

Table 1.1 Several Illustrative Longitudinal Data Sets

Data Title	Subject Area	File Name	Unit of Analysis	Description
Airline	Finance	Airline	Subjects are $n=19$ airlines over $T=11$ years: 1970-1980. $N=187$ observations.	Examine characteristics of airlines to determine total operating costs.
Bond Maturity	Finance	BondMat	Subjects are $n=328$ firms over $T=10$ years: 1980-1989. $N=3,280$ observations.	Examine the maturity of debt structure in terms of corporate financial characteristics.
Capital Structure	Finance	Capital	Subjects are $n=361$ Japanese firms over $T=15$ years: 1984-1998. $N=5,415$ observations.	Examine changes capital structure before and after the market crash for different types of cross holding structures.
Charitable Contributions	Accounting	Charity	Subjects are $n=47$ taxpayers over $T=10$ years; 1979-1988. $N=470$ observations.	Examine characteristics of taxpayers to determine factors that influence the amount of charitable giving.
Divorce	General social science	Divorce	Subjects are $n=51$ states over $T=3$ years: 1965, 1975 and 1985. $N=150$ observations.	Assess socioeconomic variables that affect the divorce rate.
Electric Utilities	Economics	Electric	Subjects are $n=68$ (electric) utilities over $T=12$ months. $N=816$ observations.	Examine the average cost of utilities in terms of the price of labor, fuel and capital.
Group Term Life Data	Insurance	GLife	Subjects are $n=106$ credit unions over $T=7$ years. $N=742$ observations.	Forecast group term life insurance claims of Florida credit unions.
Housing Prices	Real estate	HPrice	Subjects are $n=36$ metropolitan statistical areas (MSAs) over $T=9$ years: 1986-1994. $N=324$ observations.	Examine annual housing prices in terms of MSA demographic and economic indices.
Lottery Sales	Marketing	Lottery	Subjects are $n=50$ postal code areas over $T=40$ weeks.	Examine effects of area economic and demographic characteristics on lottery sales.
Medicare Hospital Costs	Social Insurance	Medicare	Subjects are $n=54$ states over $T=6$ years: 1990-1995. $N=324$ observations.	Forecast Medicare hospital costs by state based on utilization rates and past history.
Property and Liability Insurance	Insurance	PDemand	Subjects are $n=22$ countries over $T=7$ years: 1987-1993. $N=154$ observations.	Examine the demand for property and liability insurance in terms of national economic and risk aversion characteristics.
Student Achievement	Education	Student		Examine student math and reading achievement based on student and school demographic and socioeconomic characteristics.
Tax Preparers	Accounting	TaxPrep	Subjects are $n=243$ taxpayers over $T=5$ years: 1982, 1984-1988. $N=1,215$ observations.	Examine characteristics of taxpayers to determine the demand for a professional tax preparer.
Tort Filings	Insurance	TFiling	Subjects are $n=19$ states over $T=6$ years: 1984-1989. $N=114$ observations.	Examine demographic and legal characteristics of states that influence the number of tort filings.
Worker's Compensation	Insurance	WorkerC	Subjects are $n=121$ occupation classes over $T=7$ years. $N=847$ observations.	Forecast worker's compensation claims by occupation class.

Airline Industry

There are 187 total observations; these consist of 19 airlines over 11 years (1970 - 1980).

Variable	Description
LOGCOST	log of total costs
OUTPUT	log of available ton miles
PLANES	log of flight equipment
WAGES	log of average per employee wage rate
FUEL	log of average per gallon fuel cost
MATERIALS	log of per employee materials
STGLEN	stage length - log of average distance traveled on scheduled flights
LOADF	load factor - ratio of revenue passenger miles to available seat miles
DBTWGT	book value of the (long-term) debt/asset ratio
BNKDBT	ratio of the book value of debt owed to banks to long-term debt
VRTBL	ratio of the book value of convertible debt to long-term debt
MGRSHR	managerial share of equity ownership multiplied by the equity/asset ratio
VRTBLE	interaction term for mgrshr and vrtbl
FCFPCT	ratio of free cash flow to total assets
TIME	Time identifier, 1-11.
CODESN	Airline identifier, 1-19.

