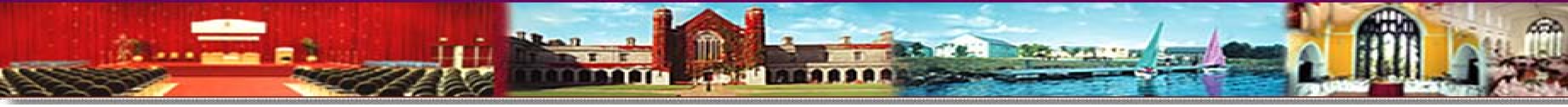
$$y_{it} = \alpha + \sum_{k=1}^{K}\beta_k x_{kit} + \sum_{m=1}^{M}\gamma_m z_{mi} + \delta h_i + \varepsilon_{it}$$

Notice that α_i in the original model has become $\alpha + \delta h_i$ in the revised model.



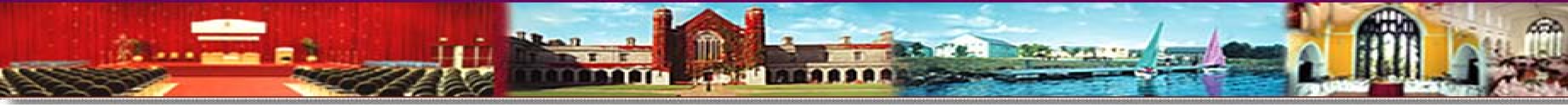
Step 1 (Based on full sample)

```
These  3 variables have no within group variation.
FEM      ED        BLK
F.E. estimates are based on a generalized inverse.
--------+---------------------------------------------------------
        |                  Standard            Prob.       Mean
   LWAGE| Coefficient        Error       z    z>|Z|       of X
--------+---------------------------------------------------------
     EXP|     .09663***      .00119    81.13  .0000    19.8538
     WKS|     .00114*        .00060     1.88  .0600    46.8115
     OCC|    -.02496*        .01390    -1.80  .0724     .51116
     IND|     .02042         .01558     1.31  .1899     .39544
   SOUTH|    -.00091         .03457     -.03  .9791     .29028
    SMSA|    -.04581**       .01955    -2.34  .0191     .65378
   UNION|     .03411**       .01505     2.27  .0234     .36399
     FEM|       .000    .....(Fixed Parameter).....     .11261
      ED|       .000    .....(Fixed Parameter).....    12.8454
     BLK|       .000    .....(Fixed Parameter).....     .07227
--------+---------------------------------------------------------
```



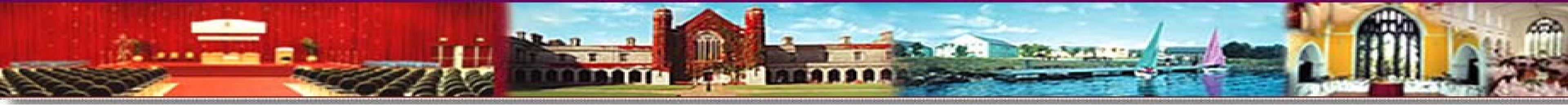
Step 2 (Based on 595 observations)

```
--------+--------------------------------------------------------------
        |                  Standard            Prob.           Mean
     UHI| Coefficient        Error        z    z>|Z|           of X
--------+--------------------------------------------------------------
Constant|    2.88090***      .07172    40.17   .0000
     FEM|    -.09963**       .04842    -2.06   .0396          .11261
      ED|     .14616***      .00541    27.02   .0000         12.8454
     BLK|    -.27615***      .05954    -4.64   .0000          .07227
--------+--------------------------------------------------------------
```



Step 3!

```
--------+--------------------------------------------------------
        |                  Standard            Prob.       Mean
   LWAGE|  Coefficient       Error         z    z>|Z|       of X
--------+--------------------------------------------------------
Constant|     2.88090***      .03282     87.78  .0000
     EXP|      .09663***      .00061    157.53  .0000    19.8538
     WKS|      .00114***      .00044      2.58  .0098    46.8115
     OCC|     -.02496***      .00601     -4.16  .0000     .51116
     IND|      .02042***      .00479      4.26  .0000     .39544
   SOUTH|     -.00091         .00510      -.18  .8590     .29028
    SMSA|     -.04581***      .00506     -9.06  .0000     .65378
   UNION|      .03411***      .00521      6.55  .0000     .36399
     FEM|     -.09963***      .00767    -13.00  .0000     .11261
      ED|      .14616***      .00122    120.19  .0000    12.8454
     BLK|     -.27615***      .00894    -30.90  .0000     .07227
      HI|     1.00000***      .00670    149.26  .0000  -.103D-13
--------+--------------------------------------------------------
```



The Magic

Step 1

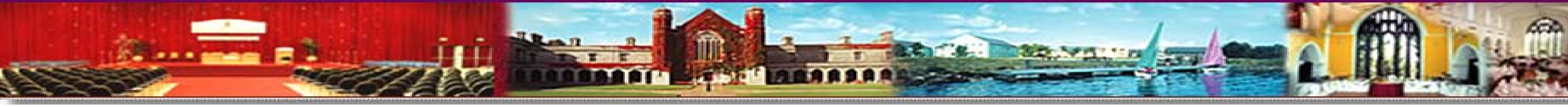
```
--------+--------------------------
        |                 Standard
   LWAGE| Coefficient       Error
--------+--------------------------
     EXP|     .09663***     .00119
     WKS|     .00114*       .00060
     OCC|    -.02496*       .01390
     IND|     .02042        .01558
   SOUTH|    -.00091        .03457
    SMSA|    -.04581**      .01955
   UNION|     .03411**      .01505
```

Step 2

```
--------+--------------------------
        |                 Standard
     UHI| Coefficient       Error
--------+--------------------------
Constant|   2.88090***      .07172
     FEM|   -.09963**       .04842
      ED|    .14616***      .00541
     BLK|   -.27615***      .05954
--------+--------------------------
```

Step 3

```
+----------------------------
|                   Standard
| Coefficient         Error
+----------------------------
|   2.88090***        .03282
|    .09663***        .00061
|    .00114***        .00044
|   -.02496***        .00601
|    .02042***        .00479
|   -.00091           .00510
|   -.04581***        .00506
|    .03411***        .00521
|   -.09963***        .00767
|    .14616***        .00122
|   -.27615***        .00894
|   1.00000***        .00670
+----------------------------
```



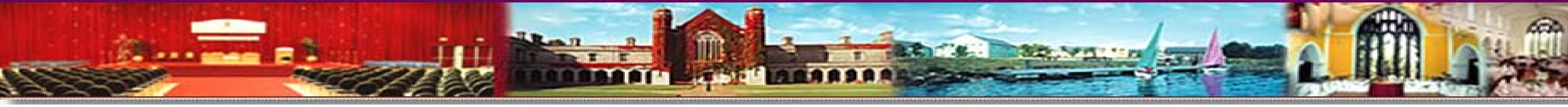
What happened here?

$$y_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k x_{kit} + \sum_{m=1}^{M} \gamma_m z_{mi} + \varepsilon_{it}$$

where α_i denote the N unit effects.

An assumption is added along the way

$\text{Cov}(\alpha_i, Z_i) = \mathbf{0}$. This is exactly the number of orthogonality assumptions needed to identify γ. It is not part of the original model.



The Random Effects Model

- The random effects model

  $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + \varepsilon_{it}$, observation for person i at time t

  $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{c}_i + \boldsymbol{\varepsilon}_i$, T_i observations in group i

  $= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{c}_i + \boldsymbol{\varepsilon}_i$, note $\mathbf{c}_i = (c_i, c_i, ..., c_i)'$

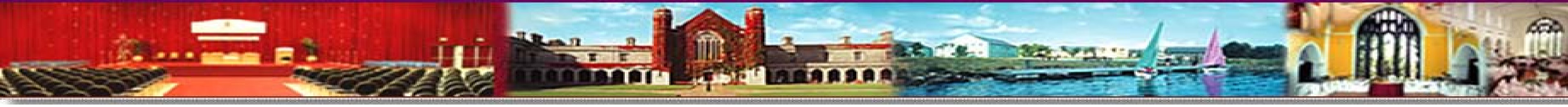
  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{c} + \boldsymbol{\varepsilon}$, $\Sigma_{i=1}^{N} T_i$ observations in the sample

  $c = (\mathbf{c}_1', \mathbf{c}_2', ...\mathbf{c}_N')', \Sigma_{i=1}^{N} T_i$ by 1 vector

- c_i is uncorrelated with $\mathbf{x}_{it}$ for all t;

  $E[c_i \,|\mathbf{X}_i] = 0$

  $E[\varepsilon_{it}|\mathbf{X}_i, c_i] = 0$



Error Components Model

A Generalized Regression Model

$$y_{it} = \mathbf{x}_{it}'\mathbf{b} + \varepsilon_{it} + u_i$$

$$E[\varepsilon_{it} \mid \mathbf{X}_i] = 0$$

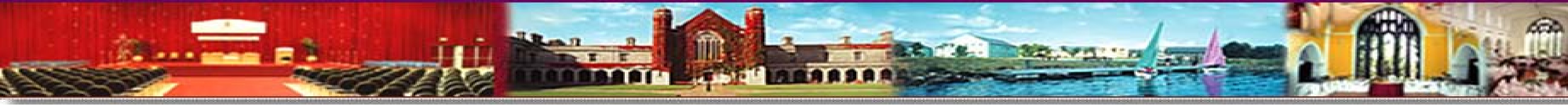
$$E[\varepsilon_{it}^2 \mid \mathbf{X}_i] = \sigma_\varepsilon^2$$

$$E[u_i \mid \mathbf{X}_i] = 0$$

$$E[u_i^2 \mid \mathbf{X}_i] = \sigma_u^2$$

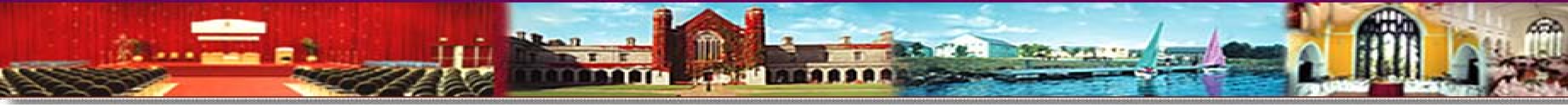
$$\mathrm{Var}[\boldsymbol{\varepsilon}_i + u_i\mathbf{i}] = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_\varepsilon^2 + \sigma_u^2 & \cdots & \sigma_u^2 \\ \cdots & \cdots & \ddots & \cdots \\ \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_\varepsilon^2 + \sigma_u^2 \end{bmatrix} = \mathbf{\Omega}_i$$

$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i + u_i\mathbf{i}$ for T_i observations



Random vs. Fixed Effects

- Random Effects
  - Small number of parameters
  - Efficient estimation
  - Objectionable orthogonality assumption ($c_i \perp \mathbf{X}_i$)
- Fixed Effects
  - Robust – generally consistent
  - Large number of parameters



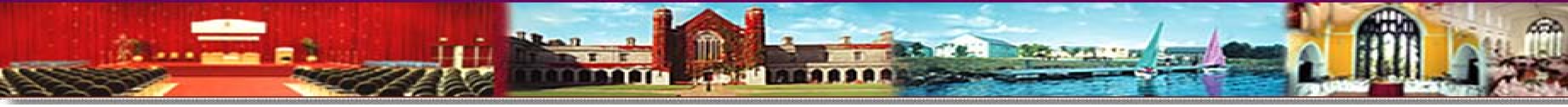
Ordinary Least Squares

- Standard results for OLS in a GR model
  - Consistent
  - Unbiased
  - Inefficient
- True variance of the least squares estimator

$$\text{Var}[\mathbf{b}\mid\mathbf{X}] = \frac{\mathbf{1}}{\Sigma_{i=1}^{N} T_i}\left[\frac{\mathbf{X\Omega X}}{\Sigma_{i=1}^{N} T_i}\right]^{-1}\frac{\mathbf{XX}}{\Sigma_{i=1}^{N} T_i}\left[\frac{'}{\Sigma_{i=1}^{N} T_i}\right]^{-1}$$

$$\rightarrow\ \mathbf{0}\ \times\ \rightarrow \mathbf{Q}^{\mathbf{-1}} \times \rightarrow \mathbf{Q}\ ^{*} \times \rightarrow \mathbf{Q}^{\mathbf{-1}}$$

$$\rightarrow\ \mathbf{0}\ \text{as } N \rightarrow \infty$$



Estimating the Variance for OLS

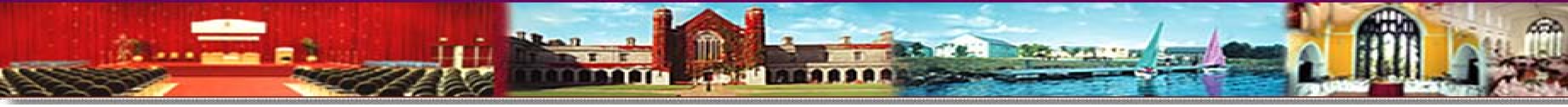
$$\text{Var}[\mathbf{b}\,|\,\mathbf{X}] = \frac{\mathbf{1}}{\Sigma_{i=1}^{N} T_i}\left[\frac{\mathbf{X\Omega X}}{\Sigma_{i=1}^{N} T_i}\right]^{-1}\left(\frac{\mathbf{XX}}{\Sigma_{i=1}^{N} T_i}\right)\left[\frac{'}{\Sigma_{i=1}^{N} T_i}\right]^{-1}$$

