Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Econometrics 127 (2005) 131–164
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Panel data analysis of U.S. coal productivity

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

MIT Sloan School of Management, E52-444, 50 Memorial Drive, Cambridge, MA 02142, USA

Accepted 30 June 2004
Available online 25 August 2004

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 in linear and nonlinear models.

JEL classification: C23; D24; L11; L72

Keywords: Labor productivity; Coal mining; Heterogeneity; Economies of scale; Diagnostics

1. Introduction

The coal-mining industry in the United States (U.S.) is a remarkably dynamic industry. Labor productivity grew steadily at an annual rate of 5.36% from 1978 to

*Corresponding author. Tel.: +1-617-253-2625; fax: +1-617-258-6855.
E-mail address: tstoker@mit.edu (T.M. Stoker).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2004.06.006

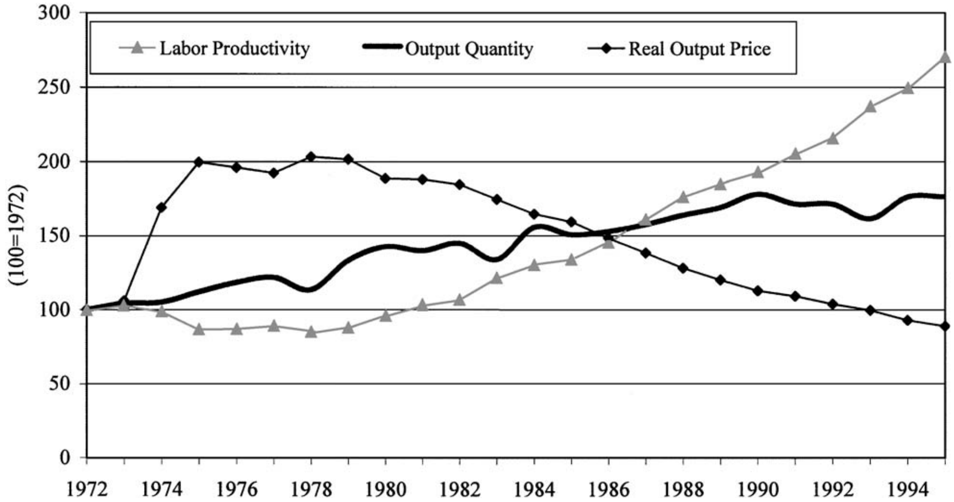


Fig. 1. Price, quantity and labor productivity U.S. coal industry, 1972–95.

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

The technology for mining coal varies greatly across the U.S., which gives rise to many possibilities 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 categorized by mining process; the traditional continuous process or the more recent longwall process. Moreover, 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. But with extensive heterogeneity, it is not clear how to interpret aggregate productivity growth without

[1]This is a conservative estimate based on averaging productivity from the eleven coal-mining groups defined below. If instead one simply uses data on total tons and total labor hours in the U.S., the productivity growth rate from 1978 to 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, sulfur content would be an important differentiating feature, which we do not distinguish here.

understanding its sources, and it is not clear how to measure those sources. Here we present an econometric analysis of U.S. coal mining, and we define 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 U.S. from 1972 to 1995. In addition to mine location, we identify the type of production technology used in each mine; namely surface mining, underground continuous mining (CM) and underground longwall mining (LW). 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 employ methods that allow for substantial heterogeneity across mines.

We model labor productivity separately for groups of mines defined by geographic region and type of technology. Mine-specific fixed effects are included to 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. Our results give an intriguing depiction of smooth, uniform technological progress in coal mining over the period 1972–1995, as well as an assessment of the importance of scale economies and technological improvements embodied in physical capital.

Our modeling rests on an important specification assumption, that mine output is predetermined relative to labor, due to contracting practice in coal mining. We present extensive diagnostic analysis to assess how sensitive our results are to that assumption. We adapt traditional (linear) errors-in-variables bounds to our productivity analysis. We examine (linear) instrumental variable estimates of productivity effects, applying recent methods from the "weak instruments" literature, and follow with the generalized method of moments estimates that account for many kinds of model structure. Further, we examine the interplay of errors-in-variables with the nonlinear structure of our model by computing polynomial estimates adjusted for the presence of such errors. Through all of these methods, we do not find sufficient evidence to contradict our basic empirical findings. Our applications illustrate nicely the usefulness of these recent diagnostic techniques, which may be helpful in validating other applied research.

Section 2 describes our data, spells out our modeling assumptions and provides our overall results. Section 3 presents our diagnostic work on the sensitivity of our results to key assumptions. Section 4 offers some concluding remarks.

[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 et al. (1987), and Jorgenson (1990). For studies of data from states and individual mines, see Kruvant et al. (1982), Baker (1981), Byrnes et al. (1984) and Boyd (1987). Boyd (1987) gives a detailed analysis of Illinois strip mines and documents substantial economies of scale.

2. Panel data analysis of productivity

2.1. U.S. coal-mining data

2.1.1. Data specifics and mining groups

The data on coal mine output and labor input are collected by the Mine Safety and Health Administration (MSHA) as part of its mandated regulatory effort since 1972.[4] Coal output is measured in clean short tons, and for aggregating output across regions, coal output is (quality) adjusted for heat content.[5] Labor is measured in hours, and we do not distinguish different types of labor.

We observe mine location and whether the mine is a surface mine or an underground mine. Surface mining involves a substantially different 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 CM or LW methods, depending on the nature of the coal deposit. Continuous mines use (grinding) 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. LW 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). CM is the traditional technique, which is suitable for many types of mining sites. LW is a more recent technique, that had been introduced in Europe and was then adopted in the U.S. over the time frame of our data.[6]

The basic MSHA data do not identify which underground mines are longwall mines, and so we identified longwall mines by matching specific mine locations with longwall installations reported in *Coal Age* magazine. The MSHA data contain a few details of mine facilities (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 is seam height, but that data appeared to be of very poor quality and were not used in the analysis. We constructed an annual coal price index for each region, and used a national wage series to proxy labor cost changes.[7]

We segment the data into 11 groups of mines, and analyze each group separately. We categorize mines based on their location in three geographical regions—Appalachia (APP), Interior (INT), and West (WST). We categorize mines based on

[4]Ellerman et al. (2001) gives much more detail on the specifics of the coal industry and the data.

[5]These adjustments of each coal type for Btu (heat) content are listed in Table 1. It is important to note that these adjustments do not affect our models or estimates, but are only applied for aggregating across mining groups.

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

[7]The price data are constructed from annual mine-mouth coal prices by state as collected by the Energy Information Adminstration of the U.S. Department of Energy. Wage data are from the *Employment, Hours and Earnings* series published by the Bureau of Labor Statistics. These data are deflated to real prices and wages using the consumer price index.

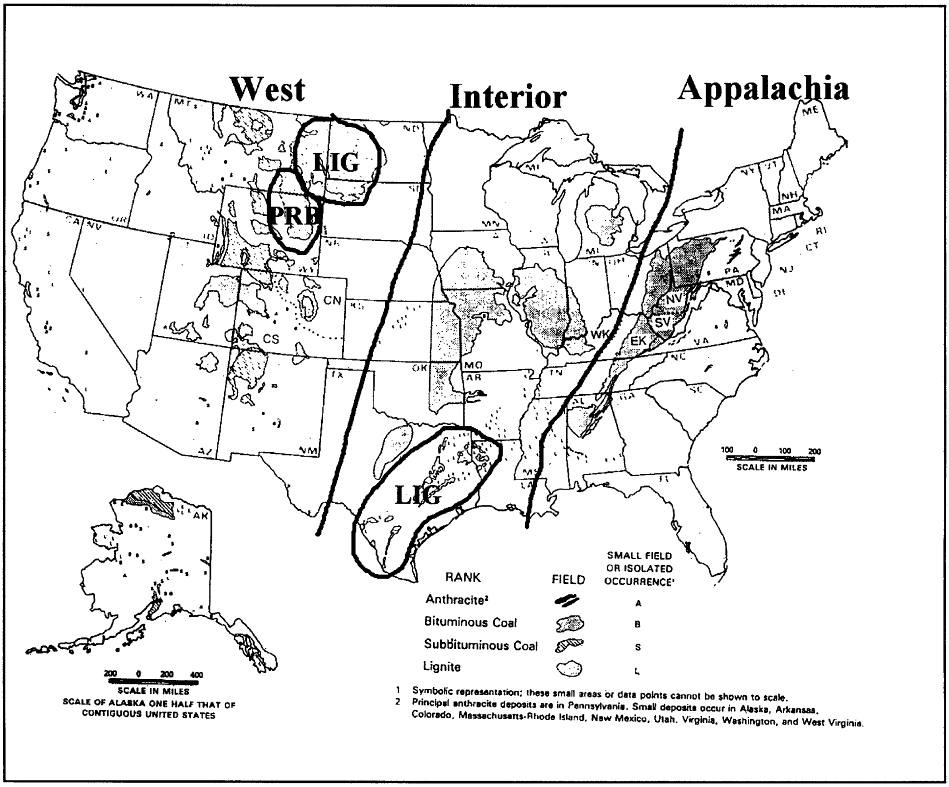


Fig. 2. Coal-producing regions.

type of production technology—surface mining (S), underground CM and LW. These two dimensions determine 9 groups (APP-S, APP-CM, ..., etc.). We further separate out two special surface-mining groups, the Powder River basin (PRB) and lignite coal (LIG), which make for 11 groups in total.[8] Fig. 2 shows a map of the U.S. with the three major regions, the PRB and the lignite-producing areas. All in all, there are 85,968 total annual observations on 19,221 individual mines. Table 1 provides the composition of the sample in terms of the 11 groups. All estimation is performed within each group.

