IS EDUCATIONAL ATTAINMENT A SIGNIFICANT DETERMINANT OF WHERE FIRMS DECIDE TO LOCATE OR EXPAND OPERATIONS?

by

Sharon D. Hoke

Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

in

Economics

APPROVED: ______________________________

Roger N. Waud, Chairman

______________________ ______________________

Thomas Lutton Nancy A. Wentzler

____

Brian K. Reid

September 1998

Falls Church, Virginia

Keywords: Industrial location, urban economics, and education.

IS EDUCATIONAL ATTAINMENT A SIGNIFICANT DETERMINANT OF WHERE FIRMS DECIDE TO LOCATE OR EXPAND OPERATIONS?

by

Sharon D. Hoke

Roger N. Waud, Chairman

Economics

(Abstract)

Firms seeking sites for new or expansion plants rely on their ability to assess the benefits and costs generated by locating operations in a given state. State governments strive to understand the issues important to firms who are seeking a site for new operations or branch plants. They do so because attracting branch plants and new firms is critical to their economic growth. In addition to factors traditionally considered important to industrial location decisions (energy prices, wage levels, unionization, taxes, and public services), this study also considers the impact of the average level of education attained by the population of a state. Specifically, this study hypothesizes that the average level of education attained by a state's labor force significantly affects the location decisions of firms oriented toward local inputs.

Results indicate that educational attainment is a significant determinant of where firms in the transportation industry decide to locate or expand operations. However, educational attainment is not a significant determinant for the electronic equipment industry. Wage and unemployment levels are significant factors for both industries. With the exception of educational attainment and tax levels, the results do not vary across industries. Since industries such as transportation equipment seem to place a premium on states with a relatively

high level of education attainment, states wishing to attract this type of companies may find it beneficial to encourage higher education, both for the well-being of its residents and for the economy of the state.

Acknowledgments

I wish to thank research and thesis committee professors Roger Waud, Nancy Wentzler, Thomas Lutton, and Brian Reid for their guidance during the research process. Their insightful comments and suggestions enhanced the quality of research considerably.

Is Educational Attainment a Significant Determinant of Where Firms Decide to Locate Or Expand Operations?

Is Educational Attainment a Significant Determinant of Where Firms Decide to Locate Or Expand Operations?

1. <u>Introduction</u>

Firms seeking sites for new or expansion plants rely on their ability to assess the benefits and costs generated by locating operations in a given state. State governments strive to understand the issues important to firms who are seeking a site for new operations or branch plants. They do so because attracting branch plants and new firms is critical to their economic growth.[1] This economic growth translates into job growth. Moreover, newly-created jobs often pay higher wages than the jobs currently available. Firms study the factors associated with industrial location because they must understand how to maximize profit in order to remain competitive. In addition to factors traditionally considered important to industrial location[2] decisions (energy prices, wage levels, unionization, taxes, and public services), this study also considers the impact of the average level of education attained by the population of a state. Specifically, this study hypothesizes that the location decisions of firms oriented toward local input are significantly impacted by the average level of education attained by a state's labor force.

The remainder of this section provides a framework for the empirical discussion. Section 1.1 contains a discussion of industrial location theory; Section 1.2 provides a specific explanation of local input orientation; and Section 1.3 explains the process undertaken by firms deciding where to locate operations. Section 2 provides a review of relevant literature. Section 3 describes the

---

[1] Branch plants are defined as one plant location of a multi-plant firm. Firms are defined as a single establishment, i.e., no branch operations exist.
[2] Industrial location theory is also referred to as development economic or urban economic theory.

methodology supporting this paper, and Section 4 details the corresponding empirical results.

1.1 Industrial Location Theory

Theory resulting from years of research provides a classification of firms by their orientation. Firms are generally oriented towards transport costs, costs of local inputs, or the amenities of the potential worksite (O'Sullivan.) Table 1 provides a summary of orientation and lists key characteristics. By recognizing their orientation, firms can focus their efforts on finding plant sites that complement their operations. Since the orientation of a firm provides insight as to which resources are most critical to a firm's location decision, the selection of variables to include in a model can be improved. For instance, instead of simply including data for all factors theoretically believed to affect location decision, researchers can use the insight gained from knowledge of firm orientation to select the most important variables for a given industry. Since the emphasis of this paper is the impact of educational attainment on industrial location, firms oriented toward local inputs will be the focus group. (Educational attainment is a defining characteristic of labor inputs.)

Table 1. Summary of Orientation Characteristics

Orientation	Relevant Characteristic	Example
Transfer Orientation	Transport costs are large fraction of total costs.	
Resource Orientation	Weight loss Bulk loss Perishability loss Hazard (fragility) loss	Bat Factory Cotton baling Canning Skunk deodorizing
Market Orientation	Weight loss Bulk loss Perishability loss Hazard (fragility) loss	Bottling Auto assembly Baking Explosives
Local-input Orientation	Transport costs are small fraction of total costs.	
Energy	Energy-intensive production	Aluminum
Labor	Labor-intensive production	Textiles
Intermediate inputs		
Specialized inputs	Localization economies	Dressmakers
Business services	Urbanization economies	Corporate HQ
Amenity Orientation	Workers are sensitive to weather and recreation.	R&D

Source: O'Sullivan, Arthur. *Urban Economics*. Boston: Irwin Publishing, 1993. (Page 60)

1.2 Firms Oriented Toward Local Inputs

When evaluating plant sites, firms oriented toward local inputs are most concerned with energy costs, labor costs, availability of intermediate goods, and the level of public services provided. Energy costs are important to firms whose production process requires a considerable amount of energy. Accordingly, these firms will locate in areas where energy costs are relatively low. When evaluating labor costs, firms must consider other factors in addition to wage level--the extent of unionization, and the educational attainment of the workforce. Each of these characteristics affect the total cost of labor. Education attainment is also an indication of the quality of the available labor force. The availability of intermediate goods is important because a firm must have easy-access to raw materials. If an area has a concentration of firms in the same industry, additional benefits may exist. For example, by sharing the suppliers of intermediate goods, firms in a cluster can benefit from scale economies in the production of their goods. Public services are also an important determinant of the location decision. The level of public services impacts transportation of finished products and the ability to attract new employees. A developed infrastructure reduces the cost and difficulty of transporting goods, and amenities such as quality schools are attractive to employees and entices them to move to an area.

1.3 The Process Involved When Making a Location Decision

The progression through the location decision-making process does not different significantly from firm to firm (Schmenner 1982.) However, the degree of centralization and the departments assigned responsibilities leading up to and at the final decision point varies. As a result, the process a corporation undergoes to select a site for new operations should be considered before developing a model for location decisions. For example, the centralized process will involve fewer "hands-on" line management than a decentralized process; therefore, the information used to make the decision may not be as specific to operations. For example, instead of the polling of existing industry that is frequently part of a decentralized process, the centralized process involves a more general evaluation of the location, such as favorable land and energy prices. Identification of a current trend toward decentralization or centralization could provide insight into the stages of and importance of variables in the decision-making process. Table 2 outlines the decision-making process by type of organization.

Table 2. Decision-Making Process By Type of Organization

Type of Organization	Explanation
Centralized Organization Schemes	Location search initiated by corporate level managers and studied by corporate staff. Any affected division is brought in to approve.
Large Corporate Group Study	Location search initiated by a senior management committee and studied by a staff group drawn from a variety of functions (e.g., real estate, planning industrial relations, logistics, plant engineering etc. Study usually chaired by a real estate planning function. Study group may be a semi-permanent body. Studies can be very detailed and quite sophisticated in analysis. The division for which the new location is sought usually has final selection rights, or at least veto power over choice and is responsible for supplying data on size, product, process technology and other needs for the study. Final site selection within a town may be left to real estate function.
Decentralized Organization Scheme	Location search entirely the responsibility of the division.
Division-only study	Manager-designate may be named early on to direct the team and lead it through the plant start-up phase. Operates much like the division-controlled study except there is no available cooperate staff to invite into the study. Any expertise must either be "home-grown" within the division or must be brought in from the outside. Thus, this organizational scheme is the most likely to purchase the services of specialized location consultants. An informal network of discussion and approvals predominate. Divisions must sell their ideas to upper management. Previous "track record" becomes an important factor in evaluating the proposal and the division general manager who champions it through the approval process.

Source: Schmenner, Roger. *Making Business Location Decisions.* Englewood-Cliffs, NJ: Prentice-Hall, 1982.

In his 1982 book, Roger Schmenner provides examples of the decision making process for location decisions of several Fortune 500 companies (67-70.) The examples utilize the terms of Table 2 to explain and contrast the processes. Schmenner included Brown Shoe Company as an example of a decentralized, division-only study. The example is especially relevant to this study because it involves a labor-intensive operation.

Example: Brown Shoe Company

Brown Shoe is a member of the Brown Group, Incorporated, and its sales account for half of the sales of the Group. Brown Shoe's decision to expand was initiated by a business plan. Various departments provided input into the plan, e.g., the marketing department provided forecasts.

Once a need for expansion is identified, Brown Shoe Company senior management decides whether a new plant will be built or the existing plants will be expanded. If a new plant will be built, location search responsibilities go to the Senior Vice President (SVP) of Manufacturing, who undertakes it as a one-person effort. The SVP then contacts state economic development agencies; specifies Brown Shoe's needs in a plant site; and waits for a response from the agencies. The SVP then selects the best location alternatives from the responses. The next step in the decision-making process is to assess labor force quality. This step is especially important since shoe manufacturing is a labor-intensive industry. This assessment is accomplished by consulting the plant managers who are closest to the sites being considered, and by surveying existing firms in the industry. The results are reported to the SVP. If reports are positive, the personnel or industrial relation department conducts a formal labor survey and the SVP for manufacturing visits the areas to form an impression. The SVP makes the decision, but must provide support of the decision to higher levels of management and the board of directors. Appendix A outlines the information gathered during a firm's search for a location and provides insight into data sources available to firms.

2. **Review of Literature**

There is a wealth of literature on industrial location. Studies are divided along three lines: 1) those that use micro or individual firm data, 2) those that examine aggregate trends in employment, and 3) those that examine aggregate trends in capital investment (Waslyenko 1991, page 26.) Perhaps the critical distinction is whether aggregate or firm data are used, because the type of data used largely determines the specification of the model.[3] This section discusses two groups of literature: studies utilizing firm-level data and studies utilizing aggregate data. Dennis Carlton was one of the first to apply an econometric model to firm data. Timothy Bartik adopted Carlton's approach and provides more recent industrial location research. Thomas Plaut and Joseph Pluta provide a detailed examination of business activity using aggregate data. Wasylenko and McGuire provide more recent research and offer improvements upon Plaut and Pluta's approach.