In the spirit of the White estimator, use

$$\frac{\mathbf{X\hat{\Omega}X}}{\Sigma_{i=1}^{N} T_i} = \Sigma_{i=1}^{N} f_i \frac{\mathbf{X}_i'\hat{\mathbf{w}}_i\hat{\mathbf{w}}_i'\mathbf{X}_i}{T_i},\ \ \hat{\mathbf{w}}_i = \mathbf{y}_i - \mathbf{X}_i\mathbf{b},\ \ f_i = \frac{T_i}{\Sigma_{i=1}^{N} T_i}$$

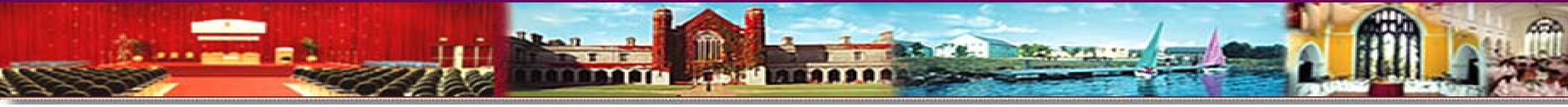
Hypothesis tests are then based on Wald statistics.

THIS IS THE 'CLUSTER' ESTIMATOR



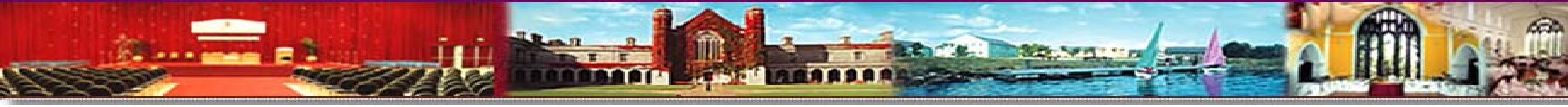
OLS Results for Cornwell and Rupert

```
+----------------------------------------------------+
| Residuals    Sum of squares        =   522.2008    |
|              Standard error of e   =   .3544712    |
| Fit          R-squared             =   .4112099    |
|              Adjusted R-squared    =   .4100766    |
+----------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant      5.40159723      .04838934   111.628   .0000
 EXP            .04084968      .00218534    18.693   .0000    19.8537815
 EXPSQ        -.00068788    .480428D-04   -14.318   .0000    514.405042
 OCC          -.13830480      .01480107    -9.344   .0000     .51116447
 SMSA           .14856267      .01206772    12.311   .0000     .65378151
 MS             .06798358      .02074599     3.277   .0010     .81440576
 FEM          -.40020215      .02526118   -15.843   .0000     .11260504
 UNION          .09409925      .01253203     7.509   .0000     .36398559
 ED             .05812166      .00260039    22.351   .0000    12.8453782
```



Alternative Variance Estimators

```
+---------+--------------+----------------+--------+---------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] |
+---------+--------------+----------------+--------+---------+
 Constant       5.40159723        .04838934   111.628   .0000
 EXP             .04084968        .00218534    18.693   .0000
 EXPSQ          -.00068788      .480428D-04   -14.318   .0000
 OCC            -.13830480        .01480107    -9.344   .0000
 SMSA            .14856267        .01206772    12.311   .0000
 MS              .06798358        .02074599     3.277   .0010
 FEM            -.40020215        .02526118   -15.843   .0000
 UNION           .09409925        .01253203     7.509   .0000
 ED              .05812166        .00260039    22.351   .0000
Robust – Cluster
 Constant       5.40159723        .10156038    53.186   .0000
 EXP             .04084968        .00432272     9.450   .0000
 EXPSQ          -.00068788      .983981D-04    -6.991   .0000
 OCC            -.13830480        .02772631    -4.988   .0000
 SMSA            .14856267        .02423668     6.130   .0000
 MS              .06798358        .04382220     1.551   .1208
 FEM            -.40020215        .04961926    -8.065   .0000
 UNION           .09409925        .02422669     3.884   .0001
 ED              .05812166        .00555697    10.459   .0000
```



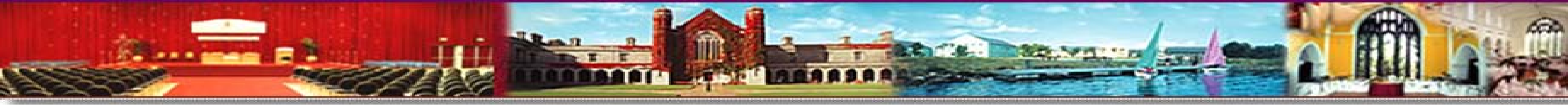
Generalized Least Squares

GLS is equivalent to OLS regression of

$y_{it}* = y_{it} - \theta_i \overline{y}_{i}.$ on $\mathbf{x}_{it}* = \mathbf{x}_{it} - \theta_i \overline{\mathbf{x}}_{i}.$,

where $\theta_i = 1 - \dfrac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T_i\sigma_u^2}}$

$\text{Asy.Var}[\hat{\boldsymbol{\beta}}] = [\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}]^{-1} = \sigma_\varepsilon^2[\mathbf{X}'*\mathbf{X}*]^{-1}$



Estimators for the Variances

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \varepsilon_{it} + u_i$$

Using the OLS estimator of **β**, $\mathbf{b}_{OLS}$,

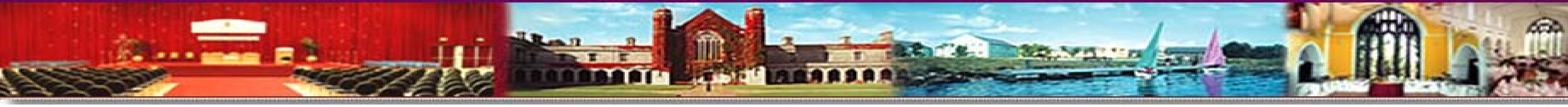
$$\boxed{\frac{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - a - \mathbf{x}_{it}'\mathbf{b})^2}{\left(\Sigma_{i=1}^{N} T_i\right) - 1 - K} \text{ estimates } \sigma_\varepsilon^2 + \sigma_U^2}$$

With the LSDV estimates, a_i and $\mathbf{b}_{LSDV}$,

$$\boxed{\frac{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - a_{\mathbf{i}} - \mathbf{x}_{it}'\mathbf{b})^2}{\left(\Sigma_{i=1}^{N} T_i\right) - N - K} \text{ estimates } \sigma_\varepsilon^2}$$

Using the difference of the two,

$$\boxed{\left[\frac{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - a - \mathbf{x}_{it}'\mathbf{b})^2}{\left(\Sigma_{i=1}^{N} T_i\right) - 1 - K}\right] - \left[\frac{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - a_{\mathbf{i}} - \mathbf{x}_{it}'\mathbf{b})^2}{\left(\Sigma_{i=1}^{N} T_i\right) - N - K}\right] \text{ estimates } \sigma_U^2}$$



Practical Problems with FGLS

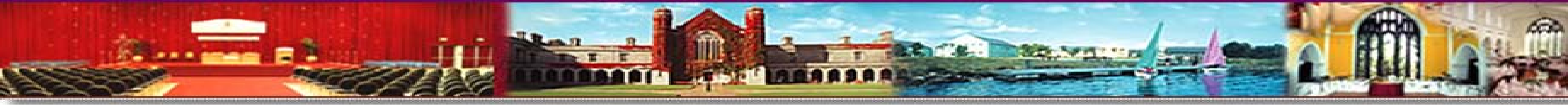
- The preceding regularly produce negative estimates of σ_u^2.
- Estimation is made very complicated in unbalanced panels.

A bulletproof solution (originally used in TSP, now NLOGIT and others).

From the robust LSDV estimator: $\hat{\sigma}_\varepsilon^2 = \dfrac{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - a_i - \mathbf{x}_{it}'\mathbf{b}_{LSDV})^2}{\Sigma_{i=1}^{N} T_i}$

From the pooled OLS estimator: $\text{Est}(\sigma_\varepsilon^2 + \sigma_u^2) = \dfrac{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - a_{OLS} - \mathbf{x}_{it}'\mathbf{b}_{OLS})^2}{\Sigma_{i=1}^{N} T_i} \geq \hat{\sigma}_\varepsilon^2$

$$\boxed{\hat{\sigma}_u^2 = \frac{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - a_{OLS} - \mathbf{x}_{it}'\mathbf{b}_{OLS})^2 - \Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - a_i - \mathbf{x}_{it}'\mathbf{b}_{LSDV})^2}{\Sigma_{i=1}^{N} T_i} \geq 0}$$



Stata Variance Estimators

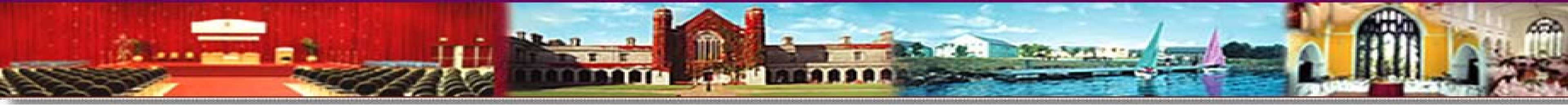
$$\hat{\sigma}_{\varepsilon}^2 = \frac{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - a_i - \mathbf{x}_{it}'\mathbf{b}_{LSDV})^2}{\Sigma_{i=1}^{N}T_i - K - N} > 0 \text{ based on FE estimates}$$

$$\hat{\sigma}_u^2 = \text{Max}\left[0, \frac{\text{SSE(group means)}}{N-A} - \frac{(N-K)\hat{\sigma}_{\varepsilon}^2}{(N-A)\bar{T}}\right] \geq 0$$

where A = K or if $\hat{\sigma}_u^2$ is negative,

A=trace of a matrix that somewhat resembles $\mathbf{I}_K$.

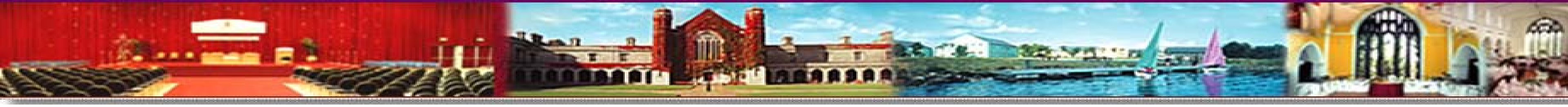
Many other adjustments exist. None guaranteed to be positive. No optimality properties or even guaranteed consistency.



Application

```
+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]              =   .231188D-01  |
|             Var[u]              =   .102531D+00  |
|             Corr[v(i,t),v(i,s)] =   .816006      |
| Variance estimators are based on OLS residuals.  |
+--------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 EXP            .08819204       .00224823    39.227   .0000   19.8537815
 EXPSQ         -.00076604     .496074D-04   -15.442   .0000   514.405042
 OCC           -.04243576       .01298466    -3.268   .0011     .51116447
 SMSA          -.03404260       .01620508    -2.101   .0357     .65378151
 MS            -.06708159       .01794516    -3.738   .0002     .81440576
 FEM           -.34346104       .04536453    -7.571   .0000     .11260504
 UNION          .05752770       .01350031     4.261   .0000     .36398559
 ED             .11028379       .00510008    21.624   .0000   12.8453782
 Constant      4.01913257       .07724830    52.029   .0000
```

No problems arise in this sample.



