ARE FIXED EFFECTS FIXED?
Persistence in Plant Level Productivity

Douglas W. Dwyer*
Department of Economics
Columbia University

CES 96-3 May 1996

Division, Center for Economic Studies, Washington Plaza II, Room 211, U.S. Bureau of the Census, Washington, D.C. 20233-6101, (301) 457-1882) or INTERNET address: snguyen@info.census.gov

Abstract

Estimates of production functions suffer from an omitted variable problem; plant quality is an omitted variable that is likely to be correlated with variable inputs. One approach is to capture differences in plant qualities through plant specific intercepts, i.e., to estimate a fixed effects model. For this technique to work, it is necessary that differences in plant quality are more or less fixed; if the "fixed effects" erode over time, such a procedure becomes problematic, especially when working with long panels. In this paper, a standard fixed effects model, extended to allow for serial correlation in the error term, is applied to a 16-year panel of textile plants. This parametric approach strongly accepts the hypothesis of fixed effects. They account for about one-third of the variation in productivity. A simple non-parametric approach, however, concludes that differences in plant qualities erode over time, that is plant qualities N-mix. Monte Carlo results demonstrate that this discrepancy comes from the parametric approach imposing an overly restrictive functional form on the data; if there were fixed effects of the magnitude measured, one would reject the hypothesis of N-mixing. For textiles, at least, the functional form of a fixed effects model appears to generate misleading conclusions. A more flexible functional form is estimated. The "fixed" effects actually have a half life of approximately 10 to 20 years, and they account for about one-half the variation in productivity.

Keywords: fixed effects; plant level productivity; textile

industries

[*]Columbia University and William M. Mercer Inc. Address correspondence to 512 West 112th St #3B, New York, NY 10025 or dwd4@columbia.edu. The author thanks Andrew Caplin, Vivek

Dehejia, Phoebus Dhrymes, Richard Ericson, Zvi Griliches, Boyan Jovanovic, Ariel Pakes, and the seminar participants of Columbia University's Applied Microeconomics Workshop and the NBER Productivity Program's Lunch Seminar Series for their questions and ideas. The assistance of Bob Bechtold, Robert McGuckin and Arnold Reznek at the Census Bureau's Center for Economic Studies is highly appreciated. Any remaining errors, however, are mine alone.

I. Introduction

Estimates of production functions provide the framework that economists employ to study the sources of economic growth (cf. Solow, 1957; Jorgenson, Gollop and Fraumeni, 1987; Bartelsman, 1992; Olley and Pakes, 1992; Gort, Bahk and Wall, 1993). Until recently, most of this work has been based on nation-wide or industry-wide information. The increasing availability of plant and firm level data in manufacturing has generated research in the micro-foundations of productivity growth and the relationship between plant level productivity dynamics and job flows (Bartelsman and Dhrymes, 1991; Davis and Haltiwanger, 1992; Baily, Hulten and Campbell, 1992; Caballero and Hammour, 1994; Campbell, 1994; Cooley, Greenwood and Yorukoglu, 1994; Atkeson, Khan and Ohanian, 1995). As old as the literature on the analysis of economic growth, however, is the debate on how to estimate production functions. That is, how to create the framework with which to perform the analysis (see Griliches and Mairesse, 1995 for a recent review of this issue).

Techniques for estimating production functions fall into two groups, econometric and non-econometric. The non-econometric (or accounting) method requires the assumptions of constant returns to scale and static cost minimization to take the cost shares as estimates of the output elasticities (cf. Solow, 1957). Under the econometric method, one regresses the log of output onto the logs of the different types of inputs (cf. Dhrymes, 1991). The residual of this equation (times 100) is the percentage of output produced, above and beyond expected output given the inputs used, i.e., productivity. This suggests the intrinsic problem with the econometric approach: as long as the manager knows more about his plant than the econometrician, and the manager uses this knowledge in his choice of inputs, the error term will be correlated with the independent variables resulting in biased parameter estimates (hereafter the simultaneity problem).

The most common "solution" to this problem is a fixed effects model: one assumes that there are unobserved, permanent productivity differences across plants and estimates these differences through plant specific intercepts. The null hypothesis of no fixed effects is routinely rejected via an F-test. For a panel that is short in the time dimension, this approach may be reasonable. Over time, however, it is certainly possible that these "permanent" productivity differentials change because the plant retools, the manager retires, and/or the product mix changes. In order for a fixed effects model to solve the simultaneity problem in a long panel, it is necessary that the "fixed" effects are relatively fixed over the length of the sample. This brings us in a rather round about way to a recent non-parametric test of alternative models of industry dynamics (Pakes and Ericson, 1995). Pakes and Ericson test an active exploration model (Ericson and Pakes, 1995) against a passive learning model (Jovanovic, 1982). In the active exploration model, firms invest to improve their quality, and hence their quality may change permanently. In the passive learning model, in contrast, firms are born with a fixed quality that they learn over time. Therefore, under passive learning there are permanent differences between plants, i.e., fixed effects.

The sharpest distinction between the two models involves the concept of N-mixing--N-mixing is one definition of a stochastic sequence ($x_t, x_{t+1}, x_{t+2}, ...$) becoming independent of the initial value (x_0) as t becomes large. In the active exploration model, the effect of being a certain quality today will erode over time (plant quality N-mixes). In the passive learning model, in contrast, it cannot (plant quality does not N-mix). Pakes and Ericson use firm size (number of employees) as a proxy for firm quality. They find that there are permanent differences in the size of Wisconsin retail firms, which is consistent with the passive learning model. The active exploration model, in contrast, is consistent with Wisconsin manufacturing firms, because manufacturing firms do not have permanent size differences. In this paper, I present a non-parametric test for N-mixing of plant productivity levels.[1]

---

[1]My test is an application of the general linear model, it is asymptotically equivalent to Pakes and Ericson's unconstrained test when I allow for heteroscedasticity of the error term.

The log of total factor productivity (tfp), the residual of a production function, is arguably a better measure of firm or plant quality than size in a competitive market.[2] In this paper, I measure tfp at the plant level in the textile industry using a 16-year panel.[3] I then test for the presence of fixed effects using a conventional parametric method, extended to allow for serial correlation in the error term, and find that there are permanent differences in tfp. In fact, the permanent differences account for about one-third the variation in productivity. From the non-parametric viewpoint, in contrast, plant level tfp's N-mix, which implies that there are no permanent differences in tfp. Monte Carlo techniques resolve the discrepancy between the two approaches; if there actually were fixed effects in the data of the magnitude measured, the non-parametric test would reject

---

[2]Economists have long been interested in the degree of persistence in measures of plant/firm quality. Mueller (1986; see also Pakes' review, 1987) and Rumuelt (1991) look at profit rates. Roberts and Supina examine output price (1994). Finally, Dhrymes (1991), Bartelsman and Dhrymes (1991), Baily, Hulten and Campbell (1992), and Dwyer (1995b) study the persistence in plant productivity levels. The technique developed in Section VI may provide a useful method for describing the nature of persistence in all of these measures.

[3]The data are an extract of the Longitudinal Research Database (LRD). This extremely rich data base is based on the Annual Survey of Manufactures and the Census of Manufactures. Unfortunately, access to the data requires financial support and a security clearance. Furthermore, the research must be performed in residence and removing statistics requires a clearance procedure to ensure that confidentiality is maintained. Further information is available on the Census Bureau's world wide web site (http://www.census.gov/ces.html).

the hypothesis of N-mixing. It appears that the parametric form of the fixed effects model is too restrictive and consequently leads to misleading results. This suggests that parametric results that find fixed effects in panel data may be misleading.

