CEEPR Center for Energy and Environmental Policy Research

Panel Data Analysis of U.S. Coal Productivity

by

Thomas M. Stoker, Ernst R. Berndt, A. Denny Ellerman, Susanne M. Schennach

00-004 WP March 2000

A Joint Center of the Department of Economics, Laboratory for Energy and the Environment, and Sloan School of Management

Panel Data Analysis of U.S. Coal Productivity

Thomas M. Stoker Ernst R. Berndt A. Denny Ellerman
Susanne M. Schennach[*]

March, 2000

Abstract

We analyze labor productivity in coal mining in the United States using indices of productivity change associated with the concepts of panel data modeling. This approach is valuable when there is extensive heterogeneity in production units, as with coal mines. We find substantial returns to scale for coal mining in all geographical regions, and find that smooth technical progress is exhibited by estimates of the fixed effects for coal mining. We carry out a variety of diagnostic analyses of our basic model and primary modeling assumptions, using recently proposed methods for addressing `errors-in-variables' and `weak instrument bias' problems, as well as a new method for studying errors-in-variables in nonlinear contexts.

1. Introduction

The coal mining industry in the United States is a remarkably dynamic industry. In particular, labor productivity grew steadily at an annual rate of 5.36% from

[*] Sloan School of Management, MIT, 50 Memorial Drive, Cambridge, MA 02142 USA. Berndt and Stoker are Professors, Sloan School of Management, Ellerman is Assistant Director, MIT Center for Energy and Environmental Research (CEEPR) and Schennach is a PhD student in the Department of Economics, MIT. We are grateful to seminar audiences at several MIT CEEPR meetings, the NBER-CRIW Conference on Productivity, and the IFS in London, England. We have received valuable comments from William Bruno, William Dix, Larry Rosenblum and Dale Jorgenson. Finally, we thank Zvi Griliches for comments that motivated much of the diagnostic work in Section 3.

1978-1995,[1] after some decline in the early 1970's. This high rate of productivity growth makes coal mining's experience comparable to sectors whose advances are more well known, such as consumer electronics. As shown in Figure 1, the rate of improvement has been accompanied by the strong growth of coal output from 1972-1995, in the face of falling coal prices from 1975-1995.

The technology for mining coal varies greatly across the United States, which gives rise to several possible sources for explaining the dramatic productivity growth. At the most basic level, coal[2] deposits vary in size, shape and accessibility depending on the specific geology of each mine location. In terms of overall technology, mines are either surface mines or underground mines. Underground mines are further categoried by mining process; the traditional continuous process or the more recent longwall mining process. In addition to these overall categorizations, each mine location has specific characteristics that affect mining technique, equipment design and plant configuration, depending on the nature of the coal deposit itself. The size of the mining deposit, as well as the life of a mine in a particular location, varies from site to site.

To analyze the sources of productivity growth in U.S. coal mining, we believe it is extremely important to account for heterogeneity across mines. In addition, when there is extensive heterogeneity, it is not immediately clear how to measure sources of productivity growth. This paper discusses an econometric analysis of U.S. coal mining, and defines interpretable sources of productivity growth consistent with concepts drawn from panel data analysis.[3]

We employ a data set that is in some ways extremely rich and in other ways very limited. We observe the annual output and labor input of every coal mine in the United States from 1972-1995. In addition to mine location, we have identified

[1] This is a conservative estimate based on averaging productivity from the eleven coal mining groups defined below. In terms of total tons and total labor hours in the United States, the productivity growth rate from 1978-1995 was 6.81% per year.

[2] We view coal as a homogeneous commodity, after controlling for heat content. In a study of the demand for coal, sulphur content would be an important differentiating feature. Here we do not explicitly separate out coal type relative to production (aside from regional origin and lignite, a particularly low quality coal).

[3] The importance of recognizing heterogeneity in coal mining dates back at least to the use of (British) regional data in Leser (1955) or (U.S.) state data in Madalla (1965). For work on productivity using aggregate data, see Jorgenson, Gollop and Fraumeni (1987), and Jorgenson (1990). For studies of data from states and individual mines, see Kruvant, Moody and Valentine (1982), Baker (1981), Byrnes, Fare and Grosskopf (1984) and Boyd (1987). Boyd (1987) gives a detailed analysis of Illinois strip mines, that documents substantial economies of scale.

the overall technology used in each mine; namely suface mining, underground continuous mining and underground longwall mining. However, we do not observe measures of capital in use at each mine, nor do we observe details on local geology or the configuration of specific production facilities. For these reasons, we focus on labor productivity of individual mines, and develop methods that take into account heterogeneity in the data.[4]

We model labor productivity separately for groups of mines defined by regional location and overall type of technology. Mine-specific fixed effects are included to further capture the myriad of heterogeneous features (geology and different types of capital configuration), and time effects are included to capture group-wide productivity variation. We define indices of productivity change in line with the panel model concepts; fixed effects, scale effects and time effects. With this framework, our results give an interesting depiction of smooth, uniform technological progress in coal mining as it affects labor productivity over the period 1972-1995, as well as an assessment of the importance of scale economies and embodied technological improvements in physical capital. Our basic modeling does rest on an important specification assumption, and we examine how sensitive our results are to that assumption.

Section 2 gives describes our data, spells out our modeling assumptions and gives our overall results.[5] Section 3 follows with numerous diagnostic methods to judge the sensitivity of our results to key assumptions, including analysis of weak instruments and nonlinear errors-in-variables. Section 4 gives some concluding remarks.

2. Panel Data Analysis of Productivity

2.1. Data Specifics

The data on coal mine output and labor input is collected by the Mine Safety and Health Administration (MSHA) as part of its mandated regulatory effort since 1972. Coal output is measured in clean short tons, and for aggregating

[4] We discuss many of the salient aspects of our data below. See Ellerman, Stoker and Berndt (1999) for more details on the construction of our data. They also list the various government publications used in the references.

[5] Ellerman, Stoker and Berndt (1999) gives much more detail on the specifics of the coal industry and the data, as well as results for specific mining regions.

output across regions, coal output is (quality) adjusted for heat content.[6] Labor is measured in hours, and we do not distinguish di®erent types of labor.

We observe mine location, as well whether the mine is a surface mine or an underground mine. Surface mining involves substantially di®erent technology than underground mining. In a surface mine, the overburden (earth) is stripped back to reveal the coal seam, and the overburden is put back in place after the coal is mined. This makes surface mining similar to modern road building or other surface development projects. Underground mines employ either continuous mining or longwall mining methods, depending on the nature of the coal deposit. Continuous mines employ machines that remove coal from the seam and pass it back to a shuttle car or conveyor belt system. This system requires tunnels, with some coal left in place as pillars to support the roof of the mine. Longwall mining uses an elaborate shearing device that operates along an extended face (a \long wall"); with the entire device moving through the coal seam (and the roof capsizing behind it). Continuous mining is the traditional technique, which still plays an important role because of its suitability under many mining conditions, whereas longwall mining is a more recent technique that had been introduced in Europe and adapted in the United States over the time frame of our data.[7]

Unfortunately, the basic MSHA data does not identify which underground mines are longwall mines, and so we carried out an identi¯cation by the (arduous) matching of speci¯c mine locations with longwall installations reported in Coal Age magazine and other industry publications. In addition, in the MSHA data, a few details of mine facilities are observed (e.g. presence of a preparation plant), but there is no information on overall capital inputs (plant and equipment) to the mines. The only geological feature observed was seam height, but that data appeared to be of very poor quality and was not used in the analysis. Finally, we constructed an annual coal price index for each region, and used a national wage series to proxy labor cost changes.[8]

[6] These adjustments are indicated in Table 1. It is important to note that these adjustments do not impinge on our statistical modeling and estimation, but only apply when output is aggregated across regions.

[7] See DOE/EIA 0588 (95), Longwall Mining, for more details.

[8] The price data is contructed from annual mine-mouth coal prices by state as collected by the Energy Information Adminstration of the U.S. Department of Energy. Wage data is from the Employment, Hours and Earnings series published by the Bureau of Labor Statistics. These data are de°ated to real prices and wages using the consumer price index. See Ellerman, Stoker and Berndt (1999) for more details.

Because of the importance of location and the overall mining techology, we segmented the data into eleven group of mines, defined as follows. Nine groups are formed by classification of the three regions — Appalachia (APP), Interior (INT), and West (WST) — along with the three overall technologies — surface mining (S), underground continuous mining (CM) and longwall mining (LW). We separated out two special surface mining groups for the analysis, the Powder River Basin (PRB) and lignite coal (LIG). As indicated below, the PRB has experienced the most spectacular growth (with somewhat inferior coal) and lignite is a substantially inferior coal in terms of heat content. Figure 2 shows a map of the United States with the three major regions, the PRB and the lignite producing areas. All in all, there are 85,968 total annual observations on 19,230 individual mines. Table 1 provides the composition of the sample in terms of the eleven mine groups. All estimation is performed within each group, and for combining results across groups, tons of coal are weighted by the average Btu content given in Table 1.

The U.S. coal industry has changed in a dramatic fashion from 1972-1995. Figure 3 shows the composition of the of overall output growth. In terms of mining regions (Figure 3A), there has been truly spectacular output growth in the PRB. Output has increased in all other mining regions except for the interior. In terms of mining technique (Figure 3B), there is strong growth in output from longwall mines, and strong growth in output from surface mines. Continuous mines have held at roughly the same overall output throughout the period.

Figure 4 displays labor productivity in coal mining. The productivity levels (Figure 4A) vary considerably across groups. Clearly, the most productive region is the PRB, and the surface mining groups are more productive than other groups. Hence, part of the increase in overall labor productivity is due to increased output from the PRB, lignite and other surface mining. However, the normalized productivities (Figure 4B) indicate that the groups with the most change in productivity are the underground mines. Part of the increase in aggregate productivity is due to these changes, since those groups did not decrease in coal output share. Finally, Figure 5 illustrates average annual mine output for the mining groups. The PRB has the largest mines, and has shown exceptional growth in per-mine output. The other mining groups have average increases in output scale, although not as pronounced as the PRB.

While compositional shifts clearly play a role in the increasing labor productivity, it is clear that productivity changes within groups are relevant as well.

Each mining group is changing in its character over time. New mines open and older mines close, and dramatic scale changes occur within individual mines. In order to understand the productivity process at the level of individual mines, we now turn to our empirical modeling and results.

2.2. The Empirical Model

In the analysis of firm level data from a competitive industry, it is natural to assume that output and inputs are endogenously determined given prevailing output and input prices. Not only is this approach infeasible with our data, due to lack of data on wages at specific locations or output prices net of transportation costs, we also believe it would seriously misrepresent institutional features of the market for coal output. In particular, the majority of coal output is set in advance by contracts with specific buyers.[9] As such, we assume that output is predetermined, and that labor (and other inputs) are set endogenously to produce the necessary output at minimum cost.[10] This is a key assumption of our approach, and its failure would lead to biases in estimation. In Section 3, we examine the sensitivity of our main findings to this assumption as well as other similar issues, such as mis-measurement of output.

Because of our overall aims, we focus on modeling labor productivity. Let Q_{it} and L_{it} denote observed output and labor hours input for mine i at time t, giving labor productivity as Q_{it}/L_{it}. Our analysis is based on the model

$$\ln\left(\frac{Q_{it}}{L_{it}}\right) = \tau_t + \alpha_i + F(\ln Q_{it}) + \varepsilon_{it} \tag{2.1}$$

where $t = T_i^O, \ldots T_i^C;\ \ i = 1, \ldots, N$, and we assume that ε_{it} has mean zero and variance σ^2 conditional on τ_t, α_i, $\ln Q_{it}$. Here $T_i^O, \ldots T_i^C$ denotes the years that mine

[9] See Joskow (1987, 1990). The role of multi-year contracts has decreased over time, but it is still very large. In 1994, 78% of all coal deliveries to electric utilities were under contracts of greater than one year's duration. Electric utility deliveries account for about 80% of total production, but the arrangements for coal sold in the export, metallurgical and industrial markets are similar. Although there is some variability in the quantities to be delivered under these contracts, that variability reflects the demand for electricity from the powerplant being supplied, which in turn reflects the level of economic activity and the weather.

[10] This assumption also neglects potential endogeneity due to the choice of whether to open a new mine or not, or shut down an existing mine. We discuss below how this feature is partially accommodated by mine fixed effects, but we do not model this process explicitly.

i is in operation (positive output). The time e®ect τ_t and the mine e®ect α_i are treated as ¯xed e®ects in estimation. The unknown function $F(\cdot)$ relates output scale changes to productivity changes, and will be treated nonparametrically in estimation. Recall that estimation is group speci¯c (so that N, τ_t, $F(\cdot)$ and σ^2 vary by mine group).

The time e®ects τ_t are designed to capture group-wide changes in the level of overall productivity over time. The mine-speci¯c ¯xed e®ects α_i provide our accounting for geological formations (or ease of mining at site i) and speci¯c features of capital used at site i. In particular, a new mine will typically make use of the best available technology for the site — aside from the overall decision of what form of mining (e.g. underground continuous or underground longwall), the capital will embody the current state of technology on several other dimensions, such as delivery systems for transporting the coal from the seam face to outside the mine mouth. While some features can evolve over a mine's life (and arguably could be proxied by the scale Q_{it}), the ¯xed e®ects capture embodied technical change in new mine capital as well as speci¯c geological features.

Another issue partially addressed by the ¯xed e®ect modeling is the phenomena of turnover in the mining industry. For example, average mine life in our data is 4.5 years, and our equation for mines in operation could include a term for selection bias based on mine pro¯tability. However, we have no information on speci¯c depletion pro¯les, which are determined by mine geology, nor do we have a clear accounting on external concerns of the investment environment that would lead to closing down a mine. As such, we cannot model a selection term directly. However, to the extent that the probability of continuing operation is determined by mine speci¯c factors, or factors common across mines in each time period, such a selection term will be subsumed into the ¯xed e®ects α_i and τ_t.

The time e®ect τ_t could capture many di®erent phenomena. A substantial amount of safety regulation was applied to the coal industry in the early 1970's, and that regulation applied di®erentially to underground and surface mines. In addition, there is common variation in coal prices, which could impact mining practice, with a consequent impact on productivity. In order to portray how the time e®ects vary with regard to prices and other phenomena, we adopt a two-stage modeling approach. The ¯rst stage is the main estimation is of model (2.1), which produces estimates $\hat{\tau}_t$ of the time e®ects τ_t. The second stage relates the estimates of time e®ects to observed prices.