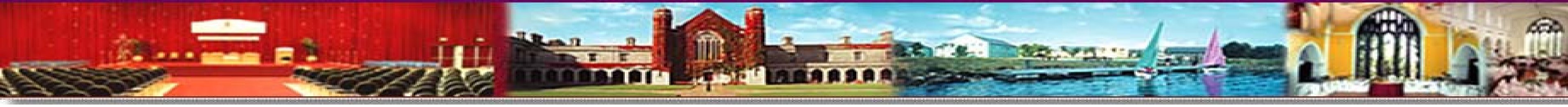
Testing for Effects: An LM Test

Breusch and Pagan Lagrange Multiplier statistic

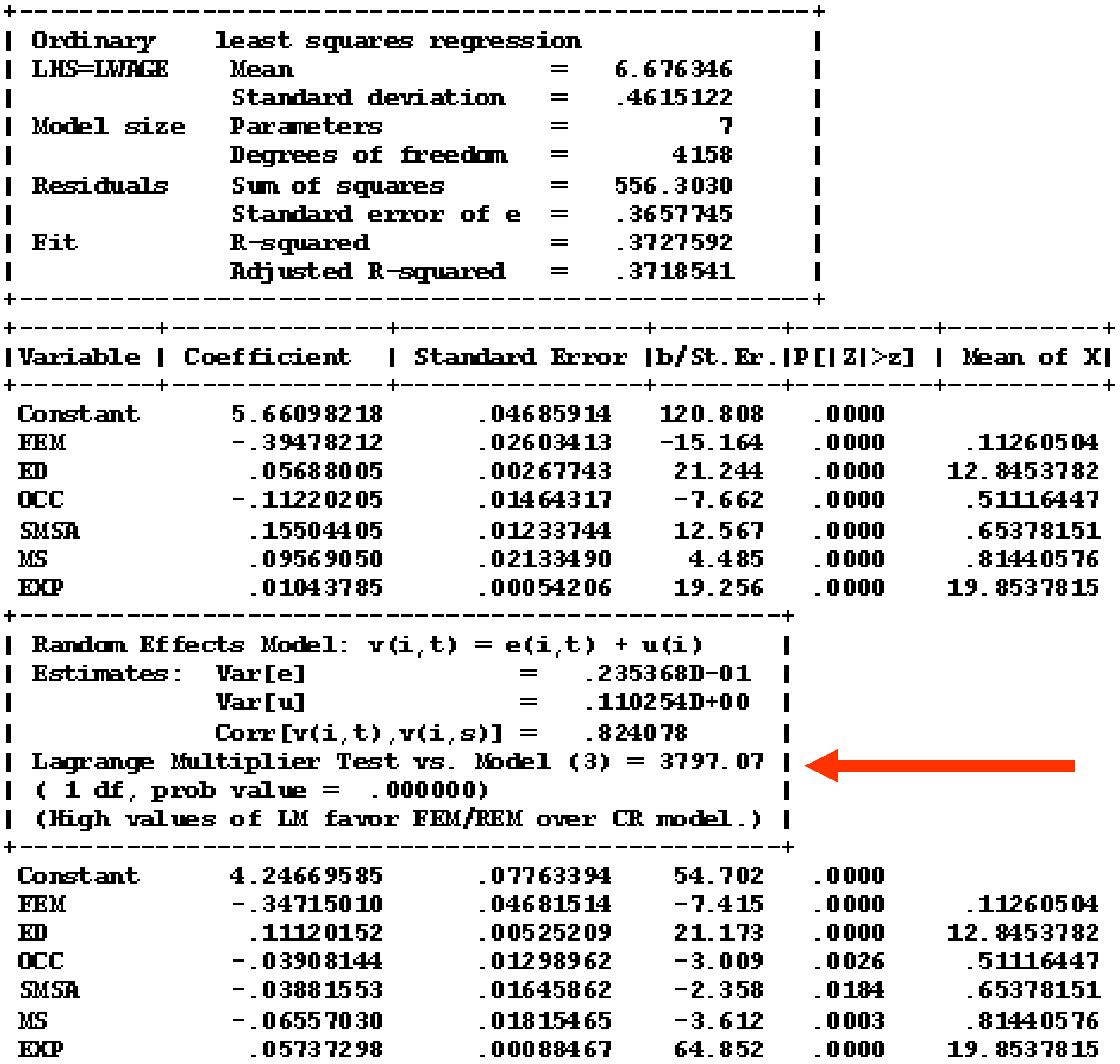
$$y_{it} = \beta' x_{it} + u_i + \varepsilon_{it}, \ u_i \text{ and } \varepsilon_{it} \sim \text{Normal}\left[\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\sigma_u^2 & 0\\ 0 & \sigma_\varepsilon^2\end{pmatrix}\right]$$

$$H_0 : \ \sigma_u^2 = 0$$

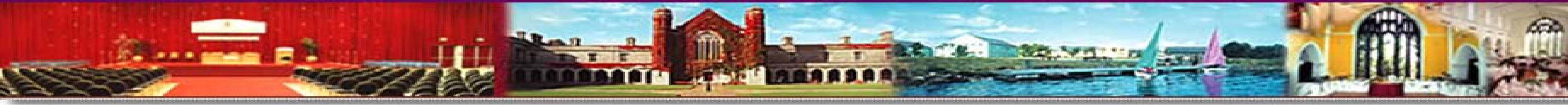
$$LM = \frac{(\Sigma_{i=1}^N T_i)^2}{2\Sigma_{i=1}^N T_i (T_i - 1)}\left[\frac{\Sigma_{i=1}^N (T_i \bar{e}_i)^2}{\Sigma_{i=1}^N \Sigma_{t=1}^T e_{it}^2} - 1\right]^2 \longrightarrow \chi^2[1]$$



Application: Cornwell-Rupert

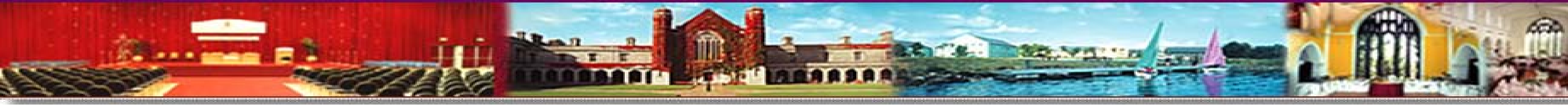


```
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=LWAGE     Mean                 =    6.676346   |
|               Standard deviation   =    .4615122   |
| Model size    Parameters           =           7   |
|               Degrees of freedom   =        4158   |
| Residuals     Sum of squares       =    556.3030   |
|               Standard error of e  =    .3657745   |
| Fit           R-squared            =    .3727592   |
|               Adjusted R-squared   =    .3718541   |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant      5.66098218       .04685914    120.808   .0000
 FEM           -.39478212       .02603413    -15.164   .0000     .11260504
 ED             .05688005       .00267743     21.244   .0000   12.8453782
 OCC           -.11220205       .01464317     -7.662   .0000     .51116447
 SMSA           .15504405       .01233744     12.567   .0000     .65378151
 MS             .09569050       .02133490      4.485   .0000     .81440576
 EXP            .01043785       .00054206     19.256   .0000   19.8537815
+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]              =   .235368D-01  |
|             Var[u]              =   .110254D+00  |
|             Corr[v(i,t),v(i,s)] =   .824078      |
| Lagrange Multiplier Test vs. Model (3) = 3797.07 |
| ( 1 df, prob value =  .000000)                   |
| (High values of LM favor FEM/REM over CR model.) |
+--------------------------------------------------+
 Constant      4.24669585       .07763394     54.702   .0000
 FEM           -.34715010       .04681514     -7.415   .0000     .11260504
 ED             .11120152       .00525209     21.173   .0000   12.8453782
 OCC           -.03908144       .01298962     -3.009   .0026     .51116447
 SMSA          -.03881553       .01645862     -2.358   .0184     .65378151
 MS            -.06557030       .01815465     -3.612   .0003     .81440576
 EXP            .05737298       .00088467     64.852   .0000   19.8537815
```



Hausman Test for FE vs. RE

Estimator	Random Effects $E[c_i\|\mathbf{X}_i] = 0$	Fixed Effects $E[c_i\|\mathbf{X}_i] \neq 0$
FGLS (Random Effects)	Consistent and Efficient	Inconsistent
LSDV (Fixed Effects)	Consistent Inefficient	Consistent Possibly Efficient



Computing the Hausman Statistic

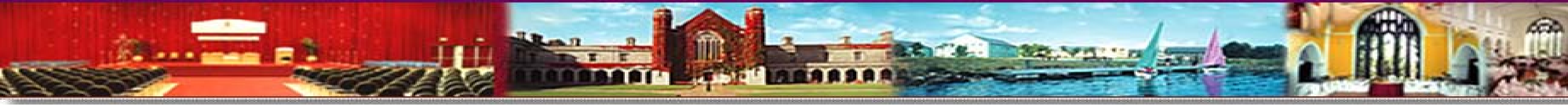
$$\text{Est.Var}[\hat{\boldsymbol{\beta}}_{\mathbf{FE}}] = \hat{\sigma}_{\varepsilon}^{2}\left[\Sigma_{i=1}^{N}\mathbf{X}_{i}'\left(\mathbf{I}-\frac{1}{T_{i}}\mathbf{ii}'\right)\mathbf{X}_{\mathbf{i}}\right]^{-1}$$

$$\text{Est.Var}[\hat{\boldsymbol{\beta}}_{\mathbf{RE}}] = \hat{\sigma}_{\varepsilon}^{2}\left[\Sigma_{i=1}^{N}\mathbf{X}_{i}'\left(\mathbf{I}-\frac{\hat{\gamma}_{i}}{T_{i}}\mathbf{ii}'\right)\mathbf{X}_{\mathbf{i}}\right]^{-1},\ 0\ \leq\ \hat{\gamma}_{i}\ =\ \frac{T_{i}\hat{\sigma}_{u}^{2}}{\hat{\sigma}_{\varepsilon}^{2}+T_{i}\hat{\sigma}_{u}^{2}}\ \leq\ 1$$

As long as $\hat{\sigma}_{\varepsilon}^{2}$ and $\hat{\sigma}_{u}^{2}$ are consistent, as N $\rightarrow\infty$, $\text{Est.Var}[\hat{\boldsymbol{\beta}}_{\mathbf{FE}}]-\text{Est.Var}[\hat{\boldsymbol{\beta}}_{\mathbf{RE}}]$ will be nonnegative definite. In a finite sample, to ensure this, both must be computed using the same estimate of $\hat{\sigma}_{\varepsilon}^{2}$. The one based on LSDV will generally be the better choice.

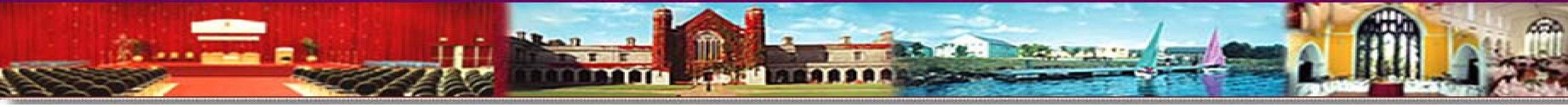
Note that columns of zeros will appear in $\text{Est.Var}[\hat{\boldsymbol{\beta}}_{\mathbf{FE}}]$ if there are time invariant variables in **X**.

β does not contain the constant term in the preceding.