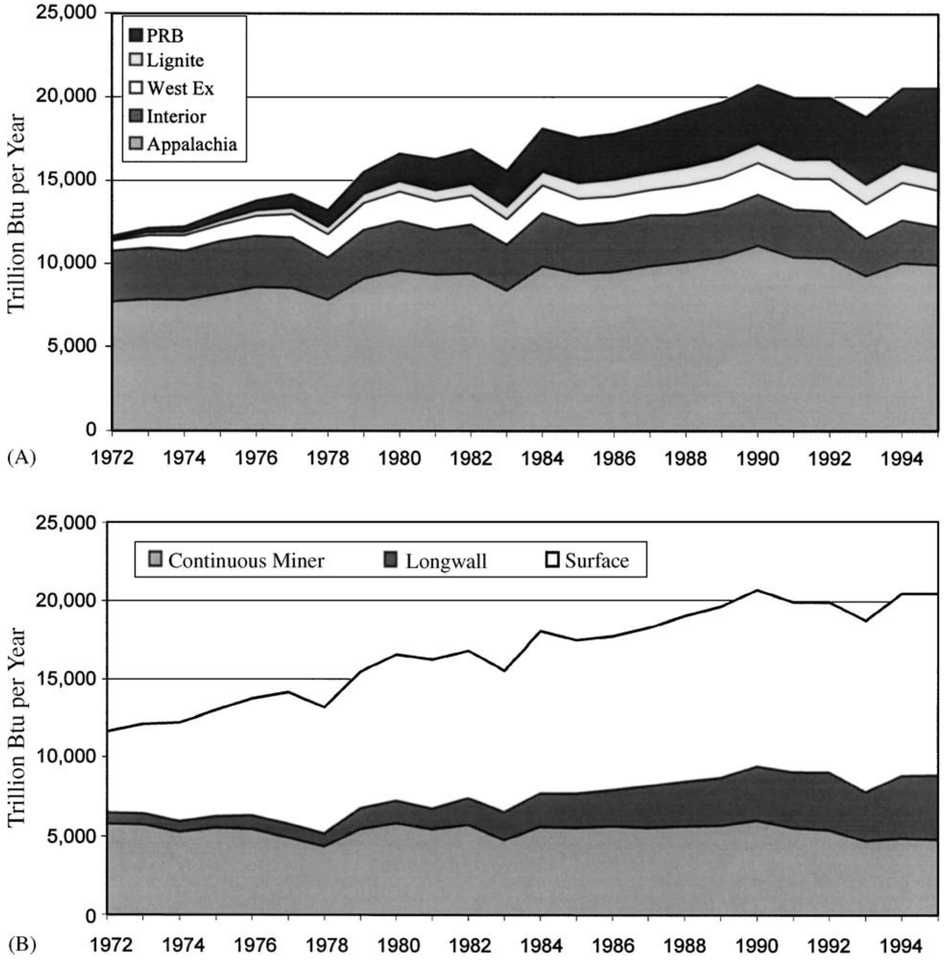
2.1.2. Recent trends in U.S. coal mining

U.S. coal mining changed dramatically from 1972 to 1995. Fig. 3 shows the composition of overall output growth. Output has increased in all mining regions

[8]The PRB had very rapid growth. Lignite is a very inferior coal in terms of heat content. Also, type of technology remains the same for all mines in the data. It is true that some continuous mines become longwall mines, and we include them as two different mines (the continuous mine closes down and then the longwall mine opens) in the observed data.

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	37,161	9019	4.120	23
	Longwall	APP-LW	1216	111	10.955	23
	Continuous	APP-CM	38,100	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			85,968	19,221	4.473	



(A) (B)

Fig. 3. (A) Coal production by geographic region. (B) Coal production by mining technique.

except for the interior, and there was truly spectacular growth in the PRB (Fig. 3A). There is strong growth in output from longwall mines and in surface mines, whereas continuous mines have shown little growth (Fig. 3B).

Table 2 displays labor productivity and scale in U.S. coal mining. By far, the most productive mines are in the PRB, but the largest increases in productivity occur for underground mines with longwall technology.[9] The overall increase in labor productivity reflects the shift in overall output to the PRB and surface mines, and the increase in productivity of underground mines. Finally, all groups show increases in

[9]It is worth nothing that in the aggregate, longwall mines are much larger but not higher in productivity than either continuous or surface mines. We isolate the impact of mine size as the “scale effect” in our models.

Table 2
Average productivity and scale

Mining technology	Average mine productivity (mmBtu/h)			Average annual mine output (Trillion Btu/year)		
	1972	1995	Growth rate (%)	1972	1995	Growth rate (%)
PRB	275.30	512.61	2.70	33.91	208.45	7.90
Surface (exc. PRB)	85.27	103.40	0.84	3.66	6.55	2.52
Continuous	33.52	74.27	3.46	3.89	5.22	1.28
Longwall	28.62	90.42	5.00	21.09	64.16	4.84

per-mine output, with the PRB having the largest mines and the largest (per-mine) growth.

In addition to these broad trends, each mining group is changing over time, as new mines open and older mines close, and operating mines change in scale. Empirical modeling of individual mines is required to understand the productivity process. We now turn to our model 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—we lack data on wages, output prices and transportation costs at specific locations—but we also believe it would seriously misrepresent institutional features of the coal market. In particular, the majority of coal output is set in advance by contracts with specific buyers.[10] 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.[11] This is a key assumption of our approach, and its failure can lead to biases in

[10]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 reflects the weather and overall level of economic activity, which are exogenous to individual coal mines.

[11]This assumption neglects potential endogeneity due to the choice of whether to open a new mine, or shut down an existing mine. We discuss below how fixed effects can accommodate this, but we do not model the entry–exit process explicitly. We also do not address issues that could arise with common ownership of mines, such as outputs that are adjusted endogenously with inputs across several mines. It is useful to note that electric utilities do not own mines as a rule; even when a mine's output is dedicated to a power plant, it is typically owned by separate parties with the relationship covered by long-term contracts. Mine-mouth powerplants may be an exception; they exist but represent at most 5–15% of total coal produced.

estimation. In Section 3, we examine the sensitivity of our main findings to this assumption.[12]

We focus on labor productivity in coal mining. Let Q_{it} and L_{it} denote the 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

$$\ln\left(\frac{Q_{it}}{L_{it}}\right) = \tau_t + \alpha_i + F(\ln Q_{it}) + \varepsilon_{it}, \tag{1}$$

where $i = 1, \ldots, N$ indexes mines, and $t = T_i^{\text{Open}}, \ldots, T_i^{\text{Close}}$ denotes the years that mine i is in operation.[13] The time effect τ_t and the mine effect α_i are treated as fixed effects 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 separate for each mine group (so that N, τ_t, and $F(\cdot)$ vary by mine group). We assume that ε_{it} has mean zero and (possibly heteroskedastic) variance σ_{it}^2 conditional on τ_t,α_i,ln Q_{it}.

For the within panel estimator to be consistent for large N and fixed width T, mine output Q_{it}would have to be assumed *strictly exogenous*, or independent from $\varepsilon_{it'}$, for all t'. Here, we consider mine output Q_{it} to only be predetermined, with strict exogeneity a rather stronger restriction. This means that the "within" coefficient estimates may contain a bias of order $1/T$ (since estimation subtracts out mine-specific averages, which have a correlation of order $1/T$ with $\varepsilon_{it'}$ for all t'). For large T, this bias is minor. Since for many mine groups our panel dataset is relatively wide (see Table 1), there may be some basis for "large T" properties. We note that our diagnostic work in Section 3 can account for these potential biases.

The time effects τ_t capture group-wide variation in mine productivity at time t.[14] Such variation could arise from the safety regulations that were applied to the U.S. coal industry in the early 1970s (regulations that applied differently to underground and surface mines), as well as common variation in coal and input prices that affect mining practice.

The mine-specific fixed effects α_i account for geological factors (or ease of mining at site i) and specific features of capital at site i. Given the type of technology (surface, underground continuous or longwall), a new mine will typically make use of the best available capital—machinery, equipment, delivery system for transporting coal to outside the mine, etc.— so that capital embodies the current state of technology. While some technology can evolve over a mine's life (and arguably would be proxied by scale Q_{it}), the fixed effects α_i will reflect "new mine" embodied technology together with specific geological factors.

[12]Various issues, such as departures from strict exogeneity (discussed below), can be viewed as measurement error problems. In Section 3 we discuss coefficient bounds as well as consistent estimation using instrumental variables.