2.1 Research Utilizing Firm-Level Data

Dennis Carlton (1983) simultaneously models both the location and employment choice of new branch plants across Standard Metropolitan Statistical Areas (SMSAs) [now referred to Consolidated Metropolitan Statistical Areas[4] (CMSAs.)] Carlton's research goal was a broad examination of the impact several variables have on the firm's location decisions.

Carlton uses a restricted profit function in a logit model. This model is a common approach and was first developed by Daniel McFadden as a qualitative choice

---

[3] Firm-level data are observations of the location decisions made by individual firms, whereas aggregate data represent the total level of business activity at a state level.
[4] CMSA are combined primary metropolitan areas with a population of more than one million.

model. It is used when the dependent variable is a dummy variable (1/0 outcome.) The functional form of the model is based on the Cobb-Douglas production function. Carlton utilizes the duality of the profit function and the production function to predict the size of establishment.

The model suggests that probability that firm (i) will locate in state (j) can be measured by the model shown below (assuming firms are profit-maximizing.)

$$Pr(j) = \frac{exp \sum_k \beta_k \ ln \ X_k(j)}{\sum_s exp \sum_k \beta_k \ ln \ X_k(j)}$$

β_k is a vector of coefficients for the exogenous variables at location j. The probability of firm (i) locating in state (j) is then equal to the expected profits in (j), divided by the sum of the profits available in all other states. Coefficients of all variables except tax are elasticities.[5]

Carlton uses industry data at the four-digit level. He selected the industries for the study using the following criteria: 1) industry is not tied to location by local supply and demand factors (e.g., transportation costs were low enough for the industry to produce for a national market); 2) the number of new businesses being created is large; 3) industries vary with respect to the amount of energy used; and 4) industries differ in the level of technological sophistication required of labor. The industries he selected were Fabricated Plastic Products (3079), Communication Transmitting Equipment (3662), and Electronic Components (3679). His dataset includes the following variables: wages, price of electricity,

[5]Elasticity is the percent change in probability resulting from a one-percent change in the independent variable.

price of natural gas, number of employees, unionization, corporate tax, property tax, unemployment rate, and production hours.

Carlton includes wages for production workers in the two digit SIC code as a measure of the cost of labor. Carlton justifies this by asserting that workers of similar skills will be desired by the firm and that these workers comprise the entire labor supply. This assumes that the available pool of labor must come from the same industry. Unless the labor skills required are very specialized, the firm should be able to draw from the entire labor force. Moreover, a firm requirement of a skilled labor force should be reported in the education variable. In addition, his use of engineers available in a state as an indication of the quality of the labor force is not clearly supported. It is not reasonable to assume that all industries will need engineers and will subsequently place a high value on their availability. Furthermore, other skills which may be more important such as vocational training as a technician are ignored. For example, skilled technicians are extremely valuable in the transportation equipment industry because they possess specialized training for testing, handling, and assembling components.

Carlton concludes that by exploiting the link between firm location and firm size, one can obtain better estimates of the location model and accurately predict the employment variable. He also found that: 1) the size of plants (number of employees) can be predicted; 2) the wage effect cannot be measured very precisely; 3) energy costs have a large effect; 4) taxes and state incentive programs do not seem to have major effects; 5) existing concentrations of employment matter a great deal with the effect being stronger for industries with smaller average plant size; and 6) available technical expertise is likely to be important for highly sophisticated industries. However, these results are questionable. Carlton's sample size ranges from 84 to 290 and is not large enough to employ the logit model. The maximum likelihood estimator he used tends to work best with large samples.[6] Further, both Carlton and Timothy Bartik utilize firm-level data. The source of their data is the Dun and Bradstreet data files. This database has many flaws which have not been denied by either author. In fact, Bartik noted that 34% of the new business starts from 1972 and 1978 were not reported in the data files. Bartik attempts to correct for the flaws; Carlton does not.

Timothy Bartik (1985) employs the logit model used by Carlton and others. He continues on Carlton's approach and adds two advances in modeling branch plant location. First, he improves the Independence of Irrelevant Alternatives assumption inherent in Carlton's model. The Independence of Irrelevant Alternatives Assumption states that if a plant has higher profits in one state, there is no implication that it will in another state. Since states located in the same region share many of the same characteristics, this assumption may not be valid. Bartik modifies the assumption by including a fixed correlation factor in the error terms for states within a region, but he includes no such correlation across regions. As a result, the Independence of Irrelevant Alternatives Assumption is restated to mean that the higher profits in one state may indicate higher profits could be earned in another state, but not in another region. Second, Bartik includes information about the number of plant sites available. Carlton and most researchers have used state level data on characteristics such as educational level etc., instead of data for the specific sites that are available. Bartik includes land area as a proxy for the number of sites available within a state and subsequently considers plant site information. For example, even if two states have similar wage and education levels, Bartik's model can make a distinction by the number of sites available.

[6] When used with large samples, the error term tends to have a normal distribution and statistical inferences such as t-tests can be utilized.

Bartik includes variables representing land area, unionization, corporate tax rate, property tax rate, unemployment, workers compensation, road miles, manufacturing hours, wages, education level, construction costs, population density, and energy prices. He finds that a 10% increase in land area causes almost a 10% increase in new plants, and that unionization is very important. However, he does not find educational attainment, as well as several other variables, to be significant. In fact, the signs of the variables are incorrect as well.

Bartik applies his model to all industries. Industries are so diverse that they cannot be accurately modeled this way. Specifically, industries differ in their orientation. As described in Section 1.1, the orientation of a firm dictates which variables are important to a firm. Unless every variable important to each type of orientation is included, the importance of the variables cannot be accurately measured. Stated differently, empirical problems such as bias from omitted variables are incurred. When Bartik applied his model to two digit industries as an aside, it performed poorly, and he offers this fact as justification for examining all industries at once. The poor performance may be an indication of his misspecification, rather than justification for applying the model to all industries at once. Given that the model generated incorrect signs and failed to be significant for several important variables, misspecification is even more likely.

2.2 Research Utilizing Aggregate Data

Most researchers that utilize aggregate data employ disequilibrium models (Wasylenko and McGuire, Plaut and Pluto, Helms, and Newman.) However, the models utilized by Carlton and Bartik also model disequilibrium. (The models differ in functional form, however.) Robert Newman (1983) explains that there are four types of equilibrium models: 1) cross sections of locations as a static factor market equilibrium, 2) single cross section of locations as factor disequilibria, with factor mobility as a means of equilibration, 3) cross sections at two or more points in time, whereby the observed factor mobility is treated as a comparative static response to observed changes in the economy, and 4) pooled[7] cross sections and time series data, whereby differential factor mobility is treated as the effect of differential rates of change in the determinants of factor market equilibrium.

The first type of equilibrium model assumes that the market is in equilibrium state. This approach requires that existing firm locations be treated as the effect of levels of current locational factors (Newman, 1983.) As a result, all factors must be included. This is a tremendous task since numerous variables affect industrial location and many are jointly determined. Fortunately, this model is not a necessary approach—alternative models exist which require examination of only the changes that result from factor disequilibrium. The second type of model hypothesizes that locational change (e.g., growth in the number of establishments) is the effect of disequilibrium at the beginning of the period examined. Central to this approach is the requirement that excess demand or supply be identified and measured. The interpretation of results then becomes, for example, wages are low, relative to productivity measured by education attainment. The third model examines changes as a function of changes in independent variables. Variables that do not exhibit change during the period can be omitted from this type of model. The fourth model is essentially a pooled version of the third model.

As mentioned above, Timothy Bartik and Dennis Carlton utilize disequilibrium models. However, their models are most often discussed in terms of logit

---

[7] Pooled datasets result when datasets for two different time periods are combined. For instance, Thomas Plaut and Joseph Pluta (1983) used a pooled dataset of industrial growth measures and explanatory variables from 1967-1967 and 1972-1977.

analysis, since that is the specific form of their disequilibrium model. The form stems from a restricted profit function. Equilibrium analysis models of aggregated data generally use linear models. Profit functions which must be translated into logs cannot be employed because activity can and is often negative. Wasylenko and Plaut and Pluto (1983) both use linear models. Thomas Plaut and Joseph Pluta (1983) looked at aggregate changes in economic variables such as total employment.

Plaut and Pluta's model consisted of three separate regressions with the following as dependent variables: 1) percent change in employment, 2) percent change in real value added, and 3) percent change in real capital stock. They use a pooled dataset consisting of the time periods 1967-72 and 1972-77. The variable coefficients are forced to be equal for each period and ordinary least squares is used to estimate the equations. They propose that simultaneous equations are not needed because all variables are measured at the beginning of the period. The independent variables include market accessibility, wage rate, unemployment rate, union activity, inherent productivity, energy prices, land prices, climate, business climate, tax effort, corporate taxes, personal taxes, sales tax, property tax, education expenditures, and welfare expenditures, and 1972 dummy. Many of these variables were combined to reduce multicollinearity by using principal components analysis.

Plaut and Pluta found market accessibility, wage rate, unemployment rate, union activity, climate, business climate, tax effort, property taxes, and expenditures on education to be significant determinants of the change in employment. Climate is the largest determinant of change in employment levels. They also found that some variables are more important to different types of growth, e.g., employment versus capital stock growth.

Plaut and Pluta have been criticized because the time period they chose was one of considerable manufacturing cycles and no variables were included to capture this phenomenon (Newman, 1985.) They acknowledge this, but explain that the number of relevant cyclical variables makes the model too cumbersome to estimate. Another shortcoming of their study is their choice of metric for education. Expenditures is a common metric, however, it does not directly measure labor force quality and is only a mediocre indicator of the quality of public schools.

Wasylenko and McGuire (W&M) chose the percentage change in state employment to examine location decisions from 1973 through 1980. The industries chosen were very broad: manufacturing, transportation, wholesale trade, retail trade, financial services, and other services. W&M include variables representing wage rate, lost working time, percentage of state population that are working age, median years of education, average electric bill for industries, education expenditures, welfare expenditure, tax effort, corporate income tax rate, personal income tax rate, effective corporate tax rate, effective personal income tax rate, percentage of general revenue from sales taxes, population density, per capita state income, high and low average temperatures, and employment in industry divided by land area. W&M found: 1) higher wages, utility prices, personal income tax rates, and an increase in the overall level of taxation discourage employment growth in several industries, 2) variables are not significant across all industries, and 3) higher spending on education and per capita income favorably impact job growth. Other significant variables include median education level, percentage change in state tax effort, wages (at 10% level), and temperatures (at 10% level.)