This decidedly empirical strategy can be understood as follows. We employ the model

$$\tau_t = \kappa + \gamma_p \ln p_t + \gamma_w \ln w_t + \delta D_t + \eta_t, \quad t = 1, \ldots, T \tag{2.2}$$

where p_t is the real coal price, w_t is the real wage rate, and D_t is a dummy variable for 1972-1973,[11] and where the panel estimates $\hat{\tau}_t$ are used in place of the true values τ_t. One way of viewing our results (and the way that underlies any interpretation of standard errors) is that the term η_t is a standard homoskedastic disturbance with mean 0 conditional on $\ln p_t$, $\ln w_t$ and D_t, and that our two stage approach is potentially inefficient because it does not impose the structure of (2.2) in the estimation of the main model (2.1). Another way to view this approach is to view our second stage as just giving an OLS decomposition of the time effects for interpretation. Namely, the time effects are estimated independently of equation (2.2); we use OLS to decompose the time effects as

$$\begin{aligned} \hat{\tau}_t &= \text{"Price Effects"} + \text{"Other Effects"} \\ &= \hat{\kappa} + \hat{\gamma}_p \ln p_t + \hat{\gamma}_w \ln w_t + \hat{\delta} D_t \quad + \quad \hat{\eta}_t \end{aligned} \tag{2.3}$$

where $\hat{\kappa}$, $\hat{\gamma}_p$, $\hat{\gamma}_w$, $\hat{\delta}$ are the OLS estimates and $\hat{\eta}_t$ is the OLS residual. In any case, we do not have a specific model of how prices cause changes in mining productivity, but our results give a summary of the price "effects" constructed in this way.

2.3. Estimation Details

Estimation of the parameters of the panel model (2.1) is entirely standard, aside from the unknown function $F(\cdot)$. To give flexible treatment of this function, we approximate it by a polynomial in log output. We choose the order of the polynomial (for each group of mines) by least squares cross validation. Namely, we choose the order d of the polynomial to minimize

$$SS(d) = \sum_{i=1}^{N} \sum_{t=T_i^O}^{T^C} \left[\hat{\tau}_t^{(-it)} + \hat{\alpha}_i^{(-it)} + \hat{F}_d^{(-it)}(\ln Q_{it}) - \ln\left(\frac{Q_{it}}{L_{it}}\right)\right]^2$$

where $\hat{*}^{(-it)}$ refers to the least squares estimator computed by omitting the it^{th}

[11] We found D_t to be empirically necessary, and interpret it as accounting for the change in coal outlook from the four-fold increase of oil prices in late 1973.

observation, and F_d is a polynomial of degree d.[12] This process led to the choice of polynomials at most of order 3, with $F(\cdot)$ specified as

$$F(\ln Q_{it}) = \beta_1 \ln Q_{it} + \beta_2 (\ln Q_{it})^2 + \beta_3 (\ln Q_{it})^3$$

More specifically, for six of the groups, a cubic polynomial was chosen; for three groups, a quadratic polynomial was indicated ($\beta_3 = 0$) and for two groups, a linear function was indicated ($\beta_2 = \beta_3 = 0$). Having determined the order of the polynomial for each mine group, we estimate the polynomial coefficients by OLS.

The scale estimates are presented in Table 2. While it is clear that all coefficients are estimated precisely, it is difficult to interpret what the estimated pattern of scale effects are from the polynomial coefficients. A good method is to plot the estimated functions $\hat{F}$, and we include such plots later in Figure 7 of the diagnostic section. It is worthwhile mentioning here that all estimates are consistent with substantial economies of scale,[13] and that cubic estimates have the same S shape for different regions, implying that an intermediate range of scales is associated with greatest productivity improvement.

The results of the second stage of estimation, regressions of estimated time effects on log prices and wages, are given in Table 3. We see that price effects are estimated to be negative for all but one region, and that wage effects are typically positive, although everywhere imprecise. This is consistent with the notion that high real coal prices will allow less efficient operations to be profitable, as will low real wages. Also, we show the results of testing the restriction $\gamma_p = -\gamma_w$ for each region, or whether the coal price and wage effects are adequately summarized by the impact of the price/wage ratio. While these estimates are clearly reduced form in nature, we view the patterns as interesting and informative. Specifically, as real coal prices increase (decrease), ceteris paribus, less (more) productive mines are in operation.

[12] Least squares cross-validation is a common method for choosing parameters of nonparametric estimators of density and regression; see Silverman (1986) among others. We made use of the computational algorithms given in Green and Silverman (1994, p. 3–35), and considered polynomials up to order five.

[13] For the log-linear specifications (APP-LW and INT-LW), overall scale elasticities are substantially greater than one (1.471 and 1.333 respectively).

2.4. Panel Model Decomposition of Productivity Change

The estimates of the model (2.1) - (2.2) give a full empirical description of productivity in the U.S. coal industry. However, the estimates themselves are not very helpful in understanding what are the predominant influences on coal productivity. Our approach is to define indices that are conceptually aligned with the panel model structure. Our objective is to obtain a clear depiction of the sources of productivity growth from the indices.

For each mining group, overall labor productivity is expressed as

$$\frac{\sum_i Q_{it}}{\sum_i L_{it}} = \frac{\sum_i L_{it}\exp\left[\ln \frac{Q_{it}}{L_{it}}\right]}{\sum_i L_{it}} = \frac{\sum_i L_{it}\exp\left[\hat{\lambda}_t + \hat{\alpha}_i + \hat{F}(\ln Q_{it}) + \hat{\varepsilon}_{it}\right]}{\sum_i L_{it}}$$

$$= \frac{\sum_i L_{it}\exp\left[\hat{\alpha}_i + \hat{F}(\ln Q_{it}) + \hat{\varepsilon}_{it}\right]}{\sum_i L_{it}} \cdot \exp(\hat{\lambda}_t) \qquad (2.4)$$

where "^"s denote the panel data estimates. Overall labor productivity thus decomposes into two factors; one for mine-specific productivity factors and the other for common time-varying trends. We now examine these two factors in more detail.

The first factor of (2.4) reflects elements that vary across mines; namely geology and embodied capital technology, efficiencies associated with scale, and all other features of productivity that vary over mines. This term does not decompose exactly, and so we approximate it in a fashion consistent with $\hat{\alpha}_i$, $\ln Q_{it}$ and $\hat{\varepsilon}_{it}$ being independently distributed across mines (weighted by labor hours). In particular, we consider

$$\frac{\sum_i L_{it}\exp\left[\hat{\alpha}_i + \hat{F}(\ln Q_{it}) + \hat{\varepsilon}_{it}\right]}{\sum_i L_{it}} \simeq FE_t \cdot SC_t \cdot MR_t \qquad (2.5)$$

where

$$FE_t = \frac{\sum_i L_{it}\exp[\hat{\alpha}_i]}{\sum_i L_{it}} \qquad (2.6)$$

defines the Fixed Effect Index,

$$SC_t = \frac{\sum_i L_{it}\exp\left[\hat{F}(\ln Q_{it})\right]}{\sum_i L_{it}} \qquad (2.7)$$

de¯nes the Scale E®ect Index and

$$MR_t = \frac{\sum_i L_{it} \exp[\hat{"}_{it}]}{\sum_i L_{it}} \qquad (2.8)$$

de¯nes the Residual Microheterogeneity Index.

As the ¯xed e®ects ®$_i$ represent the base levels of productivity for each mine, the index FE_t of (2.6) re°ects how those base levels vary over time. For instance, if coal mining technology were stable over time, and the more productive sites were mined ¯rst, then FE_t would decline over time. Alternatively, if site selection were unrelated with mine productivity (say dictated by changing demands from population migration and transportation costs), then FE_t would increase as the (embodied) technology of new mining capital increased. Since ®$_i$ captures both geology and initial technology levels, FE_t summarizes how those conditions vary over the time period of interest.

Productivity improvements associated speci¯cally with increases in scale are indicated by the index SC_t. It is natural to think of scale e®ects as a combination of technology and mine-speci¯c learning e®ects. Namely, it can take time to learn the most e®ective way of mining a given site, such as in designing the system for conveying coal out of the mine and away from the site, and such processes can di®er for a young mine versus a more mature mine.[14]

The index MR_t summarizes the role of the residual in the log-productivity regression. We include it primarily as a check on whether overall impacts of ¯xed e®ects and scale e®ects are large relative to the residual.[15]

The second term of (2.4) represents the transformation of the time e®ect relevant for comparing to the above indices; we could de¯ne the Time E®ect Index directly as

$$TE_t = \exp(\hat{¿}_t) \qquad (2.9)$$

[14] It is possible, although we feel unlikely (given our estimates), for the scale index to capture the adverse productivity e®ects of depletion of coal at a given site. The main reason for this is that as reserves are depleted at a given site, it is typical for smaller contractors to take over the mining, using di®erent techniques for isolated pockets of coal. In the MSHA data, this is accounted for as the closing of the original mine, and the opening of a new mine associated with the smaller contractor { i.e., the depletion e®ects are not retained in a single mine's observations. In any case, we have no information on the initial size of reserves at a given site, which would be necessary to isolate the depletion e®ect.

[15] MR_t typically will re°ect changes in the variance of $"_{it}$ over time. For instance, if the (labor weighted) distribution of $"_{it}$ were normal with mean 0 and variance $¾_t^2$ at time t, then up to sampling error, $MR_t \simeq \exp(¾_t^2/2)$.

From the earlier decomposition (2.3), we express the time effect index in terms of price effects and other effects, as

$$
\begin{aligned}
TE_t &= \exp\left(\hat{\gamma} + \hat{\gamma}_p \ln p_t + \hat{\gamma}_w \ln w_t + \hat{\delta} D_t\right) \cdot \exp(\hat{\gamma}_t) \\
&= P_t \cdot R_t \qquad (2.10)
\end{aligned}
$$

where,

$$P_t = \exp\left(\hat{\gamma} + \hat{\gamma}_p \ln p_t + \hat{\gamma}_w \ln w_t + \hat{\delta} D_t\right) \qquad (2.11)$$

defines the Price Effect Index and

$$R_t = \exp(\hat{\gamma}_t) \qquad (2.12)$$

defines the Residual Time Effect Index. Again, these indices permit the relative size of the time effects versus fixed and scale effects to be judged.

These various indices constitute an empirically-based method of assessing the importance of the different factors: scale, fixed effects, prices, residual, etc. in the overall labor productivity changes observed in the coal mining industry. To assess the accuracy of our approximation, we define the Predicted Productivity Index as the product

$$PP_t = FE_t \cdot SC_t \cdot MR_t \cdot TE_t = FE_t \cdot SC_t \cdot MR_t \cdot P_t \cdot R_t \qquad (2.13)$$

The difference between observed labor productivity and the predicted index is the approximation error in (2.5).

2.5. Sources of Labor Productivity Changes in U.S. Coal Mining

The results from applying our productivity indices to coal mining in the United States are given in Figure 6.[16] All indices are normalized to 1 in 1972. One initial conclusion is that the approximation error in (2.5) seems of little concern; while there are some differences, the predicted productivity index (dashed line) has the same time pattern as the observed labor productivity (solid line).

The most interesting time pattern in Figure 6 is that of the fixed effect index FE_t. Despite the large oscillation in observed productivity, FE_t grows smoothly through the sample time period. Since this index represents geological conditions

[16] Specifically, the indices are computed for each mining group and are then aggregated in the same way as labor productivity values (using Btu weights, etc.).

and the level of technology of new mines, and since it is unlikely that inferior geological sites are chosen before superior sites, the FE_t index gives a plausible rendition of continuous (embodied) technical improvements in mining capital over the full period 1972-1995.[17]

The scale index SC_t drops slowly through the early years (possibly because of decreased output associated with new environmental regulations), but then begins a steady increase over the period 1978-1995. Finally, the price index P_t shows substantial variation, initially dropping rapidly, leveling out, and then increasing steadily from 1981 onward.

The residual indices MR_t and R_t show relatively minor variation. MR_t decreases gradually and then increases gradually and returns to its initial level. R_t varies substantively over the first three years but then hovers around 1; after 1976 or so the impact of the time effects is given by the price index P_t. At any rate, the main movements in aggregate productivity seem well captured by the three indices FE_t, SC_t and P_t.

It is straightforward to carry out an analysis of the productivity indices at the level of the mining groups; such an analysis is summarized in Ellerman, Stoker and Berndt (2000), and we do not go into details here. We look at one feature that shows how the productivity indices can aid insight into the process of technological advance. In particular, we have interpreted the fixed effect index as reflecting technical progress in capital of new mines, and the scale effect index as associated with improvements in capital that are associated with scale increases. It is natural to hypothesize that these features are related to average mine life. Namely, for regions with short mine lives, there is less scope for improvements associated with scale enhancements than for regions with long mine lives. In Table 4, we show the growth of the fixed effect index and of the scale effect index by region over the period 1972-1995. We find precisely the hypothesized relationship - namely scale-related productivity improvements predominate in groups with long mine lives, whereas improvements in initial capital (fixed effects) predominate in areas with shorter mine lives.

[17] If inferior sites were chosen prior to superior sites, the rate of embodiment of technical improvements would be even larger.

3. Diagnostics on the Log Productivity Relationship

The model underlying our productivity analysis is decidedly simple, and the results are interesting. In part because of the simplicity of our model, one can envision several potential problems with the results. Given that we have estimated the model using a goodness-of-fit criterion, such problems center on the interpretation of our results, which relies on our assumption that output is predetermined in our estimation procedure. As part of judging this assumption, we applied many new (and old) techniques for studying endogeneity problems, such as those in the literature on weak instrument bias. We now discuss much of this work. However, it is worth stating at the front that we did not find compelling evidence against our basic results, using any of the diagnostic methods.

