WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

AN ECONOMETRIC POLICY MODEL OF NATURAL GAS*

by

Paul W. MacAvoy

Robert S. Pindyck

635 - 72

December, 1972

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

AN ECONOMETRIC POLICY MODEL OF NATURAL GAS*

by

Paul W. MacAvoy   Robert S. Pindyck

635 - 72   December, 1972

* This paper was presented at the Winter Meetings of the Econometric Society in Toronto, December, 1972. The research was funded by the National Science Foundation under Research Grant # GI-34936.

ACKNOWLEDGEMENTS

The results of this project would not have been possible without the support of our research assistants at M.I.T. We would like to extend our sincere appreciation to Robert Brooks, Krishna Challa, Ira Girshkoff, and Marti Subrahmanyam for their energetic and enthusiastic help in developing a computerized data base during the summer of 1972, and estimating and simulating the model over the past few months. They worked many long hours, and often under considerable time pressure.

We received considerable support from the National Bureau of Economic Research Computer Research Center in the use of the TROLL system for the estimation and simulation of the model, as well as the maintenance of our data base. We would like to thank Mark Eisner, Fred Ciarmaglia, Richard Hill, and Jonathan Shane for their help in using TROLL.

We would also like to thank Morris Adelman and Gordon Kaufman of M.I.T., and Robert Fullen and Wade Sewell of the Federal Power Commission for their comments and suggestions.

1. Reasons for a Policy Model

Much has been made of an energy "crisis" in the United States, and the need for planning to meet "needs" of consumers in the coming decades. Economists have not joined the legions of those proclaiming "emergency," except perhaps in the case of natural gas. In this industry, there is a "crisis" in the economist's sense of non-clearing of demands. There exists a shortage of natural gas in this country, to the extent that non-delivery on contract sales was approximately 3 percent of total sales of the interstate pipelines in 1971-1972, and 2 percent of national sales by all sources [4, Chapter 4]. This crisis has been precipitated by direct control of gas field contract prices by the Federal Power Commission in the 1960's [18]. Policy changes are needed to solve this "crisis."

One of the purposes of a policy model of natural gas will be to test that assertion. By building an econometric model of gas markets with explicit policy controls, and simulating over time, we expect to be able to examine the effects of past regulatory policy, as well as to perform experiments to find the effects of alternative regulatory policies. Finally, by performing simulations of the model into the future using different assumptions about what regulatory practices of the Federal Power Commission might be, we plan to derive a set of alternative forecasts for excess demand at both the producer and the wholesale levels.

Most previous econometric studies of natural gas have investigated either demand or supply of gas, but have neglected the simultaneous interaction of these two sides of markets. Balestra [2], for example, in his

classic study of the demand for natural gas by residential and commercial consumers, assumed a perfectly elastic supply curve for production. This assumption was probably justified during the 1950's and 1960's since production of gas for final consumers took place on an "as needed" basis from large stocks of reserves. This modeling approach would not continue to be valid during the 1970's, however, as total gas demand exceeds the constraints on production imposed by smaller reserve levels. The supply studies of Erickson and Spann [6] and Khazzoom [14], similarly, are certainly admirable attempts at defining and testing some of the relationships that exist in the gas industry, particularly those accounting for reduced reserve levels under price controls. But, to the extent that policies are changed in the future so that markets clear, and demand is once again observed, models of only the supply side of markets will be inadequate to represent the effects of policy. If the "crisis" is to be ameliorated, in other words, the production and reserve supply levels of the industry have to be analyzed as a simultaneous system.

The starting point for this model is a recent simultaneous equation study by MacAvoy [18] which focuses primarily on demonstrating how price regulation at the wellhead has led to reserve shortages and a consequent inability of field reserve markets to be cleared of excess demand. His model treats only the market for new reserves (i.e., at the wellhead) with the pre-regulation wellhead price acting as a demand variable that is determined in part by exogenous variables (population, income, all-fuels price index) that in fact determine the demand for ultimate gas consumption. It

serves to demonstrate that the production market, and excess demand in that market, can be at least roughly described by a simple simultaneous equation econometric model.

The policy model developed here treats simultaneously the regulation-induced disequilibrium in the field market for reserves (gas producers selling to pipeline companies at the wellhead price) and equilibrium in the wholesale market for production (pipeline companies selling to public utilities and to industrial consumers at different wholesale prices). The linking of these two markets is an important characteristic of the natural gas industry: production in the wholesale market is a determinant of pipelines' demand for gas reserves in the field market, and price on new reserve contracts in the field market is a determinant of pipeline delivery costs and thus price in the wholesale market. Disequilibrium in the wholesale market occurs when meeting total consumers demands would result in total production that would exceed the constraints imposed by reserve levels in the field market.

The structure of the model is described in some detail in the next section, as a set of econometric relationships between several aggregated market and policy variables. Variables which are endogenous to the field market include, on the supply side, non-associated and (oil) associated discoveries of reserves, extensions and revisions of reserves, and wells drilled. On the demand side, wellhead prices would be endogenous if demands were cleared, but after ceiling prices were set by the F.P.C. in the 1960's, this variable became an exogenous policy variable. Endogenous variables in the wholesale market include production of gas and wholesale prices for two

sectors, residential-commercial, and industrial. These are derived from a price markup equation linking field contract prices and transportation costs to wholesale prices in the residential and industrial sectors, and from demand equations in the two sectors. The field and wholesale markets operate to provide reserves, new production, and wholesale prices for new production, once ceiling prices for reserves are set by F.P.C. policy.

The model is estimated by pooling cross-section and time series data from the 1960's for the wholesale market, and for supply in the field market, over regional break-downs that are particular to each part of the model. Our method for pooling, and the estimation techniques as a whole, are treated in the third section. Section 4 describes the statistical results and the fifth section describes preliminary results from a simulation of the model for different price increase or "deregulation" policies of the Federal Power Commission.

2. Structure of the Model

The model is a policy model, since relationships describing field price regulation are introduced and serve to change the structure. Both the well-head price in the field market and the wholesale price of gas sold by pipeline companies are determined in a different way by regulation. When regulation does not interfere with the equilibrium of supply and demand, prices and quantities in both markets are endogenous variables. However, if the regulated ceiling price is lower than the equilibrium market price, we no longer observe both the demand curve and the supply curve (depending on the market) and price is exogenous in one or both sets of markets.

In the field market the supply variable is the change in reserves, ΔR, that can be "committed" to pipelines and local industrial buyers for production over a ten to twenty year period. The change in reserves comes about from additions in both non-associated gas, and that associated with oil reserves, that result from new discoveries and extensions and revisions of earlier discoveries.[1] The demand variable in the field market is the wellhead price that pipelines are willing to pay for new reserves. After 1961, however, this price variable became the exogenous F.P.C. ceiling price, since it was lower than the price that would have resulted from a market equilibrium (cf.[4] Chapter 4, and [10]).

In the wholesale market the demand variable is production of additional gas ΔQ for delivery to public utilities and industrial consumers as a function of wholesale delivery prices, the prices of alternative fuels, and economy-wide variables that determine the size of markets for energy. Because the wholesale market is oligopolistic,[2] there is no explicit "supply" variable, but rather wholesale prices are endogenous and determined by pipeline markups on the field price and the marginal costs of transmission.

---

[1] The findings are "supply" variables. There are other accounting definitions of reserves different from ΔR, however, that are important for assessing the stock available at any one time. The stock is equal to ΔR, less subtractions due to production and any net change in storage. The levels are affected by the existence of repressuring (which adds to reserves), field use, transportation losses, etc. A flow diagram for non-associated reserve levels in shown in Figure 1 (page 6). The same diagram would hold for associated reserves, the two being linked at Marketed Production, as shown in the diagram.

[2] There are actually several wholesale "markets" (just as there are several production markets) in different parts of the country that are geographically defined, and the degree of monopoly power in each market is different. This is explained in more detail in Section 3.

FLOW DIAGRAM FOR NON-ASSOCIATED RESERVES

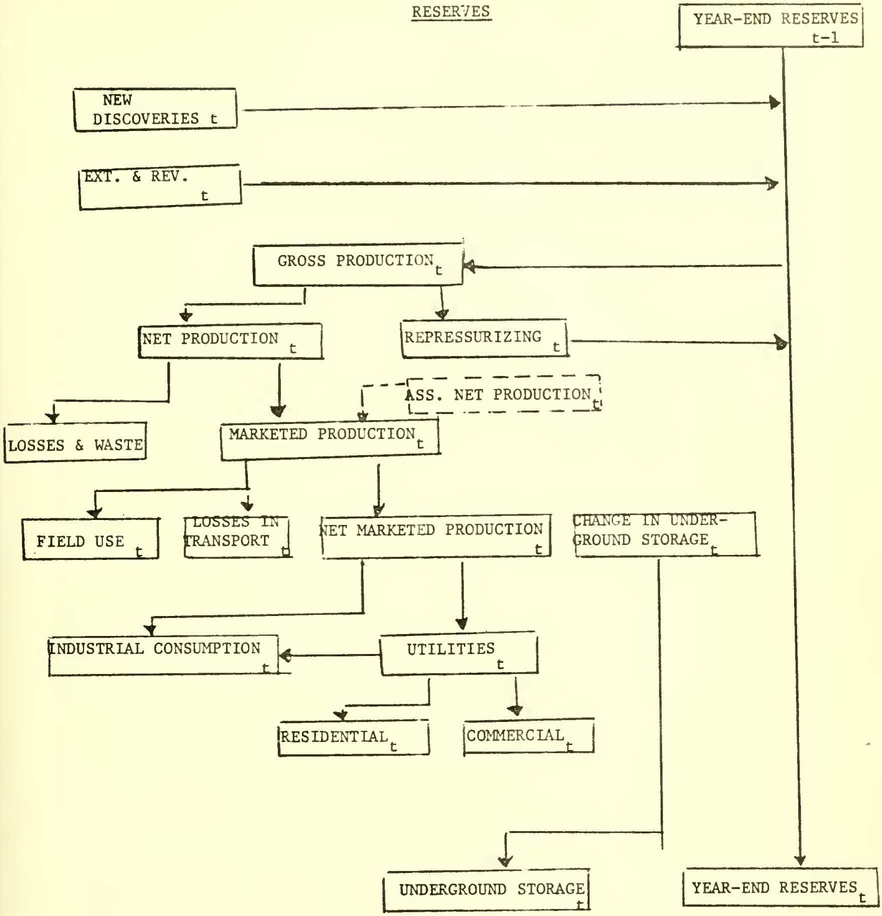


FIGURE 1

The quantity of gas sold must be consistent with the constraints imposed by existing reserve levels. Pipeline companies can only deliver additional gas based on the commitments of new reserves made to them by producers. These commitments are usually made for ten to twenty year periods, and therefore the increment of gas produced to meet new sales commitments can be only approximately one-tenth to one-twentieth of the change in the reserve level. If demand for production for an extended period exceeds this allowable level of extraction from new reserves, rationing has to take place (as has indeed been the case in recent years).

A. The Field Market

Probably the sector of the natural gas industry most difficult to capture in a conceptual model is the supply of new reserves. Most of the current controversy over regulatory policy centers on this sector and whether or not supply has been too low as a result of past regulatory policy. Actual additions to reserves through new discoveries are realized by a complicated process involving a large number of technological factors, and it may seem naive to try and model the process using a set of simple econometric relationships. Structural equations can be formulated, however, that do link several economic variables that are important in gas reserve additions and also describe most simply and directly the regulatory effects. Whether or not the econometric description is adequate can be judged in part from the logic of the interrelationships and in part from the statistical results.

A.1 Supply of Reserves

The major component of additions to reserves is new discoveries of both non-associated and associated gas (associated includes "dissolved" gas recovered from oil production, as well as "free" natural gas forming a cap in contact with crude oil).

The process of discovering gas begins with, and is driven by, the drilling of wells. Discovery is a process for obtaining productive inputs the same as any other oil and gas industry process. The results follow when there are sales of new reserves ΔR; to obtain these reserves at least costs, the exploratory firm minimizes costs -- expenditures on labor L and well drilling W -- subject to the production function $\Delta R = f(L,W)$ as a constraint. The least-cost combination of inputs results in supply $\Delta R = f(PG,PW,CW,X_i)$ where PG is the field contract price for gas, PW is the industry wage rate, CW is unit costs for wells, and X_i is a parameter i in the production function.[1] In modeling new discoveries specifically, however, we draw from and expand on the work of Erickson and Spann [6], Khazzoom [14] and MacAvoy [18].

There is a distinction between the supply curve of "exploratory effort" and supply curve of actual new discoveries (cf. Erickson and Spann). Economic incentives not only influence the amount of exploration but also determine

---

[1]Assume, for simplicity, that exploratory activity can be described by the Cobb-Douglas function $\Delta R = \alpha L^{\beta} W^{\gamma}$, so that the minimum cost combination of L,W is found by the first-order conditions for minimizing costs $C = P_L \cdot L + C_W \cdot W$ subject to $\Delta R = \alpha L^{\beta} W^{\gamma}$. The level of minimum marginal costs from these conditions is $\partial C/\partial \Delta R = f(\Delta R, P_L, C_W, \alpha, \beta, \gamma)$ and with the supply of reserves set at the level at which price of new reserves PG equals marginal costs $\partial C/\partial \Delta R$, then supply $\Delta R = f(PC, P_L, C_W, \alpha, \beta, \gamma)$. Here $x_1 = \alpha$, $x_2 = \beta$, and $x_3 = \gamma$.

its characteristics since an increase in incentives not only leads to more wildcat drilling but also drilling of poorer quality prospects. Any measure of the results of exploration such as the success ratio, SR (ratio of productive to total wildcats) or the average size of discovery S, is a function of both the distribution of prospects found in nature and the risk preferences of the "wildcatters." Thus, total new discoveries depend on (a) the number of wildcats drilled, (b) the success ratio, and (c) the average size of discovery per successful wildcat, and since the latter two factors decline with a rise in economic incentives, the response of new discoveries to such inventives will be less than the response of wildcatting itself.

The supply curve of wildcat drilling can be formulated, however, in terms close to the general characterization of the exploratory process. The "production function" process consists of using wells and labor to select from the prospects those able to provide lowest cost reserves ΔR. The prospects set the parameters of the production function. Since data on the characteristics of prospects is not available, it is necessary to use surrogates -- the characteristics of drilling in the preceding year. The values of the lagged variables W_{t-1}, SR_{t-1}, S_{t-1}, in a region are indicative of the year's prospects in that region, because they measure the information available to wildcatters regarding current results in that area. Drilling depends on the "incentive" variables -- costs per foot drilled, DC, as well as the price of oil, PO,[1] and the price of gas PG they expect

[1] As Khazzoom [14] points out, the price of oil comes in through two effects: directionality, i.e., gas discovered in the search for oil, as well as discovery of associated gas.

to receive.[1] With discovery wells designated as WXG for those proving out new fields as well as for extensions to the size of known fields,[2] and with the exact functional forms for the equation yet to be determined, we expect the following:

$$WXG_{t,j} = f(PG_{t,j}, PO_{t,j}, DC_{t,j}, SR_{t-1,j}, S_{t-1,j}, WXG_{t-1,j}) \quad . \tag{1}$$

Here t is the time index and j is the district index (a total of 19 Texas Railroad Commission and other substate or state districts are used to obtain a cross section, as discussed in Section 3).

