<u>A Model of Coal Supply and Demand</u>

Appendix B

to

A Special Report to the Senate Finance Committee

Senator Oshel Craigo
Chair

Please Direct Questions to:
Michael Hicks, Ph.D.
Director, Applied Research

Center for Business and Economic Research
Lewis College of Business
400 Hal Greer Boulevard
Huntington, West Virginia 25755
(304) 696-6251
hicksm@marshall.edu

Introduction

This study attempts to simulate the short run impact of a variety of economic changes on coal production in southwestern West Virginia counties. A centerpiece of this effort was the construction of an econometric model of coal supply and demand that would capture key variables that influence the sale and mining of bituminous coal. This effort provides a basis for formulating simulations of the impact of changes in these key variables on coal production in individual counties in southern West Virginia. These include changes in cost, price, imports, exports and other factors affecting the mining and sale of West Virginia coal. By linking this model to individual counties we are able to simulate changes in coal production, wages and employment by adjusting these external variables. Changes in the external economic factors of supply and demand are provided by projecting trends externally and applying them to the model. The final step in this process is to estimate the county level impact on overall jobs, income and output as changes in coal production, employment and wages occur. This appendix outlines the modeling process for coal supply and demand presenting both the theoretical and econometric issues involved in its construction.

A key limitation in this effort was the dearth of monthly or quarterly time series data for several important variables. Similarly, data error, lumpiness and outright absence of critical data components suggest a variety of estimation and modeling techniques be employed to overcome these challenges. Two guiding principles aided in this process. The first was to employ conservative, defensible assumptions. The second was to adhere to existing, obvious institutional conditions where possible. These principles allowed us to impose restrictions on parameter estimates which realistically reflected effects that we observe, but cannot empirically model. When we apply restrictions generated by these assumptions we explicitly describe these restrictions in the text.[1] Since we are adopting this technique, we also performed a fragility analysis of the critical coefficients in the model.

[1]This techniques is commonly termed Bayesian estimation. The primary elements of Bayesian Estimations we employed involve the non-positive restriction on the price coefficient to adhere to institutional details which we have discussed in the text of this study. This turned out to be moot since this was among the consistent statistical results.

The Data

The selection of annual variables was necessitated by the data, see Table B-1.

Table B-1
Supply and Demand Variables

Variable	Description
WVCOALQ	county level coal production in tons
BTU	the BTU estimate, in SO_2 per ton of West Virginia coal, all pre-1986 values held constant, a proxy for quality
ELECD	Per capita electricity use in the United States in Kwh, a proxy for end use demand
BTUprice	The price per BTU of coal
Import	U.S. imports of coal in tons
Export	U.S. exports of coal in tons
Bondrate	The real rate on 6-month Commercial Bonds, a proxy for per capita capital costs
Minewage	The real annual wage of coal miners in West Virginia, a proxy for per unit labor costs
Tech	The residual from the basic underground production function, a proxy for technology shocks.
Umine	The number of underground mines, per county
Smine	The number of surface mines, per county
AR(n)	The n lagged autoregressive component

These data are available from a variety of sources noted in the reference section of this report.

Model Specification

We specified the following supply and demand model:

(Equation B-1)

$$Q_D = f(\text{BTU, BTUprice, ELECD, IMPORT, EXPORT})$$

$$Q_S = f(\text{BTUprice, Tech, Bondrate, Minewage, Umine, Smine})$$

$$Q_S^* = Q_D^*$$

In this specification, we assume that the quantities Q_D and Q_S are total industry output, which is further defined as:

(Equation B-2)

$$Q^* = \sum_{i=1}^{n} q_i$$

where total output in the industry is the sum of individual producer output. We assume that firms in West Virginia face a competitive market in which they are price takers.[2] This assumption of competitive markets permits us to estimate a partial equilibrium model. In this case total demand Q_D is a function of production, q_i in West Virginia. This suggests we introduce West Virginia coal production as an explanatory variable in our demand and supply equations above:

[2]The shift from long term production contracts to a futures based commodification offers some credible anecdotal evidence of this assumption. Similarly the rapid technological diffusion, homogeneity of product and large numbers of buyers and sellers suggests a high degree of at least *effective competition* in this market.