3.1. Traditional Linear Methods

3.1.1. Interpretation of the Productivity-Scale Effect

While we estimate log labor productivity equations that are nonlinear in the log of output for some groups, we begin with some diagnostics appropriate for log linear specifications (we return to the nonlinear versions in Section 3.2). For this, it is useful to consider our model in the context of familiar Cobb-Douglas formulae. Suppose that the production function for a coal mine is specified as

$$Q^* = A(L^*)^{\omega}(K_1^*)^{\rho_1}\cdots(K_M^*)^{\rho_M} \tag{3.1}$$

where L^* is labor hours, $K_1^*, \ldots, K_M^*$ represent small equipment and other variable inputs, and A can include fixed inputs. The "scale elasticity" for all variable inputs is

$$\eta = \omega + \rho_1 + \cdots + \rho_M. \tag{3.2}$$

Minimizing total cost $TC = wL^* + \sum_{i=1}^{M} r_j K_j^*$ subject to predetermined output in (3.1) gives the following expression for log labor:

$$\ln L^* = -a + \left(\frac{1}{\eta}\right)\ln Q^* \tag{3.3}$$

where a depends on A and input prices.[18] The resulting expression for log labor productivity is

$$\ln\left(\frac{Q^*}{L^*}\right) = a + \left(1 - \frac{1}{\gamma}\right)\ln Q^*. \tag{3.4}$$

Thus, returns-to-scale with regard to variable inputs is captured in the coefficient of log output; returns are decreasing, constant or increasing if $\beta_1 = (1 - 1/\gamma)$ is negative, zero or positive, respectively. In log linear form, our model is an implementation of (3.4)[19], and our strongly positive estimates of β_1 are consistent with substantial economies of scale.

3.1.2. Errors-in-Variables and Bracketing

Our first diagnostic procedure is to examine whether traditional bracketing results are consistent with economics of scale.[20] We begin by examining the most basic implications of errors-in-variables in this framework. Suppose that true log output and labor are denoted $q^* = \ln Q^*$, $l^* = \ln L^*$ respectively, and true log labor productivity is $pr^* = q^* - l^*$. Write (3.4) as

$$pr^* = \alpha + \beta_1 q^* + \varepsilon \tag{3.5}$$

where α is an intercept and ε is a homoskedastic disturbance obeying $E(\varepsilon|q^*) = 0$ (i.e. we have set $a = \alpha + \varepsilon$). Suppose that observed log output $q = \ln Q$ and log labor $l = \ln L$ are given as

$$\begin{aligned} q &= q^* + v \\ l &= l^* + \varepsilon^* \end{aligned} \tag{3.6}$$

where v, ε^* are homoskedastic errors that have mean 0 conditional on q^* and l^*. We set $\eta = \varepsilon - \varepsilon^*$ and assume that $\mathrm{Cov}(v;\eta) = 0$.

[18] Specifically

$$a = \frac{\ln A + \sum_j \rho_j \ln\left(\frac{w\rho_j}{r_j}\right)}{\gamma}.$$

[19] There a is specified with effects for time, mine, and the disturbance as $a = \tau_t + \alpha_i + \varepsilon_{it}$.

[20] Bracketing results are well known in econometrics, since at least Frisch (1934). See Griliches and Ringstad (1971) for applications of bracketing results to production problems similar to ours.

For summarizing traditional error-in-variables results, denote percentages of error variance as follows:

$$\lambda_q = \frac{Var(v)}{Var(q)}; \quad \lambda_l = \frac{Var(\varepsilon)}{Var(l)} \tag{3.7}$$

Suppose that $\hat{b}_{lq}$ denotes the OLS coefficient of l on q. Then the traditional bias result is

$$\text{plim}\, \hat{b}_{lq} = \left(\frac{1}{\eta}\right)(1-\lambda_q) = (1-\beta_1)(1-\lambda_q) \tag{3.8}$$

and for $\hat{b}_{ql}$,

$$\text{plim}\, \hat{b}_{ql} = \eta(1-\lambda_l) = \left(\frac{1}{1-\beta_1}\right)(1-\lambda_l) \tag{3.9}$$

These give rise to the standard bracketing formula as

$$\text{plim}\, \hat{b}_{lq} \le 1-\beta_1 \le \frac{1}{\text{plim}\, \hat{b}_{ql}} \tag{3.10}$$

For studying the log productivity regression, we have that $\hat{b}_{pr;q} = 1-\hat{b}_{lq}$, so that

$$\text{plim}\, \hat{b}_{pr;q} = 1 - \text{plim}\, \hat{b}_{lq} = \beta_1 + \lambda_q(1-\beta_1). \tag{3.11}$$

Since it is natural to assume $\beta_1 < 1$, errors in observed output values bias the log productivity coefficient upward. If there are constant returns to scale, then $\beta_1 = 0$, and $\text{plim}\, \hat{b}_{pr;q} = \lambda_q$; in that case, errors in observed log output could give a spurious finding of estimated increasing returns. In general, from (3.10), we have a bracketing relationship for the true coefficient, β_1:

$$1 - \frac{1}{\text{plim}\, \hat{b}_{ql}} \le \beta_1 \le 1 - \text{plim}\, \hat{b}_{lq} \tag{3.12}$$

Our equations contain fixed mine and time effects, and so even with a log-linear scale specification, they would not fit within the simple bivariate regression framework above. We examine the bracketing relationships by using the residuals of $\ln L$ and $\ln Q$ regressed on all mine and time effects in the role of l and q above, which is associated with assuming that errors in $\ln L$ and $\ln Q$ are uncorrelated with the mine and time effects.[21] Further, in addition to computing the estimates of the lower and upper bounds of (3.12):

$$LB = 1 - \frac{1}{\hat{\beta}_{ql}} \; ; \quad UB = 1 - \hat{\beta}_{lq} \tag{3.13}$$

we also compute bounds that are adjusted (widened) to include sampling error in the regression coefficients, namely

$$ALB = 1 - \frac{1}{\hat{\beta}_{ql}} - c \cdot s_{\hat{\beta}_{ql}} \left(\frac{1}{\hat{\beta}_{ql}^2} \right) \; ; \quad AUB = 1 - \hat{\beta}_{lq} + c \cdot s_{\hat{\beta}_{lq}} \tag{3.14}$$

where $s_{\hat{\beta}_{ql}}$, $s_{\hat{\beta}_{lq}}$ are the estimated standard errors of $\hat{\beta}_{ql}$, $\hat{\beta}_{lq}$, the lower bound adjustment follows from the delta method, and $c = 1.96$ is chosen for an approximate 95% confidence interval.

The bounding results are presented in Table 5. The bracketing bounds are fairly wide, which is consistent with the overall goodness-of-fit of the equations. However, on the question of the evidence on returns to scale, the constant returns value $\beta_1 = 0$ is contained in the intervals for only two of the eleven mine groups, namely APP-S and WST-LW. Even in these two cases the bounding intervals contain mostly positive values,[22] and we view it as reasonable to conclude that our finding of increasing returns is not spurious.

Nevertheless, the bracketing bounds are quite wide, and so we now turn to various methods of estimating the scale effect directly.

3.1.3. Instrumental Variables Estimates

We begin with some simple instrumental variables estimations of the scale effect. Since we do not observe a separate indicator of output, we make the assumption that any measurement error is uncorrelated across time periods, and use linear

[21] This is not equivalent to just subtracting within-mine and within-time averages, because of the unbalanced nature of our panel data.

[22] For instance, consider the implications for error variances in the two groups APP-S and WST-LW. If we ignore sampling error, the value of $\beta_1 = 0$ is consistent with error variance percentages of $\lambda_l = .0457$ and $\lambda_q = .286$ for APP-S, and $\lambda_l = .00396$ and $\lambda_q = .373$ for WST-LW. Thus the vast majority of measurement error must be in log quantity to give constant returns.

combinations of lagged outputs as instruments.[23] It is tempting to first orthogonalize the left- and right-hand sides with respect to the time and fixed effects, as above, before proceding with the instrumental variable estimation. However, orthogonalization with respect to the fixed effect could tend to induce correlation in measurement errors across time periods, which could invalidate the use of lagged quantities as intrumental variables. Consequently, we remove the fixed mine effects by estimating the model in first differenced form, and keep the time effects as regressors.

We therefore have the model:

$$\Delta pr^*_{it} = \beta_1 \Delta q^*_{it} + \Delta \tau_t + \Delta \varepsilon_{it} \tag{3.15}$$

where as above, $pr^*_{it} = \ln(Q^*_{it}/L^*_{it}) = q^*_{it} - l^*_{it}$, and Δ denotes the first difference operator ($\Delta x_{i,t} = x_{i,t} - x_{i,t-1}$). Observed log-output $q_{it} = \ln Q_{it}$ and log-labor $l_{it} = \ln L_{it}$ are measured with error, as

$$\begin{aligned} q_{it} &= q^*_{it} + \nu_{it} \\ l_{it} &= l^*_{it} + \varepsilon^*_{it} \end{aligned} \tag{3.16}$$

where errors are uncorrelated over time and over mines,

$$E[\varepsilon^*_{js}\varepsilon^*_{it}] = 0 \text{ and } E[\nu_{js}\nu_{it}] = 0 \text{ when either } i \neq j \text{ or } s \neq t; \tag{3.17}$$

uncorrelated across log labor and log output,

$$E[\varepsilon^*_{js}\nu_{it}] = 0 \text{ for all } i,j \text{ and } s,t; \tag{3.18}$$

and errors are uncorrelated with true values of log-output and log-labor,

$$E[l^*_{js}\varepsilon^*_{it}] = 0;\ E[q^*_{js}\varepsilon^*_{it}] = 0;\ E[q^*_{js}\nu_{it}] = 0;\ E[l^*_{js}\nu_{it}] = 0; \tag{3.19}$$

for all s,t,i,j.

Under these assumptions, potential instruments for Δq_{it} are any linear combinations of:

$$q_{is} \text{ for } s < t - 1 \text{ and } s > t. \tag{3.20}$$

Clearly all these variables satisfy the requirement that their measurement errors are uncorrelated with the measurement error on the variable they instrument.

[23] See Keane and Runkle (1992) among many others.

Various assumptions can guarantee that these instruments are also correlated with the variable they instrument. For instance, if q^*_{it} follows a stationary process of the form:

$$\left(q^*_{i,t} - \bar{q}^*_i\right) = \rho\left(q^*_{i,t-1} - \bar{q}^*_i\right) + \xi_{it}$$

where ξ_{it} are iid, then q_{it} is correlated with q_{is}; and as long as the process is stationary ($\rho < 1$), q_{is} (with $s < t - 1$) is a valid instrument for Δq_{it}. Note that these instruments become weaker as the process for q_{it} becomes more persistent ($\rho \to 1$). In principle, the future log-outputs q_{is} with $s > t$; could also be used be used as instruments.[24]

From the choice of possible instruments for Δq_{it} discussed above, we began by using the twice lagged difference

$$\Delta q_{i,t-2}$$

and also by using the associate log-output levels

$$q_{i,t-2} \text{ and } q_{i,t-3}$$

We do not consider longer lags in order to avoid dropping too many observations, which is a problem with our sample since many mines have short lives.

The 2SLS estimates of the scale effect $\beta_1$are given in Table 6. For some regions, the point estimates are very close to the OLS estimates, and in others the 2SLS estimates are very imprecise. On the issue of increasing returns, in five regions (APP-S, APP-CM, LIG, WST-S, WST-LW) the 95% confidence intervals for β_1 clearly exclude $\beta_1 = 0$; using both lagged levels and lagged first differences as instruments. In one region (WST-CM) only the lagged level instruments exclude $\beta_1 = 0$ at a 90% confidence level. For the other five regions the scale effect estimate is very imprecise and the possibility of the constant returns to scale is not excluded. The problem originates from the weakness of the instruments: for those regions, q_{is} appears to follow a very persistent process (ρ near to 1), which makes the correlation between Δq_{it} and $\Delta q_{i,t-2}$ or $q_{i,t-2}; q_{i,t-3}$ small (as shown in Table 3). This weak correlation can translate into imprecise estimation of the scale effect.

To obtain more efficient IV estimators, we include powers of the lagged levels or powers of the lagged first-differences as instruments. For these additional

[24] Note that the case $s > t$ becomes problematic when the right-hand side of the model contains lagged values of the left hand-side variable, which is one reason why instruments with $s > t$ are seldom used in the literature.

instruments to be valid, we need the following additional assumption:

$$E\left[q_{is}^{d}\varepsilon_{it}\right] = 0 \text{ for } t \neq s \text{ and } d = 0, 1 \ldots$$

A sufficient condition for this is that $q_{is}^{*}, \varepsilon_{is}$ and ε_{it} are mutually statistically independent.[25]

Estimates of the scale effect using the expanded instrument sets are also presented in Table 6. These estimates are generally much more precise. In all but three regions (INT-S, INT-CM, and PRB), the hypothesis $\beta_1 = 0$ is rejected using power series in both lagged differences and lagged levels. In the cases of the PRB and INT-CM regions, only the estimates with power series of lagged levels as instruments reject the hypothesis, while in the case of the INT-S region the hypothesis is not rejected for either set of instruments.

Within the context of the log-linear model, we have uncovered no substantial evidence to doubt our finding of increasing returns in mining on the basis of errors-in-variables as modelled above. We obtain fairly precise estimates of the scale effect for all but one region (INT-S). However, the correlation coefficients between the instruments and log output appear to be quite small,[26] and the increasing precision from using the lagged log-output powers as instruments is a bit surprising. One possibility has been a currrent focus of the literature, namely that we may be in a situation of weak instrument bias, as discussed in Nelson and Startz (1990). In particular, the concern is that instruments may exhibit small sample correlations with the measurement error, which impart a small sample bias of IV estimates toward the OLS estimated values. Such a bias is exacerbated when the number of instruments is increasing (holding their joint explanatory power constant). We now examine this issue in our data.

Bound, Jaeger and Baker (1995) present approximations to the finite-sample bias of IV estimates, while Stock and Staiger (1997) derive an asymptotic 2SLS bias when the correlation between the instruments and the endogenous variables is modeled to be zero. In either case, the F-statistic of the first step regression (here log-output regressed on the instruments) is found to provide a good indication of whether weak instrument bias is present. An F-statistic close to 1 indicates that

[25] Note that the same independence assumption validates both the use of powers of lagged levels as well as powers of lagged differences.

[26] A table of all correlations between log output and all instruments used (lags and powers) is available from the authors. In any case, a large percentage of the simple correlations are smaller than 0.1 in absolute value.

the bias of IV estimates relative to the OLS bias is significant; namely, that

$$\frac{E\left[\hat{\beta}_{1;2SLS} - \beta_1\right]}{E\left[\hat{\beta}_{1;OLS} - \beta_1\right]} \qquad (3.21)$$

is significantly different from 0.

To address this, in Table 7 we report the F-statistics from first step regressions for our estimates in Table 6. We first demeaned the data to remove time effects, so these are the appropriate partial F-statistics. However, the F-statistic criterion is strictly justified only in the absence of heteroskedasticity and other error problems, which we have not ruled out, so these results may best be viewed as suggestive. Neverthless, for three of the four cases when the 2SLS estimates indicate statistically significant returns to scale (INT-LW, INT-CM, PRB), the first stage F-statistic is quite low even for the power series. There may be a problem for two of the regions, PRB and INT-LW, where the F-statistics are close to 1, which suggests that 2SLS may not lead to much improvement on any bias in OLS estimates. It is also interesting to note that for several of the other mining groups, the F-statistic actually decreases as more instruments are added.