Similar to Plaut and Pluta, W&M utilize the linear model which is shown below. Since heteroscedasticity was detected, they used generalized least squares to correct for the bias.

$$PCT(i) = a + F(b) + t(c) + P(d) + A(e) + \varepsilon$$

Where:

PCT(i): Percent change in employment
a: Constant
ε: Error term
F: Independent variables for labor, productivity, and energy cost
t: Independent variables for fiscal climate
P: Independent variables for markets
A: Location-specific attributes
b,c,d,e: Coefficients

3. <u>Materials and Methods</u>

The focus of this research endeavor was to examine specific industries, since to examine all industries would yield results from which few supportable conclusions could be drawn. Given that the goal of this research was to isolate the impact of educational attainment on firm location, industries were chosen based upon the skill level of the labor used. (The model is to be applied separately to each industry.) Specifically, chosen industries employ medium to highly-skilled labor. The SIC code and name of the industries (two-digit SIC code) selected are listed below:

36: Electronic and other electric equipment
37: Transportation equipment

Note: Observations containing missing values or having an insufficient sample were deleted.

3.1 <u>Theoretical Foundation of Approach</u>

The dataset underlying this research is a cross-sectional dataset for changes in business activity from 1993 to 1994. Since the data utilized in this study are aggregated, the corresponding model differs from the one used by Carlton and Bartik. The approach, however, closely follows that of Plaut and Pluta and Wasylenko and McGuire (1985.)

This study focuses on the level of business activity measured as the percent change in the number of establishments in year t (1994) to year t-1 (1993). A disequilibrium model provides the basis for the functional form. The markets must be in disequilibrium since profit-seeking firms are choosing to locate where they perceive above-average profits are possible. Location literature supports this approach (Plaut and Pluta 1983, Wasylenko and McGuire 1985, and Newman 1988.)

The changes/levels model or the second model described in Section 2.2 most appropriately models the hypothesis that firms are attracted by existing levels of education attainment. Specifically, firms are hypothesized to decide to locate in states where the education attainment level is high relative to other states. This subsequently generates increases in business activity levels. Equation 1 shows the functional form of the model.[8]

Equation 1. Disequilibrium Model of Changes in Business Activity

$$\frac{Activity_{it} - Activity_{it-1}}{Activity_{it-1}} = \lambda\beta X_{it} - \lambda Activity_{it-1} + e_{it}$$

Where:
Activity: Measured by the number of establishments
X: Matrix of independent/explanatory variables
E: Disturbance
μ: Random Disturbance
$e = \lambda E_{it} + \mu_{it}$
λ: Speed of adjustment parameter

The variables included in Equation 1 were selected because industrial location theory lists them as important locational determinants for firms oriented toward local inputs.

Business activity is the percent change in the number of total establishments in year t (1994) to year t-1 (1993). This is the dependent variable. In actuality, this is a measure of changes in business activity. Alternatively, the percent change in the number of production workers measures business activity. During the course of this research both the percent change in the number of establishments and the number of production workers were used as the dependent variables.

[8] Bartik provides an excellent explanation of the various equations for disequilibrium models on pages 79-80 of *Industry Location and Public Policy*. Herzog, Henry W, and Alan M. Schlottman, eds. Knoxville, Tennessee: University of Tennessee Press, 1991.

For the chosen model and variables, a suitable fit could not be obtained with the percent change in the number of employees as the dependent variable.

The unit price of energy per state represents the cost of energy. Energy prices vary considerably by state, so this is an important variable. The expected sign is negative, since increases in costs are deterrents to plants seeking a location for operations.

Unionization is measured as the percent of private workers that are union members (%LFUnion.) Unfortunately, the same data for manufacturing workers were not available. This variable is expected to have a large negative sign, per Bartik's research (1985.) Firms with labor-intensive operations have typically sought cheaper labor in the South and West since unionization is lower. Many disadvantages are associated with a unionized firm: 1) the inability of the managers to pursue enhanced productivity via technological advances because of a contract clause agreed to years prior, and 2) wage levels tend to be higher than those of non-union firms. Even capital-intensive industries shy away from unionized areas because construction of new facilities may be delayed.

The average annual pay level of employees by state (Wages) is an important factor to labor-intensive firms. For instance, apparel and shoe industries seek low labor costs to remain competitive. Randall Eberts found that these firms search through isolated rural areas that have not previously been industrialized, or areas industrialized by firms that employ overwhelmingly more men than women. Eberts also found that the characteristics of a labor force such as educational attainment only partially explain the differences in wages across regions; wage levels are also determined by the value the regional labor markets place on the characteristics. Wage is hypothesized to have a negative sign.

The quality of the labor force is measured by the percentage of a state's population with a college degree (EDUC90.) This measure indicates the importance of attaining education <u>beyond</u> the high school level. Recognizing that advanced skills make a labor force more attractive, many southern Atlantic states provide labor training programs such as on-the-job (OTJ) or CETA programs. This variable typically is hypothesized to have a positive sign. However, many researchers have found a negative coefficients (Bartik, 1985.)

The impact of unemployment rates is somewhat unclear (URate.) Unemployment rates could reflect a low local demand for the product (Carlton, 1983.) A high unemployment rate can also reduce the costs of assembling and maintaining a work force. Given the uncertainty of unemployment's effect, no sign has been hypothesized.

The amount of value added to goods by processing done by firms in a given state (ValAdd) is used to indicate the availability of raw materials to firms that locate in a state. Having raw materials readily available reduces the cost of production. This variable also serves as a measure of agglomeration. As a result, value added is expected to have a positive effect on business activity.

The ratio of taxes paid on corporate income to total tax collections (Tax) measures the impact of taxes. Government incentives are likely to be reflected by lower business taxes, so a small ratio may indicate the existence of incentives. Since public services are funded by tax revenue, above average tax rates could indicate above average levels of public services. The expected sign of tax rates is unclear and many researches have found mixed results (Papke; Carlton, 1983.)

The lag of business activity (estab93) measures the impact of a prior period on the current level of business activity. λ (the speed of adjustment parameter) quantifies the importance of past activity and indicates of the degree of disequilibrium.

The data sources for the variables listed above and their expected sign are shown in Table 3.

Table 3. Data Sources and Expected Signs

Variables/ Expected Sign	Disaggregate	Report	Agencies
Energy (-)	State	Annual Energy Price and Expenditure Report	Department of Energy (DOE)
%LFUnion (-)	State	Union Membership and Earnings Data Book: Compilations from the Current Population Survey	Bureau of National Affairs, Inc.
Wage (-)	State	Annual Pay By State and Industry	Bureau of Labor Statistics (BLS)
EDUC90 (+)	State	Educational Attainment in the United States	BLS & Bureau of Census (BOC)
ValAdd (+)	State, 2-digit SIC code	Annual Survey of Manufactures	BOC
URate (uncertain)	State, Civilian Noninst. LF	Geographic Profile of Employment and Unemployment	BLS
Tax (uncertain)	State	US Statistical Abstract	BOC
% change in # estab. (dep. variable)	State, 2 digit SIC code	County Business Patterns	BOC

3.2 <u>Comparison to Other Research</u>

The model incorporates most of the variables included by Carlton. Since this research utilizes aggregate data, Carlton's model was not used. This study is similar to Wasylenko and McGuire and Plaut and Pluto. However, this study looks at a much shorter time period—one year instead of five or more years. In addition, Wasylenko and McGuire and Plaut and Pluto do not include an adjustment parameter (λ). The partial adjustment model accounts for the degree of disequilibrium in the markets and subsequently more theory is incorporated into regression results.

In general, this study uses the same metric of variables as Plaut and Pluta and Wasylenko and McGuire, but differs from those chosen by Carlton for a few variables. The percent of a state's population with a college degree is included as an indication of labor force quality, where Carlton used the number of engineers and Plaut and Pluta used expenditures. As mentioned in Section 2, the percent of the state population with a college degree is believed to be a more reliable and direct indication of the skill level of the general labor force. In addition, this study measures wages differently from Carlton. As mentioned in Section 2, a general measure is believed to be a better measure of the wages of the labor pool for new firms and branch plants. It could be argued that a firm must pay the prevailing industry wage. However, this argument assumes that firms intend to lure existing workers in the industry away from their current positions. This argument lacks support because there is no evidence that the workers comprising the labor supply must have experience in a specific industry. Another difference is the measure of agglomeration. This research uses value-added by manufacturers, whereas Carlton uses production man-hours. The variables tend to be extremely correlated, so the difference is not a concern. The most notable difference between this study and those done by Plaut and Pluto and Wasylenko and McGuire is the number of variables included. Since this study did not utilize a pooled dataset, the number of variables was limited to preserve degrees of freedom. This is not a concern because this study consolidated many of the variables included in their studies. Another difference is that this study uses percent change in the number of establishments instead of employment as the dependent variable.

4. **Results**

Several tests were used to examine potential empirical concerns before and after the generating regression results. The following sections address the problems of multicollinearity, omitted variables, and heteroscedasticity. Appendix B includes the output of these tests.

4.1 Tests For Correlation

The final step before utilizing the model was to examine variables for correlation. To check for correlation among variables, the correlation matrix shown in Table 4 was calculated. Correlations of .80 or higher were considered indicative of causing multicollinearity problems.[9] No variables met this criteria. During the modeling process, the lag of the number of establishments (estab93) was found to be a linear combination of the other variables. Principal components analysis was used to combine this variable with the variable representing value added since the two variables are generated by the same survey and are likely to be related.

[9] Thomas Plaut and Joseph Pluta (1983) use a similar correlation threshold to determine if multicollinearity poses a serious problem.

Table 4. Correlation Matrix

SIC 36									
	energy	wages	lfunion	valadd	urate	tax	pc_nestb	Educ90	estab93
Energy	1.00								
Wages	0.31	1.00							
Lfunion	(0.04)	0.61	1.00						
valadd	0.01	(.30)	(.32)	1.00					
urate	0.06	0.29	0.20	(0.17)	1.00				
tax	0.39	0.47	0.29	(0.06)	0.07	1.00			
pc_nestb	(0.17)	(0.11)	0.01	(0.05)	(0.22)	(0.27)	1.00		
educ90	0.47	0.54	0.01	(0.27)	(0.17)	0.15	0.01	1.00	
estab93	0.24	0.47	0.23	(0.12)	0.39	0.31	(0.19)	0.31	1.00

SIC 37									
	energy	wages	lfunion	valadd	urate	tax	pc_nestb	Educ90	estab93
energy	1.00								
wages	0.25	1.00							
lfunion	(0.08)	0.62	1.00						
valadd	(0.08)	0.27	0.29	1.00					
urate	0.06	0.28	0.12	(0.10)	1.00				
tax	0.17	0.52	0.59	0.27	0.17	1.00			
pc_nestb	(0.42)	(0.18)	0.10	(0.15)	0.05	(0.20)	1.00		
educ90	0.50	0.47	(0.02)	(0.23)	(0.17)	0.05	(0.36)	1.00	
estab93	0.14	0.46	0.19	0.23	0.38	0.36	(0.25)	0.13	1.00

4.2 Test for Omitted Variables

Omitted variables is a concern shared by most empirical studies. To test for omitted variables, the underlying theory was re-visited to identify any potential variables to include. The only possible additional variable to include is transportation costs. Quality data on transportation costs are not available. With this in mind, one of the criteria for the selection of industries was that transportation costs were not significant. The orientation theory discussed in Section 1.1 supports this approach. The Ramsey test (Studenmund) for omitted variables was also conducted. The results did not indicate the problem of omitted variables.

