New York Economic Review
Vol. 34 Fall 2003

CONTENTS

ARTICLES

Microeconomic Modeling And Analysis Of Commodity Chemical Production In A Simple Plant
Arthur S. Gow 4

A Count Panel Data Study Of The Schumpeterian Hypothesis
James J. Jozefowicz 21

Using Nlsy-Geocode Data To Determine The Effects Of Taxes And Minimum Age Laws On The Alcoholic Beverage Demand Of Young Adults
Mark Paul Gius 38

The Influence Of Illiquid Assets On Prices
Michael N. Baur, Omar Benkato, and Socorro M. Quintero 51

Rent Control And Its Reform In China
Anthony Yanxiang Gu 63

Referees 75

Final Program (55th Annual Convention – October 18-19, 2002) 76

EDITORIAL

The *New York Economic Review* is an annual journal, published in the Fall. The *Review* publishes theoretical and empirical articles, and also interpretive reviews of the literature. We also encourage short articles. The *Review's* policy is to have less than a three month turnaround time for reviewing articles for publication.

MANUSCRIPT GUIDELINES

1. Please submit three copies of a manuscript.

2. All manuscripts are to be typed, double spaced and proofread. Prepared on a IBM PC/compatible computer in Microsoft Word format, the computer disk should be submitted in addition to the three hard copies.

3. All charts and graphs *must* be reproduction quality (Microsoft Word or Excel).

4. Footnotes should appear at the end of the article under the heading of "Endnotes."

5. Citations in the text should include the author and year of publication, as found in the references, in brackets. For instance (Marshall, 1980).

6. A compilation of bibliographic entries should appear at the very end of the manuscript under the heading "References."

Manuscript submissions should be sent to the editor, William O'Dea.

Division of Economics and Business
State University of New York College at Oneonta
Oneonta, New York 13820

The NYER is cited in:

- Ulrich's International Periodicals Directory
- Cabell's Directory of Publishing Opportunities
  in Business and Economics

ISSN NO: 1090-5693

MICROECONOMIC MODELING AND ANALYSIS OF COMMODITY CHEMICAL PRODUCTION IN A SIMPLE PLANT

Arthur S. Gow*

ABSTRACT

A four-input (capital, labor, material and energy) production theory is applied to a representative chemical reaction occurring in a model plant, which captures the general features of large-scale chemical production processes. Engineering model and *bridge equations* (links between engineering and economic variables) are numerically solved to obtain feasible input combinations for a given production rate. Labor and material flows are fixed for a constant production rate, such that the capital-energy isoquant/isocost map gives the technically efficient region and (cost minimizing) optimum output expansion path for planned plants (*ex ante* case). Model plant total capital investment versus plant capacity is in excellent agreement with capital investment costs for actual polymerization plants. Finally, short- and long-run total, average and marginal cost curves exhibit theoretically correct behavior, and an example of static equilibrium analysis of the firm in the chemical product market is presented using short-run cost and postulated product demand and marginal revenue curves.

INTRODUCTION

Simple empirical economic models of industrial production processes are desirable from both applied and theoretical points of view. Models formulated in terms of a few meaningful variables (i.e., capital stock, and flows of labor, raw material, energy, and product output) facilitate analysis and assessment of alternative process technologies and yield useful general conclusions regarding process *trade-offs* and economies of scale. Such models provide useful tools with which to optimize processes, plan future plant additions, and to perform static equilibrium analyses of the firm in various market settings. A novel approach has recently been advanced (Gow and Gow, 2003; Gow 2003a, b; Gow 2002) to derive economic models from engineering models for single unit chemical production processes. This approach has yielded useful technical and economic information for the processes studied. This paper extends the same methodology to an entire chemical plant consisting of a chemical reactor and separation cascade to produce a common commodity chemical product.

BACKGROUND

A production function expresses the relationship(s) between the inputs of resources (i.e., capital, labor, material, etc.) to an industrial process and the output of a desired good. The general production

*Department of Chemistry and Chemical Engineering, University of New Haven, West Haven, CT 06516-1999, (203) 932-7173, agow@newhaven.edu
Winner of Outstanding Conference Paper Award at 55th Annual NYSEA Conference, Buffalo, NY, October 2002.

function is often of the form

$$q = f(x_1, x_2, \ldots, x_n) \qquad (1)$$

where q is the output flow (units/time) of the desired product, and x_n is the stock (units) or flow (units/time) of the nth input to the process. The particular functional form of equation 1 (which may actually consist of a set of equations) may be either: (1) chosen from a list of empirical models by statistical tests (econometric or *top-down* approach); or (2) derived analytically or numerically from the detailed production technology (engineering or *bottom-up* approach). The most basic form of empirical production function is the Cobb-Douglas two-factor model (Cobb and Douglas, 1928)

$$q = A\, k^{\alpha}\, l^{\beta} \qquad (2)$$

where k and l are the capital and labor inputs respectively to the production process, plant, firm, industry, sector or economy, α and β are the capital and labor output elasticities respectively, and A is the "time-dependent" index of factor productivity. A simple (unitary elasticity of substitution) form of the Cobb-Douglas model is obtained if $\beta=1-\alpha$ is specified.

Variations of the Cobb-Douglas model and several more-flexible functional forms (Zellner and Ryu, 1998; Diewert and Wales, 1995; Bairam, 1994; Pollak and Wales, 1987; Arrow et. al., 1961) have been successfully used to correlate industrial or sectoral level data for various product groups including consumer durable goods, food, clothing, paper, agricultural products and chemicals. However, generalizations regarding returns to scale and substitution elasticities reached from the results of *time-series* and *cross-sectional* studies (Giannakas et. al., 2000; Hsieh, 1995; Sato, 1975) are statistically supported by industrial level data without documented support at the process level. Furthermore, while application of generic economic models to industries and sectors is somewhat useful, difficulties are encountered regarding the division and units of input factors of production, the method of aggregation from firm-to-industry level of production, and assumptions about returns to scale and technological progress.

A useful method for obtaining meaningful process or plant production functions is the engineering (*bottom-up*) approach (Gow and Gow, 2003; Gow 2003a, b; Gow 2002; Barsan and Ignat, 2001; Sav, 1984; Chenery, 1949). Here, a deterministic model is derived from scientific laws, process constitutive relationships, and aggregation rules (*bridge equations* which provide the link between engineering and economic variables). Input unit cost data are incorporated to obtain a complete economic model (production and cost functions) from which the optimum combination of inputs may be determined. This paper applies the above methodology to a typical chemical production process occurring in a simple plant. The reaction and simple plant studied have features common to a wide range of intermediate (commodity) chemical production processes, and can yield useful general microeconomic information about large-scale chemical production.

CHEMICAL PROCESS FUNDAMENTALS AND PLANT MODEL

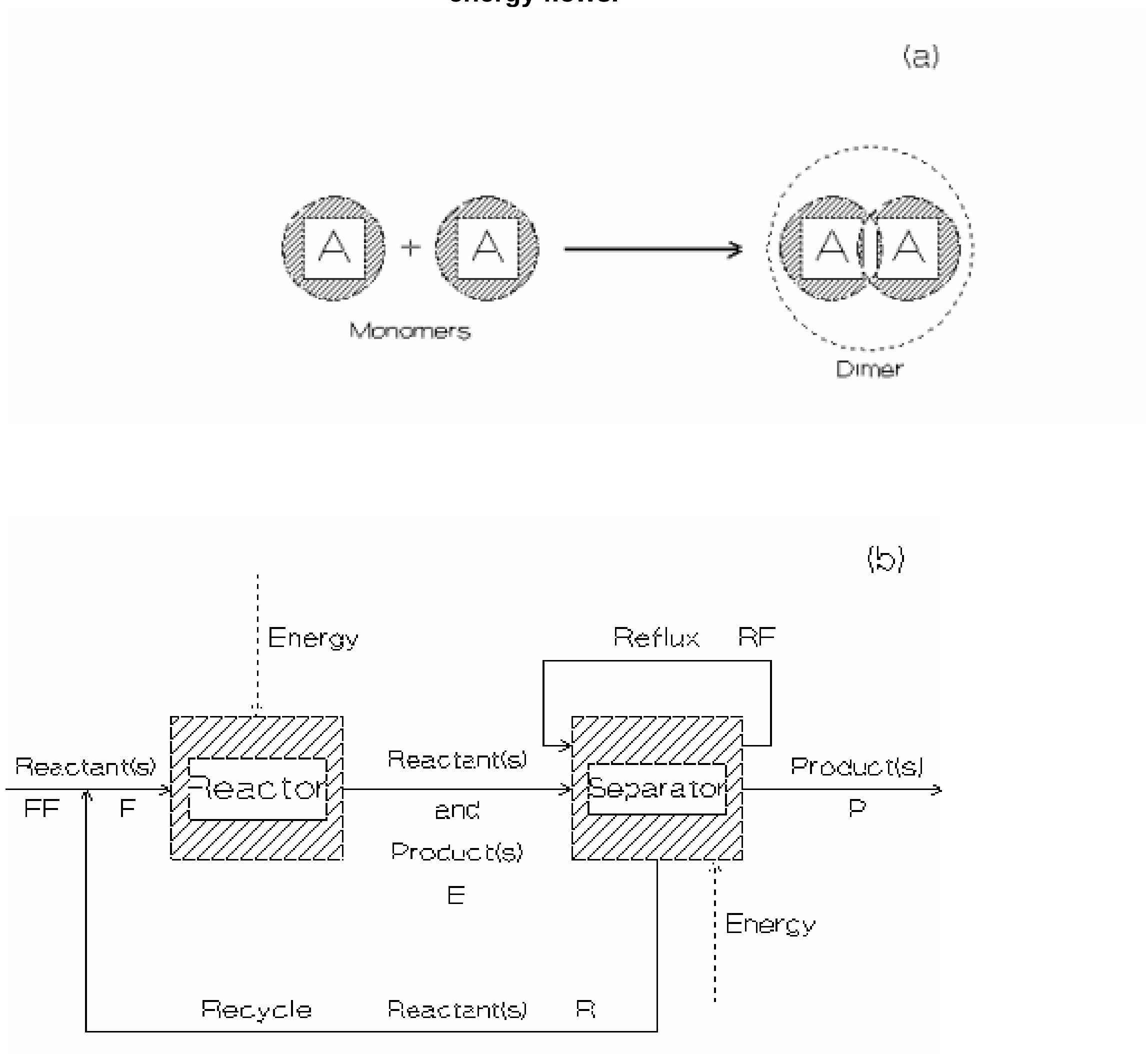
A broad spectrum of consumer products and intermediates (used in the production of other finished goods) are obtained from chemical production processes. These products, which account for nearly ten percent of the U.S. gross national product, include bulk commodity chemicals, pharmaceuticals, electronic materials, polymers and various other substances (Luyben and Wenzel, 1987). Chemical substances (elements and compounds) undergo reactions to produce new substances. *Chemists* are involved in the discovery and explanations (mechanisms) of chemical reactions, whereas *Chemical Engineers* are primarily engaged in the development of large-scale commercial processes to safely, ecologically and economically produce chemical products.

Industrial chemical production generally consists of two types of processing steps: (1) *reactions* - chemical reactants (reagents or raw materials) are converted to products; and (2) *separations* - products and unreacted raw materials are recovered in nearly pure form. Reactions are frequently incomplete (i.e., a fraction of raw materials remains and/or multiple products are produced); hence, separations are almost always required. Products are isolated, and unreacted reagents are recovered for further processing (reaction). Reactions take place in vessels called *reactors*, and separations occur in vessels called *separators* (or *separation cascades*). All plants consist of various types and combinations of reactors and separators. A very simple reaction and separation network comprised of one reaction unit and one separation unit is used here to illustrate the novel microeconomic modeling approach and the process and economic *trade-offs* typically encountered in industrial chemical production practice.

Model Reaction - Organic (carbon-based) intermediates are a large and vital segment of the chemical process industries (CPI). These substances are the building blocks for a wide range of end products including polymers, petrochemicals, and pharmaceuticals. Combination of small identical organic molecules to produce a larger intermediate molecule is a common class of industrial reactions. The simplest case of combination is *dimerization*, in which two molecules of reactant are combined to produce a product with twice the molecular mass of the reactant (see Figure 1a).

Model Plant - The model plant *process flow diagram* (PFD), which is shown in Figure 1b, captures the major technical features of many commercial chemical production processes. The model *dimerization* plant consists of a chemical reactor, in which raw material A is converted to product B, and a separation cascade (series of identical units called *stages*) which recycles the unreacted *monomer* (i.e., the unreacted raw material is recovered in the separator and sent back to the reactor for further processing). Reactant A enters the plant at a flow rate FF tons/day, mixes with recycle A at a flow rate R tons/day, and the combined stream enters the reactor at F tons/day (F=FF+R). However, only a fraction of the combined stream, f x F, is converted to the desired product B. Thus, the reactor effluent stream (stream E leaving the reactor) consists of a mixture of desired product B and unreacted reactant A. The separation cascade also has its own internal recycle stream, RF tons/day called *reflux*, the reason for which is discussed below.

Figure 1. (a) Dimerization reaction, and (b) simple plant showing components and material and energy flows.



There are a few key relationships, which describe the process behavior of the plant. First, total material is conserved. That is, there is no build-up or depletion of material within the plant over time. Thus, if 1,000 tons/day of monomer A enter the plant in stream FF, then 1,000 tons/day of dimer B are produced within the plant and must leave the plant in stream P. This is called the *law of conservation of mass*, which states that matter is neither created nor destroyed, but may be converted from one form to another (i.e., new substances may be formed in a chemical reaction). Another important relationship pertains to the chemical reactor, which facilitates intimate contact of reactant molecules required for reaction to occur. The longer that reactant molecules spend in the reactor, the higher the fraction of them that will be converted to product molecules. This is accomplished by either using a larger reactor or a lower feed flow rate, F, for a set production rate, P, of the dimer product. Finally, some general principles apply to the separation cascade. Here, higher flow rate through the cascade (throughput) results in more efficient (faster) separation and hence, requires a smaller separation cascade (i.e., smaller number of repetitive units or *stages*); however, a higher flow rate also requires larger diameter stages to accommodate the increased flow rate. The internal flow rate (cascade throughput) is adjusted by increasing or decreasing the flow of stream RF. Note that increasing the flow rate RF has the effect of reducing the number of stages (more efficient separation) while increasing the stage diameter due to the increased throughput.

In summary, there is a key process trade-off for both the reactor and the separator. First, a smaller reactor may be used for a given dimer production rate, P (tons/day), which results in more A per day being recovered and recycled back to the reactor (higher R); or a larger reactor may be used, which results in lower R. The internal reflux of the separation cascade, RF, may be made large (low number of stages/large diameter) or small (high number of stages/low diameter) for a specified production rate, P. The essential long-run optimization problem is *what are the optimum sizes of the reactor and separation unit for a given production rate of B?*

ECONOMIC MODEL

The simplest production function for a chemical process is of the general form

$$q = f(k_T, l, m, h_T) \qquad (3)$$

where q is the product output flow (mass/time), k_T is the capital stock (i.e., reaction and separation units and plant infrastructure), l is the process labor flow (workerhours/time), m is the raw material input flow to the process (mass/time), and h_T is the total energy flow to the process (energy/time). Chemical process model equations may be combined with aggregation rules (*bridge equations*), which provide the link between engineering and economic variables. However, a fundamental problem encountered in the econometric (*top-down*) approach to production is how to aggregate heterogeneous capital equipment items.

Fortunately, engineering equipment cost correlations provide a consistent means for obtaining an accurate measure of the chemical plant capital stock using an engineering (*bottom-up*) approach. An equipment cost correlation expresses the purchased cost of an item as a function of the size of the item according to

$$C_{pur} = a\,(size)^b \qquad (4)$$

where C_{pur} is the purchased cost of the equipment item ($) at a specific point in time (base month/year), and the constants a and b are for a particular type of equipment constructed of materials to withstand specified extremes of operating conditions (temperature, pressure, corrosiveness of reagents, etc.). The installed cost of the equipment item in the base year, C_{ins} ($), is given by

$$C_{ins} = F_i\, C_{pur} \qquad (5)$$

where F_i accounts for the labor and materials involved in the installation of the item. Purchased equipment cost correlations are available from published sources[1] (e.g., Peters and Timmerhaus, 1991).

Capital stock should be expressed in physical units (capital units) rather than financial units (dollars) since capital stock is physical material. A method was proposed (Gow and Gow, 2003; Gow, 2003a, b; Gow 2002) for expressing heterogeneous capital equipment items in terms of a common "physical" unit, which allows easy aggregation of capital. This approach, used here, is now briefly reviewed. First, all conditions and equipment sizes are determined from the process model equations[2]. Next, the dollar value (in a base month/year) of an installed equipment item of specified size is related to the base month/year dollar value of an installed reference equipment item of specified size. A capital unit is defined as an arbitrary physical amount of a reference capital good, where it is assumed that the nominal prices of all capital goods change by the same proportion over time, such that the total capital is the same regardless of base month/year used.

The chosen (arbitrary) capital standard is that 1 capital unit is equivalent to 10 feet of 6-inch schedule 40 carbon steel welded pipe, which cost $122 in January 1990 (base month/year) (Peters and Timmerhaus, 1991). The January 1990 value of an installed capital equipment item, C_{ins} ($), is determined and converted to capital units using the conversion factor 1 capital unit equals $122 (Jan. 1990) or

$$k_i = 0.0082\ C_{ins} \qquad (6)$$

where k_i is the units of capital stock (capital units) for equipment item i. The total process capital, k_T, is then obtained from

$$k_T = 3.37\ \Sigma_i\ k_i \qquad (7)$$

where the multiplier 3.37 accounts for the necessary plant infrastructure (i.e., land, buildings and grounds, electrical and plumbing systems, controls, office and distribution facilities, etc.). The breakdown of plant infrastructure items for a typical chemical plant, which leads to the multiplier value of 3.37, is presented in Gow (2003b).

Plant labor requirement is also obtained from published correlations in the engineering literature. Published process labor studies (Guthrie, 1970; O'Connell, 1962; Haines, 1957; Isard and Schooler, 1955) propose that process labor for a chemical manufacturing operation depends on: (1) the total mass throughput, *m* (tons/day), (2) the *subdivision* or number of repetitive or distinct major process units, and (3) the degree of process *automation*. The model used here is built on the findings of these earlier studies and is given by

$$l = a'\ m^{0.25} \qquad (8)$$

where *l* is the process labor requirement (workerhours/day) and *m* is the process mass throughput (tons/day), and a' is a constant related to the degree of process subdivision (a' is higher for a larger number of process units) and the degree of process automation (a' is lower for a more highly automated process). Here, a'=11.0 for a moderately automated plant consisting of a few unique process units.

The total material flow to the plant (mass throughput) is simply given by the mass production rate of dimer B (i.e., since mass flow entering the plant equals mass flow leaving the plant)

$$m = q = FF = P \qquad (9)$$

Finally, the required energy input, h_T, for the plant includes electrical energy supplied to the reactor agitator and liquid pumps required to move fluids through the plant, and heat supplied to the separation cascade[3]. Furthermore, the fluid motion energy supplied to the agitator, pumps, and separator is "sunk" (lost) because this energy is not recovered in any useful form. Thus, the energy requirement for the model commodity chemical plant is

$$h_T = \Sigma_i h_i \qquad (10)$$

where *h* is in units of horsepower (HP). Application of Equation 10 requires constitutive equations, which express the terms h_i as functions of process variables[4].

RESULTS

Simulation Specifications and Methods – Long-run (planned plant) simulations were made for a large number of dimerization rates covering a wide range of production; however, detailed results are presented for three production rates (i.e., *q*=1,500, 2,000, and 2,500 tons/day), which cover the typically observed output range for real plants. Furthermore, one short-run simulation was made for the optimum plant (plant size of k_T=77,200 capital units) designed for a capacity of 2,000 tons/day of product. The *process model* is solved to determine the reactor size (volume) required to achieve a particular fraction, f, of the monomer A in the reactor feed stream F converted to dimer B product for a given pure A plant feed, FF (tons/day), and the separator size (i.e., number of stages and cascade diameter) required to achieve the desired B production rate, P (tons/day) (long-run or *ex ante* case); or to determine the production rate, P (tons/day), for specified reactor and separation cascade sizes and varying fresh A feed flow rate, FF (tons/day) (short-run or *ex post* case). Data produced in the simulation runs are presented in various useful forms and the following analyses are performed: (1) technical analysis - examination of long-run (*ex ante*) isoquant maps; (2) economic analysis - determination of the optimum output expansion path (minimum cost path across an isoquant map) and study of short- and long-run total, average and marginal cost curves; and (3) static equilibrium analysis - investigation of plant profitability in a typical short-run market setting.

Technical Analysis - Long-run technical analysis investigates the input factor relationships for planned plants and identifies the region of technically efficient production (i.e., decreases in at least one production factor and increases in other factors to maintain constant production rate). The optimum point in a technically efficient region is determined from an economic analysis of the process using unit factor costs. Technical analysis for the four-factor chemical process/plant model considers relationships (isoquant maps) between various pairs of the four input factors (i.e., capital stock, k_T, labor, l, material, m, and energy, h_T, flows). A fortuitous consequence of the simple chemical plant model is that labor and raw material inputs are constant for any constant production rate of B considered[5]. Hence, the capital versus energy isoquant map captures all of the essential features of the simple dimerization production process and yields valuable new information about the economics of chemical production.

The capital versus energy isoquant map for the model dimerization plant was constructed using data from a series of constant output sub-processes for which the fraction of monomer A reacted, f, was constant (e.g., f=0.30, 0.40, 0.50, etc.) and the reflux flow of monomer A in the separator, RF, was varied to obtain the number of separation stages, N_{act}, required to recover nearly pure A recycle and pure B product streams. Each set of data is for constant reactor size and energy input and variable separator size and energy input. Figure 2a illustrates the construction of a capital-energy isoquant from a series of constant conversion capital-energy *isoquantlets* for a dimer B production rate of q=1,500 tons/day[6]. *Isoquantlets*, shown for f=0.830 and 0.943, exhibit a small range of capital-energy substitution possibilities, and the envelope of *isoquantlets* defines the capital-energy isoquant for the given production rate, which in this case is q=1,500 tons/day.