The supply functions for reserves are formulated to show the results from drilling. In the case of non-associated gas, new discoveries (DN) are a function of the number of gas wells drilled for new discoveries, but also of the geological productivity of drilling in that region. As an indicator of these geological conditions, we have formulated the variable MM_j as the average level of discoveries per well in district j over the entire time period t. When this level has been large over the period, it can be inferred

---

[1]These in turn may be weighted distributions of past prices they did receive.

[2]In general, wells are classfied according to intent when they are drilled or by the result after drilling is completed (or abandoned). The classification according to intent by the American Association of Petroleum Geologists (AAPG) consists of drilling for new fields (new field wildcats), drilling for new pools (new pool wildcats, shallower-pool tests, deeper-pool tests) or to extend presently known pools (outposts). After drilling, wells are classified as "dry holes" if they are unsuccessful, or as new field or new pool-discovery wells if a new field or pool is discovered or extension wells if they extend the size of presently known pools. We center attention on new field extension wells, as representative of those that take the development of the field from the level of a single wildcat well to the level of reasonably known or "discovered" reserves.

that potential reservoir content has been greater there than elsewhere. New discoveries of non-associated gas, then, will have the form:

$$DN_{t,j} = f(WXG_{t,j}, MM_j, DN_{t-1,j}) \quad . \tag{2}$$

Depending on the relative prices of gas and oil, associated gas can be viewed as an input factor for producing oil or as a separate output. As Khazzoom indicates [14], if the price of oil is considerably greater than the price of gas, then gas is to some extent an input for producing oil (since gas pressure forces the oil to the surface, and the higher the gas pressure, the higher is the oil extraction rate). If oil prices were to be reduced relative to gas prices, the gas-to-oil production ratio should increase. In many regions, however, state conservation laws constrain the gas-to-oil production ratio to be within some range to conserve the "gas-input" resources so as ultimately to produce more oil. This effectively puts upper bounds on production. In this case, the supply of associated gas will be linked to the production, QO, of oil. On the whole, we would expect discoveries of associated gas to be more a function of oil production but to also depend somewhat on well drilling for oil that produces both oil and gas:

$$DA_{t,j} = f(QO_{t-1,j}, WXG_{t,j}, DA_{t-1,j}) \quad . \tag{3}$$

Extensions and revisions for both non-associated and associated gas are determined by previous discoveries, previous or present production of both gas and oil, and perhaps even prices (if prices induce further activities beyond drilling that add to reserves <u>in these categories</u>). Extensions and revisions for associated gas specifically should depend on the rate of

production of oil (QO), particularly in those regions that constrain the gas-to-oil production ratio. Our equations for extensions and revisions will thus be of the form:[1]

(Non-associated)

$$XRN_{t,j} = f(PG_{t,j}, \sum_j \Delta R_{t,j}, DA_{t-1,j}, WXG_{t-1,j} XRN_{t-1,j}) \tag{4}$$

(Associated)

$$XRA_{t,j} = f(PG_{t,j}, QO_{t,j}, \sum_j \Delta R_{t,j}, DA_{t-1,j}, WXG_{t-1,j}, XRA_{t-1,j}) \tag{5}$$

where prices are indicators of previous discoveries ($\sum_j \Delta R_{t,j}$, $DA_{t-1,j}$ and $WXG_{t-1,j}$) and previous extensions and revisions are included. For accounting purposes, the net change in reserves is identically equal to total new discoveries plus total extensions and revisions less marketed production (Q), losses (L) and change in underground storage ΔUS:

$$\Delta R_{t,j} = DN_{t,j} + XRN_{t,j} + DA_{t,j} + XRA_{t,j} - Q_{t,j} - L_{t,j} - \Delta US_{t,j}. \tag{6}$$

The cross-sectional index for the change in underground storage, ΔUS, corresponds to the production regions. Pipeline companies often store gas during the off-peak season in the vicinity of major consumers for eventual use to satisfy peak demand. Gas inventories thus serve as a substitute for peak-load transmission capacity. Although the gas is stored in the consumption areas (and the amount of storage is cost-limited by the extent of geological conditions favoring inground insertion of gas in those areas), an increase in inventories is a drain on the reserves of the production region which supply the gas. We model inventory change (as well as losses

---

[1] In the econometric work that follows, separate equations are fitted for extensions and for revisions. It should be noted that they all follow from this general formulation.

L) as a fixed proportion of production that is particular to the production region. For a monthly or quarterly model this would be unrealistic because of the seasonal fluctuation in inventory change that depends largely on the geographical and climactic conditions of the consumption area. Since our model is annual, though, these seasonal fluctuations are not important and it is only the long-term trend in inventory change that we need concern ourselves with.

A.2 Demand for Reserves

The demand for new reserves by pipelines is specified by an equation for the wellhead price, PG. Some time after 1961, however, the regulated ceiling price (PC) has to be below the price that would have resulted from a supply-demand equilibrium since otherwise the ceiling would not prevail. When ceilings became effective, excess demand became a necessary result and we no longer observe the demand curve after that point. This is shown graphically in Figure 2a (before ceilings) and 2b (after ceilings).

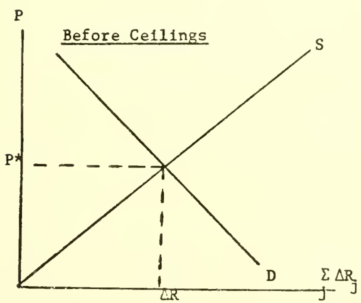


Figure 2a

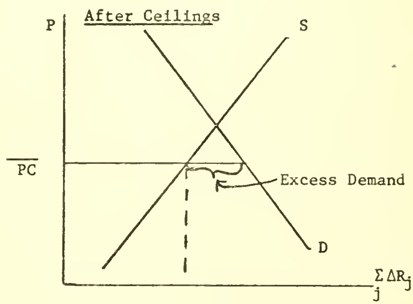


Figure 2b

Before ceiling price controls, the endogenous wellhead price in region j is a function of total new reserves $\sum_j \Delta R_{t,j}$ (as shown in Figure 2a), the change in reserves in "j" alone (because costs of transmission from j vary with the amount of reserves in that region), the average mileage (M) between the center of the production region and the distribution area (because costs of transmission vary with distance), and the total quantity of new production from the region $\Delta Q_{t,j}$. But the buying markets are not the producing regions "j." The districts have to be aggregated to regional "market" areas, i, each of which consists of a grouping of the production or supply districts j which are identified by the state conservation commissions and the F.P.C. to be geological units. The wellhead price, indexed as $PG_{t,j}$, in each of the nineteen production districts depends upon reserves there and is aggregated over i purchase markets,[1] as given by:

$$
\begin{aligned}
PG_{t,j} &= f(\textstyle\sum_{j=i} \Delta R_{t,j}, \Delta R_{t,j}, M_{t,j}, \Delta Q_{t,j}, PG_{t-1,j}); \quad t \leq \text{ceiling year} \\
&= PC_{t,j} \qquad ; \quad t \geq \text{ceiling year} .
\end{aligned} \tag{7}
$$

In some markets, even without price controls, demands would not be observable, however. Just after World War II, the few existing pipeline companies had monopsony power in gas field markets (cf. MacAvoy [17], as summarized by Erickson and Spann [6]). In a monopsony market, individual buyers have an effect upon the prices at which they buy, and as a result equilibrium price and quantity lie on the supply curve but not at the intersection with the demand curve as in Figure 2a. If a profit-maximizing pipeline is a monopsony buyer of gas, it will equate demand with its marginal expenditure

---

[1] This is to say that prices on individual wellhead sales at j are a function of the total amount sold in the market i made up of this group of regions j.

on gas so as to find ΔR^* -- but it pays no more than the supply price P* for this reduced volume (see Figure 3 below). Thus, the equilibrium price and quantity lie on the supply curve but below the demand curve.

Monopsony Market

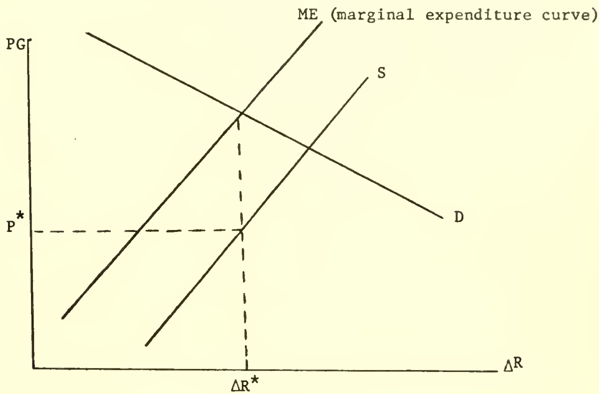


Figure 3

After the early 1950's, however, this condition changed as the network of pipelines grew extensively. Pipeline companies no longer enjoyed a monopsony market in gas fields, except perhaps in supply regions for gas to the West Coast in which El Paso Natural Gas Company was a very large source of demand. But to the extent that monopsony prevails, prices are determined exogenous to the system. Whether control is administered by the F.P.C. or by pipeline companies, prices $PG_{t,j} = PC_{t,j}$ in a "ceiling year."

B. The Wholesale Market

Pipeline companies that contract for reserve commitments from producers sell gas on the wholesale market directly to industry, and to local public utilities for resale to residential, commercial and industrial consumers. We have identified wholesale market regions across the continental United States, with each region representing a geographical area in which from one to several pipeline companies compete for consumer sales. This regional break-down is different from those of the field discovery and field purchase markets, primarily because wholesale buyers have access to pipelines traversing large geographical areas and they operate in "markets" covering large regions. Pipelines from the same field demand markets may feed into two or more regional wholesale markets, and pipelines from several field demand markets may all feed into a single regional wholesale market. (This regional break-down is discussed further in Section 3.)

Demands in these markets can be classified. There are three broad uses of natural gas by industry, and for each use the quality required of the gas (and thus the price the buyers are willing to pay) is different. Gas used for chemical processes should be of extremely pure quality and may be sufficiently unique to that process that there are few substitutes. A second use for industrial gas is for boiler fuel, and here the gas need not be very pure and competes with oil and coal. The third (and smallest) use of industrial gas is for electricity generation, and here too, the quality of the gas need not be very high nor the substitutability low, since this is again for boiler use. Contracts for industrial gas are also made

on either a firm or an interruptible basis. Firm contracts require that gas be supplied throughout the year at a more or less constant flow rate, while interruptible gas is supplied only in the off-peak season when there is excess capacity.

At the first stage of model construction, the three industrial gas demands are not classified separately. One reason for consolidating different demands is that it is difficult to obtain data on industrial gas sales broken down by class of user; pipeline companies must report to the F.P.C. gas sales to each industrial firm, but they do not report the ultimate use of the gas. It is necessary even at the first stage, on the other hand, to be able to separate "interruptible" from "firm" sales, particularly since the proportions of interruptible to firm sales are different in the different regional markets (the greater part of the firm industrial sales are made in the Southwest, close to field sources), so that industrial demand elasticities are not homogeneous across different regions. For now, this classification is made on the basis of treating the firm-interruptible ratio as one source of cross-sectional heterogeneity.

During most of the period before 1968, it would appear that wholesale markets cleared and there was no excess demand for gas by consumers.[1] The equilibrium quantity of gas sold by pipelines, and therefore produced from reserves, is determined by the intersection of the consumer demand curve

---

[1]After about 1968, wholesale markets no longer cleared and the demand curve is no longer observable. It can, however, be extrapolated from pre-1968 data and in this way we can obtain a measurement of excess demand. We shall return to this point later.

with a curve that essentially specifies the pipelines' mark-up on the field price. The price mark-up is a function of field purchase and regional transmission costs, and of the "competitiveness" of both gas sales and interfuel substitution.

B.1 Demand for Gas

In this model the demand for wholesale gas will be represented by equations that relate the quantity demanded to wholesale price, the price of alternative fuels, and "market size" variables such as population, income, and investment that determine the number of potential consumers. Two demand equations are estimated, one for pipeline sales to retail utilities and the other for pipelines sales to direct industrial buyers. In both cases, the dependent variable is new demand, δQ, rather than the total level of demand. In the short run, as Balestra has shown for residential gas [2], the level of demand should be relatively price inelastic and would depend on stock variables that do not change much in time (e.g., the total stock of gas-burning appliances for residential gas). New demand, however, should respond to the price of gas and to the price of competing fuels (decisions to buy new appliances, for example, are affected by fuel prices). The new demand for gas, δQ, is made up of the increment in gas consumption $\Delta Q = Q_t - Q_{t-1}$ and of replacement for continuation of old consumption. In any period, total residential and commercial gas demand could be considered to be a function of the stock of gas-burning appliances, A:

$$Q_t = \lambda \cdot A_t \qquad (8)$$

where λ is the (constant) utilization rate. Then, if r is the average rate at which the stock of appliances depreciates, the new demand for gas includes

$r\lambda A_{t-1}$. The total new demand is $\Delta Q + r\lambda A_{t-1}$, or

$$\delta Q_t = \lambda A_t - (1-r)\lambda A_{t-1} \ . \tag{9}$$

Now substituting (8) into (9) gives us:

$$\delta Q_t = Q_t - (1-r)Q_{t-1} \tag{10}$$

or

$$\delta Q_t = \Delta Q_t + rQ_{t-1} \ . \tag{10a}$$

Thus the new demand for gas is the sum of the incremental change in total gas consumption (ΔQ_t) plus the demand resulting from the replacement of old appliances.

Our _a priori_ assumption is that _new_ demand depends on prices and total income (through purchases of new appliances), but the _level_ of _total_ demand is itself a function of income and population. Thus, we have for residential and commercial demand:

$$\delta QSR_{t,k} = f(PSR_{t,k}, PF_{t,k}, Y_{t,k}, \delta Y_{t,k}, \delta N_{t,k}, \delta QSR_{t-1,k}) \tag{11}$$

where PSR is the wholesale price of residential and commercial gas, PF is a price index of competing fuels, Y is disposable income, N is population, in regions k, and

$$\delta Y_{t,k} = \Delta Y_{t,k} + rY_{t-1,k} \tag{12}$$

and

$$\delta N_{t,k} = \Delta N_{t,k} + rN_{t-1,k} \ . \tag{13}$$

Balestra [2] distinguishes between two depreciation rates, one for gas appliances and the other for alternative fuel-burning appliances, since

lifetime for appliances using alternative fuels may be different. He then estimates the two depreciation rates by estimating an equation of the form:

$$QSR_t = a_0 + a_1 PSR_t + a_2 \Delta N_t + a_3 N_{t-1} + a_4 \Delta Y_{t-1} + a_5 Y_{t-1} + a_6 QSR_{t-1} \quad . \tag{14}$$

The depreciation rate for gas appliances is then given by $(1-a_6)$. (His results, however, gave an estimated a_6 that was always greater than 1, which cannot be justified theoretically.) The all-fuel depreciation rate comes out of equation (14) as either the ratio a_3/a_2 or a_5/a_4. Thus, the equation is over-identified, and the depreciation rate can be obtained only by estimating (14) subject to the constraint of $a_3/a_2 = a_5/a_4$. (The resulting estimation problem is non-linear, but Balestra uses an iterative method suggested by Houthakker and Taylor [13] to obtain an estimated depreciation rate equal to 0.11.) The two depreciation rates needed here will be estimates in keeping with these procedures.