This result is consistent with an active exploration model for plant dynamics in the textile industry; the textile plants that survive are those that are able to reinvent themselves. In terms of our understanding of dynamic industry equilibria, this finding leads to two more questions. First, how strong is the survival condition? Second, how long does it take plants to reinvent themselves? The survival condition is strong but not overwhelming; sixteen year survival rates range from 30 to 60 percent across the 21 different four-digit textile industries.[4] The answer to the latter question requires one to estimate a more flexible functional form for the time series properties of plant level tfp.

It appears as though there are at least two components to plant level tfp. One component is transitory while another is highly persistent. The transitory component may result from idiosyncratic supply or demand shocks and/or transitory measurement error. The persistent component may be the product of managerial and technological differences and/or persistent

[4]The survival rate is the percentage of plants that were in the industry in 1972 and were in any manufacturing industry in the 1987 Census.

measurement error. Such a decomposition can be represented as the sum of two AR(1) processes, with the persistent and transitory components having a large and a small autocorrelation coefficient, respectively. Consistent estimates of these two components reveal a great deal of persistence in the latter component; if a plant is 50 percent above average today, net of the transitory component, you would expect it to be 25 percent above average 10 to 20 years from now. This degree of persistence suggests that either the market is working rather slowly or there is measurement error in the persistent component.[5] Furthermore, the persistent component explains about 45 percent of the variation in productivity, which is more than suggested by the fixed effects model. This suggests that forcing the persistent component to be fixed understates its importance.

The remainder of this paper is organized as follows. In the next section, I present the econometric background behind the fixed effects model as a solution to the simultaneity problem. In section III, I discuss my data and how I measure plant level tfp. Sections IV and V present the parametric case for fixed effects and the non-parametric case for N-mixing, respectively. Section V resolves the discrepancy in the two results through

[5]Not all of the dispersion in productivity levels is the product of measurement error. Plants with above average productivity levels expand faster and are less likely to exit (Dwyer, 1995b). I will come back to this issue in the conclusion.

Monte Carlo techniques. Section VI presents consistent estimates of a more flexible functional form, which suggest that the fixed effects actually have a half life of 10 to 20 years. A brief consideration of the econometric and economic implications of these results finishes out the paper.

II. Econometric Background

The argument for a simultaneity problem in estimates of production functions is simple and old.[6] Its simplest form is the following. If the quality of one plant is higher than that of another, then the owner of the more productive plant will choose to use more inputs. This implies that the OLS estimate of the elasticity of output with respect to the input has an upward bias, as can be analytically illustrated.

Suppose the true production function is:

$$y_{it} = \alpha x_{it} + a_{it} + \epsilon_{it},$$

where y is the log of output, x is the log of an input, *a* is a plant specific effect that is known to the firm but unknown to the econometrician (allowing for more than one production input is straightforward), " is the elasticity of output with respect to the input, and , is an idiosyncratic shock unknown to both the manager and the econometrician. Suppose the manager chooses

---

[6]For a detailed history of this argument see Griliches and Mairesse (1995).

x to maximize profits:

$$X^{*}_{it} = argmax_{[x]}\ E(A_{it}X^{\alpha}e^{\epsilon_{it}} - wX_{it}) \Rightarrow$$
$$x^{*}_{it} = \left(\frac{1}{1-\alpha}\right)\left(\log\left(\frac{\alpha E(e^{\epsilon_{it}})}{w}\right) + a_{it}\right),$$

where lower case letters denote logarithms, w is the input cost per unit output and x^* denotes the optimal level of the input (In order for x^* to be finite, " must be less the one, i.e., decreasing returns to scale.). Note that x^* is positively correlated with a. The OLS estimate of " is given by:

$$\hat{\alpha} = \frac{Cov(y,x)}{Var(x)} = \frac{Cov(\alpha x, x) + Cov(a,x)}{Var(x)} = \alpha + \frac{Var(a)}{Var(x)(1-\alpha)}.$$

Therefore, " has an upward bias (^ > "); actually ^ will always equal one in this framework (Atkeson, Khan, and Ohanian, 1995).[7] A possible "solution" is to suppose that a_{it} is constant over time ($a_{it}=a_i$ for all t). Then one either includes plant dummies

---

[7] If there were errors in static optimization, i.e., $x_{it} = x_{it}^* + v_{it}$, then

$$E(\hat{\alpha}) = \alpha + \left(\frac{1}{1-\alpha}\right)\left(\frac{Var(a)}{Var(x)}\right).$$

Therefore, the E(^) goes to " and 1 as Var(v)/Var(a) goes to 4 and 0, respectively.

or estimates the regression as deviations from the time mean (y_{it}-$y_{i.}$ on x_{it}-$x_{i.}$), which is algebraically equivalent (for a recent exposition of this technique see Green, 1993). Provided the rest of the Gauss Markov assumptions apply, the estimate of " is now unbiased. Even if a_i were truly constant over time, such a methodology is not without costs. If there is transitory measurement error in x_{it}, then subtracting out the time mean of x increases the noise to signal ratio and aggravates the downward bias on " associated with the measurement error.[8] Nevertheless, the fixed effects solution is commonly utilized in the literature; sometimes the technique is contrasted with the between and total estimates and sometimes instrumental variables are used (cf. Irwin and Klenow, 1994; Jones and Kato, 1995).

Early on, the fixed effects technique was applied to agricultural data (Griliches and Mairesse, 1995). The hypothesis that a_{it} is fixed over time seems reasonable when looking at farms and short time periods. The quality of the land, which is relatively fixed over time, is known to the farmer before

[8]This hypothesis is rather consistent with what happens to the capital coefficient when estimating a fixed effects model; it becomes implausibly small both in my data and others (cf. Mairesse and Griliches, 1990; Olley and Pakes, 1992; Jones and Kato, 1995). Errors in measurement of capital include non-uniform capacity utilization and the fact that when new investment actually comes on line is not observed. It is reasonable to suspect that these errors in measurement are much more of a ***within*** (x_{it} - $x_{i.}$) than a ***between*** ($x_{i.}$) phenomenon, that is they will average out over time (Griliches, 1986 and for further elaboration see Griliches and Mairesse 1995).

choosing the level of production inputs. The idiosyncratic shock, ,, is then viewed as weather, which is unknown to both the econometrician and the framer until after the inputs have been purchased. In manufacturing, however, this interpretation becomes rather problematic, especially as the time dimension of panels becomes longer. For a certain period of time, a plant may have a fixed technology. Over time, however, there is always the possibility of a plant retooling. If the a_{it} is time varying, then the fixed effect method can not fully account for the simultaneity problem. Of course, the extent to which the fixed effects method partially accounts for the simultaneity problem is going to depend on how variable a_{it} is over the sample period.

III. Estimates of Productivity at the Plant Level.

My database, an extract of the Longitudinal Research Database (LRD), includes plants in 21 different four-digit textile industries from 1972 to 1987. The panel is highly unbalanced. This results from plants entering and exiting as well as the fact that small plants are sampled with a probability of less than 1 in non-census years. The appendix contains a description of the sampling methods as well as a discussion of the construction of the each variable. Table 1 reports the number of plants and firms ever present in each industry.

Suppose production in a four digit textile industry can be represented by a value added Cobb-Douglas production function:

$$y_{it} = \alpha l_{it} + \beta k_{it} + \epsilon_{it},$$

where y is value added, l is total employment, k is gross book value of capital, and , is the residual, i.e., the log of total factor productivity. This paper will examine the stochastic process behind the estimated residual, tfp. In order to estimate the residual, I first need consistent estimates of " and $, i.e., I need to perform the first stage of GLM. In order to obtain efficient estimates of " and $, one would then use the knowledge of the stochastic process behind the residual to estimate the second stage of a GLM procedure (for example, a random effects model).[9]

I use the OLS estimates of " and $ after including time and time region dummies to take into account potential simultaneity problems resulting from aggregate shocks.[10] Dwyer (1994) argues that this technique does reduce the simultaneity bias in the estimates of " and $. It is likely, however, that these dummy

---

[9]If there are simultaneity problems and measurement error in the data, correcting for heteroscedasticity and serial correlation does not lead to unbiased estimates. Since I believe that there are both simultaneity and measurement error in the data, it is not clear what one would gain by correcting for heteroscedasticity and serial correlation.