[13]Positive output begins in (opening) year $T_i^{\text{Open}}$and continues through (closing) year T_i^{Close}. For identification, we set the time effect to zero for the first year $t = 1972$.

[14]It is important to stress that t is the current year, not the time elapsed since the opening of a mine, and that τ_t is a common "macro" productivity effect in year t.

The fixed effect modeling can partially address issues of turnover in mining. Since we only observe mines in operation, our productivity equation could include a term for selection based on mine profitability. We cannot model such a term directly, since we have no information on specific depletion profiles (site geology), or on accounting features of the investment environment that would lead to mine closure. However, to the extent that the probability of continuing operation is determined by mine-specific factors or time-specific factors, a selection impact will be captured by the fixed effects α_i and τ_t.[15]

2.3. Estimation details

Estimation of the parameters of the panel model (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

$$SS(d) = \sum_{i=1}^{N} \sum_{t=T_i^{\text{Open}}}^{T_i^{\text{Close}}} \left[\tilde{\tau}_t^{(-it)} + \tilde{\alpha}_i^{(-it)} \tilde{F}_d^{(-it)}(\ln Q_{it}) - \ln\left(\frac{Q_{it}}{L_{it}}\right)\right]^2,$$

where $\tilde{*}^{(-it)}$ refers to the least-squares estimator computed by omitting the itth observation (mine i at time t), and F_d is a polynomial of degree d.[16] 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 3. While it is clear that all coefficients are estimated precisely,[17] 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 Fig. 5 of the diagnostic section. It is

[15] The relevance of fixed effects is clear; mines with a high fixed effect could withstand greater negative shocks, and would tend to exit later. Also, we thank a reviewer for pointing out how turnover could lead to underestimation of scale effects. If firms operating at higher scale tend to have higher productivity, then those firms will tend to exit later, since a more negative productivity shock can be absorbed. There would then be a negative correlation between the errors of surviving firms and their scale, leading to an underestimate of the scale effects.

[16] For clarity, the itth term of $SS(d)$ is constructed as follows: estimate the model with all data except for the itth observation, and compute the error of predicting the itth observation with those estimates. Adding up the squared prediction errors across all data points gives $SS(d)$. 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, pp. 3–35), and considered polynomials up to order five.

[17] Standard errors are heteroskedasticity consistent estimates.

Table 3
Scale coefficients

Mine group	OLS coefficient of			Sample size	R^2 within	R^2 overall
	ln Q	(ln Q)^2	(ln Q)^3			
APP-S	1.686 (0.0685)	−0.128 (0.0074)	0.0037 (0.00026)	37,161	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)	38,100	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

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

worthwhile mentioning here that all estimates are consistent with substantial economies of scale,[18] and that cubic estimates have the same S shape for different mine groups, implying that an intermediate range of scales is associated with greatest productivity improvement.

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

For aggregate productivity analysis, we also do an OLS decomposition of the estimated time effects, to examine how they relate to prices. That is, we fit

$$\begin{aligned}\hat{\tau}_t &= \text{" Price Effects" } + \text{" Other Effects" },\\ &= \hat{\kappa} + \hat{\gamma}_p \ln p_t + \hat{\gamma}_w \ln w_t + \hat{\delta} D_t + \hat{\eta}_t,\end{aligned} \tag{2}$$

where $\hat{\tau}_t$ is the estimated time effect, p_t is the real coal price, w_t is the real wage rate, D_t is a dummy variable for 1972–1973,[19] $\hat{\kappa}$,$\hat{\gamma}_p$, $\hat{\gamma}_w$, $\hat{\delta}$ are the OLS estimates and $\hat{\eta}_t$ is the OLS residual. The estimated coefficients are given in Table 4.[20] We do not have a specific model of how prices and wages affect labor productivity, but our results will give a summary of price "effects"constructed in this way. It is worth noting that price coefficients are negative (as expected) for all but one mine group, and that wage effects are typically positive. This is consistent with the notion that high real coal prices will allow less efficient mines to be in operation (be profitable), as will low real wages.

2.4. Panel model decomposition of productivity change

The estimates of the model (1) give a full empirical description of productivity in the U.S. coal industry, but the estimates themselves are not very helpful in understanding the predominant influences on coal productivity. We implement an approach that defines indices that are conceptually aligned with the panel model structure, to obtain a clear depiction of the sources of productivity growth.

For each mining group, overall labor productivity is expressed as

$$\begin{aligned}\frac{\sum_i Q_{it}}{\sum_i L_{it}} &= \frac{\sum_i L_{it} \exp\left[\ln \frac{Q_{it}}{L_{it}}\right]}{\sum_i L_{it}} = \frac{\sum_i L_{it} \exp\left[\hat{\tau}_t + \hat{\alpha}_i + \hat{F}(\ln Q_{it}) + \hat{\varepsilon}_{it}\right]}{\sum_i L_{it}}\\ &= \frac{\sum_i L_{it} \exp\left[\hat{\alpha}_i + \hat{F}(\ln Q_{it}) + \hat{\varepsilon}_{it}\right]}{\sum_i L_{it}} \exp(\hat{\tau}_t),\end{aligned} \tag{3}$$

where "^"s denote the panel data estimates. Overall labor productivity decomposes into two factors: one for mine-specific productivity factors and the other for common time-varying trends.

The first factor of (3) reflects elements that vary across mines; namely geology and embodied capital technology, efficiencies associated with scale, and all other features of productivity that vary across 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

[19] We found D_t to be empirically necessary, and interpret it as change related to the four-fold increase of oil prices in late 1973.

[20] The first-stage sampling errors were neglected, as they are of order $1/N$, while the second-stage errors are of order $1/T$, with $T \ll N$. Also note that our two-step estimation method is inefficient, but has the advantage that consistency of the first step is not affected by potential misspecification of Eq. (2).

Table 4
Time effect regressions

Mine group	OLS coefficient of			R^2
	$\ln p$: price effect	$\ln\ w$: wage effect	D: 72_73 dummy	
APP-S	0.174 (0.0527)	−0.5 (0.2778)	0.389 (0.0486)	0.742
APP-LW	−0.858 (0.0397)	0.202 (0.2020)	−0.407 (0.03633)	0.968
APP-CM	−0.399 (0.0296)	−0.011 (0.1100)	−0.13 (0.0271)	0.923
INT-S	−0.252 (0.1575)	0.105 (0.5250)	0.259 (0.0996)	0.538
INT-LW	−1.449 (0.1628)	1.14 (0.600)	NA[a]	0.877
INT-CM	−0.759 (−0.1116)	0.321 (0.4012)	−0.287 (0.0718)	0.802
WST-S	−0.11 (.02750)	−0.379 (0.7580)	0.517 (0.1261)	0.53
WST-LW	−1.432 (0.4475)	1.474 (1.340)	−0.735 (0.2227)	0.38
WST-CM	−1.071 (0.2550)	0.87 (0.7909)	−0.079 (−0.1317)	0.554
PRB	−0.399 (0.1023)	0.002 (0.1818)	0.263 (0.0848)	0.688
LIG	−0.617 (0.0685)	−0.36 (0.2250)	−0.031 (0.0517)	0.877

Twenty-four annual observations, except for INT-LW which has 20 observations. Standard errors in parentheses.
[a]No production prior to 1976.

distributed across mines (weighted by labor hours). In particular, we consider

$$\frac{\sum_i L_{it} \exp\left[\hat{\alpha}_i + \hat{F}(\ln Q_{it}) + \hat{\varepsilon}_{it}\right]}{\sum_i L_{it}} \cong FE_t\ SC_t\ MR_t, \tag{4}$$

where

$$FE_t = \frac{\sum_i L_{it} \exp[\hat{\alpha}_i]}{\sum_i L_{it}} \tag{5}$$

defines the *Fixed Effect Index*,

$$SC_t = \frac{\sum_i L_{it} \exp\left[\hat{F}(\ln Q_{it})\right]}{\sum_i L_{it}} \tag{6}$$

defines the *Scale Effect Index* and

$$MR_t = \frac{\sum_i L_{it} \exp[\hat{\varepsilon}_{it}]}{\sum_i L_{it}} \tag{7}$$

defines the *Residual Microheterogeneity Index*.[21]

As the fixed effects α_i represent the base levels of productivity for each mine, the index FE_t reflects how those base levels vary over time. If coal-mining technology were stable over time, and more productive sites were mined first, then FE_t would decline. Alternatively, if (embodied) technology of new mining capital improved over time but site selection were unrelated with mine productivity (say dictated by population migration), then FE_t would increase. If FE_t is stable over time, that would indicate a balance between selection of less productive sites and more productive technology. In short, FE_t summarizes how geological and initial (embodied) technology levels vary over time.