The demands of direct industrial purchasers should not differ greatly from those of utility purchasers shown in (11). But there are some distinctive variables. As mentioned above, sales of industrial gas are made on either a firm or interruptible basis, with the price for interruptible contracts considerably lower than that for firm contracts. Because in this model we do not disaggregate firm from interruptible sales and because the ratio of firm to interruptible sales is different in different parts of the country, we will introduce into our demand equation for industrial gas an exogenous variable, $FIR_{t,k}$, for the ratio of firm to interruptible sales of industrial gas during year t in wholesale market region k. Since our price variable for wholesale industrial gas $PI_{t,k}$, is an average of

firm and interruptible prices, we would expect the variable $FIR_{t,k}$ to appear in the demand equation with a positive coefficient.

Although some industrial uses of gas require a high level of chemical purity, on the average the cleanliness and ease of handling of industrial gas are not as important as in residential and commercial consumption. As a result, industrial gas can usually be sold only at prices equivalent to or lower than the price of competing fuels. We would expect, then, that the exogenous price index of competing fuels, PF, will be an important variable in the industrial wholesale demand equation.

As was the case for the residential and commercial sector, our demand equation for industrial gas will use new demand, δQI, as the dependent variable. We cannot expect the depreciation rate, r_I, however, to be the same as it was in the residential and commercial sector. Total industrial demand depends also on market size variables, such as population and income. But, since new gas in the industrial sector is for use in producing more output, it might be expected that industrial capital investment would be the most important determinant of new energy demands. Our demand equation will be of the form:

$$\delta QI_{t,k} = f(PI_{t,k}, PF_{t,k}, FIR_{t,k}, \delta_I N_{t,k}, \delta_I Y_{t,k}, K_{t,k}, \delta QI_{t-1,k}) \quad (15)$$

with

$$\delta QI_{t,k} = \Delta QI_{t,k} + r_I QI_{t-1,k} \quad (16)$$

$$\delta_I Y_{t,k} = \Delta Y_{t,k} + r_I QI_{t-1,k} \quad (17)$$

$$\delta_I N_{t,k} = \Delta N_{t,k} + r_I N_{t-1,k} \quad (18)$$

and $K_{t,k}$ is corporate capital investment in wholesale market k in year t.

B.2 Gas Pricing at Wholesale

With large numbers of connecting pipelines between producing districts and consumers developed during the 1960's, the industrial and utility buyers in the wholesale markets were faced with a variety of "supply" conditions. In some markets, there were one or two sources of new gas, so that prices were fixed by the pipelines to meet demands. That is, strictly speaking, there was no "supply function," but rather a price mark-up on field purchases set by costs and Federal Power Commission regulation. In other markets, there were more transporters of gas and the mark-ups were as limited by the conditions of inter-pipeline rivalry as by those conditions set by regulatory procedures. But even under extensive inter-pipeline competition, the process is the same: the wholesale prices are set by the pipelines equal to the wellhead price plus a mark-up proportional to the marginal costs of transmission and distribution. The "proportionality" factor varies by the extent of the rivalry, and the nature of regulation.

The marginal costs of transmission can be found, at least through an "averaging process" similar to that in computing regression equations. We assume to begin with that in an existing pipeline system average variable costs of transmission are constant up to some capacity, and then rise with infinite slope. This lets us net out the marginal cost of purchasing gas at the wellhead, as long as pipelines' contracts for new reserves can satisfy final demand for new production. Thus, there are two factors that can result in infinite marginal costs: limited transmission capacity, and a limitation on the ability of pipeline companies to deliver enough gas production from the available new contracts for

field reserves.[1] We take the marginal costs of transmission to be a function of mileage M_k, capacity of the lines transmitting from region k that year $v_{t,k}$, and the ratio of firm to interruptible sales $FIR_{t,k}$. Initially, we will estimate some very simple wholesale price equations of the form:

$$PRC_{t,k} - \overline{PG}_{t,k} = f(M_{t,k}, v_{t,k}, FIR_{t,k}) \qquad (19)$$

and

$$PI_{t,k} - \overline{PG}_{t,k} = f(M_{t,k} mv_{t,k}, FIR_{t,k}) \quad . \qquad (20)$$

Here $\overline{PG}_{t,k}$ is a "wellhead" price for region k determined by averaging the wellhead prices from those production regions which feed pipelines that transmit gas to that wholesale market k. The mileage variable M_k and capacity variable $v_{t,k}$ are similarly defined, i.e., in terms of the average values of these variables from production regions to the wholesale market region k. The "mark-up" is thus the excess of the price over field costs allowed by competition and the Commission, but as shown by the size of coefficients of the independent variables in marginal transmission costs. The "determined" price is approximated by the regression equation coefficients.

When the wholesale price is regulated to be a mark-up over the wellhead price and the marginal cost of transmission (as it is in the case of residential and commercial gas), excess demand can exist

---

[1]This is in the tradition of models of pipeline costs such as that of Wellisz [20], where transmission capacity is the limiting factor. Furthermore, in recent years, when there was excess demand in wholesale markets, it was new reserves that placed a limitation on wholesale production capacity.

in the wholesale market. This is shown in Figure 4. When demand increases from DD to D′D′, the wholesale price is regulated to be a mark-up over marginal costs (MC_f, the marginal cost of purchase in the field, is just the wellhead price, PG), and marginal costs become infinite at ΔQ_{max}, then there will be excess demand as shown. The question is, at what point ΔQ_{max} does the marginal cost curve become vertical? This occurs at the point at which new production is constrained by the lack of productivity from the given level of new reserves. Generally, we would expect to have

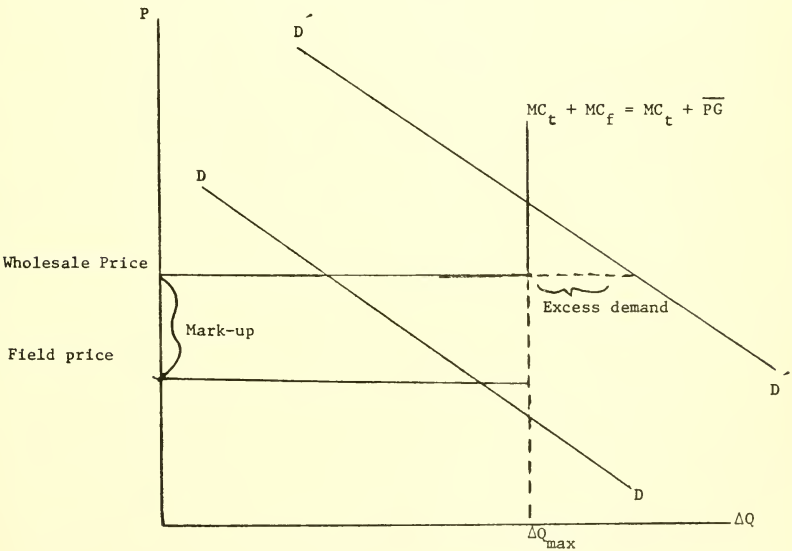


Figure 4

$$\alpha \Delta R = \Delta Q_{max} \tag{21}$$

with α set at a limit by the full flow characteristics of the new reservoir put under contract. The problem is that it is difficult to use past data to measure α since up until about 1968, at least, production demands were such that $\Delta Q < \Delta Q_{max}$. We will attempt, however, to use cross-sectional data over production regions in 1969 and 1970 to run regressions of the form

$$\Delta Q_{t,j} = \alpha \, \Delta R_{t,j} + \beta R_{t,j} \tag{22}$$

under the assumption that in those years $\Delta Q = \Delta Q_{max}$ (we will exclude those districts for which this was not believed to be the case).

Once we have a measure of ΔQ_{max}, we can extrapolate the demand curves estimated using pre-1968 data (since equilibrium production and price did not lie on these curves after 1968) to determine excess demand in recent years. In other words, we can simulate the model past 1968 (thus the demand curve $D'D'$ is just the demand curve DD simulated into the future) using a constraint on production:

$$\sum_k \Delta QSR_{t,k} + \sum_k \Delta QI_{t,k} = \Delta Q_t^{max} \quad ; \; t \geq 1968 \quad . \tag{23}$$

This, unfortunately, does not take into account the complicated interconnections between regional field markets and regional wholesale markets. Our constraint

$$\sum_k \Delta Q_k \leq \alpha \sum_j \Delta R_j \tag{24}$$

is actually somewhat incorrect. In fact, there should be a separate constraint for each wholesale market region k:

$$\Delta Q_k \leq \Sigma \frac{1}{\alpha_{i,k}} \beta_{i,k} \Delta R_i \qquad (25)$$

where the above summation is taken over those i's that feed into the particular k, and $\beta_{i,k}$ is the fraction of ΔR_i going to region k.

3. The Use of Cross-Section and Time-Series Data for Estimation

The equations of this model describe the dynamic equilibria of two distinct (but interrelated) groups of regional markets, namely, field markets and wholesale markets. The supply side of each regional field market is determined by the discovery and accumulation of new reserves in several different production districts. In fact, approximately 20 production districts determine the supply of gas reserves in 6 or so regional field markets. The demands for gas on the wholesale level are determined in about 5 different regional markets. As well, the field and wholesale markets are themselves interrelated. Thus the cross-sectional break-down will be different in the different sectors of the model.

The time-bounds over which equations can be estimated will also be different for different equations. This is the case for two reasons. First, the structure of the models changes over time. To begin with, the degree of monopsony and monopoly power held by pipelines in different regional markets has been changing in time. But even more fundamental is the fact that when regulation becomes effective (i.e., when the ceiling price is below the equilibrium price) the demand curve is no longer observable and new reserves are defined by the supply curve and the ceiling price. The second reason for the use of different time-bounds is that the

availability of statistical data is different for different variables. For many disaggregated variables, particularly related to field supply, data has been collected and compiled only after 1965; this imposes constraints on the time-bounds over which we can estimate particular sectors of the model, but not other sectors. The equations of this model, then, will be estimated by pooling cross-section and time-series data, but using cross-sections and time-bounds that are different for different parts of the model.

The differences in coverage and time-lengths is extensive. Table 1 contains a summary of the model's equations (identified by dependent variable) grouped according to cross-sectional break-down and time-bounds. The remainder of this section will explain the geographical break-down in more detail, and will also discuss some of the statistical aspects of pooling heterogeneous cross-sections and time-series.

A. Identification of Regional Markets

Although data are available for new reserves and field price for production districts, these districts do not by themselves constitute a "market" in the usual sense of the word, since many districts may be selling gas to the same group of pipeline companies. We define a "field market" as a distinct geographical region i where gas producers in all of the producing regions j in competition with each other sell reserve commitments to pipeline companies also in competition with each other as buyers.

As one would expect, the identification of regional field markets is not a straightforward matter. The AGA Reserve Committee has identified ten

Table 1:

THE MODEL'S STOCHASTIC EQUATIONS GROUPED BY

CROSS-SECTIONAL BREAK-DOWN AND TIME-BOUNDS

1. Cross-Section Break-down by Production District (j's)

Equations:	Time Bounds
Wells (WXG)	1954 - 1971
Additions to Reserves (DN, DA, XN, XA, RN RA)	1966 - 1971

2. Cross-Section Break-down by Field Markets (i's)

Wellhead Price (PG)	1954 to date ceiling price is effective (approximately 1961)

3. Cross-Section Break-down by Wholesale Market (k's)

Wholesale Demand (QRC, QI)	1954 - 1971
Wholesale Price (PRC, PI)	1954 - 1971

regional field markets, but these markets do not seem to be distinct, i.e., there are reserve commitments across markets. Therefore, we make an identification of six field markets based on geography, but also on judgments about where competition between districts does and does not exist. The two alternative classifications of field markets are summarized in Table 2. Estimation of the model's field market demand equations can be made using both classifications, although on the first round of construction only the M-P classifications are used; eventually, contrasting results will be used to help choose the most reasonable classification.

Regional wholesale markets are defined and identified in much the same way. We define a "wholesale market" as a distinct geographical region where pipeline companies in competition with each other sell gas to public utilities and industrial consumers who also are in competition with each other as buyers. The identification of these markets, also not straightforward, is based on an examination of pipeline locations and regional data on sales of gas. Our break-down of wholesale markets is shown in Table 3.

Table 2:

ALTERNATIVE CLASSIFICATIONS OF FIELD MARKETS

AGA	MacAVOY/PINDYCK[1]
1. Appalachian	
Kentucky	
Maryland	
New York	
Ohio, Pennsylvania	
Tennessee	
Virginia	
West Virginia	
2. Southeastern	
Alabama	
Florida	
Mississippi	
3. South Central Area	Field Market 2
Arkansas	Louisiana Southeast
Louisiana North	Texas 1
	Texas 2
	Texas 3
	Texas 4
	Texas 5 & 6
4. South Louisiana Area	Field Market 3
Louisiana South	Louisiana South
Louisiana Offshore	Louisiana Offshore
5. Texas Gulf Coast Area	
Texas 1	
Texas 2	
Texas 3	
Texas 4	

[1]This classification excludes states without significant additional reserves in the 1960's.

Table 2 (continued)

AGA	MacAVOY/PINDYCK
6. Northeast Texas Area	
Texas 5	
Texas 6	
Texas 7B	
Texas 9	
7. West Texas-Southeast New Mexico Area	Field Market 4
Texas 7C	Texas 7C & 8
Texas 8	New Mexico Southeast
Texas 8A	
New Mexico Southeast	
8. Mid-Continent Area	Field Market 1
Illinois	Kansas
Indiana	Oklahoma
Iowa	Texas 10
Kansas	Texas 7B, 8A & 9
Oklahoma	
Michigan	
Missouri	
Texas 10	
9. Rocky Mountain Area	Field Market 6
Colorado	Colorado & Utah
Montana	Montana
Nebraska	New Mexico Northwest
New Mexico Northwest	Wyoming
North Dakota	
10. Pacific Coast Area	Field Market 5
Alaska	California
Arizona	
California	
Oregon	
Washington	

Table 3: District Breakdown for Wholesale Markets

A. SALES FOR RESALE

1. Alabama
2. Arizona
3. Arkansas, Missouri
4. California
5. Colorado, New Mexico, Utah (Rocky Mountain)
6. Connecticut, Massachusetts, Rhode Island (New England)
7. Florida
8. Georgie
9. Idaho, Wyoming
10. Indiana
11. Illinois
12. Iowa
13. Kansas
14. Kentucky
15. Louisiana
16. Maryland, Virginia
17. Michigan
18. Minnesota
19. Mississippi
20. Nebraska
21. Nevada, Oregon, Washington
22. New Jersey
23. New York
24. Ohio
25. Oklahoma
26. Pennsylvania
27. Tennessee
28. Texas
29. West Virginia
30. Wisconsin

B. MAINLINE SALES

1. Alabama, Georgia
2. Arizona
3. Arkansas
4. Colorado, Wyoming
5. Florida
6. Illinois
7. Iowa, Nebraska
8. Kansas
9. Kentucky, Ohio, Pennsylvania, West Virginia (Appalachian)
10. Louisiana
11. Minnesota
12. Mississippi
13. Missouri
14. Oklahoma, Texas
15. Tennessee

B. Estimation Using Pooled Data

The cross-sectional breakdowns described above were chosen in order to identify functional economic markets, and no attempt was made to insure that the resulting districts will be homogeneous. It is very likely, in fact, that observations across districts are not homogeneous. The wholesale market in the northeastern U.S. probably has different characteristics from that in the southeast due to different weather conditions and different degrees of industrialization, and gas discovery rates in one Texas Railroad Commission district may be different from those in another due to geological differences. If district heterogeneity is not accounted for when specifying the equations of our model, we can expect the additive error terms to be autocorrelated across districts, just as time trends that are not accounted for will result in error terms that are autocorrelated across time.