4.3 Tests For Heteroscedasticity

Heteroscedasticity is common in cross-sectional datasets because variance tends not to be constant. For instance, the variance of personal income is likely to be quite different for a large state like California versus a small state like Rhode Island. The variables comprising the dataset used for this study were defined in relative terms, e.g., value-added was measured per establishment. This was designed to reduce the possibility of heteroscedasticity; however, the problem persisted. The Cook-Weisberg (STATA) and White tests (Studenmund) were used to assess heteroscedasticity. Results of tests showed heteroscedasticity was indeed a problem. In addition, graphs of the residuals with respect to variables believed to cause the heteroscedasticity were created. Based upon the graphs, the number of establishments appears to be the likely cause of nonconstant variance. (These graphs are included as Appendix B.) To correct for heteroscedasticity, weighted least squares regression analysis was utilized. The number of establishments was used as the weight.

4.4 Empirical Results and Conclusions

The results for each industry are summarized below in Table 5. Results indicate that educational attainment is a significant determinant of where firms in the transportation industry decide to locate or expand operations. However, educational attainment is not a significant determinant for the electronic equipment industry. One possible explanation for this result is that the nature of work performed in the transportation equipment industry requires more skill than the work performed in the electronic equipment industry.

Wage and unemployment levels are significant factors for both industries. With the exception of educational attainment and tax levels, the significance of variables does not vary across industries.[10] The coefficients for education and wages had the expected signs. However, corporate tax levels and unemployment rates had positive signs which is somewhat surprising, even though previous studies have found both positive and negative coefficients for both variables. Since public services are funded by tax collections, the positive sign could be a reflection of a higher level of public services. As mentioned previously, the unemployment rates could have a positive impact on location decisions, if high unemployment rates reduce the cost of assembling and maintaining a workforce.

Since principle components analysis was necessary, the adjustment parameter (λ) could not be separated. As a result, the degree of disequilibrium cannot be determined.[11] Appendix B includes the statistical program used to generate the results as well as the complete output.

---

[10] The magnitude of the coefficients shown in Table 5 is misleading. Weighted least squares was used to estimate the coefficients, so β for each variable is actually β for variable/weight.

[11] Technical note: The coefficients for the independent variables generated from Equation 1 are actually $\lambda*\beta$. As a result of principal component analysis, the adjustment parameter (λ) could not be separated from the coefficients. Table 5 shows the coefficients without factoring out λ. Only the sign of λ is known. Research shows that the sign is negative (Bartik, 1991.) This means that λ changes the sign of the coefficients. As a result, the coefficients generated by the model were multiplied by (-1.) Table 5 reflects this change.

Table 5. Empirical Results

	SIC36		SIC37	
	Coefficient (Standard Error)	t-statistic (P-Value)	Coefficient (Standard Error)	t-statistic (P-Value)
Energy	-34.11 (53.97)	-0.63 (0.53)	64.03 (77.95)	0.82 (0.42)
%LFUnion	-1709.10 (2807.30)	-0.61 (0.55)	1772.60 (3619.00)	0.49 (0.63)
Wages	-0.07 (0.35)	-1.99 *(0.05)	-0.15 (0.03)	-4.91 *(0)
EDUC90	40.50 (28.39)	1.43 (0.16)	93.97 (19.84)	4.74 *(0)
Urate	17769.12 (5820.47)	3.05 *(0.00)	13750 (6028.28)	2.28 *(0.03)
Tax	5782.98 (2868.04)	2.02 *(0.05)	2172.91 (4074.77)	0.53 (0.60)
PCA of ValAdd and estab93	1.32 (.69)	1.90 **(0.07)	1.134.22 (0.75)	1.51 (0.14)
Intercept	-0.48 (1.01)	-0.05 (0.96)	2.2 (1.00)	2.19 *(0.04)
R^2	0.71		0.74	
Number of obs.	43		38	

*Significant at the 95 percent level.
**Significant at the 90 percent level.

5. Summary

Firms that seek sites for both new plants or expansion plants can benefit from an understanding of industrial location research. This research shows that firms consider many descriptive location factors such as wage levels and education in their location decision. This research utilizes industrial location theory and literature to quantify the impact the level of education attained by a state's population on a firm's decision locate as a new firm or branch plant in that state. As discussed in Sections 1.1 and 1.2, the firm's orientation weighs heavily in the factors important to a firm. Accordingly, this examination of factors important to firms oriented toward local inputs considers the characteristics of their specific orientation.

The results indicate that industries such as transportation equipment seem to place a premium on states whose labor force has a relatively high level of education attainment. States wishing to attract this industries may find it beneficial to encourage higher education, both for the well-being of its residents and for the economy of the state.

References

Bagchi-Sen, Sharmistha, "Foreign Direct Investment in U.S. Metropolitan Areas, 1979-83." *Urban Geography* March/April (1989): 121-37.

Bartik, Timothy J. "Business Locations Decisions in the United States: Estimates of the Effects of Unionization, Taxes, and Other Characteristics of States." *Journal of Business and Economic Statistics* January (1985): 14-22.

Bartik, Timothy J. "The Effects of Property Taxes and Other Local Public Policies on the Intrametropolitan Pattern of Business Location." *Industry Location and Public Policy*. Eds. Herzog, Henry W, and Alan M. Schlottman. Knoxville, Tennessee: University of Tennessee Press, 1991. 13-30.

Bartik, Timothy J. "Small Business Start-ups in the United States: Estimates of the Effects of characteristics of States." *Southern Economic Journal* April (1989): 1004-18.

Bartik, Timothy J. *Who Benefits From State And Local Economic Development Policies?* Kalamazoo, Mich.: W. E. Upjohn Institute for Employment Research, 1991.

Bartik, Timothy J., Charles Becker, Steve Lake, and John Bush. "Saturn and State Economic Development." *Forum for Applied Research and Public Policy.* Spring (1987): 29-40.

Calzonetti, F.J. and Robert T. Walker. "Factors Affecting Industrial Location Decisions: A Survey Approach." *Industry Location and Public Policy.* Eds. Herzog, Henry W, and Alan M. Schlottman. Knoxville, Tennessee: University of Tennessee Press, 1991. 221-41.

Carlton, Dennis W. "The Location and Employment Choices of New Firms: An Econometric Model." *Interregional Movements and Regional Growth.* Ed. William C. Wheaton. Washington, DC: The Urban Institute, 1979: 13-50.

Carlton, Dennis W. "The Location and Employment Choices of New Firms: An Econometric Model with Discrete and Continuous Endogenous Variables." *Review of Economics and Statistics* 65.3 (1983): 440-49.

Deich, Michael. "An Empirical Analysis of State Taxation and Manufacturing Plant Location." Dissertation. University of Michigan, 1990.

Deich, Michael. “State Taxes and Manufacturing Plant Location.” *Proceedings of the Eighty-Second Annual Conference on Taxation*, *Atlanta, GA, October 8-11*. Ed. Frederick D. Stocker. Columbus, Ohio: National Tax Association-Tax Institute of America, 1989. 240-49.

Eberts, Randall and Joe Stone. *Wage And Employment Adjustment In Local Labor Markets.* Kalamazoo, Michigan: W. E. Upjohn Institute for Employment Research, 1992.

Fortune. “500 Largest Corporations.” Publication on the Internet. 1995.

Helms, L. Jay. “The Effect of State and Local Taxes on Economic Growth: A time Series Cross Section Approach.” *Review of Economics and Statistics* 67.4 (1985): 574-582.

Herzog, Henry W, and Alan M. Schlottman. *Industry Location and Public Policy*. Knoxville, Tennessee: University of Tennessee Press, 1991.

Hirsch, Barry and David MacPherson. *Union Data Book.* Washington, DC: Bureau of National Affairs, 1995.

Levy, John M. *Economic Development Programs For Cities, Counties And Towns.* Second edition. New York: Praeger Publishing, 1990.

McFadden, Daniel. “Conditional Logit Analysis of Qualitative Choice Behavior.” In *Frontiers in Econometrics*. Ed. Paul Zarembka. New York: Academic Press, 1974. 105-142.

Newman, Robert J. and Dennis H. Sullivan. “Economic Analysis of Business Tax Impacts on Industrial Location: What Do We Know, and How Do We Know It?” *Journal of Urban Economics* 23 (December 1988): 215-234.

O’Sullivan, Arthur. *Urban Economics*. Boston: Irwin Publishing, 1993.

Plaut, Thomas and Joseph Pluta. “Business Climate, Taxes and Expenditures, and State Industrial Growth in the United States.” *Southern Economic Journal* July (1983): 99-119.

Schmenner, Roger, Joel C. Huber, and Randall L. Cook. “Geographic Differences and the Location of New Manufacturing Facilities.” *Journal of Urban Economics* 21 (1987): 83-84.

Schmenner, Roger. “Energy and the Location of Industry.” *Energy Costs, Urban Development, and Housing*. Ed. Anthony Downs and Katharine L. Bradbury. Washington DC: The Brookings Institution, 1984. 188-225.

Schmenner, Roger. *Making Business Location Decisions.* Englewood-Cliffs, NJ: Prentice-Hall, 1982.

STATA. *Reference A-F*. Release 5.0. College Station, TX: STATA Press, 1997.

Studenmund, A. H. *Using Econometrics: A Practical Guide.* Second edition. New York: Harper Collins, 1992.

United States. Department of Energy. *Annual Energy Price and Expenditure Report*. Washington, DC: GPO, 1994-95.

United States. Bureau of Census. *Annual Survey of Manufactures*. Washington, DC: GPO, 1993-95.

United States. Bureau of Census. *County Business Patterns*. Washington, DC: 1992-95.

United States. Bureau of Labor Statistics and Bureau of Census. *Educational Attainment in the United States*. Washington, DC: GPO, 1994-95.

United States. Bureau of Labor Statistics. *Annual Pay By State and Industry*. Washington, DC: GPO, 1994.