Source: Zwecher, M. (1991). *Empirical Test of Agency Theory: An Analysis of the US Airline Industry* –unpublished PH.D. dissertation, University of Wisconsin-Madison.

Bond Maturity

These data consists of observations of 328 nonregulated firms over the period 1980-1989. The goal is to assess the debt maturity structure of a firm. Source: *The Determinants of Corporate Debt Maturity Structure*, unpublished PH.D. Dissertation, by Mark Stohs, 1992, University of Wisconsin-Madison. See also “The determinants of corporate debt maturity structure,” by M. H. Stohs and D.C. Mauer, 1996, *Journal of Business* 69, no. 3, 279-312.

Variable	Description
SIC	Standard Industrial Classification (SIC) of the firm.
FIRMID	Subject (firm) identifier, 1-328
TIME	Time identifier, 1-10
SFDMAT	
DEBTMAT	
MVBV	The market value of the firm (proxied by the sum of the book value of assets and the market value of equity less the book value of equity) scaled by the book value of assets.
SIZE	The natural logarithm of the estimate of firm value measured in 1982 dollars using the PPI deflator.
CHANGEEPS	The difference between next year's earnings per share and this year's earnings per share scaled by this year's common stock price per share.
GTAXRATE	Equals TAXRATE if TAXRATE is between zero and one and otherwise equals zero.
BTAXRATE	Equals zero if TAXRATE is between zero and one and otherwise equals TAXRATE. That is, $BTAXRATE = TAXRATE - GTAXRATE$.
ASSETMAT	The book value-weighted average of the maturities of current assets and net property plant and equipment.
VAR	Ratio of the standard deviation of the first difference in earnings before interest, depreciation and taxes to the average of assets over the period 1980-1989.
ADVRD	
INTANG	
TERM	The difference between the long-term and short-term yields on government bonds.
BONDRATE	The firm's cardinalized S&P bond rating.
RATEDUM	An indicator variable that is equal one if the firm has a bond rating and is zero otherwise.
TAXRATE	Ratio of income taxes paid to pretax income.
LOWBOND	An indicator variable that equal one if the firm has an S&P bond rating of CCC or is not rated and is zero otherwise.
HIGHBOND	An indicator variable that equal one if the firm has an S&P bond rating of AA or higher and is zero otherwise.
LEVERAGE	Ratio of total debt to the market value of the firm.

Capital Structure

During the 1980s, Japan's real economy was exhibiting a healthy rate of growth. The onset of the crash in the stock and real estate markets began at the end of December, 1989, and the financial crisis soon spread to Japan's banking system. After more than ten years, the banking system is still weak and the economy struggles.

These data provide information on 361 industrial Japanese firms before and after the crash. 355 of the 361 firms in the sample are from the First Section of the Tokyo Stock Exchange; the remaining six are from the Second Section. Together, they constitute about 33 percent of the market capitalization of the Tokyo Stock Exchange.

Based on *Industrial Groupings in Japan* (1992), the sample firms are classified as keiretsu or non-keiretsu. That is, the main bank system is often part of a broader, cross share-holding structure that includes corporate groups called "keiretsu." A useful function of such corporate groupings and the main bank system is to mitigate some of the informational and incentive problems in Japan's financial markets. An important feature of the study is to identify changes in financial structure before and after the crash, and to see how these changes are affected by whether or not a firm is classified as "keiretsu."