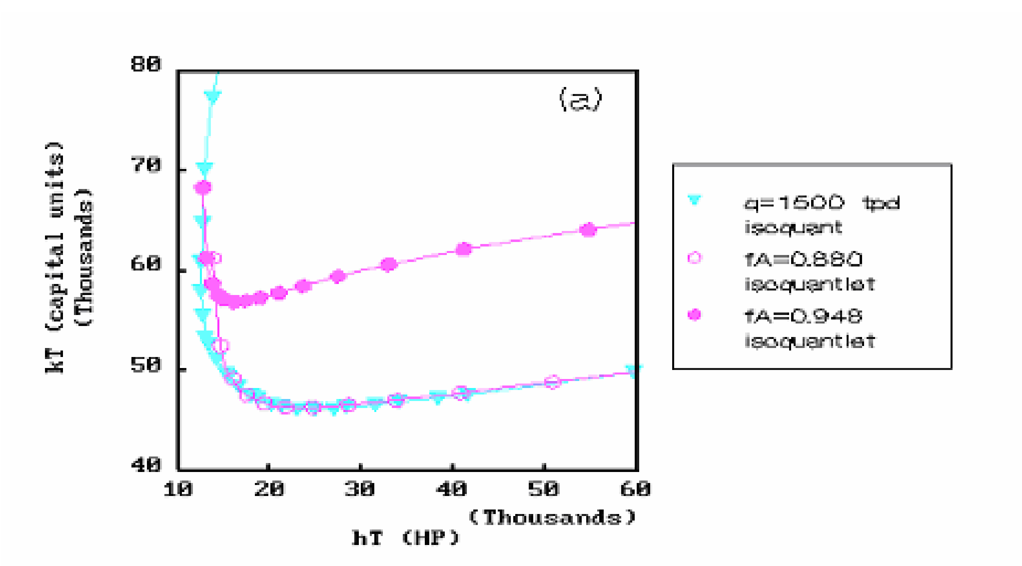
The *ex ante* capital versus energy isoquant map for the simple dimerization plant, including isoquants for q=1,500, 2,000, and 2,500 tons/day of dimer B output, is shown in Figures 2b and c. First, the isoquants have the theoretically correct convex-to-the-origin shape. Furthermore, the isoquant map exhibits ridge curves, which are the loci of points of zero or infinite slope [i.e., $(\partial k_T/\partial h_T)_q$=0 or ∞] on adjacent isoquants. The lower and upper ridge curves, which define the technically efficient region of production, have been added to the capital-energy isoquant map in Figure 2b. It is significant that that capital and energy are substitutes to a limited extent. Moreover, the isoquant map clearly shows that simple Cobb-Douglas or CES models are incapable of describing input factor relationships for the simple chemical plant studied here. These models assume a defined returns to scale behavior and monotonically declining marginal rate of technical substitution along an isoquant, which are not observed in the isoquant maps of Figures 2b and c. Similar results have been obtained in other studies of a variety of chemical production processes (Gow and Gow, 2003; Gow, 2003a, b; Gow, 2002), which especially support using the *engineering approach* advanced in this paper for any type of chemical process at the single unit or plant level. Finally, it is difficult to determine returns to scale behavior from the simulation results because it is not possible to simultaneously vary all inputs by the same proportion.

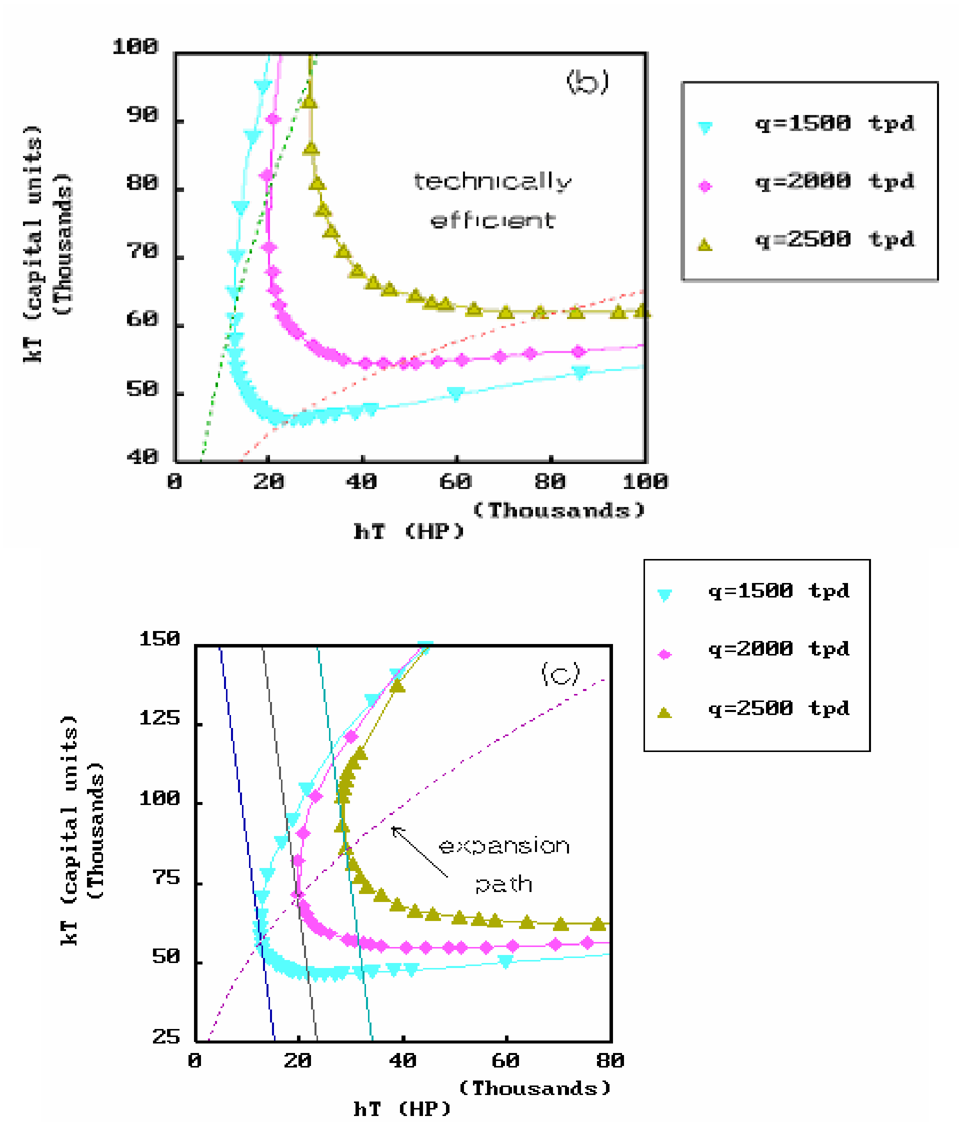
By far, the most useful feature of the capital-energy isoquant map is that it can be used to optimize the combination of inputs for planned plants at any level of production. A series of (parallel) capital-energy isocost (dashed) lines have been plotted on the capital-energy isoquant map in Figure 2c.

The unit costs of productive inputs (capital, labor, material and energy) are given the economic analysis section of this paper. The endpoints of the isocost lines are capital and energy equivalents, which may be purchased with a fixed budget after accounting for the constant required labor and raw material input costs at the given dimer B production rate. There are an infinite number of isocost lines of the same slope (each for a different total budget), which may be plotted. However, only the isocost lines that are tangent to the capital-energy isoquants for dimer B production rates *q*=1,500, 2,000, and 2,500 tons/day have been plotted in Figure 2c. The points of tangency between isocosts and isoquants give the economic optimum (cost minimizing) combination of inputs at each production rate. The optimum output expansion path (dashed curve in Figure 2c, which is the locus of least cost points) suggests that plant size (capital stock) should be increased at a lesser rate than energy consumption to increase production, based on current capital and energy unit costs. A change in either capital or energy unit cost would shift the location of the optimum output expansion path.

The observed capital-energy relationship in Figures 2b and c may be compared with findings from other published studies. Curiously, the relationship between capital and energy inputs in production is a widely studied unresolved issue in the economics literature. Berndt and Woods' (1979) seminal study suggests that capital and energy are complements (i.e., *kinked* isoquants with two positively sloped branches and zero substitution possibilities). The results of some studies support Berndt and Woods' claim. Mahmud and Chishti (1990) concluded that capital and energy are complements in Pakistani manufacturing, and Caloghirou et. al. (1997) found that electrical energy and capital are long-run production complements in the Greek economy.

Figure 2. (a) Construction of isoquant from *isoquantlets*, and isoquant map showing (b) technically efficient region, and (c) output expansion path.





However, other studies present evidence, which supports capital-energy substitutability. Rushdi's (1991) investigation of the South Australian electrical supply industry, Chang's (1994) study of Taiwanese manufacturing, and Hisnanick and Kyers' (1995) and Ziari and Azzams' (1999) studies of U.S. manufacturing conclude that capital and energy are substitutes. Finally, one paper proposes a capital-energy duality depending on the time frame (short- or long-run) considered. Apostolakis (1990) found that cross-section (long-run) studies support substitutability between capital and energy, whereas time series (short-run) studies support the capital-energy complementary hypothesis. An important result of this study is that capital and energy are clearly shown to be weak-to-moderate substitutes in the long-run. It is significant to note that all of the interesting input behavior in chemical production processes is for capital and energy. Capital and energy may both be varied (along isoquants) in the long-run, whereas energy is the only important variable input in the short-run.

Finally, it is useful to compare key benchmarks such as capital investment cost of model dimerization and real polymerization plants[7]. Figure 3 shows profiles of the total capital investment versus plant capacity for model dimerization and commercial polymerization plants from which excellent agreement is observed. Thus, it appears that the engineering approach, presented in this paper, can be successfully used to estimate the capital investment requirements of new chemical plants.

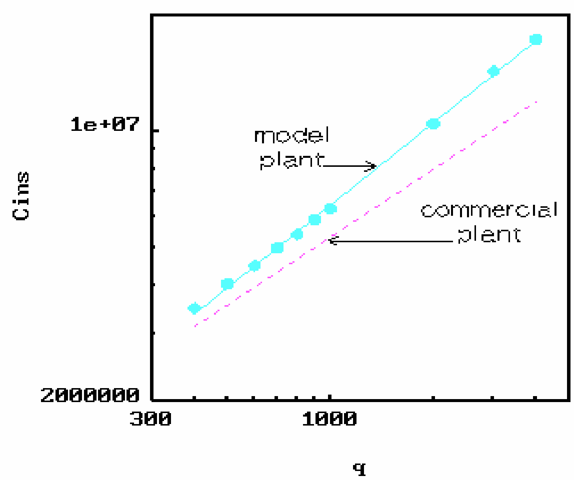
Economic Analysis – Economic analysis *optimizes* a process within the technically efficient region of production. Isocost lines are superimposed on the isoquant map to permit graphical determination of the optimum output expansion path in the two-dimensional case (two variable process inputs) as was previously illustrated for the simple dimerization plant. However, additional insight is gained by examining the behavior of both long- and short-run cost curves.

The total cost, TC ($/day), is given by

$$TC = C_{kT}\, k_T + C_l\, l + C_m\, m + C_{hT}\, h_T \qquad (11)$$

where C_{kT} ($/capital unit day), C_l ($/workerhour), C_m ($/ton) and C_{hT} ($/HP day) are the unit cost of capital,

Figure 3. Capital investment cost versus capacity for model and commercial plants.



labor, material and energy respectively. The values of these unit cost components are determined from practical considerations. The unit cost of capital, C_{kT}, is obtained by dividing the base month/year dollar value of the defined capital unit (1 capital unit equals $122, Jan.1990) by the average useful life of chemical process equipment, which is about ten years (3,600 days) (Peters and Timmerhaus, 1991). Moreover, the capital unit cost is corrected from 1990 to 2000 dollars using a common cost index, which accounts for inflation (Peters and Timmerhaus, 1991). Finally, the capital cost is multiplied by a factor of 1.10, which reflects the opportunity cost of investing in the next most attractive venture (assumed to yield an annual 10% return for a period of ten years). Thus, C_{kT}=0.0426 $/capital unit day represents the daily depreciation rate of total plant capital. The term $C_{kT}k_T$ in Equation 11 is the average daily cost of maintenance and improvements necessary to keep the value of plant capital equipment constant at k_T. 3,600$C_{kT}k_T$ is the dollar amount required to replace the existing plant when it is completely depreciated. The unit cost of labor, C_l=22.50 $/workerhour, is the inflation corrected average of wage rates for semi-skilled plant operating and supervisory labor (Peters and Timmerhaus, 1991). C_m, is unit cost of reactant (monomer) A, which is taken to be $500/ton in January 2000[8]. Finally, the unit energy cost, C_{hT}=0.80 $/HP day, is the inflation corrected average electrical cost per kilowatt-hour for purchased and self-produced electricity (Peters and Timmerhaus, 1991).

Equation 11 may be applied to either the long-run (*ex ante*/planned plant) or short-run (*ex post*/fixed plant) case of production in the simple chemical plant. The unit factor costs, C_{kT}, C_l, C_m and C_{hT}, are used in Equation 11 to compute total cost, TC ($/day) along isoquants (fixed dimer B production rate, *q*) for technically feasible input factor combinations (long-run case). Conversely, the short-run case involves determining the total cost, TC ($/day), for the given plant size and feasible combinations of raw material, energy and labor input flows to achieve a given dimer B production rate. Here, the first term on the right-hand side of Equation 11 is constant due to the fixed plant size (i.e., reactor volume and number of separator stages). The long- or short-run average cost, LAC or SAC ($/ton) respectively, is computed from

$$LAC(SAC) = (TC)_{min}/q \qquad (12)$$

where $(TC)_{min}$ is the minimum (optimum) total cost determined by considering all feasible input combinations at a particular production rate for either the long- or short-run case.

Results of the application of Equations 11 and 12 to the general long-run case and one short-run case of production in the simple chemical plant are presented in Figure 4a. Here, short- and long-run average cost, SAC and LAC ($/ton) respectively, are plotted versus dimer B output (tons/day). Note that the short-run average cost versus log *q* for the fixed plant size (k_T=77,200 capital units for *q*=2,000 tons/day) is a narrow U-shaped curve, which is within and tangent to the wider U-shaped LAC curve.

The long- or short-run marginal cost, LMC or SMC ($/ton) respectively, is the output derivative of the total cost

$$LMC(SMC) = \partial(TC)_{min}/\partial q \qquad (13)$$

The results of Equations 12 and 13 for LAC and LMC versus production rate, *q*, are plotted in Figure 4b, from which it is clear that the LMC intersects the minimum in the LAC from below as expected. The LAC is flat at just over 505 $/ton for an output range extending from around *q*=700 to 2,200 tons/day of dimer B output, which indicates a considerable region of constant returns to scale for the simple plant.

Finally, short-run cost curves including the SAC and SMC ($/ton) as computed from Equations 12 and 13 respectively, and the short-run average variable cost

$$AVC = (STC - C_{kT}k_T)/q \qquad (14)$$

(\$/ton), which gives the plant shutdown conditions, are computed for a plant of 2,000 tons/day capacity (i.e., a plant of k_T=77,200 capital units in size) and are presented in Figure 4c. Short-run average cost declines sharply and levels off at around SAC=505 \$/ton over the production range from approximately q=1,500 to 2,300 tons/day. SMC is less than SAC at low output, is nearly constant in the range 503 to 505 \$/ton from approximately q=1,500 to 2,300 tons/day, and intersects SAC at q=2000 tons/day from below, which is theoretically consistent. Furthermore, AVC falls sharply to values just one to two \$/ton less than SAC for production rates ranging from around q=1,500 to 2,300 tons/day, and then nearly coincides with SAC at production rates above q=2,300 tons/day. It is noteworthy that the long- and short-run total cost curves, LTC and STC (\$/day) respectively (not shown) are economically consistent. Both profiles have the theoretically correct shape and are tangent at the same output as for the average cost curves in Figure 4a. Furthermore, the zero-output total cost of a plant with a capacity of 2,000 tons/day is 3,290 \$/day, which is the fixed cost of keeping the plant open with no production.

Static Equilibrium Analysis - It is useful to conduct simple equilibrium analyses of a typical industry firm in a typical market setting using derived cost curves and estimated product demand and marginal revenue curves. The analysis presented here is for a firm which holds a portfolio of equally sized plants of capacity q=2,000 tons/day (k_T=77,200 capital units). The short-run cost curves from the previous section (Figure 4c) are re-plotted along with estimated short-run demand (average revenue or AR) and marginal revenue (MR) curves in Figure 5. The equation of the linear demand curve is

$$P = a + bq \qquad (15)$$

where the slope (b=-0.1366 \$ day/ton^2) and the intercept (a=1,011 \$/ton) were determined by considering a plant with a typical 50% markup over average cost of production at the equilibrium output. Furthermore, it follows that the marginal revenue curve, MR (\$/ton), is given by the relation

$$MR = a + 2bq \qquad (16)$$

where the slope is twice the negative magnitude of the demand curve slope.

The intersection of the marginal cost (MC) and marginal revenue (MR) curves gives the optimum (profit maximizing) rate of output for the fixed plant, which is q=1,850 tons/day (92.5% capacity utilization of a q=2,000 tons/day capacity plant). The plant profit (approximately 462,500 \$/day) is given by the difference between the demand curve, AR (\$/ton) and short-run average cost curve, SAC (\$/ton), multiplied by the output at which MC intersects MR (q=1,850 tons/day).

Figure 4. (a) Construction of long-run average cost curve, (b) long-run cost curves, and (c) short-run cost curves for a plant capacity of *q*=2000 tons/day.

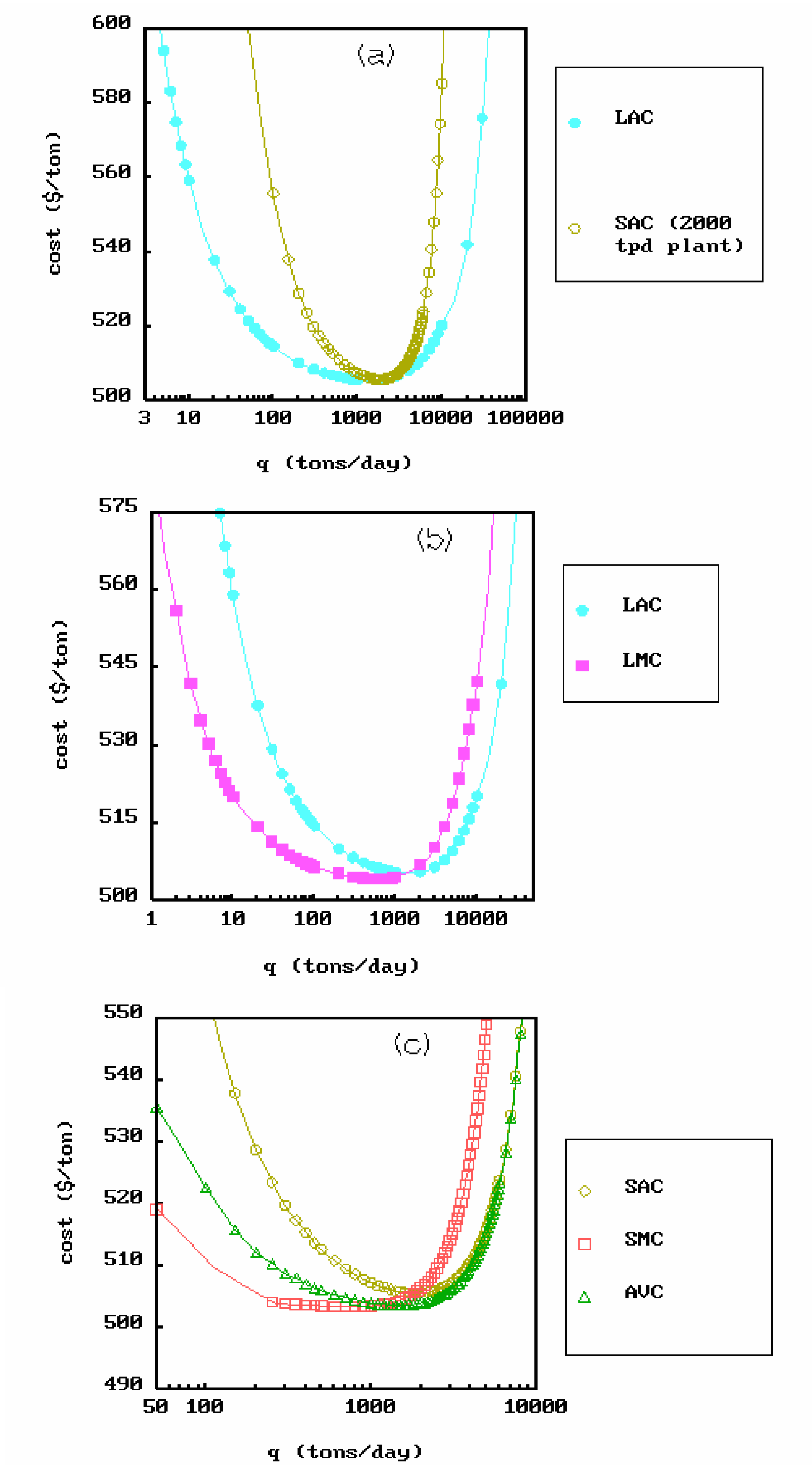
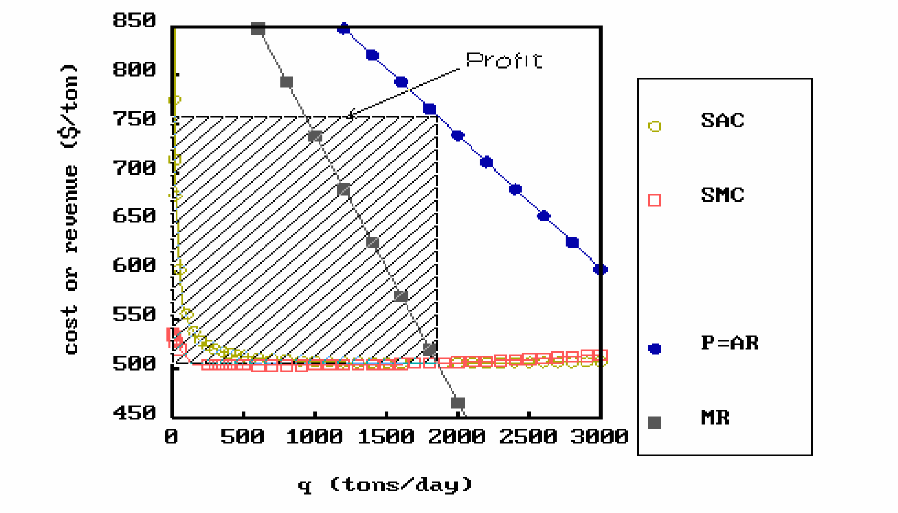


Figure 5. Static equilibrium analysis of a firm holding equally-sized plants of capacity *q*=2000 tons/day in an oligopolistic market setting.