Hausman Test

```
+----------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)       |
| Estimates:  Var[e]              =   .235368D-01  |
|             Var[u]              =   .110254D+00  |
|             Corr[v(i,t),v(i,s)] =   .824078      |
| Lagrange Multiplier Test vs. Model (3) = 3797.07 |
| ( 1 df, prob value =  .000000)                   |
| (High values of LM favor FEM/REM over CR model.) |
| Fixed vs. Random Effects (Hausman)     = 2632.34 |
| ( 4 df, prob value =  .000000)                   |
| (High (low) values of H favor FEM (REM).)        |
+----------------------------------------------------+
```



Variable Addition

A Fixed Effects Model

$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$

LSDV estimator - Deviations from group means:

To estimate $\boldsymbol{\beta}$, regress $(y_{it} - \overline{y}_i)$ on $(\mathbf{x}_{it} - \overline{\mathbf{x}}_i)$

Algebraic equivalent: OLS regress y_{it} on $(\mathbf{x}_{it}, \overline{\mathbf{x}}_i)$

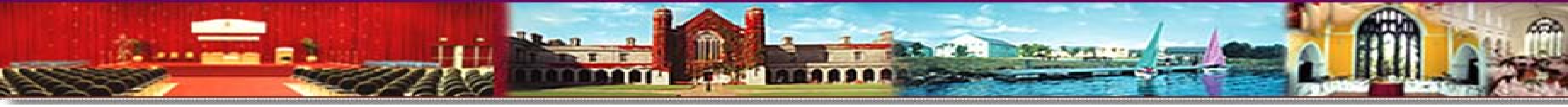
Mundlak interpretation: $\alpha_i = \alpha + \boldsymbol{\delta}'\overline{\mathbf{x}}_i + u_i$

Model becomes $y_{it} = \alpha + \boldsymbol{\delta}'\overline{\mathbf{x}}_i + u_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$

$= \alpha + \boldsymbol{\delta}'\overline{\mathbf{x}}_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i$

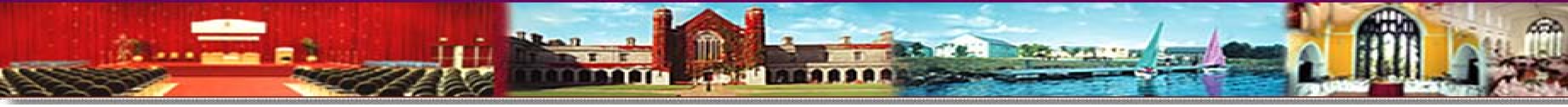
= a random effects model with the group means.

Estimate by FGLS.



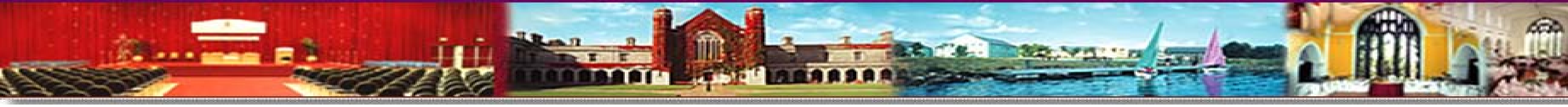
A Variable Addition Test

- Asymptotic equivalent to Hausman
- Also equivalent to Mundlak formulation
- In the random effects model, using FGLS
  - Only applies to time varying variables
  - Add expanded group means to the regression (i.e., observation i,t gets same group means for all t.
  - Use Wald test to test for coefficients on means equal to 0. Large chi-squared weighs against random effects specification.



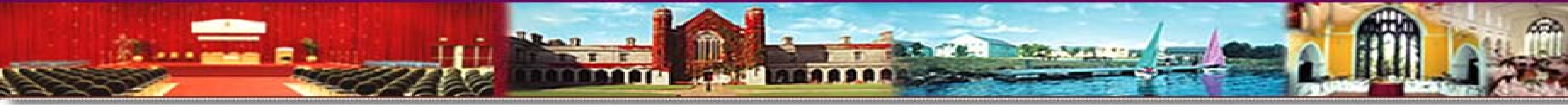
Fixed Effects

```
+----------------------------------------------------+
| Panel:Groups   Empty       0,   Valid data     595 |
|                Smallest    7,   Largest          7 |
|                Average group size             7.00 |
| There are  3 vars. with no within group variation. |
| ED       BLK      FEM                              |
| Look for huge standard errors and fixed parameters.|
| F.E. results are based on a generalized inverse.   |
| They will be highly erratic. (Problematic model.)  |
| Unable to compute std.errors for dummy var. coeffs.|
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
|WKS     |     .00083          .00060003     1.381   .1672   46.811525|
|OCC     |    -.02157          .01379216    -1.564   .1178    .5111645|
|IND     |     .01888          .01545450     1.221   .2219    .3954382|
|SOUTH   |     .00039          .03429053      .011   .9909    .2902761|
|SMSA    |    -.04451**        .01939659    -2.295   .0217    .6537815|
|UNION   |     .03274**        .01493217     2.192   .0283    .3639856|
|EXP     |     .11327***       .00247221    45.819   .0000   19.853782|
|EXPSQ   |    -.00042***     .546283D-04    -7.664   .0000   514.40504|
|ED      |       .000        ......(Fixed Parameter).......           |
|BLK     |       .000        ......(Fixed Parameter).......           |
|FEM     |       .000        ......(Fixed Parameter).......           |
+--------+------------------------------------------------------------+
```



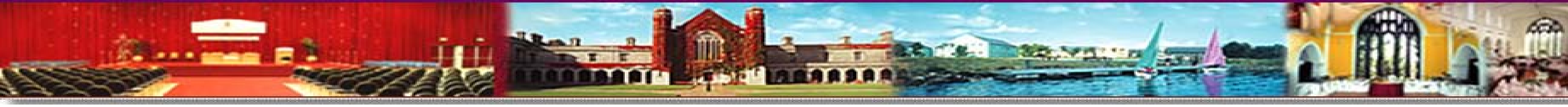
Random Effects

```
+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]              =    .235368D-01 |
|             Var[u]              =    .110254D+00 |
|             Corr[v(i,t),v(i,s)] =    .824078     |
| Lagrange Multiplier Test vs. Model (3) = 3797.07 |
| ( 1 df, prob value =  .000000)                   |
| (High values of LM favor FEM/REM over CR model.) |
+--------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
|WKS     |     .00094          .00059308    1.586   .1128   46.811525|
|OCC     |    -.04367***        .01299206   -3.361   .0008    .5111645|
|IND     |     .00271          .01373256     .197   .8434    .3954382|
|SOUTH   |    -.00664          .02246416    -.295   .7677    .2902761|
|SMSA    |    -.03117*         .01615455   -1.930   .0536    .6537815|
|UNION   |     .05802***        .01349982    4.298   .0000    .3639856|
|EXP     |     .08744***        .00224705   38.913   .0000   19.853782|
|EXPSQ   |    -.00076***      .495876D-04  -15.411   .0000   514.40504|
|ED      |     .10724***        .00511463   20.967   .0000   12.845378|
|BLK     |    -.21178***        .05252013   -4.032   .0001    .0722689|
|FEM     |    -.24786***        .04283536   -5.786   .0000    .1126050|
|Constant|    3.97756***        .08178139   48.637   .0000            |
+--------+------------------------------------------------------------+
```



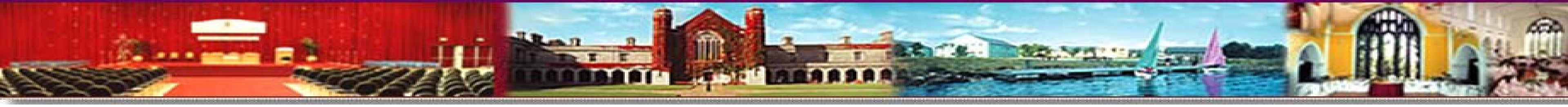
The Hausman Test, by Hand

```
--> matrix;  br=b(1:8) ; vr=varb(1:8,1:8)$
--> matrix ; db = bf - br ; dv = vf - vr $
--> matrix ; list ; h =db'<dv>db$

Matrix H        has  1 rows and  1 columns.
               1
        +--------------
       1| 2523.64910

--> calc;list;ctb(.95,8)$
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 Result  =     15.507313
```



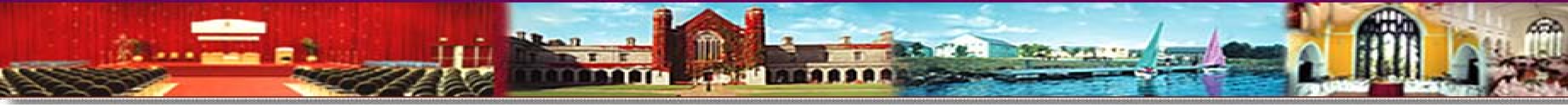
Means Added to REM - Mundlak

```
+--------+-------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+----------------+--------+--------+----------+
|WKS     |    .00083           .00060070     1.380   .1677   46.811525|
|OCC     |   -.02157           .01380769    -1.562   .1182    .5111645|
|IND     |    .01888           .01547189     1.220   .2224    .3954382|
|SOUTH   |    .00039           .03432914      .011   .9909    .2902761|
|SMSA    |   -.04451**         .01941842    -2.292   .0219    .6537815|
|UNION   |    .03274**         .01494898     2.190   .0285    .3639856|
|EXP     |    .11327***        .00247500    45.768   .0000   19.853782|
|EXPSQ   |   -.00042***     .546898D-04    -7.655   .0000   514.40504|
|ED      |    .05199***        .00552893     9.404   .0000   12.845378|
|BLK     |   -.16983***        .04456572    -3.811   .0001    .0722689|
|FEM     |   -.41306***        .03732204   -11.067   .0000    .1126050|
|WKSB    |    .00863**         .00363907     2.371   .0177   46.811525|
|OCCB    |   -.14656***        .03640885    -4.025   .0001    .5111645|
|INDB    |    .04142           .02976363     1.392   .1640    .3954382|
|SOUTHB  |   -.05551           .04297816    -1.292   .1965    .2902761|
|SMSAB   |    .21607***        .03213205     6.724   .0000    .6537815|
|UNIONB  |    .08152**         .03266438     2.496   .0126    .3639856|
|EXPB    |   -.08005***        .00533603   -15.002   .0000   19.853782|
|EXPSQB  |   -.00017           .00011763    -1.416   .1567   514.40504|
|Constant|   5.19036***        .20147201    25.762   .0000             |
+--------+-----------------------------------------------------------+
```



Wu (Variable Addition) Test

```
--> matrix ; bm=b(12:19);vm=varb(12:19,12:19)$
--> matrix ; list ; wu = bm'<vm>bm $

Matrix WU       has  1 rows and  1 columns.
               1
        +--------------
       1| 3004.38076
```



A Hierarchical Linear Model Interpretation of the FE Model

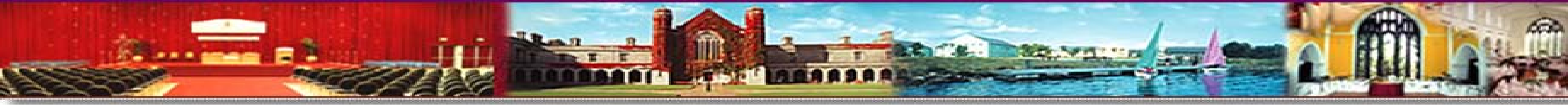
$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + c_i + \varepsilon_{it}$$ ($\mathbf{x}_{it}$ does not contain a constant)

$$E[\varepsilon_{it}|\mathbf{X}_i, c_i] = 0, \text{Var}[\varepsilon_{it}|\mathbf{X}_i, c_i] = \sigma_\varepsilon^2$$

$$c_i = \alpha + \mathbf{z}_i'\boldsymbol{\delta} + u_i,$$

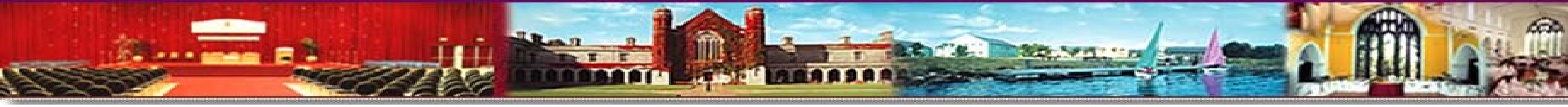
$$E[u_i|\mathbf{z}_i'] = 0, \ \text{Var}[u_i|\mathbf{z}_i'] = \sigma_u^2$$

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + [\alpha + \mathbf{z}_i'\boldsymbol{\delta} + u_i] + \quad {}_{it}$$



Hierarchical Linear Model as REM

```
+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]              =   .235368D-01  |
|             Var[u]              =   .110254D+00  |
|             Corr[v(i,t),v(i,s)] =   .824078      |
|             Sigma(u)            =  0.3303        |
+--------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 OCC     |    -.03908144       .01298962    -3.009   .0026    .51116447
 SMSA    |    -.03881553       .01645862    -2.358   .0184    .65378151
 MS      |    -.06557030       .01815465    -3.612   .0003    .81440576
 EXP     |     .05737298       .00088467    64.852   .0000   19.8537815
 FEM     |    -.34715010       .04681514    -7.415   .0000    .11260504
 ED      |     .11120152       .00525209    21.173   .0000   12.8453782
 Constant|    4.24669585       .07763394    54.702   .0000
```