Variable	Description
SECTOR	
NAME	
TANG	Tangibility. Net total fixed assets as a proportion of the book value of total assets. Net total fixed assets include land and buildings, plant and machinery, construction in progress, and any other fixed assets. Assets leased to others are excluded. The tangibility variable proxies for the availability of collateral to the borrowing firm.
MTB	Market-to-Book. Ratio of total assets at market value to total assets at book value. The market value of total assets is found by replacing the book value of total equity with its market value. This variable proxies for growth and investment opportunities which are not directly observable.
LS	Logarithmic Sales. The natural logarithm of the amount of sales of goods and services to third parties, relating to the normal activities of the company. This amount is net of trade discounts. Logarithmic sales proxies for the firm size.
PROF	Profitability. Earnings before interest and taxes plus depreciation, all divided by the book value of total assets. Profitability provides a measure of the availability of internally generated cash flows.
STD	Volatility. The standard deviation of weekly, unlevered stock returns during the year. This variable proxies for the business risk facing the firm. It is market base and reflects the market's perception of risk, rather than the risk itself. The main advantage of a market-based measure, versus a book based measure, is that its frequency is not limited to several observations the year.
Dependent Variables	
LVB	Total Leverage (Book). Total debt as a proportion of total debt plus book equity. Total debt is the sum of short-term and long-term debt
LVB1	Short-term Leverage. Short-term debt as a proportion of total debt plus book equity. Short-term debt is defined as having a remaining maturity less than or equal to a year. Trade payables and accruals are not included in short-term debt.
LVB2	Long-term Leverage. Long-term debt as a proportion of total debt plus book equity. Long-term debt includes straight bonds as well as convertible bonds, bonds with warrants, and any loans with remaining maturities in excess of one year.

Source: Paker, B. S. (2000). *Corporate Fund Raising and Capital Structure in Japan during the 1980s and 1990s*. Unpublished Ph.D. dissertation, University of Wisconsin-Madison.

Charitable Contributions

These data consist of individual income tax returns data from the 1979-1988 Statistics of Income (SOI) Panel of Individual Returns. The SOI Panel is a subset of the IRS Individual Tax Model File and represents a simple random sample of individual income tax returns filed each year. Based on the individual returns data, the goal is to investigate whether a taxpayer's marginal tax rate affects private charitable contributions, and secondly, if the tax revenue losses due to charitable contributions deductions is less than the gain of charitable organizations. To address these issues, we consider a price and income model of charitable contributions.

Define price as the complement of an individual's federal marginal tax rate, using taxable income prior to contributions. Income of an individual is defined as the adjusted gross income. The dependent variable is total charitable contributions, which is measured as the sum of cash and other property contributions, excluding carry overs from previous years. Other covariates included in the model are age, marital status and the number of dependents of an individual taxpayer. Age is a dichotomous variable representing whether a taxpayer is over sixty four years or not. Similarly, marital status represents if an individual is married or single.

The population consists of all U.S. taxpayers who itemize their deductions. Specifically, these are the individuals who are likely to have and to record charitable contribution deductions in a given year. Among the 1413 taxpayers in our subset of the SOI Panel, approximately 22% itemized their deductions each year during the period 1979-1988. A random sample of 47 individuals was selected from the latter group. *Source:* Banerjee and Frees (1995).

Variable	Description
SUBJECT	Subject identifier, 1-47.
TIME	Time identifier, 1-10.
CHARITY	The sum of cash and other property contributions, excluding carry overs from previous years.
INCOME	Adjusted gross income.
PRICE	One minus the marginal tax rate. Here, the marginal tax rate is defined on income prior to contributions.
AGE	An indicator variable that equals one if a taxpayer is over sixty four years and equals zero otherwise.
MS	An indicator variable that equal one if a taxpayer is married and equals zero otherwise.
DEPS	Number of dependents claimed on the taxpayer's form.

Divorce

These are data describing the divorce rate in each state. In addition, there is other socioeconomic information about a state that may be related to the divorce rate. In particular, data concerning the number of marriages and births, unemployment and crime rates, and AFDC (Aid to Families with Dependent Children) payments are available. In this file, data are available for the years 1965, 1975, 1985 and 1995. The information provided by this study is potentially useful for governing agencies in budgeting for social needs such as judicial and welfare services that are affected by divorce.