Table 7 also addresses another question on the specification of instruments. When estimating a panel data model with instrumental variables, it is fairly common practice to include each year of observation as a distinct column in the instrument matrix. This allows for a flexible (time-varying) correlation structure between instruments and regressors (as opposed to fixed correlation), and seemingly provides a greater number of instrumental variables.[27] The 2SLS results presented in Table 6 did not process the instruments in this way, and indeed, when we applied this approach to our data, the estimates of scale effects were broadly similar but we observed a dramatic reduction in the asymptotic standard errors of our estimates. However, such an apparent increase in efficiency comes at the cost of much greater scope for problems from weak instrument bias. The final column of Table 7 shows that using a separate instrument for each year typically implies decreases in the F statistics;[28] we observe sharp declines for 5 regions, declines for 7 and a substantial increase in the F-statistic for only 2 regions. In

[27] For instance, if l is the maximum lag used as an instrument, distinguishing each year of observation implies that one (original) instrumental variable is associated with $\left(T_i^C - T_i^O - l + 1\right)$ columns in the final instrument matrix.

[28] The instrument here is the twice lagged level of output.

any case, we have chosen to focus our reporting on estimates using the original instruments (smaller set, not separated by year).

We could continue to add other instruments, such as other powers or more lags (and decreasing the estimation sample size), to try to increase the explanatory power of the instrumental variables. Instead, we appeal to a recent solution to this problem due to Blundell and Bond (1998), namely enhancing the instrument list systematically with a GMM approach.

3.1.4. Generalized Method of Moments

We begin by considering our basic model

$$pr_{it} = \beta_1 q_{it} + \alpha_i + \tau_t + \varepsilon_{it} + \nu^*_{it} \tag{3.22}$$

where ε_{it} and ν^*_{it} are homoskedastic and uncorrelated over time, and q_{it} is uncorrelated with ε_{it}, but q_{it} is potentially correlated with the measurement error term ν^*_{it}. Later we consider the estimation of a model with more general disturbance structure.

The traditional 2SLS approach for estimating model (3.22), using lagged levels as instruments for equations in first-differenced form, is based on moment conditions of the form

$$E\left[q_{is} \Delta \nu^*_{it}\right] = 0 \text{ for } s < t - 1. \tag{3.23}$$

The approach of Blundell and Bond (1998) is to introduce the following additional moment conditions:

$$E\left[\Delta q_{is} \nu^*_{it}\right] = 0 \text{ for } s < t, \tag{3.24}$$

which amounts to using lagged first differences as instruments for equations in level form. Abstracting from the time effects in (3.22), these two sets of conditions are combined via a linear system GMM estimator as:[29]

$$\hat{\beta} = \left[\left(\sum_i W_i'Z_i\right)\left(\sum_i Z_i'Z_i\right)^{-1}\left(\sum_i Z_i'W_i\right)\right]^{-1}\left(\sum_i W_i'Z_i\right)\left(\sum_i Z_i'Z_i\right)^{-1}\left(\sum_i Z_i'y_i\right) \tag{3.25}$$

where

$$\begin{aligned}
W_i &= \begin{bmatrix} \Delta q_{i,\underline{T}_i+l} & \cdots & \Delta q_{i,\overline{T}_i} & q_{i,\underline{T}_i+l} & \cdots & q_{i,\overline{T}_i} \end{bmatrix}' \\
Z_i &= \begin{bmatrix} q_{i,\underline{T}_i} \;\cdots\; q_{i,\overline{T}_i-l} & 0 \\ 0 & \Delta q_{i,\underline{T}_i+1} \;\cdots\; \Delta q_{i,\overline{T}_i-l+1} \end{bmatrix}' \\
y_i &= \begin{bmatrix} \Delta pr_{i,\underline{T}_i+l} & \cdots & \Delta pr_{i,\overline{T}_i} & pr_{i,\underline{T}_i+l} & \cdots & pr_{i,\overline{T}_i} \end{bmatrix}';
\end{aligned} \tag{3.26}$$

and l is the maximum lag used as an instrument. In this way, the regressor matrix W_i, the dependent variable matrix y_i and the instrument matrix Z_i contain the information pertaining to equations in first-differenced and in level form. In the actual estimation procedure, we also include time effects by including time dummies in both W_i and Z_i (the time dummies are both regressors and instruments).[30]

We computed estimates using the DPD98 software of Arellano and Bond (1998), using $l = 2; 3$ lags, and the results are presented in Table 8. The GMM estimates are by and large more precise versions of our previous 2SLS results: the GMM estimates are not statistically significantly different from our previous 2SLS estimates and have much smaller standard errors.[31] As well, the GMM estimates are not systematically smaller than our OLS estimates of the log productivity equations in first-differenced form, which suggests that measurement error has

[29] The system GMM estimator considered by Blundell and Bond (1998) is slightly more general and asymptotically more efficient. The term $Z_i'Z_i$ is replaced by $Z_i'\hat{H}_iZ_i$, where $\hat{H}_i$ is a consistent estimate of the correlation structure of the residuals which can be obtained from the residuals of a first-step estimation where H_i is initially set to an identity matrix. However, as the same authors also noted, at typical sample sizes, the estimates obtained from the first-step tend to provide more reliable standard errors than the two-step estimates. Hence, we will use first step estimators in what follows, setting H_i to an identity matrix.

[30] As discussed previously, we did not introduce separate columns for each instrumental variable for each year, because of the potential for weak instrument bias.

[31] Two of the regions (WST-LW and INT-LW) are omitted because they posed numerical difficulties associated with the small number of observations available.

not significantly biased our results. Finally, the hypothesis of constant returns to scale can be clearly rejected in all regions.

Since the additional moment restrictions of the GMM estimator makes the system overidentified, the validity of the instruments can be tested. Our diagnostic approach is based on the assumption that measurement errors are serially uncorrelated, so our main concern here is for the additional moment conditions used in the system GMM approach. Blundell and Bond (1998) show that stationarity of the q_{it} process is sufficient for those moment conditions, which is a rather strong condition. In any case, we can directly assess the validity of the conditions via Sargan tests, as presented in Table 9. Clearly, the Sargan test fails to invalidate the instruments at the 90% level in all but three regions (APP-CM, INT-S, INT-CM). Note, however, that even if the test indicates rejection, there might not be a significant bias imparted to the estimates of the scale effect; that is, correlations between instruments and residuals could be statistically detectable while the absolute magnitude of the bias induced could still be small. Since GMM results are so similar to the 2SLS results, it appears that the bias potentially introduced by the new moment conditions is small relative to the scale effect itself, and likely represents a small price to pay for the enhanced precision of the estimates.

Table 9 also presents the results of testing for the presence of second-order serial correlation, for which we see substantive evidence in three regions. In itself, the presence of such serial correlation is only a serious problem if the serially correlated component of the error term is itself correlated with q_{it}, in violation of our basic error assumptions. We now examine this issue in more detail.

The possibility of the presence of a moving average error term correlated with the regressors can be investigated by using longer lags as instruments. Table 6 shows that using instruments lagged by one more year still clearly rejects the hypothesis of constant returns to scale in most mining groups. As such, there is no strong evidence against our conclusions here.

We can examine the same issue with autocorrelated errors by generalizing the model directly, as

$$\begin{aligned} pr_{it} &= \beta_1 q_{it} + \alpha_i + \tau_t + \varepsilon_{it} + \nu^*_{it} \\ \varepsilon_{it} &= \rho \varepsilon_{it-1} + \xi^*_{it} \end{aligned} \tag{3.27}$$

where ξ^*_{it} and ν^*_{it} are homoskedastic and uncorrelated over time, and q_{it} is potentially correlated with the productivity shock ξ^*_{it} and with the measurement error ν^*_{it}: This equation can be rewritten in a so-called "dynamic form" to make the error term MA(1):

$$pr_{it} = \rho \cdot pr_{it-1} + \beta q_{it} + \gamma q_{it-1} + (1 - \rho)\alpha_i + \tau_t - \rho\tau_{t-1} + \nu^*_{it} - \rho\nu^*_{it-1} + \xi^*_{it} \quad (3.28)$$

where

$$\gamma = -\rho\beta \quad (3.29)$$

is a nonlinear restriction on the parameters.

The GMM estimates of model (3.28) are also presented in Table 8. While the estimates of returns to scale differ somewhat from the estimates of the basic model (Equation (3.22)), they still clearly exclude constant returns to scale. More importantly, the estimates of the more general model are not systematically smaller than estimates for the basic model, suggesting that the problem of potential correlation between the autoregressive component of the error term and q_{it}, even if it were present, does not lead to spurious returns to scale.

It is worthwhile noting a few features of the model (3.27). First, this may not be the best way to generalize - namely, retaining a log-linear model but generalizing the error structure. We, in fact, do find some serial correlation in the log-linear model residuals, but that could easily arise from nonlinearities that exist in the true data relationships. We study nonlinearity with error-in-variables directly in the next section. Second, given log-linearity, this model may be more general than necessary; the estimates of (3.27) that we compute permit possible correlation between log-quantity q_{it} and the error ε_{it}, an issue with which we have hitherto not been concerned.

In any case, by adding additional moment conditions, the system GMM estimator allows us substantively to improve the efficiency of our estimates, without dramatically increasing the number of instruments. The estimates of returns to scale are not systematically smaller than the results for OLS on first differences, suggesting that measurement error is not a major issue in our data. The additional moment conditions introduced to improve efficiency are not contradicted for most of the mining groups, which supports foundation of the GMM system estimates. The finding of increasing returns to scale is not affected by more general models that allow, for instance, for autocorrelation and endogeneity.

In sum, while measurement errors may be present, there is no evidence that they are of sufficient magnitude to alter our basic finding of substantial economies of scale in every coal mining group.

3.2. Non-Linear Model Diagnostics

The above diagnostic section would su±ce if all of our estimated productivity relationships were log-linear. However, in our basic estimation results we found log-linearity to hold for only two mining groups, with three groups exhibiting a quadratic relationship and six groups a cubic relationship in log-output. In a non-linear context, the diagnosis of potential problems from measurement errors and the like is, if anything, quite daunting. The main known solutions for nonlinear models require su±cient assumptions and information to precisely measure the amount of measurement error, such as an independent measurement on the regressor of interest. We do not have such an instrument, and while we discuss this later, we are precluded from obtaining consistent point estimates of the quadratic and cubic models here.

However, we are able to obtain a clear understanding of some implications of measurement error to our results. Namely, we begin by assuming that the quadratic and cubic equations speci¯cations are, in fact, the true speci¯cations of the productivity relationship. For an assumed levels of measurement error, we can disentangle that error from our estimates, and learn what the productivity relationship would be without the measurement error. We develop this procedure in some detail next, and illustrate what occurs with several di®erent amounts of measurement error.

3.2.1. The Impact of Measurement Error

In a non-linear model, the simple bracketing technique of Section 3.1.2 (using reverse regression) is not directly applicable. Klepper and Leamer(1984) have generalized the idea of bounding the true regression coe±cients in the case of multiple regressors with uncorrelated measurement errors. Bekker, Kapteyn and Wansbeek (1987) have extended their ¯ndings to the case of correlated measurement errors in the variables, which would be necessary for a cubic speci¯cation such as ours, since a positive error in log-output q implies a positive error in all regressors of the form q^d. However, this method does not handle the case of polynomial regressions in a fully satisfactory manner: the knowledge that the regressors are powers of the same variable measured with error is not used. Only information about the correlation between the errors on each regressor is used; higher order moments are ignored.

For this reason, we study measurement error issues using results available from polynomial regression. In particular, given a level of error variance, we can adjust the original least squares estimates to consistent estimates. We study the impact of measurement error by examining the adjusted estimates for several different levels of measurement error.

We use derivations similar to Hausman, Ichimura, Newey and Powell (1991) and Chescher (1998) for adjusting the estimates. We do not have repeated observations for estimating measurement error, or particularly compelling instruments, so we carry our adjustments by assuming the distribution of the measurement errors.[32]

We present the specific adjustment formulae in the Appendix. The framework is as follows: we begin with a polynomial model;

$$pr = \sum_{i=1} \beta_i (q^*)^i + \sum_{l=1} \delta_l z_l + \varepsilon. \quad (3.30)$$

where the z's are regressors, which we will later set to all time and mine-specific fixed effects. For our purposes, observed log-output q is true log-output q^* measured with error as

$$q = q^* + \nu .$$

We must make the following specific assumptions about the measurement error:

A1: ε is independent of q^*, z and ν and is such that $E[\varepsilon] = 0$,

A2: ν is independent of q^* and z

A3: ν is distributed as a $N(0; \sigma^2)$, where σ^2 is known.

Assumption A3 could easily be replaced by another distributional assumption (with known polynomial moments).

To adjust our original estimates for measurement error, we express the observed (polynomial) moments of q in terms of variation of q^* and ν. For different levels of error variance, we can solve for the relevant moments of q^*, and then

[32] Our approach is reminiscent of Griliches and Ringstad (1970). See Hausman, Newey and Powell (1995) for approaches using instruments. Other references include Newey (1993), Lewbel (1996), Wang and Hsiao (1995) and Li (1998). Schennach (2000) develops an error adjustment process for general models using fourier transforms.

compute the polynomial coefficients that would have arisen if q^* were observed; giving the adjustments for the assumed level of error variance. The details of this calculation are given in the Appendix.[33] Clearly, this will give a consistent estimation scheme in a large sample (when the error variance is known).

We apply this method with q as observed log output and z the set of time and fixed effect dummies. We assume that the true model for each mining group is a polynomial of the order estimated by cross-validation, and examine how the estimates would be adjusted if a known amount of measurement error were introduced. In particular, we set the measurement error variance to be 0%, 5% and 10% of the variance of the observed log-output deviations (orthogonal to the fixed and time effects).

Figure 7 gives the log-productivity - log-output relationships adjusted for measurement error. The heaviest line gives the relationship from our basic (OLS) results (namely 0% measurement error), and the other lines give values adjusted for measurement error. For the mine groups with log-linear models (APP-LW, INT-LW), we see the downward slope adjustment as implied by (3.11). For the nonlinear models, the adjustments are particularly interesting. Specifically, while there are some differences for low scales, the main differences in shape occur at high output levels. For high output levels, the relationships adjusted for measurement error approach constant returns to scale (zero slope). It is clear that with 5% measurement error, there is no range of output for any mining group where constant returns to scale exists, but 10% error does show constant returns for high output levels in some groups.