[10]I do not employ plant specific intercepts when estimating " and $, because they lead to implausibly low estimates of $, the output elasticity of capital.

variables do not fully eliminate the simultaneity problems and certainly do not eliminate the measurement error in the independent variables. I have, therefore, conducted robustness tests of the key results through alternative estimates of " and $. I use the labor share and one minus the labor share as estimates of " and $, which are consistent and unbiased under the assumptions of static cost minimization and constant returns to scale. Additionally, I use labor productivity, which arbitrarily sets " to 1 and $ to 0.[11] Regardless of the measure of productivity, the results are substantively the same.

For each four-digit industry, I estimate:

$$y_{it} = a + \sum_{t=72}^{87} \sum_{r=0}^{2} a_{tr} I_{itr} - a_{72\,0} I_{i72\,0} + \alpha l_{it} + \beta k_{it} + \epsilon_{it}.$$

The subscripts, itr, denote the plant, time period, and region respectively. The indicator variable, I_{irt}, is defined as:

$$I_{irt} = \begin{cases} 1 & \text{if year} = t \text{ and region} = r, \\ 0 & \text{otherwise,} \end{cases}$$

[11]Labor share is computed as the weighted average of total compensation divided by value added over the whole sample, where the weights are the real value added of each plant. The labor share is consistently smaller than the econometric estimate of ". Therefore, this measure of productivity places a larger weight on capital productivity. Labor productivity, measured as real value added per employee, places no weight on capital productivity. Most other measures of productivity, therefore, should lie "within" these three measures.

where region 1 is the mid-atlantic states (NY, NJ and PA), region 2 is the southern states (VA, WV, NC, SC, GA, FL, KY, TN, AL, MS) and region 0 is all other states. Table 2 summarizes the results of these regressions. Observe that the coefficient estimates are plausible (the capital coefficient is always greater than 0) and the production functions exhibit constant returns to scale or close to constant returns to scale. The residual, estimated tfp, is then computed as

$$tfp = \hat{e} = y - \hat{\alpha} l - \hat{\beta} k.$$

The estimated tfp is a random variable. This paper's results, with respect to statistical inference, pertain to this random variable rather than its true value.[12] The reader may be concerned with how my results regarding the stochastic process behind the estimated tfp, $\hat{,}_{it}$, relate to the true value of $,_{it}$. The fact that my results are robust to many measures of productivity, however, makes it unlikely that these results are the product of the measurement error in $\hat{"}$ and $\hat{\$}$.

[12] $tfp_{it} = ,_{it} + (\hat{"}-")l_{it} + (\hat{\$}-\$)k_{it}$.

Table 1: Number of Firms and Plants Ever Present in Each Industry

SIC	Number of Firms	Number of Plants

2211	(Broad woven fabric mills, cotton)	334	496
2221	(Broad woven fabrics mills, man made	531	776
	fiber and silk)	233	249
2231	(Broad woven fabric mills, wool)	422	460
2241	(Narrow fabrics and other smallwares	325	376
	mills)	541	609
2251	(Women's hosiery above the knee)	1583	1645
2252	(Women's hosiery below the knee)	139	167
2253	(Knit outerwear mills)	922	1008
2254	(Knit underwear mills)	499	548
2257	(Circular knit fabric mills)		
2258	(Lace goods and warp knit fabrics, an	180	177
	aggregation see appendix)	447	471
2259	(Knitting mills NEC)	468	523
2261	(Finishers of broad woven cotton	321	337
	fabrics)	678	733
2262	(Finishers of broad woven man-made	380	432
	fiber and silk)	586	858
2269	(Finishers of textiles NEC)	344	355
2273	(Carpets, an aggregation see appendix)	22	34
2282	(Yarn texturizing, throwing, twisting	217	249
	and winding mills)	249	267
2283	(Yarn and thread mills, an	885	931
	aggregation see appendix)		
2295	(Coated fabrics, not rubberized)		
2296	(Tire cord and fabric)		
2297	(Nonwoven fabrics)		
2298	(Cordage and twine)		
2299	(Textile goods NEC, an aggregation		
	see appendix)		

Table 2: Estimates of Production Functions

SIC	"	$	"+$	R^2
2211	0.8242	0.1739	0.9981	0.88
	(.0164)	(.0131)	(.0090)	
2221	0.8013	0.1720	0.9732*	0.86
	(.0117)	(.0093)	(.0071)	
2231	0.6936	0.2773	0.9709	0.86
	(.0274)	(.0224)	(.0151)	
2241	0.7740	0.1845	0.9585*	0.83
	(.0185)	(.0136)	(.0123)	
2251	0.8550	0.1665	1.0215	0.85
	(.0226)	(.0188)	(.0145)	
2252	0.8678	0.1849	1.0527*	0.84
	(.0177)	(.0135)	(.0103)	
2253	0.6332	0.3303	0.9635*	0.83
	(.0114)	(.0091)	(.0076)	
2254	0.8579	0.1369	0.9948	0.84
	(.0358)	(.0265)	(.0195)	
2257	0.7718	0.1859	0.9577*	0.80
	(.0144)	(.0113)	(.0089)	
2258	0.7811	0.2374	1.0185	0.83
	(.0210)	(.0161)	(.0124)	
2259	0.5732	0.3632	0.9363*	0.87
	(.0393)	(.0328)	(.0225)	
2261	0.8333	0.1929	1.0262	0.89
	(.0265)	(.0214)	(.0143)	
2262	0.8152	0.1776	0.9928	0.89
	(.0192)	(.0152)	(.0104)	
2269	0.8457	0.1784	1.0242	0.82
	(.0282)	(.0222)	(.0169)	
2273	0.7585	0.2467	1.0052	0.80
	(.0198)	(.0162)	(.0100)	
2282	0.7805	0.1992	0.9798	0.81
	(.0220)	(.0165)	(.0135)	
2283	0.8845	0.1319	1.0164*	0.79
	(.0132)	(.0101)	(.0081)	
2295	0.8193	0.2048	1.0241	0.82
	(.0258)	(.0197)	(.0143)	
2296	0.9080	0.1934	1.1014*	0.72
	(.0743)	(.0716)	(.0507)	
2297	0.7182	0.2739	0.9921	0.82
	(.0303)	(.0204)	(.0190)	
2298	0.8304	0.1753	1.0057	0.86
	(.0271)	(.0219)	(.0153)	
2299	0.7451	0.2559	1.0010	0.84
	(.0167)	(.0131)	(.0102)	

The standard errors are in parentheses, which should be interpreted with caution, because the procedure does not take into account the serial correlation in the error term. The * in column four denotes that the hypothesis of constant returns to scale can be rejected with 95 percent certainty.