The index SC_t indicates productivity improvements associated specifically with increases in scale. It is natural to think of scale effects as a combination of technology and mine-specific learning effects. It can take time to learn the most effective way of mining a given site, including optimizing the delivery system for conveying coal out of the mine, and such processes can differ for a young mine versus a more mature mine.[22]

The index MR_t summarizes the role of the residual in the log-productivity regression. We include it primarily as a check on whether the overall impacts of fixed effects and scale effects are large relative to the residual.[23]

The second factor of (3) represents the time effect relevant for comparing to the above indices; we can define the Time Effect Index directly as

$$TE_t = \exp(\hat{\tau}_t). \tag{8}$$

[21]It is clear that the LHS of (4) is the mean of the product of $\exp(\hat{\alpha}_i)$, $\exp(\hat{F}(\ln Q_{it}))$ and $\exp(\hat{\varepsilon}_{it})$, and that the RHS is the product of their separate means, each using (probability) weights $L_{it}/\sum_i L_{it}$. Correlation between the terms has a predictable effect: for instance, positive correlation between $\hat{\alpha}_i$ and $\hat{F}(\ln Q_{it})$ will cause the RHS to be smaller than the LHS.

[22]It is possible, although we believe unlikely (given our estimates), for the scale index to capture the adverse productivity effects of depletion of coal. This is because as coal is depleted at a given site, it is typical for smaller contractors to take over the mining. The MSHA data record this as the closing of the original mine and the opening of a new mine associated with the smaller contractor, so that the depletion effects are not retained in a given mine's data.

[23]MR_t typically will reflect changes in the variance of ε_{it} over time. For instance, if the (labor-weighted) distribution of ε_{it} were normal with mean 0 and variance σ_t^2 at time t, then up to sampling error, $MR_t \cong \exp[\sigma_t^2/2]$.

From the OLS decomposition (2), we express this index as

$$TE_t = \exp(\hat{\kappa} + \hat{\gamma}_p \ln p_t + \hat{\gamma}_w \ln w_t + \hat{\delta} D_t)\exp(\hat{\eta}_t) \\ = P_t \cdot R_t, \tag{9}$$

where

$$P_t = \exp(\hat{\kappa} + \hat{\gamma}_p \ln\, p_t + \hat{\gamma}_w \ln w_t + \hat{\delta} D_t) \tag{10}$$

defines the Price Effect Index and

$$R_t = \exp(\hat{\eta}_t) \tag{11}$$

defines the Residual Time Index. These indices permit the relative size of the time effects versus fixed and scale effects to be judged.[24]

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

$$PP_t = FE_t\ SC_t\ MR_t\ TE_t = FE_t\ SC_t\ MR_t\ P_t\ R_t \tag{12}$$

The difference between observed labor productivity and the predicted index is the approximation error in (4).

2.5. Sources of labor productivity changes in U.S. coal mining

Fig. 4 gives the productivity indices for coal mining.[25] All indices are normalized to 1 in 1972. One initial conclusion is that the approximation error in (4) seems of little concern; the predicted productivity index (dashed line) has nearly the same time pattern as the observed labor productivity (solid line).

The most interesting time pattern in Fig. 4 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 and the level of technology of new mines, and since it is unlikely that inferior 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.[26]

The scale index SC_t drops slowly through the early years (possibly because of decreased output associated with new environmental regulations), and then begins a steady increase over the period 1978–1995. The price index P_t shows substantial

[24] By our two-step procedure, sampling error in the coefficients of prices and wages affect only P_t and R_t but not TE_t.

[25] 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.).

[26] We cannot rule out the possibility that some inferior sites were mined before superior ones, since there is anecdotal evidence that struggling steel conglomerates unloaded valuable lands, some with coal deposits, in the 1980s and 1990s. The smooth trend in fixed effects over time suggests that this is not a significant issue.

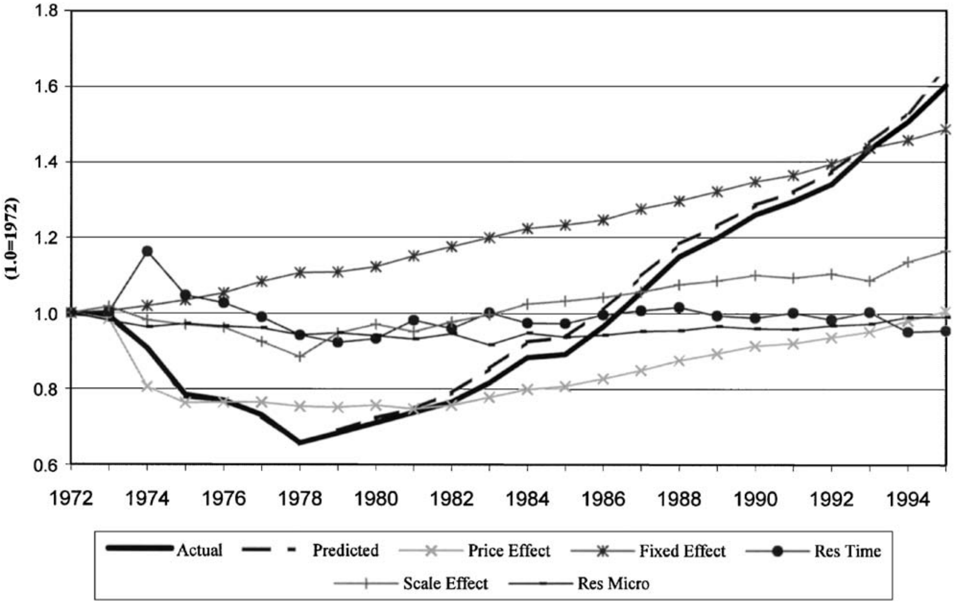


Fig. 4. Contributions of scale, price, fixed and time effects.

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 initially decreases gradually, and then slowly increases, eventually returning to its initial level. R_t varies substantively over the first 3 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.

We do not present productivity indices for each mining group separately; see Ellerman et al. (2001) for detailed analyses. Such group-specific indices can aid insight into the process of technological advance; for instance, Table 5 relates the growth of productivity to average mine life. Productivity improvements show up as scale effects for groups with long mine lives, and they show up as improvements in initial capital (fixed effects) for groups with shorter mine lives.[27] This finding is plausible and concurs with our interpretations of the indices.

3. Diagnostics on the log productivity relationship

The model underlying our productivity analysis is decidedly simple, and the results are interesting. As noted above, the interpretation of our results relies on the

[27] Mine vintage effects can be seen from the average of fixed effects for surviving mines—for instance, see Fig. 12 of Ellerman et al. (1998).

Table 5
Relation of mine life to scale and fixed effects

	Average Mine Life	Average Annual Growth Rate, 1972–1995		
		Scale Effects Index	Fixed Effects Index	Combined
PRB	15	2.77	0.46	3.23
LIG	12.65	3.43	−2.37	1.06
APP-LW	10.96	2.21	−0.09	2.12
WST-S	8.99	0.77	0.09	0.86
WST-LW	7.72	2.97	0.6	3.57
INT-LW	7.57	1	−0.55	0.45
INT-CM	7.49	0.46	1.61	2.07
WST-CM	7	0.46	3.15	3.61
APP-CM	4.57	−0.30	3.17	2.87
INT-S	4.14	−0.72	2.39	1.67
APP-S	4.12	0.87	2.16	3.02

assumption that *output is predetermined* in our estimation procedure. To judge this assumption, we applied a variety of new (and old) techniques for studying endogeneity problems, such as those in the literature on weak instrument bias, as discussed below. However, it is worth stating at the outset that we did not find compelling evidence against our results using any of the diagnostic methods.

3.1. Traditional linear methods

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, where the diagnostic methods are more developed. We return to nonlinear specifications in Section 3.2.

3.1.1. Interpretation of the productivity-scale effect

It is useful to consider our estimates in the context of familiar Cobb–Douglas formulae. Suppose that the production function for a coal mine is

$$Q^* = A(L^*)^{\omega}(K_1^*)^{\rho_1} \cdots (K_M^*)^{\rho_M}, \tag{13}$$

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

$$\eta = \omega + \rho_1 + \cdots + \rho_M. \tag{14}$$

Minimizing total cost $TC = wL^* + \sum_{i=1}^{M} r_j K_j^*$ subject to predetermined output in (13) gives log-labor as

$$\ln L^* = -a + \left(\frac{1}{\eta}\right) \ln Q^*, \tag{15}$$

where a depends on A and input prices.[28] This implies that log-labor productivity is

$$\ln\left(\frac{Q^*}{L^*}\right) = a + \left(1 - \frac{1}{\eta}\right) \ln Q^*. \tag{16}$$

Returns-to-scale with regard to variable inputs is captured in the coefficient of log output; returns are decreasing, constant or increasing if $\beta_1 = (1 - 1/\eta)$ is negative, zero or positive, respectively. In log-linear form, our model implements (16),[29] and our strongly positive estimates of β_1 are consistent with substantial economies of scale.