The similarity of the shape of the log-productivity — log scale relationship is a bit surprising given the amount of error assumed by design. Because of the polynomial forms, 5% or 10% measurement error in q is not as small as it seems. For instance, for 10% measurement error, we have

$$\frac{Var\left[(q^*)^2\right]}{Var\,(q)^2} = .80 \text{ and } \frac{Var\left[(q^*)^3\right]}{Var\,(q)^3} = .71$$

so the induced measurement error throughout all regressor terms is much higher than 10%. Of course, the joint variation in q, q^2 and q^3 is complicated, so these variances do not tell the whole story.

In any case, we see that measurement error, as structured above, could have a significant impact on our results. Unfortunately, our data does not include

[33] The Appendix derivation is similar to the derivation given in Cheng and Schneeweiss (1998).

sufficient information to estimate the error variance, and therefore we cannot settle this issue once and for all.

3.2.2. Direct Estimation of the Coefficients

As mentioned above, we are unaware of an instrumental variables solution to the estimation of the coefficients of a nonlinear model in the presence of measurement error. As was first noted by Amemiya (1985), traditional instrumental techniques are unable to tackle measurement errors in non-linear specifications. To estimate a nonlinear function $g(\cdot)$, the problem that arises is that the error $g(q) - g(q^*)$ is typically correlated with q^*, unless $g(\cdot)$ happens to be linear. Moreover, the productivity relationship of interest here is $g(q^*)$, not the mean of log productivity conditional on log-output q.[34] As shown by Hausman, Newey and Powell (1995), this problem can be circumvented in polynomial regressions, where the correlation between powers of a variable and errors on these power takes a known functional form, but this technique requires an indicator or an instrument that is linearly related to the variable measured with error.

In our case, we used power series as instruments to obtain sufficiently efficient estimators in Section 3. Since the quantities we are using as instruments are just as likely to be measured with error, we face the problem of a non-linear regression with measurement error in the instrumental equation as well. It is not clear at the moment how this problem could be solved. With other, independent indicators of log-output, Hausman, Newey and Powell's (1995) method would provide a solution.

3.3. Variations in Productivity Measurements

In the various specifications studied as part of our diagnostic exercise, we found some differences with our main results. We have interpreted these differences as being fairly minor, and not indicative of any serious problems with our original results. However, as before, looking directly at the estimated scale effects may not be the best way of judging the differences we have seen. In this section we examine the implications of those differences for our overall depiction of productivity change in the coal industry.

[34] Namely, the techniques of Newey, Powell and Vella (1999) would be applicable if the conditional mean were of interest and q were endogenous. Those techniques do not apply in the measurement error problem.

Figure 8 shows the evolution of our productivity indices for ¯ve di®erent sets of estimation results. The \Within" estimates refer to our basic (nonlinear) estimates from Table 2 (indices presented earlier in Figure 6), and serve as a benchmark for comparison. \No scale" refers to indices constructed by assuming no scale e®ect on productivity; namely constant returns to scale in all mining groups. Two sets of results from the log linear panel speci¯cation for all groups are presented; \Linear 1st Di®." refers to OLS estimates of the linear model in ¯rst-di®erenced form (estimates from ¯rst column of Table 6), and \Linear GMM" refers to the basic Blundell-Bond estimates (¯rst column of Table 8[35]). Finally, \Nonlinear Meas. Error 10%" refers to the scale e®ects adjusted for 10% independent measurement error, as displayed in Figure 8.

In broad terms, the di®erent estimates are not associated with dramatically di®erent interpretatations of productivity change in coal mining. Somewhat surprisingly, the time pattern of the ¯xed e®ects indices are quite similar over time, exhibiting growth rates in the narrow range of 1.67% - 1.91% per year. The growth patterns of the scale indices are as follows; most growth with the log-linear estimates, followed by the nonlinear models (original results as well as the measurement error results), followed ¯nally by the no growth \no scale" simulation. In particular, the log-linear models tend to overstate the role of scale relative to nonlinear models.[36] These ¯gures also show a tendency for o®sets between the scale e®ect indices and the time e®ect indices. Speci¯cally, estimates associated with greatest growth in the scale e®ect indices are associated with the least variation in the time e®ects indices, and vice versa. Indices computed from our nonlinear estimates fall into the middle range of growth for scale indices and time e®ect indices.

4. Concluding Remarks

This paper has presented an empirical analysis of labor productivity in U.S. coal mining. The overall motivation for this work is the explanation of observed changes in labor productivity from 1972 through 1995, and particularly the strik-

[35] These are computed from the 9 mine groups for which estimates were obtained.

[36] One interesting feature to note is how there is no drop in the scale index for the nonlinear model adjusted for measurement error. Since the only substantive di®erence in those estimates was for large scale, this implies that the drop for other estimates arises from a pull back in larger scale mines in the early 1970's.

ing productivity increase after 1978. We began with data coverage of annual output and labor input for every coal mine in the U.S., and studied productivity with panel regression methods. Panel methods provide straightforward channeling of heterogeneity into fixed effects for mines, and time effects. We then proposed the use of productivity indices based on the parameter estimates from the panel model analysis. Of particular interest was the fixed effect index, that showed how (average) fixed effect values for mines in operation increased uniformly over the time period, which we interpreted as representing progress embodied in capital available for mines at their start date. The scale index reflected the productivity gains associated directly with output scale increases. Between 1972 and1995, we found that virtually all the change in observed labor productivity was captured by those two indices (Figure 6). This is true but a bit misleading; when examining the period 1978 to 1995 of rapid productivity increase, we find comparable, essentially uniform increases in fixed effect, scale effect and time (price) effect productivity indices.

Our model of labor productivity was nonlinear but reasonably simple, in part because of lack of information on capital for each mine. Because of the simplicity, we found that many recent proposals for model diagnostics were applicable, and so we carried out many such tests and analyses. While we did not find any strong evidence against our original estimates, we believe that the application of a battery of checks; bounds, weak instruments, improved point estimates, and nonlinear adjustments is sufficiently illustrative to benefit researchers facing similar kinds of modeling/data situations.

References

[1] Amemiya, T. (1985), "Instrumental Variables Estimator for the Nonlinear Errors-in-Variables Model," Journal of Econometrics, 28, 273-289.

[2] Arellano, M. and S.R. Bond (1998), "Dynamic Panel Data Estimation using DPD98 for GAUSS," draft, Institute for Fiscal Studies, London.

[3] Baker, J. G. (1981), "Sources of Deep Coal Mine Productivity Change, 1962-1975," The Energy Journal, 2, 95-106.

[4] Bekker, P., A. Kapteyn and T. Wansbeek (1987), \Consistent Sets of Estimates for Regressions with Correlated or Uncorrelated Measurement Errors in Arbitrary Subsets of all Variables," Econometrica, 55, 1223-1230.

[5] Blundell, R. and S. Bond (1998), \Initial conditions and moment restrictions in dynamic panel data models," Journal of Econometrics, 87, 115-44.

[6] Bound, J., D.A. Jaeger and R.M. Baker (1995), \Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," Journal of the American Statistical Association, 90, 443-450.

[7] Boyd, G.A. (1987), \Factor Intensity and Site Geology as Determinants of Returns to Scale in Coal Mining," Review of Economics and Statistics, 69, 18-23.

[8] Byrnes, P., R. Fare and S. Grosskopf (1984), \Measuring Productive E±ciency: An Application to Illinois Strip Mines," Management Science, 30, 671-681.

[9] Cheng, C-L and H. Schneeweiss (1998), \Polynomial Regression with Errors in the Variables," Journal of the Royal Statistical Society B, 60 (part 1), 189-199.

[10] Chescher, A. (1998), \Polynomial Regression with Covariate Measurement Error," Discussion Paper 98/448, University of Bristol.

[11] Coal Age, \U.S. Longwall Census." Annual feature in this and predecessor publications, Intertec Publishing Corporation, Chicago.

[12] Ellerman, D., T.M. Stoker and E. Berndt (2000), \Sources of Productivity Growth in the American Coal Industry: 1972-1995," in Dean, E., M. Harper and C. Hulten (eds.), New Directions in Productivity Analysis, National Bureau of Eeconomic Research, (forthcoming), presently available as MIT-CEEPR Working Paper 98-004.

[13] Frisch, R. (1934), Statistical Con°uence Analysis by Means of Complete Regression Systems, Oslo, University Economics Institute, Publication No. 5.

[14] Griliches, Z. and V. Ringstad (1970), \Errors in Variables in Nonlinear Context," Econometrica, 38, 368-370.

[15] Griliches, Z. and V. Ringstad (1971), Economies of Scale and the Form of the Production Function, Amsterdam, North Holland.

[16] Green, P.J. and B.W. Silverman (1994), Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach, London, Chapman and Hall.

[17] Hausman, J., H. Ichimura, W.K. Newey and J.L. Powell (1991), \Measurement Errors in Polynomial Regression Models," Journal of Econometrics, 50, 271-295.

[18] Hausman, J., W.K. Newey and J.L. Powell (1995), \Nonlinear Errors in Variables: Estimation of Some Engel Curves," Journal of Econometrics, 65, 205-233.

[19] Jorgenson, D.W. (1990), \Productivity and Economic Growth," in E.R. Berndt and J.E. Triplett, eds., Fifty Years of Economic Measurement: The Jubilee of the Conference of Research in Income and Wealth, Chicago, University of Chicago Press, 19-118. Reprinted in Jorgenson, D.W., Productivity: International Comparisons of Economic Growth, vol 2, Cambridge, The MIT Press, 1995, 1-98.

[20] Jorgenson, D.W., F.M. Gollop and B.M. Fraumeni (1987), Productivity and U.S. Economic Growth, Cambridge, Harvard University Press.

[21] Joskow, P.L. (1987), \Contract Duration and Relationship-Speci¯c Investments: Empirical Evidence for Coal Markets," American Economic Review, 77, 168-185.

[22] Joskow, P.L. (1990), \The Performance of Long-Term Contracts: Further Evidence from Coal Markets," Rand Journal of Economics, 27, 251-274.

[23] Keane, M.P. and D.E. Runkle (1992), \On the Estimation of Panel-Data Models With Serial Correlation When Instruments Are Not Strictly Exogenous," Journal of Business and Economic Statistics, 10, 1-29.

[24] Keystone Coal Industry Manual (1996). Intertec Publishing Company, Chicago, IL.

[25] Klepper, S. and E.E. Leamer(1984), \Consistent Sets of Estimates for Regressions with Errors in All Variables," Econometrica, 52, 163-184.

[26] Kruvant, W.J., C.E. Moody and P.L. Valentine (1982), \Sources of Productivity Decline in U.S. Coal Mining, 1972-1977," The Energy Journal, 3, 53-70.

[27] Leser, C.E.V. (1955), \Production Functions and British Coal Mining," Econometrica, 23, 442-446.

[28] Lewbel, A. (1996), \Demand Estimation with Expenditure Measurement Errors on the Left and Right Hand Side," The Review of Economics and Statistics, 78, 718-725.

[29] Li, T. (1998), "Estimation of Nonlinear Errors-in-Variables Models: A Semiparametric Minimum Distance Estimator," Working Paper, Washington State University.

[30] Madalla, G.S. (1965), \Productivity and Technological Change in the Bituminous Coal Industry, 1919-1954," Journal of the Political Economy, 73, 352-365.

[31] Nelson, C.R. and R. Startz (1990), \The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument is a Poor One," Journal of Business, 63, S125.

[32] Newey, W.K. (1993), \Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models," MIT Department of Economics Working Paper 93-18.

[33] Newey, W.K., J.L. Powell and F. Vella (1999), \Nonparametric Estimation of Triangular Simultaneous Equations Models," Econometrica, 67, 565-603.

[34] Schennach, S. M. (2000), \Estimation of Nonlinear Models with Measurement Error," mimeo, Department of Economics, MIT.

[35] Silverman, B. W. (1986), Density Estimation for Statistics and Data Analysis, London, Chapman and Hall.

[36] Staiger, D. and J.H. Stock (1997), \Instrumental Variables Regression with Weak Instruments," Econometrica, 65, 557-586.

[37] U.S. Department of Energy, Energy Information Administration, Coal Industry Annual, (various years, 1993-1995), Washington, D.C.

[38] U.S. Department of Energy, Energy Information Administration, Coal Production, (various years, to1993), Washington, D.C.

[39] U.S. Department of Energy, Energy Information Administration, Longwall Mining, (DOE/EIA-TR-0588), March 1995.

[40] U.S. Department of Interior, Bureau of Mines, Minerals Yearbook, (various years, to 1976), Washington, D.C.

[41] U.S. Department of Labor, Bureau of Labor Statistics, Employment, Hours and Earnings, United States, (1994, 1996), Washington, D.C.

[42] U.S. Department of Labor, Mine Safety and Health Administration, Part 50 Coal Mining Address/Employment and Accident/Injury Files, Division of Mining Information Systems, Denver, CO.

[43] Wang, L. and C. Hsiao (1995), \A Simulation-Based Semiparametric Estimation of Nonlinear Errors-in-Variables Models," Working Paper, University of Southern California.

[44] Wang, J. and E. Zivot (1998), \Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments," Econometrica, 66, 1389-1404.

A. Appendix A: Adjusting Polynomial Models for Measurement Error

With reference to (3.30), if the exact value of log-output q¤ were observable, an estimate of the ¯'s and the ±'s could be obtained by solving the following (normal) equations:

$$E[y(q^*)^n] = \sum_{i=1}^{r} \beta_i E\left[(q^*)^{i+n}\right] + \sum_{l=1}^{s} \delta_l E[z_l (q^*)^n] \quad \text{for } n = 1 \ldots r$$

$$E[yz_n] = \sum_{i=1}^{r} \beta_i E\left[(q^*)^{i} z_n\right] + \sum_{l=1}^{s} \delta_l E[z_l z_n] \quad \text{for } n = 1 \ldots s, \qquad \text{(A.1)}$$

where all expectations $E[\cdot]$ can be estimated from sample moments.