CONCLUSIONS

A novel microeconomic framework was applied to a representative simple chemical plant for the production of a common commodity chemical substance (i.e., an intermediate, which is a raw material in the production of other chemical substances and finished goods). A four factor (capital, labor, material, and energy) production model was derived for a simple dimerization plant consisting of a chemical reactor and a multistage separation column. Labor and material inputs are constant for a fixed production rate, such that the capital-energy isoquant map gives all relevant input behavior for long-run (planned) plants. Simulation runs were made for both short- and long-run cases, and fundamental technical, economic and static equilibrium analyses of the results were presented. Useful results include identification of the technically efficient region for capital and energy inputs, the output expansion path and capital investment cost versus capacity for planned plants, the cost behavior of both fixed and planned plants, and the profitability of a firm comprised of equally sized plants in a typical market setting.

The observations made regarding technical, economic and static equilibrium analysis suggest some general conclusions for the economic behavior of chemical process and plant models. First, apparent *ex ante* capital-energy substitution possibilities reconcile the large body of conflicting published evidence on this topic, for which both input complementarity and substitutability are frequently reported. Moreover, the behavior of short- and long-run average and marginal cost curves is economically consistent and indicates economies of scale for the simple dimerization plant. In sum, the methodology applied in this paper provides chemical manufacturing firms with an essential tool for making decisions regarding the type, size, and timing of capacity additions. It is possible that the concisely developed theory at the plant level will ultimately facilitate simple aggregation to describe production relationships at the firm, industry, and sector levels.

REFERENCES

Apostolakis, B. E., 1990. "Energy-capital substitutability/complementarity: the Dichotomy." *Energy Economics*, 12: 48-58.

Arrow, K. J., H. B. Chenery, B. S. Minhas and R. M. Solow, 1961. "Capital-labor substitution and economic efficiency." *Review of Economics and Statistics*, 43: 225-250.

Bairam, E. I., 1994. *Homogeneous and nonhomogeneous production functions: theory and applications.* Avebury Press.

Barsan, E. and C. Ignat, 2001. "Technical-economic analysis of water distribution networks." *Water Science and Technology: Water Supply*, 1: 210-219.

Berndt, E. R. and D. O. Wood, 1979. "Engineering and economic interpretations of energy-capital complementarity." *American Economic Review*, 69: 342-354.

Caloghirou, Y. D., A. G. Mourelatos and H. Thompson, 1997. "Industrial energy substitution during the 1980's in the Greek economy." *Energy Economics*, 19: 476-491.

Chang, K. -P., 1994. "Capital-energy substitution and the multi-level CES production function." *Energy Economics*, 16: 22-26.

Chenery, H. B., 1949. "Engineering production functions." *Quarterly Journal of Economics*, 63: 307-331.

Cobb, C. W. and Douglas, P. H., 1928. "A theory of production." *American Economic Review*, Supplement 18: 139-165.

Diewert, W. E. and T. J. Wales, 1995. "Flexible functional forms and tests of homogeneous separability." *Journal of Econometrics*, 67: 259-302.

Giannakas, K., C. T. Tran, and V. Tzouvelekas, 2000. "Efficiency, technological change and output growth in Greek olive growing farms: a Box-Cox approach." *Applied Economics*, 32: 909-916.

Gow, A. S., III and A. S. Gow, Jr., 2003. "Microeconomic theory of chemical production processes: application to VOC air stripping operations." *Advances in Environmental Research*, in press.

Gow, A. S., III, 2003a. "A microfoundations approach to input-output modeling and analysis of simple chemical processes." *Journal of Applied Input-Output Analysis*, in review.

Gow, A. S., III, 2003b. "Renaissance of the engineering production function approach: application to simple chemical reaction processes." *Advances in Theoretical Economics*, in review.

Gow, A. S., III, 2002. "Engineering production functions for technical and economic analysis of popular alkylation technologies." *Proceedings of the 5th International Conference on Refinery Processing*, AIChE Spring National Meeting, New Orleans, LA, March 2002, 543-553.

Guthrie, K. M., 1970. "Capital and operating costs for 54 chemical processes." *Chemical Engineering*, 77(13): 140-156.

Haines, T. B., 1957. "Direct operating requirement for chemical processes." *Chemical Engineering Progress*, 53(11): 556-562.

Hisnanick, J. J. and B. L. Kyer, 1995. "Assessing a disaggregated energy input using confidence intervals around translog elasticity estimates." *Energy Economics*, 17: 125-132.

Hsieh, W.-J., 1995. "Test of variable output and scale elasticities for 20 U.S. manufacturing industries." *Applied Economics Letters*, 2: 284-287.

Isard, W. and E. W. Schooler, 1955. *Location factors in the petrochemical industry*, U. S. Department of Commerce.

Luyben, W. L. and L. A. Wenzel, 1988. *Chemical process analysis: mass and energy balances.* Prentice Hall: Englewood Cliffs, NJ.

Mahmud, F. and S. Chishti, 1990. "The demand for energy in the large scale manufacturing sector of Pakistan." *Energy Economics*, 12: 251-254.

O'Connell, F. P., 1962. "Chart gives operating labor for various plant capacities." *Chemical Engineering*, 69(4): 150-152.

Peters, M.S. and K.D. Timmerhaus, 1991. *Plant design and economics for chemical engineers*, 4th ed., McGraw-Hill, New York.

Pollak, R. A. and T. J. Wales, 1987. "Specification and estimation of nonseparable two-stage technologies: the Leontief CES and the Cobb-Douglas CES." *Journal of Political Economy*, 95: 311-333.

Rushdi, A. A., 1991. "Economies of scale and factor substitution in electricity supply industry: a case study of South Australia." *Energy Economics*, 13: 219-229.

Sato, K., 1975. *Production functions and aggregation, Contributions to Economic Analysis*, North-Holland Publishing Company, Amsterdam.

Sav, G. T., 1984. "The engineering approach to economic production functions revisited: an application to solar processes." *Journal of Industrial Economics*, 33: 21-35.

Zellner, A. and H. Ryu, 1998. "Alternative functional forms for production, cost and returns to scale functions." *Journal of Applied Econometrics*, 13: 101-127.

Ziari, H. and A. Azzam, 1999. "Parametrizing nonparametric translog models: a goal programming/constrained regression study of U.S. manufacturing." *Empirical Economics*, 24: 331-339.

ENDNOTES

1. Cost correlations are periodically updated to account for both technological changes and inflationary effects. Inflationary effects are accounted for using a cost index, which relates the purchased cost of a bundle of capital goods in the present year to the purchased cost of the same bundle in a specified base year.
2. The detailed mathematical description of the *engineering process model* for the simple chemical plant is beyond the scope of this paper; however, a complete listing of model equations for all process equipment, the process variables and parameters and their units, the method of process model solution, and executable software are available from the author upon request.
3. The separator here is a fractionator in which heat is supplied to produce a vapor, which is enriched in the lighter monomer A component.

4. The author will provide a list of the constitutive equations (and assumptions) for each contribution in Equation 10 upon request.
5. See Equations 8 and 9.
6. An *isoquantlet* is the capital versus energy relationship for a special case of constant output, in which the reactor size is held constant while the separator size is varied.
7. Dimerization is the combination of two monomer molecules to form one dimer molecule, whereas polymerization is the combination of n monomer molecules to form one polymer molecule (n-mer).
8. The estimated monomer cost of 500 $/ton is the typical cost of raw materials used in polmerization processes.

A COUNT PANEL DATA STUDY OF THE SCHUMPETERIAN HYPOTHESIS

James J. Jozefowicz*

ABSTRACT

This study estimates the patent-R&D relationship using count panel data. The data is an original panel of 318 firms making R&D investments and applying for patents during the period from 1984 to 1993. A negative binomial model with fixed effects is estimated, taking into account both the discrete nature of the count dependent variable and firm-specific unobserved heterogeneity as well as overdispersion in the data. Firm-level R&D capital, concentration ratios, and various firm size proxies are used as independent variables. Analysis of the data fails to reveal support for the basic tenets of the Schumpeterian Hypothesis. In particular, firm size has a significant negative impact on innovation while industry concentration is statistically insignificant. (JEL O3, L0, C0)

1. INTRODUCTION

A firm's economic environment is very likely to have a significant impact on its innovative activity. Schumpeter advanced the notion that innovative advantage belonged to large firms as opposed to small firms and to industries characterized by imperfect competition. These are the two main tenets of the Schumpeterian hypothesis. That is, market imperfections account for the relative innovative superiority of large firms over their smaller counterparts because they allow the larger firms to retain the returns from R&D. These larger firms may also have greater access to capital markets and a stronger ability to secure funding for their research endeavors.

Griliches (1979) explained the relationship between innovative output and innovative inputs in the context of the knowledge production function (KPF), which basically stated that innovative output is the product of innovative inputs. In many studies, patent applications have served as a measure of innovative output. The United States has experienced a surge in patenting uniformly distributed across technologies. Between 1900 and the mid-1980s, 40,000 to 80,000 patent applications were submitted per year. In 1995, however, more than 120,000 patent applications were submitted to the Patent Office. A firm's R&D expenditures are the most common innovative input to the KPF. It is reasonable to assume, however, that other forces affect the R&D relationship. In particular, the degree to which R&D expenditures produce innovative output is conditioned by the market structure characteristics of the industry in question.

Empirical studies of the relationship between market structure and innovation have found that large firms tend to have higher rates of R&D spending and innovation (e.g. Scherer, 1967). In general, the rationale is that the static efficiency losses associated with monopoly are offset by gains in dynamic efficiency. However, on the whole, the empirical work in this area has been inconclusive.

---

*Department of Economics, 213 McElhaney Hall, Indiana University of Pennsylvania, Indiana, PA 15705, jimjozef@iup.edu

Industry level studies (e.g. Geroski, 1990) suggest that concentration has a dampening effect on innovation. Blundell *et al.* (1993) find that, while higher market share firms innovate more, firms in competitive industries tend to have a greater probability of innovating. Thus, the lack of competition combined with a high level of industry concentration depresses the aggregate level of innovative activity.

Research by Gopinath and Vasavada (1999) on the U.S. food processing industry demonstrates a positive effect of market share on patenting, but a negative impact of concentration on it. Work by Blundell *et al.* (1995), and Acs and Audretsch (1987) also contradicts Schumpeter's belief by finding a similar dampening effect of industry concentration on innovation. Furthermore, Acs and Audretsch and others have shown that small firms do have an innovative advantage over large firms in some industries and under certain conditions.

Smythe (2001) obtained tentative support for the Schumpeterian Hypothesis studying electric power utilization at the turn of the century in the U.S. Higher degrees of industry concentration were found to be conducive to rapid innovation. Likewise, Hall and Ziedonis (2001) found that larger firms in the U.S. semiconductor industry submit more patent applications than smaller ones.

The issues surrounding the link between market structure and innovation need further investigation. Empirically, it has been found that while large firms are more innovative in a number of industries, the opposite is true in other cases. Coupling these findings with a lack of supporting evidence regarding the effect of industry concentration on innovation calls the Schumpeterian Hypothesis into question.

In this paper, the effect of market structure on patenting as summarized by the Schumpeterian Hypothesis is studied. The study borrows from Blundell *et al.* (1995) who used British firm-level panel data on the "technologically significant and commercially important" innovations commercialized during the period from 1972 to 1982, while controlling for market structure and firm size. It also widens the scope of Gopinath and Vasavada (1999) who employed U.S. firm-level panel data for the food processing industry to investigate the relationship between market structure and patent applications, and Hall and Ziedonis (2001) who investigated patenting behavior in the U.S. semiconductor industry using panel data.

This paper is presented in six sections: Section 2 presents the data, its main characteristics, and the construction of some variables. Section 3 explores the econometric models for count data, using the basic Poisson as a benchmark model. Past research is presented in Section 4. The empirical findings are presented in Section 5. Section 6 provides a brief conclusion.

2. DATA

The data are an original panel of 318 U.S. manufacturing firms with concentration ratio data available for their SICs, which invested in R&D and applied for patents between 1984 and 1993. The relevant explanatory variables for this analysis include firm-level R&D capital, firm size proxies, and industry-level concentration ratios. The dependent variable is patent application counts serving as a proxy for innovative output.

Patents are assumed to be an indicator of innovative output or the "success" of R&D rather than just the input of R&D. The validity of this assumption has been investigated by Pakes (1985) and Griliches (1981) using the market value of the firm as an additional indicator of R&D success. In regressions of the rate of return of market value on the R&D history of the firm, contemporaneous patenting is moderately significant. Although the results are somewhat inconclusive, their result suggests that patents measure something more than the input of R&D, which can be considered the "success" or output of R&D.

Patents have been criticized as a measure of innovative output because not all patented inventions prove to be innovations and many innovations are never patented. Nevertheless, Schmookler (1966, p. 56) states, "We have the choice of using patent statistics cautiously and learning what we can from them, or not using them and learning nothing about what they alone can teach us."

In addition, the findings of Ernst (2001) support the significance of patent data as an objective output indicator for R&D efforts. The paper also points out that patent data are easily obtained. Patent data provide an aggregate indication of patenting activity by R&D-conducting firms even though the information conveyed by a single patent may be very inconsequential. Further, they provide a long time series for study in contrast to the Innovation Citation Database published by the U.S. Small Business Administration in 1984, which only included data for innovations in 1982.[1]

Concentration ratios were collected as measures of market structure. They were taken from the Census of Manufacturing bulletin, *Concentration Ratios in Manufacturing*. The ratios are defined as the percentage of total industry sales accounted for by the largest 4, 8, 20, or 50 firms.

Controlling for market structure is important because various changes, such as mergers and acquisitions, in the composition of an industry over the course of a ten-year period can have an impact on the patenting activity of firms. In addition, the Schumpeterian Hypothesis asserts that firms in industries characterized by imperfect competition tend to be more innovative than firms in industries more closely resembling the competitive model. Industry concentration ratios can be used to measure the competitiveness of an industry and help control for the role of market structure.

The information collected for each firm from Standard and Poor's Compustat includes its corporate name, primary SIC code, annual R&D expenditures, assets, capital expenditures, employment, market value, and net sales. In addition, the capital expenditure-employment, the R&D-sales, and the R&D-employment ratios were created. The total number of patent applications submitted by year and U.S. Patent Office company codes were collected from the PATSIC file on CD-ROM from the United States Patent and Trademark Office. The Compustat and patent data were matched by company code for each firm. The R&D investments at the firm level and all other firm-level dollar amount variables have been converted to constant 1990 dollars.

Following Crepon and Duguet (1997a, b), the firm's own R&D capital was created as an input to the KPF. The formula for the variable is

$$k_{it} = (1-\delta)k_{it-1} + r_{it} \qquad \textbf{(1)}$$

where k_{it} is R&D capital for firm *i* at time *t*, δ is the annual depreciation rate, and r_{it} is real R&D for firm *i* at time *t*. The depreciation rate was set at 15 percent in line with Crepon *et al.* (1998), Crepon and Duguet (1997a, b), Klette (1996), and Encaoua *et al.* (1998). All of these studies used depreciation rates in the range of 15-20 percent. Rates of 20 percent, 25 percent, 30 percent, and 50 percent were experimented with, but the exact depreciation rate made very little difference in the results as discovered by Blundell *et al.* (1995).

The descriptive statistics reported in Table 1 indicate that the mean number of patent applications submitted by a firm in a given year in this sample is 15.5. The overdispersion typical of count data is evident in the variance-to-mean ratio of 174.45 for the patent count variable. The minimum and maximum for this variable also reveal a skewed distribution, which is characteristic of non-negative data.

Table 1. Descriptive Statistics

VARIABLE	MEAN	STD. DEV.	MINIMUM	MAXIMUM
PATENT APPLICATION COUNT	15.5	52.0	0.00	1139.0
R&D ($ millions)	57.1	152.5	0.00	1727.9
ASSETS ($ millions)	2152.1	10910.2	0.03	181416.6
CAPITAL EXPENDITURE ($ millions)	148.9	696.8	0.00	10541.2
EMPLOYMENT (thousands)	10.2	30.9	0.00	383.7
CAPITAL EXPENDITURE/EMPLOYMENT	8.9	13.0	0.00	214.7
MARKET VALUE ($ millions)	2342.3	8477.2	0.03	113061.1
NET SALES ($ millions)	2080.0	9113.7	0.00	112011.8
R&D/EMPLOYMENT	5598	6604.1	0.00	104820.0
R&D/NET SALES	27.45	2126.4	0.00	348014.7

3. ECONOMETRIC MODELS

The integer-valued patent data possess some unique attributes that must be addressed econometrically using models appropriate for count data. Due to the difficulties and uncertainties associated with R&D activities, firms do not always apply for patents, resulting in a nonnegligible number of zero values. Linear regression models are not recommended to analyze this type of data because it is unlikely that the basic assumptions of normal residuals and linearity will be satisfied.

3.1 Basic Poisson Model

The simplest non-linear regression model to accommodate the discrete, non-negative nature of the patent application count variable is the Poisson model. The Poisson model requires the first two conditional moments to be equal and allows for the straightforward treatment of the zero outcomes since they are a natural outcome of the Poisson specification. Estimation of unknown parameters is straightforward and proceeds by either an iterative weighted least squares (WLS) technique or maximum likelihood estimation (MLE). Since the log-likelihood function is globally concave, maximization routines converge rapidly. In addition, the heteroskedastic and skewed distributions natural to non-negative data are accounted for by the equality of the first two conditional moments (sometimes called the "equidispersion" property).

According to the Poisson regression model, each y_{it} is drawn from a Poisson distribution with parameter λ_{it} which is related to the explanatory variables x_{it}. The primary equation for this model is

$$\Pr(Y_{it} = y_{it}) = \frac{e^{-\lambda_{it}} \lambda_{it}^{y_{it}}}{y_{it}!}, y_{it} = 0,1,... \quad \textbf{(2)}$$

where

$$\lambda_{it} = \exp(x_{it}\beta) \quad \textbf{(3)}$$

The λ_{it} is a deterministic function of x_{it} and the randomness in the model comes from the Poisson specification for the y_{it}.[2] How the mean number of events changes as a consequence of changes in one or more of the regressors is the point of interest.

As mentioned before, the first two conditional moments for the Poisson model are equal. That is,

$$E[y_{it} \mid x_{it}, \beta] = Var[y_{it} \mid x_{it}, \beta] = \lambda_{it} \quad \textbf{(4)}$$

The log-likelihood function of a panel data sample for the Poisson model is

$$L(\beta) = \sum_{i=1}^{N} \sum_{t=1}^{T} [y_{it}! - \exp(x_{it}\beta) + y_{it} x_{it}\beta]. \quad (5)$$

The basic Poisson model with its "equidispersion" property suffers from some limitations, however. Overdispersion is a concern with patent data, where the conditional variance exceeds the conditional mean. Thus, the variance of the estimator will be larger than expected and a possible efficiency loss will result. In the case of patent data, unobserved effects such as the inherent uncertainty of R&D activities, different appropriability conditions, the ability of engineers to discover new products, strategies of secrecy, or the obvious commercial risk of selling an invention result in only a few successful firms applying for a large number of patents in a given period of time, while the majority of firms may find patenting holds little or no importance for them.

In addition, the "equidispersion" property does not allow for unobserved heterogeneity, i.e., individual firm-specific effects. This is definitely a concern with firm-level data where the heterogeneity is not necessarily fully represented by the observed individual characteristics summarized by the regressors. If this restriction is inappropriately imposed, spuriously small estimated standard errors of $\hat{\beta}$ may result.

3.2 Basic Negative Binomial Model

The negative binomial model represents a more general formulation than the Poisson model. It attempts to improve on the Poisson model by including a firm unobserved effect, ε_i, in the λ_{it} parameters. The negative binomial model arises from a natural formulation of cross-section heterogeneity and is essentially an "apparent contagion" model in which individuals have constant, but unequal probability of experiencing an event.[3]

Ideally, the negative binomial model would permit the variance to grow with the mean while simultaneously allowing a conditional fixed effect, which could be correlated with the independent variables, in particular R&D. In other words, firms, which are better at producing patents for unobserved reasons, may make larger R&D expenditures than others because their return to the expenditures is higher. Support for the existence of such a correlation can be found in Duguet and Kabla (1998), who analyze data from the French technological appropriation survey (EFAT). They point out that the highest R&D budgets belong to firms that possess a technical advantage in their industry, which enables them to patent a larger fraction of their innovations.

With the fixed effect specification, it is not necessary to assume away a correlation between the firm-specific effect and the right-hand-side variables because the individual effects are conditioned out and are not estimated. This is an especially attractive feature of the fixed effects approach.

The negative binomial model assumes that the Poisson parameter λ_{it} follows a gamma distribution with parameters (γ,δ) where $\gamma = exp\ (x_{it}\beta)$ with δ common both across firms and across time. With this specification, the mean and variance of λ_{it} are $E\ [\lambda_{it}] = exp\ (x_{it}\ \beta)/\delta$ and $Var\ [\lambda_{it}] = \exp\ (x_{it}\ \beta)/\delta^2$. Note that λ_{it} can still vary even if x_{it} remains constant for a firm over time.

3.3 Fixed Effects Negative Binomial Model

To add the firm-specific effects, assume the parameters of the underlying model are $(\gamma_{it},\ \delta_i) = (e^{xit\beta},\ \phi_i /\ e^{\mu i})$ where both ϕ_i and μ_i are allowed to vary across firms. The mean is

$$\tilde{\lambda}_{it} = \exp(x_{it}\beta + \mu_i)/\phi_i \qquad \textbf{(6)}$$

while the variance is

$$Var[\lambda_{it}] = \exp(x_{it}\beta + 2\mu_i)/\phi_i^2\,. \qquad \textbf{(7)}$$

Here, the mean has been multiplied by $exp\ (\mu_i)$ and so has the standard deviation. With respect to the corresponding unconditional negative binomial model, calculate

$$E[y_{it}] = \exp(x_{it}\beta + \mu_i)/\phi_i \qquad (8)$$

with

$$Var[y_{it}] = \{e^{x_{it}\beta+\mu_i}/\phi_i\}\{1 + e_i^{\mu_i/\phi_i}\} \qquad \textbf{(9)}$$

so that the variance-to-mean ratio is $(e^{\mu i}+\phi_i)/\phi_i$. This allows for both overdispersion, which is lacking in the Poisson model, as well as a firm-specific variance-to-mean ratio, which the basic negative binomial model does not.