The data for the study were collected from various U.S. Statistical Abstracts. Divorce rate is defined as the number of divorces and annulments per thousand population per state. The independent variables include the number of marriages and live births per thousand population, the total unemployment rate as percent of total work force, the average monthly AFDC payments per family, and the total number of criminal offenses known to the police (murder, rape, robbery, aggravated assault, burglary, larceny, and motor vehicle theft). Some of the data points contain missing observations due to unavailability, and Nevada is unusual due to its uniquely high and unrepresentative marriage and divorce rates. Source: *U.S. Statistical Abstract*, various issues.

Name	Description
DIVORCE	Number of divorces and annulments per state per one thousand population.
BIRTH	Number of live births per state per one thousand population.
MARRIAGE	Number of marriages per state per one thousand population.
UNEMPLOY	Total unemployment rate as a percentage of the total work force.
CRIME	Total number of criminal offenses (murder, rape, robbery, aggravated assault, burglary, larceny and motor vehicle theft) known to police per one hundred thousand population.
AFDC	Average monthly AFDC (Aid to Families with Dependent Children) payments per family.
STATE	State identifier, 1-51.
TIME	Time identifier, 1-4.

Electric Utilities

Economists are interested in understanding relationships between an industry's average (over firms) cost and quantity produced. If the relation between average cost and quantity is declining, then mechanisms for increasing quantities produced, such as takeovers, should be desirable. In particular, the electric utilities enjoy a form of quasi-monopolistic markets, so there is interest in understanding cost-quantity relationships.

Classical economic theory suggests that a firm's total cost is related to quantity produced, the price of capital, the price of labor (wages) and the price of fuel, all in logarithmic units. This is known as a "translog cost model." The data here are 68 privately-owned electric utility firms considered over 12 months in 1994.

These data are based on a student project conducted by Sylvia Kuo. Source: Energy Information Association, Department of Energy (Web site: <http://www.eia.doe.gov/fuelectric.html>).

Variable	Description	Source
UTILNAME	Name of the utility.	
ID	Firm identifier, 1-68.	
MONTH	Month identifier, 1-12.	
PCAPITAL	Price of capital	From <i>Financial Statistics of Major US Investor-Owned Electric Utilities 1995</i> .(Table 41) and <i>Moody's Public Utility Manual (1994)</i> .
PFUEL	Price of fuel, a weighted average of the prices for oil, gas and coal per month. This variable is in dollars per BTU.	<i>1994 Monthly Cost and Quality of Fuels for Electric Plants Data</i> .
WAGE	Wage, in dollars per hour. The total salaries and wages divided by the number of full-time employees and ½ times the part-time employees. This variable is time-invariant.	From <i>Financial Statistics of Major US Investor-Owned Electric Utilities 1995</i> .(Table 41).
TC	Total Cost (response variable). The sum of the cost of capital, labor and fuel.	
Q	Quantity, in millions of kilowatt hours.	The variable "TOTSALES" under the publication called <i>1994 (utility level) Monthly Electric Utility Sales/Revenue Data</i> .
FXK		
FXW		
KXW		
QXF		
QXK		
QXW		

Group Term Life Data

We analyze claims data provided by a Wisconsin-based credit insurer. The data contains claims and exposure information for 106 Florida credit unions. These are “life savings” claims from a contract between the credit union and their members that provides a death benefit based on the member’s savings deposited in the credit union. *Source:* Frees, E. W., Young, V. and Y. Luo (2001). Case studies using panel data models. *North American Actuarial Journal*, 4, No. 4, 24-42.

Variable	Description
CONT ID	Occupation Class identifier, 1-106
PERIOD	Year identifier, 1-7
LP RPED	Loan protection rating period ending date.
LP UECC	Loan protection upcoming expected claim cost (adjusted for contract changes)
LP CACC	Loan protection credibility adjusted claim cost
LS RPED	Life saving rating period ending date.
LS UECC	Life saving upcoming expected claim cost (adjusted for contract changes)
LS CACC	Life saving credibility adjusted claim cost

Housing Prices

In this problem, we will examine models of housing prices in US metropolitan areas. Many studies have addressed the housing market, see, for example, Green and Malpezzi (1996, Wisconsin Working Paper) for an introduction. The prices of houses are influenced by demand-side factors such as income and demographic variables. Supply-side factors, such as the regulatory environment of a metropolitan area, may also be important.