This section will briefly review some of the standard methods in dealing with heterogeneity when estimating with pooled data and will then suggest an approach to be used in this study.

B.1 Problems of Autocorrelated Errors

We are estimating equations of the form

$$Y_{jt} = X_{jt}\beta + \varepsilon_{jt} \tag{26}$$

where j is a cross-section index (j=1,...,N) and t is the time index (t=1,...,T). If the equation is estimated by ordinary least squares (OLS), the Durbin-Watson statistic might indicate the presence of autocorrelation in the error terms, but it will not tell us what part of the autocorrelation is across time and what part is between cross-sections. Furthermore, the standard correction techniques, such as Hildreth-Lu [11], cannot be used since the

autocorrelation is two-dimensional. If the errors are autocorrelated and OLS is used, we can expect that the resulting estimates will at best be consistent and unbiased, but inefficient. If, however, the equation also contains lagged dependent variables (e.g., if $Y_{j,t-1}$ is one of the components of $X_{j,t}$), or independent variables referenced across districts, then the estimates will also be inconsistent [10].

The problem of autocorrelation in the cross-section dimension is often the result of a mis-specification that can be anticipated. Suppose, for example, that new discoveries of gas (ND) is believed to be linearly related to the number of wells drilled (W), so that the equation to be estimated is

$$ND_{j,t} = \beta_0 + \beta_1 W_{j,t} + \varepsilon_{j,t} \quad . \tag{27}$$

It is reasonable, however, to believe that geological differences will make some regions richer in gas than others, and therefore the wells in those regions will have a higher average "output." Perhaps in any given year, the same number of wells per district in each of two different districts j and j' can be expected to result in different amounts of discoveries. It is easy to see that this situation will result in autocorrelated errors in equation (27).

Consider two different districts, j and j', with average "output ratios" given by

$$\frac{1}{T} \sum_{t=1}^{T} \left(\frac{ND_{j,t}}{W_{j,t}}\right) = \alpha_j \tag{28}$$

and

$$\frac{1}{T} \sum_{t=1}^{T} \left(\frac{ND_{j',t}}{W_{j',t}}\right) = \alpha_{j'} \quad . \tag{29}$$

Thus, if the number of wells in these two districts were always the same, we would still expect to find on the average that

$$ND_{j,t} = \frac{\alpha_j}{\alpha_{j'}} ND_{j',t} = \theta_{jj'} ND_{j't} \quad . \tag{30}$$

A model, then, that would account <u>only</u> for the geological differences between districts j and j' would be

$$ND_{j,t} = \theta_{jj'} ND_{j',t} + \varepsilon^*_{j,t} \tag{31}$$

where the error term $\varepsilon^*_{j,t}$ is independent of j. Now if this equation (31) is substituted for $ND_{j,t}$ in (27), and the resulting equation is written with $\varepsilon_{j,t}$ on the left-hand side, we have

$$\varepsilon_{j,t} = \theta_{jj'} ND_{j',t} - \beta_0 - \beta_1 W_{j,t} + \varepsilon^*_{j,t} \quad . \tag{32}$$

But $ND_{j',t} = \beta_0 + \beta_1 W_{j',t} + \varepsilon_{j',t}$, and substituting this into (32) gives us

$$\varepsilon_{j,t} = \theta_{jj'} \beta_1 W_{j',t} - \beta_1 W_{j,t} + \theta_{jj'} \beta_0 - \beta_0 + \varepsilon^*_{j,t} + \theta_{jj'} \varepsilon_{j',t} \tag{33}$$

so that

$$E[\varepsilon_{j,t} \; \varepsilon_{j',t}] = \theta_{jj'} \sigma^2_\varepsilon \tag{34}$$

and the errors are thus autocorrelated.[1]

---

[1]Errors autocorrelated in time can occur in the same way. Consider the regression equation $Y_t = \beta X_t + \varepsilon_t$ with an unexplained time trend; e.g., $Y_t = \rho Y_{t-1}$ and $X_t = \rho X_{t-1}$. Then, $\varepsilon_{t-1} = Y_{t-1} - \beta X_{t-1} = \rho Y_t - \beta \rho X_t = \rho \varepsilon_t$, so that $E[\varepsilon_t \; \varepsilon_{t-1}] = \rho \sigma^2_\varepsilon$.

B.2 Methods of Estimating with Pooled Data

Several methods exist for dealing with cross-sectional heterogeneity when estimating equations using pooled data. Probably the two most common approaches are to introduce district dummy variables, or else to use a lagged dependent variable to account for district heterogeneity.

District dummy variables have been used in natural gas studies both by Balestra [2] and by Khazzoom [14] although in both cases the statistical significance of the dummy variables was sometimes questionable. There are two basic problems with district dummies, however, that would make their use in this study unappealing. The first is that they use up a large number of degrees of freedom. Some of the equations of this model are estimated by pooling data over some 20 districts, but over only 5 or 6 years, and the introduction of district dummies would thus use some 20% of the degrees of freedom. The second problem is that district dummies do not really explain in any kind of systematic way. As we will see, heterogeneity in an equation such as (27) can better be dealt with by introducing a constructed variable that explains past differences between districts.

The use of the lagged dependent variable can be a convenient way to deal with heterogeneity, particularly if the errors are not autocorrelated in time. If the equation already contains lagged independent variables, then the use of the lagged dependent variable will not use up any extra degrees of freedom. One problem that can arise, however, is an estimated value greater than one for the coefficient of the lagged dependent variable. Aside from the difficulty of interpreting such a result, the estimated equation will be dynamically unstable. Balestra had this problem in his estimates of demand for residential gas. He frequently obtained values for the coefficient of the lagged dependent term that were greater than one,

implying a negative depreciation rate for gas-burning appliances.

Another approach is to assume that the error terms are made up of components that originate from different sources and that therefore have different variances, and then gain efficiency in the estimates by using a variation of generalized least squares (GLS). The "residual" or "error components" model was first suggested by Kuh [15], and later generalized and applied by Balestra and Nerlove [3] and Wallace and Hussain [19]. Basically, this model assumes that the error term of equation (27) is made up of three independent components, one of which is associated with time, one with the cross-sections, and the last an independent random variable across both time and cross-sections, i.e., $\varepsilon_{j,t}$ is given by

$$\varepsilon_{j,t} = u_j + v_t + w_{jt} \tag{35}$$

with variances σ_u^2, σ_v^2, and σ_w^2. It is assumed that u_j, v_t, and w_{jt} are all independent of each other and that

$$E[u_j u_{j'}] = 0 \quad \text{for } j \neq j',\ E[v_t v_{t'}] = 0 \text{ for } t \neq t', \text{ and}$$

$$E[w_{jt} w_{j't}] = E[w_{jt'} w_{jt}] = E[w_{j't'} w_{jt}] = 0 \text{ for } j \neq j' \text{ and } t \neq t'.$$

Given these assumptions about the error vector ε_{jt}, we can write its variance-covariance matrix as

$$\Omega = E[\varepsilon\varepsilon'] = \sigma_u^2 A + \sigma_v^2 B + \sigma_w^2 I_{NT} \quad . \tag{36}$$

Note that Ω is an NT×NT matrix. I_{NT} is an NT×NT identity matrix, and A and B are NT×NT matrices defined by

$$A = \begin{bmatrix} J_T & 0 & \cdots & 0 \\ 0 & J_T & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & \cdots & J_T \end{bmatrix} \tag{37}$$

where J_T is a T×T matrix of ones, and

$$B = \begin{bmatrix} I_T & I_T & \cdots & I_T \\ I_T & I_T & \cdots & I_T \\ \vdots & & & \vdots \\ I_T & I_T & \cdots & I_T \end{bmatrix} \tag{38}$$

where I_T is a T×T identity matrix.

If the variance components σ_w^2, σ_u^2, σ_v^2 are known, then the minimum variance estimate of β is given by the GLS estimate

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}Y \ . \tag{39}$$

If the variance components are not known (which would presumably be the case), then Zellner's method [21] can be used, where consistent (but inefficient) estimates of β are obtained by OLS, the residuals are used to obtain consistent estimates of σ_w^2, σ_u^2, and σ_v^2, and GLS is finally used to obtain a new (and efficient) estimate of β.

The problem with this method is that while it accounts for differences in the variances of the error components, it does not account for heteroscedasticity or autocorrelations <u>within</u> each error component. Thus, if the error component that is cross-sectionally generated is itself heteroscedastic or if its elements are autocorrelated, we will still obtain inefficient estimates for β (although the estimates will be more efficient than those generated by OLS). Also, the method is computationally expensive, since the variance-covariance matrix Ω is a function of two parameters which must be "scanned," namely the variance ratios $\rho_1 = \sigma_u^2/\sigma_w^2$ and $\rho_2 = \sigma_v^2/\sigma_w^2$.

Estimation of the equations of our model is also complicated by the existence of simultaneity and the need, therefore, to use two-stage least squares (2SLS) in obtaining final values for coefficients. We need a method, then, that will:

a) account for cross-sectional heterogeneity

b) correct for possible autocorrelation in time

c) allow us to use 2SLS, since the equations of the model are simultaneous.

B.3 <u>A Hybrid Approach</u>

In deciding on a procedure to estimate the coefficients of this model, it must be kept in mind that:

a) A large number of regressions may be run for each equation in the model, using different specifications and different regional breakdowns, and therefore the estimation procedure must be convenient and reasonably inexpensive.

b) One purpose of this model is to provide insight into the structure and functioning of the natural gas industry. Therefore, if the cross-sections are heterogeneous (and in most cases we can expect that they will be) it is desirable to introduce regional variables (geological or economic) to <u>explain</u> those differences. If this is done properly, most of the autocorrelation across districts can be removed. An equation for new discoveries such as (27), for example, should be re-specified in the form:

$$ND_{j,t} = \beta_0 + \beta_1 W_{j,t} + \beta_2 \alpha_j + \varepsilon_{j,t} \tag{40}$$

where α_j is a geographical "output" variable defined as in equation (28).

c) We can expect that in some equations the errors will be autocorrelated in time. This should be accounted for.

d) We are estimating a simultaneous equation model, and therefore 2SLS will have to be used to ensure that the estimates are consistent and unbiased.

Our approach, then, will be to use regional variables whenever possible -- as in equation (40) -- to account for cross-sectional heterogeneity. Lagged dependent variables will be used to pick up remaining heterogeneity but will also serve to modify the dynamic response of the dependent variable. In addition, when errors are autocorrelated in time, we will use as a time-dependent correction a simple adaptation of the Hildreth-Lu procedure. An equation such as (26) would be tranformed as follows:

$$Y_{j,t} - \rho Y_{j,t-1} = (X_{j,t} - \rho X_{j,t-1})\beta + \varepsilon_{j,t} - \rho\varepsilon_{j,t-1} \quad . \tag{41}$$

The value of the parameter ρ, of course, is not known. A grid search can be performed on ρ, however, and that value chosen which minimizes the sum of squared residuals.

Equation (41) must in the end still be estimated by two-stage least squares. The direct application of 2SLS would lead to inconsistent estimates, however, since the equation is an autoregressive transformation and may also contain a lagged dependent variable. In order to obtain consistent estimates, then, we will use Fair's method [7]. The first-stage regression

of 2SLS is done as usual, resulting in a "constructed" series, $\hat{X}_{j,t}$. In the second stage, instead of replacing both $X_{j,t}$ and $X_{j,t-1}$ with the constructed series, only $X_{j,t}$ is replaced, so that OLS is performed on the equation:

$$Y_{j,t} - \rho Y_{j,t-1} = (\hat{X}_{j,t} - \rho X_{j,t-1})\beta \quad . \tag{42}$$

As an example, equation (40) would be estimated finally by applying OLS to

$$ND_{j,t} - \rho ND_{j,t-1} = \beta_0(1-\rho) + \beta_1(\hat{W}_{j,t} - \rho W_{j,t-1}) + \beta_2\alpha_j(1-\rho) \tag{43}$$

Actually, some 10 or 20 OLS regressions are performed on (43), each of which uses a different value of ρ.

4. Estimation of the Model

One of the difficulties in constructing a model of this sort is that one must work under the constraints imposed by data limitations. Data for many variables is either difficult or else impossible to obtain, particularly for years prior to 1966. In addition, much of the data is extremely noisy. As a result, a good deal of compromise was often required in estimating equations between functional forms that are theoretically pleasing and those that lend themselves to the existing data. This should be kept in mind when interpreting the estimation results.

A. List of Variables

A good deal of effort was involved in developing the data base represented by the variables listed below. The list, which includes the sources of data, is divided up into field market variables and wholesale market variables.

A.1 Field Market Variables

Wells

WXO: Exploratory wells completed for oil. Consists of new field wildcats, new pool wildcats, outposts, shallower pool tests and deeper pool tests (each completed for oil). Oil wells are defined by a minimum oil-to-gas production ratio of (1 bbl oil/100,000 c.f. gas). Annual data in the most disaggregated form (our 32 production sub-districts) available only from 1966. F.P.C. Bureau of Natural Gas: Well Drilling Statistics. Docket No. AR 67-1, and AAPG Bulletin, annual wells issue (June before 1971, July after).

WXG: Exploratory wells completed for gas or condensate. See WXO for details.

XF: Total footage drilled in exploratory wells. Includes gas, oil, condensate, and dry. Available only from 1966. AAPG Bulletin 1967-72 (data for 1966-1971).

FWGXF: A geographical variable designed to act as a surrogate for average drilling costs in each production district. It is the average exploratory footage per gas well completed.

$$(FWGXF)_j = \frac{1}{6} \sum_{t=1966}^{1971} \left(\frac{XF_{t,j}}{WXG_{t,j}}\right)$$

FWXF: Similar to FWGXF, but this is the average exploratory footage per total (gas and oil) wells completed. ("Completed" implies that the well is successful.)

SR: A geographical variable defined as the average (taken over the years 1966-1971) success ratio of oil and gas wells drilled. Defined for each district separately.

$$SR_j = \frac{1}{6} \sum_{t=1966}^{1971} \left[\frac{\text{successful gas wells + successful oil wells}}{\text{total gas and oil wells drilled}}\right]_{j,t}$$

MM1: Another geographical variable which is a surrogate for average size of discoveries in the particular district.

$$MM1_j = \frac{1}{5} \sum_{t=1967}^{1971} \left[\frac{ND_{t,j}}{WXG_{t-1,j}}\right]$$

Reserves: All data from AGA/API/CPA Reserves of Crude Oil, etc. Only available in disaggregated (i.e., ass-diss, non-ass. by production district) from 1966 except for year-end reserves (YN, YA, US) which we have from 1965.[1] In millions of cubic feet; see Figure 1.

[1] No data on Louisiana Offshore.

YN: Year end non-associated reserves. Reserves as defined by the AGA.

YA: Year end associated-dissolved reserves. See YN.

YO: Year end oil reserves.

US: Year end reserves in underground storage other than in their original locations.

XN: Extensions of non-associated gas. Includes any newly proved reserves already established in pools and fields.

XA: Extensions of associated-dissolved gas. See XN.

RN: Revisions of non-associated gas. Includes any proved <u>decreases</u> in size of proved reserves discovered by drilling of extension wells or changes (+ or −) resulting from better engineering estimates of economically recoverable reserves in established pools.

RA: Revisions of associated-dissolved gas. See RN.

FN: New field discoveries of non-associated gas (discovered by new field wildcats).