United States. Bureau of Labor Statistics. *Geographic Profile of Employment and Unemployment*. Washington, DC: GPO, 1995.

United States. *US Statistical Abstract*. Washington, DC: GPO, 1994-96.

Wasylenko, Michael and Therese McGuire. "Jobs and Taxes: The Effect of Business Climate on States' Employment Growth Rates." National Tax Journal 38.4 (1985): 497-512.

Wasylenko, Michael J. "Empirical Evidence on Interregional Business Location Decisions and the Role of Fiscal Incentives in Economic Development." *Industry Location and Public Policy*. Eds. Herzog, Henry W, and Alan M. Schlottman. Knoxville, Tennessee: University of Tennessee Press, 1991. 13-30.

Woodward, Douglas P. "Locational Determinants of Japanese Manufacturing Start-ups in the United States." Working paper. University of South Carolina, 1990.

Appendix A

Table A-1. Summary of Information Gathered for a Location Decision (Schmenner, 1982)

Item	Quanti-fiable	Data Typically Developed From	Company Unit Typically Respons-ible
Access To Markets/Distribution Centers			
Cost of serving present markets/distribution centers	Estimate possible	• Location of markets/distribution centers, • Quantities of each market/ distribution center, shipping mode, freight rates, handling charges	Logistics
Trends in sales by area/ability to generate sales by company presence in area	Somewhat	Estimates of the same data listed above	Marketing
Access To Supplies			
Cost of transporting supplies	Estimating possible	• Location of suppliers for which transport expenses will be incurred by plant, • Quantities shipped from each location by which mode, freight rates, handling charges	Purchas-ing, logistics
Trends in supplier by area	Somewhat	Estimates of the same data listed above	Purchas-ing
Community Government Aspects			
Ambiance/charisma of community	No	• Site visit • Hearsay polls of people's preferences	Location team

Table A-1. Summary of Information Gathered for a Location Decision (Schmenner, 1982)

Item	Quanti-fiable	Data Typically Developed From	Company Unit Typically Respons-ible
Cost of living	Yes	BLS area figures	Location team
Cooperation with established local industry	No	Site visit	Location team
Community price (appearance, activity, and citizen views)	No	Site visit	Location team
Housing (availability and prices)	Somewhat	• Site visit, • Discussions with realtors	Location team
Schools, cultural attractions, recreation activities	Somewhat	• Site visit, • Program offerings from state/local sources, • % going to college	Location team
Colleges nearby and graduate programs	Somewhat	State/local sources	Location team
Churches and civic groups	Somewhat	Site visit	Location team
Competitive Considerations			
Competition's location	Yes	Industry sources	Division
Likely competition reaction to this new site	No	Industry sources, own knowledge	Division
Environmental Considerations			
Government "attainment" area or note?	Yes	Federal/state environmental protection agencies	Environ-mental affairs
Are pollution rights of other companies available for purchase?	Yes	State EPA development agencies, companies in area	Staff, location team
Ease and speed of compliance	No	• Federal and state EPA, • Militancy of local conservation efforts	Staff

Table A-1. Summary of Information Gathered for a Location Decision (Schmenner, 1982)

Item	Quanti-fiable	Data Typically Developed From	Company Unit Typically Respons-ible
Interaction With Rest Of Corporation			
Is new plant to be a satellite of another plant or not?	Yes	Own knowledge	Division
Is plant to be supplied by or to supply other company plants?	Yes	Own knowledge	Division
Extent of expected engineering/ management trouble-shooting form headquarters.	No	Own knowledge	Division
Labor			
Prevailing wage rates	Somewhat	• BLS, • State/local publications, • Poll of manufacturers in the area	Personnel Location team
Extent and militancy of labor unions in area	Somewhat	• Does the state have right-to-work" law? • BLS data on union workers Work stoppages • NLRB certification/ Decert. Elections, • Poll of area manufacturers	Personnel Location team
Productivity (absenteeism, turnover, worker attitudes)	Somewhat	Poll of area manufacturers	Personnel Location team
Availability (population, area unemployment, commuting distances)	Somewhat	• BLS data, • Census data	Personnel Location team
Skill levels available	Somewhat	• Poll of manufactures in area, • Local training programs	Personnel Location team

Table A-1. Summary of Information Gathered for a Location Decision (Schmenner, 1982)

Item	Quanti-fiable	Data Typically Developed From	Company Unit Typically Respons-ible
Site Itself			
Area of site, sq. ft. and layout of each structure	Yes	• State/local development agencies, • Railroads, power companies, • Developers, realtors, • Site visit	Location team, especially real estate function
Price of site and any structure	Yes	Seller	Location team
Ability to option site, length, cost	Yes	Seller	Location team, especially real estate function
Condition of site and any structures (including structural assessment, topography, geology, and other concerns for construction and improvement)	Yes	Site visit especially once option is taken	Location team, especially real estate and engineer-ing functions
Area parking and traffic	Somewhat	Site visit	Location team
Construction, remodeling costs, insurance	Yes	• Site inspections, • Engineering plans	Project engineer-ing team
Taxes and Financing			
State income tax (corporate and personal)	Yes	State government office	Tax depart-ment
Local property and local income tax (if any)	Yes	Local government office	Location team
Unemployment and workmen's compensation	Yes	State government office	

Table A-1. Summary of Information Gathered for a Location Decision (Schmenner, 1982)

Item	Quanti-fiable	Data Typically Developed From	Company Unit Typically Respons-ible
Other state/local business taxes	Yes	State/local offices	
Tax incentives and/or concessions (holidays, abatements, exemptions, credits, accelerated depreciation and the like)	Yes	State/local offices	
Industrial and/or pollution control revenue bonds	Yes	State/local offices	
Transportation			
Trucking service (number and reputation of truckers, rates charged, quality of service)	Somewhat	Area trucking companies, other manufacturers in area	Location team, logistics
Rail service (number and reputation of railroads serving site, frequency of service, rates)	Somewhat	• Railroads, • Other manufacturers in area	Location team, logistics
Utilities services			
Availability quality and price of water sewerage, electricity power and natural gas	Somewhat	• Utility companies, • Poll of manufactures in area	Location team
Quality of roads, police, fire, medical, and other services	Somewhat	• State visit, • Poll of manufactures in area	Location team