Evolution: Correlated Random Effects

Unknown parameters

$y_{it} = \alpha_i + \beta' \mathbf{x}_{it} + \varepsilon_{it}, \quad \Theta = [\alpha_1, \alpha_2, ..., \alpha_N, \beta, \sigma_\varepsilon^2]$

Standard estimation based on LS (dummy variables)

Ambiguous definition of the distribution of y_{it}

Effects model, nonorthogonality, heterogeneity

$y_{it} = \alpha_i + \beta' \mathbf{x}_{it} + \varepsilon_{it}, \quad E[\alpha_i \,|\, \mathbf{X}_i] = g(\mathbf{X}_i) \neq 0$

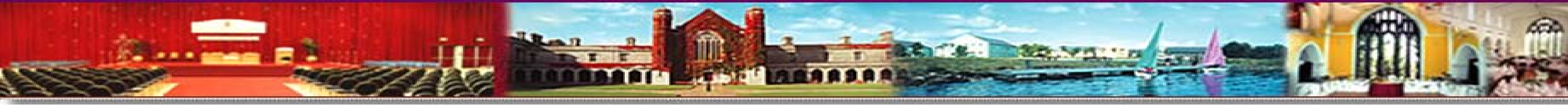
Contrast to random effects $E[\alpha_i \,|\, X_i] = \alpha$

Standard estimation (still) based on LS (dummy variables)

Correlated random effects, more detailed model

$y_{it} = \alpha_i + \beta' \mathbf{x}_{it} + \varepsilon_{it}, \quad P[\alpha_i \,|\, \mathbf{X}_i] = g(\mathbf{X}_i) \neq 0$

Linear projection? $\alpha_i = \boldsymbol{\theta}' \mathbf{x}_i + u_i \quad \text{Cor}(u_i, \mathbf{x}_i) = 0$



Mundlak's Estimator

Mundlak, Y., "On the Pooling of Time Series and Cross Section Data, Econometrica, 46, 1978, pp. 69-85.

Write $c_i = \bar{\mathbf{x}}_i\boldsymbol{\delta} + u_i$, $\boxed{E[c_i|\mathbf{x}_{i1},\mathbf{x}_{i1},..\mathbf{x}_{iT_i}] = \bar{\mathbf{x}}_i\boldsymbol{\delta}}$

Assume c_i contains all time invariant information

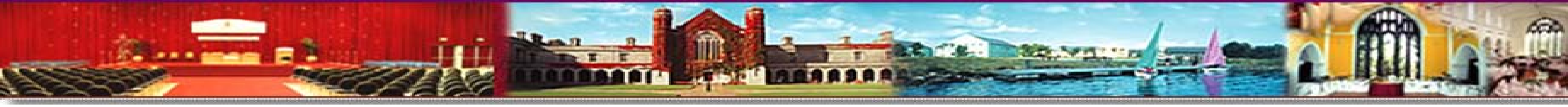
$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{i}c_i + \boldsymbol{\varepsilon}_i$, T_i observations in group i

$= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{i}\bar{\mathbf{x}}_i\boldsymbol{\delta} + \boldsymbol{\varepsilon}_i + u_i\mathbf{i}$

Looks like random effects.

$\text{Var}[\boldsymbol{\varepsilon}_i + u_i\mathbf{i}] = \boldsymbol{\Omega}_i + \boldsymbol{\sigma}_u^2\mathbf{i}\mathbf{i}'$

This is the model we used for the Wu test.



Correlated Random Effects

Mundlak

$$c_i = \bar{\mathbf{x}}_i\boldsymbol{\delta} + u_i, \quad E[c_i|\mathbf{x}_{i1},\mathbf{x}_{i1},..\mathbf{x}_{iT_i}] = \bar{\mathbf{x}}_i\boldsymbol{\delta}$$

Assume c_i contains all time invariant information

$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + c_i\mathbf{i} + \boldsymbol{\varepsilon}_i$, T_i observations in group i

$$= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{i}\bar{\mathbf{x}}_i\boldsymbol{\delta} + \boldsymbol{\varepsilon}_i + u_i\mathbf{i}$$

Chamberlain / Wooldridge

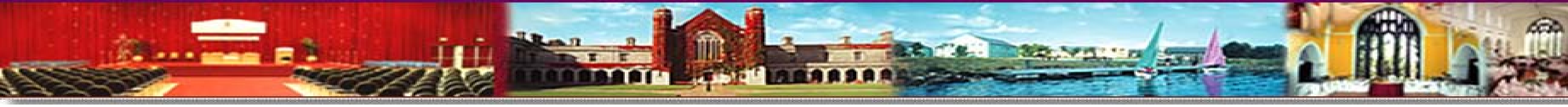
$$c_i = \mathbf{x}_{i1}'\boldsymbol{\delta}_1 + \mathbf{x}_{i2}'\boldsymbol{\delta}_2 + \ldots + \mathbf{x}_{iT}'\boldsymbol{\delta}_T + u_i$$

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{i}\mathbf{x}_{i1}'\boldsymbol{\delta}_1 + \mathbf{i}\mathbf{x}_{i1}'\boldsymbol{\delta}_2 + ..\mathbf{i}\mathbf{x}_{iT}'\boldsymbol{\delta}_T + \boldsymbol{\varepsilon}_i + u_i\mathbf{i}$$

$$\underbrace{\text{TxK}} + \underbrace{\text{TxK}} + \underbrace{\text{TxK}} + \qquad \underbrace{\text{TxK}} \quad \text{etc.}$$

Problems: Requires balanced panels

Modern panels have large T; models have large K



Mundlak's Approach for an FE Model with Time Invariant Variables

$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\delta} + c_i + \varepsilon_{it}$, ($\mathbf{z}$ does not contain a constant)

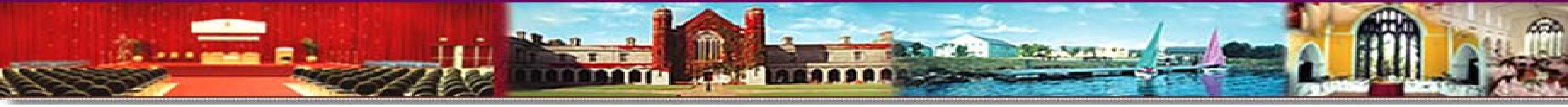
$$E[\varepsilon_{it}|\mathbf{X}_i, c_i] = 0, \text{Var}[\varepsilon_{it}|\mathbf{X}_i, c_i] = \sigma_\varepsilon^2$$

$$c_i = \alpha + \bar{\mathbf{x}}_i\boldsymbol{\theta} + w_i,$$

$$E[w_i|\mathbf{X}_i, \mathbf{z}_i] = 0, \ \text{Var}[w_i|\mathbf{X}_i, \mathbf{z}_i] = \sigma_w^2$$

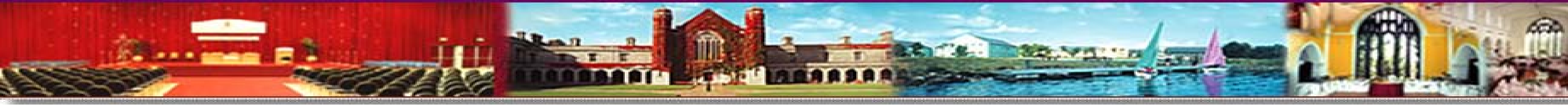
$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\delta} + \alpha + \bar{\mathbf{x}}_i\boldsymbol{\theta} + w_i + \varepsilon_{it}$$

$=$ random effects model including group means of time varying variables.



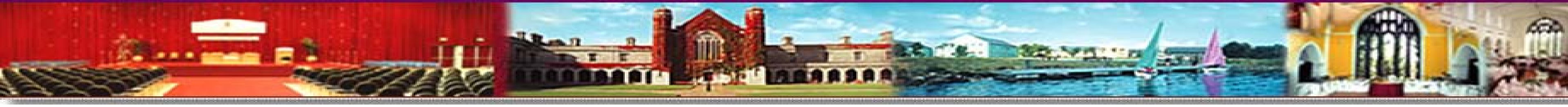
Mundlak Form of FE Model

```
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
x(i,t)=================================================================
 OCC     |    -.02021384       .01375165    -1.470   .1416    .51116447
 SMSA    |    -.04250645       .01951727    -2.178   .0294    .65378151
 MS      |    -.02946444       .01915264    -1.538   .1240    .81440576
 EXP     |     .09665711       .00119262    81.046   .0000   19.8537815
z(i)===================================================================
 FEM     |    -.34322129       .05725632    -5.994   .0000    .11260504
 ED      |     .05099781       .00575551     8.861   .0000   12.8453782
Means of x(i,t) and constant===========================================
 Constant|    5.72655261       .10300460    55.595   .0000
 OCCB    |    -.10850252       .03635921    -2.984   .0028    .51116447
 SMSAB   |     .22934020       .03282197     6.987   .0000    .65378151
 MSB     |     .20453332       .05329948     3.837   .0001    .81440576
 EXPB    |    -.08988632       .00165025   -54.468   .0000   19.8537815
Variance Estimates=====================================================
   Var[e]|      .0235632
   Var[u]|      .0773825
```



Panel Data Extensions

- Dynamic models: lagged effects of the dependent variable
- Endogenous RHS variables
- Cross country comparisons– large T
- More general parameter heterogeneity – not only the constant term
- Nonlinear models such as binary choice



The Hausman and Taylor Model

$$y_{it} = \mathbf{x1}_{it}'\boldsymbol{\beta}_1 + \boxed{\mathbf{x2}_{it}'\boldsymbol{\beta}_2} + \mathbf{z1}_i'\boldsymbol{\alpha}_1 + \boxed{\mathbf{z2}_i'\boldsymbol{\alpha}_2} + \varepsilon_{it} + \boxed{u_i}$$

Model: **x2** and **z2** are correlated with u.

Deviations from group means removes all time invariant variables

$$y_{it} - \overline{y}_i = (\mathbf{x1}_{it} - \overline{\mathbf{x1}}_i)'\boldsymbol{\beta}_1 + (\mathbf{x2}_{it} - \overline{\mathbf{x2}}_i)'\boldsymbol{\beta}_2 + \varepsilon_{it}$$

Implication: $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ are consistently estimated by LSDV.

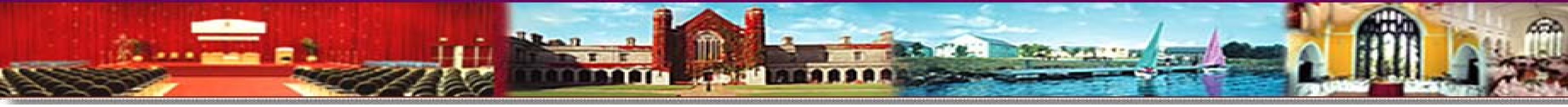
$(\mathbf{x1}_{it} - \overline{\mathbf{x1}}_i)$ = K_1 instrumental variables

$(\mathbf{x2}_{it} - \overline{\mathbf{x2}}_i)$ = K_2 instrumental variables

$\mathbf{z1}_i$ = L_1 instrumental variables (uncorrelated with u)

? = L_2 instrumental variables (where do we get them?)

H&T: $\overline{\mathbf{x1}}_i$ = K_1 additional instrumental variables. Needs $K_1 \geq L_2$.



H&T’s 4 Step FGLS Estimator

(1) LSDV estimates of $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma^2_\varepsilon$

(2) $(\mathbf{e}*)' = (\bar{e}_1, \bar{e}_1, \ldots, \bar{e}_1), (\bar{e}_2, \bar{e}_2, \ldots, \bar{e}_2), \ldots, (\bar{e}_N, \bar{e}_N, \ldots, \bar{e}_N)$

IV regression of **e *** on **Z *** with instruments

$\mathbf{W}_i$ consistently estimates $_1$ and $_2$.