We can express the moments of observed log outputs in terms of the moments of true log-output as

$$\begin{aligned} E[q^n] &= E[(q^* + \varepsilon)^n] \\ &= \sum_{j=0}^{n} \binom{n}{j} E\left[(q^*)^j \varepsilon^{n-j}\right] \\ &= E[(q^*)^n] + \sum_{j=0}^{n-1} \binom{n}{j} E\left[(q^*)^j\right] E\left[\varepsilon^{n-j}\right] \end{aligned}$$

where we have used the independence of q^* and ε. We can now isolate $E[(q^*)^n]$ and obtain a recursive relation which gives us all the true moments in terms of the observed ones:

$$E[(q^*)^n] = E[q^n] - \sum_{j=0}^{n-1} \binom{n}{j} E\left[(q^*)^j\right] E\left[\varepsilon^{n-j}\right]$$

Note that the $E[\varepsilon^{n-j}]$ are assumed to be known since ε has a known distribution. Similarly, we have:

$$E[(q^*)^n z_l] = E[q^n z_l] - \sum_{j=0}^{n-1} \binom{n}{j} E\left[(q^*)^j z_l\right] E\left[\varepsilon^{n-j}\right]$$

where we have used the independence between ε and q^*, z.

We now rewrite expressions analogous to equations A.1 but involving only observed moments or the true moments we have already determined above:

$$E[yq^n] = \sum_{i=1} \beta_i E\left[(q^*)^i q^n\right] + \sum_{l=1} \delta_l E[z_l q^n]$$

$$= \sum_{i=1} \beta_i E\left[(q^*)^i (q^* + \epsilon)^n\right] + \sum_{l=1} \delta_l E[z_l q^n]$$

$$= \sum_{i=1} \beta_i \sum_{j=0} \binom{n}{j} E\left[(q^*)^{i+j}\right] E\left[\epsilon^{n-j}\right] + \sum_{l=1} \delta_l E[z_l q^n] \quad \text{for } n = 1 \ldots r$$

and

$$E[y z_n] = \sum_{i=1} \beta_i E\left[(q^*)^i z_n\right] + \sum_{l=1} \delta_l E[z_l z_n] \quad \text{for } l = 1 \ldots s.$$

The regression coefficients can thus be obtained from these modified normal equations, by isolating β and δ in:

$$\begin{pmatrix} b \\ d \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \beta \\ \delta \end{pmatrix}$$

where

$$\begin{aligned}
b &= E[y q^n] \text{ for } n = 1 \ldots r \\
d &= E[y z_l] \text{ for } l = 1 \ldots s \\
A_{ni} &= \sum_{j=0} \binom{n}{j} E\left[(q^*)^{i+j}\right] E\left[\epsilon^{n-j}\right] \text{ for } n = 1 \ldots r \text{ and } i = 1 . r \\
B_{nl} &= E[z_l q^n] \text{ for } n = 1 \ldots r \text{ and } l = 1 \ldots s \\
C_{li} &= E\left[(q^*)^i z_l\right] \text{ for } l = 1 \ldots s \text{ and } i = 1 \ldots r \\
D_{nl} &= E[z_l z_n] \text{ for } n = 1 \ldots s \text{ and } l = 1 \ldots s
\end{aligned}$$

and

$$\begin{aligned}
E\left[(q^*)^n\right] &= E[q^n] - \sum_{j=0}^{n-1} \binom{n}{j} E\left[(q^*)^j\right] E\left[\epsilon^{n-j}\right] \\
E\left[(q^*)^n z_l\right] &= E[q^n z_l] - \sum_{j=0}^{n-1} \binom{n}{j} E\left[(q^*)^j z_l\right] E\left[\epsilon^{n-j}\right]
\end{aligned}$$

Figure 1. Price, Quantity and Labor Productivity U.S. Coal Industry, 1972–95

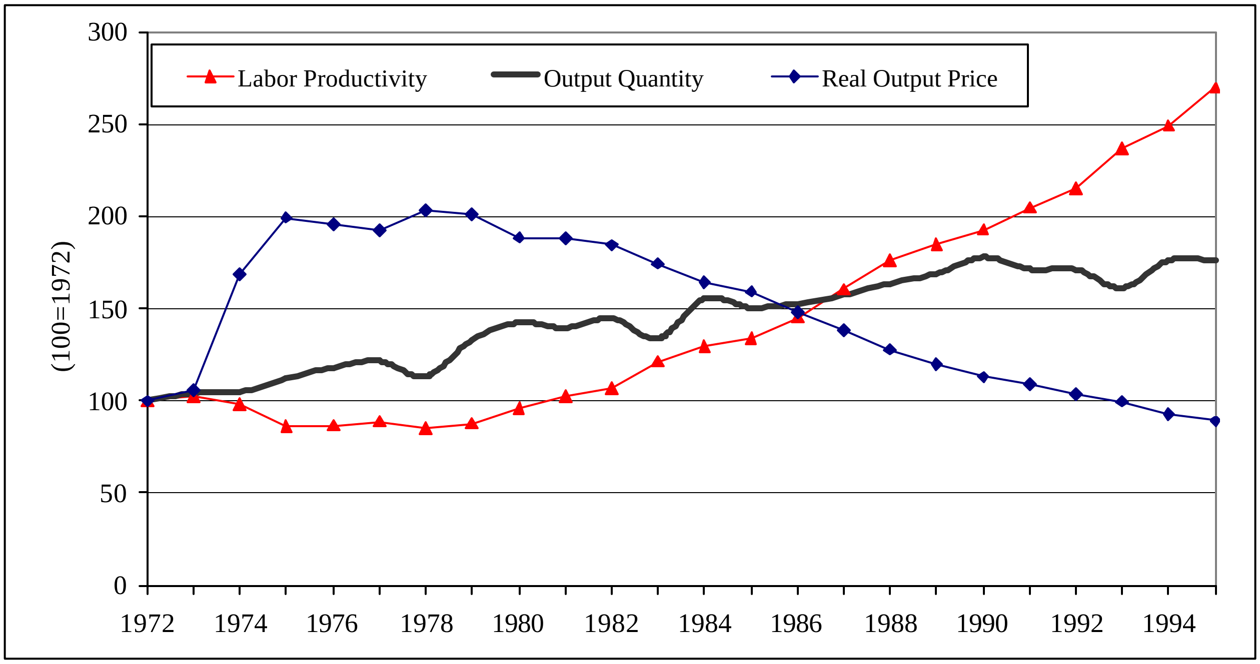

Figure 2: Coal Producing Regions

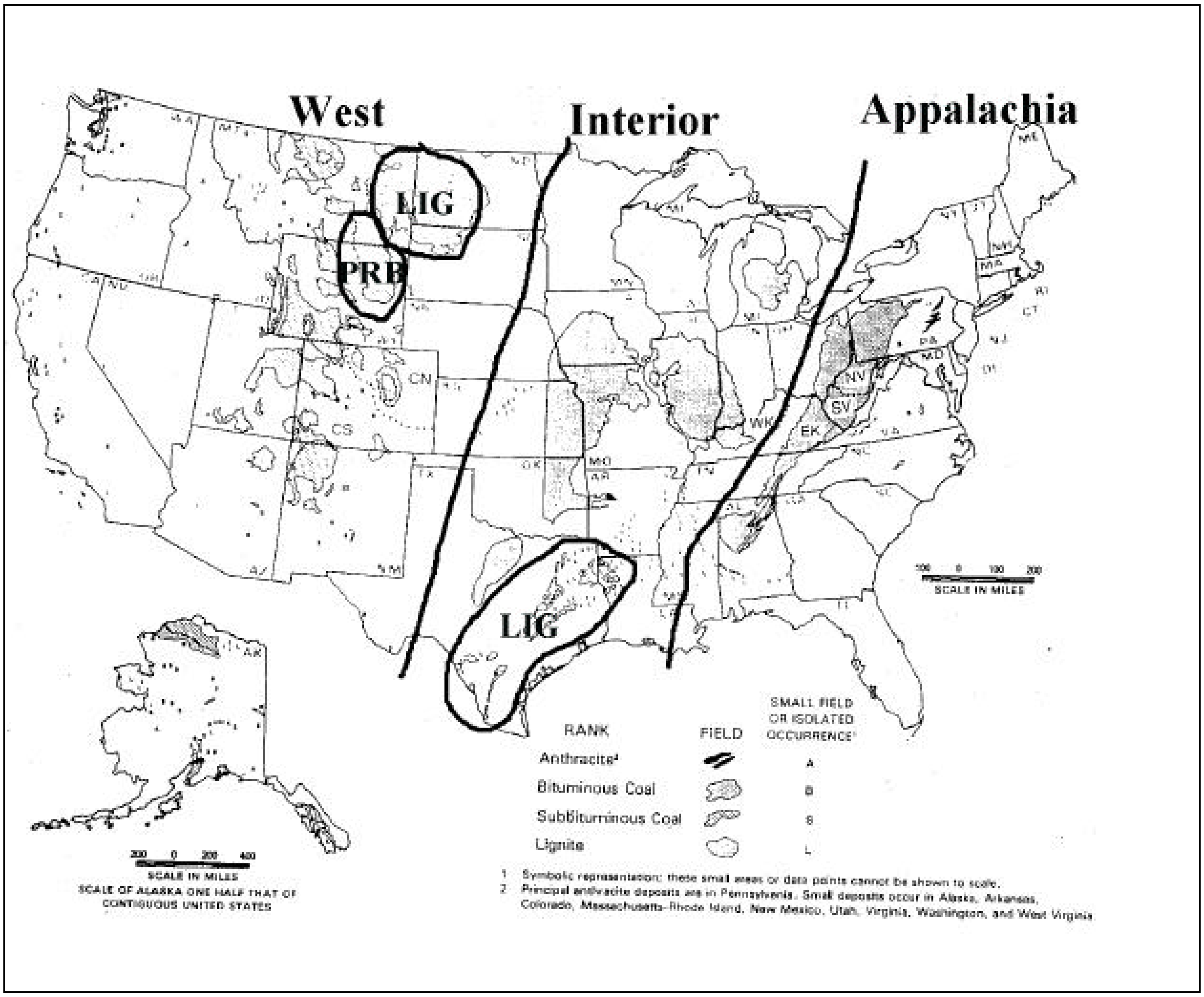

Table 1: Sample Composition and Mine Groups

Region	Technology	Abbreviation	Number of Observations	Number of Mines	Observations per Mine	Average Btu Content
Appalachia	Surface	APP-S	37161	9019	4.120	23
	Longwall	APP-LW	1216	111	10.955	23
	Continuous	APP-CM	38100	8339	4.569	23
Interior	Surface	INT-S	5219	1260	4.142	22
	Longwall	INT-LW	106	14	7.571	22
	Continuous	INT-CM	1295	173	7.486	22
Western	Surface	WST-S	789	87	9.069	20
	Longwall	WST-LW	224	29	7.724	22
	Continuous	WST-CM	902	128	7.047	22
Powder River Basin	Surface	PRB	450	28	16.071	17
Lignite	Surface	LIG	506	33	15.333	13
Total			85968	19221	4.473	