IV. The Case for Fixed Effects

Previous work suggests that each plant's productivity has a permanent component and an idiosyncratic component that contains serial correlation (Dwyer, 1995b). Suppose that estimated tfp can be characterized by:

$$tfp_{it} = v_t + a_i + \varepsilon_{it}; \quad \varepsilon_{it} = \rho\varepsilon_{it-1} + \mu_{it},$$

where v_t is a non-stochastic time shock, a_i is a non-stochastic fixed effect, D is the autocorrelation coefficient, and $:_{it}$ is i.i.d. across time and plants. Let T and N be the time dimension and the cross-sectional dimension of the panel, respectively. Estimating the parameters of such a data-generating process is problematic in the context of an unbalanced panel. The first problem arises in the estimate of the serial correlation coefficient. Regressing the residual on the lag of a residual introduces a simultaneity problem, because the lagged endogenous variable is not independent of the error term by construction.[13]

---

[13]For expositional purposes, suppose v_t is known to be 0 for all t. The OLS estimate of D is obtained by regressing $,_{it}$ - $,_{i.}$ on $,_{it-1}$ - $,_{i.}$. The OLS estimate is therefore biased because both the right hand side and left hand side of the equation contain $,_{i.}$. The equation one is actually estimating is:

$$\varepsilon_{it} = \rho(\varepsilon_{it-1} - \varepsilon_{i.}) + \mu_{it} + \varepsilon_{i.}$$

Note that the measurement error in the dependent variable, $,_{i.}$, is negatively correlated with the error term, ($:_{it}$ + $,_{i.}$), which yields a downward bias on $\hat{D}$. The estimate of D will converge in probability to the true value with T for fixed N, but not with N for a fixed T. This is a problem, because in panel data T tends

This problem, however, goes away as the time dimension goes to infinity *for every plant*. For four industries, a balanced panel can be constructed, that is a panel in which at least 30 plants are observed in every year over the 16 year time period.[14] This time dimension is viewed as sufficiently large to allow for the estimation of a fixed effects model with a lagged endogenous variable.

I estimate:

$$x_{it} - x_{i.} = \rho(x_{it-1} - x_{i.}) + \mu_{it},$$

where $x_{it} = tfp_{it} - tfp_{.t}$ for the four industries, via OLS. The results of this procedure are in the fourth column of Table 3.

An alternative methodology is to construct dummy variables for each plant and to pre-multiply both the independent variable (x_{it}) and the dummy variables by H, where H'H equals the inverse of the variance covariance matrix of the error term, which is a function of D. Then one searches for the D that minimizes the sum of squared residuals resulting from the OLS regression on the transformed variables. The results of the search procedure are

---

to be small.

[14]Balancing the panel, of course, introduces the possibility of sample selection bias. Dwyer (1995b) compares results from a "robust method of performing analysis of variance on an unbalanced panel with serial correlation and substantial reporting error" to the results of the conventional methodology executed on these balanced panels, and argues that they are remarkably similar.

in the fifth column of Table 3. In practice, the search procedure's estimate of D is bigger than the OLS estimate.

It is shown in Appendix II that

$$E(ESS1) = T(N-1)(\sigma_a^2+\sigma_e^2)\text{; } and$$

$$E(ESS2)=\left(TN-T-N+1-\left(\frac{N-1}{T}\right)\left(\frac{2(T-1)\rho-\rho^{T+1}}{(1-\rho)^2}\right)\right)\sigma_e^2,$$

where ESS1 and ESS2 are the sum of square residuals, when one regresses tfp_{it} onto time dummies and time and plant dummies, respectively. The term $(2(T-1)D+D^{T+1})/((1-D)^2)$ corrects for the fact that serial correlation "looks" like a fixed effect in a panel with a short time dimension. Columns 2, 3 and 7 of Table 3 use these two equations and the OLS estimate of D to compute $F^2_{,}$, F^2_a, and the percentage of variation explained by the fixed effects $(100(F^2_a/(F^2_{,}+F^2_a))$ for each of these four industries. The percentage of the variation explained by fixed effects is about one-third. While I compute these numbers using an estimate of D, rather than its true value, the estimates are not sensitive to one standard deviation changes in D. Therefore using the estimate of D rather than its true value does not appear to introduce a substantial bias into the estimates of $F^2_{,}$, F^2_a, and %fixed.

One can test the hypothesis that $a_i = a_j$ for all i and j, via a chi-square test. One transforms the dependent variable and the dummy variables according to the transformation in the search procedure. Then one computes the difference between the sum of squared residuals of the transformed independent variable regressed on a transformed constant variable, and the sum of squared residuals when the transformed independent variable is regressed on the transformed dummy variables. The data is, of course, transformed with the D that minimizes the sum of square residuals of the unconstrained model. This statistic converges in distribution (in T) to a O^2 distribution with degrees of freedom equal to the number of plants minus one.[15] The null hypothesis is conclusively rejected at any traditional level of significance for all four industries. One should be concerned, however, with invoking asymptotic properties in T, when working with a 16 year time dimension, given the potential for a downward bias in $\hat{D}$. Nevertheless, the p-value of the test only rises above .05 for large and unreasonable estimates of D.

Table 3: Analysis of Variance and the Autocorrelation Coefficient

---

[15]This is a special case of the GLM as exposited in Dhrymes, 1978, chapter 3.

SIC	F^2_a	$F^2_,$	D (OLS)	D̂ (search based)	Pvalue that $a_i = a_j$ for all i,j.	$100(F^2_a/(F^2_a+F^2_,))$	Number of Plants
2211	.047	.12	0.23 (.047)	.24	0	28	30
2221	.054	.097	0.25 (.030)	.27	0	36	73
2273	.156	.207	0.31 (.041)	.35	0	43	38
2283	.053	.123	0.34 (.026)	.38	0	30	86

Imposing a conventional fixed effects model, extended to allow for serial correlation in the error term, leads one to conclude that there are fixed effects in the data. That is, there are permanent quality differences across plants. Furthermore, these differences are large. A plant with a fixed effect that is one standard deviation above average expects to produce between 21% (SIC 2211) and 40% (SIC 2273) more output than the average plant, with the same inputs.

V. The Case for M-Mixing

This section presents a test of N-mixing. For a random variable to be N-mixing over-time, means that with the passage of time the distribution of the random sequence becomes independent of the initial condition. If there truly are fixed effects, plants do not N-mix. Formally, let $\{x_t\}$ be a stochastic process, where x is an element of a compact subset of the reals with a continuous density function.[16] Let M[a,b] be the F-algebra generated by possible realizations of $[x_a,...,x_b]$. Let P be the probability measure defined on M[0,4]. Then x_t is said to N-mix at a geometric rate if for all b > 0:

$$sup_{E_1 \in M[1,y]\ s.t.\ P(E_1)>0,\, E_2 \in M[y+b,\infty]} |P(E_2|E_1)-P(E_2)|<\Delta_b \phi^b$$

with $)_b$ finite and N < 1. This definition implies that for any * there exists a t such that:

$$sup_{x_0}|E(x_t|x_0)-E(x_t)|<\delta,$$

Testing this condition, however, may require an unreasonably long time period.

A more powerful implication of N-mixing at a geometric rate is:

$$p_{x_{i0}}\{|E(x_{it+b+1}|x_{it+b},...,x_{it}) - E(x_{it+b+1}|x_{it+b},...,x_{it},x_{i0})|\} \le \Delta_b ¢$$

with $)_b$ finite (for all b) and N < 1. Intuitively, the expectation

[16] These are sufficient conditions to ensure that the expectations will always exist.

conditional on the lags becomes arbitrarily close to the expectation conditional on both the lags and the initial value, as the number of lags becomes large (for details see Pakes and Ericson, 1995). Another powerful implication of N-mixing at a geometric rate is:

$$sup_{x_{i0}}\{|E(x_{it+b+1}|x_{it+b},...,x_{it}) - E(x_{it+b+1}|x_{it+b},...,x_{it},x_{i0})|\}\le\Delta_t\phi^t,$$

with $)_t$ finite (for all t) and N < 1.

This implies that for N-mixing, the observations close to the current year may be informative, but as the observations move farther away from the current year they contain less information. This property clearly does not hold for a fixed effects model.

Test Procedure

The objective is to test the N-mixing hypothesis; that is, two plants with the same recent history have the same expected value of x_{it} even though their initial values of x_{i0} differ. Under the hypothesis of N-mixing, this independence property becomes stronger as the distance between t and 0 becomes bigger and as the number of lags become larger. I define $x_{it} = tfp_{itj} - tfp_{.tj}$, where j denotes that four digit industry. Note that x_{it} is a standardized tfp; if $x_{it} = .35$ then plant i is 35% more productive than the average plant in its industry in that time period.