FA: New field discoveries of associated-dissolved gas. See FN.

PN: New pool discoveries of non-associated gas.

PA: New pool discoveries of associated-dissolved gas.

XRN: = XN + RN. Total extensions and revisions, non-associated.

XRA: = XA + RA. Total extensions and revisions, associated-dissolved.

DN: = FN + PN. Total new discoveries, non-associated.

DA: = FA + PA. Total new discoveries, associated-dissolved.

<u>Production</u>:[1] All data from AGA/API/CPA, <u>Reserves of Crude Oil, Etc</u>.

Available in disaggregated form from 1965 for gas (and about 1940 for oil).

In 10^6 cubic feet and 10^3 barrels.

QN: Production of non-associated gas. This is net production. See Figure 1.

QA: Production of associated-dissolved gas. See QN.

QO: Production of oil.

L: Losses and waste. API Petroleum Yearbook. By states, Texas, Louisiana, New Mexico undivided.

---

[1] No data on Louisiana Offshore.

Prices:[1]

PG: New contract price of interstate sales of gas at the wellhead. Table F, F.P.C., Sale of Natural Gas, Ftc., 1966-1970 (Cents/thousand cubic feet).

PO: Wellhead price of oil ($/bbl). Bureau of Mines, Minerals Yearbook, 1954-1971. Texas divided into 4 sub-districts. Otherwise, the same as our breakdown.

A.2 Wholesale Market Variables

The wholesale market is presently divided into two parts, mainline sales and sales for resale. Later, intra-state wholesale demand may be added. All variables can be written in the forms A.B.C or B.C . The first form is defined in terms of the second as A.B.C = B.C - (1-A)*B.C(-1) and hence is the depreciated change in B.C, where A is the depreciation rate (see p. 19). B is the variable name (discussed below) and C refers to the number corresponding to a particular disaggregation of the United States into wholesale markets (see below).

Variable Names (B)

QI: Mainline industrial sales volume by interstate pipeline companies by state and year (1962-1970) in Mcf (10^6 cu.ft.). From National Coal Association, Main Line Natural Gas Sales, 1964 to 1971.

PI: Price of mainline industrial sales made on a firm basis, by state and year (1964-1970). Estimated for 1962-1963 and for 1964-1965 for states with largely unspecified sales. Estimates based on linear extrapolation of following years' data. See QI for source.

VAM: Value added in manufacturing in millions of dollars by state and year, 1958-1969 (1968 is missing and is interpolated). From U.S. Department of Commerce, Bureau of the Census, Annual Survey of Manufacturers.

K: New capital expenditures in millions of dollars. See VAM for source.

PALT: Price of alternate fuels. Weighted average (over kilowatt-hours generated) of prices of fuel oil and coal consumed by electric utility industry in generating electrical power, by state and year (1962-1969). Weighting factors extrapolated for 1969 from previous data (1961-1968). From Edison Electric Institute, Statistical Annual of the Electric Utility Industry.

---

[*] In the sales for resale equations, we have converted D.SS.6 and D.SS.7 to MMcf by dividing this by 1000.

[1] No data on Louisiana Offshore.

QSR: Sales for resale volume of natural gas by class A and B interstate pipeline companies, by state and year, 1956-1970 in Mcf. Some companies have not reported for 1966-1968 and some for 1970-1971. Data needs to be cleaned up. Compiled from F.P.C. Form 2 reports.

PSR: Price of sale for resale by state and year (see QSR) in $/Mcf.

VCC: Value of construction contracts by states in which work was done, 1956-1970, in millions of dollars. From Statistical Abstract of the U.S. and F.W. Dodge Corporation, Dodge Construction Contract Statistics Service.

Y: Personal income by state, 1956-1969, in millions of dollars. Source: same as PSR, and U.S. Bureau of Economic Analysis, Survey of Current Business.

N: Population by state and year, 1955-1970, in millions. Source: same as above, and U.S. Department of Commerce, Bureau of the Census, Current Population Reports.

Wholesale Market Breakdowns (C)

C=3: Breakdown by state. Includes twenty-two of the most important states in terms of volume of mainline sales: Alabama, Arizona, Arkansas, Colorado, Florida, Georgia, Illinois, Iowa, Kansas, Kentucky, Louisiana, Minnesota, Mississippi, Missouri, Nebraska, Ohio, Oklahoma, Pennsylvania, Tennessee, Texas, West Virginia, and Wyoming. Others were omitted because of little or no volume, or as in the case of small but substantial states such as New Mexico, Nevada, Indiana and Michigan, they could not be integrated into a nearby market because of extreme price differences.

C=4: Here we integrated small volume states in C=3 if they were contiguous and had approximately the same price. Our breakdown is then
Alabama, Georgia
Colorado, Wyoming
Kentucky, Ohio, Pennsylvania, West Virginia
Iowa,Nebraska
Texas, Oklahoma
the rest of the states in C=3 being by themselves market regions.
This leaves us 15 regions.

C=6: This is the 41-state breakdown of sale for resale which leaves out only Alaska, Hawaii, Montana, Maine, Vermont, North and South Dakota, and North and South Carolina, for lack of data (or sales).

C=7: This combines several states in C=6 to get 30 districts; i.e.,
Arkansas, Missouri
New Hampshire, Massachusetts, Connecticut, Rhode Island
New Mexico, Utah, Colorado
Nevada, Oregon, Washington
Virginia, Maryland
Wyoming, Idaho.
Delaware is deleted because its price is too different from other states nearby to justify including its small volume of sales.

B. Statistical Results

All of the regression results described below were obtained using ordinary least squares (OLS). It was not necessary to use two-stage least squares (2SLS) -- unless otherwise indicated -- because the final equation forms that best fit the data contained no endogenous variables that were not lagged. This was not the result of a predisposition to eliminate simultaneity from the model, but merely a statistical result.

In addition, we did not use tle modified version of Hildreth-Lu described in section 3.B in obtaining these results. It was our feeling that this initial effort at estimating the model should be preliminary and experimental, designed more to test the data and the variety of relationships that might fit the data. We do expect to use that method, however, in estimating a future version of this model.

All of our equations are linear in form with the exception of the wells equation (for WXG) which is in log-log form. The log version of this equation fit far better than did any of a whole variety of linear versions. It is interesting to note that for all of the other equations of the model none of the log versions fit as well as the linear versions. Future work on this model will include the testing of other equation forms, including some that are non-linear. Also, we expect to be able to obtain reserves data for the years 1964 and 1965 (our present reserves data only covers the years 1966-1971), so that the supply equations for the field market can be estimated over a longer time period.

B.1 Field Market Equations

The field market portion of our model contains eight stochastic equations together with an identity to define changes in reserves. Seven of the stochastic equations (wells, non-associated discoveries, associated discoveries, non-associated extensions, associated extensions, non-associated revisions, and association revisions) determine the supply of new reserves, while the last (explaining the price level before 1962) determines demand.

The price equation, of course, can only be estimated using data before 1962. We were able to obtain data for the price equation beginning in 1956, and the cross-sections were very heterogeneous. As a result, our initial estimation results for this equation were rather unsatisfactory, and rather than extrapolate a poorly fitted equation from 1962 to 1972 (and then to 1980 for a forecast), we decided to leave the equation out of this preliminary version of the model. The wellhead price is exogenous anyway after 1962, and therefore leaving this equation out does not affect our simulation experiments and forecasts for new reserves and production in the future. It does make it impossible, however, to obtain a direct estimate of excess demand in the field market. An indirect estimate can be obtained though by equilibrating the wholesale market with the field market under the assumption of some minimum required reserves-to-production ratio. This point will be discussed in more detail later.

Our final equation for wells is in log form and relates completed exploratory wells to the wellhead price of gas (which is the main incentive to drill), the production of oil (also an incentive to drill, since oil may be discovered), and previous drilling. The equation is shown below together with

the R^2, F-statistic, standard error of the regression, Durbin-Watson statistic, and statistics (in parentheses).

$$\log WXG = \underset{(-2.10)}{-3.335} + \underset{(2.31)}{1.022}\ PG_{t-1} + \underset{(1.61)}{0.1212}\ QO_{t-1} + \underset{(6.21)}{0.6684}\ WXG_{t-1} \quad (44)$$

$R^2 = .7096$ $F = 40.7$ $S.E. = 0.591$ $DW = 1.75$

Other versions of this equation were estimated both in log form and in linear form (none of the linear forms were satisfactory), and the results are summarized in Table 4. It is interesting to note that the geographical variable MM1 which measures the "richness" of each production district was not significant in any of the versions, and as a result was dropped from the final form. Other specifications not listed in Table 4 were tested but were found to be inferior. It was thought that lagged new discoveries could be used as an explanatory variable since recent discoveries in a region might stimulate further drilling, but the relationship was not found to be significant.

Our equation for new discoveries of non-associated gas (DN) does contain an unlagged endogenous variable, namely wells, WXG, and therefore was estimated using two stage least squares (a fitted series for WXG was formed by regressing it on a set of instrumental variables). Our geographical variable MM1 (the average past size of discoveries per well on a regional basis) was used in this equation and proved quite significant. This final form of the 2SLS regression was:

$$DN = \underset{(-2.66)}{-141773.-} + \underset{(3.11)}{25.085}\ MM1 + \underset{(1.78)}{4189.3}\ WXG + \underset{(1.54)}{0.261}\ DN_{t-1} \quad (45)$$

$R^2 = .721$ $F = 43.0$ $S.E. = 2.2 \times 10^5$ $DW = 1.16$

Table 4: Alternative Wells Equations

(t-statistics in parentheses)

A. LOG FORM

R^2	const.	PG	PG_{t-1}	QO	QO_{t-1}	PO_{t-1}	MM1	WXG_{t-1}
.695	-1.693 (-1.11)		0.6055 (1.10)			0.4568 (0.340)	0.0111 (0.129)	0.7640 (7.719)
.711	-3.096 (-1.86)		1.015 (2.27)		0.1367 (1.69)		-.047 (-.547)	0.6734 (6.19)
.711	-3.063 (-1.85)		1.019 (2.267)	0.132 (1.66)			-.046 (-.532)	0.6706 (6.072)
.709	-3.298 (-2.08)		1.025 (2.29)	0.118 (1.584)				0.666 (6.093)
.693	-2.690 (-1.44)	0.8265 (1.54)			0.103 (1.312)			0.6905 (5.854)

B. LINEAR FORM

R^2	const.	PG	PG_{t-1}	QO	QO_{t-1}	PO_{t-1}	MM1	WXG_{t-1}
.552	-9.43 (-0.85)	0.985 (1.61)		.0000168 (.855)			.000573 (1.26)	0.388 (2.81)
.545	-5.812 (-0.22)		0.968 (1.08)			-0.910 (-.08)	.000765 (1.91)	0.436 (3.31)
.554	-8.14 (-0.79)		0.942 (1.60)		.000021 (0.99)		.000546 (1.20)	0.393 (2.90)

Other forms of this equation use a first difference in wells, multiply wells by the geographical variable MM1, and include the price of gas, but these did not fit as well. An attempt was also made initially to divide DN up into pools (PN) and fields (FN), but we were unable to obtain an equation that had a reasonable fit for the latter.

Our equation for associated discoveries (DA) also contained the wells variable WXG unlagged, and was therefore also estimated using two stage least squares. The geographical variable MM1 was not significant in this equation, but that was not surprising since the "payoff" for associated discoveries is linked so closely to both oil discoveries and oil production (many oil wells also contain gas which can be tapped after much of the oil has been produced). As a result, we included as an exogenous variable in this equation the production of oil QO. The final form of the equation is:

$$\underset{}{DA} = \underset{(-2.78)}{-10784.2} + \underset{(3.40)}{0.0685}\ QO + \underset{(2.33)}{381.64}\ WXG + \underset{(4.01)}{0.426}\ DA_{t-1} \tag{46}$$

$$R^2 = .833 \qquad F = 83.2 \qquad S.E. = 1.5 \times 10^4 \qquad DW = 1.73$$

Again an attempt was made to divide associated discoveries into those from pools (PA) and those from fields (FA):

$$PA = \underset{(-1.68)}{-4984.8} + \underset{(0.66)}{.0436}\ QO + \underset{(0.71)}{76.35}\ WXG_{t-1} + \underset{(6.14)}{0.586}\ PA_{t-1} \tag{47}$$

$$R^2 = .843 \qquad F = 89.7 \qquad S.E. = 1.25 \times 10^4 \qquad DW = 1.94$$

$$FA = \underset{(0.668)}{1002.3} + \underset{(3.35)}{.0253}\ QO + \underset{(0.36)}{18.90}\ WXG_{t-1} + \underset{(0.74)}{0.104}\ FA_{t-1} \tag{48}$$

$$R^2 = .368 \qquad F = 9.71 \qquad S.E. = 7.0 \times 10^3 \qquad DW = 1.80$$

As can be seen, these results are not as satisfactory, much of the difficulty arising from the instability in the series for FA. Thus, the equation in the aggregated form (i.e., for DA) was used.

Extensions of non-associated gas (XN) are related positively to the price of gas, negatively to the price of oil (since a higher oil price is a disincentive for companies which explore for both oil and gas to spend money extending gas fields), and positively to the number of wells drilled:

$$\underset{(1.16)}{XN = 8.816 \times 10^5} + \underset{(2.18)}{57813.0\ PG} - \underset{(-1.81)}{6.1 \times 10^5\ PO} + \underset{(1.99)}{5269.8\ WXG_{t-1}} + \underset{(10.96)}{0.615\ XN_{t-1}} \tag{49}$$

$$R^2 = .774 \qquad F = 42.0 \qquad S.E.\ 4.4 \times 10^5 \qquad DW = 1.54$$

Note that it is not necessary to use 2SLS to estimate this equation since the price of gas PG is exogenous after 1962, i.e., during the period over which the equation is estimated.

Extensions of associated gas (XA) depend on the production of oil QO (dissolved-associated gas fields are extended to allow for increased oil production), and also the previous year's new discoveries of associated gas:

$$XA = \underset{(-2.07)}{24521.9} + \underset{(4.36)}{0.4195\ QO} + \underset{(0.60)}{0.2671\ DA_{t-1}} + \underset{(1.62)}{0.2327\ XA_{t-1}} \tag{50}$$

$$R^2 = .757 \qquad F = 52.0 \qquad S.E. = 6.3 \times 10^4 \qquad DW = 1.65$$

Revisions of non-associated gas (RN) depend positively on the previous year's total reserves of non-associated gas (YN), and negatively on the change in the previous year's reserves of non-associated gas. Basically, large short-run increases in reserves usually limit revisions in the following year (which apply more to established discoveries), but in the long run a higher level of reserves results in a higher average level of revisions.

Our equation is[1]:

$$RN = \underset{(-1.71)}{-1.946 \times 10^5} + \underset{(2.11)}{0.0143}\ YN_{t-1} - \underset{(-4.09)}{0.1726}\ \Delta YN_{t-1} + \underset{(3.24)}{0.4946}\ RN_{t-1} \qquad (51)$$

$$R^2 = .511 \qquad F = 17.4 \qquad S.E. = 6.03 \times 10^5 \qquad DW = 1.52$$

Note that the coefficient of the autoregressive term (RN_{t-1}) is close to .5, and thus the long run impact on revisions resulting from an increase in reserves of 1 Mcf is approximately equal to

$$\frac{1}{1 - .5}(.0143) = .0286 \text{ Mcf}$$

since there is no long-run effect from the ΔYN_{t-1} term. Thus a one Mcf increase in the stock of non-associated reserves implies about a .03 Mcf per year increase in the flow of non-associated revisions.