The data consists of annual observations from 36 metropolitan statistical areas (MSAs) over the nine-year period 1986-1994. The response variable is NARSP, an MSA's average sale price based on transactions reported through the Multiple Listing Service, National Association of Realtors. As part of a preliminary analysis, the response variable has been transformed using a natural logarithm. For this problem, the demand-side variables are time varying yet the supply-side variables do not vary with time.

Response variable	
NARSP	an MSA's average sale price, in logarithmic units. It is based on transactions reported through the Multiple Listing Service.
Demand side explanatory variables	
YPC	Annual Per Capita income, from the Bureau of Economic Analysis
POP	Population, from the Bureau of Economic Analysis
PERYPC	Annual percentage growth of per capita income
PERPOP	Annual percentage growth of population
Supply side explanatory variables	
REGTEST	Regulatory index from Wharton team.
RCDUM	Rent control dummy variable
SREG1	Sum of American Institute of Planners state regulatory questions regarding use of environmental planning and management.
AJPARK	Indicates whether the MSA is adjacent to a coastline
AJWTR	Indicates whether the MSA is adjacent to one or more large parks, military bases or reservations.
Additional Variables	
MSA	Subject (MSA) identifier, 1- 36.
TIME	Time identifier, 1-9.

Lottery Sales

State of Wisconsin lottery administrators provided weekly lottery sales data. We consider online lottery tickets that are sold by selected retail establishments in Wisconsin. These tickets are generally priced at \$1.00, so the number of tickets sold equals the lottery revenue. We analyze lottery sales (OLSALES) over a forty-week period, April, 1998 through January, 1999, from fifty randomly selected ZIP codes within the state of Wisconsin. We also consider the number of retailers within a ZIP code for each time (NRETAIL).

A budding literature suggest variables that influence lottery sales. Table 4.1 lists economic and demographic characteristics that we consider in this analysis. Much of the empirical literature on lotteries is based on annual data that examines the state as the unit of analysis. In contrast, we examine much finer economic units, the ZIP code level, and examine weekly lottery sales. The economic and demographic characteristics were abstracted from the United States census. These variables summarize characteristics of individuals within ZIP codes at a single point in time and thus are not time varying. *Source:* Frees, E.W. and Miller, T. W. (2003). Sales forecasting using longitudinal data models. To appear in the *International Journal of Forecasting*.

Table 4.1. Lottery, Economic and Demographic Characteristics of 50 Wisconsin ZIP Codes

Lottery Characteristics	
OLSALES	Online lottery sales to individual consumers
NRETAIL	Number of listed retailers
Economic and Demographic Characteristics	
PERPERHH	Persons per household
MEDSCHYR	Median years of schooling
MEDHVL	Median home value in \$1000s for owner-occupied homes
PRCRENT	Percent of housing that is renter occupied
PRC55P	Percent of population that is 55 or older
HHMEDAGE	Household median age
MEDINC	Estimated median household income, in \$1000s
POPULATN	Population, in thousands

Medicare Hospital Costs

We consider $T=6$ years, 1990-1995, of data for inpatient hospital charges that are covered by the Medicare program. The data were obtained from the Health Care Financing Administration, Bureau of Data Management and Strategy. To illustrate, in 1995 the total covered charges were \$157.8 billions for twelve million discharges. For this analysis, we use state as the subject, or risk class. Thus, we consider $n=54$ states that include the 50 states in the Union, the District of Columbia, Virgin Islands, Puerto Rico and an unspecified “other” category. *Source*: Frees, E. W., Young, V. and Y. Luo (2001). Case studies using panel data models. *North American Actuarial Journal*, 4, No. 4, 24-42.