The alternative hypothesis is that x_{it} is not N-mixing. It is illustrative, however, to consider a special case of the alternative hypothesis. Suppose the x_{it} is generated according to the fixed effects model with serial correlation, as estimated in the previous section. For fixed effects, as the number of lagged values becomes large our estimate of the permanent parameter converges to its true value and therefore the information content of x_{i0} becomes negligible. If we fix the number of lags, however, the information content of x_{i0} is non-decreasing in the distance between 0 and t; if there is serial correlation, then the further away the initial observation is from the current observation the more independent the error term becomes and consequently the observation becomes more informative. Under fixed effects, x_{i0} always provides information on the expectation of x_{it} given a fixed number of lags, regardless of the distance between t and 0.

I start with observations of tfp at the plant level for all of textiles. I select the plants that are observed in the same industry in 1987, the initial year and for all of the lagged years.

Define q_1, q_2, and q_3 such that

$\text{prob}(tfp_0 < q_1) = \text{prob}(q_1<tfp_0<q_2) = \text{prob}(q_2<tfp_0<q_3) = \text{prob}(q_3<tfp_0) = .25$.

Define the function $F(x_{it})$: ú 6 {1,2,3,4} such that

$Q_{it} =$ 1 if $tfp_{it} < q_1$,

2 if $q_1 < tfp_{it} < q_2$,

3 if $q_2 < tfp_{it} < q_3$,

4 if $q_3 < tfp_{it}$.

Note that this function maps x_{it} into ***quartiles*** only in the initial year. If t … 0, then more or less than 25% of the plants can end up in one ***quartile***.

Define

$I_{ijt} =$ 1 if $Q_{it} = j$,

0 otherwise.

That is, $I_{ijt} = 1$ if plant i was in cell j in year t and 0 otherwise. In order to test that the expectation of tfp_{i87} given tfp_{i86} and tfp_{i0} is equal to the expectation of tfp_{i87} given tfp_{i86} (vs. the alternative hypothesis that it is not) I estimate the parameters of the equation below via OLS:

$$tfp_{i87} = \sum_{j=1}^{4} \beta_j \; I_{ij86} + \sum_{j=1}^{4} \sum_{k=2}^{4} \beta_{jk} \; I_{ij86} \; I_{ik0},$$

and test the null hypothesis that $\$_{jk} = 0$ for all j and k (vs. the alternative that $\$_{jk}$ does not equal 0 for some j and k). The interpretation of $\$_{jk}$ is the difference between the expectation of x_{it} given $q_{i86} = j$ and $q_{i0} = 1$ and the expectation of x_{it} given $q_{i86} = j$ and $q_{i0} = k$. Therefore, the null is N-mixing and the alternative is that it is not.

The extension to additional lags is straightforward. For two lags I estimate:

$$tfp_{i87} = \sum_{j=1}^{4} \beta_j \; I_{ij86} + \sum_{j=1}^{4}\sum_{k=2}^{4} \beta_{jk} \; I_{ij86} \; I_{ik85} + \sum_{j=1}^{4}\sum_{k=1}^{4}\sum_{l=2}^{4} \beta_{jkl} \; I_{ij86} I_{ik85} I_{il0}$$

and test the hypothesis that $\$_{jkl} = 0$ for all j, k, and l. For three lags, I estimate:

$$I_{ij86} + \sum_{j=1}^{4}\sum_{k=2}^{4} \beta_{jk} \; I_{ij86} \; I_{ik85} + \sum_{j=1}^{4}\sum_{k=1}^{4}\sum_{l=2}^{4} \beta_{jkl} \; I_{ij86} I_{ik85} I_{il84} + \sum_{j=1}^{4}\sum_{k=1}^{4}\sum_{l=1}^{4}\sum_{m=2}^{4} \beta_{jklm} I_{ij8}$$

and test the hypothesis that $\$_{jklm} = 0$ for all j, k, l and m. When one of the independent variables is always zero, i.e., when a cell is empty, it is dropped from the regression.

I estimate this regression for all possible initial years (for one lag the initial year ranges from 72 to 85, for two lags it ranges from 72 to 84, and so forth) via OLS.[17] For one lag, the null hypothesis (of N-mixing) can only be rejected, with 90 percent confidence, when the initial year is greater than 75 (Table 4). For two lags the null hypothesis can only be rejected when the initial year is after 1982 (Table 5). For three lags, the null

[17]It is certainly possible that the variance of x differs across cells, which would imply that an F-test is not valid. This problem can be overcome via an application of GLM; divide the dependent and independent variables by the standard deviation of the cell and run OLS on the transformed data. Now an F-test is asymptotically valid. In fact, the corresponding chi-squared test is equivalent to the Pakes and Ericson test. I have executed this procedure and the results are substantively the same.

hypothesis is only rejected in 1981 (Table 6). The fact that the null hypothesis is rejected only as the initial year moves closer to 1987 strongly suggests N-mixing; N-mixing says that recent observations may matter, but that the more distant the observation the less information it provides.

Table 4: Test of tfp_{87} Being Independent of tfp_0 Given tfp_{86}

$year_0$	R^2ur	R^2r	DFN	DFD	Ftest	Pvalue
72	0.254	0.243	12	763	0.923	0.522
73	0.263	0.250	12	675	1.032	0.416
74	0.262	0.245	12	693	1.295	0.216
75	0.270	0.260	12	709	0.818	0.631
76	0.256	0.237	12	729	1.566	0.096
77	0.269	0.245	12	908	2.476	0.003
78	0.276	0.259	12	741	1.454	0.136
79	0.264	0.237	12	677	2.062	0.017
80	0.261	0.228	12	692	2.557	0.002
81	0.244	0.224	12	714	1.630	0.078
82	0.273	0.250	12	1170	3.117	0.000
83	0.254	0.224	12	799	2.757	0.001
84	0.289	0.260	12	1142	3.931	0.000
85	0.286	0.251	12	1191	4.829	0.000

Table 5: Test of tfp_{87} Being Independent of tfp_0 Given tfp_{86} and tfp_{85}

$year_0$	R^2ur	R^2r	DFN	DFD	Ftest	Pvalue

72	0.323	0.288	47	665	0.730	0.910
73	0.318	0.286	47	582	0.584	0.988
74	0.336	0.280	47	591	1.061	0.366
75	0.362	0.299	45	608	1.346	0.069
76	0.322	0.272	48	626	0.975	0.521
77	0.337	0.289	46	802	1.267	0.113
78	0.355	0.301	46	639	1.146	0.239
79	0.334	0.270	45	574	1.232	0.147
80	0.329	0.275	45	589	1.057	0.374
81	0.335	0.266	48	605	1.300	0.089
82	0.317	0.280	48	1057	1.197	0.170
83	0.310	0.243	45	693	1.493	0.021
84	0.349	0.291	47	1076	2.070	0.000

Table 6: Test of tfp_{87} Being Independent of tfp_0 Given tfp_{86}, tfp_{85}, tfp_{84}.

$year_0$	R^2ur	R^2r	DFN	DFD	Ftest	Pvalue
72	0.470	0.369	125	524	0.798	0.937
73	0.445	0.340	120	450	0.714	0.986
74	0.476	0.357	107	468	0.995	0.500
75	0.476	0.374	109	486	0.865	0.821
76	0.466	0.339	122	490	0.954	0.616
77	0.477	0.365	118	658	1.196	0.092
78	0.481	0.359	115	504	1.029	0.409
79	0.439	0.331	106	455	0.822	0.888
80	0.426	0.337	97	478	0.767	0.944
81	0.482	0.333	106	485	1.314	0.029
82	0.419	0.334	141	879	0.912	0.749
83	0.421	0.321	105	578	0.952	0.613

In the previous section, I estimated parameters of a data generating process that allows for both fixed effects and serial correlation in plant level tfps. The variance of the fixed effects is statistically discernible from zero, which contradicts N-mixing. These contradictory results beg the question: what is the power of this test? For a fixed number of observations, the probability of accepting a wrong null hypothesis (N-mixing) will become arbitrary close to .95 (the probability of rejecting a correct null hypothesis) as the variance of the fixed effects becomes arbitrary small. One can ascertain the power of this test by asking the following question: if the parametric estimates of the fixed effects were the true data generating process behind plant productivity levels, what would be the probability of rejecting the null hypotheses of N-mixing under the above methodology?