Associated revisions (RA) are extremely erratic and difficult to explain using a simple linear regression model. There seems to be no relationship between this variable and lagged reserves of associated gas, mostly because associated revisions are linked more closely to oil reserves (which we do not include in this model). We modeled associated revisions by relating it to the previous year's associated discoveries and the previous year's associated extensions, although the relationship is somewhat dubious:

$$RA = \underset{(-1.30)}{-99078.6} + \underset{(1.56)}{4.766}\ DA_{t-1} + \underset{(0.89)}{0.844}\ XA_{t-1} - \underset{(-0.39)}{0.0543}\ RA_{t-1} \qquad (52)$$

$$R^2 = .236 \qquad F = 5.2 \qquad S.E. = 4.73 \times 10^5 \qquad DW = 1.47$$

---

[1] Note that $\Delta YN_{t-1} = YN_{t-1} - YN_{t-2}$.

B.2 Wholesale Market Equations

There are currently four stochastic equations in our representation of the wholesale market for gas. Consumption demand equations are estimated for mainline sales (QI) and for sales for resale (QSR), and price markup equations are also estimated for both mainline sales (PI) and sales for resale (PSR). Note that we are describing the market only for interstate sales of gas, and the production identity

$$Q = QI + QSR \tag{53}$$

defines only interstate production. This must be kept in mind when Q is used later to calculate the reserves-to-production ratio.

Before the mainline demand equation could be estimated, an estimate had to be made of the depreciation rate for industrial gas-burning equipment (r). This was done by estimating the rate simultaneously with the other coefficients of a total industrial demand equation that did not contain an autoregressive term:

$$QI = \underset{(2.95)}{(1-\ 0.104)}\ QI_{t-1} + \underset{(0.86)}{17175.8} - \underset{(-1.62)}{97554.8}\ PI_{t-1} + \underset{(1.18)}{9144.6}\ FIR_{t-1} \tag{54}$$

$$+ \underset{(0.46)}{1.576}\ K_{t-1} + \underset{(2.88)}{79150.3}\ PALT_{t-1}$$

$$R^2 = .317 \qquad F = 7.8 \qquad DW = 2.11$$

Here FIR is the firm-to-interruptible ratio, K is new capital expenditures, and PALT is a price index of alternative fuels. Note that the depreciation rate r came out to be about 10% and was significant (t = 2.95). A depreciation rate of 0.1 was then used to calculate our "new" demand variable $\delta QI = QI - (1-r)QI_{t-1}$ for use in the final equation for mainline demand.

Our mainline demand equation contains the same explanatory variables as equation (54) above, except that it also contains an autoregressive term. This is added to help explain the dynamic response of new demand (i.e., it imposes a Koyck distribution on the explanatory variables) and also to account partially for whatever heterogeneity might remain in the cross-sections. Our final equation was:

$$\underset{}{\delta QI} = \underset{(0.64)}{10951.0} - \underset{(-1.69)}{74609.8}\, PI_{t-1} + \underset{(1.10)}{7611.9}\, FIR_{t-1} + \underset{(0.35)}{1.063}\, K_{t-1} \qquad (55)$$

$$+ \underset{(2.85)}{69467.0}\, PALT_{t-1} + \underset{(1.62)}{0.181}\, \delta QI_{t-1}$$

$$R^2 = .325 \qquad F = 8.1 \qquad S.E. = 1.8 \times 10^4 \qquad DW = 2.11$$

Note that the R^2 and t-statistics are low, but this reflects the scaling of the dependent variable. If the depreciation rate r was set to 0.18 instead of 0.10, the mean of the dependent variable would be larger and the R^2 also larger. The regression result using r = 0.18 is shown below:

$$\delta QI = \underset{(1.38)}{24418.3} - \underset{(-2.73)}{1.30 \times 10^5}\, PI_{t-1} + \underset{(2.05)}{13631.6}\, FIR_{t-1} + \underset{(1.01)}{3.154}\, K_{t-1} \qquad (56)$$

$$+ \underset{(3.40)}{85583.6}\, PALT_{t-1} + \underset{(2.73)}{0.290}\, \delta QI_{t-1}$$

$$R^2 = .569 \qquad F = 22.2 \qquad S.E. = 1.8 \times 10^4 \qquad DW = 2.14$$

Note that the standard error in both cases is the same (1.8×10^4).

An attempt to estimate the depreciation rate for the sales for resale demand equation was unsuccessful, since the rate that resulted was negative. We therefore used the same depreciation rate (0.10) as was used in the mainline demand equation. Explanatory variables include the wholesale price,

the price of oil (oil is the main competing fuel for residential and commercial demand) and total personal income (on a state-wide basis). The autoregressive term was also included. The time-bounds for this equation went from 1962 to 1969, and 30 districts were used (whereas an aggregated set of 15 districts was used in the mainline demand equation -- see Table 4). The resulting equation is shown below:

$$\underset{(0.3C)}{\delta QSR = 6.32 \times 10^6} - \underset{(-3.67)}{1.762 \times 10^8}\, PSR + \underset{(4.95)}{1.121 \times 10^8}\, PO + \underset{(8.07)}{15599.5}\, Y \qquad (57)$$

$$+ \underset{(1.46)}{0.103}\, \delta QSR_{t-1}$$

$$R^2 = .312 \qquad F = 26.6 \qquad S.E. = 6.35 \times 10^7 \qquad DW = 2.38$$

Again, the low R^2 reflects the dimensioning of the dependent variable.

In keeping with the arguments on page 23, the wholesale price markup was seen to depend on volumetric capacity and average mileage from production regions to the particular wholesale market region. Both of these explanatory variables are assumed to be roughly constant in time, so that our price markup regressions are actually cross-section regressions resulting in constant markups. For mainline sales, the markup is given by:

$$PI = PG + \underset{(3.71)}{3.071} + \underset{(23.46)}{1.053} \times 10^{-4}\, M - \underset{(-3.25)}{8.959 \times 10^{-5}}\, v \qquad (58)$$

$$R^2 = .861 \qquad F = 463.6 \qquad S.E. = 2.71 \qquad DW = 0.17$$

Note that the coefficient of volumetric capacity v is negative, since a larger capacity implies lower average costs. For sales for resale, our markup equation is given by:

$$PSR = PG + \underset{(14.82)}{13.672} + \underset{(13.88)}{0.5973}\, M - \underset{(-13.56)}{4.828 \times 10^{-4}}\, v \qquad (59)$$

$$R^2 = .688 \qquad F = 426.9 \qquad S.E. = 5.05 \qquad DW = 0.23$$

This completes the estimation of the model. In the next section, the equations are combined (together with the appropriate accounting identities), and the complete model is simulated.

5. Use of the Model for Policy Simulations

Before the model can be simulated into the future, a mechanism must be specified by which gas is distributed from production regions to consumption regions, since the equations as they now stand cannot be simply combined and simulated directly. The wellhead price that appears in the field market equations, for example, applies to a different set of regions than does the wellhead price that appears in our wholesale price markup equations. Similarly, additions to reserves (from discoveries, etc.) are made in field market regions, while depletions of reserves (resulting from production) are indexed by wholesale regions. Furthermore, the wholesale regions themselves are different for mainline sales and sales for resale. Thus the model cannot be equilibrated unless there is some means by which reserve additions from field markets can be added together and then redistributed to wholesale markets.

To do this we have divided the U.S. into three "zones" and then postulated three large regional field markets (East, North, and West), in which gas produced in the various production districts can be accumulated and then distributed to the various regional markets for mainline sales and sales for resale. The regional breakdown that we have selected is shown in Table 5. Note, for example, that gas from the Texas Gulf Coast, East Texas, and Louisiana

Table 5: Regional Interrelationships for Simulation

Production Districts	Regional Field Markets	Mainline Markets	Sales for Resale Markets
Texas RRC #1		Alabama, Georgia	Alabama
Texas RRC #2			New England
Texas RRC #3		Florida	Florida
			Georgia
Texas RRC #4		Appalachian	Kentucky
Texas RRC #5,6	East	Louisiana	Louisiana
Louisiana North			Maryland, Virginia
		Mississippi	Mississippi
Louisiana South (including Off-shore)			New Jersey
			New York
			Pennsylvania
		Tennessee	Tennessee
			West Virginia
		Arkansas, Missouri	Arkansas, Missouri
			Indiana
		Illinois	Illinois
		Iowa,Nebraska	Iowa
Texas RRC #7B,8A,9		Kansas	Kansas
			Michigan
Texas RRC #10		Minnesota	Minnesota
	North		Nebraska
Oklahoma			Ohio
		Oklahoma, Texas	Oklahoma
Kansas			Texas
			Wisconsin
Texas RRC #7C, 8		Arizona	Arizona
			California
New Mexico Northwest	West	Colorado, Wyoming	Rocky Mountain
			Idaho, Wyoming
New Mexico Southwest			Nevada, Oregon, Washington
Montana, Colorado, Utah Calif, Wyoming, West Virginia	Other		

is sold to pipelines which begin in the Regional Field Market East and then transmit the gas east to New England, the Appalachian region, the Eastern Seaboard, and the Southeast. Gas from North Texas, Oklahoma, and Kansas is sold in the North Regional Field Market to pipelines distributing gas to the North Central and South Central U.S. Gas produced in West Texas and New Mexico is sold in the West Regional Field Market to pipeline companies which distribute it to the West and Northwest U.S. Since gas produced in Montana, Colorado, Utah, California, Wyoming, and West Virginia does not contribute significantly to interstate mainline sales or sales for resale, we have omitted them from the interactive part of the model, though they are still included in terms of their contribution to year-end reserves through new discoveries, extensions, revisions, etc.

Note that this breakdown of field markets need not correspond to actual economic markets. We have not yet been able to statistically identify the geographical boundaries of actual field or wholesale markets, and thus for purposes of simulation we are simply trying to aggregate districts in a "reasonable" way. An extension of this research will include an attempt at better identifying regional markets and their interrelationships. As we have explained earlier in this paper, the interaction of regional markets is an important aspect of the natural gas industry.

A. Initial Policy Simulations

In the policy simulations that follow, we have not included equation (49) for non-associated extensions, since it is unrealistic in explaining the behavior of this variable when the price of gas at the wellhead

is fixed to be constant and the price of oil is growing. Thus, in these initial simulations, non-associated extensions (XN) is taken to be exogenous and allowed to grow at 5% from its 1970 value.

Table 6 shows our assumptions about the growth of the exogenous variables. The price of oil is assumed to grow at 6%, which is probably on the high side and is the cause of the poor performance of the XN equation (with oil prices rising that rapidly and gas prices fixed, it does not pay to extend non-associated gas fields). In the next set of simulations, we will reduce the growth rate of the price of oil to 2% per year, and then include the XN equation in the model.

We ran six simulations using this version of the model together with the exogenous variable forecasts summarized in Table 6. In the first simulation, we hold the ceiling price of gas fixed at its 1970 values (on a region-by-region basis). In the second, we allow the ceiling price to increase by one cent per year, so that in 1980 it is ten cents higher than its 1970 level in each production district. In the remaining simulations it increases by 2¢ per year, 3¢ per year, 4¢ per year, and 5¢ per year. The results are summarized graphically for several key aggregate variables in Figures 5 through 10, and on an East-West-North regional basis in Figures 11 through 15.

Note that the drilling of wells is extremely sensitive to the ceiling price of gas, and therefore additions to reserves (through new discoveries and extensions and revisions) is indirectly sensitive, although with a time lag of about two years. With <u>no change in the ceiling price of gas</u> during

Table 6: Forecast Assumptions for Exogenous Variables

Price of Oil (PO)	grows at 6% per year
Quantity of Oil (QO)	grows at 6% per year
Price of Alternate Fuels (PALT)	grows at 6% per year
Personal Income by State:	
East:	grows at 7.2% per year
North:	grows at 7.5% per year
West:	grows at 7.7% per year
Capital Expenditures (K) (Business Fixed Investment)	grows at 6% per year
Firm-to-Interruptible Ratio (FIR)	assumed constant
Mileage (M)	assumed constant
Volumetric Capacity (v)	assumed constant
Non-Residential Structures	grows at 3.5% per year

the decade 1970-1980, the number of wells drilled declines slightly, total new discoveries remain about constant, and extensions and revisions decline rapidly. Total demand for gas (i.e., production), however, increases by about 11% per year, for a total increase of 150% over the decade. This increase in production demand is to be expected since the wholesale price is constant (the markup is constant), while the price of alternate fuels, GNP, and capital expenditures are all increasing. If this demand were to be somehow satisfied (by depleting reserves), we would run out of reserves by 1980. This demand, of course, would not be satisfied. By the time the reserves-to-production ratio dropped to about 7 or 8 (i.e., by 1974), considerable rationing would occur, and many consumers would simply not be able to purchase gas.

Even with a ceiling price increase of 2¢ every year, demand for reserves would increase faster than the supply, and the reserves-to-production ratio would drop to about 2.3 by 1980. Many more wells would be drilled in this case, and both new discoveries and extensions and revisions would increase accordingly, but this increase in reserves would be outstripped by the more rapidly growing increase in production. Even with a 2¢ per year increase in the price of gas, increases in the prices of other fuels, in GNP, and in capital expenditures would result in a doubling of wholesale gas demand by 1980.

A ceiling price increase of 4¢ per year or 5¢ per year would result in increases in reserve additions that would be larger than the increases in wholesale demand, so that the reserves-to-production ratio would increase

over the decade, reaching, by 1980, 23 for a 4¢ annual price increase and 41 for a 5¢ increase. This is the result of a dramatic increase in well drilling together with an increase in wholesale demand of only 4% to 6% per year.

If F.P.C. objectives are to keep the reserves-to-production ratio approximately constant at around 12, then these results indicate that the ceiling price should be increased by about 3-3.5¢ per year. In this case, additions to reserves would be commensurate with increases in production demand, and the reserves-to-production ratio would remain about the same.

Note that all of these results make certain assumptions:

(1) That exogenous variables will grow at certain rates (for example, the price of oil is assumed to grow at 6% per year).

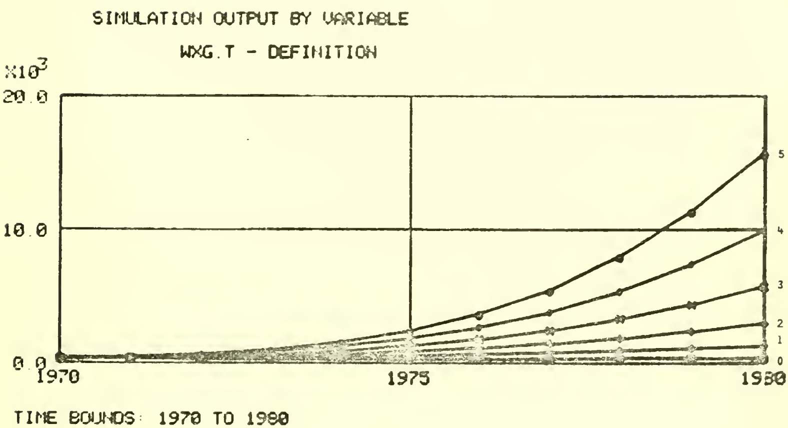
(2) Geological conditions will remain the same, i.e., we will not simply begin to run out of gas. Thus we assume that wells drilled in 1975 will produce as much gas as those drilled in 1970.