4. PREVIOUS WORK

Hausman *et al.* (1984) were the first to investigate differences in the propensity to patent across firms in the context of the patent-R&D relationship and explicitly account for the discrete, nonnegative nature of the patent count variable in a panel data setting. Their work reveals the importance and role of lagged R&D spending in the innovation process and develops the models for count panel data utilized in this study. Theirs was the seminal research of the patent-R&D relationship using count panel data.

The Schumpeterian Hypothesis emphasizes the importance of firm size in the production of innovations. In particular, large firms are supposed to be more innovative than small firms. Measuring firm size as a source of heterogeneity in the propensity to patent across firms is very important in this study since the firms are drawn from various manufacturing sectors.

To study the disparate effects of firm size and industry structure on innovation, Acs and Audretsch (1987) use cross-sectional data from the U.S. Small Business Administration's Innovation Citation Database to test a modified Schumpeterian Hypothesis. Specifically, they test the hypothesis that large firms hold an innovative advantage in markets characterized by imperfect competition while small firms have the innovative advantage in markets more reminiscent of the competitive model.

Acs and Audretsch find that large firms tend to have the relative innovative advantage in markets, which are capital-intensive, highly unionized, concentrated, and produce a differentiated product. On the other hand, their results indicate that small firms hold the innovative advantage over large firms in industries that are highly innovative, utilize a large component of skilled labor, and tend to be composed of a relatively high proportion of large firms. Thus, these results generally support the modified Schumpeterian hypothesis Acs and Audretsch posit.

Using the same data as in their previous paper at a more aggregated level, Acs and Audretsch (1988) test two more hypotheses related to the Schumpeterian Hypothesis. Specifically, that the degree to which R&D expenditures produce innovative output is tempered by market structure characteristics, and that the innovative activity of small firms and large firms responds to particular technological and economic regimes. They study 247 four-digit SIC industries bringing forth innovations in 1982.

Acs and Audretsch conclude that the number of innovations increases with increased industry R&D spending, but at a decreasing rate. Innovation is also positively related to skilled labor and the degree to which large firms comprise the industry. Further, they find that industry concentration dampens industry innovation, and that unionization is negatively associated with innovation as well.

Smythe (2001) conducts a study of electric power utilization using a sample of 197 U.S. manufacturing firms over the years 1899-1909. Concerned with merger activity and the diffusion of electric power in the early 1900s, Smythe uses ordinary least squares (OLS) and instrumental variable (IV) estimation to reveal tentative support for the Schumpeterian Hypothesis. Specifically, the results indicate that a high degree of industry concentration fostered rapid innovation in U.S. manufacturing firms during the study period.

Duguet and Kabla (1998) analyze an international cross-section of 299 firms representing the U.S., Japan, and Europe. The number of patents is specified to be a function of R&D spending, sales, average Herfindahl concentration index, industry dummies, and other variables. They employ pseudo maximum likelihood estimation (PMLE) of a heterogeneous Poisson model in addition to other estimation routines.

The logarithm of average concentration does not achieve statistical significance in any of the regressions run by Duguet and Kabla. The sign of the coefficient on this variable is almost always negative throughout their study. The logarithm of sales has a positive sign, but is not statistically significant using pseudo maximum likelihood estimation of a heterogeneous Poisson model.

With a cross-section of 4,164 French manufacturing firms, Crepon *et al.* (1998) specify the number of patents per employee as a function of R&D capital, the number of employees, demand pull variables, and technology push variables. They estimate the patent equation using asymptotic least squares (ALS), and find a significant positive effect of R&D capital, but an insignificant negative impact of the number of employees as a firm size proxy. Crepon *et al.* conclude that firm innovative output increases with its research effort as well as demand-pull and technology push variables.

Van Cayseele (1998) criticizes the use of cross-sectional data to test the Schumpeterian Hypothesis, and attributes much of the inconclusiveness of the results of this research to the use of cross-sectional data. Van Cayseele calls for studies that, ideally, employ panel data in the study of innovation and market structure.

Blundell *et al.* (1995) use a panel of 375 British manufacturing firms to analyze the wide range of innovative activity across firms resulting, in part, from permanent unobservable differences across them. Concerned with the effects of market structure on innovation, Blundell *et al.* model innovation as a function of market structure measures at the firm and industry levels, tangible capital stock, knowledge capital stock at the industry level, the firms' accumulated knowledge stock, a firm-specific effect, and a time-specific effect.

Blundell *et al.* find a positive, significant effect for market share while concentration enters negatively (impact of competition on aggregate innovation is positive). Since the effects of the recession dummies are negative, they conclude that firms innovate more in booms to capture increased demand. With measured fixed effect variables, the effect of market share becomes smaller as does the effect of knowledge stock. Basically, controlling for unobserved firm heterogeneity indicates that dominant firms have a higher propensity to innovate, but competitive industries tend to generate higher aggregate number of innovations as an offsetting effect.

Gopinath and Vasavada (1999) study a panel of 32 U.S. food-processing firms over the period 1965-1981. Their aim is to investigate the impact of market structure and R&D spending on the number of patent applications submitted by firms using fixed and random effects Poisson and negative binomial models for count panel data. They consider patent applications to be a function of R&D expenditures, market share, and industry concentration variables, among others.

Gopinath and Vasavada discover a positive impact of R&D capital on patent applications using both Poisson and negative binomial models. When taking market structure into account, a positive and statistically significant coefficient for market share is obtained with the random effects Poisson model and the random effects negative binomial model. While positive, the coefficient is not significant at a 5 percent level with the fixed effects negative binomial model. Alternatively, a negative sign is obtained for the coefficient on the number of establishments variable, but it is only statistically significant with the random effects Poisson model. A Hausman test favors both random effects models over the fixed effects model indicating the absence of correlation between the firm-specific effect and the regressors.

Hall and Ziedonis (2001) utilize a panel of 95 U.S. publicly traded semiconductor firms for the period 1979-1995 to examine their patenting behavior. They attempt to explain the surge in semiconductor patenting since the mid-1980s, which contradicts more recent survey data claiming a lack of dependence upon patents to capture returns to R&D in the industry. Patent applications submitted by firms are modeled as a function of R&D spending, firm size, time dummies, firm age, and firm type.

Using Poisson models, their results show a positive and significant effect of R&D expenditures and firm size on the propensity of semiconductor firms to patent. Hall and Ziedonis conclude that a strengthening of U.S. patent rights in the 1980s has led to “patent portfolio races” and entry of specialized design firms.

In contrast to Blundell *et al.* (1995), Gopinath and Vasavada (1999), and Hall and Ziedonis (2001), this work utilizes U.S. firm-level panel data on patents from 1984 to 1993 for a variety of industries. It is well known that considerable merger and acquisition activity in the U.S. as well as the emergence of a plethora of innovations by firms of all sizes characterized the 1980s. As such, this data set provides a unique opportunity to analyze the Schumpeterian Hypothesis.

5. RESULTS

Since the regressors are taken in logs, the estimated coefficients of the R&D capital and the firm size proxies can be interpreted as elasticities. In the event that the regressors are not taken in logs then the parameters can be directly interpreted as semi-elasticities. Thus, the estimate gives the proportionate change in the conditional mean when the regressor in question changes by one unit. This is the case for the industry concentration ratios.

Likelihood ratio (LR) and Wald tests for overdispersion due to Cameron and Trivedi (1998) strongly reject the null hypothesis of equidispersion at the 1 percent level throughout the study.[4] The presence of overdispersion in the data supports the use of the negative binomial model for the analysis because its variance is proportional to the mean. In addition, the log-likelihood values obtained with the fixed effects negative binomial model are higher than those of the fixed effects Poisson model in all regressions, providing further support for the negative binomial model.

5.1 Industry Concentration

Using the four-firm concentration ratio as the measure of industry concentration in Table 2, the fixed effects negative binomial model finds R&D capital significant at the 1 percent level with an elasticity of 0.15. Gopinath and Vasavada (1999), Crepon *et al.* (1998), and Crepon and Duguet (1997 a, b) also obtained positive and significant estimates for R&D capital. In fact, Crepon and Duguet (1997a) obtained a coefficient on R&D capital of about 0.11. The concentration ratio estimate of 0.0009 is not statistically significant. Duguet and Kabla (1998) also found industry concentration to be insignificant in their analysis of French manufacturing data.

However, the estimate for the four-firm concentration ratio here disputes the findings of Blundell *et al.* (1995) as well as others. Using a British five-firm concentration ratio as an independent variable in their study, they found a significant, negative effect of industry concentration on innovation. On the other hand, Acs and Audretsch (1987), and Smythe (2001) obtained significant, positive coefficients for concentration.

Table 2. Fixed Effects Negative Binomial Regression with Four-Firm Concentration Ratio

Variable	Coefficient	Standard Error	t-Statistic
LN R&D Capital	0.1508*	0.0073	20.691
CR4	0.0009	0.0015	0.577

* Significant at 1%; ** Significant at 5%; ***Significant at 10%.

Table 3 shows that the eight-firm concentration ratio is not significant in the fixed effects negative binomial regression when used in place of the four-firm concentration ratio. Its elasticity of 0.0011 is only slightly larger than that of the four-firm concentration ratio. The twenty-firm concentration ratio also fails to show statistical significance in Table 4 with a coefficient of 0.0010. The effect of the fifty-firm concentration ratio is 0.0007 and insignificant in Table 5. The elasticity on R&D capital remains roughly 0.15 regardless of the concentration ratio used.

Table 3. Fixed Effects Negative Binomial Regression with Eight-Firm Concentration Ratio

Variable	Coefficient	Standard Error	t-Statistic
LN R&D Capital	0.1490*	0.0081	18.479
CR8	0.0011	0.0014	0.804

* Significant at 1%; ** Significant at 5%; ***Significant at 10%.

Table 4. Fixed Effects Negative Binomial Regression with Twenty-Firm Concentration Ratio

Variable	Coefficient	Standard Error	t-Statistic
LN R&D Capital	0.1484*	0.0090	16.529
CR20	0.0010	0.0013	0.738

* Significant at 1%; ** Significant at 5%; ***Significant at 10%.

Table 5. Fixed Effects Negative Binomial Regression with Fifty-Firm Concentration Ratio

Variable	Coefficient	Standard Error	t-Statistic
LN R&D	0.1491*	0.0010	15.505
CR50	0.0007	0.0012	0.571

* Significant at 1%; ** Significant at 5%; ***Significant at 10%.

Clearly, the coefficient of R&D capital is considerably less than one, indicating decreasing returns to scale as discovered by Jaffe (1986), Crepon and Duguet (1997a), and Blundell *et al.* (2002). Specifically, a doubling of R&D capital leads to increases of about 15 percent in the number of patents. This can be expected, though, as investments in R&D do not always yield useful technologies. In addition, Encaoua *et al.* (1998) point out that only one-third of innovations are patented, on average.

Overall, the uniform lack of statistical significance of the concentration ratios in the regressions echoes the results of Duguet and Kabla (1998), and casts doubt on the Schumpeterian Hypothesis. From these results, it appears that industry concentration does not exert an appreciable influence on innovation. Perhaps market structure does not represent the important determinant of innovation that it did in the past. These findings run counter, however, to the work of Blundell *et al.* (1995) who discovered that greater concentration has a dampening effect on innovation, and Acs and Audretsch (1987), and Smythe (2001) who found a positive impact of concentration.

5.2 Firm Size

Using the fixed effects Negative Binomial model, different candidates for firm size measures are considered and their findings are reported in Tables 6-13. The firm size variable in question is included with R&D capital on the right-hand-side in each case. Most of the regressions show a small, negative effect of firm size on patenting significant at the 10 percent level. In fact, the elasticities for all of the statistically significant firm size proxies are in the vicinity of –0.03. The elasticities on employment, net sales, and the R&D-sales ratio are not significant. The capital expenditure estimate is the only proxy significant at the 5 percent level.

Table 6. Fixed Effects Negative Binomial Regression with Assets

Variable	Coefficient	Standard Error	t-Statistic
LN R&D Capital	0.1998*	0.0190	10.532
LN ASSETS	-0.0342***	0.0182	-1.885

* Significant at 1%; ** Significant at 5%; ***Significant at 10%.

Table 7. Fixed Effects Negative Binomial Regression with Capital Expenditures

Variable	Coefficient	Standard Error	t-Statistic
LN R&D Capital	0.1921*	0.0141	13.590
LN CAPITAL EXP	-0.0346**	0.0169	-2.051

* Significant at 1%; ** Significant at 5%; ***Significant at 10%.

Table 8. Fixed Effects Negative Binomial Regression with Employment

Variable	Coefficient	Standard Error	t-Statistic
LN R&D Capital	0.1917*	0.0176	10.913
LN EMPLOYMENT	-0.0418	0.0267	-1.568

* Significant at 1%; ** Significant at 5%; ***Significant at 10%.

Table 9. Fixed Effects Negative Binomial Regression with Capital Expenditure-Employment Ratio

Variable	Coefficient	Standard Error	t-Statistic
LN R&D Capital	0.1960*	0.0153	12.830
LN CAP EXP/EMPL	-0.0402***	0.0210	-1.919

* Significant at 1%; ** Significant at 5%; ***Significant at 10%.

Table 10. Fixed Effects Negative Binomial Regression with Market Value

Variable	Coefficient	Standard Error	t-Statistic
LN R&D Capital	0.1920*	0.0213	8.998
LN MARKET VALUE	-0.0365***	0.0197	-1.853

* Significant at 1%; ** Significant at 5%; ***Significant at 10%.

Table 11. Fixed Effects Negative Binomial Regression with Net Sales

Variable	Coefficient	Standard Error	t-Statistic
LN R&D Capital	0.1698*	0.0134	12.708
LN NET SALES	-0.0025	0.0122	-0.207

* Significant at 1%; ** Significant at 5%; ***Significant at 10%.

Table 12. Fixed Effects Negative Binomial Regression with R&D-Employment Ratio

Variable	Coefficient	Standard Error	t-Statistic
LN R&D Capital	0.1879*	0.0127	14.804
LN R&D/EMPL	-0.0309***	0.0163	-1.896

* Significant at 1%; ** Significant at 5%; ***Significant at 10%.

Table 13. Fixed Effects Negative Binomial Regression with R&D-Sales Ratio

Variable	Coefficient	Standard Error	t-Statistic
LN R&D Capital	0.1823*	0.0118	15.485
LN R&D/NET SALES	-0.0229	0.0160	-1.437

* Significant at 1%; ** Significant at 5%; ***Significant at 10%.

The firm size results reflect the findings of other studies, including Duguet and Kabla (1998), Acs and Audretsch (1987), and Crepon *et al.* (1998). Duguet and Kabla used sales as their size proxy and obtained an insignificant coefficient for it. Crepon *et al.* (1998) found employment to be a statistically insignificant size proxy in their research. A negative impact of firm size on patenting in some industries was also obtained by Acs and Audretsch (1987). However, Hall and Ziedonis (2001), Blundell *et al.* (1995), and Gopinath and Vasavada (1999) found a positive effect of firm size on innovation.

Basically, the negative albeit small effects of firm size proxies on patenting indicate that larger firms are less innovative than smaller firms. This would seem to contradict the Schumpeterian Hypothesis view of innovation.

5.3 Industry Concentration and Firm Size

Since part of the Schumpeterian Hypothesis asserts that larger firms are more innovative than smaller firms, regressions are run adding the capital expenditure variable as a firm size proxy while controlling for market structure, and their results are reported in Tables 14-17.[5] The elasticity on R&D capital is consistently 0.19 throughout this analysis.

Table 14.

Fixed Effects Negative Binomial Regression with Four-Firm Concentration Ratio and Capital Expenditures

Variable	Coefficient	Standard Error	t-Statistic
LN R&D Capital	0.1918*	0.0148	12.925
LN CAPITAL EXP	-0.0347**	0.0170	-2.045
CR4	0.0001	0.0017	0.062

* Significant at 1%; ** Significant at 5%; ***Significant at 10%.

Table 15.

Fixed Effects Negative Binomial Regression with Eight-Firm Concentration Ratio and Capital Expenditures

Variable	Coefficient	Standard Error	t-Statistic
LN R&D Capital	0.1934*	0.0151	12.777
LN CAPITAL EXP	-0.0344**	0.0170	-2.020
CR8	-0.0004	0.0016	-0.224

* Significant at 1%; ** Significant at 5%; ***Significant at 10%.

Table 16.

Fixed Effects Negative Binomial Regression with Twenty-Firm Concentration Ratio and Capital Expenditures

Variable	Coefficient	Standard Error	t-Statistic
LN R&D Capital	0.1962*	0.0153	12.789
LN CAPITAL EXP	-0.0339**	0.0171	-1.976
CR20	-0.0008	0.0014	-0.576

* Significant at 1%; ** Significant at 5%; ***Significant at 10%.

Table 17.

Fixed Effects Negative Binomial Regression with Fifty-Firm Concentration Ratio and Capital Expenditures

Variable	Coefficient	Standard Error	t-Statistic
LN R&D Capital	0.1939*	0.0154	12.566
LN CAPITAL EXP	-0.0343**	0.0173	-1.983
CR50	-0.0003	0.0013	-0.240

* Significant at 1%; ** Significant at 5%; ***Significant at 10%.

Looking at the results for firm size, the elasticities on capital expenditure are consistently significant at the 5 percent level across models with a magnitude of approximately –0.03. Alternatively, the concentration ratios are never significant, regardless of model. The signs on all concentration ratio estimates are negative except for that of the four-firm concentration ratio.

From these results, it appears that firm size does have a significant dampening effect on patenting. Industry concentration, however, does not significantly influence innovative activity. Thus, this study does not yield support for the Schumpeterian Hypothesis.

6. CONCLUSION

The Schumpeterian Hypothesis is called into question in this analysis. An insignificant effect of industry concentration on innovation is discovered in agreement with Duguet and Kabla (1998). This runs counter to the work of Blundell *et al.* (1995) who found a significant and negative impact of concentration, and Acs and Audretsch (1987), and Smythe (2001) who estimated significant, positive effects of concentration. Thus, more concentrated industries, which are characterized by less competition, do not foster more patent applications, according to this work.

Turning to another tenet of the Schumpeterian Hypothesis – firm size – this paper finds that larger firms tend to have fewer patent applications than smaller firms. Acs and Audretsch (1987) echo this finding in certain industries. This effect is most significant when firm size is measured by capital expenditure. It is remarkable that, regardless of measure, the coefficients on the firm size proxies are in the neighborhood of –0.03. This means that doubling the value of a particular firm size proxy will result in 3 percent fewer patent applications submitted on average.

The government should consider channeling R&D funds to small firms in an effort to maximize the return on its investment. These firms could complement the government financing with their own R&D resources in the pursuit of innovations and, possibly, secure more patent applications as a result. The magnitude of the changes indicated by this analysis, though, are relatively small, and it is not clear that enough additional patenting would be fostered for such programs to be worthwhile. If sufficient patenting activity occurs and the estimated value of the benefits from these innovations exceed the costs of R&D grants, then such policies should be implemented.

In considering the impact of market structure and firm size on patent applications, as summarized in the Schumpeterian Hypothesis, evidence contradictory to Schumpeter's assertions is obtained. Firm size has a negative role in the promotion of innovation, while industry concentration plays no significant role at all based on these results. Thus, smaller firms are found to hold the innovative advantage over larger firms in this study. Coupling these findings with those of Hall and Ziedonis (2001), Gopinath and Vasavada (1999), and Blundell *et al.* (1995) who obtained opposite results, however, dilutes the relevance and strength of the Schumpeterian Hypothesis in the modern era.

ACKNOWLEDGEMENTS

I gratefully acknowledge the suggestions and support of Derek Bandera, Michael Collins, Maria Phelan, Natalie Reaves, David Yerger, and an anonymous reviewer. Kenneth McDermott and Jason Kelley provided invaluable research assistance and their contributions are greatly appreciated. I am grateful to Debbie Bacco for her superb editorial assistance. I would, especially, like to thank Robert Viégas for his role in the completion of this study. Finally, my greatest debt is to Stephanie Jozefowicz.

REFERENCES

Acs, Zoltan J. and Audretsch, David B. "Innovation, Market Structure, and Firm Size." *Review of Economics and Statistics*, November 1987, 69(4), pp. 567-574.

Acs, Zoltan J. and Audretsch, David B. "Innovation in Large and Small Firms: An Empirical Analysis." *American Economic Review*, September 1988, 78(4), pp. 678-690.

Blundell, Richard; Griffith, Rachel and Van Reenen, John. "Knowledge Stocks, Persistent Innovation and Market Dominance: Evidence from a Panel of British Manufacturing Firms." *IFS Working Paper No. W93/19*, 1993.

Blundell, Richard; Griffith, Rachel and Van Reenen, John. "Dynamic Count Data Models of Technological Innovation." *Economic Journal*, March 1995, 105(429), pp. 333-344.

Blundell, Richard; Griffith, Rachel and Windmeijer, Frank. "Individual Effects and Dynamics in Count Data Models." *Journal of Econometrics,* May 2002, 108(1), pp. 113-131.

Cameron, A. Colin and Trivedi, Pravin K. *Regression Analysis of Count Data*. Cambridge: Cambridge University Press, 1998.

Crepon, Bruno and Duguet, Emmanuel. "Estimating the Innovation Function from Patent Numbers: GMM on Count Panel Data." *Journal of Applied Econometrics*, May-June 1997a, 12(3), pp. 243-263.

Crepon, Bruno and Duguet, Emmanuel. "Research and Development, Competition and Innovation Pseudo-Maximum Likelihood and Simulated Maximum Likelihood Methods Applied to Count Data Models with Heterogeneity." *Journal of Econometrics*, August 1997b, 79(2), pp. 355-378.

Crepon, Bruno; Duguet, Emmanuel and Mairesse, Jacques. "Research, Innovation and Productivity: An Econometric Analysis at the Firm-level." *Economics of Innovation and New Technology*, 1998, 7(2), pp. 115-158.

Duguet, Emmanuel and Kabla, Isabelle. "Appropriation Strategy and the Motivations to Use the Patent System: An Econometric Analysis at the Firm-level in French Manufacturing." *Annales d'Economie et de Statistique*, January-June 1998, 0(49/50), pp. 289-327.

Edwards, K. and Gordon, T. J., *Characterizations of Innovations Introduced on the U. S. Market in 1982, Report Prepared for the U. S. Small Business Administration*, Washington, D.C.: The Futures Group, 1984.