Variable	Description
STATE	State identifier, 1- 54
YEAR	Year identifier, 1-6
TOT_CHG	Total hospital charges, in millions of dollars.
COV_CHG	Total hospital charges covered by Medicare, in millions of dollars.
MED_REIM	Total hospital charges reimbursed by the Medicare program, in millions of dollars.
TOT_D	Total number of hospitals stays, in days.
NUM_DSHG	Number discharged, in thousands.
AVE_T_D	Average hospital stay per discharge in days.

Property and Liability Insurance

During the 1980s and early 1990s, the world insurance market grew substantially. In 1993 world insurance premiums were approximately \$1.8 trillion, accounting for about eight percent of world Gross Domestic Product compared to four percent in 1984. What factors affect the demand for insurance?

This study examines property-liability insurance consumption across OECD member countries. Recently compiled OECD data, which reports insurance premiums on a “by line basis,” were gathered. The study focuses on two lines of insurance – motor vehicle and general liability. Further, six additional lines of insurance are available in the data set. The data span twenty-two countries and seven years, 1987-1993. Source: *International Property-Liability Insurance Consumption*, unpublished Ph.D. Dissertation, Jaewook Chung, University of Wisconsin, 1996. See also “International property-liability insurance consumption,” by M. Browne, J. Chung, and EW. Frees, submitted to the *Journal of Risk and Insurance*.

Variable	Description
COUNTRY	Subject (country) identifier, 1-24.
TIME	Time identifier, 1-7.
GNP_CAP	Gross national product, in US dollars per capita.
NEWMEAS	A new measure of wealth produced by the World Bank. It is a composite measure that includes human resources, produced or manufactured assets and natural resources. This variable is time-invariant. It is wealth per capita, in thousands of US dollars.
RISKAVER	Risk aversion, which is proxied by level of education. This is measured by the enrollment ratio of third-level education, that is, the ratio of total enrollment in third-level education institutions to the total population age 20 to 24. Education at the third level is provided by different types of institutions, including universities, teacher-training institutions and technical institutes.
PROTECT	Protective measures may reduce competition and thus raise prices. Trade barriers are proxied by the insurance market share of foreign firms. Specifically, this is the market share of branches or agencies of foreign undertakings in total domestic non-life insurance.
POPDEN	Population density, the average number of people living within a square kilometer.
URBAN	Urbanization. The percentage of people living in urban areas.
LEGALSYS	Legal system. This is an indicator variable that is equal to one if the country has a common law system and is zero otherwise (statutory law system). This variable is time-invariant.
CPI	Consumer Price Index, as a percentage.
MRATE	Motor vehicle ownership per capita.
<i>Dependent Variables</i>	
AUTO	Automobile premium density, computed as total direct gross automobile insurance premiums divided by the country’s population. It includes damage or loss to land vehicles as well as liability arising out of the use of motor vehicles. The measure is in US dollars per capita.
TRANSPRT	Transport premium density. Transport insurance includes railway loss, aircraft loss and liability and ship loss and liability.
FREIGHT	Freight premium density. It includes all damage to or loss of goods in transit or baggage.
FIREPROP	Fire and other property damage premium density. It includes damage or loss of property due to fire, explosion, storm, other natural forces, nuclear energy and land subsidence as well as other damage to property.
PECLOSS	Pecuniary loss premium density. It includes credit loss, surety loss and other miscellaneous financial losses
GENLIAB	General liability premium density. It includes all liability other than motor vehicle, aircraft and ship liability.
ACCSICK	Accident and sickness premium density.
OTHERNL	Other non-life premium density. It includes legal expenses, assistance and other miscellaneous insurance.

Data	Source	Publishers
------	--------	------------

Premium Density dependent variables	Insurance Statistics Yearbook (1986-1993)	OECD
Income (GNP/Capita)	World Tables 1995	IBRD/The World Bank
Consumer Price Index	International Financial Statistics	IMF
Protective Measures	Insurance Statistics Yearbook (1986-1993)	OECD
Third Level Education	UNESCO Statistical Yearbook	UNESCO
Population Density	IMD&S 1996	BPC Wheatons Ltd.
Urbanization	World Tables 1995	IBRD/The World Bank
Wealth	Monitoring Environmental Progress	World Bank
Legal System	Statistical Abstract of the World	Gale Research Inc.
Motorization Rate		

Student Achievements

Source: 1988 *National Educational Longitudinal Study*, National Center for Education Statistics. This was a GB 806 project by Jennifer Imazeki and Kaijie Pan.