(3) We will not run out of oil or see any rationing of oil.

(4) The basic determinants of the demand for gas will not change in time.

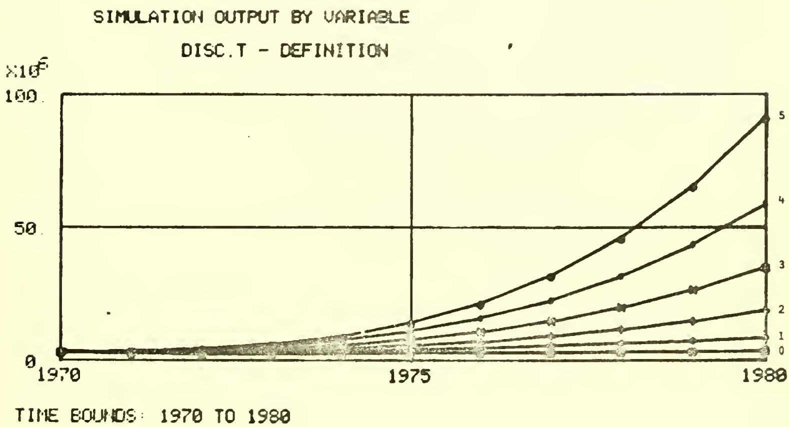
This last assumption may be the most questionable, at least in the short-run. We may have begun to see in the past year or two a change in "ecological consciousness" in the U.S. On the whole, gas is a somewhat cleaner fuel than oil, and this may result in additional increases in the demand for gas.

Figure 5: Total Wells Drilled and Completed



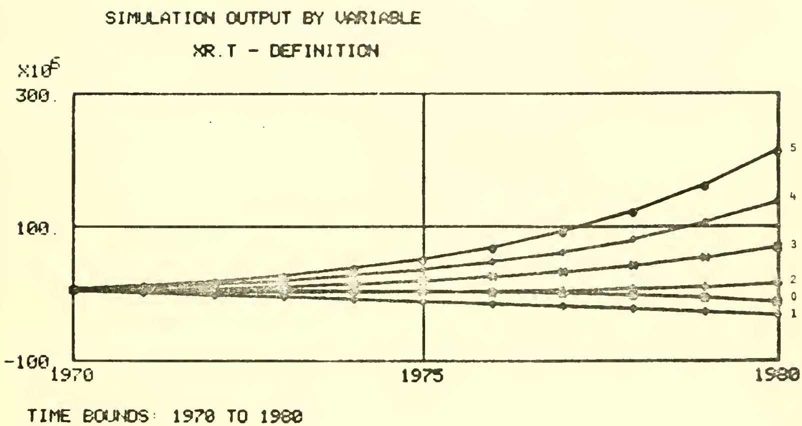
Numbers at the far right indicate the price increase in cents per year.

Figure 6: Total New Discoveries



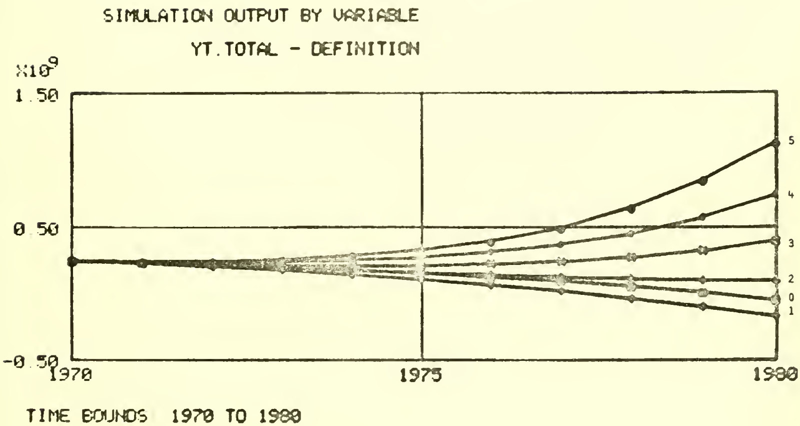
numbers at the far right indicate the price increase in cents per year.

Figure 7: Total Extensions and Revisions



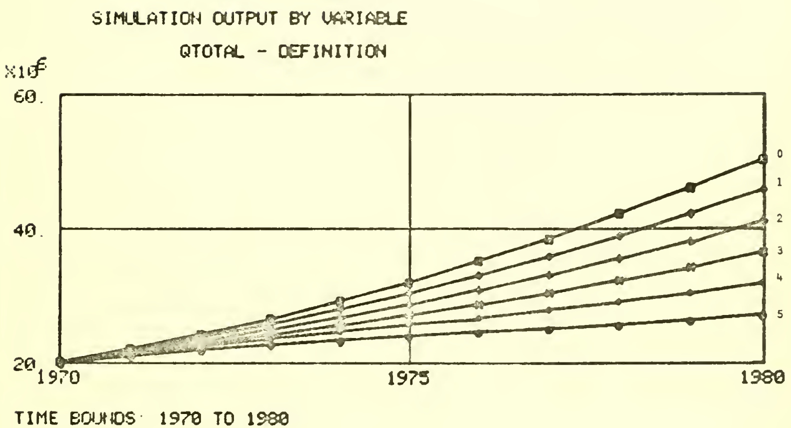
Numbers at the far right indicate the price increase in cents per year.

Figure 8: Total Year-End Reserves



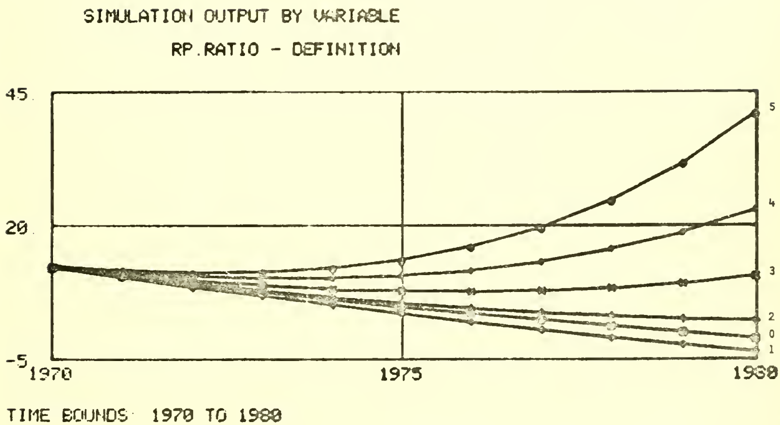
Numbers at the far right indicate the price increase in cents per year.

Figure 9: Total Production



Numbers at far right indicate the price increase in cents per year.

Figure 10: Reserves-to-Production Ratio



Numbers at far right indicate the price increase in cents per year.

Figure 11: Wells Drilled and Completed, by Region

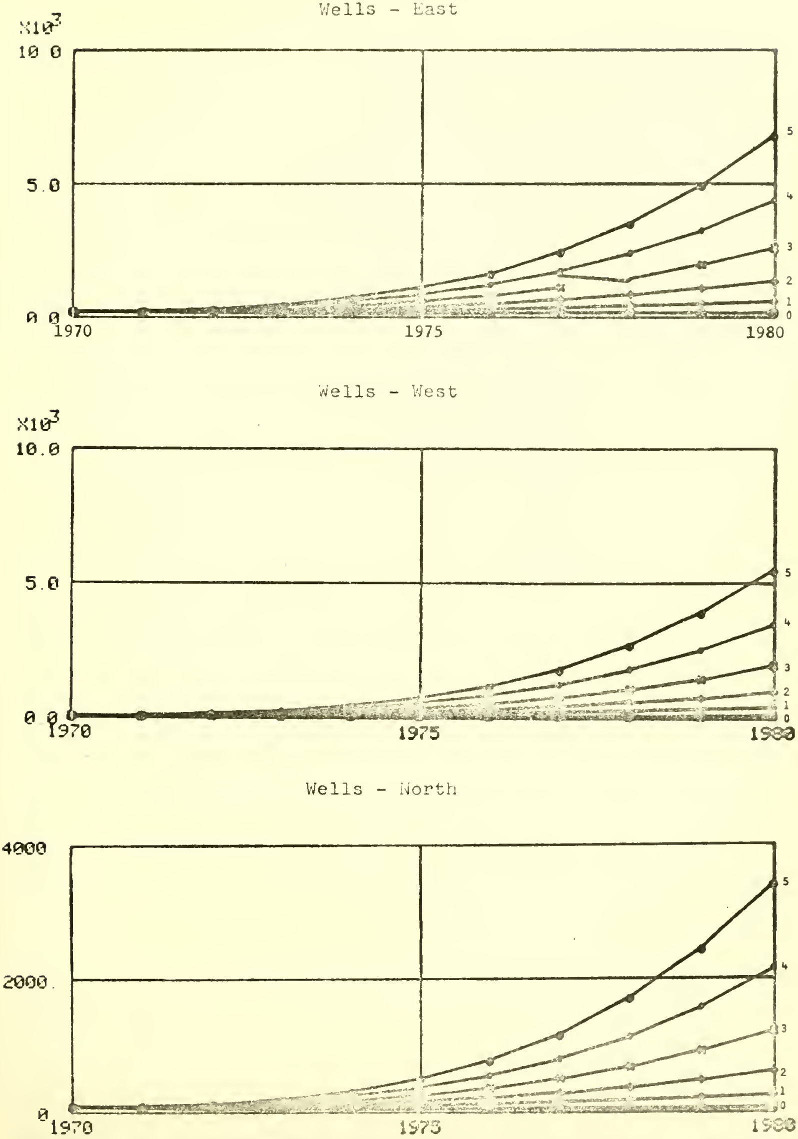
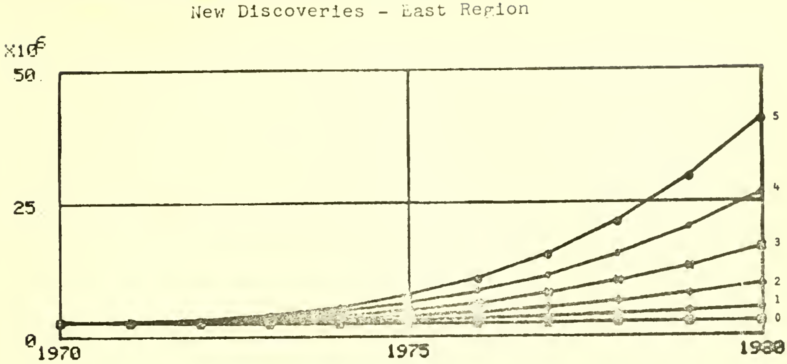
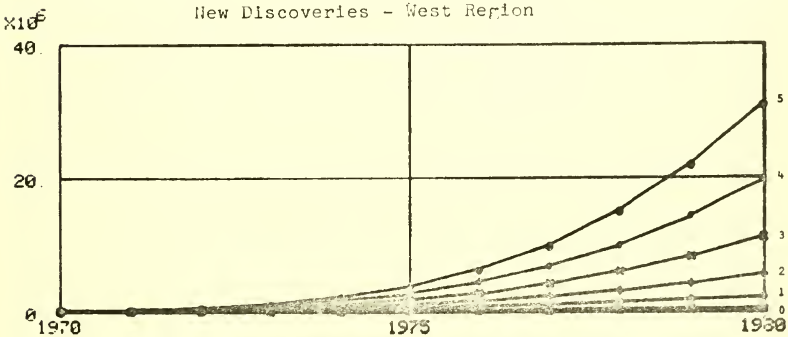


Figure 12: New Discoveries, by Region





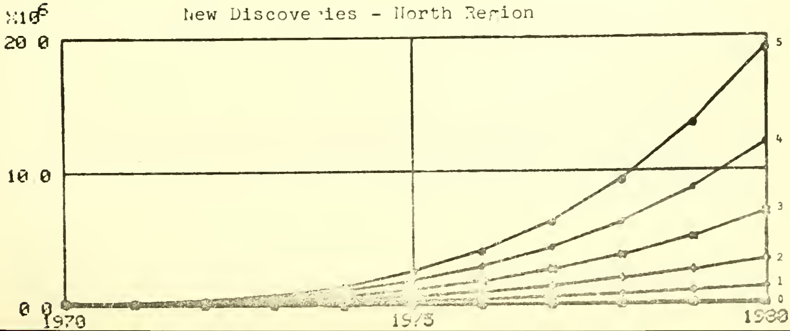
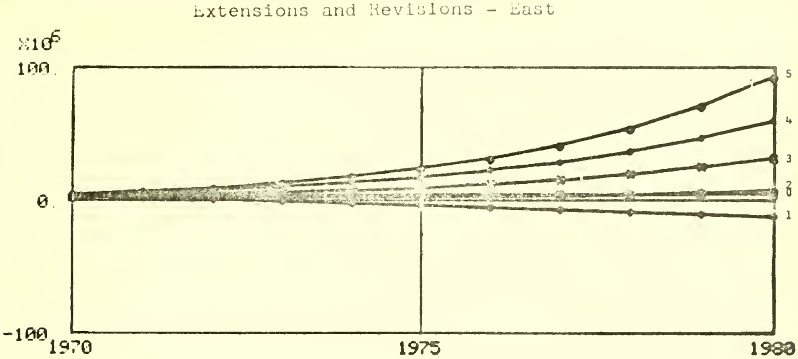
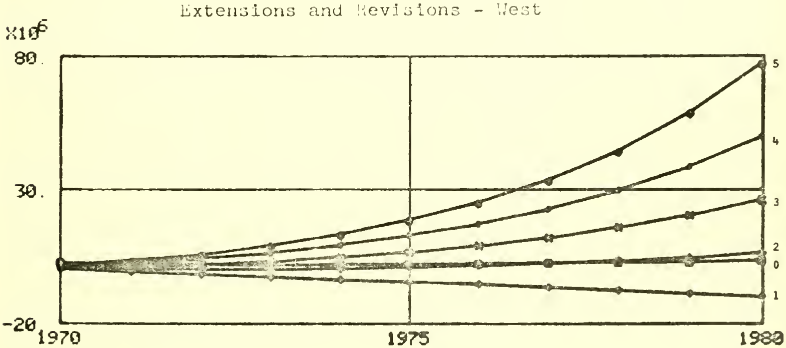