Encaoua, David; Hall, Bronwyn H.; Laisney, Francois and Mairesse, Jacques. "The Economics and Econometrics of Innovation." *Annales d'Economie et de Statistique*, January-June 1998, 0(49/50), pp. 27-51.

Ernst, Holger. "Patent Applications and Subsequent Changes of Performance: Evidence from Time-Series Cross-Section Analyses on the Firm-level." *Research Policy*, 2001, 30(6), pp. 143-157.

Geroski, Paul A. "Innovation, Technological Opportunity and Market Structure." *Oxford Economic Papers*, July 1990, 42(3), pp. 586-602.

Gopinath, Munisamy and Vasavada, Utpal. "Patents, R&D, and Market Structure in the U.S. Food Processing Industry." *Journal of Agricultural and Resource Economics*, July 1999, 24(1), pp. 127-139.

Griliches, Zvi. "Issues in Assessing the Contribution of R&D to Productivity Growth." *Bell Journal of Economics*, Spring 1979, 10(1), pp. 92-116.

Griliches, Zvi. "Market Value, R and D, and Patents." *Economics Letters*, 1981, 7, pp. 183-187.

Hall, Bronwyn H. and Ziedonis, Rosemarie H. "The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995." *The RAND Journal of Economics*, Spring 2001, 32(1), pp. 101-128.

Hausman, Jerry; Hall, Bronwyn H. and Griliches, Zvi. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." *Econometrica*, July 1984, 52(4), pp. 909-938.

Jaffe, Adam B. "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Value." *American Economic Review,* December 1986, 76(5), pp. 984-1001.

Klette, Tor Jakob. "R&D, Scope Economies, and Plant Performance." *The RAND Journal of Economics*, Autumn 1996, 27 (3), pp. 502-522.

Pakes, Ariel. "Patents, R and D, and the Stock Market Rate of Return." *Journal of Political Economy*, April 1985, 93(2), pp. 390-409.

Scherer, Frederic M. "Market Structure and the Employment of Scientists and Engineers." American Economic Review*, June 1967, 57(3), pp. 524-531.*

Schmookler, J. *Invention and Economic Growth*. Cambridge: Harvard University Press, 1966.

Smythe, Donald J. "The Great Merger Movement and the Diffusion of Electric Power Utilization in American Manufacturing, 1899-1909: A Simple Test of the Schumpeterian Hypothesis." *Eastern Economic Journal*, Summer 2001, 27(3), pp. 253-265.

Van Cayseele, P.J.G. "Market Structure and Innovation: A Survey of the Last Twenty Years." *De Economist*, October 1998, 146(3), pp. 391-417.

ENDNOTES

1. See Edwards and Gordon (1984) and Acs and Audretsch (1988) for a description of this data.
2. The exponential function is used to ensure the non-negativity of y_{it}.
3. There are also the "true contagion" model in which all individuals have the same probability of experiencing an event initially, but this is modified by prior occurrences of events; the "proneness" model in which individuals are heterogeneous in terms of their proneness to certain events, with the heterogeneity attributed to individual or environmental effects; and the "spells" model in which events occur in clusters and are dependent.
4. The LR test statistic is calculated as 2[Poisson log-likelihood – Negative Binomial log-likelihood]. Its critical value was $\chi^2_{0.98}(1) = 5.41$. The Wald test statistic is calculated as the t-statistic for α in the Negative Binomial model. Its critical value was $z_{0.99} = 2.33$. Test statistics of 71143 and 28 for the LR and Wald tests, respectively, were typical in the analysis.
5. The results of fixed effects Negative Binomial regressions using assets as the firm size proxy provide estimates fairly comparable in size and significance to these. The same is also true of regressions run using the capital expenditure-employment ratio and the R&D-employment ratio. However, the latter two obtained a negative and statistically insignificant coefficient for the four-firm concentration ratio.

USING NLSY-GEOCODE DATA TO DETERMINE THE EFFECTS OF TAXES AND MINIMUM AGE LAWS ON THE ALCOHOLIC BEVERAGE DEMAND OF YOUNG ADULTS

Mark Paul Gius*

ABSTRACT

In the present study, OLS and logit regression analysis are used to determine the effect that minimum age laws and taxes have on alcohol consumption and binge drinking. NLSY-Geocode data are used in order to construct individual-level demand equations. The use of this data allows for the identification of the individual's state of residence and thus enables the researcher to properly match the individual to the appropriate state alcohol tax rate. Results indicate that taxes have a negative effect on alcohol consumption but no effect on binge drinking. Minimum age laws, however, are effective in reducing both the total quantity of alcohol consumed and binge drinking.

1. INTRODUCTION

There have been numerous attempts by both the Federal and state governments to curtail the consumption of alcoholic beverages by young adults. These efforts usually took the form of increases in the minimum legal ages and higher taxes on alcoholic beverages. For example, all state-level minimum legal drinking ages were raised to 21 in the 1980s, and, in 1991, the Federal excise taxes on beer and wine increased from 16 cents per six-pack of beer and 3 cents per 750 ml bottle of wine to 32 cents and 21 cents respectively. Even given these increases in taxes, however, the tax as a share of the price is still very low (less than one percent in some cases).

Much research has been devoted to the effect of these public policy changes on the consumption of alcoholic beverages (Mast, Benson, and Rasmussen, 1999; Pacula, 1998; Laixuthai and Chaloupka, 1993; Heien and Pompelli, 1989; Coate and Grossman, 1988).

Mast et al. (1999) conducted an extensive study on the effect of beer taxes on alcohol-related traffic fatalities. The authors find that prior studies employing reduced-form demand equations may have significant missing variable bias and that when additional public policy control variables are used or two-equation recursive models are employed, the statistical significance of taxes is reduced substantially. This result indicates that the relationship between beer taxes and alcohol-related traffic fatalities is debatable and may not be as robust as prior research would have suggested. Other public policy control measures may be more effective than taxes at reducing consumption and hence traffic fatalities.

*Professor of Economics, Department of Economics, Quinnipiac University, Hamden, CT 06518, Ph: 203-582-8576, Fax: 203-582-8664, e-mail: gius@quinnipiac.edu

Pacula (1998) focused on the substitutability between alcohol and marijuana. The purpose of her study was to determine if increases in excise taxes on alcoholic beverages would result in consumers, primarily young adults, switching from alcoholic beverages to marijuana. Pacula begins with the premise that excise taxes reduce the demand for alcoholic beverages. Using data from the National Longitudinal Survey of Youth, she finds that alcohol and marijuana are complements rather than substitutes. Hence, increases in taxes on alcohol would not only reduce alcohol consumption but would also reduce the consumption of marijuana. Pacula used data from 1984 and used the real federal and state tax on a case of beer as a proxy for the price of alcohol.

Laixuthai and Chaloupka (1993) examined the effects of minimum legal drinking ages and excise taxes on alcohol consumption by teenagers. Using survey data from the years 1982 and 1989, the authors find that increasing excise taxes on alcoholic beverages and increasing the minimum legal drinking age both reduce teen consumption of alcoholic beverages.

Heien and Pompelli (1989) used data obtained from the 1977-78 Household Food Consumption Survey administered by the US Department of Agriculture. Creating demand structures for alcoholic beverages, the authors find that demand is price inelastic for all classes of alcoholic beverages. Given these results, the authors conclude that tax increases have little or no effect on the demand for alcoholic beverages.

Coate and Grossman (1988) used data from the National Health and Nutrition Examination Survey (1979-1980) in order to determine if public policy measures have any effect on alcohol consumption by youths ages 16 to 21. Their results indicate that both minimum age laws and excise taxes have negative effects on alcohol consumption by young adults. They conclude that increasing the minimum legal drinking age and the federal excise tax on beer would reduce alcohol consumption by those 16-21 years of age. Their results also suggest that tax policy may be somewhat more effective in deterring alcoholic beverage consumption than minimum age laws.

The present study estimates alcohol demand functions at the individual level in order to determine the effect of taxes and minimum age laws on alcoholic beverage demand. A unique feature of the present study is that it combines two data sets: the National Longitudinal Survey of Youth and data on alcoholic beverage taxes at the state level. This combination of data sets improves the analysis of the effect of taxes on alcohol consumption because it allows the present study to capture the effect of taxes and minimum age laws on what is essentially an individual decision, the decision of whether or not to consume alcoholic beverages. In addition, the present study attempts to determine if taxes not only affect the consumption of alcohol, but also whether taxes affect destructive and risky behaviors, such as binge drinking.[1]

Using ordinary least squares and binomial logit regression analyses, results of the present study indicate that alcoholic beverage taxes have a negative effect on alcoholic beverage consumption but no effect on binge drinking. However, results indicate that minimum age laws are effective in reducing not only overall alcoholic beverage demand but also binge drinking. Hence, the results of the present study corroborate the results of many prior studies in this area (Pacula, 1998; Laixuthai and Chaloupka, 1993; Coate and Grossman, 1988).

2. EMPIRICAL TECHNIQUE

In order to test the effect of taxes and minimum age laws on alcohol consumption, an individual-level demand function for alcoholic beverages is estimated. Guidance was obtained from several studies in the construction of this demand function (Pacula, 1998; Gao, Wailes, and Cramer, 1995; Thies and Register, 1993; Lee and Tremblay 1992; Adrian and Ferguson 1987; Uri 1986; McCornac and Filante 1984; Duffy 1981; Johnson and Oksanen 1974; Simon 1969).

The first empirical equation estimated in the present study is based on consumer theory models of utility maximization. Theory suggests that consumers maximize utility subject to prices and income.

$$\max U = U(A,C) \quad s.t.\ P_A, I \tag{1}$$

where A denotes the quantity of alcoholic beverages consumed, C denotes a composite good with a price of $1 per unit, P_A denotes the price of alcoholic beverages, and I is income. Obtaining first order conditions and solving for A and C, one obtains the following implicit individual-level demand function:

$$A = f\,(P_A, I, \mathrm{T}, \mathrm{X}) \tag{2}$$

where **T** is a vector of taste-influencing variables, **X** is a vector of public policy variables (taxes and minimum age laws) and all other variables are as defined previously. Theory suggests that alcohol consumption is negatively related to its own price and any factor that increases the price of consuming alcohol, and positively related to income, tastes, and other socioeconomic variables. It is reasonable to assume that taxes and minimum age laws, at least for underage individuals, would be viewed as factors that increase the price of alcoholic beverages; hence both should be negatively related to alcohol consumption. This hypothesis will be tested in the present study.

In order to determine the effects of taxes and minimum age laws on binge drinking, the empirical technique must capture the probability that a person will drink an excessive number of alcoholic beverages in a limited period of time. In order to model that behavior, it is reasonable to assume that individuals view binge drinking as a risky behavior, similar to criminal activities or unprotected sex. Hence, there are returns, albeit sometimes rather implicit returns, from engaging in such risky behaviors. These returns may take the form of increased pleasure or admiration from one's peers. These positive returns from binge drinking may be reduced by a variety of factors, two of which may be the public policy control variables of taxes and minimum age laws. Taxes may reduce the incidence of binge drinking because an increases in taxes increases the price of alcoholic beverages, thus reducing the net returns from drinking copious amounts of alcohol. Minimum age laws may also reduce the net return from binge drinking primarily because such laws impose fines and penalties on those underage individuals who engage in such behavior.

Using this theory as a guide, the binge drinking behavior of young adults may be modeled as follows:

$$Y = f\,(P_A, \mathrm{T}, \mathrm{Z}, \mathrm{X})$$

where Y denotes the number of times that a person binge drinks, **Z** is a vector of socioeconomic factors that influence binge-drinking, and all other variables are as defined previously. It is reasonable to assume that, for a certain segment of the population, binge drinking is desirable ; hence when there is a perception that binge drinking is desirable, an individual will binge drink more often. Certain socioeconomic and individual characteristics, such as being a male or being heavily influenced by one's peers, may increase the desirability of this risky behavior. It is reasonable to assume that taxes and minimum age laws reduce the desirability of binge drinking; hence the net return from binge drinking falls, and the individual binge drinks less as taxes and minimum ages increase.

Unfortunately, Y is not observed. A binary variable that takes the value of one if a person has engaged in binge drinking during the past month and zero otherwise is instead used as a proxy for the unobservable variable Y. In order to estimate equation (4), assuming a binary dependent variable, the following logit regression is employed:

$$Prob(Y = 1) = \frac{e^{\mathrm{B}'\mathrm{X}}}{1 + \mathrm{e}^{\mathrm{B}'\mathrm{X}}}$$

where Y is the dependent variable that equals one if the person binge drinks and zero otherwise, **X** is a vector of explanatory variables, and B is a vector of parameters.

Given the above theoretical foundations, the following equations are estimated in the present study:

$$DRINKS = \alpha_0 + \alpha_1 LTAX + \alpha_2 MARITAL + \alpha_3 LINCOME +$$

$$\alpha_4 WHITE + \alpha_5 MALE + \alpha_6 LAGE +$$

$$\alpha_7 URBAN + \alpha_8 SOUTH + \alpha_9 PEER + \alpha_{10} FAMILY +$$

$$\alpha_{11} LPRICE + \alpha_{12} LMINAGE + u$$

$$BINGE = \alpha_0 + \alpha_1 TAX + \alpha_2 MARITAL + \alpha_3 INCOME +$$

$$\alpha_4 WHITE + \alpha_5 MALE + \alpha_6 AGE + \alpha_7 AGE2 + \alpha_8 GRADE +$$

$$\alpha_9 URBAN + \alpha_{10} SOUTH + \alpha_{11} PEER + \alpha_{12} FIRST + \alpha_{13} FAMILY +$$

$$\alpha_{14} PRICE + \alpha_{15} MINAGE + u$$

Variables are defined as follows: DRINKS is the log of the number of alcoholic drinks consumed during the past month; BINGE takes a value of one if person drank more than six drinks at least one time in the past month and zero otherwise; TAX is the weighted average tax rate in percentage terms on alcoholic beverages[2]; INCOME is the respondent's income[3]; MARITAL denotes marital status where a value of one indicates that the person is married and a value of zero indicates otherwise; WHITE denotes the race of the individual where a value of one indicates that the person is white and zero otherwise; MALE denotes sex; AGE is the age of the individual; AGE2 is age squared; GRADE denotes years of education; URBAN has a value of one if the person lives in an urban area and zero otherwise; SOUTH takes a value of one if person lives in the South and zero otherwise; PEER indicates if the person was most influenced as a child by a peer; FIRST is the age at which the individual first started drinking; FAMILY takes a value of one if the person has a family member who has a drinking problem; PRICE is the price of alcoholic beverages[4]; MINAGE is the minimum legal drinking age in the respondent's state of residence; and u denotes a normally-distributed random error term. The variables in equation (5) prefixed by "L" are the logs of the respective variables; hence, equation (5) is estimated as a log-log model.

The equations were estimated using the National Longitudinal Survey of Youth - Geocode (NLSY) data set. Two years of data were examined: 1982 and 1994. The use of these two different data sets allows us to determine if the factors that affect alcohol consumption differ by age; the average age of the 1982 data set is 19, while the average age of the 1994 data set is 31. In addition, since in 1982 minimum legal drinking ages differed by state, the use of these two years of data allow us to determine if minimum age drinking laws have any effect on alcohol consumption. These two data sets were combined in order to obtain a pooled data set with 2067 observations. Equation (5) was estimated using ordinary least squares; equation (6) was estimated using a logit regression analysis.

3. Data and Results

Data for the present study was obtained from a variety of sources. State-level, alcoholic beverage tax rates were obtained from The Book of the State, The Brewer's Almanac, and the Tax Administrator's web-site.[5]

Most of the data used in the present study was obtained from the National Longitudinal Survey of Youth - Geocode (NLSY). The NLSY was constructed to be a nationally representative sample of the civilian non-institutionalized population at the time of the initial survey in 1979. The NLSY consisted of 12,686 young men and women who were between the ages of 14 and 22 when they were first surveyed in 1979. Interviews with NLSY respondents have been conducted annually since 1979, and retention rates have been relatively high, averaging over 90 percent. Each age-sex cohort is represented by a multi-stage probability sample drawn by the Bureau of the Census from a list of sampling areas that had been constructed for the Monthly Labor Survey. The NLSY employed extensive household interviews in the selected sampling areas in order to obtain as random and as representative a sample as possible.

The Geocode data set used in the present study is especially important because it provides detailed geographic information concerning the residences of the respondents. This detailed residential information is not available on the regular NLSY and allows one to identify the particular state of residence for each respondent. Hence, one is able to match the appropriate state tax rate to the respondents who reside in that state. To my knowledge, no other study employs this approach.

Although the NLSY-Geocode surveys over 12,000 individual every year, due to a variety of problems, there are sometimes missing responses. After eliminating all of the observations with missing responses, the 1982 sample had 1174 observations, and the 1994 sample had 893 observations; the total number of observations is 2067.

In addition, not all states in the US are examined in the present study. Some states in the US are known as "control" states; in these states, the state governments have monopolized some or all of the alcoholic beverage retail market or have imposed price controls on alcoholic beverage products. Given the difficulty in determining the true tax rates in these states, the control states have been eliminated from the present study. For the 1994 sample, 32 states were included; for the 1982 sample, 31 states were included. Due to data constraints, Hawaii was excluded from the 1982 sample. See Table 1 for a list of those states included in the present study.

Descriptive statistics for all variables used in the present study are presented on Table 2. Regression results are shown on Tables 3 and 4.

Given that equation (5) is estimated as a log-log model, the coefficients in table 3 can be interpreted as elasticities. Hence, according to the results, a one percent increase in the tax rate would result in a 2.75 percent decline in alcohol consumption. This is a rather elastic demand relationship, much more elastic than many other studies have found (Pacula, 1998; Laixuthai and Chaloupka, 1993;

Coate and Grossman, 1988). Regarding minimum age laws, results also indicate that increases in minimum ages reduces alcohol consumption; once again, this result may be interpreted as an elasticity. Hence, a one percent increase in the minimum age would reduce alcohol consumption by 1.6 percent, not as robust as the tax result, but still a rather elastic relationship.

Regarding the other variables in equation (5), price is significant and positive although inelastic; MARITAL and SOUTH are significant and negative; WHITE and MALE are significant and positive. Hence, single, white males living in areas outside the South are more likely to drink than others.

Concerning the binge drinking results presented on Table 4, it is important to note that the coefficients are interpreted as follows: If a logit coefficient is 0.4, then for a unit increase in the value of that explanatory variable, the log of the odds of engaging in binge drinking increase by 0.4. Given that, results indicate that taxes have no statistically-significant effect on binge drinking, but minimum age laws do reduce the incidence of binge drinking. In fact, a one year increase in the minimum age would reduce the log of the odds of binge drinking by 0.11. This result also corroborates those of earlier works (Laixuthai and Chaloupka, 1993; Coate and Grossman, 1988). This result suggests that increases in the minimum age laws enacted in the 1980's were significant in curbing alcohol consumption. It is reasonable to assume that if minimum ages were rolled back, holding all other factors constant, binge drinking would increase at a statistically-significant and substantial rate.

Regarding other variables that have effects on binge drinking, WHITE, MALE, AGE, PEER and FIRST are positively and significantly related to binge drinking, while MARITAL and GRADE are negatively related to binge drinking. Two noteworthy results are as follows: First, it appears that if a person starts drinking later in life then that actually increases the probability that a person will engage in binge drinking; there is no a priori reason for this result. Second, the influence of peers, while having no statistically-significant effect on an individual's decision about drinking alcoholic beverages, does have a very significant effect on an individual's decision about whether or not to binge drink. This result is even more interesting given the fact that the average age of the sample used in the present study is 24, with a minimum age of 17 and a maximum age of 33. This result indicates that teens are not the only individuals subject to peer pressure in alcohol-laden environments.

4. CONCLUDING REMARKS

The present study extends the work of earlier researchers examining the role of taxes and minimum legal ages on the consumption of alcoholic beverages. The purpose of the present study was to determine the effect that taxes and minimum age laws have on alcoholic beverage demand and binge drinking. Using NLSY-Geocode and state-level tax data, the present study is one of the first studies to link individual-level data with the appropriate state-level alcohol tax rate. OLS and logit regression analyses were used to estimate individual-level alcoholic beverage demand equations. Results of the present study indicate that taxes have a negative effect on alcohol consumption but no effect on binge drinking. In fact, results suggest that alcohol demand is relatively elastic with regards to taxes. In addition, minimum age laws, by reducing the net return of risky behaviors, reduce both the total quantity of alcohol consumed and the incidence of binge drinking. These results corroborate the results of earlier studies.

Table 1

States Included in Sample

Alaska	Arizona	Arkansas
California	Colorado	Connecticut
Delaware	Florida	Georgia
Hawaii*	Illinois	Indiana
Kansas	Kentucky	Louisiana
Maryland	Massachusetts	Minnesota
Missouri	Nebraska	Nevada
New Jersey	New Mexico	New York
North Dakota	Oklahoma	Rhode Island
South Carolina	South Dakota	Tennessee
Texas	Wisconsin	

*Hawaii was excluded from youth sample

Table 2

Descriptive Statistics

Variable	Mean	Standard Deviation
BINGE	0.509	0.5
MARITAL	0.3091	0.4623
WHITE	0.7417	0.4378
MALE	0.5239	0.4995
INCOME	8644	10256
AGE	24	6
GRADE	12	2
URBAN	0.8495	0.3576
SOUTH	0.3096	0.4625
PEER	0.1659	0.3721
FAMILY	0.4306	0.4953
FIRST	14.6	5.3
MINAGE	20	1.2
TAX	0.0563	0.0336
PRICE	120.37	27.152

N = 2067

Table 3

Alcoholic Beverage Demand Regression Results

Variable	Coefficient	Test Statistic
CONSTANT	18.595	9.538**
MARITAL	-0.277	-4.004**
WHITE	0.403	6.292**
MALE	0.718	12.67**
LINCOME	0.0212	2.316**
LAGE	0.209	0.325
URBAN	-0.308	-0.409
SOUTH	-0.127	-2.128**
PEER	-0.101	-0.14
FAMILY	0.0747	1.334
LMINAGE	-1.59	-2.67**
LTAX	-2.751	-3.798**
LPRICE	0.163	2.987**

Note:
R^2 = .342
F=89.00
* = denotes variable is significant at 95% level
** = denotes variable is significant at 99% level

Table 4

Binge Drinking Regression Results

Variable	Coefficient	Test Statistic
CONSTANT	-1.016	-0.375
MARITAL	-0.664	-5.146**
WHITE	0.4169	3.499**
MALE	1.0705	10.323**
INCOME	0.0000083	1.347
AGE	0.4135	1.879
AGE2	-0.00844	-1.988*
GRADE	-0.1368	-4.705**
URBAN	-0.0427	-0.305
SOUTH	-0.2027	-1.438
PEER	0.4489	3.293**
FAMILY	0.1206	1.15
FIRST	0.0569	5.17**
MINAGE	-0.112	-1.984*
TAX	-0.717	-0.383
PRICE	-0.0112	-0.947

Note:
Chi-Squared = 502.08
* = denotes variable is significant at 95% level
** = denotes variable is significant at 99% level

REFERENCES

Adrian, M. and Fergus on, B. "Demand for Domestic and Imported Alcohol in Canada." *Applied Economics* 19: 531-540, 1987.