3.1.2. Errors-in-variables and bracketing

We first examine whether traditional bracketing results are consistent with economies of scale.[30] Denote true log output and log labor as $q^* = \ln Q^*$, $l^* = \ln L^*$, respectively, and true log-labor productivity is $pr^* = q^* - l^*$. Write (16) as

$$pr^* = \alpha + \beta_1 q^* + \varepsilon, \tag{17}$$

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

$$\begin{aligned} q &= q^* + v, \\ l &= l^* + \varepsilon^*, \end{aligned} \tag{18}$$

where v, ε^* are homoskedastic errors that have mean 0 conditional on $q^*$and l^*. We set $\zeta = \varepsilon - \varepsilon^*$and assume that $Cov(v, \zeta) = 0$. Denote the percentages of error (variance) in the observed variables as

$$\lambda_q = \frac{Var(v)}{Var(q)}, \quad \lambda_l = \frac{Var(\zeta)}{Var(l)}. \tag{19}$$

Denote the OLS coefficient of l on q as $\hat{b}_{lq}$. The standard bias result is

$$\operatorname{plim} \hat{b}_{lq} = \left(\frac{1}{\eta}\right)(1 - \lambda_q) = (1 - \beta_1)(1 - \lambda_q), \tag{20}$$

and for $\hat{b}_{ql}$, the OLS coefficient of q on l,

$$\operatorname{plim} \hat{b}_{ql} = \eta(1 - \lambda_l) = \left(\frac{1}{1 - \beta_1}\right)(1 - \lambda_l). \tag{21}$$

[28] Specifically $a = (\ln A + \sum_j \rho_j \ln (w\rho_j / r_j \omega))/\eta$.

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

[30] 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.

These give rise to the well-known bracketing formula as

$$\operatorname{plim} \hat{b}_{lq} \leqslant 1 - \beta_1 \leqslant \frac{1}{\operatorname{plim} \hat{b}_{ql}}. \tag{22}$$

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

$$\operatorname{plim} \hat{b}_{pr,q} = 1 - \operatorname{plim} \hat{b}_{lq} = \beta_1 + \lambda_q(1 - \beta_1). \tag{23}$$

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 $\operatorname{plim} \hat{b}_{pr,q} = \lambda_q$; in that case, errors in the observed log output could give a spurious finding of estimated increasing returns. To bracket β_1, (22) transforms to

$$1 - \frac{1}{\operatorname{plim} \hat{b}_{ql}} \leqslant \beta_1 \leqslant 1 - \operatorname{plim} \hat{b}_{lq}. \tag{24}$$

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 framework above. We compute the bracketing formulae using the residuals of ln L and ln Q regressed on all mine and time effects in the roles of l and q above, which take the errors in ln L and ln Q to be uncorrelated with the mine and time effects. In addition to estimating the bounds of (24):

$$LB = 1 - \frac{1}{\hat{b}_{ql}}, \quad UB = 1 - \hat{b}_{lq}, \tag{25}$$

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

$$ALB = 1 - \frac{1}{\hat{b}_{ql}} - cs_{\hat{b}_{ql}}\left(\frac{1}{\hat{b}_{ql}^2}\right), \quad AUB = 1 - \hat{b}_{lq} + cs_{\hat{b}_{lq}}, \tag{26}$$

where $s_{\hat{b}_{ql}}$, $s_{\hat{b}_{lq}}$ are the estimated standard errors of $\hat{b}_{ql}$, $\hat{b}_{lq}$, and $c = 1.96$.[31]

The bounding results are presented in Table 6. The bracketing bounds are fairly wide, which is consistent with the overall goodness-of-fit of the equations. However, on the question of returns to scale, the value $\beta_1 = 0$ (constant returns) is contained in the intervals for only two of the 11 mine groups, APP-S and WST-LW. Even in these two cases the bounding intervals contain mostly positive values,[32] and we view it as reasonable to conclude that our finding of increasing returns is not spurious. Nevertheless, the bracketing bounds are wide, and so we now turn to other methods of estimating the scale effect.

[31] Appendix A shows how these adjusted bounds give a conservative 95% confidence interval asymptotically.

[32] 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.

Table 6
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

3.1.3. Instrumental variable estimates of the scale effect

We begin with some simple instrumental variable estimations of the scale effect. Since we do not observe a separate indicator of output, we assume that any measurement error is uncorrelated across time periods, and use linear combinations of lagged outputs as instruments.[33] We focus on the model in log-linear form, where, for simplicity, we have first-differenced to remove the mine fixed effects:

$$\Delta pr_{it}^* = \beta_1 \, \Delta q_{it}^* + \Delta \tau_t + \Delta \varepsilon_{it}, \tag{27}$$

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

$$\begin{aligned} q_{it} &= q_{it}^* + v_{it}, \\ l_{it} &= l_{it}^* + \varepsilon_{it}^*, \end{aligned} \tag{28}$$

where errors are uncorrelated over time and over mines,

$$\mathrm{E}[\varepsilon_{js}^* \varepsilon_{it}^*] = 0 \text{ and } \mathrm{E}[v_{js} v_{it}] = 0 \quad \text{when either } i \neq j \text{ or } s \neq t, \tag{29}$$

uncorrelated across log labor and log output,

$$\mathrm{E}[\varepsilon_{js}^* v_{it}] = 0 \quad \text{for all } i,j \text{ and } s,t, \tag{30}$$

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

$$\mathrm{E}[l_{js}^* \varepsilon_{it}^*] = 0, \;\; \mathrm{E}[q_{js}^* \varepsilon_{it}^*] = 0, \;\; \mathrm{E}[q_{js}^* v_{it}] = 0, \;\; \mathrm{E}[l_{js}^* v_{it}] = 0, \tag{31}$$

for all s, t, i, j.

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

Potential instruments for Δq_{it} include any linear combinations of q_{is} for $s<t-1$, as measurement error is uncorrelated over time. Various assumptions guarantee that these instruments are also correlated with Δq_{it}. For instance, if q^*_{it} follows a stationary process of the form:

$$(q^*_{i,t} - \overline{q}^*_i) = \rho(q^*_{i,t-1} - \overline{q}^*_i) + \xi_{it}, \tag{32}$$

where ξ_{it} are i.i.d. and $|\rho|<1$, then q_{is} is correlated with q_{it}, and q_{is} (with $s<t-1$) is a valid instrument for Δq_{it}. For our estimates, we began with the twice lagged difference as instrument $\Delta q_{i,t-2}$ and then estimated using the associated log-output levels as instruments, $q_{i,t-2}$ and $q_{i,t-3}$.

The 2SLS estimates of the scale effect β_1 are given in Table 7. 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 groups (APP-S, APP-CM, LIG, WST-S, WST-LW) the 95% confidence intervals for β_1 clearly exclude $\beta_1 = 0$, using lagged first differences or lagged levels as instruments. In one group (WST-CM), $\beta_1 = 0$ is excluded at a 90% confidence level with the lagged levels as instruments. For the other five mining groups the scale effect estimate is very imprecise and $\beta_1 = 0$ is not rejected. The problem appears to be due to the weakness of the instruments: for those groups, the correlation between Δq_{it} and $\Delta q_{i,t-2}$ or $q_{i,t-2}$, $q_{i,t-3}$ is small.

We expanded the instrument set to include powers of the lagged levels and powers of the lagged first differences; to justify this, we assume that

$$\mathrm{E}[q^d_{is} v_{it}] = 0 \quad \text{for } t \neq s \text{ and } d = 0, 1 \ldots,$$

which holds if q^*_{is}, v_{is} and v_{it} are mutually independent, for instance. Table 7 also presents these estimates, and they are generally much more precise. The hypothesis $\beta_1 = 0$ for constant returns is rejected for at least one set of estimates for every mining group except INT-S.

We have uncovered no evidence to cast doubt on our finding of increasing returns in mining, and we have obtained fairly precise estimates of the scale coefficient for all but one region (INT-S). However, the correlations between the instruments and log output do appear to be small.[34] This raises the possibility that we have weak instrument bias: instruments that exhibit sample correlations with measurement error, which biases IV estimates toward the OLS estimates.[35] For this, we examine the F-statistic of the first step regressions: log output regressed on the instruments. Bound et al. (1995) and Staiger and Stock (1997) note that an F-statistic value of 1 or less indicates that

$$\frac{\mathrm{E}\left[\hat{\beta}_{1,2SLS}\right] - \beta_1}{\mathrm{E}\left[\hat{\beta}_{1,OLS}\right] - \beta_1} \tag{33}$$

[34] A table of these correlations is available from the authors. Many of the simple correlations between current log output and its lags are below 0.1 in absolute value.

[35] See Nelson and Startz (1990).

Table 7
IV estimates of scale effect for log-linear model

Mine group	OLS first differences		2SLS with instruments			Sample size	Number of mines
	Δ (ln Q_t)	Δ (ln Q_{t-2})	ln (Q_{t-2}), (ln (Q_{t-3})	Powers of Δ (ln Q_{t-2})	Powers of ln (Q_{t-2}), ln(Q_{t-3})		
APP-S	0.363 (0.0113)	0.300 (0.0892)	0.254 (0.0703)	0.251 (0.0846)	0.253 (0.0547)	15,943	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)	16,046	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

2SLS estimation samples smaller than observation sample due to availability of instruments. Standard errors in parentheses.

is significantly different from 0, so that the bias in IV estimates is comparable to OLS bias. From Table 8, we note that we obtain uniformly low F-statistics for several regions (INT-LW and PRB, for instance) and in many cases, we see the F-statistic values declining as more instruments are added.[36]

[36]Strictly speaking, the F-statistic criterion requires the absence of heteroskedasticity, which we have not ruled out. Also, it is fairly common practice in panel data analysis to separate instruments by time period

Table 8
F-Statistics from first-step regressions

Main group	2SLS with instruments			
	$\Delta(\ln Q_{t-2})$	$\ln(Q_{t-2}), \ln(Q_{t-3})$	Powers of $\Delta(\ln Q_{t-2})$	Powers of $\ln(Q_{t-2}), \ln(Q_{t-3})$
APP-S	49.77	36.02	13.77	12.07
APP-LW	3.31	1.66	3.43	2.23
APP-CM	55.75	43.26	26.34	18.97
INT-S	4.47	3.02	1.18	1.57
INT-LW	0.06	0.06	1.28	0.71
INT-CM	1.08	0.97	2.87	1.63
WST-S	2.41	1.21	4.02	2.22
WST-LW	12.05	9.97	5.91	4.54
WST-CM	0.08	2.49	2.35	1.8
PRB	0.8	0.55	0.38	1.42
LIG	13.89	7.4	3.11	2.64

F-statistic from first-step regressions for estimates of Table 7.