Appendix B

I. Statistical program (STATA) to create dataset, generate correlation matrix, and run the model.

II. Graphs used to examine residuals for heteroscedasticity

STATA Code Used To Create Dataset, Run the model, and Generate Results

```
. *oooooooooooooooooooooooooooooooooooooooooooooooooooo*
.
. *SIC 36*
.
. insheet state busact energy lfunion wages edsc valadd newcap urate tax numemp
>  prodhrs emprat pc_nemp c_numemp c_nestab pc_nestb educ90 emp93 estab93 wager
> ate using c:\data\sic36_8.csv
(note: variable names in file ignored)
(21 vars, 50 obs)

.
. *This program for SIC36 and the purpose is to create a complete STATA dataset
>  that has all possible variables, show summary statistics, generate a correla
> tion matrix, and generate regression results for a disequilibrium model*
.
. label data "Initial input dataset:  SIC 36"

. label var busact "bus. Activity (#estab ratio)"

. label var state "state"

. label var energy "price of energy"

. label var wages "Annual pay, prod. workers"

. label var lfunion "% priv. sector workers in unions"

. label var edsc "% population with some coll. ed."

. label var valadd "value-added per establishment"

. label var newcap "expend. per estab. on new plant & equip."

. label var urate "unemployment rate for state"

. label var tax "corp. income tax as a % of total tax coll."

. label var numemp "# of production workers"

. label var prodhrs "# of production hours"

. label var emprat "change in # of production workers"

. label var pc_nemp "%change in # of prod. workers"

. label var c_numemp "change in # of prod. workers"

. label var c_nestab "change in # of estab."

. label var pc_nestb "%change in # of estab."

. label var educ90 "% of pop. with a coll. degree, 1990"

. label var emp93 "# of employees, 1993"

. label var estab93 "# of establishments, 1993"

. label var wagerate "labor cost per production hr."

.
```

```
. mvdecode _all, mv(-999.99)

state:   string variable ignored
valadd:  3 missing values
newcap:  4 missing values
numemp:  3 missing values
prodhrs: 3 missing values
emprat:  3 missing values
pc_nemp: 4 missing values
c_numemp:4 missing values
emp93:   3 missing values
wagerate:7 missing values

.
. describe

Contains data
  obs:            50                          Initial input dataset:  SIC 36
 vars:            21
 size:         4,650 (99.3% of memory free)
-------------------------------------------------------------------------------
   1. state       str14  %14s                 state
   2. busact      float  %9.0g                bus. Activity (#estab ratio)
   3. energy      float  %9.0g                price of energy
   4. lfunion     float  %9.0g                % priv. sector workers in unions
   5. wages       long   %12.0g               Annual pay, prod. workers
   6. edsc        float  %9.0g                % population with some coll. ed.
   7. valadd      float  %9.0g                value-added per establishment
   8. newcap      float  %9.0g                expend. per estab. on new plant
   9. urate       float  %9.0g                unemployment rate for state
  10. tax         float  %9.0g                corp. income tax as a % of total
  11. numemp      float  %9.0g                # of production workers
  12. prodhrs     float  %9.0g                # of production hours
  13. emprat      float  %9.0g                change in # of prod. workers
  14. pc_nemp     float  %9.0g                %change in # of prod. workers
  15. c_numemp    float  %9.0g                change in # of prod. workers
  16. c_nestab    byte   %8.0g                change in # of estab.
  17. pc_nestb    float  %9.0g                %change in # of estab.
  18. educ90      float  %9.0g                % of pop. with a coll. deg. 1990
  19. emp93       float  %9.0g                # of employees, 1993
  20. estab93     int    %8.0g                # of establishments, 1993
  21. wagerate    float  %9.0g                labor cost per production hr.
-------------------------------------------------------------------------------
Sorted by:
     Note:  data has changed since last save

.
. summarize


Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
   state |       0
  busact |      50    1.019682     .064253   .9185185   1.333333
  energy |      50       8.597    1.516298       5.55       11.9
 lfunion |      50      .09696    .0452494       .024       .193
   wages |      50    30350.72    4900.952      21648      41982
    edsc |      50      .60304    .0418223       .515       .697
  valadd |      47    7.546391    5.168907   4.17e-13   24.71667
  newcap |      46    .5833747    .7774426   4.17e-13   5.135417
   urate |      50      .05634    .0127993       .029       .089
     tax |      50    .0744672    .0964483   5.48e-16   .6202965
  numemp |      47    19.22553    21.83128   1.00e-11      122.8
```

```
 prodhrs |        47    38.83617    45.86163    1.00e-11      268.7
  emprat |        47    1.028965    .1987554   -1.00e-14    1.512195
 pc_nemp |        46    5.133345    12.78355   -18.18182    51.21951
c_numemp |        46    .8913043    2.046007        -3.3        10.4
c_nestab |        50         .34    14.57228         -82          21
pc_nestb |        50    1.968232    6.425304   -8.148149    33.33333
  educ90 |        50      26.002    4.515664        16.1        34.4
   emp93 |        47    19.22553    21.83128    1.00e-11       122.8
 estab93 |        50      341.42    552.8269           6        3623
wagerate |        43    11.57781    2.042687    5.773585    15.89655

.
. drop if valadd ==.
(3 observations deleted)

. drop if newcap ==.
(1 observation deleted)

. drop if tax ==.
(0 observations deleted)

. drop if prodhrs ==.
(0 observations deleted)

. drop if pc_nemp ==.
(1 observation deleted)

. drop if emprat ==.
(0 observations deleted)

. drop if pc_nestb ==.
(0 observations deleted)

.
. *drop observations for Hawaii and Alaska*
. drop if state=="Alaska"
(1 observation deleted)

. drop if state=="Hawaii"
(1 observation deleted)

.
. *drop variables that are not included in the model*
. drop busact

. drop edsc

. drop newcap

. drop numemp

. drop prodhrs

. drop emprat

. drop pc_nemp

. drop c_numemp

. drop c_nestab

. drop emp93
```

```
. drop wagerate

.
. *generate a correlation matrix*
. correlate energy wages lfunion valadd urate tax pc_nestb educ90 estab93
(obs=43)

        |   energy    wages  lfunion   valadd    urate      tax pc_nestb
--------+---------------------------------------------------------------
  energy|   1.0000
   wages|   0.3052   1.0000
 lfunion|  -0.0436   0.6053   1.0000
  valadd|  -0.0095  -0.3001  -0.3276   1.0000
   urate|   0.0593   0.2869   0.2016  -0.1719   1.0000
     tax|   0.3915   0.4658   0.2877  -0.0569   0.0691   1.0000
pc_nestb|  -0.1729  -0.1065   0.0082   0.0492  -0.2209  -0.2677   1.0000
  educ90|   0.4742   0.5412   0.0062  -0.2728  -0.1696   0.1543   0.0055
 estab93|   0.2422   0.4659   0.2252  -0.1201   0.3926   0.3142  -0.1919

        |   educ90  estab93
--------+------------------
  educ90|   1.0000
 estab93|   0.3063   1.0000


.
. *run tests for heteroscedasticity and graph residuals as another test*
. fit pc_nestb energy lfunion wages educ90 valadd urate tax estab93

  Source |       SS       df       MS                  Number of obs =      43
---------+------------------------------               F(  8,    34) =    0.69
   Model |  259.447045     8  32.4308806               Prob > F      =  0.6939
Residual |  1588.30328    34  46.7148025               R-squared     =  0.1404
---------+------------------------------               Adj R-squared = -0.0618
   Total |  1847.75033    42  43.9940555               Root MSE      =  6.8348

------------------------------------------------------------------------------
pc_nestb |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  energy |   -.385846   .9281003     -0.416   0.680      -2.271973    1.500281
 lfunion |   23.14063   36.84067      0.628   0.534      -51.72862    98.00987
   wages |  -9.55e-06   .0004748     -0.020   0.984      -.0009746    .0009555
  educ90 |   .1490688   .4244214      0.351   0.728      -.7134593    1.011597
  valadd |   .0990724   .2500786      0.396   0.694      -.4091485    .6072933
   urate |   -84.7761   111.1674     -0.763   0.451      -310.6954    141.1432
     tax |  -54.97505   45.06411     -1.220   0.231      -146.5563    36.60623
 estab93 |  -.0009458   .0022787     -0.415   0.681      -.0055766     .003685
   _cons |   6.927398   11.62522      0.596   0.555      -16.69789    30.55269
------------------------------------------------------------------------------
.
. rvpplot energy, border yline(0)

. gphprint, saving(c:\data\final\enrgy36.wmf)

.
. rvpplot wages, border yline(0)

. gphprint, saving(c:\data\final\wage36.wmf)

.
. rvpplot valadd, border yline(0)

. gphprint, saving(c:\data\final\vadd36.wmf)
```

```
.
. rvpplot estab93, border yline(0)

. gphprint, saving(c:\data\final\est93_36.wmf)

.
. hettest

Cook-Weisberg test for heteroscedasticity using fitted values of pc_nestb
     Ho: Constant variance
         chi2(1)      =     17.56
         Prob > chi2  =      0.0000

.
. white

White's test for Ho: homoscedasticity
         against Ha: unrestricted heteroscedasticity

    test statistic W =            43
    Pr(chi2(41) > W) =        0.3856

.
. *apply weights to perform weighted least squares*

. gen weenrgy=energy/estab93

. gen welfunn=lfunion/estab93

. gen wewage=wages/estab93

. gen weeduc90=educ90/estab93

. gen wevadd=valadd/estab93

. gen weurate=urate/estab93

. gen wetax=tax/estab93

. gen weestb93=estab93/estab93
.
. *Principal component analysis to reduce collinearity*
. factor valadd estab93, pc
(obs=43)

            (principal components; 2 components retained)
Component    Eigenvalue     Difference    Proportion    Cumulative
------------------------------------------------------------------
     1        1.12011         0.24022      0.5601         0.5601
     2        0.87989               .      0.4399         1.0000

            Eigenvectors
 Variable |      1          2
----------+--------------------
   valadd |  -0.70711    0.70711
  estab93 |   0.70711    0.70711
```

```
. score f1
            (based on unrotated principal components)
            (1 scoring not used)

            Scoring Coefficients
 Variable |      1
----------+----------
   valadd |  -0.70711
  estab93 |   0.70711

. egen stdvadd=std(valadd)

. egen stdest93=std(estab93)

. gen pca= -0.70711*stdvadd + 0.70711*stdest93
.
. label var pca "Combination of estab93 and valadd"

.
. *disequilibrium model depvar is % change in #estab*
. fit pc_nestb weenrgy welfunn wewage weeduc90 weurate wetax pca

  Source |       SS       df       MS                  Number of obs =      43
---------+------------------------------               F(  7,    35) =   12.69
   Model |  1325.4079     7  189.343986                Prob > F      =  0.0000
Residual | 522.342427    35  14.9240693                R-squared     =  0.7173
---------+------------------------------               Adj R-squared =  0.6608
   Total | 1847.75033    42  43.9940555                Root MSE      =  3.8632
------------------------------------------------------------------------------
pc_nestb |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 weenrgy |    34.1157   53.97086      0.632   0.531      -75.45098    143.6824
 welfunn |   1709.097   2807.302      0.609   0.547      -3990.029    7408.223
  wewage |   .0711538   .0357205      1.992   0.054      -.0013626    .1436702
weeduc90 |  -40.49956   28.38736     -1.427   0.163      -98.12896    17.12985
 weurate |  -17769.12   5820.465     -3.053   0.004      -29585.29   -5952.943
   wetax |  -5782.975   2868.037     -2.016   0.051       -11605.4    39.44982
     pca |  -1.315991   .6934624     -1.898   0.066      -2.723794    .0918129
   _cons |   .0478608   1.009746      0.047   0.962      -2.002033    2.097755
------------------------------------------------------------------------------

.
. ovtest

Ramsey RESET test using powers of the fitted values of pc_nestb
       Ho:  model has no omitted variables
                  F(3, 32) =      0.50
                  Prob > F =      0.6823

.
. clear

.
. *ooooooooooooooooooooooooooooooooooooooooooooooooo*
.
. *SIC 37*
.
. insheet state busact energy lfunion wages edsc valadd newcap urate tax numemp
>  prodhrs emprat pc_nemp c_numemp c_nestab pc_nestb educ90 emp93 estab93 wager
> ate using c:\data\sic37_8.csv
(note: variable names in file ignored)
(21 vars, 50 obs)
```

```
. *This program for SIC37 and the purpose is to create a complete STATA dataset
>  that has all possible variables, show summary statistics, generate a correla
> tion matrix, and generate regression results for a disequilibrium model*
.
. label data "Initial input dataset:  SIC 37"

. label var busact "bus. Activity (#estab ratio)"

. label var state "state"

. label var energy "price of energy"

. label var wages "Annual pay, prod. workers"

. label var lfunion "% priv. sector workers in unions"

. label var edsc "% population with some coll. ed."

. label var valadd "value-added per establishment"

. label var newcap "expend. per estab. on new plant & equip."

. label var urate "unemployment rate for state"

. label var tax "corp. income tax as a % of total tax coll."

. label var numemp "# of production workers"

. label var prodhrs "# of production hours"

. label var emprat "change in # of production workers"

. label var pc_nemp "%change in # of prod. workers"

. label var c_numemp "change in # of prod. workers"

. label var c_nestab "change in # of estab."

. label var pc_nestb "%change in # of estab."

. label var educ90 "% of pop. with a coll. degree, 1990"

. label var emp93 "# of employees, 1993"

. label var estab93 "# of establishments, 1993"

. label var wagerate "labor cost per production hr."

.
. mvdecode _all, mv(-999.99)
state:   string variable ignored
valadd:  4 missing values
newcap:  8 missing values
numemp:  4 missing values
prodhrs: 4 missing values
emprat:  4 missing values
pc_nemp: 5 missing values
c_numemp:5 missing values
emp93:   4 missing values
wagerate:9 missing values


.
```

```
. describe

Contains data
  obs:            50                          Initial input dataset:  SIC 37
 vars:            21
 size:         4,650 (99.3% of memory free)
-------------------------------------------------------------------------------
   1. state       str14  %14s                 state
   2. busact      float  %9.0g                bus. Activity (#estab ratio)
   3. energy      float  %9.0g                price of energy
   4. lfunion     float  %9.0g                % priv. sector workers in unions
   5. wages       long   %12.0g               Annual pay, prod. workers
   6. edsc        float  %9.0g                % population with some coll. ed.
   7. valadd      float  %9.0g                value-added per establishment
   8. newcap      float  %9.0g                expend. per estab. on new plant
   9. urate       float  %9.0g                unemployment rate for state
  10. tax         float  %9.0g                corp. income tax as a % of total
  11. numemp      float  %9.0g                # of production workers
  12. prodhrs     float  %9.0g                # of production hours
  13. emprat      float  %9.0g                change in # of prod. workers
  14. pc_nemp     float  %9.0g                %change in # of prod. workers
  15. c_numemp    float  %9.0g                change in # of prod. workers
  16. c_nestab    byte   %8.0g                change in # of estab.
  17. pc_nestb    float  %9.0g                %change in # of estab.
  18. educ90      float  %9.0g                % of pop. with a coll. deg. 1990
  19. emp93       float  %9.0g                # of employees, 1993
  20. estab93     int    %8.0g                # of establishments, 1993
  21. wagerate    float  %9.0g                labor cost per production hr.
-------------------------------------------------------------------------------
Sorted by:
     Note:  data has changed since last save
.
. summarize

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
   state |       0
  busact |      50    1.010479   .0776344   .8918919   1.296296
  energy |      50      8.5794   1.517579       5.55       11.9
 lfunion |      50      .09696   .0452494       .024       .193
   wages |      50    30350.72   4900.952      21648      41982
    edsc |      50      .60304   .0418223       .515       .697
  valadd |      46    10.73727   9.800639   2.08e-13   42.95484
  newcap |      42   -.0221757   6.050069  -37.03667   9.517742
   urate |      50      .05634   .0127993       .029       .089
     tax |      50    .0744672   .0964483   5.48e-16   .6202965
  numemp |      46    21.51087   29.44844   1.00e-11      153.3
 prodhrs |      46    44.15652   62.72944   1.00e-11      338.1
  emprat |      46    .9972907   .1752006  -1.00e-14      1.225
 pc_nemp |      45    1.945271   9.102891     -31.25       22.5
c_numemp |      45    .4133333   2.109567       -4.6        8.1
c_nestab |      50       -3.44   13.90956        -77         14
pc_nestb |      50    1.047931   7.763442  -10.81081   29.62963
  educ90 |      50      26.002   4.515664       16.1       34.4
   emp93 |      46    21.51087   29.44844   1.00e-11      153.3
 estab93 |      50      228.36   274.9213          5       1564
wagerate |      41    15.05974   2.965195       8.05    21.4895

.
. drop if valadd ==.
(4 observations deleted)
```