Figure 3A: Coal Production by Geographic Region

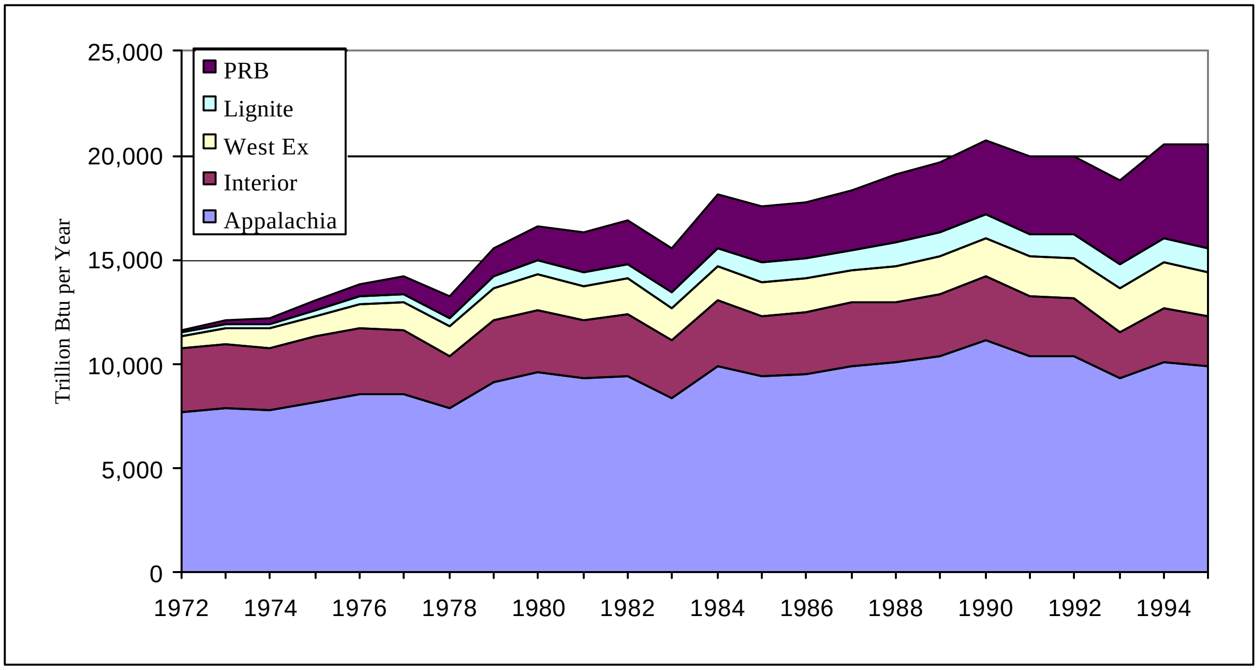

Figure 3B. Coal Production by Mining Technique

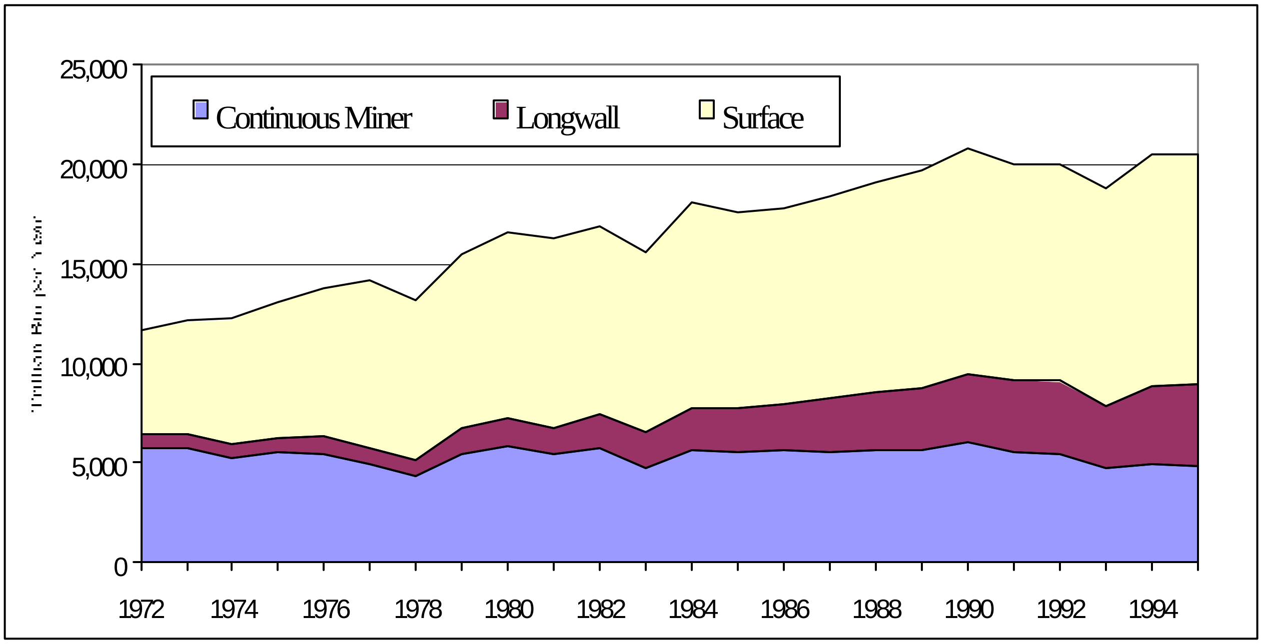

Figure 4A: Observed Labor Productivity by Group

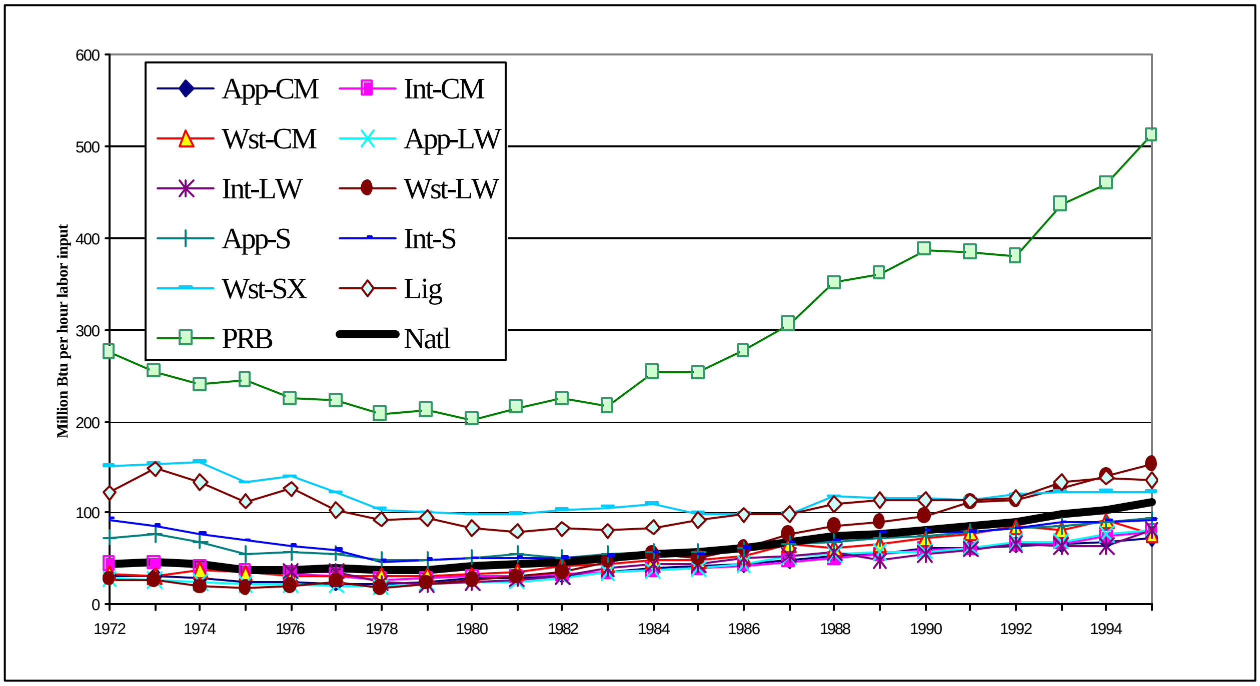

Figure 4B: Indices of Labor Productivity, 1972-95

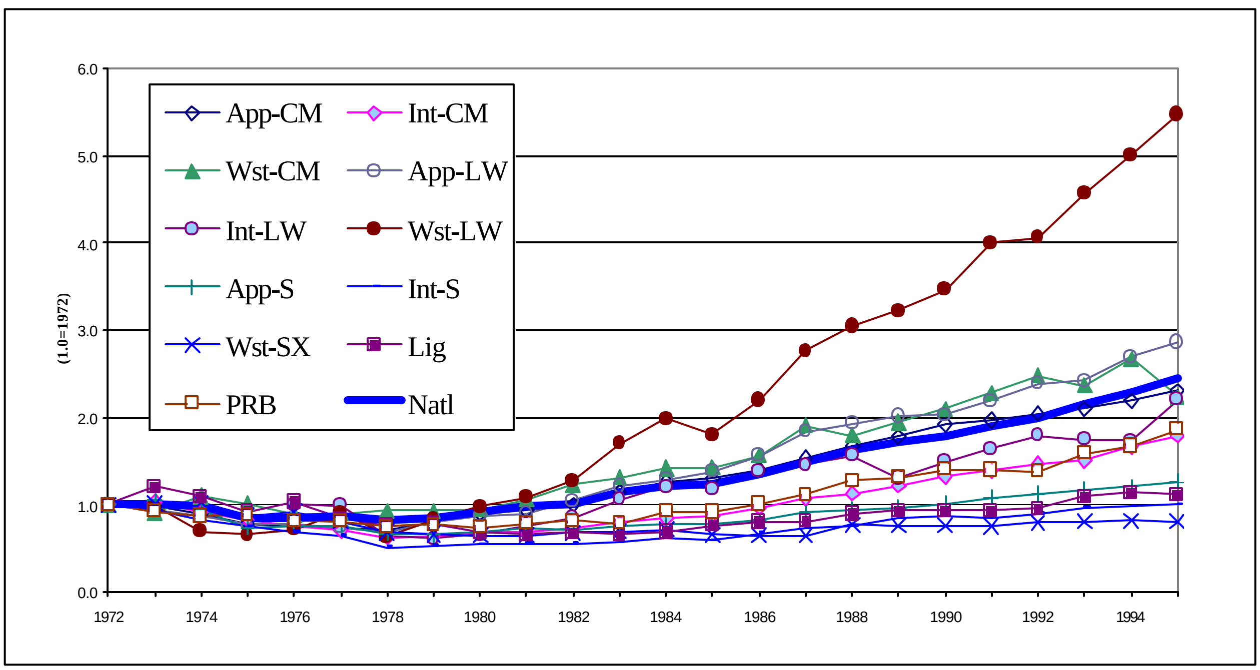

Figure 5: Average Annual Mine Output

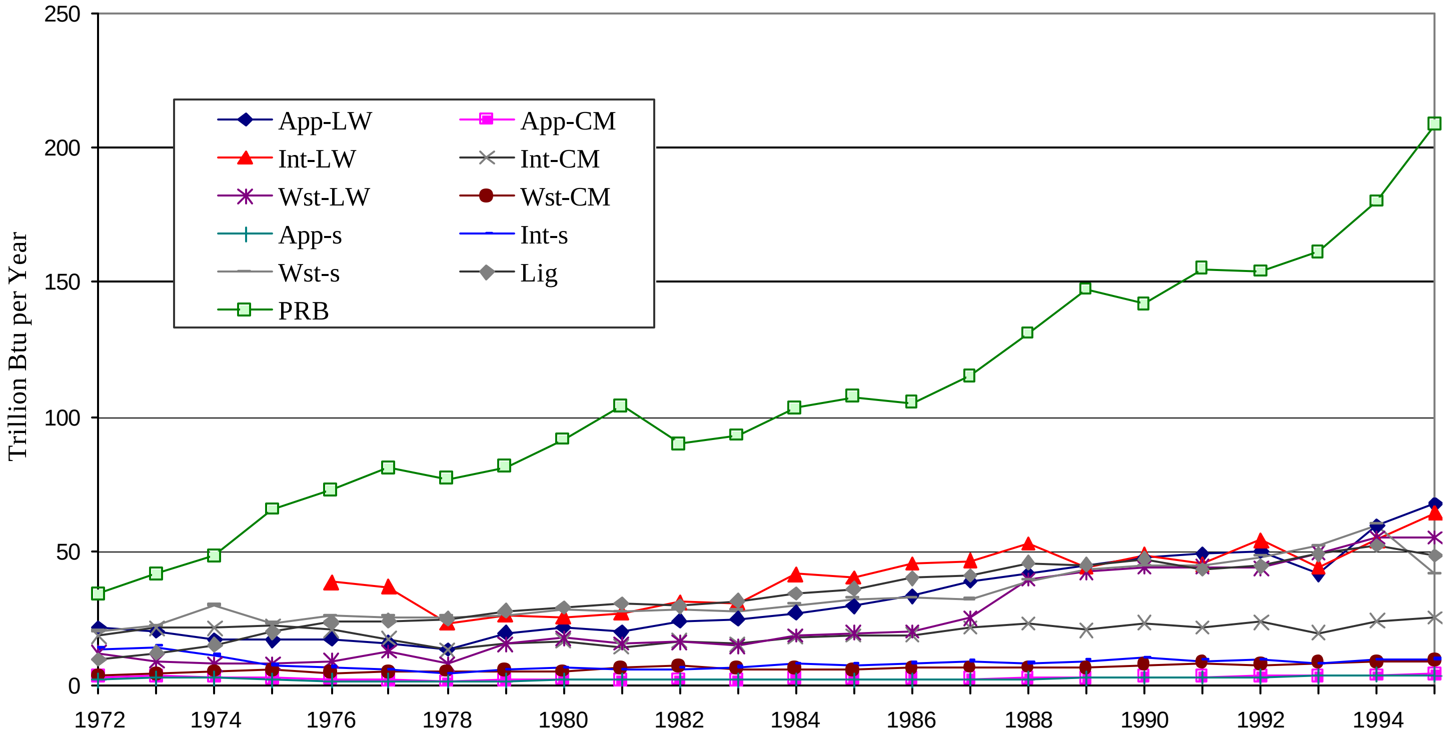

Table 2: Scale Coefficients

	OLS Coefficient of				R^2	
Mine Group	ln *Q*	$(\ln Q)^2$	$(\ln Q)^3$	Sample Size	Within	Overall
APP-S	1.686 (0.0685)	-0.128 (0.0074)	0.0037 (0.00026)	37161	0.302	0.177
APP-LW	0.471 (0.0140)			1216	0.774	0.745
APP-CM	1.784 (0.0497)	-0.158 (0.0055)	0.00519 (0.0002)	38100	0.335	0.265
INT-S	1.502 (0.1894)	-0.114 (0.01862)	0.00348 (0.0006)	5219	0.391	0.179
INT-LW	0.333 (0.0435)			106	0.923	0.865
INT-CM	5.223 (0.4622)	-0.411 (0.0411)	0.0113 (0.0012)	1295	0.634	0.439
WST-S	1.207 (0.0757)	-0.0314 (0.0034)		789	0.673	0.619
WST-LW	9.212 (3.1017)	-0.731 (0.2553)	0.0199 (0.0070)	224	0.573	0.797
WST-CM	2.192 (0.3195)	–0.165 (0.0343)	0.00467 (0.0012)	902	0.556	0.609
PRB	1.801 (0.1268)	-0.0447 (0.0048)		450	0.828	0.609
LIG	1.178 (0.0860)	-0.0222 (0.0035)		506	0.767	0.529

Notes: Polynomial order chosen by cross validation, and standard errors in parentheses.

Table 3: Time Effect Regressions

Mine Group	OLS Coefficient of ln *p*: Price Effect	ln *w*: Wage Effect	*D*: 72_73 Dummy	R^2	Test: Price Effect = - Wage Effect (p-value)
APP-S	0.174	-0.5	0.389	0.742	0.2006
	(0.0527)	(0.2778)	(0.0486)		
APP-LW	-0.858	0.202	-0.407	0.968	0.0021
	(0.0397)	(0.2020)	(0.03633)		
APP-CM	-0.399	-0.011	-0.13	0.923	0.0075
	(0.0296)	(0.1100)	(0.0271)		
INT-S	-0.252	0.105	0.259	0.538	0.7561
	(0.1575)	(0.5250)	(0.0996)		
INT-LW	-1.449	1.14	NA*	0.877	0.5341
	(0.1628)	(0.600)			
INT-CM	-0.759	0.321	-0.287	0.802	0.2098
	(-0.1116)	(0.4012)	(0.0718)		
WST-S	-0.11	-0.379	0.517	0.53	0.433
	(.02750)	(0.7580)	(0.1261)		
WST-LW	-1.432	1.474	-0.735	0.38	0.9691
	(0.4475)	(1.340)	(0.2227)		
WST-CM	-1.071	0.87	-0.079	0.554	0.7477
	(0.2550)	(0.7909)	-0.1317		
PRB	-0.399	0.002	0.263	0.688	0.4468
	(0.1023)	(.1818)	(0.0848)		
LIG	-0.617	-0.36	-0.031	0.877	0.0001
	(0.0685)	(0.2250)	(0.0517)		

Notes: 24 annual observations, except for INT-LW which has 20 observations.
Standard errors in parentheses.
* No production prior to 1976.