This is sufficient reason to seek improvement in the instruments. Instead of adding more lags, etc., we enhance the instrument list systematically with the GMM approach of Blundell and Bond (1998). Recall that our basic model is

$$pr_{it} = \beta_1 q_{it} + \alpha_i + \tau_t + \varepsilon_{it} + v_{it}^*, \tag{34}$$

where ε_{it} and v_{it}^* are uncorrelated over time, and q_{it} is uncorrelated with ε_{it}, but q_{it} is potentially correlated with the measurement error term v_{it}^*.[37] Estimation based on first-differencing the model and using lagged levels as instruments coincides with implementing the moment conditions:

$$\mathrm{E}[q_{is}\,\Delta v_{it}^*] = 0 \quad \text{for } s < t-1. \tag{35}$$

Blundell and Bond (1998) add the moment conditions

$$\mathrm{E}[\Delta q_{is} v_{it}^*] = 0 \quad \text{for } s < t, \tag{36}$$

which coincide with estimating the model in levels, using first differences as instruments. The combined set of restrictions define a linear GMM estimator.[38]

(*footnote continued*)
(that is, if a variable is observed for 7 years, to define seven instruments, with its value in each of the 7 years.) We have not processed our instruments in this way, in part because it would frequently lower the first-stage F-statistics. These calculations are available from the authors.

[37]Later we allow ε_{it} to be serially correlated.

[38]The estimator essentially pools IV estimates of the level equation and first difference equation with respective instruments given in the moment conditions. Estimates were computed using the DPD98 software of Arellano and Bond (1998). We used 'first step' estimators that do not weight for cross correlations in residuals—while less efficient, Blundell and Bond (1998) noted that 'first step' estimators appeared to provide more reliable standard error estimates.

Table 9
System GMM estimates of scale effect

Mine group	Basic model (34) GMM with instrument		Model with autocorrelated error (37) GMM with instrument	
	$\ln(Q_{t-2})$	$\ln(Q_{t-3})$	$\ln(Q_{t-3})$	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)

Standard errors in parentheses.

The first two columns of Table 9 contain the GMM estimates using log-quantity lagged 2 and 3 periods as instruments.[39] The GMM estimates are not statistically significantly different from our previous IV estimates and have much smaller standard errors. Moreover, the GMM estimates are not systematically smaller than the OLS estimates of Table 7, so it does not appear that measurement error has

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

Table 10
Various specification tests: GMM estimates

Main 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

P-values for all tests.

significantly biased our results. The hypothesis of constant returns to scale is clearly rejected in all regions.

Various diagnostic statistics are given in Table 10. Sargan tests of the (overidentifying) moment restrictions show rejection at 90% level for three regions; while there may be some detectable correlation between instruments and residuals, the associated estimate bias may be small. Since the GMM estimates are close to the IV estimates but much more precise, such potential bias appears to be a small price to pay. Despite the lack of empirical evidence rejecting the GMM estimates, there are stationarity restrictions in Blundell and Bond (1998) that may be problematic for the short time series of output values (starting when the mine opens). To justify this, one can argue that productivity, once purged of fixed effects, is determined by a long series of exogenous shocks and management practice and expertise that predates the mine opening. Blundell and Bond (1998) provide further discussion of how their conditions can hold with entry period disequilibrium.

In addition, Table 10 presents test statistics for the presence of second-order serial correlation, for which there is evidence in three regions. This is an issue if the serially correlated part of the error is correlated with q_{it}, in violation of our basic error assumptions. We examine this possibility[40] by generalizing the model to have an autocorrelated error, as

$$\begin{aligned} pr_{it} &= \beta_1 q_{it} + \alpha_i + \tau_t + \varepsilon_{it} + v^*_{it}, \\ \varepsilon_{it} &= \rho \varepsilon_{it-1} + \xi^*_{it}, \end{aligned} \tag{37}$$

[40]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 7 shows that using instruments lagged by one more year still clearly rejects the hypothesis of constant returns to scale, so there is no strong evidence against our conclusions here.

where ξ^*_{it} and v^*_{it} are homoskedastic and uncorrelated over time, and q_{it} is potentially correlated with the productivity shock ξ^*_{it} and with the measurement error v^*_{it}.

The GMM estimates of model (37) are also presented in Table 9. While the estimates of returns to scale differ somewhat from the estimates of the basic model (Eq. (34)), 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 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 fair to say at this point that we may have pushed the log-linear analysis too far. While we find some serial correlation in the log-linear model residuals, that could easily arise from the nonlinearities that exist in the true data relationships. We turn to those considerations in the next section. In summary, the GMM approach improves the precision of our estimates, and the GMM estimates of returns to scale are not substantially smaller than the OLS estimates. As such, we believe we are justified in concluding that our findings of increasing returns to scale in every mining group are not altered by more general estimates that allow for autocorrelation and endogeneity.

3.2. Nonlinear models and measurement error

Our diagnostics above would suffice if all of our estimated productivity relationships were log linear. However, in our basic 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 nonlinear context, the diagnosis of potential problems from measurement errors and the like is, if anything, quite daunting. Solutions with nonlinear models require sufficient assumptions and additional information to assess the amount of measurement error, such as an independent measurement of the regressor of interest.[41] We do not have such additional information, so that if measurement error is a serious problem we will have difficulty in obtaining consistent estimates for the quadratic and cubic models.

In a nonlinear model, the simple bracketing methods of Section 3.1.2 are not directly applicable. Klepper and Leamer (1984) have developed more general bounds on regression coefficients in situations of multiple regressors with uncorrelated measurement errors. Bekker et al. (1987) have extended their methods to the case of correlated measurement errors. Such correlation is relevant with polynomial regressors: a positive error in log output q implies a positive error in all regressors of the form q^d. However, these methods do not address the case of polynomial regressions in a fully satisfactory manner, because they do not use the knowledge that the different regressors are powers of the same variable (measured with error).

[41] Amemiya (1985) first noted how traditional instrumental variable techniques are unable to address measurement error problems in nonlinear specifications. See Hausman et al. (1991), Hausman et al. (1995), Newey et al. (1999) and Wang and Hsiao (2003) for related discussion.

For this reason, we study measurement error issues using results available from polynomial regression. We assume that the quadratic and cubic specifications are, in fact, the true specifications of the productivity equation. Then for an assumed level of measurement error, we can adjust our estimates, showing what the productivity relationship would be without the error. This gives some useful insights.

Specifically, we assume the true model is

$$pr = \sum_{i=1}^{r} \beta_i (q^*)^i + \sum_{l=1}^{s} \delta_l z_l + \varepsilon, \tag{38}$$

where the z's are regressors free of measurement error (our time and mine-specific fixed effects). The observed log output q is true log output q^* measured with error as

$$q = q^* + v.$$

We assume that ε is independent of q^*, z and v (with $\mathrm{E}[\varepsilon] = 0$), that v is independent of q^* and z. We further assume that v is distributed as $\mathrm{N}(0, \sigma_v^2)$, although we could use any distribution (with known polynomial moments).

It is clear that the polynomial regression coefficients estimated using q will be (asymptotically) biased. We derive[42] the appropriate adjustments to the true coefficients, those that would be estimated from polynomial regression using q^*. Those adjustments depend on the value of the error variance σ_v^2. Unfortunately, we do not have repeated or independent observations on output (or particularly compelling instruments) for measuring this variance. Instead, we compute the adjustments for different levels of error, to see how our results would be affected.[43]

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 (see Section 2.3), and examine how the estimates would be adjusted if known amounts of measurement error existed in the data. In particular, we set the measurement error variance to be 0%, 5% and 10% of the variance of observed log-output deviations (orthogonal to the fixed and time effects).[44]

Fig. 5 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 adjusted values for 5% and 10% error. For the mine groups with log-linear models (APP-LW, INT-LW), we see the downward slope adjustment as implied by (23). For the nonlinear models, the adjustments are particularly interesting. While there are some differences for low scales, the main differences in shape occur at high output levels. For high output

[42]Our derivations are given in the appendices, and are similar to Hausman et al. (1991), Chescher (1998) and Cheng and Schneeweiss (1998).