Coate, D. And M. Grossman. "Effects of Alcoholic Beverage Prices and Legal Drinking Ages on Youth Alcohol Use." *Journal of Law and Economics* 31: 145-171, 1988.

Duffy, M. "The Influence of Prices, Consumer Incomes and Advertising upon the Demand for Alcoholic Drink in the United Kingdom: An Econometric Study." *British Journal of Alcohol and Alcoholism* 16: 200-208, 1981.

Gao, X., E. Wailes, and G. Cramer. "A Microeconomic Model Analysis of US Consumer Demand for Alcoholic Beverages." *Applied Economics* 27: 59-69, 1995.

Heien, D. And G. Pompelli. "The Demand for Alcoholic Beverages: Economic and Demographic Effects." *Southern Economic Journal* 55: 759-770, 1989.

Johnson, J. and Oksanen, E. "Socio-Economic Determinants of the Consumption of Alcoholic Beverages." *Applied Economics* 6: 293-301, 1974.

Laixuthai, A. And F. Chaloupka. "Youth Alcohol Use and Public Policy." *Contemporary Policy Issues* 11: 70-79, 1993.

Lee, B. and Tremblay, V. "Advertising and the U.S. Market Demand for Beer." *Applied Economics* 24: 69-76, 1992.

Mast, Brent, Bruce Benson, David Rasmussen "Beer Taxation and Alcohol-Related Traffic Fatalities." *Southern Economic Journal* 66: 214-249, 1999.

McCornac, D. C. and Filante, R.W. "The Demand for Distilled Spirits: An Empirical Investigation." *Journal of Studies on Alcohol* 45: 176-178, 1984.

Pacula, R. "Does Increasing the Beer Tax Reduce Marijuana Consumption." *Journal of Health Economics* 17: 557-585, 1998.

Simon, J. L. "The Effect of Advertising on Liquor Brand Sales." *Journal of Marketing Research* 6: 303-313, 1969.

Thies, C. And C. Register. "Decriminalization of Marijuana and the Demand for Alcohol, Marijuana and Cocaine." *Social Science Journal* 30: 385-399, 1993.

Uri, J. "The Demand for Beverages and Interbeverage Substitution in the United States." *Bulletin of Economic Research* 38: 77-85, 1986.

ENDNOTES

1. Binge drinking is defined as having 6 or more drinks on one occasion.
2. TAX was calculated in the following fashion. First, the state-level tax per gallon for the three types of alcoholic beverages (beer, wine, and liquor) were obtained from a variety of sources. Second, average price data for each of the three types of alcoholic beverages were obtained from the Bureau of Labor Statistics. These data were only available on a regional level. It was assumed that this would be a suitable proxy for the state-level average prices of alcoholic

beverages. The earliest year for which this data is available is 1995. Since the focus of the present study is on the years 1994 and 1982, the 1995 average price data was deflated to 1994 and 1982 prices using the appropriate CPI for alcoholic beverages. Third, in order to obtain the state-level tax rate in percentage terms, the state-level tax per gallon is divided by the state-level price; this procedure results in a tax rate in percentage terms for liquor, beer, and wine. In order to obtain a single tax rate for alcoholic beverages, a weighted average is then taken of the three tax rates, with the weights being their share of consumption.

3. Income was deflated using the CPI - All Urban Consumers, base year 1982-1984.
4. The Consumer Price Index for alcoholic beverages is used as a proxy for the price of alcoholic beverages.
5. Wine tax rates for 1982 and 1995 were extrapolated from 1994 data.

THE INFLUENCE OF ILLIQUID ASSETS ON PRICES

Michael N. Baur*
Omar Benkato**
Socorro M. Quintero***

ABSTRACT

There are fundamental differences between assets held by closed and open-end funds. Past research indicates that closed-end funds are less liquid than open-end funds. For example, a larger percentage of closed-end funds hold less liquid international securities. This paper surveys the existing evidence of the impact of illiquid assets on the prices of open and closed-end funds. Specifically, it surveys existing literature on whether the price discount typically observed on closed-end funds is the expected result given the inescapable costs associated with managing relatively illiquid assets in closed-end funds. This paper synthesizes the various observations into one coherent theme: illiquid assets appear to be a significant source of the differences in the behavior we observe between open and closed-end funds. For a broader audience, this sheds light on valuation issues related to illiquid assets by looking at the case study of closed-end funds.

INTRODUCTION

This paper reviews the existing literature on the pricing differences between open and closed-end funds. After reading this literature, our conclusion is that illiquid assets appear to be a significant source of these pricing differences. In addition to the long-standing issue of why closed-end funds do not sell at net asset value, the issue and valuation problems of illiquid assets are important to a wider audience. This is the reason for our emphasis on the problem of pricing illiquid assets, and what we can learn about this topic from the closed-end fund literature.

The structure of the paper is as follows. First, we provide a brief history and comparison of open and closed-end mutual funds. Second, we draw the connection between manufacturing corporations and closed-end funds, and why the pricing problems of closed-end fund may be applicable to manufacturing firms. Third, we review the literature about illiquid assets with regard to closed-end funds. Fourth, we discuss illiquid assets and management costs. Fourth, we discuss illiquid assets and methods to collect transaction costs. Finally, we draw some conclusions.

*Department of Finance and Insurance, Ball State University, Muncie, Indiana 47306 U.S.A. Phone (765) 285- 5214 email: mbaur@bsu.edu

**Department of Finance and Insurance, Ball State University, Muncie, Indiana 47306 U.S.A.

***Department of Finance and Accounting, Oklahoma City University, Oklahoma City, OK 73106 U.S.A.
(Please address all correspondence to Michael N. Baur)

THE HISTORY AND FEATURES OF CLOSED AND OPEN-END MUTUAL FUNDS

Closed-end mutual funds were the first investment companies, and their history dates back to an investment trust created by King William I of the Netherlands in 1822 (Herzfeld, 1992). A closed-end fund is similar to a corporation that restricts the assets on its balance sheet to marketable securities only. Developing in Britain in the nineteenth century, these specialized corporations were called investment trusts, and from their beginning, they advertised their diversification services to small investors. Although many nineteenth century British investment trusts invested in American stocks, the first American investment trust was the closed-end Boston Personal Property Trust created in 1893. It was not until the 1920s that the U.S. experienced a boom in closed-end investment trusts.

The great bull markets of the 1920s -- and 1980s -- provided fertile soil for mutual funds. In their first incarnation and heyday, mutual funds played a central role in the robust stock market of the 1920s. After the 1929 stock market crash -- and in American history textbooks today -- these closed-end investment trusts were characterized as the "evil trusts" that manipulated the stock market and had a hand in causing the Great Crash of 1929. These charges added to the flourish of securities regulation that took place in the 1930s, which created the Securities and Exchange Commission (SEC).

Until the advent of "open-end" funds in 1924, all mutual funds were modeled after the traditional corporation. These new open-end funds distinguished themselves from the traditional corporate form by offering an innovative new feature: the continuous redemption of shares at net asset value (net asset value is defined as the book value of a firm's assets minus its liabilities). Although novel, the idea of redeemable shares did not catch on until the 1930s. In more recent times, however, the popularity of open-end funds has far outstripped the popularity of closed-end funds.

Besides the continuous redemption feature, two other features distinguish closed-end funds from open-end funds: the use of leverage and fixed capitalization. Open-ended funds stress simple equity-based capital structures, prohibitions on bank borrowing, and the provision of detailed audited financial reports at regular intervals to investors. Like a corporation, closed-end funds are frozen with a fixed amount of capitalization until they formally issue or redeem stock or bonds. Open-end funds, on the other hand, issue and redeem shares continuously at net asset value and can grow or decline quickly according to recent sales or redemptions.

CLOSED-END FUND SIMILARITIES TO MANUFACTURING FIRMS

Fundamental differences exist between open-end funds and closed-end funds. Arbitrage-based pricing theories have thus far suggested that the appropriate pricing model for closed-end funds should be one based on net asset value.[1] Since market prices exist for all the assets of a closed-end fund, arbitrage dictates that the price of a closed-end fund should at least equal its net asset value.

Arbitrage, however, is not the only method to value assets, and the prices of many assets do not match the prices suggested by arbitrage techniques. For example, because arbitrage is considered a poor valuation model for manufacturing corporations, few people expect the prices of manufacturing firms to equal book or net asset value. For example valuation techniques for manufacturing firms include future earnings, as well as asset values. Physical assets and financial assets present different valuation problems. However, when financial assets are illiquid, they begin to take on some of the characteristics of physical assets. The closed-end fund with illiquid assets begins to behave like a hybrid of a manufacturing firm and an open-end fund with liquid assets. Illiquid financial assets have some of the same characteristics as physical assets. As such, they should be valued based on their future earnings and how they are employed -- not strictly on the current value of the assets.

If closed-end funds share these similarities with traditional corporations, closed-end funds need different valuation methods than the pricing models used for open-end funds. In areas such as asset illiquidity, institutional organization and government regulation, corporations and closed-end funds have similar structures. Like traditional corporations, closed-end funds cannot be accurately valued by looking solely at net asset value. In contrast, open-end funds *can* be accurately valued and priced by looking solely at the net asset value as recorded on the fund's balance sheet. We suspect this is possible because of the unique characteristics of the assets in an open-end fund, and the legal mandate that these funds sell for their net asset value. In contrast, closed-end funds are more difficult to price because of their illiquid assets, more expensive management services, transaction costs, and uncertainty over asset values. This framework suggests that closed-end fund discounts are the expected result given the inescapable costs associated with managing illiquid assets.

THE CLOSED-END FUND LITERATURE ON ILLIQUID ASSETS

Malkiel (1977) and Anderson and Born (1987a and 1987b) report a positive connection between illiquid assets and closed-end price discounts. In a test of seven explanations of closed-end fund discounts, Malkiel found significantly positive correlations among four explanations. Two of these four explanations involved potentially illiquid investments: foreign stocks and restricted stock. Similarly, Anderson and Born (1987a) construct an illiquid asset index and a restricted asset index which are found to exhibit a significantly positive association with closed-end fund discounts. Patro (2001) finds evidence that the *risk-adjusted* performance of 45 international closed-end funds matchs the performance of respective local market indices. This is consistent with the notion of rational investors properly adjusting for (illiquidity) risk and using valid valuation and pricing fundamentals.

A concentration of illiquid assets in closed-end funds is also seen in Baur, Benkato and Sundaram (1994), who document statistical differences in the types of assets held by open and closed-end funds. Table 1 reproduces their results. First, this table shows -- at a 99 percent probability level -- that open-end funds hold different assets than closed-end funds. Second, the table suggests that open-end funds do not specialize. They hold "diversified" portfolios much more frequently than would be expected from a random selection. Closed-end funds, on the other hand, tend to hold far fewer "diversified" portfolios and far more non-diversified specialty portfolios (senior securities, international, government securities, etc.) than random selection would predict.

Table 1

Panel 1: Types of Assets in Open-End Funds

Observed versus Expected Frequency

	Observed	Expected
Diversified Funds	427 (66%)	338 (52%)
Senior Security Funds	94 (15%)	120 (19%)
Specialized Funds	53 (8%)	54 (8%)
International Funds	28 (4%)	53 (8%)
Gov't Security Funds	45 (7%)	76 (12%)
Dual-Purpose Funds	0 (0%)	6 (1%)
Total Number of Funds	647	647

Chi-square = 59*
Probability = 0.0001

Panel 2: Types of Assets in Closed-End Funds

Observed versus Expected Frequency

	Observed	Expected
Diversified Funds	23 (11%)	111 (52%)
Senior Security Funds	65 (31%)	39 (19%)
Specialized Funds	19 (9%)	18 (8%)
International Funds	40 (19%)	17 (8%)
Gov't Security Funds	57 (27%)	25 (12%)
Dual-Purpose Funds	8 (4%)	2 (1%)
Total Number of Funds	212	212

Chi-square = 177*
Probability = 0.0001

*Indicates that the observed frequencies are significantly different from the expected frequencies at a 95% confidence level.

The expected frequencies are calculated from the percentage of all funds in each Weisenberger (1990) category. For example, the expected percentage of diversified funds(52%) is calculated from adding the total number of open and closed-end diversified funds(427+23), and dividing by the total number of all open and closed-end funds(647+212). The expected number of, say, closed-end diversified funds(111) is then calculated by multiplying the total number of closed-end funds(212) by the expected percentage(52%). The categories include the following types of assets:

•Diversified funds include: maximum capital gain funds, long-term growth funds, growth and current income funds, balanced funds, and stock (and bond) income funds.

•Senior security funds are funds that concentrate in bonds and preferred stock.

•Specialized funds include technology funds, gold and precious metals funds, industry specific funds, and "other" funds.

•International and government security funds include international and government securities, respectively. Among the closed-end government security funds, 40 of the 57 are municipal bond funds.

•Dual-purpose funds are funds with two types of stock. Income shares receive all dividends, and capital shares receive all capital gains.

Source: Baur, Benkato and Sundaram (1994).

Other non-statistical evidence of differences between open and closed-end funds includes the composite balance sheets of these two different types of funds. For purposes of comparison between an individual firm and the industry average, Dun and Bradstreet provides composite balance sheets for many industries based on SIC codes. Table 2 presents these composite balance sheets for open and closed-end funds. Of particular note is the larger number and bigger size of open-end funds. On average, open-end funds are five times the size of closed-end funds, and three open-end funds exist for every closed-end fund.

ILLIQUID ASSETS AND MANAGEMENT COSTS

Rational investors should be expected to price the shares of closed-end funds *net* of management costs (and other expenses). This price may not be net asset value. Since management costs cannot be avoided by people who buy and sell individual securities to form their own portfolios, comparing closed-end fund share prices to net asset values is an inappropriate comparison. Investigating managerial costs and benefits, Chay and Trzcinka (1999) find that closed-end price premiums are positively correlated with future managerial performance. Akhigbe and Madura (2001) and Bers and Madura (2000) address the unique managerial characteristics of closed-end funds and why these characteristics may lead to a persistence of price performance.[2] Other researchers have documented a statistical relation between closed-end fund discounts and management fees, see Malhotra and McLeod (2000), Kumar and Noronha (1992), Anderson and Born (1987a, 1987b), and Crawford and Harper (1985). However, Malkiel (1977) found no relation between management fees and closed-end fund discounts.

Table 2

Composite Balance Sheets of Open and Closed-end Funds

	Open-end Funds (164 funds)		Closed-end Funds (57 funds)	
	$ (in thousands)	%	$ (in thousands)	%
Cash	4,348	7.6	502	4.7
Accounts Receivable	6,407	11.2	438	4.1
Notes Payable	57	0.1	214	2.0
Inventory	57	0.1	150	1.4
Other Current Assets	27,003	47.2	5,022	47.0
Total Current Assets	37,873	66.2	6,325	59.2
Fixed Assets	2,746	4.8	673	6.3
Other Non-current Assets	16,591	29.0	3,686	34.5
Total Assets	57,210	100.0	10,686	100.0
Accounts Payable	3,719	6.5	214	2.0
Notes Payable	457	0.8	737	6.9
Other Current Liabilities	6,464	11.3	1,047	9.8
Total Current Liabilities	10,641	18.6	1,998	18.7
Long-term Liabilities	3,947	6.9	449	4.2
Deferred Credits	458	0.8	53	0.5
Net Worth	42,163	73.7	8,185	76.6
Total Liabilities and Net Worth	57,209	100.0	10,685	100.0

Source: Dun and Bradstreet Credit Services, *Industry Norms and Key Business Ratios*, 1987-1988 Edition, p. 169. The SIC codes for open-end and closed-end investment companies are 6722 and 6723, respectively.

Brauer (1984) presents evidence that assets with uncertain market values require more management and transaction costs than assets with readily available market prices. If this illiquidity indirectly causes increased management and other transaction costs, this adds to the explanation of the pricing differences of open versus closed-end funds. If the different funds hold different assets, the services provided by open-end managers are different than the services provided by the closed-end fund manager.

Some investment companies hold extremely liquid asset portfolios. For example, unmanaged index funds advertise superior returns to shareholders from low management expenses and a random selection of representative issues. Because of free and competent oversight by a presumed efficient capital market, index fund promoters believe there is little diminution in the quality of stock selection from the more expensive stock selection techniques that employ costly portfolio managers. In this light, less extensive in-house oversight structures are needed for assets that have active markets and extensive market-based oversight. Indeed, Malhotra and McLeod (2000) find that U.S. closed-end funds -- which hold relatively liquid portfolios -- have notably lower expense ratios than foreign closed-end funds.

Traditional manufacturing corporations hold less liquid assets than open and closed-end investment companies. These investment companies (or mutual funds) limit their holdings to one asset (i.e., financial securities). Manufacturing corporations need extensive oversight structures to monitor their (illiquid) asset portfolios. If all physical assets had well-functioning liquid secondary markets like those for many financial assets, the concept of an unmanaged "index" corporation might be plausible. However, because of the lack of market-based oversight of most (illiquid) assets, there remains the need for extensive management oversight in traditional manufacturing corporations.

Given their relatively illiquid assets, closed-end funds need more expensive management oversight than open-end funds. Brauer (1984) presents evidence of this. Brauer compared the expense ratios of both closed and open-end funds and found statistically significant differences. Brauer's results are reproduced in Table 3. He found that closed-end funds had systematically higher expense ratios than open-end funds in his sample of 826 fund pairs during the 1965-1981 time period.

ILLIQUID ASSETS AND METHODS TO COLLECT TRANSACTION COSTS

At some cost, investors can duplicate the diversification, management, and brokerage services provided by an open-end fund manager. Taking into account these transaction costs, arbitrage pricing models suggest that an open-end fund should sell at a price close to net asset value. If arbitrageurs are less able to duplicate the management services of closed-end funds, then arbitrage pricing models suggest that closed-end funds should sell at a price further removed from net asset value. In other words, arbitrage will force the price of open-end funds to net asset value *plus transaction costs*. The arbitrage of the less liquid, less marketable assets of closed-end funds becomes more difficult and costly.

If one includes transaction costs, the statement that open-end funds trade at net asset value is misleading. A trade at net asset value implies arbitrage pricing and zero transaction costs. This statement creates an unrealistic benchmark when comparing the prices of open and closed-end funds. A better

Table 3

A Paired Comparison Test of the Hypothesis that Closed-End Funds have on Average the Same Expense Ratios as Similar Open-End Funds

Years	Number of closed-end fund and open-end fund pairs[1]	Average of closed-end fund expense ratio minus open-end expense ratio (in percent)	t-score	Significance level[2]
1965	42	0.0929	0.99	0.1631
1966	42	0.0612	0.76	0.2266
1967	41	0.0527	0.61	0.2741
1968	37	0.0549	0.62	0.2706
1969	39	0.1167	1.13	0.1322
1970	39	0.0664	0.82	0.2093
1971	36	0.1939	2.09	0.0218
1972	32	0.1322	1.29	0.1035
1973	51	0.2525	2.67	0.0051
1974	67	0.3199	2.92	0.0024
1975	66	0.2895	2.92	0.0024
1976	61	0.3236	2.50	0.0075
1977	56	0.3070	2.90	0.0027
1978	54	0.1863	2.02	0.0240
1979	57	0.2632	2.35	0.0112
1980	55	0.2744	2.01	0.0249
1981	51	0.2541	2.83	0.0034
1965-1981	826	0.2088	8.05	0.0001

[1]Number of closed-end funds for which a 'meaningful' expense ratio was reported in Weisenberger's *Investment Companies*.

[2]To reject the null hypothesis of no difference in expense ratios in favor of the alternative hypothesis that closed-end fund expense ratios are greater.

Source: Brauer (1984).

comparison would compare prices *net of transaction costs* for both open and closed-end funds. Including transaction costs for closed-end funds but not for open-end funds magnifies the perceived pricing differences between the two types of funds. According to the Investment Company Act of 1940, only open-end funds must sell for net asset value. This legal restriction, however, does not prohibit open-end funds from charging fees or loads in addition to net asset value. This load percentage varies. There are also no load funds. An open question is whether -- in the absence of the SEC regulation to sell at net asset value -- open-end funds would have developed alternative fee mechanisms to pay for transaction costs. The 1940 law largely codified the pricing practices of the 1930s. If effectively enforced, the 1940 law (and 1970 amendments) may have limited the development of new methods to collect transaction fees.

Discounts that fall within the bounds of transaction costs cannot be profitably exploited and may persist for long periods. For example, for funds charging loads on the date of the sample, the average load on 200 open-end funds was 5 percent of net asset value with a standard deviation of 1.5 percentage points. This suggests that transaction costs typically fell within a range of 3.5 to 6.5 percent for these funds. If these costs are similar for closed-end funds, discounts ranging up to 6.5 percent fall within the bounds of transaction costs. If one includes other fund management fees and annual 12b-1 fees, discounts in closed-end funds beyond this 6.5 percent could persist for long periods.

If the typical load or bid-ask spread for open-end funds is about 5 percent of net asset value, the ask price for an open-end fund would be $100, and the average bid price would be $95.[3] Because of these load or bid-ask spread charges, one could say that open-end funds are typically bought at net asset value and sold at a discount. In contrast, with a closed-end fund, an investor typically buys the fund at the typical 8 percent discount and pays a brokerage commission. When selling the closed-end fund, the investor will sell the fund at the same 8 percent discount and pay another brokerage commission. For example, to *sell* $100 of net asset value in a closed-end fund, an investor would get $90.50 (i.e., $100 in net asset value minus the typical $8 discount minus a $1.50 brokerage commission).[4] Also note that to *buy* a closed-end fund, the investor pays $93.50 ($100 - $8 + $1.50) to purchase $100 in net asset value. From an investor's perspective, it is ambiguous whether the selling price discounts of closed-end funds are preferable to high purchase prices and load charges of open-end funds. After transaction costs, neither fund is bought and sold at net asset value.