(Equation Set B-3)

$$Q_D = f(\text{BTU, BTUprice, ELECD, IMPORT, EXPORT, WVCOALQ}_d)$$

$$Q_S = f(\text{BTUprice, Tech, Bondrate, Minewage, Umine, Smine, WVCOALQ}_s)$$

$$Q_S^* = Q_D^*$$

This simple specification, in general form, is consistent with most modeling approaches for energy supply and demand in a partial equilibrium setting (see Varian, 1992, Silberburg, 1994). Since our efforts involve simulation of regional impacts to a variety of shocks, we are not interested in estimating demand and supply coefficients individually. We are instead searching for the reduced form of the equation which would yield a sensitivity coefficient of changes in external variables on the equilibrium quantity of coal. The parameter estimates provide some insight to the net impact of variables through their magnitude and direction. The specification will be in the first differences of the natural logarithm, so omitting that notation the model takes the form:

(Equation Set B-4)

$$\begin{aligned} Q_D = \alpha_1^i + \gamma_d \text{WVCOALQ}_D + \alpha_2 \text{BTU} + \alpha_3 \text{ELECD} + \\ \alpha_3 \text{BTUprice} + \alpha_5 I\text{MPORTS} + \alpha_6 \text{EXPORTS} \end{aligned}$$

$$\begin{aligned} Q_s = b_1^i + \gamma_s \text{WVCOALQ}_D + b_4 \text{BTUprice} + b_7 \text{Tech} + \\ b_8 \text{Bondrate} + b_9 \text{Minewage} + b_{10} \text{Umine} + b_{11} \text{Smine} \end{aligned}$$

Given the equality of supply and demand in equilibrium and our desire to estimate marginal effects on the exchanged quantity of coal, not demand and supply coefficients, a reduced form equation would seem useful. For our purposes a reduced form equation yields coefficients on each variable that allow us to estimate (in log-log form) the percentage change in West Virginia coal attributable to a one percent change in each explanatory variable. To this we added an autoregressive component, ϕ_{t-n}WVCOALQ$_{t-n}$. The reduced form equation takes the form:

(Equation B-5)

$$\text{WVCOALQ}^{*} = \frac{a_1^i + b_1^i}{\gamma_S + \gamma_D} + \frac{a_2}{\gamma_S + \gamma_D}(\text{BTU}) + \frac{a_3}{\gamma_S + \gamma_D}(\text{ELECD}) + \frac{a_4 + b_4}{\gamma_S + \gamma_D}(\text{BTUprice}) +$$
$$\frac{a_5}{\gamma_S + \gamma_D}(\text{IMPORTS}) + \frac{a_6}{\gamma_S + \gamma_D}(\text{EXPORTS}) + \frac{b_7}{\gamma_S + \gamma_D}(\text{Tech}) + \frac{b_8}{\gamma_S + \gamma_D}(\text{Bondrate}) +$$
$$\frac{b_9}{\gamma_S + \gamma_D}(\text{Minewage}) + \frac{b_{10}}{\gamma_S + \gamma_D}(\text{Umine}) + \frac{b_{11}}{\gamma_S + \gamma_D}(\text{Smine}) +$$
$$\frac{\phi_{t-n}}{\gamma_S + \gamma_D}(\text{WVCOALQ}_{\text{t}-\text{n}}) + \frac{u_{i_t} + e_t}{\gamma_S + \gamma_D}$$

The final term represents the composite error term for the model which is adjusted by the sums of the coefficient estimates of the regional supply and demand variables. We rewrite the expression, compressing the rather tedious coefficient notation into the following:

(Equation B-6)

$$\text{WVCOALQ}^{*} = B_1^i + B_2(\text{BTU}) + B_3(\text{ELECD}) + B_4(\text{BTUprice}) + B_5(\text{IMPORTS}) + B_6(\text{EXPORTS}) +$$
$$B_7(\text{Tech}) + B_8(\text{Bondrate}) + B_9(\text{Minewage}) + B_{10}(\text{Umine}) + B_{11}(\text{Smine}) +$$
$$B_{12}(\text{COALQ}_{\text{t}-\text{n}}) + e_j$$

From this form we can estimate our fixed effects model preserving the obvious cross sectional specific variation of county level coal output and number of mines.[3] The fixed effects model combines variation across counties (the cross sectional component) with intertemporal

[3]The use of county level variables recommends itself, econometrically, as a method of preserving degrees of freedom. Also, from an analytical standpoint strong cross county heterogeneity in the mix of surface and underground production suggests that some disaggregation is necessary.

variation (the time series component) in a series of intercept terms (B^i_1) that vary for each county. This method is recommended for a variety of technical reasons.[4] The remaining variables are estimated in aggregate (no county level variation). The result, in first differenced, log-log form gives us parameter estimates $B_2, B_3, \ldots B_{11}$ which are directly interpreted as the percentage change in annual output for West Virginia mines. The B_1 coefficients are the fixed effect adjustments, or county specific intercepts and the B_{12} coefficients are the matrix of autoregressive parameters (3 lagged components).