(3) With fixed T, residual variance in (2) estimates $\sigma^2_u + \sigma^2_\varepsilon / T$

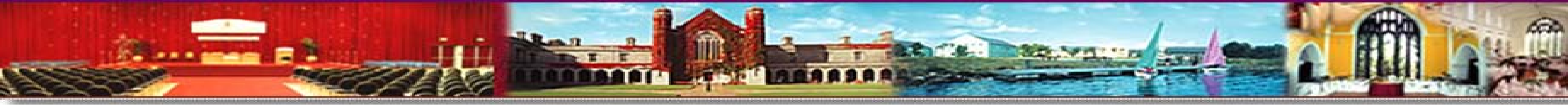
With unbalanced panel, it estimates $\sigma^2_u + \sigma^2_\varepsilon \overline{(1/T)}$ or something resembling this. (1) provided an estimate of σ^2_ε so use the two to obtain estimates of σ^2_u and σ^2_ε. For each group, compute

$$\hat{\theta}_i = 1 - \sqrt{\hat{\sigma}^2_\varepsilon / (\hat{\sigma}^2_\varepsilon + T_i \hat{\sigma}^2_u)}$$

(4) Transform $[\mathbf{x}_{it1}, \mathbf{x}_{it2}, \mathbf{z}_{i1}, \mathbf{z}_{i2}]$ to

$$\mathbf{W}_i* = [\mathbf{x}_{it1}, \mathbf{x}_{it2}, \mathbf{z}_{i1}, \mathbf{z}_{i2}] - \hat{\theta}_i [\bar{\mathbf{x}}_{i1}, \bar{\mathbf{x}}_{i2}, \mathbf{z}_{i1}, \mathbf{z}_{i2}]$$

and y_{it} to $y_{it}* = y_{it} - \hat{\theta}_i \bar{y}_i$.



H&T’s 4 STEP IV Estimator

Instrumental Variables $\mathbf{V}_i$ =

$(\mathbf{x1}_{it} - \overline{\mathbf{x1}}_i)$ = K_1 instrumental variables

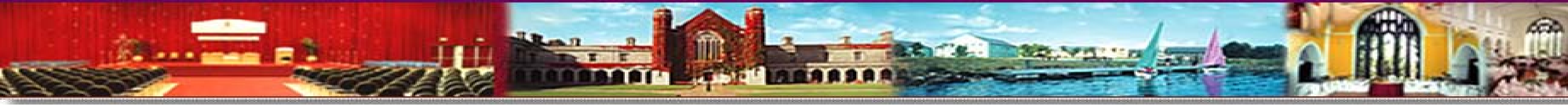
$(\mathbf{x2}_{it} - \overline{\mathbf{x2}}_i)$ = K_2 instrumental variables

$\mathbf{z1}_i$ = L_1 instrumental variables (uncorrelated with u)

$\overline{\mathbf{x1}}_i$ = K_1 additional instrumental variables.

Now do 2SLS of $\mathbf{y}*$ on $\mathbf{W}*$ with instruments $\mathbf{V}$ to estimate all parameters. I.e.,

$$[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2] = (\hat{\mathbf{W}}*' \hat{\mathbf{W}}*)^{-1} \hat{\mathbf{W}}*' \mathbf{y}*.$$



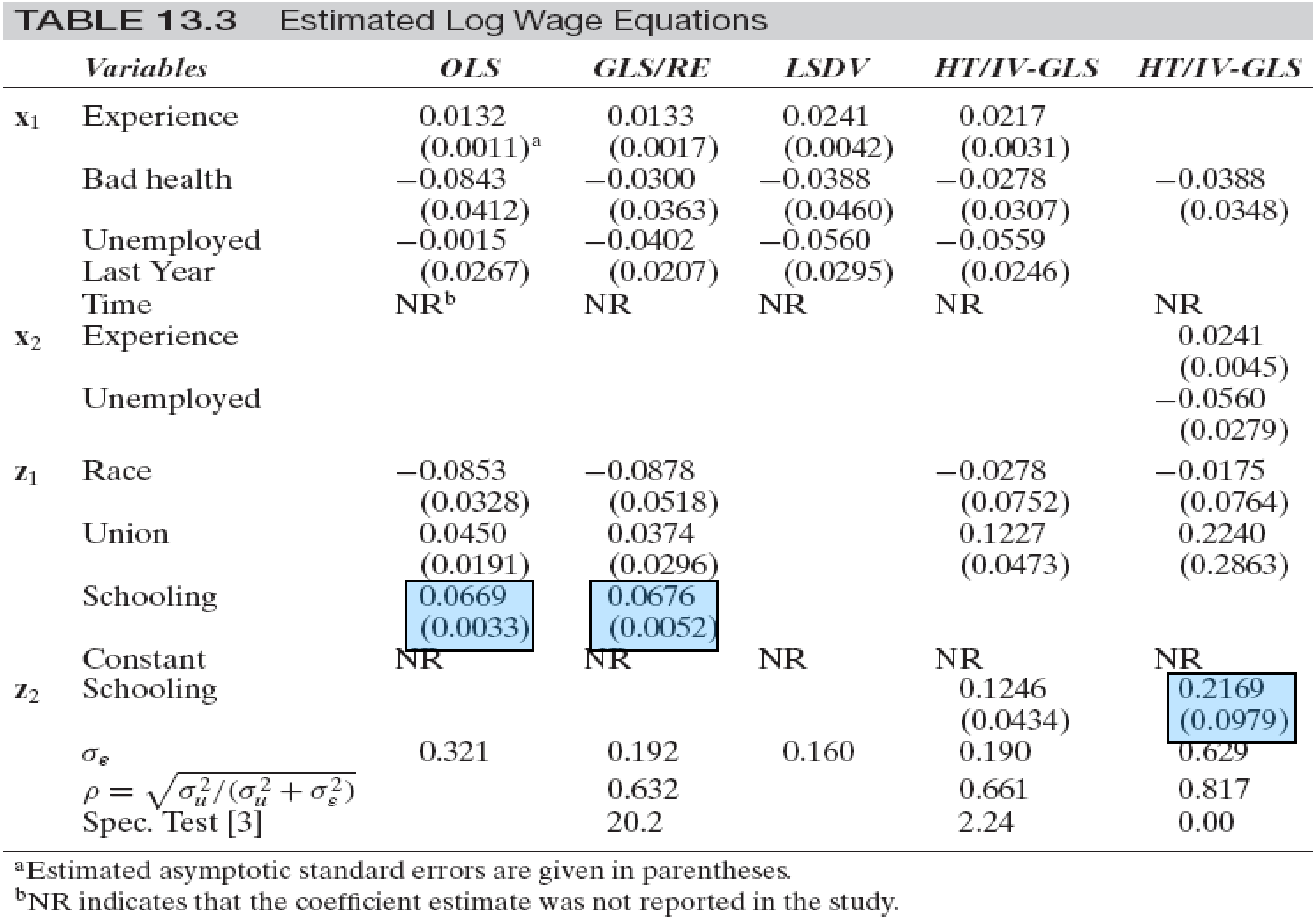
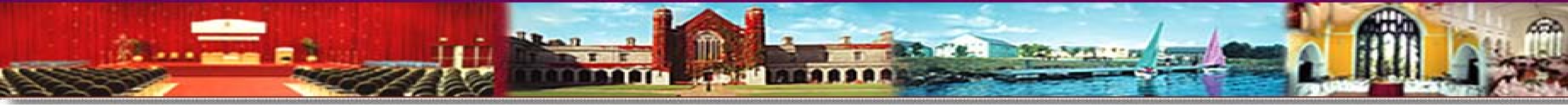


TABLE 13.3 Estimated Log Wage Equations

	Variables	*OLS*	*GLS/RE*	*LSDV*	*HT/IV-GLS*	*HT/IV-GLS*
x_1	Experience	0.0132	0.0133	0.0241	0.0217	
		(0.0011)[a]	(0.0017)	(0.0042)	(0.0031)	
	Bad health	−0.0843	−0.0300	−0.0388	−0.0278	−0.0388
		(0.0412)	(0.0363)	(0.0460)	(0.0307)	(0.0348)
	Unemployed	−0.0015	−0.0402	−0.0560	−0.0559	
	Last Year	(0.0267)	(0.0207)	(0.0295)	(0.0246)	
	Time	NR[b]	NR	NR	NR	NR
x_2	Experience					0.0241
						(0.0045)
	Unemployed					−0.0560
						(0.0279)
z_1	Race	−0.0853	−0.0878		−0.0278	−0.0175
		(0.0328)	(0.0518)		(0.0752)	(0.0764)
	Union	0.0450	0.0374		0.1227	0.2240
		(0.0191)	(0.0296)		(0.0473)	(0.2863)
	Schooling	0.0669	0.0676			
		(0.0033)	(0.0052)			
	Constant	NR	NR	NR	NR	NR
z_2	Schooling				0.1246	0.2169
					(0.0434)	(0.0979)
	σ_ε	0.321	0.192	0.160	0.190	0.629
	$\rho = \sqrt{\sigma_u^2/(\sigma_u^2 + \sigma_\varepsilon^2)}$		0.632		0.661	0.817
	Spec. Test [3]		20.2		2.24	0.00

[a] Estimated asymptotic standard errors are given in parentheses.

[b] NR indicates that the coefficient estimate was not reported in the study.



Arellano/Bond/Bover's Formulation Builds on Hausman and Taylor

$y_{it} = \mathbf{x1}_{it}'\boldsymbol{\beta}_1 + \mathbf{x2}_{it}'\boldsymbol{\beta}_2 + \mathbf{z1}_i'\boldsymbol{\alpha}_1 + \mathbf{z2}_i'\boldsymbol{\alpha}_2 + \varepsilon_{it} + u_i$

Instrumental variables for period t

$(\mathbf{x1}_{it} - \overline{\mathbf{x1}}_i) = K_1$ instrumental variables

$(\mathbf{x2}_{it} - \overline{\mathbf{x2}}_i) = K_2$ instrumental variables

$\mathbf{z1}_i = L_1$ instrumental variables (uncorrelated with u)

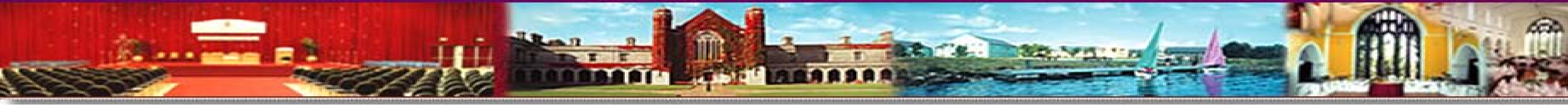
$\overline{\mathbf{x1}}_i = K_1$ additional instrumental variables. $K_1 \geq L_2$.

Let $v_{it} = \varepsilon_{it} + u_i$

Let $\mathbf{z}_{it}' = [(\mathbf{x1}_{it} - \overline{\mathbf{x1}}_i)', (\mathbf{x2}_{it} - \overline{\mathbf{x2}}_i)', \mathbf{z1}_i', \overline{\mathbf{x1}}']$

Then $E[\mathbf{z}_{it} v_{it}] = \mathbf{0}$

We formulate this for the T_i observations in group i.



Arellano/Bond/Bover's Formulation Adds a Lagged DV to H&T

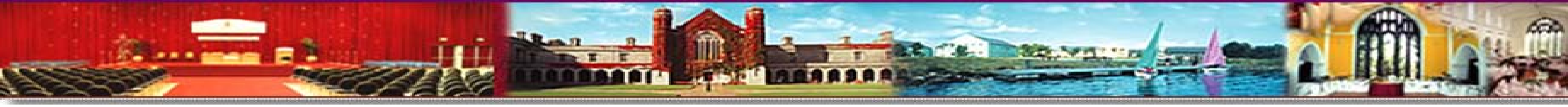
$$y_{it} = \delta y_{i,t-1} + \mathbf{x1}_{it}'\boldsymbol{\beta}_1 + \mathbf{x2}_{it}'\boldsymbol{\beta}_2 + \mathbf{z1}_i'\boldsymbol{\alpha}_1 + \mathbf{z2}_i'\boldsymbol{\alpha}_2 + \varepsilon_{it} + u_i$$

Parameters $\boldsymbol{\theta} = [\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2']'$

The data

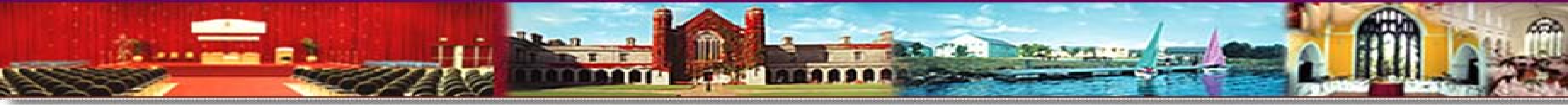
$$\mathbf{y}_i = \begin{bmatrix} y_{i,2} \\ y_{i,3} \\ \vdots \\ y_{i,T_i} \end{bmatrix}, \quad \mathbf{X}_i = \begin{array}{c} \begin{bmatrix} y_{i,1} & \mathbf{x1}_{i2}' & \mathbf{x2}_{i2}' & \mathbf{z1}_i' & \mathbf{z2}_i' \\ y_{i,2} & \mathbf{x1}_{i3}' & \mathbf{x2}_{i3}' & \mathbf{z1}_i' & \mathbf{z2}_i' \\ & & & & \\ y_{i,T-1} & \mathbf{x1}_{iT_i}' & \mathbf{x2}_{iT_i}' & \mathbf{z1}_i' & \mathbf{z2}_i' \end{bmatrix} \\ \begin{matrix} 1 & K1 & K2 & L1 & L2 & \text{columns} \end{matrix} \end{array}, \quad T_i\text{-1 rows}$$

This formulation is the same as H&T with $y_{i,t-1}$ contained in $x2_{it}$.