While some closed-end fund discounts are exploited through liquidation or open-ending, Herzfeld (1992) suspects that the cost of liquidating many funds exceeds the benefits. Liquidation requires terminating the management. This can be a costly procedure in view of management's control over the portfolio. Herzfeld (1992, p. 9) notes that "I have seen some funds . . . oppose open-end proposals for every cockeyed reason they can dream up. It is obvious that management is looking out for itself, not shareholders." Studies indicate that successful liquidations have occurred when discounts reached the neighborhood of 25 percentage points. Porter, Roenfeldt and Sicherman (1999) find that share repurchases -- when selling at a discount to NAV -- cause share prices to rise. This suggests that management intransigence can add significantly to transaction costs.

Interval Funds

A recent innovation to address the transaction costs of fund liquidation is the introduction of interval funds. This hybrid between an open and closed-end fund allows for shareholders to redeem shares at specified intervals. When shareholders are allowed to redeem their funds daily, funds cannot invest a large portion of their portfolios in illiquid assets. For example, daily liquidations effectively prohibit funds from investing in securities such as privately placed bonds and foreign stocks on thinly-traded exchanges. The traditional method to participate in these illiquid securities is through closed-end funds, since these funds are not obliged to redeem securities daily at net asset value.

Interval funds would allow shareholders to redeem fund shares at net asset value on a monthly or quarterly basis. This restriction on redemptions would allow funds adequate time to calculate an accurate assessment of net asset value and put less constraints on their investments in illiquid assets. These new interval funds allow investors to buy funds that invest in illiquid portfolios, while also allowing them to cash in their investment for their full net asset value.

CONCLUSIONS

This paper explores the consequences of closed and open-end funds having systematically different assets in their portfolios. Earlier studies have found statistically significant differences in the types of assets held by open and closed-end funds. These studies indicate that closed-end fund portfolios are illiquid relative to open-end fund portfolios. If this is true, it is not surprising that rational investors will pay different prices for (liquid) open and (illiquid) closed-end funds.

Open and closed-end funds collect for transaction and management costs using various methods. Open-end funds charge fees and loads. Closed-end funds charge fees. From the investor's standpoint, it is unclear whether the price discounts of closed-end funds are preferable to high purchase prices and extra load charges of open-end funds. After transaction costs, neither type of fund sells at net asset value. These transaction costs partially explain the anomalous and persistent discounts observed in closed-end funds. Specifically, the price discount typically observed on closed-end funds is consistent with the transaction costs associated with managing illiquid assets. A relatively new twist in fund innovation is the introduction of interval funds, which allow share redemptions at net asset value at specific intervals. As with other funds, the management costs of these new funds will depend on the types of assets these new funds hold. The collection methods for the fees needed to manage the fund will undoubtedly have an impact on the pricing of these new funds.

In contrast to open-end funds, corporations and closed-end funds have similarities in areas such as illiquid assets, institutional organization, and government regulation. By investigating the valuation issues of closed-end funds, we also shed light on the valuation of manufacturing and service firms. In particular, the presence of liquid or illiquid assets requires different valuation models and pricing fundamentals. Manufacturing and service firms and closed-end funds are more difficult to price because of their illiquid assets, more expensive management services, transaction costs, and uncertainty over asset values. If this is true, the pricing fundamentals of manufacturing and service firms -- and closed-end funds -- arise from the cost of managing illiquid assets.

REFERENCES

Akhigbe, Aigbe and Jeff Madura, "Motivation and Performance of Seasoned Offerings by Closed-End Funds, *Financial Review*, 36(3), 2001, pp. 101-22.

Anderson, Seth C. and Jeffery A Born, "Market Imperfections and Asset Pricing," *Review of Business and Economic Research,* 23(1), 1987a, pp. 14-25.

Anderson, Seth C. and Jeffery A. Born, "The Effects of Market Imperfections on Asset Pricing and Risk: An Empirical Examination," *Journal of the Midwest Finance Association,* 16, 1987b, pp. 1-17.

Anderson, Seth C. and Jeffery A. Born, *Closed-End Investment Companies: Issues and Answers*, Boston: Kluwer Academic Publishers, 1992.

Baur, Michael N., Omar Benkato and Srinivasan Sundaram, "The Mis-Measurement of Net Asset Value in Open and Closed-End Funds," *Revue Tunisienne d'Economie et de Gestion*, 9(12), 1994, pp 67-82.

Bers, Martina K. and Jeff Madura, "The Performance Persistence of Closed-End Funds," *Financial Review*, 35(3), 2000, pp. 33-51.

Brauer, Gregory A., "Open-Ending Closed-End Funds," *Journal of Financial Economics*, 13(4), 1984, pp. 491-507.

Chay, J.B. and Charles A. Trzcinka, "Managerial Performance and the Cross-Sectional Pricing of Closed-End Funds," *Journal of Financial Economics*, 52(3), 1999, pp. 379-408.

Chen, Nai-Fu, Raymond Kan and Merton Miller, "Are the Discounts on Closed-End Funds a Sentiment Index?," *Journal of Finance*, 48(2), 1993, pp. 795-800.

Chopra, Navin, Charles M. C. Lee, Andrei Shleifer and Richard Thaler, "Yes, Discounts on Closed-End Funds are a Sentiment Index," *Journal of Finance*, 48(2), 1993, pp. 808-808.

Crawford, Peggy J. and Charles P. Harper, "An Analysis of the Discounts on Closed-end Mutual Funds," *The Financial Review*, 20, August 1985, pp. 30-45.

Herzfeld, Thomas J., *Herzfeld's Guide to Closed-End Funds*, Englewood Clifts, NJ: Prentice-Hall, 1992.

Kumar, Raman and Gregory M. Noronha, "A Re-examination of the Relationship Between Closed-End Discounts and Expenses," *Journal of Financial Research*, 15(2), 1992, pp. 139-147.

Malhotra, D.K. and Robert McLeod, "Closed-End Fund Expenses and Investment Selection," *Financial Review*, 35(1), 2000, pp. 85-104.

Malkiel, Burton G., "The Valuation of Closed-End Investment Company Shares," *Journal of Finance*, 32(3), 1977, pp. 847-859.

Patro, Dilip Kumar, "Measuring Performance of International Closed-End Funds," *Journal of Banking and Finance*, 25(9), 2001, pp. 1741-67.

Porter, Gary E., Rodney L. Roenfeldt and Neil W. Sicherman, "The Value of Open Market Repurchases of Closed-End Fund Shares," *The Journal of Business*," 72(2), 1999, pp. 257-76.

Tinic, Seha M., and Richard R. West, *Investing in Securities: An Efficient Markets Approach*, Reading, MA: Addison Wesley, 1979.

Wiesenberger and Company, *Investment Companies*, New York: Warren, Gorham & Lamont Inc., 1990.

ENDNOTES

1. See Chopra, Lee, Shleifer and Thaler (1993), Chen, Kan and Miller (1993), Kumar and Noronha (1992), and Anderson and Born (1992) for a partial review of this literature.
2. A reviewer of this paper points out that asymmetric information may partially explain the differing behavior of closed and open-end funds. The paper does not investigate specifically the different asymmetric information consequences of liquid and illiquid assets. However, one would suspect that more significant asymmetric information problems are associated with illiquid assets. By their nature, illiquid assets are difficult to sell in secondary markets because on information compactness. That is, the buyers and sellers have different information about the assets.
3. The transaction costs to the fund manager of buying or selling more assets may occur on the front-end of the transaction or the rear-end of the transaction. If the ask price reflects market value, the bid price may reflect market value minus a commission. See Tinic and West (1979) for a discussion of the transaction costs incurred by brokers and dealers when making a market in financial or physical assets.
4. A 1.5% brokerage commission may be high or low depending on the dollar volume of the transaction. Not including negotiated discounts, typical full-service stock brokerage commissions (post 1989) range as follows:

Dollar Volume	Commission
Under $1000	$5 + 3% of Dollar Volume (minimum charge of $45)
$1000-$2000	$15 + 2% of Dollar Volume
$2000-$3000	$20 + 1.75% of Dollar Volume
$3000-$5000	$27.50 + 1.5% of Dollar Volume
$5000-$20000	$35 + 1.35% of Dollar Volume
$20000-$30000	$125 + 0.9% of Dollar Volume
Over $30000	$230 + 0.55% of Dollar Volume

RENT CONTROL AND ITS REFORM IN CHINA

Anthony Yanxiang Gu*

ABSTRACT

Rent controls still exist in mainland China and Western countries, and will continue for the foreseeable future. In China, rent controls have resulted in severe housing shortages, poor management and maintenance of the housing stock, and an increasingly heavy financial burden on the state. China has been reforming government rental housing for the last two decades, increasing rents and phasing out government subsidies, promoting home ownership and house commercialization. Western countries have been adjusting their rent-control and housing-subsidy policies over the last three decades.

Key words: *rent control, subsidies, privatization*

1. INTRODUCTION

Following the practice of the former Soviet Union and Eastern European socialist countries, China adopted a low rent-low salary policy in 1952, when the government stopped the wartime allowance system and started paying cash salaries to government employees. Rents for government-owned or public housing units were set at arbitrarily low levels, and were unrelated to the costs of building and maintaining the housing units. The policy caused serious problems--severe housing shortages, poor management and maintenance of the existing housing stock, an increasingly heavy financial burden on the state, and corruption among housing officials. Ironically, rent controls were not born in centrally-planned economies, but in free market economies, such as the United States and the United Kingdom. Rent controls were instituted in the U.S. and many European nations during World War I. The controls were liberalized after the war but were reintroduced at the start of World War II. In developing countries, rent controls were primarily confined to the European colonies during World War II. Since then, governments have used rent control to contain rents during periods of rapid urbanization, and to ensure affordable housing. Strict rent controls in these countries led to declines in new housing construction, deterioration in existing rental housing units, and poor services. To avoid these problems, the governments have replaced strict rent controls with moderate controls or decontrolled rents, have improved the ways in which they provide housing subsidies, and have been privatizing public housing.

This paper analyzes the disadvantages of rent control, reviews the origin of rent control and its evolution in Western countries, examines rent control and its economic consequences in China, and provides our views on rental housing reform in China.

Jones School of Business, State University of New York, 115D South Hall, 1 College Circle, Geneseo, NY 14454, Telephone: (716) 245-5368, Fax: (716) 245-5467, Email: gu@geneseo.edu

2. THEORETICAL ANALYSIS

Basic economic theory stipulates that any substantial price effect, i.e., setting a price ceiling below the equilibrium level, will yield a supply effect, which is manifested in lower quality and fewer rental units over the long run. Rent compensates owners and developers for the cost of providing housing units to tenants. Rent fluctuations signal the supply and demand conditions of the housing market; they affect future investment decisions of investors and developers and, consequently, the allocation of resources. The signaling function works only if rents are allowed to rise and fall as the balance between supply and demand changes. Rent controls distort housing markets by depressing the amount by which rents can rise in response to changes of supply and demand in the market.

Rent can be defined as the shadow price, P, of a unit of housing service, multiplied by the quantity of services, Q, generated per period (Moorhouse, 1987). Neither P nor Q is directly observable. "Housing services" include all the abstract attributes such as quality, comfort, and prestige. Rent is the revenue that landlords receive. Any change in rent caused by a change in the shadow price will be offset by the change in rent caused by a change in quantity supplied. Rent or revenue can therefore be represented by a rectangular hyperbola.

Rent control is not direct price control: it is revenue control. Thus, landlords are able to change P and Q. They can change Q by changing the temperature setting on the hot water heater, changing the frequency of garbage pick-up or floor waxing, etc. Rent is not simply the amount tenants pay for housing. There can be a substantial gap between what tenants pay and what landlords receive. This gap is primarily made up of search costs. Of course, tenants at the time rent control is imposed do not incur these search costs; all later tenants do. This does not mean that once somebody moves into a controlled unit, there will be no extra costs. Suppose there is a major change in one's life that would ordinarily cause one to move. With rent control, that move may not occur, i.e., tenants cannot optimally adjust their consumption of housing services to the change in life. In order to enjoy the artificially low rent, one pays in other ways.

Figure 1 shows the effects of rent control in a competitive market. We assume:

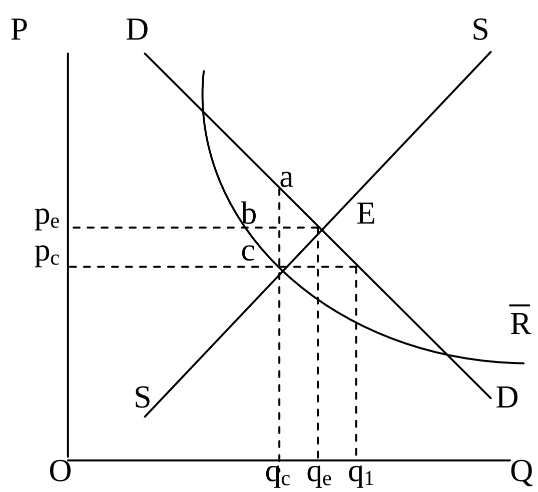
1) all housing services are homogeneous;
2) the rental market is perfectly competitive;
3) the market demand curve DD for housing services slopes downward; and
4) the market supply curve SS for housing services slopes upward.

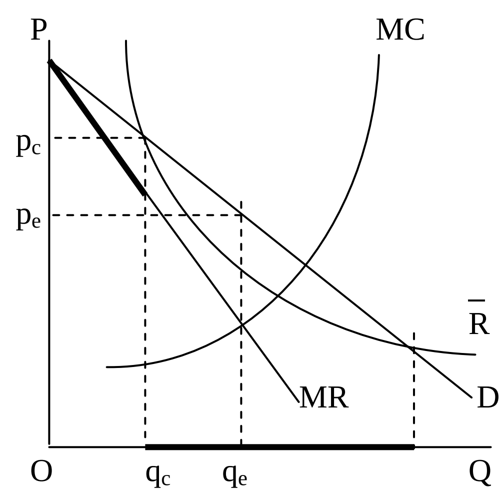
In the short run, SS may be perfectly inelastic, because we assume that no new suppliers can enter the rental market and that existing suppliers can only adjust the quality of their services. In Figure 1, the vertical axis indicates the shadow price level P per unit of housing services and the horizontal axis represents the quantity of housing services Q per period. The rectangular hyperbola R represents controlled rent. Rent control imposes a ceiling on rent; it does not control the shadow price of housing services. Without rent control, the market is in equilibrium at E with the shadow price of housing in equilibrium p_e and the quantity of housing services q_e. With strict rent control, or when the control is binding, the ceiling rent level is bound to shadow price $p_{c,}$ where the rectangular hyperbola crosses the supply curve at c and the quantity of housing services is reduced to q_c. p_c is below the equilibrium price p_e.

In equilibrium without rent control, rent revenue to the landlord is represented by the rectangle p_eEq_eO. Under rent control, the rent revenue is p_ccq_cO, which is smaller than p_eEq_eO. The landlord receives less revenue while the tenant receives less housing service, which means that rent controls impose net losses on society. These deadweight losses consist of aEb (loss in tenants' surplus) and bEc (loss in landlords' surplus). There will be excess demand, or a shortage of q_1 - q_c housing services, which can lead to discrimination against some tenants, more demand for public housing, and homelessness (Albon and Stafford, 1987). Rent controls transfer resources from landlords to tenants: The amount of this transfer is represented by the rectangle with area p_ebcp_c, which is the difference between p_eEq_eO and p_ccq_cO and bEc. Any increase in landlord costs will shift the supply curve leftward and lead to a further reduction in housing services to renters. In the long run the effects will be more obvious: the supply curve will tend to become more elastic because landlord have more options, such as converting space allocated to rental housing to other uses, and reducing services, for their strategic gaming.

Figure 2 shows the effects of rent control in a monopoly market, where MC is the marginal cost curve, and MR is the marginal revenue curve. Without rent control, the quantity of housing services supplied is q_e at shadow price p_e. Under rent control, the MR curve will become the bold segment, because the absolute value of the elasticity of a rectangular hyperbola equals to one, and the derivative of it with respect to quantity, or the marginal revenue, is zero. The quantity of housing services supplied is q_c at shadow price p_c: in comparison with a competitive market, a lower quantity of services is offered at a higher shadow price, and the deadweight losses to the society is larger. Hence, the effect of rent controls in a society with a monopoly market is even worse than the effect in a society with a competitive market.

Figure 1. **Figure 2.**





3. THE ORIGIN AND THE VARIOUS FORMS OF RENT CONTROL IN WESTERN COUNTRIES

Rent controls were instituted in the United States, the United Kingdom, France and other Western countries during World War I, and still exist today in various forms. In the U.S., rent control was first used in a few communities with severe housing shortages during World War I but was phased out gradually after the war. However, during World War II, as part of a general price control program, the federal government adopted nationwide rent control beginning in 1942. After 1949, federal rent controls were removed, and by the mid-1950s New York was the only state that retained rent controls. During the rapid inflation of the 1970s, many cities (e.g., Boston, Chicago, New York, Newark) readopted rent control, and, in 1971, the federal government imposed a 90-day price, wage, and rent freeze, followed by a flexible price-stabilization phase that lasted until January of 1973.

By 1986, more than 200 municipalities had some form of residential rent control. More than half of these were in New Jersey, one was the District of Columbia, and all the others were in the states of New York, Massachusetts, Connecticut, and California (Downs, 1988). By the middle of 1988, 14 states had passed laws or constitutional provisions against rent controls. As of 1990, rent controls were still in effect in more than 200 communities in the U. S. (Turner, 1990). The number has been stable since then.

Throughout the U.S., there have been two types of rent control--strict and moderate. Strict rent control which puts a virtual freeze on rents; was imposed during the two world wars and by New York City's rent control programs after World War II. The stabilization programs in Maryland, Massachusetts, New Jersey and Washington D.C. are considered examples of moderate rent controls. Moderate rent controls exempt new construction and "provide for annual rent adjustments to compensate for escalating costs and guarantee a fair and reasonable return on investment" (Turner, 1990). Moderate controls allow landlords to pass the escalating costs for maintenance on to their tenants. With hardship provisions, the landlords can petition for relief in cases of extraordinary cost increases or unacceptably low rates of return.

In the United Kingdom, rent control has existed since World War I. It is systematic and nationwide, and no decisions are made at the local level. The first rent control Act, which was in effect from 1915 to 1922, froze rents for most working-class dwellings at their pre-war levels. There was a period of rent decontrol from 1923 to 1939, during which rent increases were allowed either when the landlord had a vacancy or when the landlord had granted a lease fulfilling certain conditions. During World War II, this policy was abandoned, strict rent control was reintroduced, and it remained in effect from 1939 to 1957. Since then, the U.K. has made significant progress in rent decontrol, especially since 1980.

In the mid-1970s, Canada implemented strict rent controls in response to high inflation, and later tied rent increases to a price index. However, market rents are allowed for vacant units. The Socialist Government in France imposed strict rent controls in 1980, which caused a significant drop in housing construction in 1982 and 1983. In 1986, the new government abolished nationwide rent control, with the exception of high rent areas such as Paris. Finland currently allows landlords to adjust rents for certain cost increases, and both the landlords and tenants have the right to appeal. Germany and Greece have similar practices.

Rent control has also been common in developing countries. In Egypt, rent control was introduced in Cairo in 1944. Until 1952, rent controls were limited to houses built before 1944, in order to avoid discouraging housing construction. After 1952, rent controls were gradually extended to new construction (Malpezzi, 1986). India, where new construction is exempt for 10 years, has a two-tier system of controls. For units under strict control, the rent controller decides “fair” rent and has the power to allocate units to potential tenants. For units under ordinary control, rent increases are permitted but are regulated (Malpezzi and Tewari, 1991). Brazil first attempted to regulate the private rental market in 1917. Currently, rent controls in Brazil are generally less stringent than controls in other developing countries. Rent increases in Brazil are indexed to inflation, and are reset by negotiation every fifth year or when tenants change, instead of being controlled directly by legislation (Malpezzi and Ball, 1991).

Recognizing that strict rent controls can create net losses to the society, almost every country has reformed its control programs after the world wars and after severe inflation periods. Since 1980, privatization of public housing has been widespread. The U.S. Housing and Community Development Act outlined “right to buy” procedures for tenants in 1987 and the U.S. Congress passed President Bush’s Home Ownership and Opportunity for People Everywhere (HOPE) initiative in 1990. The British government has privatized over ten percent of its council housing since 1980 and, since then, the proportion of households living in state-owned housing has dropped from a peak of one third to under a quarter.

4. RENT CONTROL IN CHINA

The original purpose of rent control in China was to provide minimum and equitable housing for all employees and their families. In the mid 1950s, the central government set guidelines for urban housing rents at arbitrarily low levels that had little to do with the economic costs of housing production and housing service provision. Rents were determined based on the size of the housing unit with quality adjustments (Gu and Colwell, 1997). The local governments implemented the low rent policy. From 1952-1997, rent and utilities accounted for 1.77 percent to 3.7 percent of household income, while 20 percent to 30 percent was a typical proportion in the USA and Western Europe. Table 1 shows the percentage of household income spent on rent and utilities in China from 1952 to 1997. Because of the low rent-low salary policy, rents collected were not sufficient to cover maintenance costs, let alone the cost of new housing construction. In the early 1980s, the estimated average maintenance cost was about 0.17 *yuan* per square meter per month, while the prevailing monthly rent was 0.13 *yuan* per square meter. The estimated construction costs of building high-rise apartments ranged from 300 to 600 *yuan* per square meter (Fong, 1988a). Rent could only cover 25 percent of maintenance and management and related costs (Carlson, 1987). The cost of building could not be recouped.

Table 1. Rent and Utilities as a Percentage of Income

Year	Rent and Utilities	Year	Rent and Utilities
1952	3.7	1975	2.0
1953	2.8	1976	2.0
1954	2.8	1977	2.2
1955	3.2	1978	2.2
1956	2.8	1979	2.1
1957	2.1	1980	2.1
1958	2.4	1981	2.1
1959	2.3	1982	2.2
1960	2.1	1983	2.1
1961	2.4	1984	2.5
1962	2.5	1985	2.3
1963	2.6	1986	1.91
1964	2.6	1987	2.05
1965	2.2	1988	2.83
1966	2.1	1989	2.77
1967	2.0	1990	2.28
1968	2.0	1991	2.28
1969	2.2	1992	2.38
1970	2.1	1993	3.28
1971	2.3	1994	2.76
1972	2.3	1995	4.33
1973	2.1	1996	4.48
1974	2.1	1997	4.51

Source: 1952-1985, Zhongguo Chenzhen (China City and Town), Vol. 2, 1987, P.48;
1986-1997, Statistical Yearbook of China, 1991-1998.