I have generated ten Monte Carlo databases of 1000 plants over 16 years according to the parametric estimates of sic 2283 (carpets) from the previous section.[18] I have run the above tests for N-mixing on these data sets and report the percentage of times that the null hypothesis of N-mixing was accepted with 95% certainty in Table 7. One sees that the results are very different. For one lag, the null hypothesis is rejected for every initial year. For two lags, the null hypothesis is rejected about twenty percent of the time with the probability of rejecting being higher the earlier the initial year. For three lags, the null hypothesis is rejected about 90 percent of the time. If there were permanent plant effects of the magnitude measured, the pattern emerging from the data would be very

[18]The plant effects and the idiosyncratic error term were generated form a normally distributed random variable.

different. There is a key qualitative distinction between these results; for the real data, the null hypothesis of N-mixing was less likely to be rejected as the distance between the current year and the initial year increased, whereas for the Monte Carlo results it was more likely to be rejected.

It appears that the functional form of fixed effects with serial correlation is too restrictive to capture key features of the data. It is possible that a plant's productivity level is subject to large transitory shocks and that it occasionally retools, which is a permanent change to its productivity level. Since the fixed effects specification does not allow for these two different types of shocks to a plant's productivity level, one mistakes the latter for a fixed effect.

Table 7: Estimated Probability of a Type II Error[19]

$year_0$	P of Type II error with one lag	P of Type II error with two lags	P of Type II error with three lags

[19]Reports the percentage of times the null hypothesis was accepted with 95% confidence out of 10 trials.

72	0	0.2	0.9
73	0	0.2	0.7
74	0	0.3	1.0
75	0	0.1	0.9
76	0	0.3	1.0
77	0	0.1	0.8
78	0	0.2	0.9
79	0	0.1	0.9
80	0	0.2	1.0
81	0	0.2	0.9
82	0	0.2	0.9
83	0	0.3	1.0
84	0	0.5	
85	0		

VI. Estimates of a More Flexible Functional Form

The previous results suggest that plant level tfp (measured as the deviation from the industry mean) contain at least two components, one that is transitory and one that is persistent but not permanent. Expressing plant level tfps as the sum of two AR(1) processes provides a functional form rich enough to capture these two components:

$$tfp_{it} = v_t + a_{it} + \epsilon_{it};$$

where v_t is non-stochastic,

$$a_{it} = ra_{it-1} + \delta_{it};\ and\ \epsilon_{it} = \rho\epsilon_{it-1} + \mu_{it}.$$

Finally assume that * and : are independently distributed with means of 0 and variances of $F_a^2(1-r^2)$ and $F_,^2(1-D^2)$, respectively.

In this section, I shall compute a simple method of moments estimate of this process for each of the four balanced panels, as well as for the two digit textile industry as a whole.[20] Additionally, I shall compute these estimates for the corresponding unbalanced panels.

Once again, define

$$x_{it} = tfp_{it} - tfp_{.t},$$

which converges in probability to $tfp_{it} - v_t$. Let X be an NxT matrix where the $_{it}$th element is x_{it}.[21] Let MM = X'X/(N-1) be the sample moment matrix (TxT). This matrix has T^2 sample moments, ((T+1)T)/2 of which are distinct. Any model implies an expectation of these moments as a function of the parameters. If the number of parameters is less than ((T+1)T)/2, the model is overidentified. The idea of method of moments estimation is to choose the parameter values of the model to make the population moments as *close* as possible to the sample moments. This method yields estimates that are consistent in N, provided the model is correctly specified. That is, for a fixed T, as the number of plants goes to infinity the sample moments will converge in probability to their true values. Consequently, the parameter

[20] In the case of the textile industry as a whole, I measure productivity as a deviation from the four digit industry mean.

[21] I am adopting the convention that a y_{ij} represents the $_{ij}$ element of the matrix Y.

estimates will converge in probability to their true values. In panel data, T is typically small while N is typically large. Consistency in N, therefore, is a desirable property.

Let

$$PMM = E\left(\frac{XX'}{N-1}\right) = \begin{bmatrix} \sigma_a^2+\sigma_\varepsilon^2 & r\sigma_a^2+\rho\sigma_\varepsilon^2 & \cdots & (r^{T-1}\sigma_a^2+\rho^{T-1}\sigma_\varepsilon^2) \\ r\sigma_a^2+\rho\sigma_\varepsilon^2 & \sigma_a^2+\sigma_\varepsilon^2 & \cdots & \vdots \\ \vdots & \vdots & & \\ (r^{T-1}\sigma_a^2+\rho^{T-1}\sigma_\varepsilon^2) & \cdots & & \sigma_a^2+\sigma_\varepsilon^2 \end{bmatrix},$$

i.e., the population moment matrix given by the above model. In order to set up the minimization problem, it is useful to vectorize these matrices dropping the redundant elements. Let

$MV = (mm_{11}, mm_{12}, \ldots, mm_{1T}, mm_{22}, \ldots, mm_{2T}, mm_{33}, \ldots mm_{TT})'$

that is, the moment vector. Let

$PMV = (pmm_{11}, pmm_{12}, \ldots, pmm_{1T}, pmm_{22}, \ldots, pmm_{2T}, pmm_{33}, \ldots, pmm_{TT})'$

that is, the population moment vector.

Let $2 = (F^2_a\ r\ F^2_,\ D)'$, that is, the parameter vector, and 2_0 be its true value.

And finally let

$$g(MV, 2) = MV - PMV(2),$$

i.e., the so called orthogonality conditions, because g_i has an expectation of zero. Note that there are T(T+1)/2 such conditions.

A consistent estimate of 2_0 solves:

$$\min_{[\theta]} G \equiv g' I g,$$

where I is a T(T+1)/2 by T(T+1)/2 identity matrix. Define $\hat{2}$ to be the 2 that solves this problem, i.e., the estimate of 2_0, which is a function of MV. The derivatives of g'Ig are linear in F^2_a and $F^2_,$ and non-linear in r and D. Therefore, for a given (r,D) minimizing g'Ig with respect to F^2_a and $F^2_,$ merely involves solving two linear equations in two unknowns, which suggests a procedure for approximating the solution. I run a grid search over possible values of r and D, computing

$$\min_{[\sigma^2_a, \sigma^2_e]} g' I g$$

for each pair of r and D and choose the pair that yields the lowest min[F^2_a,$F^2_,$] g'Ig. In practice, I execute this by running a grid search over a 100 by 100 grid.[22] From a computational

[22] r' O {0.005, 0.015, 0.025, ..., 0.995} and D' O {0, 0.01, 0.02,...,0.99}. I set up the grid to avoid r' equaling D', because the values of F^2_a and $F^2_,$ that minimize g'Ig are not unique if r' = D'. I then take the largest and smallest values of r' and D' to be r and D, respectively. It should be clear from the previous results that both D and r will be positive. This prior was confirmed by running a search over a coarser grid that allowed D and r to range between -1 and 1; they consistently

perspective, the surface of G is well behaved. It has one local maximum, and small changes in the maximization algorithm do not lead to large changes in the estimates. The results for the balanced and unbalanced panels are reported in Tables 8&9.