Unfortunately, this type of reduced form specification does not permit clear theoretical expectations regarding either the magnitude or sign of the parameter estimates. This is due to the fact that individual coefficient estimates capture combined supply and demand effects. We can impose restrictions on some of the coefficients to reflect current conditions -- a Bayesian approach. The restrictions we have placed on the parameters are illustrated in Table A-3. In essence these restrictions are directional effects observed in the data. Before examining the results a discussion of the relevant econometric techniques we have employed is necessary.

Econometric Methods

Early in the data collection process it became apparent that simple *ordinary least squares* estimates would be inappropriate for a variety of reasons. Chief among these was the absence of a long time series and the use of proxy variables for quality and capital structures. A substitute for *ordinary least squares* is a *weighted least squares* estimator that minimizes a weighted horizontal and vertical deviation from the estimated linear function. The *weighted least squares* estimator appears as:

[4]The data exhausts the population (this is not a sample estimate) and there is strong evidence to suggest cross-sectional correlation. Both of these conditions recommend the use of the fixed effects model.

(Equation B-7)

$$B_{wls} = \left(X'V^{-1}X\right)^{-1} X'V^{-1}y \quad \forall V^{-1} = [\ddots]\left(1/s_{11},\ldots,1/s_{nm}\right) \otimes I_t$$

$$where \operatorname{var}\left(B_{wls}\right) = \left(X'V^{-1}X\right)^{-1}$$

The *weighted least squares* estimator is efficient and consistent, but not asymptotically unbiased in a single equation model with autocorrelated or heteroscedastic errors (see Kmenta, 1986; Kennedy, 1996). This presents additional problems which we discuss later.

The use of a panel series with a number of cross sectional invariant parameters was immediately considered and subsequently adopted. For example, while we could determine the county level production, we could not determine county level (or state level) exports, and so used a national variable as proxy. This variable was not permitted to vary across counties in this model. The panel technique selected was the fixed effects model.5 Similarly, following a visual inspection of the data a first differenced, or de-trended estimation technique appeared appropriate. This was confirmed through an exhaustive set of unit-root tests.6 Similarly, a log-log specification was initially employed for its ease of interpretation (see Varian, 1992; Greene, 1994; Kennedy, 1996).

Deviations from the classical linear model also included the potential for autocorrelated errors, heteroscedastic errors and multi-collinearity. The latter fortunately was not clearly effecting any of the final model specifications.[7] The inclusion of autoregressive components in

---

[5]The Fixed effect model is appropriate when exhausting the study population, as we have done. Other reasons including autoregressive componeents recommend this choice, with no reasonable substitutes emerging.

[6]The augmented Dickey-Fuller tests clearly rejected the hypothesis of a unit root meaning that these variables possessed a time trend, or were non-stationary. The hypothesis of a unit root in first differences for each variable could not be rejected at high levels of significance, typically .01 percent. Since this process involved well over a hundred variables we have not included these texts in the report. The authors will provide these results upon request.

[7]Use of pricing variables specific to underground or above ground coal proved to be a nearly linear combination providing textbook test statistics. This was expected, and the weighted BTU price employed in subsequent estimations prove much more fruitful.

the estimation cleared the autcorrelation problem. This also eliminated inconsistent errors in the weighted least squares estimator. Confirmation of the absence of autocorrelated errors was performed through a *Hausman* test, taking the specification:

(Equation B-8)

$$Y_H = BX + \alpha u + e$$

where the original specification $Y = BX + u$ is re-estimated with the inclusion of the original residual u and a subsequent residual e. The hypothesis tested is $\alpha \neq 0$, of which a failure to reject implies autocorrelated errors, see Hausman (1978). The selection of optimal lag length for the autoregressive component simply involved optimizing goodness of fit measures.[8] Ensuing Durbin-Watson statistics confirming this process as correct.

A similarly easy step was the use of White's heteroscedasticity invariant standard errors in estimation:

(Equation B-9)

$$X_W = \frac{T}{T-K}\left(X'X\right)^{-1}\left[\sum_{t=1}^{T} u_t^2 x_t x_t'\right]\left(X'X\right)^{-1}$$

This matrix, $\mathbf{X_W}$, is employed to calculate the standard errors. This removes the inefficiencies noted in the *weighted least squares* estimator under conditions of heteroscedastically distributed errors, see White (1980). This cleared the final hurdle. All of these empirical procedures were programmed as an *a priori* step in estimation.