As the theoretical analysis in Section 2 indicates, rent control causes a shortage of housing and housing services. The systematic rent control under the government monopoly has caused serious problems in China. Four of these are particularly relevant and well documented:

A. Severe housing shortages. A survey of 237 cities conducted in 1982 revealed that the average living space per urban resident was 4.4 square meters, even lower than the 4.5 square meters in the 1950s (Carlson, 1987). Throughout the 1980s, twenty-five percent of urban families had living space of less than 2 square meters for each member (Tolley, 1991).

B. Poor maintenance and massive deterioration in existing housing units. There were large numbers of dilapidated buildings in most cities until recently. In 1980, there were approximately 30 million square meters of dilapidated housing (Lin, 1989). In 1982, about 11 million square meters of residential housing space were demolished (Shang, 1986), further aggravating the shortage but improving average quality.

C. Reduced quality of housing accommodations. In cities, most households had to share a kitchen and a bathroom with at least one other household. Many families even had to use public toilets outside their buildings. In 1985, over 30 percent of all urban households had no kitchen or running water in their housing units, and about 70 percent of all urban households had no flush toilet in their housing units (Lin, 1989).

D. Heavy financial burden on the government. The state was solely responsible for urban housing construction and maintenance. The state also provided housing subsidies to all urban employees because their salaries were not enough even for the low rents. Consequently, as more public housing was built, more funds for subsidies were needed. Total housing subsidies, explicit as well as implicit, have been large. These are difficult to measure accurately because most of the subsidies are implicit, in the form of low rent. Although implicit subsidies are equal to the difference between market rent and actual rent, they are difficult to estimate for this period, because the market rental price was invisible. One study shows that the higher the occupational rank of the head of the household, the higher the implicit subsidy (Gu and Colwell, 1997), but no complete estimate of implicit subsidies has been found. We are left with two measurements: one measurement is the government's expenditure on housing construction; the other is the government's expenditure on maintenance. From 1950 to 1979, governmental investment in residential house construction was 37.4 billion *yuan*. In the 1980s, the state spent 24 billion to 36 billion *yuan* each year on housing construction and nearly 10 billion *yuan* on housing maintenance and rent subsidies (Lim and Lee, 1993).

5. RENTAL HOUSING REFORM IN CHINA AND ITS ACHIEVEMENTS

The serious housing problems have spurred widespread social discontent. For the last two decades, China has been reforming government rental housing as part of its broader economic reforms. The reform started with rent increases, intended to 1) at least cover the costs of construction and maintenance, 2) maintain the current level of new housing construction, 3) improve the quality of housing management and maintenance, and 4) promote home ownership. In 1990, the government started raising rents incrementally in Shanghai, Beijing, Tianjin, Shenyang, Guangzhou, and many other cities. In Shenyang, the goal was to raise monthly rent for a housing unit with a water supply, heat, a private kitchen, and bath to 1.98 *yuan* per square meter by the year 2000. As a first step, rent was increased from 0.15 *yuan* in 1991 to 0.27 *yuan* in 1993. Rent was increased to 0.38 *yuan*, in 1994, and thereafter by increments to reach the target rate of 1.98 *yuan* in the year 2000. Additional rent was charged for higher quality or special features. For example, a hard-wood floor would cost an additional 0.02 *yuan* per square meter per month; wall paper 0.01 *yuan*, and aluminum doors or windows 0.19 *yuan*. Finally, there was a surcharge if a family occupied more space than the officially established amount. As of May 1998, the average rent per square meter was 1.20 *yuan* in Shenyang, nearly as high as the average rate of 1.30 *yuan* in Beijing. Rent accounts for about 6 percent of household salary income.

Another reform measure made the implicit subsidies (through low rents) explicit. Rents were raised significantly and the subsidies became part of the regular income. In the pilot city of Yantai, monthly rent in 1987 was raised from 0.13 to 1.28 *yuan* per square meter, with a comparable increase of 23.5 percent of basic salary in the form of vouchers. If the value of the voucher was less than the rent, the household had to make up the difference; if it was more, the household could deposit the balance in an interest-bearing account and use it to pay for rent, housing construction and repairs, or the purchase of dwelling units (Fong, 1988b; Gu, 1988; Lim and Lee, 1993). In another city, Tangshan, rents were raised to 1.08 *yuan* per square meter, accompanied by a monthly housing subsidy of 24 percent of the basic salary.

To help low-income households, the government started the *Anju* [‘Secure Housing’] Project in 1995. Under the project, 150 million square meters of housing were built by 2000 on government grant land and sold to low-income households at cost. This may help to alleviate the problems of inequality and affordability, particularly for the many low-income households that do not have rationed public housing and are not able to buy even at the ration price.

Housing commercialization or privatization is the ultimate goal of the reform. The government initiated new housing sales in 1979; the buyer usually pays one-third of the construction cost, with the remainder being paid in equal parts by the State and the buyer's employer. The amount paid by the buyer goes to a fund that the State uses for new housing construction. Housing markets are regulated by municipal governments following the central government’s guidelines.

In 1995, a public Housing Saving Fund Program (HSFP) was started in most cities. Under the program, about 5 percent of an employee’s salary is set aside for the program. This amount is matched by the employer, and the state then contributes funds equal to two thirds of the combined contributions of the employee and employer. That is, if an employee contributes 30 *yuan*, the employer makes a matching contribution of 30 *yuan*, and the state provides 40 *yuan*. The funds are deposited in an interest-bearing bank account, and the employee may use the money in the account only for housing purchases within the city. A participant of the program can buy the public housing unit at a reduced price, and obtain a low rate mortgage loan from a state bank. We believe that the explicit subsidies and low-cost financing instituted by the program can alleviate the affordability problem. The program does not, however, solve the existing problem of lack of purchasing power.

Foreign investment in residential housing is encouraged, through tax deductions and exemptions. In Shanghai, as much as ten percent of the new construction in 1994 was funded by foreign investors. These units are for-profit projects, typically of higher quality than those built by the government and provide an alternative for the wealthy.

On May 15, 1998, the government decided to "stop the welfare housing ration" and announced that the "housing ration will all be commercialized” (*People's Daily*, 1998), which means that the government will no longer ration housing and households will purchase their housing in the free market. We believe that this represents the appropriate direction because free market economies have been proved to be more efficient than government controlled economies. Recently, rent increases have been accelerated in order to promote home ownership. However, the rent increase is not applied full-scale to laid-off households, which comprise over 30 percent of the households in industrial cities.

In recent years, more individuals have purchased their housing units under the HSFP, the reduced house price, the employer's match and government contribution, and the low rate mortgage loan proved to be helpful. By the end of 2000, around 60 percent of urban households had bought their formerly public apartments, while about 25 percent still lived in public rental units. Private rental markets have also emerged; about six percent of urban residents rent private housing. The average living space per urban resident reached 20 square meters. Starting on December 15, 2001, housing units purchased through government programs may be sold in the free market without limitation in Beijing. Shanghai had lifted the limitation in late 1997. Previously, homeowners had to wait at least five years to sell housing units that were bought under government programs. There has been no such limitation on housing units bought in the free market.

6. COMMENTS AND SUGGESTIONS

Raising rent is a prerequisite for market-oriented housing reform but it is not sufficient to accomplish full reform. Raising rent across the board will not transform the planned housing system into a market housing system. It may reduce price distortion to some degree since rents were artificially low, but it will still distort housing prices. Rent increases should be based on a proper estimate of rent determinants and available market data. For example, charging a flat rate per square meter does not reflect the market housing value of housing. Due to the declining marginal utility and marginal cost of additional space (the area occupied by walls does not, for instance, increase proportionally), increases in housing value are less than proportional to increases in area. Other factors being equal, if the value of a 65 m^2 apartment is 65 *yuan*, that of a 75 m^2 one should be less than 75 *yuan*.

To eliminate implicit rent subsidies and add the commensurate amount to wages is an appropriate reform because the welfare gain produced by a price subsidy can be achieved at lower costs by a cash transfer. However, in the Chinese system of state rent control, an employee's actual income includes both monetary income and implicit rent subsidies. Because the size of implicit rent subsidies varies according to occupational rank (Gu and Colwell, 1997), the effective housing price is not the same for all individuals. This means that the value of the different subsidies must be properly estimated. Since the right balance between rent changes and concomitant relative wage changes can be established as housing is privatized, an efficient housing allocation will be achieved.

The major problem in the housing market in China is the suppressed demand for housing units. Buying an apartment is still beyond the reach of many people. During the last few years, housing prices have been very soft, declining in most cites after sharp increases in the mid-1990s. While weak housing prices help make housing affordable, they reduce the investment value of owning one's home and make buying less attractive. This has caused high vacancy rates in many cities. The high vacancy rate of new housing units not only represents a huge waste of resources but also creates financial strains on the builders and financial institutions that financed these projects. Because state enterprises are laying off employees in the ongoing privatization, it is hard to envision that the government will raise rent sharply and that these laid-off workers will be able to buy their homes. Because it is politically impossible for the government to withdraw completely from the housing sector, some form of rent control, and some form of government subsidies, will remain in the future.

Secondary housing financial markets need to be developed. These will provide significant funds to the housing sector by offering a variety of investment instruments to individual and institutional investors. These financial services will create many employment opportunities, thereby increasing household income. This, in turn, will stimulate the development of the housing sector. Mortgage-backed securities will be feasible as the amount of mortgage loans increases. Bonds may be issued against the funds (provided by employee, employer, and State in a 30/30/40 ratio) in the public housing savings program. These savings accounts will provide secured sources for interest and principle payments to bond holders. Regular interest and principle payments from borrowers can be used to pay interest on, and to retire, the securities.

For these reforms to be effective, private ownership and property rights, such as right of use, right of possession, and right of disposition of the property, must be secured by law and law enforcement. Full ownership rights provide sufficient incentives to guarantee the success of privatization efforts. Existing property rights, which are ill-defined and unsecured do not encourage participation in housing privatization, because they do not secure an owner’s right to maximize expected future benefits. The private rental market, moreover, cannot function unless the right to evict non-paying tenants is secured by law and law enforcement. Property lacking the protection of these rights loses value, which in turn, jeopardizes housing privatization. Finally, there is still no provision for land ownership for private housing. Legally, the state owns all the land, homeowners rent the land from the government. This presents a major obstacle to housing privatization, and needs to be addressed by future reforms.

A variety of government-supported housing programs should be developed for low-income families. There will still be low-income families for the foreseeable future, and their housing problems could grow as housing is privatized. But the example of Western countries proves that housing programs entirely controlled by government are economically inefficient. Assistance programs that do not interfere with the market have been more efficient. For instance, governments have issued housing vouchers or coupons to the poor, who use them to pay rent to private landlords.

7. SUMMARY AND CONCLUSIONS

In China, urban housing rent control caused severe problems from the 1950s through the 1980s. The experience with rent control, decontrol, housing subsidies and housing privatization in both Western countries and non-communist developing countries can therefore provide points of reference as China continues to reform its housing system.

Housing reform in China has made undeniable achievements over the last 20 years. Private housing and rental markets are emerging, the cost of building and maintaining new housing units is no longer entirely borne by the government, more than 60 percent of households have purchased their homes, and living conditions are better than they have been during any period since the establishment of the People's Republic. However, China cannot eliminate rent control and the housing subsidy soon. Even in very affluent and developed countries, such as the United States or United Kingdom, there are sizeable populations living in rent controlled or subsidized housing.

We recommend appropriate estimation of rental determinants, development of secondary housing markets and sources of housing finance (such as mortgage-backed securities and public housing savings bonds) and legal guarantees of the security of private ownership. Because some forms of rent control and housing subsidies for low-income households will continue to exist, effective and efficient rent control and housing subsidy policies remain an important issue for further research.

ACKNOWLEDGMENTS: the author thanks the anonymous reviewer, Editor William O'Dea for helpful comments and suggestions, Peter Colwell for very helpful inspiration, suggestions and comments, and Zeng Ping, Zhu Hua for data.

REFERENCES

Albon, R. and Stafford, D. (1987). *Rent control*, New York: Croom Helm.

Carlson, E. (1987). China achieves record housing. *Habitat International* 11(4), pp. 216-67.

Downs, A. (1988). *Residential rent controls: An evaluation.* Washington, D.C.: The Urban Land Institute.

Fong, P. K. W. (1988a). The commercialization of housing in a socialist state: An attempt to solve China's urban housing problem. *Planning Quarterly* (Journal of NZPI) 89, pp. 32-36.

Fong, P. K. W. (1988b). *Housing problems in China: A case study of the management of multi-story residential developments in Shanghai*. ACSP 30th Annual Conference at Buffalo, N.Y., pp. 27-30.

Gu, R. (1988). Comparisons and analyses of housing reforms from ten cities. *Urban and Rural Construction* (Chengxiang Jianshe) 4, pp. 9-11.

Gu, A., & Colwell, P. (1997). Urban housing rent and occupational rank in Beijing and Shenyang, China. *Journal of Property Research* 14 (2), pp. 133-143.

Lim, G. C., & Lee, M. H. (1993). Housing consumption in urban China. *Journal of Real Estate Finance and Economics* 6, pp. 89-102.

Lin, H. (*1989*).* Better housing construction and improving the living conditions for urban residents. *Urban and Rural Construction* 10, pp. 4-5.

Malpezzi, S. (1986). *Rent control and housing market equilibrium: Theory and evidence from Cairo, Egypt.* Unpublished doctoral dissertation, The George Washington University, Washington, D.C.

Malpezzi, S., & Ball, G. (1991). *Rent control in developing countries*. World Bank Discussion Paper 129.

Malpezzi, S., & Tewari, V. (1991). *Costs and benefits of rent regulation in Bangalore, India.* INURD 82.

Moorhouse, J. C. (1987). Long-term rent control and tenant subsidies. *Quarterly Review of Economics and Business* 27, 6-24.

People's Daily (1998). (Overseas Edition), May 16th.

Statistical Yearbook of China, 1991-1998.

Tolley, G. S. (1991). Urban housing reform in China: an economic analysis. *World Bank Discussion Papers* 123, Geneva, Switzerland: World Bank.

Turner, Margery A. (1990). *Housing market impacts of rent control: The Washington D.C. experience.* Washington D.C.: The Urban Institute Press.

REFEREES

NEW YORK STATE ECONOMICS ASSOCIATION (NYSEA)
55th ANNUAL CONVENTION
FINAL PROGRAM

Federal Reserve Bank
Buffalo, New York
October 18 & 19, 2002

Friday, October 18
7:00-10:00 PM **NYSEA Convention Wine and Cheese Reception**
Buffalo Branch, Federal Reserve Bank of New York

Saturday, October 19
8:15-9:00 AM **Convention Registration & Continental Breakfast**

CONFERENCE PROGRAM

9:00 – 10:30	Concurrent Sessions: Group I
10:30 – 10:45	Morning Break
10:45 – 11:45	Concurrent Sessions: Group II
12:00 – 1:30	Lunch
1:45 – 2:45	Concurrent Sessions: Group III
2:45 – 3:00	Afternoon Break
3:00 – 4:00	Concurrent Sessions: Group IV
4:15 – 5:30	Business Meeting of the Board of Directors in the Buffalo Branch Boardroom (open to all members)

(Session codes A,B,C, D correspond to room assignments)

9:00 – 10:30 a.m.
Concurrent Sessions: Group I

(I-A) Macroeconomics
Chair: Doris Geide-Stevenson, Weber State University

Richard Skolnik (SUNY Oswego)
"Inflation, Operating Returns and Money Illusion"
Discussant: Doris Geide-Stevenson, Weber State University

Michael McAvoy (SUNY Oneonta)
"How Were the Federal Reserve Locations Chosen?"
Discussant: Laura Ebert, Marist College

Michael Jerzmanowski (Brown University)

" TFP Differences: Appropriate Technology vs. Efficiency"
Discussant: Arthur Gow, University of New Haven

(I-B) Economics in the Classroom
Chair: William O'Dea, SUNY Oneonta

William O'Dea (SUNY Oneonta)
"The Impact of Sample Size on Parameter Estimates in the Empirical Analysis of Student Performance"
Discussant: Elia Kacapyr, Ithaca College

Paul Romer (Hoover Institution and Stanford University)
"Electronic Experiments in the Classroom"
Discussant: Joseph Eisenhauer, Canisius College

(I-C) Economic Issues in New York State
Chair: Dale Tussing, Syracuse University

Terrence Kinal (University at Albany)
"A Report on the State of the New York State Economy from the New York Consensus Economic Forecast (NYCEF) and the New York State Establishment Survey"
Discussant: Jason Bram, Federal Reserve Bank of New York

Ramon Garcia (Federal Reserve Bank of New York – Buffalo Branch)
"Economic Restructuring in Upstate New York"
Discussant: Dale Tussing, Syracuse University

Larry Lichtenstein (Canisius College)
Mark Zaprowski (Canisius College)
"Valuing Professional Licenses and Practices in New York State"
Discussant: Richard Wall, Canisius College

(I-D) International Economics & Economic Development
Chair: Steven Onyeiwu, Allegheny College

Laura Ebert (Marist College)
"An Analysis of Financial Structure and its Impact on Financial Service Provision for Botswana"
Discussant: Steve Onyeiwu, Allegheny College

Steven Onyeiwu (Allegheny College)
"Determinants of Foreign Direct Investment in Africa"
Discussant: Wade Thomas, SUNY Oneonta

Radha Balkaransingh (University of Tsukuba, Japan)
"The Causal Nature of Public Infrastructure: Japan 1955-1993"
Discussant: Alfred Lubell, SUNY Oneonta

10:45 – 11:45 a.m.
Concurrent Sessions: Group II

(II-A) Finance
Chair: Richard Skolnik, SUNY Oswego

Thomas Kopp (Siena College)

“European Equity Markets: The Impact of Financial Flow Liberalization”
Discussant: Richard Skolnik, SUNY Oswego

Jae-Young Kim (SUNY Albany)
“Dynamic Asset Pricing in a Stable Paretian Class and Equity Premium Puzzle”
Discussant: Phillip Pfaff, Canisius College

(II-B) Applied Microeconomics
Chair: Kent Klitgaard, Wells College

Joseph Eisenhauer (Canisius College)
“The Shadow Price of Morality: An Experimental Simulation”
Discussant: John Dennis Chasse, SUNY Brockport

Pal Boring (Institute for Social Research, Norway)
“Vocationally Disabled Persons and Split Population Survivor Models”
Discussant: Kent Klitgaard, Wells College

(II-C) Labor Economics
Chair: Charles Callahan III, SUNY Brockport

Jonathan Schwabish (Syracuse University)
“Male and Female Wage Elasticities: An Exploration of Convergence”

(II-D) Student & Faculty Collaborative Research
Chair: Alfred Lubell, SUNY Oneonta

Benjamin Moody and Florence Shu (SUNY Potsdam)
“Critiques on Analytical Tools in Mankiw’s Principles of Microeconomics”
Discussant: Alfred Lubell, SUNY Oneonta

Alexi Harding and Elia Kacapyr (Ithaca College)
“IMF and the Jamaican Financial Sector”
Discussant: Amar Parai, SUNY Fredonia

David Decker and
Florence Shu (SUNY Potsdam)
“Changing World Economy: The Euro’s Role”
Discussant: Michael McAvoy, SUNY Oneonta

1:45 – 2:45 p.m.
Concurrent Sessions: Group III

(III-A) Medical and Biological Issues in Economics
Chair: Dale Tussing, Syracuse University

Dale Tussing (Syracuse University) and Martha A. Wojtowycz (SUNY Upstate Medical University)
"Peer influence in physician behavior: Method of obstetric delivery and diagnoses of dystocia and fetal distress"
Discussant: Martha Wotjowycz, SUNY Upstate Medical University

William Ganley (Buffalo State College)
“Early Evolutionary Economics and Mendel’s Revolution in Biology”

Discussant: Barbara Howard, SUNY Geneseo

(III-B) Issues in the New York State Economy
Chair: David Ring, SUNY Oneonta

Jason Bram (Federal Reserve Bank of New York)
"New York City's Economy: Before and After 9/11"
Discussant: David Ring, SUNY Oneonta

George Palumbo (Canisius College) and Craig Rogers (Canisius College)
"A Descriptive Analysis of Inner City Business in Buffalo: Findings of and Empire State Development Corporation Research Project"
Discussant: Jason Bram, Federal Reserve Bank of New York

(III-C) Labor Economics II
Chair: Robert Cunningham, Alma College

Xueda Song (State University of New York at Albany)
"Effects of Technological Changes on Experience-Earning Profiles with Endogenous Industry Choice"
Discussant: Charles Callahan, SUNY Brockport

Robert Jones (Skidmore College)
"Educational Earnings Premiums Since the 1970s"
Discussant: Robert Cunningham, Alma College

(III-D) Economics of the Family
Chair: Wade Thomas, SUNY Oneonta

Stuart Rosenberg, Dowling College
"An Economic Analysis of Child Custody Decisions"
Discussant: Florence Shu, SUNY Potsdam

3:00 – 4:00 p.m.
Concurrent Sessions: Group IV

(IV-A) Urban/ Regional Economics
Chair: Barbara Howard, SUNY Geneseo

Kent Klitgaard (Wells College)
"Environmental Racism in Cayuga County"
Discussant: Jason Bram, Federal Reserve Bank of New York

Robert Cunningham (Alma College)
"An Update on the Bailey Border Model"
Discussant: Terrance Kinal, University at Albany

(IV-B) Industrial Organization
Chair: Arthur Gow, University of New Haven

Arthur Gow (University of New Haven)
"Microeconomic Modeling and Analysis of a Simple Plant for Commodity Chemical Production"

Discussant: Joseph Eisenhauer, Canisius College

Lennart Erickson (Brown University)
“Informality, Firm Size and Growth”
Discussant: William O’Dea, SUNY Oneonta

(IV-C) Poverty and the Distribution of Income
Chair: John Dennis Chasse, SUNY Brockport

Elif Sisli (New York University)
“Inflation and Poverty”
Discussant: Paul Romer, Hoover Institution and Stanford University

John Dennis Chasse (SUNY Brockport)
“John R. Commons and His Gang: Their Attack on the Distribution of Income”
Discussant: Dale Tussing, Syracuse University

(IV-D) Issues in Applied Economics
Chair: Florence Shu, SUNY Potsdam

Doris Geide-Stevenson (Weber State University)
"Consensus on Economic Issues: A Survey of Republicans, Democrats and Economists"
Discussant: David Ring, SUNY Oneonta