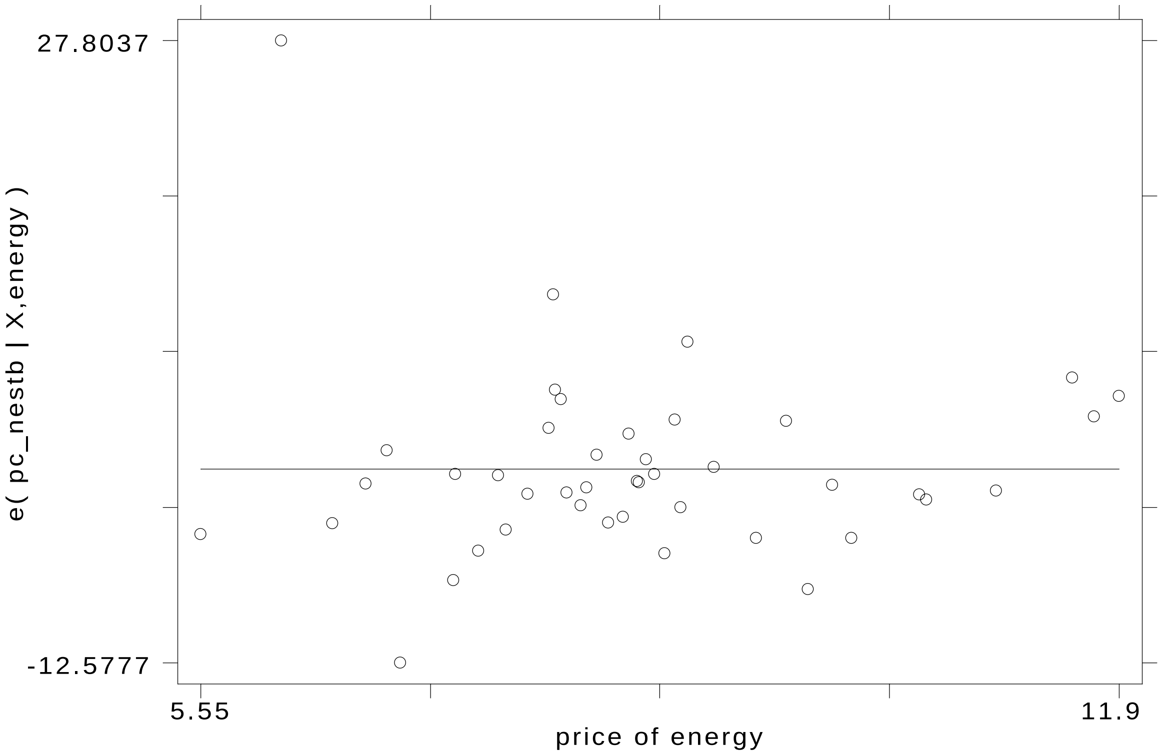
```
. drop if newcap ==.
(5 observations deleted)

. drop if tax ==.
(0 observations deleted)

. drop if prodhrs ==.
(0 observations deleted)

. drop if pc_nemp ==.
(1 observation deleted)

. drop if emprat ==.
(0 observations deleted)

. drop if pc_nestb ==.
(0 observations deleted)
.
. *drop observations for Hawaii and Alaska*
. drop if state=="Alaska"
(1 observation deleted)

. drop if state=="Hawaii"
(1 observation deleted)

. *drop variables that are not included in the model*
. drop busact

. drop edsc

. drop newcap

. drop numemp

. drop prodhrs

. drop emprat

. drop pc_nemp

. drop c_numemp

. drop c_nestab

. drop emp93

. drop wagerate
.
```

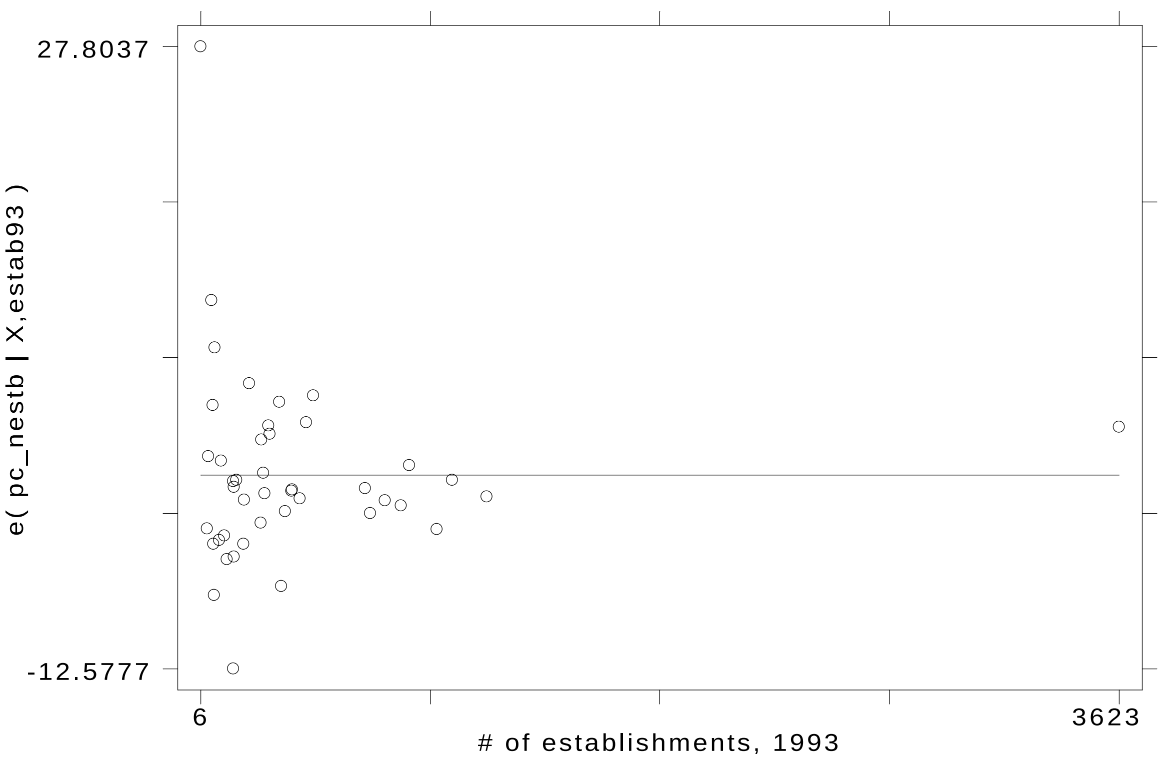
```
. *generate a correlation matrix*
. correlate energy wages lfunion valadd urate tax pc_nestb educ90 estab93
(obs=38)

        |   energy    wages  lfunion   valadd    urate      tax pc_nestb
--------+---------------------------------------------------------------
  energy|   1.0000
   wages|   0.2474   1.0000
 lfunion|  -0.0838   0.6154   1.0000
  valadd|  -0.0761   0.2654   0.2855   1.0000
   urate|   0.0617   0.2828   0.1247  -0.1030   1.0000
     tax|   0.1718   0.5208   0.5901   0.2685   0.1716   1.0000
pc_nestb|  -0.4191  -0.1827   0.1027  -0.1479   0.0454  -0.2017   1.0000
  educ90|   0.4966   0.4669  -0.0185  -0.2301  -0.1710   0.0450  -0.3648
 estab93|   0.1425   0.4638   0.1895   0.2258   0.3754   0.3567  -0.2501
        |   educ90  estab93
--------+------------------
  educ90|   1.0000
 estab93|   0.1346   1.0000

. *run tests for heteroscedasticity and graph residuals as another test*
. fit pc_nestb energy lfunion wage educ90 valadd urate tax estab93

  Source |       SS       df       MS                  Number of obs =      38
---------+------------------------------               F(  8,    29) =    1.85
   Model |  681.882439     8  85.2353048               Prob > F      =  0.1073
Residual |  1334.45573    29  46.0157149               R-squared     =  0.3382
---------+------------------------------               Adj R-squared =  0.1556
   Total |  2016.33817    37  54.4956263               Root MSE      =  6.7835

------------------------------------------------------------------------------
pc_nestb |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  energy |  -1.151329    .968373     -1.189   0.244      -3.131874    .8292163
 lfunion |   35.55891   39.88486      0.892   0.380      -46.01479    117.1326
   wages |   .0002805   .0005299      0.529   0.601      -.0008032    .0013642
  educ90 |  -.6157441   .4601401     -1.338   0.191      -1.556836     .325348
  valadd |  -.1954229   .1469114     -1.330   0.194      -.4958904    .1050445
   urate |  -6.243114   118.7807     -0.053   0.958      -249.1769    236.6907
     tax |  -68.95593      58.31     -1.183   0.247      -188.2133     50.3014
 estab93 |  -.0035363   .0047062     -0.751   0.458      -.0131617     .006089
   _cons |   21.38844   11.18572      1.912   0.066      -1.488918    44.26581
------------------------------------------------------------------------------

.
. rvpplot energy, border yline(0)

. gphprint, saving(c:\data\final\enrgy37.wmf)
.
. rvpplot wages, border yline(0)

. gphprint, saving(c:\data\final\wage37.wmf)
.
. rvpplot valadd, border yline(0)

. gphprint, saving(c:\data\final\vadd37.wmf)
.
. rvpplot estab93, border yline(0)

. gphprint, saving(c:\data\final\est93_37.wmf)

.
```