In order to convert these simple method of moment estimates into "general method of moments estimates," one would estimate the expectation of gg' on basis of the estimates of 2 in Table 8. The inverse of this matrix would then become the *optimal weighting matrix*, and one typically iterates until the parameter estimates converge (cf. Hamilton, Chapter 14, 1995). These estimates would then be efficient, given the moments. Unfortunately, working out an analytical expression for the expectation of gg' is not practical. For the balanced panel only, I can estimate it with Monte Carlo techniques (while not impossible, estimating it for the unbalanced panels is computationally rather intensive).

If the precision of consistent estimates is high, then the gain in obtaining efficient estimates is marginal. But in order to make this claim, one needs measures of the precision of the estimates, i.e., asymptotic standard errors. Asymptotic standard errors are computed for the balanced panels as follows. First, I use a first order Taylor expansion and implicit differentiation to write the parameter estimates as a linear function of the

---

came up positive.

sample moment vector. Second, I use Monte Carlo techniques to estimate the variance covariance matrix of the sample moment vector for the balanced panel. Finally, I transform this matrix on basis of the linear approximation to obtain an asymptotic variance covariance matrix of the parameter estimates. The asymptotic standard errors of the estimates are in parenthesis in

Table 8. For the complete details of this technique see Appendix III. I have not computed the standard errors for the unbalanced panel's parameter estimates due to the computational burden.

Table 8: Method of Moment Estimates (Balanced Panels)

SIC	F^2_a	r	$F^2_,$	D	$100(F^2_a/F^2_a+F^2_,)$	half life	number of plants

2211	.0688 (.0213)	.94 (.0605)	.099 (.025)	.225 (.456)	41	11.2	30
2221	.064 (.013)	.97 (.0307)	.0868 (.016)	.345 (.189)	42	22.7	73
2273	.2105 (.055)	.96 (.046)	.1595 (.066)	.305 (.109)	56	17.0	38
2283	.0703 (.014)	.96 (.032)	.1079 (.016)	.425 (.15)	39	17.0	86
22	.0977 (.0057)	.935 (.0122)	.1401 (.0066)	.30 (.041)	41	10.3	631

Table 9: Method of Moment Estimates (Unbalanced Panels)

SIC	σ_a^2	ρ	σ_ϵ^2	Δ	$100(\sigma_a^2/(\sigma_a^2+\sigma_\epsilon^2))$	half life	average number of plants[23]

[23]This is the mean number of plants used in computing each sample moment.

2211	.0823	.935	.2025	.30	28	10.3	100
2221	.0895	.995	.0446	.41	67	138	198
2273	.2135	.945	.3843	.40	36	12.8	110
2283	.08709	.92	.2055	.285	30	8.3	229
22	.1918	.935	.2486	.275	43	10.3	1700

As r approaches 1, for a finite F_a^2, this model becomes the fixed effects model estimated in Section III. Indeed, the estimated value of r is close 1 in all cases and the rest of the parameter estimates are similar to those estimated in Section III. In all the balanced panels, nevertheless, an r of less than .995 fits the data better than .995, which is largest possible value in the grid search. At the four digit industry level (where the number of observations is small) r is about one standard error away from one. When looking across all of textiles (where the number of observations is large) the estimate of r is about six standard errors away from one and therefore highly significant. These estimates are consistent with the finding of N-mixing; there is a highly persistent component to a plant's tfp, but it is not fixed. The sixth column of Table 8 presents the percentage of cross-sectional variation explained by the persistent component, which ranges from 39 percent to 56 percent. Note that these numbers are consistently larger than the corresponding numbers in Table 3. This demonstrates that by requiring the persistent component to be fixed, one understates its importance.

The picture that emerges for the unbalanced panel is somewhat different. The magnitude of the transitory shocks is considerably larger, except in SIC 2221. It is likely that this is the result of outliers, which are more of a problem in the unbalanced panel. The Census edits the responses of large plant more intensely; large plants are over-represented in the balanced panel, because small plants are sampled intermittently. The magnitudes of the persistent component, however, are similar. The degree of persistence is also substantively the same, except in SIC 2221, where r took on the largest possible value, r = .995. The estimates of r and $F^2_,$ in 2221 are both problematic; it is possible that the estimates for this industry are outlier dominated.

Note that one component of tfp is indeed highly transitory while the other is highly persistent in all four industries. Consistent estimates of the half life of a_{it} can be found by solving for J:

$$\frac{E(a_{it}|a_{it-\tau})}{a_{it-\tau}} = r^{\tau} = \frac{1}{2},$$

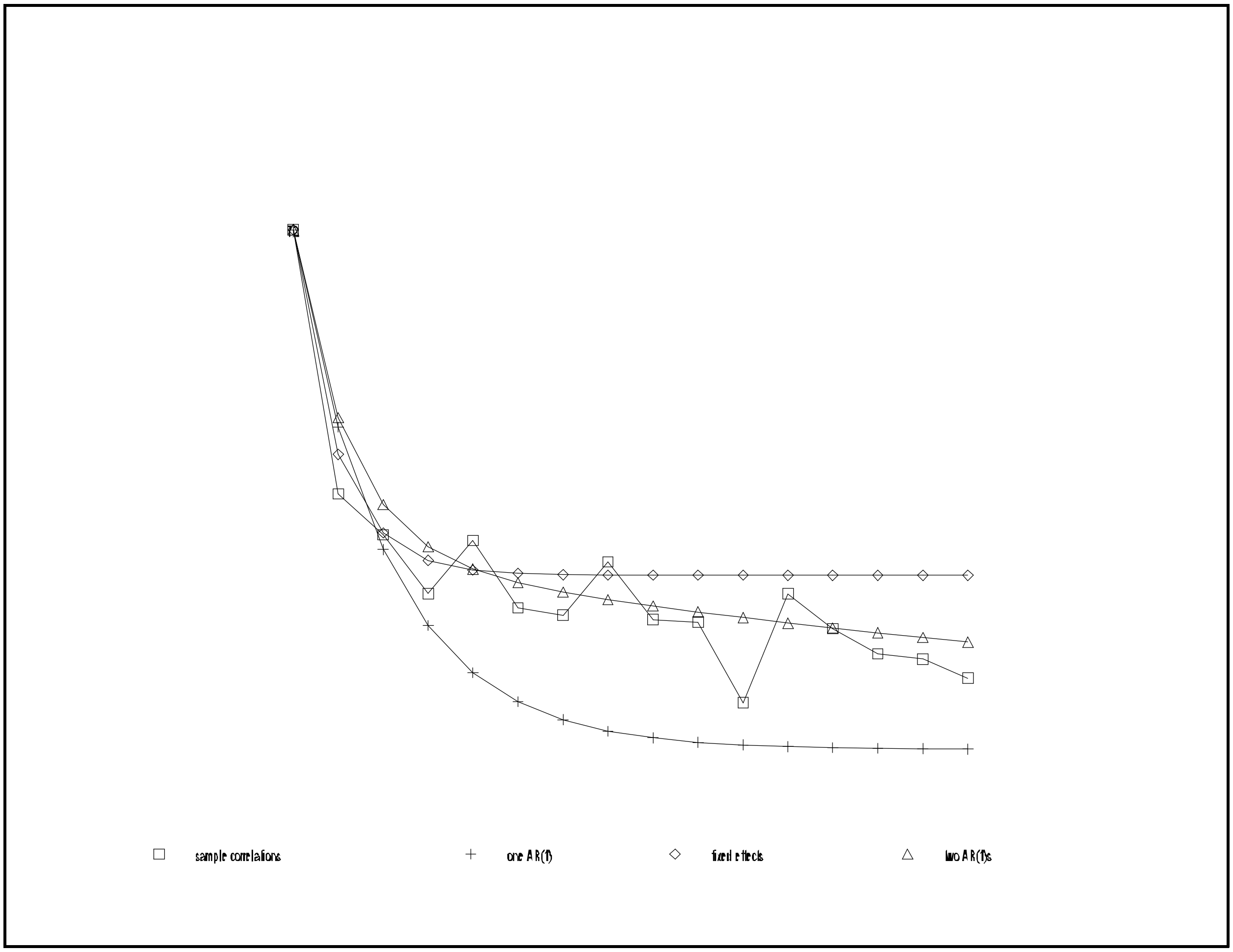
given the estimate of r. These half lives are reported in the fifth column of Table 8. In the Carpet Industry (SIC 2273), if the persistent component of a given plant is 50 percent above average today (about one standard deviation) then it expects to be 25 percent above average 15 years from now.