Estimation Results

Test statistics and Fixed effects intercepts appear in Table B-2. Results of estimation appear in Table B-3.

[8]Both the Akaike Information Criterion and Adjusted R^2 confirmed three lags as optimal for the autoregressive component.

Table B-2

Reduced Form Partial Equilibrium
Estimation-Test Statistics and Fixed Effects Intercepts

Variable	Fixed Effect Parameter Effect
Boone	-0.366112
Fayette	-0.347077
Kanawha	-0.346421
Logan	-0.376713
McDowell	-0.397479
Mingo	-0.340896
Raleigh	-0.325716
Wyoming	-0.473258
Nicholas	-0.576334
Adjusted R^2	0.951
SSR	5.139607
Durbin-Watson	1.675616
F-statistic	70.17537

Table B-3
Reduced Form Partial Equilibrium Estimation

Variable	Parameter	t-Statistic	Sign Restrictions
BTU	-7.156415	-8.905295	none
ELECD	12.11994	26.55491	+
BTUprice	-6.825508	-11.20871	-
IMPORTS	-2.646728	-4.926875	-
EXPORTS	5.631123	9.335303	+
Tech	-8.387726	-12.60774	none
Bondrate	-0.498122	-19.47004	none
Minewage	8.578024	14.91933	none
AR(1)	0.104289	2.953592	none
AR(2)	-0.428196	-5.547889	none
AR(3)	0.205078	4.847967	none
Boone-Smine	-0.264257	-4.078183	none
Fayette-Smine	1.010799	5.847004	none
Kanawha-Smine	-0.258738	-4.521325	none
Logan-Smine	-0.513246	-2.905828	none
McDowell-Smine	-0.223183	-13.71422	none
Mingo-Smine	-2.942116	-0.645253	none
Raleigh-Smine	0.006443	0.313773	none
Wyoming-Smine	0.306530	2.400216	none
Nicholas-Smine	-0.136784	-3.459549	none
Boone-Umine	-0.079204	-1.677866	none
Fayette-Umine	-1.379687	-6.754650	none
Kanawha-Umine	0.140287	10.18026	none
Logan-Umine	1.428418	3.202430	none
McDowell-Umine	-0.018499	-5.029905	none
Mingo-Umine	5.058209	0.618192	none
Raleigh-Umine	-0.218235	-2.633588	none
Wyoming-Umine	0.136997	3.393225	none
Nicholas-Umine	0.153277	2.318729	none

The *Hausman* specification test strongly rejected autocorrelation and a series of *Wald* tests on parameters strongly failed to reject misspecification. The proximal nature of the fixed effect intercepts suggests very similar specification for the individual components in the model. While this is an informal comparison, the two different counties, Nicholas and Wyoming, both experience a low level of output from primarily underground coal mining. We can infer that this production is more closely tied to metallurgical uses. Since we anticipate that metallurgical coal use is relatively price inelastic this modeling and simulation effort will not include this effect. That is why industrial use of coal for metallurgical purposes was not included in the original specification.

The strong performance of the test statistics gives us encouragement. However, a concern for robustness in this model continues our concern for mis-specification. This is the reason for the *Wald* and *Hausman* tests. Other tests including omitted variable and *Ramsey's* RESET test do not lend themselves to this type of panel estimation. We remain satisfied that the basic supply and demand specification with regional output is a reasonable method for this type of model.

Of course, the most important aspect of this model is its performance. We compared the model forecast on actual 1999 data subsequent to the construction of the model (the data was not available until April 2000). The difference between our simulation and the actual data were rather heartening. The total regional difference in actual and predicted output for 1999 was 1.061 percent. The county differences were higher, reflecting the stochastic nature of production at the disaggregated level. Not surprisingly, small counties and counties experiencing dramatic changes in the structure of firms had the largest one year forecast error. The forecast error evaluation appear in Table B-4.