Dynamic (Linear) Panel Data (DPD) Models

- Application
- Bias in Conventional Estimation
- Development of Consistent Estimators
- Efficient GMM Estimators



Dynamic Linear Model

Balestra-Nerlove (1966), 36 States, 11 Years

Demand for Natural Gas

Structure

New Demand: $G^*_{i,t} = G_{i,t} - (1 - \delta)G_{i,t-1}$

Demand Function $G^*_{i,t} = \beta_1 + \beta_2 P_{i,t} + \beta_3 \Delta N_{i,t} + \beta_4 N_{i,t} + \beta_5 \Delta Y_{i,t} + \beta_6 Y_{i,t} + \varepsilon_{i,t}$

G=gas demand

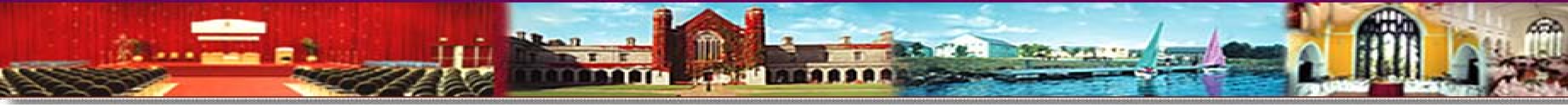
N = population

P = price

Y = per capita income

Reduced Form

$$G_{i,t} = \beta_1 + \beta_2 P_{i,t} + \beta_3 \Delta N_{i,t} + \beta_4 N_{i,t} + \beta_5 \Delta Y_{i,t} + \beta_6 Y_{i,t} + \beta_7 G_{i,t-1} + \alpha_i + \varepsilon_{i,t}$$



A General DPD model

$$y_{i,t} = \mathbf{x}\boldsymbol{\beta}_{\mathbf{i,t}} + \boxed{\delta y_{i,t-1}} + \boxed{c_i} + \varepsilon_{i,t}$$

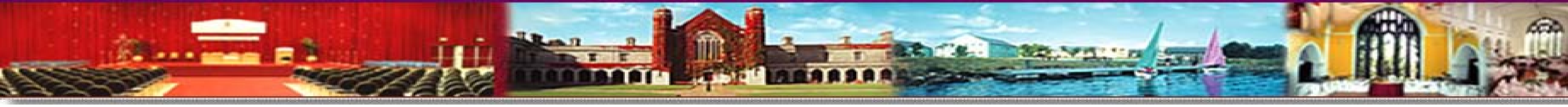
$$E[\varepsilon_{i,t} \mid \mathbf{X_i}, c_i] = 0$$

$$E[\varepsilon_{i,t}^2 \mid \mathbf{X_i}, c_i] = \sigma_\varepsilon^2,\ \ E[\varepsilon_{i,t}\varepsilon_{i,s} \mid \mathbf{X_i}, c_i] = 0 \text{ if } t \neq s.$$

$$E[c_i \mid \mathbf{X_i}] = g(\mathbf{X_i})$$

No correlation across individuals

OLS and GLS are both inconsistent.



Arellano and Bond Estimator

Base on first differences

$$y_{i,t} - y_{i,t-1} = (\mathbf{x}_{i,t}\ \boldsymbol{\beta}\ \mathbf{x}_{i,t-1})'\ \ + \delta(y_{i,t-1} - y_{i,t-2}) + (\varepsilon_{i,t} - \varepsilon_{i,t-1})$$

Instrumental variables

$$y_{i,3} - y_{i,2} = (\mathbf{x}_{i,3}\ \boldsymbol{\beta}\ \mathbf{x}_{i,2})'\ \ + \delta(y_{i,2} - y_{i,1}) + (\varepsilon_{i,3} - \varepsilon_{i,2})$$

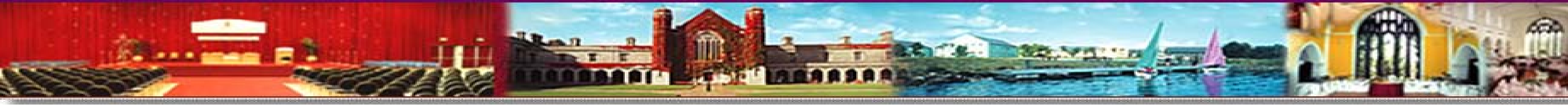
Can use y_{i1}

$$y_{i,4} - y_{i,3} = (\mathbf{x}_{i,4}\ \boldsymbol{\beta}\ \mathbf{x}_{i,3})'\ \ + \delta(y_{i,3} - y_{i,2}) + (\varepsilon_{i,4} - \varepsilon_{i,3})$$

Can use $y_{i,1}$ and y_{i2}

$$y_{i,5} - y_{i,4} = (\mathbf{x}_{i,5}\ \boldsymbol{\beta}\ \mathbf{x}_{i,4})'\ \ + \delta(y_{i,4} - y_{i,3}) + (\varepsilon_{i,5} - \varepsilon_{i,4})$$

Can use $y_{i,1}$ and y_{i2} and $y_{i,3}$



Arellano and Bond Estimator

More instrumental variables - Predetermined X

$$y_{i,3} - y_{i,2} = (\mathbf{x}_{i,3} \boldsymbol{\beta} \mathbf{x}_{i,2})' \;\; + \delta(y_{i,2} - y_{i,1}) + (\varepsilon_{i,3} - \varepsilon_{i,2})$$

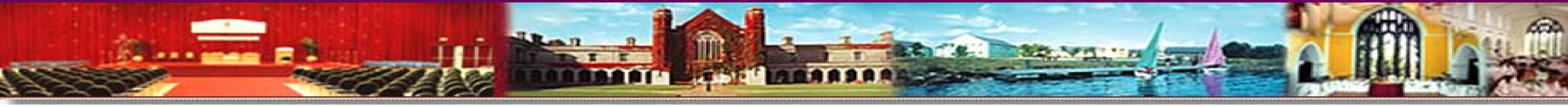
Can use y_{i1} and $\mathbf{x}_{i,1}, \mathbf{x}_{i,2}$

$$y_{i,4} - y_{i,3} = (\mathbf{x}_{i,4} \boldsymbol{\beta} \mathbf{x}_{i,3})' \;\; + \delta(y_{i,3} - y_{i,2}) + (\varepsilon_{i,4} - \varepsilon_{i,3})$$

Can use $y_{i,1}, y_{i2}, \mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \mathbf{x}_{i,3}$

$$y_{i,5} - y_{i,4} = (\mathbf{x}_{i,5} \boldsymbol{\beta} \mathbf{x}_{i,4})' \;\; + \delta(y_{i,4} - y_{i,3}) + (\varepsilon_{i,5} - \varepsilon_{i,4})$$

Can use $y_{i,1}, y_{i2}, y_{i,3}, \mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \mathbf{x}_{i,3}, \mathbf{x}_{i,4}$



Arellano and Bond Estimator

Even more instrumental variables - Strictly exogenous X

$$y_{i,3} - y_{i,2} = (\mathbf{x}_{i,3} \boldsymbol{\beta} - \mathbf{x}_{i,2})' + \delta(y_{i,2} - y_{i,1}) + (\varepsilon_{i,3} - \varepsilon_{i,2})$$

Can use y_{i1} and $\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \ldots, \mathbf{x}_{i,T}$ (all periods)

$$y_{i,4} - y_{i,3} = (\mathbf{x}_{i,4} \boldsymbol{\beta} - \mathbf{x}_{i,3})' + \delta(y_{i,3} - y_{i,2}) + (\varepsilon_{i,4} - \varepsilon_{i,3})$$

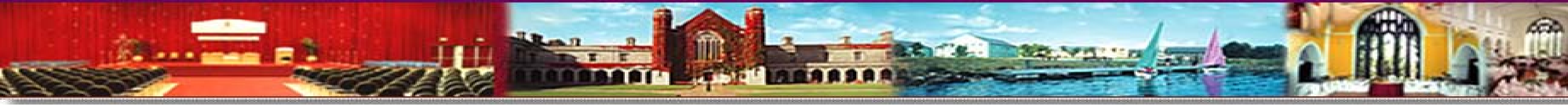
Can use $y_{i,1}, y_{i2}, \mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \ldots, \mathbf{x}_{i,T}$

$$y_{i,5} - y_{i,4} = (\mathbf{x}_{i,5} \boldsymbol{\beta} - \mathbf{x}_{i,4})' + \delta(y_{i,4} - y_{i,3}) + (\varepsilon_{i,5} - \varepsilon_{i,4})$$

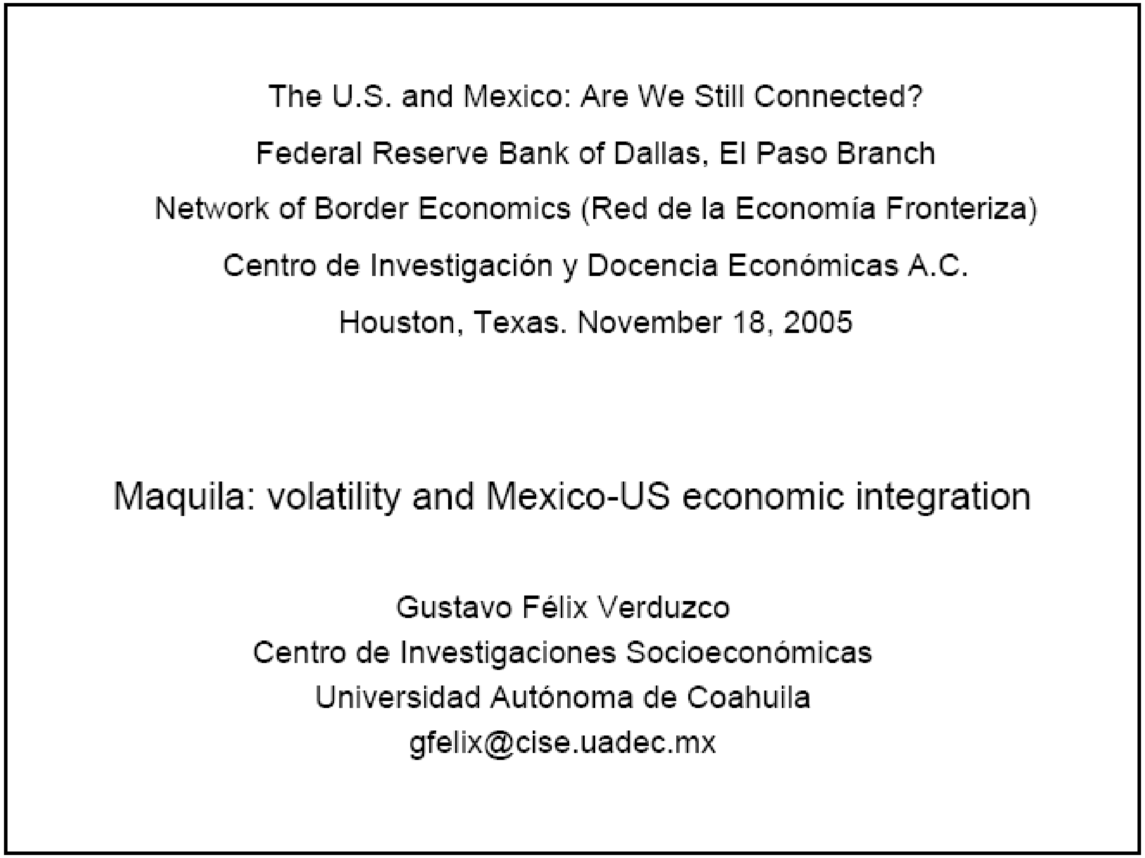
Can use $y_{i,1}, y_{i2}, y_{i,3}, \mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \ldots, \mathbf{x}_{i,T}$

The number of potential instruments is huge.