Figure 13: Extensions and Revisions, By Region





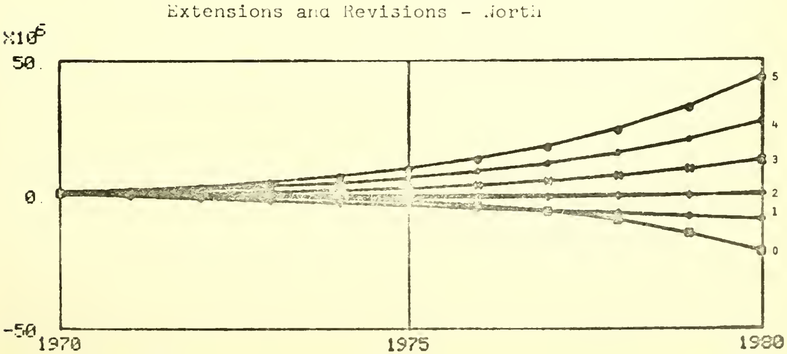


Figure 14: Year-End Reserves, By Region

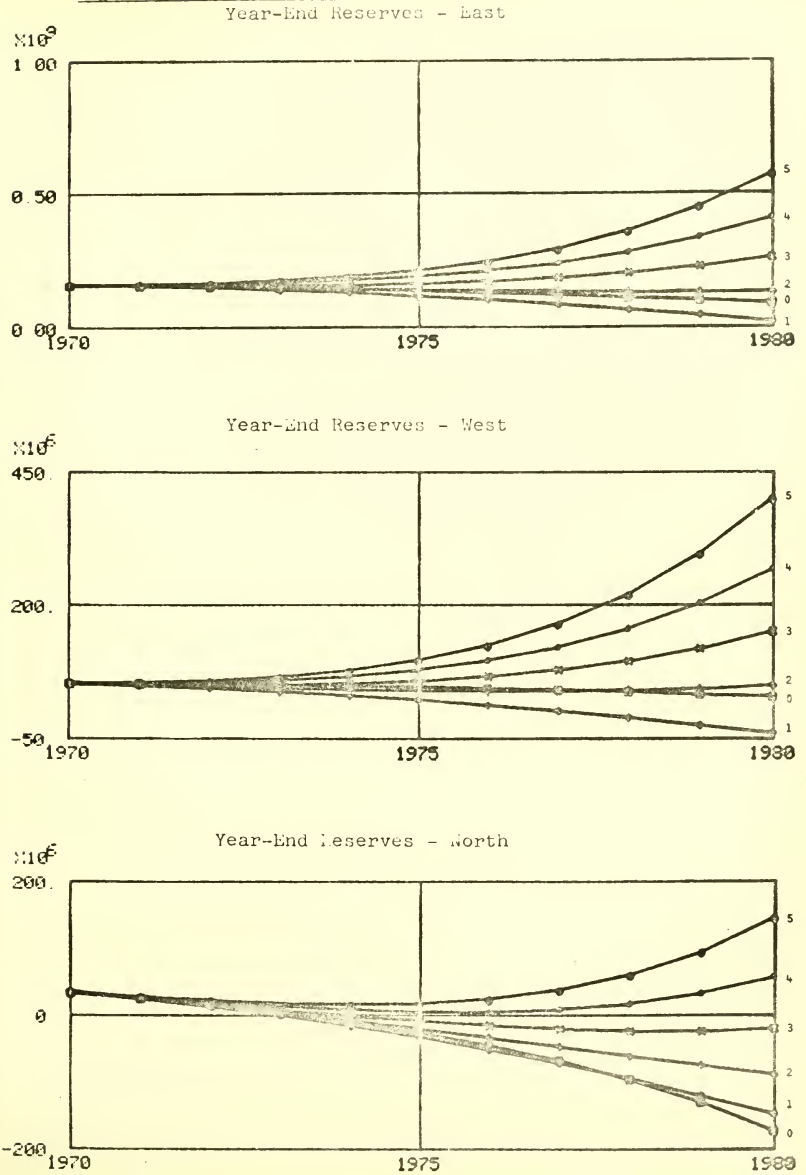
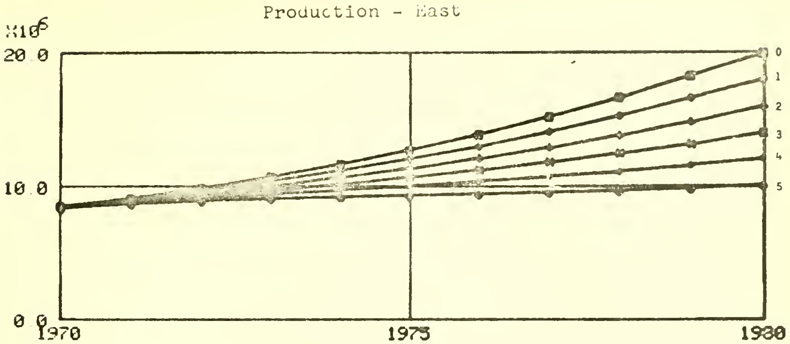
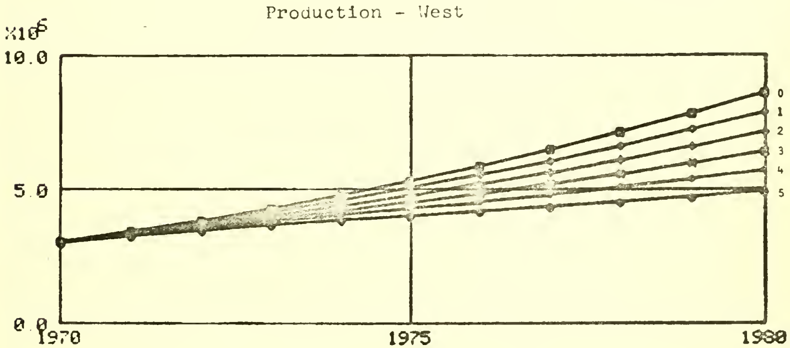


Figure 15: Production, By Regions





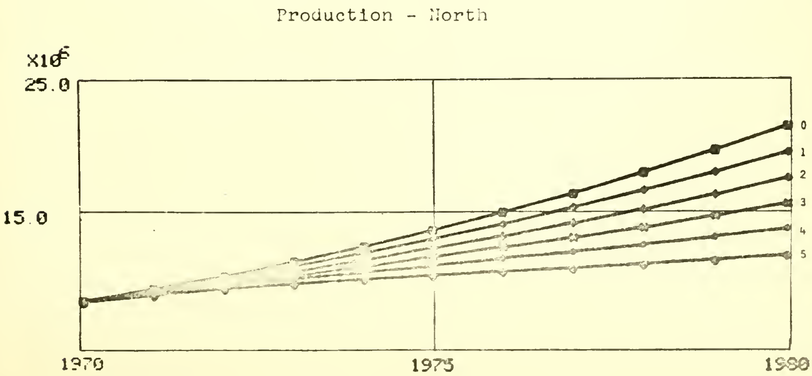


Table 7: Reserves -to- Production Ratio

	CP.0	CP.1	CP.2	CP.3	CP.4
1970	12.2318	12.1554	12.2511	12.3474	12.4444
1971	10.6626	10.4252	10.728	11.0368	11.3518
1972	9.12671	8.68331	9.2946	9.93049	10.5924
1973	7.65588	6.95993	7.96772	9.04318	10.1929
1974	6.26719	5.2726	6.75482	8.38589	10.1871
1975	4.95013	3.63378	5.66302	7.97642	10.6297
1976	3.70088	2.05219	4.7003	7.84058	11.5986
1977	2.49979	0.533869	3.87566	8.01161	13.1949
1978	1.35055	-0.917034	3.1988	8.52835	15.5393
1979	0.166315	-2.29781	2.67966	9.4329	18.7657
1980	-1.02081	-3.60682	2.32779	1[illegible].7579	23.[illegible]1[illegible]7

	CP.5
1970	12.5419
1971	11.6732
1972	11.282
1973	11.4243
1974	12.1839
1975	13.6932
1976	16.1427
1977	19.7858
1978	24.9367
1979	31.959
1980	41.2464

Table 8: New Discoveries

	CP.0	CP.1	CP.2	CP.3	CP.4
1970	3.414978E+06	3.414978E+06	3.414978E+06	3.414978E+06	3.414978E+06
1971	3.128109E+06	3.229916E+06	3.331849E+06	3.433910E+06	3.536086E+06
1972	2.970609E+06	3.261225E+06	3.566664E+06	3.886617E+06	4.220797E+06
1973	2.917502E+06	3.476808E+06	4.104810E+06	4.802168E+06	5.569549E+06
1974	2.931314E+06	3.[illegible]41849E+06	4.944653E+06	6.249346E+06	7.765068E+06
1975	2.987414E+06	4.339097E+06	6.111670E+06	8.345109E+06	1.107760E+07
1976	3.071716E+06	4.964077E+06	7.650566E+06	1.124354E+07	1.585135E+07
1977	3.176439E+06	5.719897E+06	9.618899E+06	1.512741E+07	2.249309E+07
1978	3.297322E+06	6.614250E+06	1.208360E+07	2.020344E+07	3.146430E+07
1979	3.432100E+06	7.657823E+06	1.511951E+07	2.670008E+07	4.327677E+07
1980	3.579664E+06	5.863613E+06	1.880891E+07	3.486725E+07	5.849149E+07

	CP.5
1970	3.414978E+06
1971	3.638371E+06
1972	4.568938E+06
1973	6.407604E+06
1974	9.500537E+06
1975	1.434582E+07
1976	2.157918E+07
1977	3.195866E+07
1978	4.635118E+07
1979	6.572326E+07
1980	9.113704E+07

Table 9: Total Production

	CP.0	CP.1	CP.2	CP.3	CP.4
1970	2.031741E+07	2.025398E+07	2.019054E+07	2.012712E+07	2.006368E+07
1971	2.221341E+07	2.199950E+07	2.178560E+07	2.157168E+07	2.135776E+07
1972	2.433592E+07	2.388571E+07	2.343549E+07	2.298526E+07	2.253506E+07
1973	2.669293E+07	2.592710E+07	2.516131E+07	2.439550E+07	2.362968E+07
1974	2.929091E+07	2.813755E+07	2.698418E+07	2.583085E+07	2.467749E+07
1975	3.213757E+07	3.053174E+07	2.892595E+07	2.732014E+07	2.571434E+07
1976	3.524206E+07	3.312534E+07	3.100862E+07	2.889190E+07	2.677518E+07
1977	3.861504E+07	3.593477E+07	3.325448E+07	3.057419E+07	2.789392E+07
1978	4.226851E+07	3.897728E+07	3.568605E+07	3.239482E+07	2.910358E+07
1979	4.621586E+07	4.227107E+07	3.832626E+07	3.438142E+07	3.043661E+07
1980	5.047210E+07	4.583533E+07	4.119856E+07	3.656176E+07	3.192499E+07

	CP.5
1970	2.000026E+07
1971	2.114386E+07
1972	2.206483E+07
1973	2.286387E+07
1974	2.352414E+07
1975	2.410853E+07
1976	2.465843E+07
1977	2.521362E+07
1978	2.581235E+07
1979	2.649179E+07
1980	2.728821E+07

6. Tentative Conclusions

The process of construction of an econometric model of natural resource utilization is iterative. This is a report on the first of the iterations; as improvements are made in the data base, by extending the coverage and increasing the accuracy of the observations, the model will be refitted. As mistakes in the formal structure of the model are revealed by our readers, there will be additional computations of the equations. And as the iterations occur, we expect the findings to change.

Given the first-round results in the preceding section, we can formulate preliminary conclusions, however. Following a strategy of well-publicized price increases, the Federal Power Commission can eliminate the domestic gas shortage by 1980. The best of the (feasible) price changes would be 3 cents per annum. In contrast, 1 cent per annum would eliminate the entire stock of domestic reserves, and 5 cents per annum would not be maintainable in the markets for gas because there would be a veritable flood of additional discoveries relative to additional production demand. The 3 cent per annum increase would "find" enough gas while restraining demands for a domestic 12:1 reserve-to-production ratio -- close to that before the shortage. The resulting level of wholesale prices would be less than that which would occur under a continuation of the present ceilings and importation of liquified gas at $1.15 per Mcf.

The "public strategy" set out here for solving the gas "crisis" is one of gradual, publicized increases in F.P.C. ceiling prices. An alternative

would be to deregulate, in stages or abruptly, so that markets cleared through price increases in the middle 1970's. We cannot simulate this alternative strategy at the present time, since this first model was built on 1960-1970 conditions in which prices were exogenous. An attempt will be made to anticipate the effects of endogenous prices, however, in forthcoming iterations. This will be done by formulating and testing a functional relation for "demand for reserves" based on 1955-1960 experience before ceiling prices; the addition of this function to the model "closes" the system and determines prices at equilibrium levels of reserves and production.

This approach will be _most_ tentative. Extrapolating behavioral relations fitted on data from the 1950's is likely to lead to large forecast errors. But it may provide insights into the contrasting balance of short-term market adjustments -- contrasting to the "lock-step" price increases posited in the simulations of regulation provided here.

The case for _less_ price control is made here, in these initial findings of gains for consumers from 3 cent per annum price increases. The case for deregulation is made on philosophical grounds -- we should not regulate effectively competitive markets, or use price controls to solve problems in taxing the windfall gains of resource producers. Improvements in this work may lead to quantitative findings which indicate that immediate deregulation is preferred.

REFERENCES

1. A. Alchian, "Costs and Outputs," in Allocation of Economic Resources, M. Abramowitz (ed.), Stanford University Press, 1958.

2. P. Balestra, The Demand for Natural Gas in the United States, North Holland Publishing Company (Amsterdam, 1967).

3. P. Balestra and M. Nerlove, "Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," Econometrica, July 1966.

4. S.G. Breyer and P.W. MacAvoy, Energy Regulation by the Federal Power Commission, The Brookings Institution (Washington, D.C.), forthcoming, spring 1973.

5. V.K. Chetty, "Pooling of Time Series and Cross Section Data," Econometrica, April 1968.

6. E.W. Erickson and R.M. Spann, "Supply Price in a Regulated Industry, The Case of Natural Gas," Bell Journal of Economics and Management Science, Vol. 2, No. 1 (1971).

7. R.C. Fair, "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors," Econometrica, May 1970.

8. F.M. Fisher, Supply and Costs in the U.S. Petroleum Industry: Two Econometric Studies, Johns Hopkins Press (Baltimore, 1964).

9. P.J. Garfield and W.F. Lovejoy, Public Utility Economics, Prentice-Hall, 1964.

10. Z. Griliches, "A Note of Serial Correlation Bias in Estimates of Distributed Lags," Econometrica, January 1961.

11. G. Hildreth and J.Y. Lu, "Demand Relations with Autocorrelated Disturbances," Michigan State University Agricultural Experiment Station, Technical Bulletin 276, November 1960.

12. I. Hoch, "Estimation of Production Function Parameters Combining Time-Series and Cross-Section Data," Econometrica, January 1962.

13. H.S. Houthakker and L.D. Taylor, Consumer Demand in the United States, 1929-1970, Harvard University Press (Cambridge, 1966).

14. J.D. Khazzoom, "The F.P.C. Staff's Econometric Model of Natural Gas Supply in the United States," Bell Journal of Economics and Management Science, Vol. 2, No. 1 (1971).

15. E. Kuh, "The Validity of Cross-Sectionally Estimated Behavior Equations in Time Series Applications," Econometrica, January 1962.

16. E. Kuh and J.R. Meyer, "How Extraneous are Extraneous Estimates?," Review of Economics and Statistics, November 1957.

17. P.W. MacAvoy, Price Formation in Natural Gas Fields, Yale University Press (New Haven, 1962).

18. P.W. MacAvoy, "The Regulation-Induced Shortage of Natural Gas," Journal of Law and Economics, April 1971.

19. T.D. Wallace and A. Hussain, "The Use of Error Components in Models Combining Cross-Section with Time Series Data," Econometrica, January 1969.

20. S.H. Wellisz, "Regulation of Natural Gas Pipeline Companies: An Economic Analysis," Journal of Political Economy, February 1963.

21. A. Zellner, "An Efficient Method of Estimating Seemingly Unrelated Regressions," Journal of the American Statistical Association, Vol. 57 (1962), pp. 348-368.

Date Due

~~'77~~	OCT 20 1992	
MAY 12 '77	OC 15 '89	
'77	APR 30 2001	
OCT 28 '77		
DEC 2 5		
~~JAN 11 '79~~		
FEB 12 '79		
~~'80~~		
APR 15 '80		
'81		
SEP 1 1 1982		Lib-26-67