[43]Our approach is similar to Griliches and Ringstad (1970). See Hausman et al. (1995) for approaches using instruments. Other references include Newey (2001), Lewbel (1996), Wang and Hsiao (2003) and Li (2002). Schennach (2004a, b).

[44]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 $\frac{Var[(q^*)^2]}{Var(q)^2} = .80$ and $\frac{Var[(q^*)^3]}{Var(q)^3} = .71$ so the induced measurement error throughout all regressor terms is much higher than 10%.

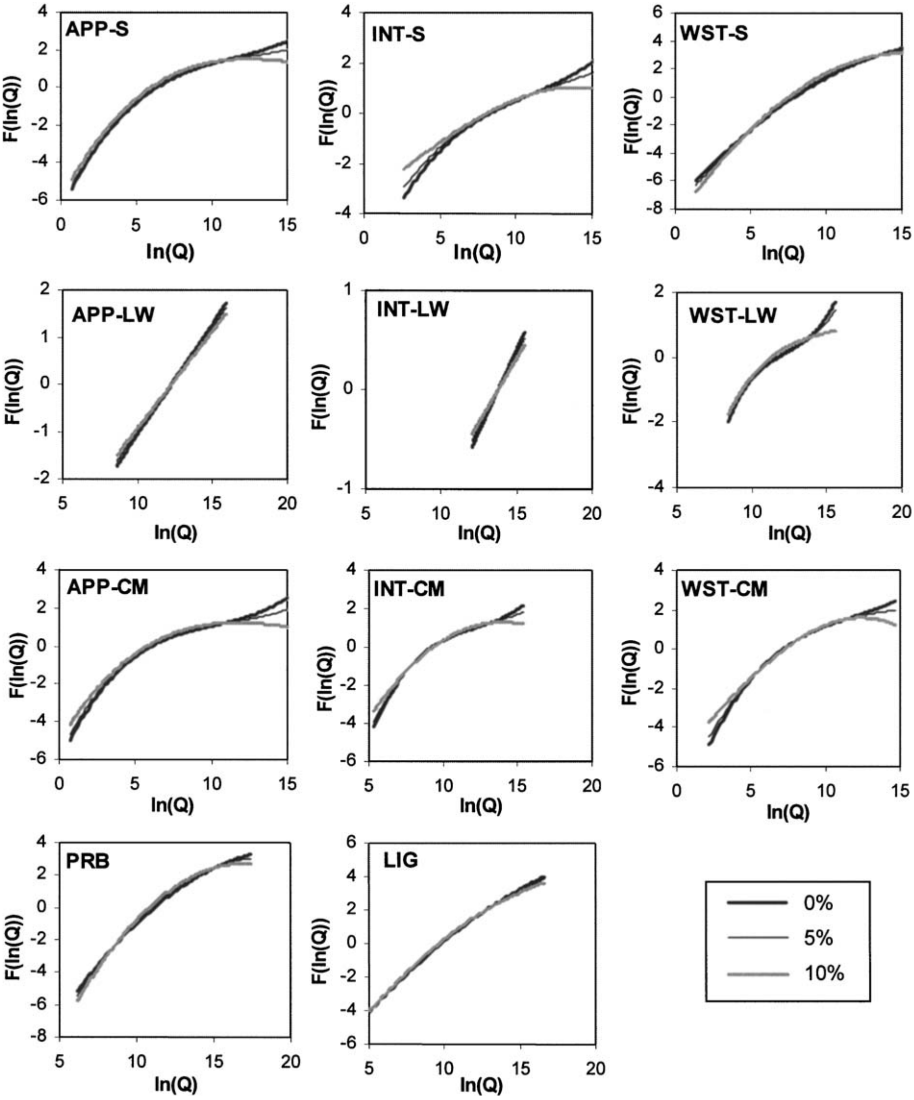


Fig. 5. Scale effects adjusted for measurement error.

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.

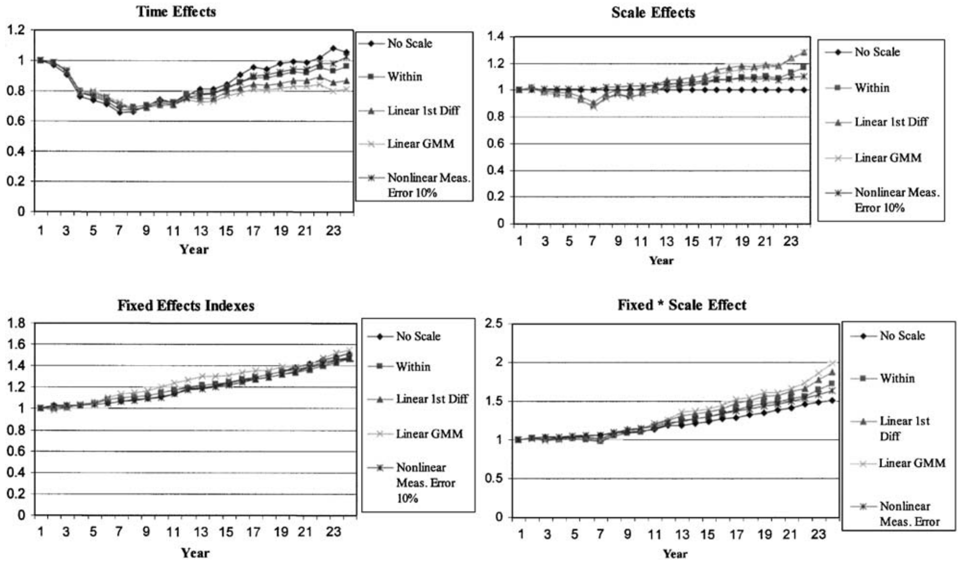


Fig. 6. Productivity decomposition for different estimates.

3.3. Implications for productivity measurements

In the various specifications studied as part of our diagnostic work, 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 found. In this section, we examine the implications of those differences for our analysis of productivity change in the U.S. coal industry.

Fig. 6 shows the evolution of our productivity indices for five different sets of estimation results. The ‘‘Within’’ estimates refer to our basic (nonlinear) estimates from Table 2 (indices presented earlier in Fig. 4), and serve as a benchmark for comparison. ‘‘No Scale’’ refers to indices constructed by assuming no scale effect on productivity; namely constant returns to scale in all mining groups. Two sets of results are presented with log-linear specifications for all groups: ‘‘Linear 1st Diff.’’ uses OLS estimates of the first-differenced model (estimates from first column of Table 7), and ‘‘Linear GMM’’ refers to the basic Blundell–Bond estimates (first column of Table 9[45]). Finally, ‘‘Nonlinear Meas. Error 10%’’ refers to the scale effects adjusted for 10% independent measurement error, as displayed in Fig. 5.

In broad terms, the different estimates are not associated with dramatically different interpretations of productivity change in U.S. coal mining. Somewhat surprisingly, the time pattern of the fixed effects indices are quite similar, exhibiting

[45]These are computed from the nine mine groups for which estimates were obtained.

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 finally by the no growth "No Scale" simulation. Broadly speaking, the log-linear models tend to overstate the role of scale relative to nonlinear models.[46] These figures also show a tendency for offsets between the scale effect indices and the time effect indices. Specifically, estimates associated with greatest growth in the scale effect indices are associated with the least variation in the time effects indices, and vice versa. Indices computed from our nonlinear estimates fall into the middle range of growth for scale indices and time effect 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 striking productivity increase after 1978. We began with data on annual output and labor input for every coal mine in the U.S., and studied productivity with panel regression methods. Panel methods permit a natural representation of heterogeneity via fixed effects for mines, and time effects.

We defined productivity indices based on the parameter estimates from the panel model analysis, and used them to delineate sources of productivity growth. The fixed effect index showed how (average) fixed effect values for mines in operation increased uniformly over the time period, which we interpreted as representing technical 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 and 1995, we found that virtually all the change in observed labor productivity was captured by those two indices (Fig. 6). This is true but a bit misleading; when examining the period 1978–1995 of rapid productivity increase, we find comparable, essentially uniform increases in fixed effect, scale effect and time effect productivity indices.

Our model of labor productivity is nonlinear but reasonably simple, partly 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 a variety of tests and analyses. We did not find any strong evidence against our original estimates. However, we believe that the application of such 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 in other contexts.

[46] 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 difference 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 1970s.

Acknowledgements

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 the reviewers, and from William Bruno, William Dix, Larry Rosenblum and Dale Jorgenson. Finally, we thank the late Zvi Griliches for comments that motivated much of the diagnostic work in Section 3.