This survey follows about 20,000 eighth grade students beginning in 1988 through two follow-ups, 1990 (tenth grade) and 1992 (twelfth grade). Further it continues them into adulthood. In addition to various survey questions and demographic information on students and their schools, students took tests designed by the Education Testing Service, that cover math reading, science and social studies.

Variable	Description
ID	Student identifier.
SCH_ID	School identifier, School public release id.
SEX	An indicator variable that equals one if a student is female and equals zero if male.
RACE	A categorical variable that describes the race of the student, 0-3.
F1SCH_ID	School identifier, F1 school id.
F2PNLWT	
F2CXTWT	
F2PNLFLG	
YEAR	Year identifier, 1-3.
MATH	Estimated number right on math test, using item-response theory to standardize scores.
READ	Estimated number right on reading test, using item-response theory to standardize scores.
SES	Socioeconomic status. Index of parents education and occupations, family income.
SKIP	Number of times student cut class last term: 1=1-6 times; 2=7-9 times; 3=10 or more times.
LATE	
SCHTYPE	An indicator variable that equals one if the school is public and equals zero otherwise (private).
URBAN	An indicator variable that equals one if the school is urban and equals zero otherwise (rural).
REGION	A categorical variable that describes the region of the country: 0=West, 1=Northeast, 2=North Central/Midwest, 3=South.
MINOR	The percent of students in the school who are white: 0=25% or less, 1=26-50%, 2=51-75%, 3=76-90%, 4=greater than 90%.
LUNCH	The percent of students in the school who receive free or reduced lunches: 1=10% or less, 2=11-50%, 3=greater than 50%.
LEP	The percent of students in the school who are Limited English Proficient: 1=10% or less, 2=11-20%, 3=21-30%, 4=31-40%, 5=greater than 40%.
SALARY	The base teacher salary, measured in thousands of dollars.
GRADDEG	
BLACK	An indicator variable that equals one if the student is black and equals zero otherwise.
ASIAN	An indicator variable that equals one if the student is Asian/Pacific Islander and equals zero otherwise.
HISP	An indicator variable that equals one if the student is Hispanic and equals zero otherwise.
AMER	An indicator variable that equals one if the student is American Indian and equals zero otherwise.

Tax Preparer

Data for this study are from the Statistics of Income (SOI) Panel of Individual Returns, a part of the Ernst and Young/University of Michigan Tax Research Database. The SOI Panel represents a simple random sample of unaudited individual income tax returns filed for tax years 1979-1990. The data are compiled from a stratified probability sample of unaudited individual income tax returns, Forms 1040, 1040A and 1040EZ, filed by U.S. taxpayers. The estimates that are obtained from these data are intended to represent all returns filed for the income tax years under review. All returns processed are subjected to sampling except tentative and amended returns.

We consider a balanced panel from 1982-1984 and 1986-1987 taxpayers included in the SOI panel; a four percent sample of this comprises our sample of 258 taxpayers. These years are chosen because they contain the interesting information on paid preparer usage. Specifically, these data include line item tax return data plus an indicator variable noting the presence of a paid tax preparer for years 1982-1984 and 1986-1987. *Source:* Frischmann and Frees (1999, *Journal of the American Taxation Association*).