These define the rows of $\mathbf{Z_i}$. These can be used for simple instrumental variable estimation.



Application: Maquiladora



The U.S. and Mexico: Are We Still Connected?

Federal Reserve Bank of Dallas, El Paso Branch

Network of Border Economics (Red de la Economía Fronteriza)

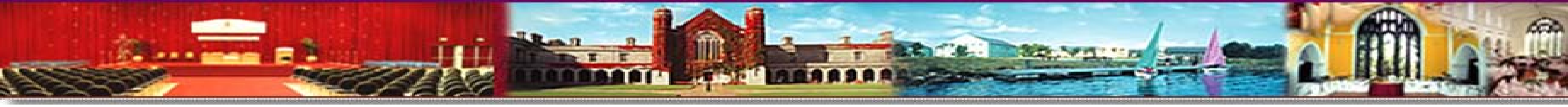
Centro de Investigación y Docencia Económicas A.C.

Houston, Texas. November 18, 2005

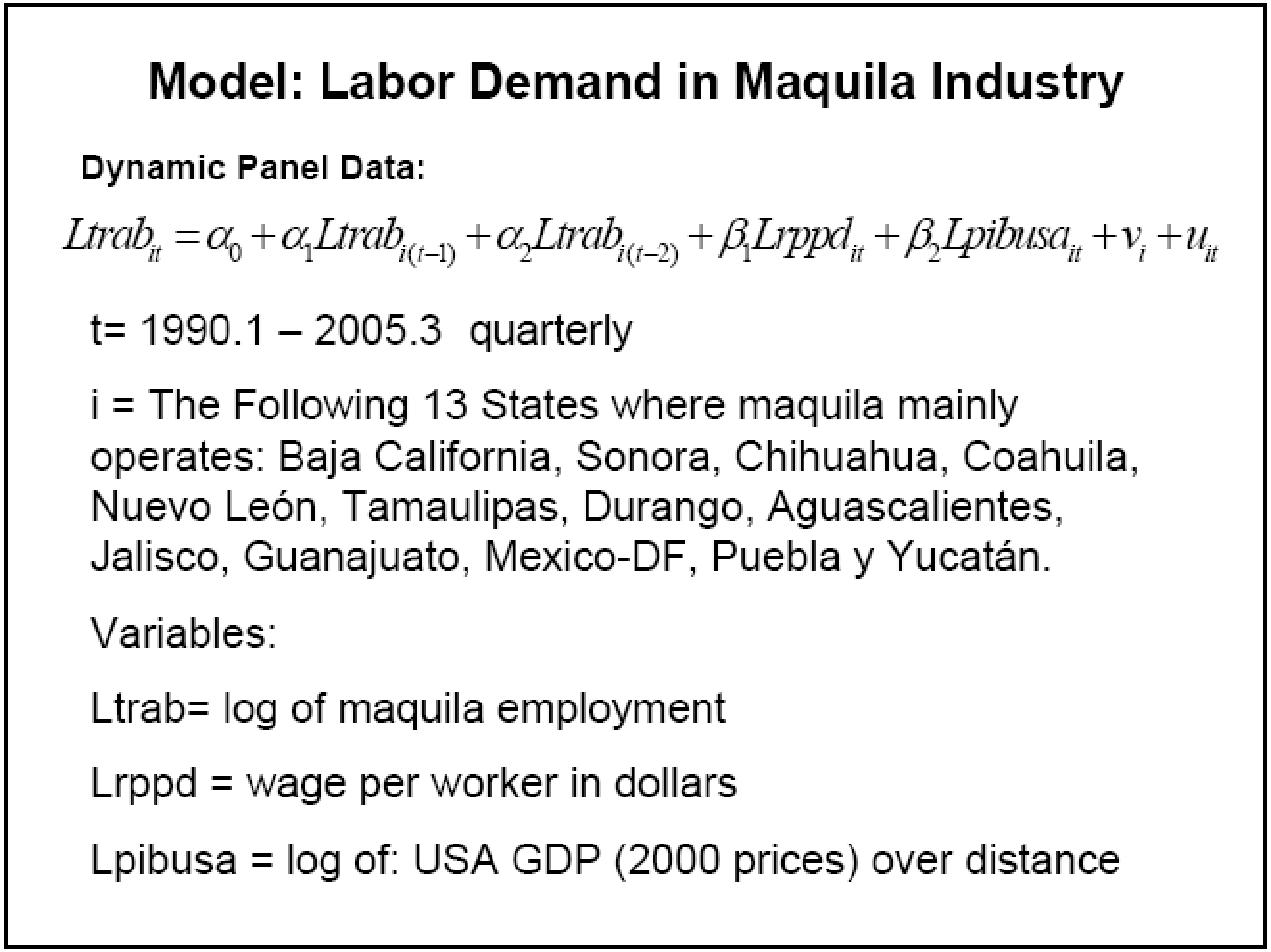
Maquila: volatility and Mexico-US economic integration

Gustavo Félix Verduzco
Centro de Investigaciones Socioeconómicas
Universidad Autónoma de Coahuila
gfelix@cise.uadec.mx

http://www.dallasfed.org/news/research/2005/05us-mexico_felix.pdf



Maquiladora



Model: Labor Demand in Maquila Industry

Dynamic Panel Data:

$$Ltrab_{it} = \alpha_0 + \alpha_1 Ltrab_{i(t-1)} + \alpha_2 Ltrab_{i(t-2)} + \beta_1 Lrppd_{it} + \beta_2 Lpibusa_{it} + v_i + u_{it}$$

t= 1990.1 – 2005.3 quarterly

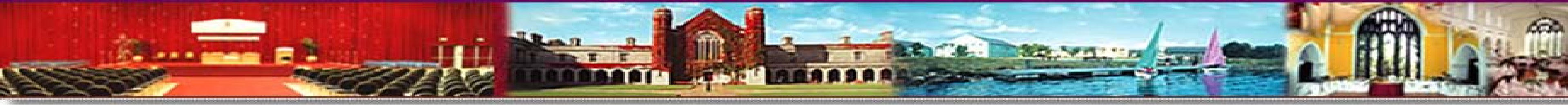
i = The Following 13 States where maquila mainly operates: Baja California, Sonora, Chihuahua, Coahuila, Nuevo León, Tamaulipas, Durango, Aguascalientes, Jalisco, Guanajuato, Mexico-DF, Puebla y Yucatán.

Variables:

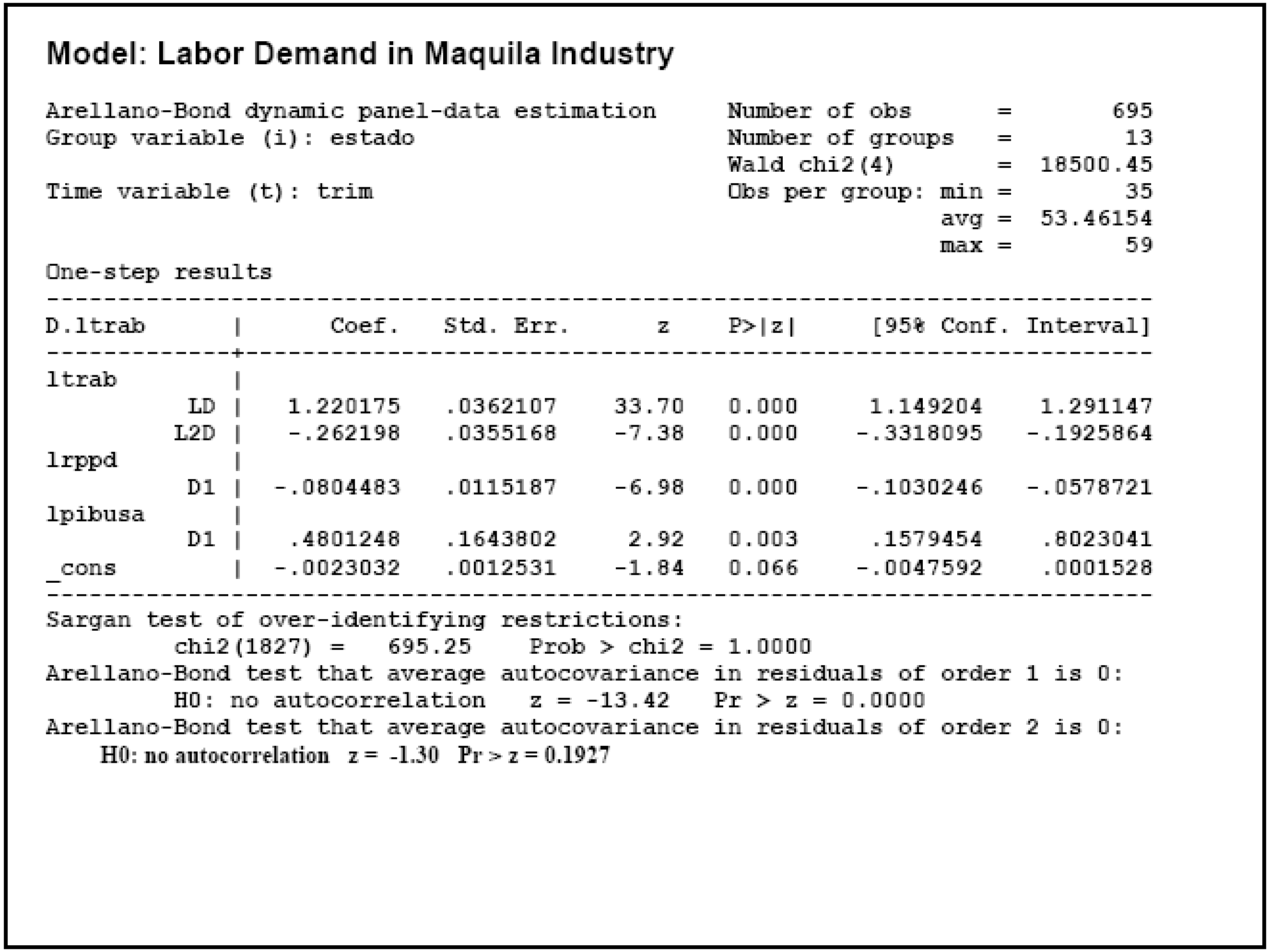
Ltrab= log of maquila employment

Lrppd = wage per worker in dollars

Lpibusa = log of: USA GDP (2000 prices) over distance



Estimates



Model: Labor Demand in Maquila Industry

```
Arellano-Bond dynamic panel-data estimation     Number of obs      =       695
Group variable (i): estado                      Number of groups   =        13
                                                Wald chi2(4)       =  18500.45
Time variable (t): trim                         Obs per group: min =        35
                                                               avg =  53.46154
                                                               max =        59
One-step results
------------------------------------------------------------------------------
D.ltrab      |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ltrab        |
          LD |   1.220175   .0362107    33.70   0.000     1.149204    1.291147
         L2D |   -.262198   .0355168    -7.38   0.000    -.3318095   -.1925864
lrppd        |
          D1 |  -.0804483   .0115187    -6.98   0.000    -.1030246   -.0578721
lpibusa      |
          D1 |   .4801248   .1643802     2.92   0.003     .1579454    .8023041
_cons        |  -.0023032   .0012531    -1.84   0.066    -.0047592    .0001528
------------------------------------------------------------------------------
Sargan test of over-identifying restrictions:
         chi2(1827) =   695.25    Prob > chi2 = 1.0000
Arellano-Bond test that average autocovariance in residuals of order 1 is 0:
         H0: no autocorrelation   z = -13.42   Pr > z = 0.0000
Arellano-Bond test that average autocovariance in residuals of order 2 is 0:
     H0: no autocorrelation  z = -1.30  Pr > z = 0.1927
```