Table B-4
Forecast Error Evaluation

	1998 Actual Production	1999 Actual Production	1999 Production Forecast	Percent Forecast Error
BooneCounty	29,420,756	30,075,908	28,929,488	-0.0389
Fayette County	3,361,209	2,018,613	3,369,061	0.4017
Kanawha County	13,754,041	15,059,145	13,795,196	-0.09189
Logan County	14,461,606	10,164,503	6,866,667	-0.2280
McDowell County	5,935,976	4,680,797	5,773,925	0.1841
Mingo County	22,645,873	20,225,684	21,557,381	0.0588
Raleigh County	12,932,085	10,646,135	13,238,539	0.2004
Wyoming County	10,936,625	9,987,079	10,444,128	0.04179
Nicholas County	2,759,970	4,523,533	2,173,820	-0.8513
Total	116,208,141	107,381,397	106,148,206	-0.0106

The discrete nature of production changes suggest that opening or closing of operations in one county may result in dramatic annual county level changes in output. As a region however, these changes appear to occur at a constant rate. That is why we observe large county variations between years, but smooth changes within the study region

There are concerns regarding this model. First, and foremost the short run nature of this specification and process cannot be overemphasized. This type of model is very appropriate for relatively brief periods of investigation. For that reason we have not extended the model beyond the short run horizon. Second, imposition of rather dramatic regulatory or economic disturbances to this system will likely disrupt the stability of the coefficient estimates. This is true of all models of this type however, and does not overly concern us, though it is worth mentioning. Finally, extending these results to other coal producing areas is not appropriate, because the fixed effects technique and the specification of this model do not lend themselves well to regional extrapolation.

Interpretation of Results

The coefficient or parameter estimate for each variable can be interpreted as the percentage change of county level output attributable to a 1 percent change in the given variable. In the main body of this study, the total output influence of this variable is illustrated as part of the core simulation. We will not repeat that step here, except to note that it is calculated by performing the partial derivative of the function with respect to the explanatory variable in question.

Since this is a reduced form model, parameter estimates do not necessarily conform to theoretical interpretation. Indeed calculating individual parameter estimates from the original supply and demand model is problematic at best. However, our inability to recover these variables is not valuable for our simulation. We seek not theoretical support for the supply and demand model but instead seek to determine the influence of a variety of factors on the production and sale of coal in these nine southwestern counties. In this section we are not trying to explain changes in actual levels of production . Such an explanation more correctly belongs in the main text of this study.

First, the effect of output on the BTU variable reflects a decline in the sale and production of WV coal as the SO_2 content increases. This suggests that environmental restrictions are slowing the sale of West Virginia coal faster than the burn quality attributes of higher BTU coal do increase the sales. This suggests that the equilibrium quantity of West Virginia coal are more affected by the regulatory environment than the desire by steam plant operators to use hotter burning fuels.

Second, electricity demand will continue to cause the sale of West Virginia coal to rise. The effect is large and statistically robust and is consistent with virtually all observations of both industries we have seen. Third, the price per BTU unit of coal (an international commodity) is negatively correlated with the sale of WV coal. We placed a non-positive restriction on this parameter estimate because we observe the demand side outpacing the supply side in this variable. Or, more clearly, in recent years output has risen while price has dropped.

Fourth, we clearly expected imports to negatively effect West Virginia coal sales and exports to increase sales. So, we placed these restrictions on both these variables, a result that was borne out empirically.

The next three variables proxy technology change, per unit capital and labor costs. In a supply function alone the first would be positive, the latter two negative, however in the reduced form model the sign cannot reasonably be theoretically determined. The sole important conclusion here is that mine wages are rising in response to sales, and that there does not appear to be a correlation between increasing wages for coal miners and reductions in output. This is likely due to productivity increases outpacing wages.

The autoregressive components are important due to the high fixed costs of coal mining. If a mine opens, it is likely to continue to operate even if the decision proves unprofitable in the years following its opening. As long as operating costs are covered, these mines are likely to remain open for some time. Similarly, the opening or closing of a mine has a heavy effect in a region. For this reason, we expect the autoregressive components to be statistically significant. There is little useful interpretation of this variable beyond its statistical significance

The remaining county specific variables simply list the number of surface and underground mines in each county. Inclusion of these variables is important for productive reasons (as part of the supply function) and permits simulations of individual mine closings across each county. The intercepts provide county level variation in output. All other interpretation of these variables is contained in the text of this study.

Conclusion

In this appendix we have specified a model of coal supply and demand in the short run. This modeling effort was directly aimed at providing a useful simulation tool, not an exhaustive study of the dynamics of coal production and demand. The short run nature of the model, data limitations and the simulation component influenced specification. Econometric tools, stability tests and modifications of errors in the classical linear model inherent in the data recommended several obvious techniques. A continued caveat of this type of modeling effort is its short run application. This is both the strength and weakness of the model. The use of this model for long run projections is strongly discouraged.