Independent Variables - Demographic Characteristics	
MS	is an indicator variable of the taxpayer's marital status. It is coded one if the taxpayer is married and zero otherwise.
HH	is an indicator variable, one if the taxpayer is a head of household and zero otherwise.
DEPEND	is the number of dependents claimed by the taxpayer.
AGE	is the presence of an indicator for age 65 or over.
Independent Variables - Economic Characteristics	
F1040A	is an indicator variable of the taxpayer's filing type. It is coded one if the taxpayer uses Form 1040A and zero otherwise.
F1040EZ	is an indicator variable of the taxpayer's filing type. It is coded one if the taxpayer uses Form 1040EZ and zero otherwise.
TPI	is the sum of all positive income line items on the return.
TXRT	is a marginal tax rate. It is computed on TPI less exemptions and the standard deduction.
MR	is an exogenous marginal tax rate. It is computed on TPI less exemptions and the standard deduction.
EMP	is an indicator variable, one if Schedule C or F is present and zero otherwise. Self-employed taxpayers have greater need for professional assistance to reduce the reporting risks of doing business.
PREP	is a variable indicating the presence of a paid preparer.
Additional Variables	
TAX	is the tax liability on the return.
SUBJECT	Subject identifier, 1- 258.
TIME	Time identifier, 1-5.
LNTAX	is the natural logarithm of the tax liability on the return.
LNTPI	is the natural logarithm of the sum of all positive income line items on the return.

Tort Filings

There is a widespread belief that, in the United States, parties have become increasingly willing to go to the judicial system to settle disputes. This is particularly true in the insurance industry, an industry designed to spread risk among individuals who are subject to unfortunate events that threaten their livelihoods. Litigation in the insurance industry arises from two types of disagreement among parties, breach of faith and tort. A breach of faith is a failure by a party to the contract to perform according to its terms. This type of dispute is relatively confined to issues of facts including the nature of the duties and the action of each party. A tort action is a civil wrong, other than breach of contract, for which the court will provide a remedy in the form of action for damages. A civil wrong may include malice, wantonness oppression or capricious behavior by a party. Generally, much larger damages can be collected for tort actions because the award may be large enough to "sting" the guilty party. Since large insurance companies are viewed as having "deep pockets," these awards can be quite large indeed.

The data that we consider below is the number of FILINGS of tort actions against insurance companies (y). Here, for each of six years (TIME), the data was obtained from 19 STATES. Thus, there are $6 \times 19 = 114$ observations available. *Source: An Empirical Study of the Effects of Tort Reforms on the Rate of Tort Filings*, unpublished Ph.D. Dissertation, Han-Duck Lee, University of Wisconsin. See also "How does joint and several tort reform affect the rate of tort filings? Evidence from the state courts" by H.D. Lee, M. J. Browne and J.T. Schmit, 1994, *Journal of Risk and Insurance* 61, 295-316.

Name	Description
FILINGS	Number of filings of tort actions against insurance companies.
POPLAWYR	The population per lawyer.
VEHCMILE	Number of automobiles miles per mile of road, in thousands.
GSTATEP	Percentage of gross state product from manufacturing and construction.
POPDENSY	Number of people per ten square miles of land.
WCMPMAX	Maximum workers' compensation weekly benefit.
URBAN	Percentage of population living in urban areas.
UNEMPLOY	State unemployment rate, in percentages.
J&SLIAB	An indicator of joint and several liability reform.
COLLRULE	An indicator of collateral source reform.
CAPS	An indicator of caps on non-economic reform.
PUNITIVE	An indicator of limits of punitive damage
TIME	Year identifier, 1-6
STATE	State identifier, 1-19.

Worker's Compensation

We consider a standard example in worker's compensation insurance, examining losses due to permanent, partial disability claims. The data are from Klugman (1992), who considers Bayesian model representations, and are originally from the National Council on Compensation Insurance. We consider $n=121$ occupation, or risk, classes, over $T=7$ years. To protect the data source, further information on the occupation classes and years is not available. *Source: Frees, E. W., Young, V. and Y. Luo (2001). Case studies using panel data models. North American Actuarial Journal, 4, No. 4, 24-42.*

Variable	Description
CL	Occupation class identifier, 1-124
YR	Year identifier, 1-4
PR	Payroll, a measure of exposure to loss, in tens of millions of dollars
LOSS	Losses related to permanent partial disability, in tens of millions of dollars.