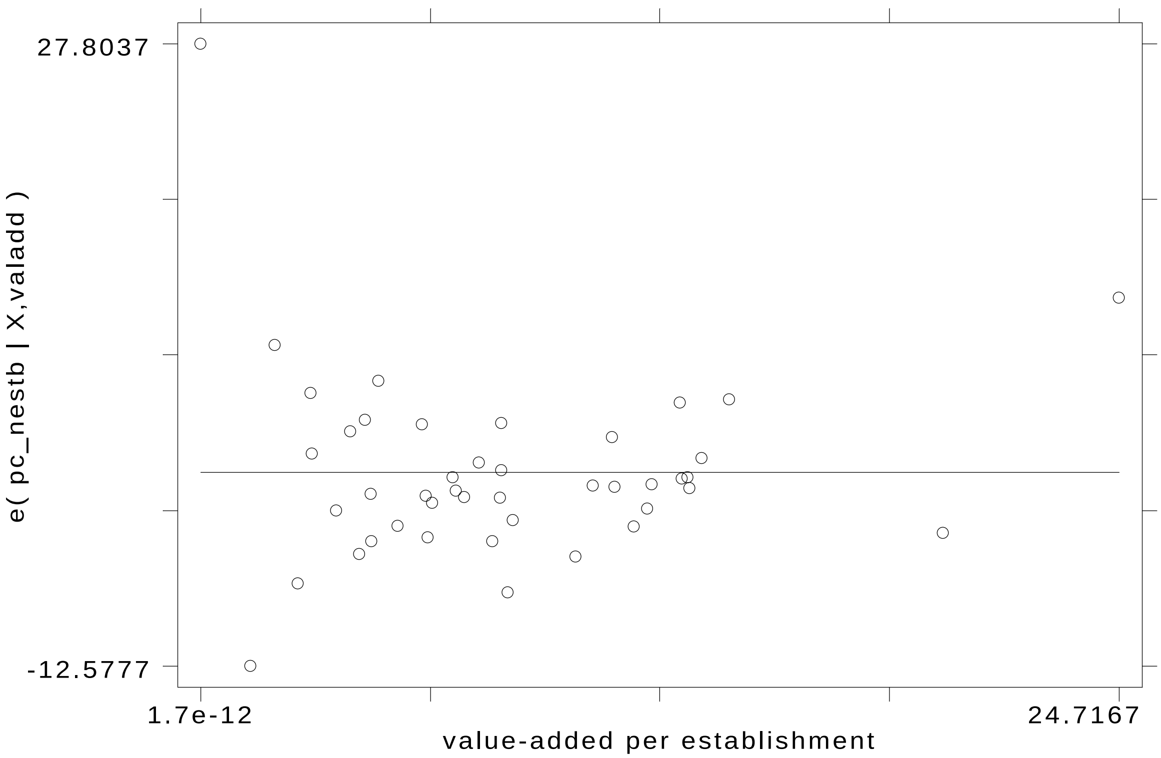
```
. hettest
Cook-Weisberg test for heteroscedasticity using fitted values of pc_nestb
     Ho: Constant variance
         chi2(1)      =     17.75
         Prob > chi2  =      0.0000
.
. white
White's test for Ho: homoscedasticity
         against Ha: unrestricted heteroscedasticity

    test statistic W =            38
    Pr(chi2(36) > W) =        0.3784


. *apply weights to perform weighted least squares*
. gen weenrgy=energy/estab93

. gen welfunn=lfunion/estab93

. gen wewage=wages/estab93

. gen weeduc90=educ90/estab93

. gen wevadd=valadd/estab93

. gen weurate=urate/estab93

. gen wetax=tax/estab93

. gen weestb93=estab93/estab93
.
. *Principal component analysis to reduce collinearity*
. factor valadd estab93, pc
(obs=38)

            (principal components; 2 components retained)
Component    Eigenvalue     Difference    Proportion    Cumulative
------------------------------------------------------------------
     1        1.22581        0.45163       0.6129         0.6129
     2        0.77419              .       0.3871         1.0000

            Eigenvectors
 Variable |      1          2
----------+---------------------
   valadd |   0.70711    0.70711
  estab93 |   0.70711   -0.70711

. score f1
            (based on unrotated principal components)
            (1 scoring not used)

            Scoring Coefficients
 Variable |      1
----------+----------
   valadd |   0.70711
  estab93 |   0.70711

. egen stdvadd=std(valadd)

. egen stdest93=std(estab93)

. gen pca= 0.70711*stdvadd + 0.70711*stdest93
```

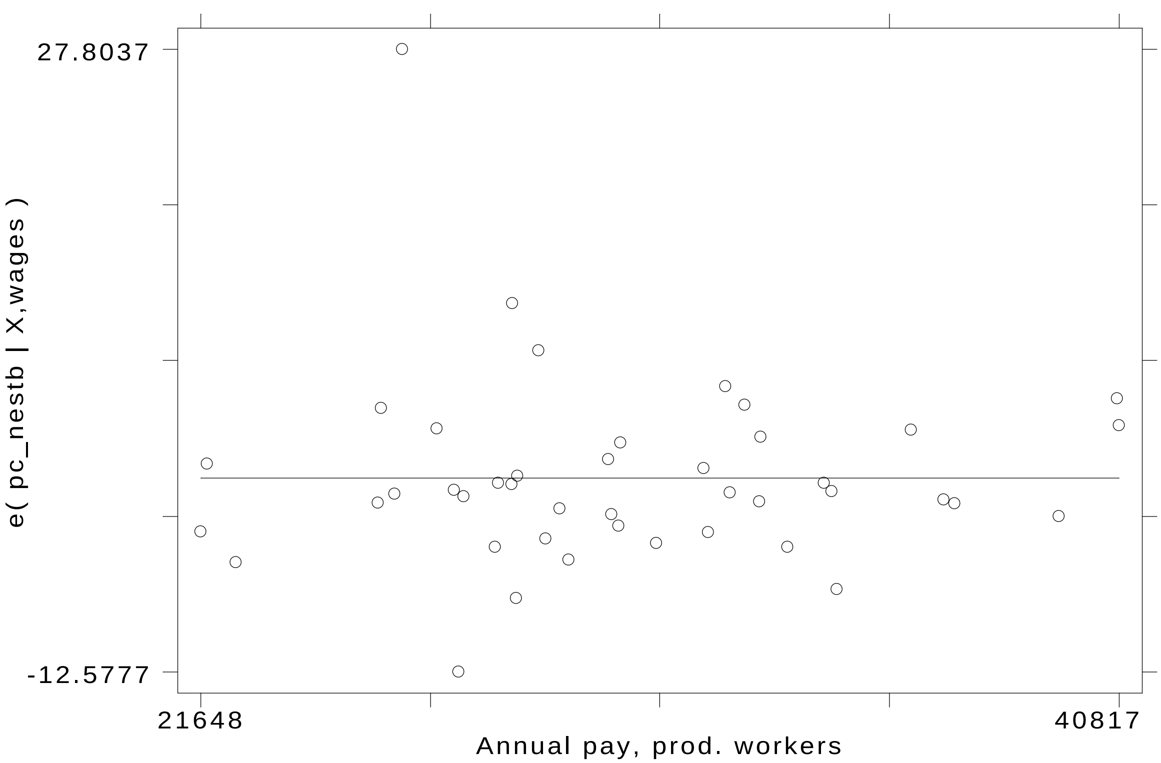
```
. label var pca "Combination of estab93 and valadd"

. *disequilibrium model depvar is % change in #estab*
. fit pc_nestb weenrgy welfunn wewage weeduc90 weurate wetax pca

  Source |       SS       df       MS               Number of obs =      38
---------+------------------------------            F(  7,    30) =   12.33
   Model |   1496.2879     7  213.755415            Prob > F      =  0.0000
Residual |   520.05027    30   17.335009            R-squared     =  0.7421
---------+------------------------------            Adj R-squared =  0.6819
   Total |  2016.33817    37  54.4956263            Root MSE      =  4.1635

------------------------------------------------------------------------------
pc_nestb |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 weenrgy |    -64.031   77.95456     -0.821   0.418      -223.2355    95.17345
 welfunn |  -1772.604   3618.998     -0.490   0.628      -9163.584    5618.376
  wewage |    .146195   .0297989      4.906   0.000       .0853375    .2070525
weeduc90 |  -93.96989   19.83822     -4.737   0.000      -134.4849   -53.45485
 weurate |  -13749.97   6028.283     -2.281   0.030      -26061.37   -1438.574
   wetax |  -2172.912   4074.772     -0.533   0.598      -10494.71    6148.883
     pca |  -1.134224   .7503088     -1.512   0.141      -2.666559    .3981113
   _cons |   -2.19776    1.00349     -2.190   0.036       -4.24716   -.1483606
------------------------------------------------------------------------------

.
. ovtest

Ramsey RESET test using powers of the fitted values of pc_nestb
       Ho:  model has no omitted variables
                 F(3, 27) =      1.50
                 Prob > F =      0.2365

.
. clear

.
. log close
```

Graphs to Examine Variance for Heteroscedasticity






27.8037
-12.5777
e(pc_nestb | X,valadd)
1.7e-12
24.7167
value-added per establishment


27.8037
-12.5777
e(pc_nestb | X,wages)
21648
40817
Annual pay, prod. workers

Vita

Sharon Hoke was born on March 25, 1971 in Sweet Springs, West Virginia. She earned a Bachelor of Arts degree in Economics and a Bachelor of Science degree in Business Administration from West Virginia State College. While an undergraduate student, Ms. Hoke worked as a research assistant for a forensic economist in Charleston, West Virginia. After graduation, she worked as a research assistant for the Economic Studies program at The Brookings Institution in Washington DC. Research provided by Ms. Hoke is acknowledged in two research publications: *Does the Bell Curve Ring True?* by William T. Dickens (forthcoming) and *Cracking the Code: Making Sense of the Corporate Alternative Minimum Tax* by Andrew B. Lyon. She is currently a Senior Economic Analyst for CEXEC, Inc.