Appendix A. Conservative confidence intervals for bounded quantities

Let β be the true value. Assume there exist β_l and β_h such that

$$\beta_l \leqslant \beta \leqslant \beta_h.$$

Assume there exists asymptotically normal estimators $\hat{\beta}_l$ and $\hat{\beta}_h$ of β_l and β_h, respectively, and $\hat{\beta}_l$ and $\hat{\beta}_h$ may or may not be correlated. Let $s_{\hat{\beta}_l}$ and $s_{\hat{\beta}_h}$ denote the standard errors of each estimate. For any given α, we want to find $\hat{l}$ and $\hat{h}$ such that

$$P\left[\beta \in \left[\hat{l}, \hat{h}\right]\right] > 1 - \alpha$$

asymptotically. The "greater than" sign makes this a *conservative* instead of an exact confidence interval. Note that

$$\begin{aligned} P\left[\beta \in \left[\hat{l}, \hat{h}\right]\right] &= 1 - P\left[\beta \notin [\hat{l}, \hat{h}]\right] \\ &= 1 - P\left[\beta < \hat{l} \text{ or } \hat{h} < \beta\right] \\ &> 1 - (P\left[\beta < \hat{l}\right] + P\left[\hat{h} < \beta\right]) \\ &> 1 - (P\left[\beta_l < \hat{l}\right] + P\left[\hat{h} < \beta_h\right]) \end{aligned}$$

for $\hat{l} = \hat{\beta}_l - c_{\alpha/2} s_{\hat{\beta}_l}$, we have that $P\left[\beta_l < \hat{l}\right] \to \alpha/2$ and, similarly, for $\hat{h} = \hat{\beta}_h + c_{\alpha/2} s_{\hat{\beta}_h}$, we have that $P\left[\hat{h} < \beta_h\right] \to \alpha/2$. This implies that

$$P\left[\beta \in \left[\hat{l}, \hat{h}\right]\right] > 1 - \alpha$$

asymptotically, as desired.

Appendix B. Adjusting polynomial models for measurement error

With reference to (38), the standard (normal) equations for estimation of the β's and the δ's are:

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

$$\mathrm{E}\left[yz_n\right] = \sum_{i=1}^{r} \beta_i \mathrm{E}\left[(q^*)^i z_n\right] + \sum_{l=1}^{s} \delta_l \mathrm{E}[z_l z_n] \text{ for } n = 1, \ldots, s, \tag{39}$$

and we could form estimates if the moments (expectations) involving true log-output q^* could be estimated. The moments of observed log output q are

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

where we have used the independence of q^* and v. We can now isolate $\mathrm{E}[(q^*)^n]$ and obtain a recursive relation:

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

Note that the $\mathrm{E}[v^{n-j}]$ are known because v has a known distribution. Similarly, we have

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

where we have used the independence between v and q^*, z. These relations allow the q^* moments to be estimated using the sample moments of q.

Now write $\mathrm{E}[yq^n]$ and $\mathrm{E}[yz_n]$ as functions of observed q moments and the true q^* moments determined above:

$$\begin{aligned}\mathrm{E}[yq^n] &= \sum_{i=1}^{r} \beta_i \mathrm{E}\left[(q^*)^i q^n\right] + \sum_{l=1}^{s} \delta_l \mathrm{E}[z_l q^n] \\ &= \sum_{i=1}^{r} \beta_i \mathrm{E}\left[(q^*)^i (q^* + v)^n\right] + \sum_{l=1}^{s} \delta_l \mathrm{E}[z_l q^n] \\ &= \sum_{i=1}^{r} \beta_i \sum_{j=0}^{n} \binom{n}{j} \mathrm{E}\left[(q^*)^{i+j}\right] \mathrm{E}\left[v^{n-j}\right] + \sum_{l=1}^{s} \delta_l \mathrm{E}[z_l q^n] \quad \text{for } n = 1, \ldots, r\end{aligned}$$

and

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

The regression adjustments follow from these modified normal equations, with β and δ isolated as

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

where

$$b = \mathrm{E}[yq^n] \quad \text{for } n = 1, \ldots, r,$$
$$d = \mathrm{E}[yz_l] \quad \text{for } l = 1, \ldots, s,$$
$$A_{ni} = \sum_{j=0}^{n} \binom{n}{j} \mathrm{E}\left[(q^*)^{i+j}\right] \mathrm{E}\left[v^{n-j}\right] \quad \text{for } n = 1, \ldots, r \text{ and } i = 1, \ldots, r,$$
$$B_{nl} = \mathrm{E}[z_l q^n] \quad \text{for } n = 1, \ldots, r \quad \text{and} \quad l = 1, \ldots, s,$$
$$C_{li} = \mathrm{E}\left[(q^*)^i z_l\right] \text{ for } l = 1, \ldots, s \quad \text{and} \quad i = 1, \ldots, r,$$
$$D_{nl} = \mathrm{E}[z_l z_n] \quad \text{for } n = 1, \ldots, s \text{ and } l = 1, \ldots, s$$

and

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

References

Amemiya, T., 1985. Instrumental variables estimator for the nonlinear errors-in-variables model. Journal of Econometrics 28, 273–289.

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

Baker, J.G., 1981. Sources of deep coal mine productivity change, 1962–1975. The Energy Journal 2, 95–106.

Bekker, P., Kapteyn, A., Wansbeek, T., 1987. Consistent sets of estimates for regressions with correlated or uncorrelated measurement errors in arbitrary subsets of all variables. Econometrica 55, 1223–1230.

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115–144.

Bound, J., Jaeger, D.A., Baker, R.M., 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.

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.

Byrnes, P., Fare, R., Grosskopf, S., 1984. Measuring productive efficiency: an application to Illinois strip mines. Management Science 30, 671–681.

Cheng, C-L, Schneeweiss, H., 1998. Polynomial regression with errors in the variables. Journal of the Royal Statistical Society B 60 (Part 1), 189–199.

Chescher, A., 1998. Polynomial regression with covariate measurement error. Discussion Paper 98/448, University of Bristol.

Ellerman, D., Stoker, T.M., Berndt, E., 1998. Sources of productivity growth in the American coal industry. MIT CEEPR Working Paper WP 98004, Massachusetts Institute of Technology, Boston.

Ellerman, D., Stoker, T.M., Berndt, E., 2001. Sources of productivity growth in the American coal industry: 1972–1995. In: Dean, E., Harper, M., Hulten, C. (Eds.), New Directions in Productivity Analysis. National Bureau of Economic Research. University of Chicago Press, Chicago.

Frisch, R., 1934, Statistical Confluence Analysis by Means of Complete Regression Systems. Publication No. 5, University Economics Institute, Oslo.

Green, P.J., Silverman, B.W., 1994. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. Chapman & Hall, London.

Griliches, Z., Ringstad, V., 1970. Errors in variables in nonlinear context. Econometrica 38, 368–370.

Griliches, Z., Ringstad, V., 1971. Economies of Scale and the Form of the Production Function. North-Holland, Amsterdam.

Hausman, J., Ichimura, H., Newey, W.K., Powell, J.L., 1991. Measurement errors in polynomial regression models. Journal of Econometrics 50, 271–295.

Hausman, J., Newey, W.K., Powell, J.L., 1995. Nonlinear errors in variables: estimation of some Engel curves. Journal of Econometrics 65, 205–233.

Jorgenson, D.W., 1990. Productivity and economic growth. In: Berndt, E.R., Triplett, J.E. (Eds.), Fifty Years of Economic Measurement: The Jubilee of the Conference of Research in Income and Wealth, University of Chicago Press, Chicago, pp. 19–118. (Reprinted in Jorgenson, D.W., 1995. Productivity: International Comparisons of Economic Growth, Vol. 2, The MIT Press, Cambridge, pp. 1–98.).

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

Joskow, P.L., 1987. Contract duration and relationship-specific investments: empirical evidence for coal markets. American Economic Review 77, 168–185.

Joskow, P.L., 1990. The performance of long-term contracts: further evidence from coal markets. Rand Journal of Economics 27, 251–274.

Keane, M.P., Runkle, D.E., 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.

Klepper, S., Leamer, E.E., 1984. Consistent sets of estimates for regressions with errors in all variables. Econometrica 52, 163–184.

Kruvant, W.J., Moody, C.E., Valentine, P.L., 1982. Sources of productivity decline in U.S. coal mining, 1972–1977. The Energy Journal 3, 53–70.

Leser, C.E.V., 1955. Production functions and British coal mining. Econometrica 23, 442–446.

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.

Li, T., 2002. Robust and consistent estimation of nonlinear errors-in-variables models. Journal of Econometrics 110, 1–26.

Madalla, G.S., 1965. Productivity and technological change in the bituminous coal industry, 1919–1954. Journal of the Political Economy 73, 352–365.

Nelson, C.R., Startz, R., 1990. The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. Journal of Business 63, S125.

Newey, W.K., 2001. Flexible simulated moment estimation of nonlinear errors-in-variables models. Review of Economics and Statistics 83, 616–627.

Newey, W.K., Powell, J.L., Vella, F., 1999. Nonparametric estimation of triangular simultaneous equations models. Econometrica 67, 565–603.

Schennach, S.M., 2004a. Estimation of nonlinear models with measurement error. Econometrica 72, 33–75.

Schennach, S.M., 2004b. Nonparametric regression in the presence of measurement error. Econometric Theory 20, 1046–1093.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London.

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.

U.S. Department of Labor, 1972–1995, Part 50 Coal Mining Address/Employment and accident Injury Files, Division of Mining Information Systems, Denver, CO, USA.

Wang, L., Hsiao, C., 2003. Identification and estimation of nonlinear errors-in-variables models. Working Paper, University of Southern California (revised January 2003).