Figure 6: Contributions of Scale, Price, Fixed and Time Effects

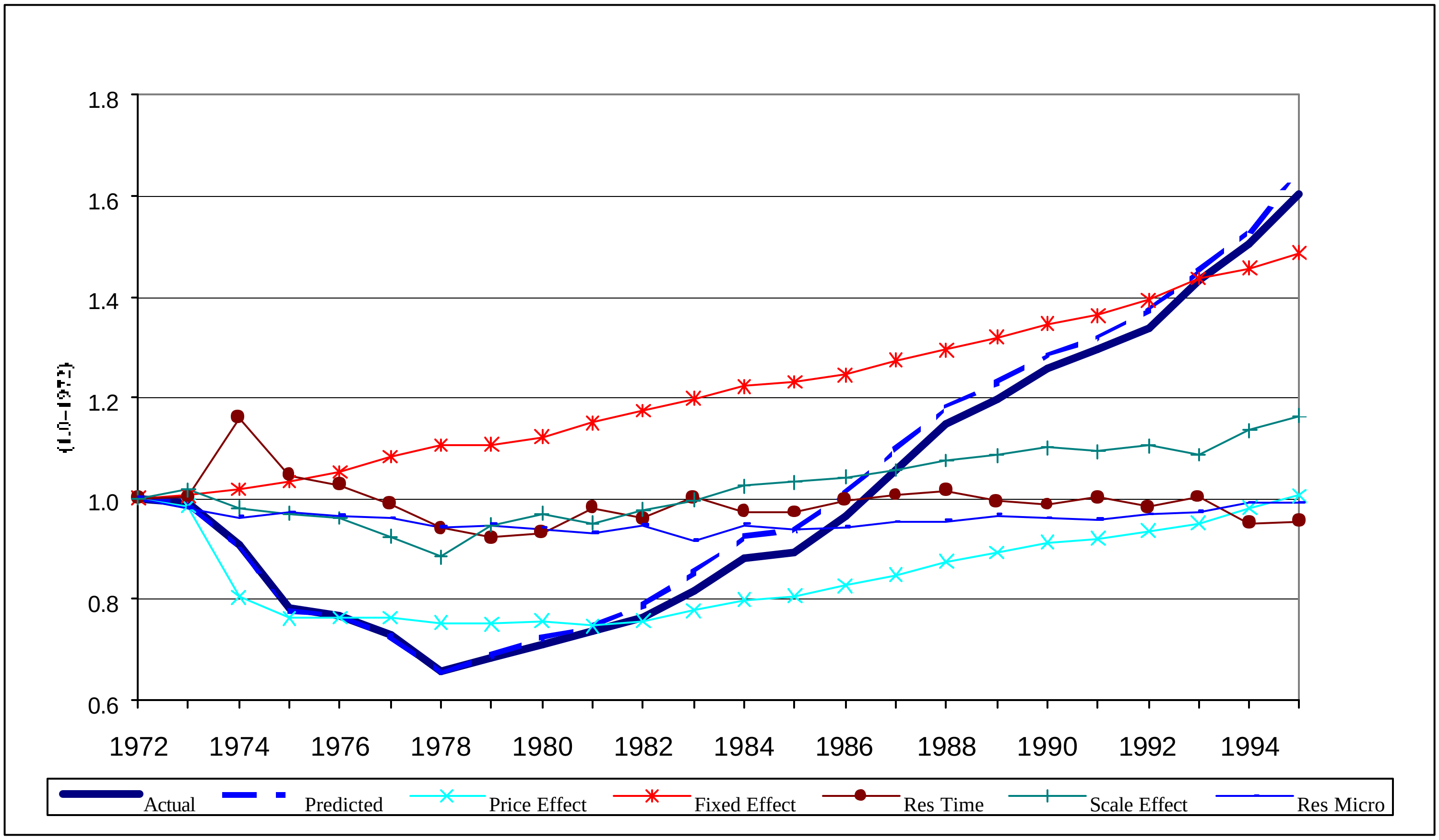

Table 4: Relation of Average Mine Life to Scale and Fixed Effects Indices

		Average Annual Growth Rate, 1972–1995		
	Average Mine Life	Scale Effects	Fixed Effects	Combined
PRB	15.00	+2.77	+0.46	+3.23
Lig	12.65	+3.43	–2.37	+1.06
AppLW	10.96	+2.21	–0.09	+2.12
WstS (Ex)	8.99	+0.77	+0.09	+0.86
WstLW	7.72	+2.97	+0.60	+3.57
IntLW	7.57	+1.00	–0.55	+0.45
IntCM	7.49	+0.46	+1.61	+2.07
WstCM	7.00	+0.46	+3.15	+3.61
AppCM	4.57	–0.30	+3.17	+2.87
IntS	4.14	–0.72	+2.39	+1.67
AppS	4.12	+0.87	+2.16	+3.02
National Total	**NA**	**+0.66**	**+1.73**	**+2.38**

Table 5: Bounds on Scale Effects

Mine Group	Adj. Lower Bound ALB	Lower Bound LB	Upper Bound UB	Adj. Upper Bound AUB
APP-S	-0.0563	-0.0479	0.2869	0.2926
APP-LW	0.0204	0.0689	0.4710	0.4985
APP-CM	0.0250	0.0308	0.2402	0.2447
INT-S	0.0316	0.0492	0.2948	0.3079
INT-LW	0.0161	0.1290	0.3331	0.4196
INT-CM	0.1092	0.1369	0.3332	0.3546
WST-S	0.1159	0.1694	0.5199	0.5508
WST-LW	-0.1214	-0.0040	0.3730	0.4463
WST-CM	0.0537	0.0951	0.3571	0.3865
PRB	0.1641	0.2400	0.6244	0.6619
LIG	0.0878	0.1767	0.6478	0.6858

Table 6: IV Estimates of Scale Effect for Long-Linear Model

	OLS First-Differences	2SLS with Instruments:					
Mine Group	D(lnQ_t)	D(lnQ_{t-2})	ln(Q_{t-2}), ln(Q_{t-3})	Powers of D(lnQ_{t-2})	Powers of ln(Q_{t-2}), ln(Q_{t-3})	Sample Size	Number of Mines
APP-S	0.363 (0.0113)	0.300 (0.0892)	0.254 (0.0703)	0.251 (0.0846)	0.253 (0.0547)	15943	2925
APP-LW	0.536 (0.0499)	0.000 (0.4733)	0.006 (0.4769)	0.444 (0.1156)	0.449 (0.1151)	861	96
APP-CM	0.293 (0.0128)	0.324 (0.0746)	0.391 (0.0578)	0.272 (0.0505)	0.297 (0.0452)	16046	3520
INT-S	0.391 (0.0250)	0.239 (0.2715)	0.205 (0.2203)	0.338 (0.2221)	0.147 (0.1308)	2328	420
INT-LW	0.408 (0.0706)	1.152 (7.1784)	-0.028 (2.1173)	0.662 (0.1205)	0.653 (0.0974)	66	9
INT-CM	0.356 (0.0396)	0.017 (0.5290)	0.226 (0.6571)	0.148 (0.1182)	0.449 (0.2230)	808	108
WST-S	0.672 (0.0554)	1.264 (0.5671)	1.258 (0.5424)	0.930 (0.1017)	0.496 (0.1862)	554	52
WST-LW	0.596 (0.0565)	0.529 (0.1646)	0.455 (0.1560)	0.407 (0.1508)	0.534 (0.1245)	131	22
WST-CM	0.399 (0.0767)	1.434 (4.1655)	0.534 (0.2732)	0.736 (0.1823)	0.692 (0.1317)	493	77
PRB	0.740 (0.1343)	0.231 (1.3236)	0.520 (0.3463)	0.176 (0.1720)	0.905 (0.1098)	361	28
LIG	0.656 (0.0439)	0.636 (0.2339)	0.623 (0.2169)	0.469 (0.1896)	0.637 (0.1144)	391	33

Notes: 2SLS estimation samples smaller than observation sample due to availability of instruments.

Table 7: F-Statistics from First Step Regressions

Mine Group	2SLS with Instruments: $D(\ln Q_{t-2})$	$\ln(Q_{t-2})$, $\ln(Q_{t-3})$	Powers of $D(\ln Q_{t-2})$	Powers of $\ln(Q_{t-2})$, $\ln(Q_{t-3})$	$\ln(Q_{t-2})$ Single Instrument	One Instrument per Year
APP-S	49.77	36.02	13.77	12.07	16.39	2.7
APP-LW	3.31	1.66	3.43	2.23	5.35	1.83
APP-CM	55.75	43.26	26.34	18.97	120.35	7.66
INT-S	4.47	3.02	1.18	1.57	0.35	2.37
INT-LW	0.06	0.06	1.28	0.71	3.15	0.71
INT-CM	1.08	0.97	2.87	1.63	2.38	2.29
WST-S	2.41	1.21	4.02	2.22	0.24	1.72
WST-LW	12.05	9.97	5.91	4.54	3.27	3
WST-CM	0.08	2.49	2.35	1.8	8.86	1.49
PRB	0.8	0.55	0.38	1.42	0.82	8.55
LIG	13.89	7.4	3.11	2.64	0	3.82

Notes: First step regressions for estimates of Table 6.

Table 8: System GMM Estimates of Scale Effect

	Basic Model (3.22)		Model with Autocorrelated Error (3.27)	
	GMM with Instruments:		GMM with Instrument	
Mine Group	$\ln(Q_{t-2})$	$\ln(Q_{t-3})$	$\ln(Q_{t-3})$	Rho
APP-S	0.3375 (0.0158)	0.3465 (0.0184)	0.2918 (0.0147)	0.3591 (0.0224)
APP-LW	0.393 (0.0939)	0.519 (0.0821)	0.874 (0.1475)	0.791 (0.0854)
APP-CM	0.411 (0.0181)	0.352 (0.0164)	0.234 (0.0134)	0.333 (0.0275)
INT-S	0.389 (0.0301)	0.200 (0.0399)	0.402 (0.0340)	0.262 (0.0689)
INT-LW	NA	NA	NA	NA
INT-CM	0.356 (0.0708)	0.069 (0.0654)	0.151 (0.0789)	0.542 (0.2347)
WST-S	0.613 (0.1959)	0.340 (0.0813)	0.369 (0.1602)	0.325 (0.2744)
WST-LW	NA	NA	0.453 (0.2239)	0.708 (0.2818)
WST-CM	0.622 (0.1435)	0.637 (0.2371)	0.479 (0.1240)	0.418 (0.1691)
PRB	1.060 (0.7467)	0.706 (0.4272)	0.406 (0.1704)	0.442 (0.0813)
LIG	0.174 (0.1062)	0.365 (0.0787)	0.799 (0.1422)	0.898 (0.1487)

Notes: Likely due to small sample size, computational difficulties occurred for INT-LW and WST-LW estimations listed as NA (failure to converge). Standard errors in parentheses.

Table 9: Various Specification Tests: GMM Estimates

Mine Group	Sargan Test	Test for Second Order Autocorrelation	
	Model (3.22)	Model (3.22)	Model (3.27)
APP-S	0.886	0	0.001
APP-LW	0.965	0.001	0.052
APP-CM	0	0	0.576
INT-S	0	0.527	0.627
INT-LW	NA	NA	NA
INT-CM	0.007	0.832	0.884
WST-S	0.114	0.305	0.875
WST-LW	NA	NA	0.353
WST-CM	0.958	0.077	0.468
PRB	0.415	0.57	0.578
LIG	0.483	0.43	0.641

Notes: p values for all tests

Figure 7: Scale Effects Adjusted for Measurement Error

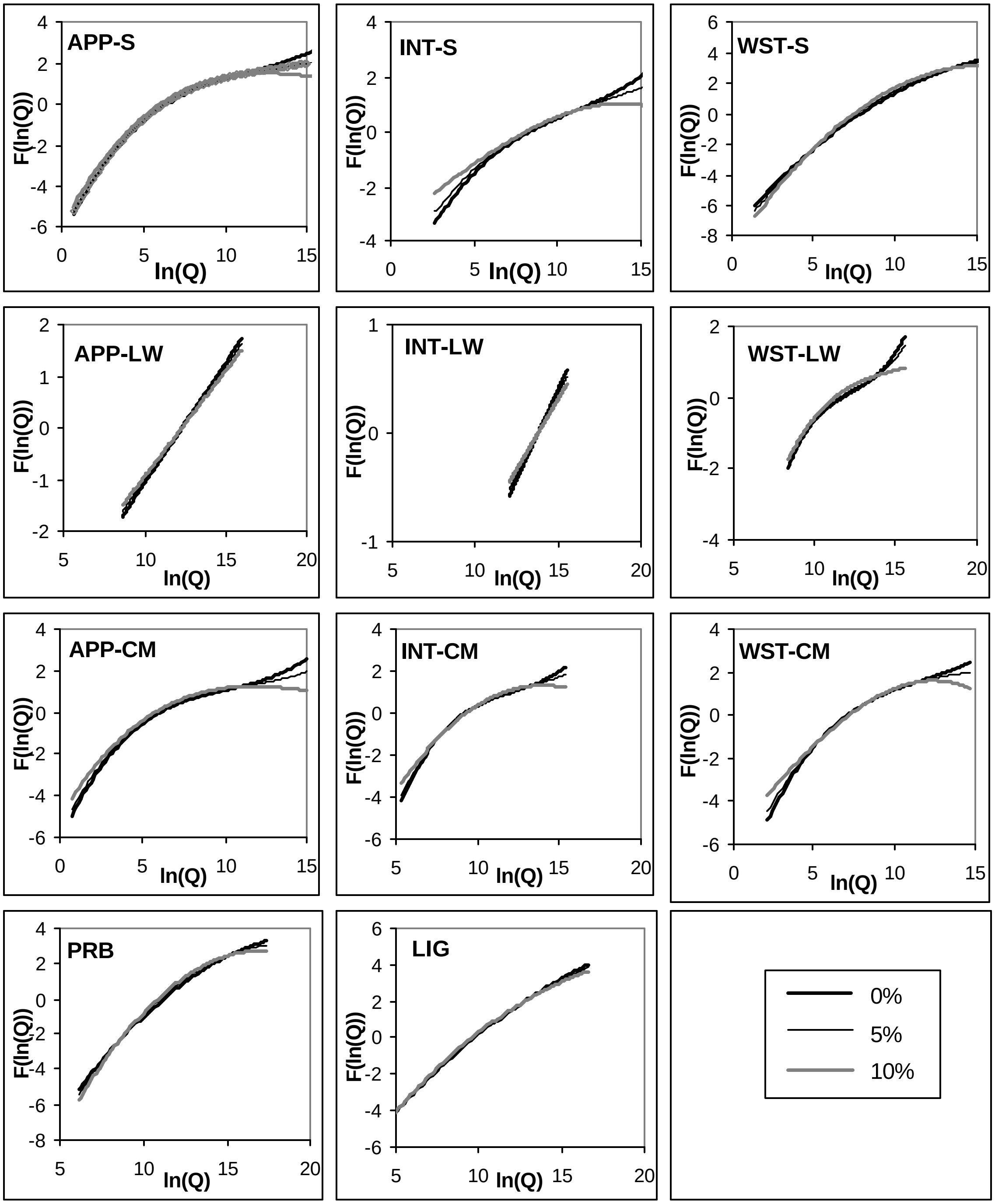

Figure 8: Productivity Decomposition for Different Estimates