This paper's results are summarized in Figure 1. The squares plot the sample correlation coefficients between tfp_{i72} and tfp_{ix}, where x ranges from 72 to 87 for SIC 2283.[24] The plus symbols plot the predicted correlation coefficients from estimates of a simple AR(1) process, i.e., estimating $x_{it} = Dx_{it-1} + ,_{it}$ via OLS. Clearly, there is too much persistence in the data for this simple model. The diamonds plot the correlation coefficients predicted by the fixed effects model estimated in Section IV. This model predicts that the correlation coefficient will asymptote a positive value that appears too large to be consistent with the data. The triangles plot the serial correlation coefficients predicted by the model estimated in this section, which fits the data well. It predicts that the correlation coefficient will rapidly fall off at first and then gradually asymptote 0. It predicts that the plant's productivity level eventually becomes independent of its initial value; plant

[24]The corresponding figures for the other industries are substantively the same, but noisier. This is to be expected, given that the balanced panel for SIC 2283 has the largest number of observations. Furthermore, corresponding graphs for the unbalanced data sets are similar.

productivity levels N-mix.

Figure #1



VII. Econometric and Economic Implications

We have long known that there is substantial persistence in plant productivity levels (cf. Dhrymes, 1991; Bartelsman Dhrymes, 1991; Baily, Hulten, and Campbell, 1992). Dwyer (1995b) demonstrates that when ranking plants according to their productivity 12 years ago, the 85th percentile plant is as much as 20% more productive than the 15th percentile plant, today. Is this persistence permanent? When imposing a conventional fixed effects model on the data, extended to allow for serial correlation in the error term, one conclusively concludes that there are fixed effects in the data. That is, the parametric methodology leads one to the conclusion that there are permanent differences in plant productivity levels, consistent with a passive learning model of plant dynamics.

A non-parametric test, however, reveals that a plant's productivity in year t becomes independent of its productivity in year 0 as t becomes large, i.e., plant productivity levels N-mix, consistent with an active exploration model of plant dynamics. This non-parametric conclusion contradicts the parametric conclusion. Monte Carlo results suggest that the functional form imposed on the data by the fixed effects model is too rigid; if there were fixed effects of the magnitude measured, the testing procedure would have rejected the hypothesis of N-mixing. If this phenomenon is true across many industries, then the "fixed-effects solution" to the simultaneity problem will become more problematic as panels become longer. There are at least two approaches to solve this problem. First, one could work with a rolling panel. Second, one could develop a *quasi-fixed* effects model, that is, estimating some sort of *moving average intercept*. I am skeptical about the value of either approach. If one is really interested in solving the simultaneity problem, one needs to "find (instrumental) variables that have genuine information about factors which affect firms differentially as they choose their input levels" (Griliches and Mairesse, 1995, page 23).

The parametric conclusion of fixed effects leads to different economic conclusions than the non-parametric conclusion of N-mixing. The parametric results suggest a passive learning model whereas, the finding of N-mixing suggests a model of active exploration (Pakes and Ericson, 1995). More recent theoretical papers emphasize the option value of an existing plant (Dixit, 1992; and Campbell, 1994). It is argued that a plant with marginal cash flows has a positive value, because there is a possibility of it becoming highly productive and its decision to exit is irreversible. The parametric results suggest that this option value would be small, because the differences in plant productivity levels are fixed, except for a transitory shock that is short lived. Whereas, the non-parametric results suggest that there are both transitory and persistent changes to a plant's productivity level and the option value associated with the persistent changes is potentially large.

A method of moments estimate of a more flexible functional form suggests that the *quasi fixed* effects erode slowly; if your manager is 50 percent more productive than average today, 15 years from now you expect him to be 25 percent above average.[25] This result begs two questions: is this persistent component mostly measurement error? if not, why are the market forces in the textile industry -- a competitive industry -- working so slowly?

There are a number of reasons to doubt that this persistence is the product of persistent measurement error alone. First, the same amount of persistence is observed in labor productivity, which should have less measurement error. Second, most explanations of measurement error in value added will imply that plants with high measured productivity would have a low material to sales ratio. For example, suppose a plant sets the intercompany transfer price of its product too high, then it will overstate its value added and productivity will be overestimated. Furthermore, the material to sales ratio will be below the industry average, *ceteris paribus*. Some work in progress reveals that plants that are measured as highly productive today do indeed have low material to sales ratios. This correlation, however, is not persistent. Therefore, measurement error that is identified by the material to sales ratio does not explain persistence in tfp.

If one rules out persistent measurement error, one is left

[25]More precisely, if your productivity today (after filtering out transitory shocks) is 50 percent above average, then you expect to be 25 percent of above average 15 years from now.

to speculate on why the market forces in a competitive industry seem to be working so slowly. Recently, there has been interest in models in which economic agents optimally chooses to periodically adopt a new technology as their current technology becomes obsolete (Parente, 1994; Dwyer, 1995a). A 10 to 20 year half life in the persistent component suggests a rather long retooling cycle. Nevertheless, there is evidence that retooling cycles are long. This is true in textiles -- diffusion of the shuttleless loom has taken over 20 years (MIT Commission on Industrial Productivity, 1989, chapter 4, page 25) -- as well as in other industries. Diffusion of the Diesel Locomotive took over forty years (Jovanovic and McDonald, 1994). Recent surveys of 61 paper mills report 1973 as the average date of the last major rebuild (Upton, 1995). It may be that the fixed costs of adoption are big, which results in a long retooling cycle.

Vintage human capital, or a *quasi-fixed* managerial effect is certainly another possibility. In the steel industry, Ichniowski and Shaw (1995) found that "through 1992, very few of these older lines (started in the 50s and 60s) made any changes to their traditional (management) practices." The firms that did make changes did so "during times of threatened job loss when new managers are brought in to make large-scale changes in the work environment." Therefore, it seems likely that vintage human capital, as modelled in Chari and Hopenhayn (1991), explains some of the persistence in productivity differentials.

REFERENCE

Atkeson, Andrew, Aubhik Khan, and Lee Ohanian, "Are Data on Industry Evolution and Gross Job Turnover Relevant for Macroeconomics?" University of Pennsylvania, June 1995.

Baily, Martin N., Charles Hulten, and David Campbell, "Productivity Dynamics in Manufacturing Plants," *Brookings Papers: Microeconomics*, pages 187-267, 1992.

Bartelsman, Eric J., *Three Essays on Productivity Growth,* 1992, Doctoral Dissertation, Columbia University.

Bartelsman, Eric J., and Phoebus J. Dhrymes, "Productivity Dynamics: US Manufacturing Plants, 1972-1986," Columbia University, Department of Economics Discussion Paper Series No. 584, September 1991.

Caballero, Ricardo J., and Mohammed L. Hammour, "The Cleansing Effect of Recessions," *American Economic Review*, pages 1350-1369, December 1994.

Campbell, Jeffrey R., "Entry, Exit, Technology and Business Cycles," University of Rochester, 1994.

Chari, V.V., and Hugo Hopenhayn, "Vintage Human Capital, Growth, and the Diffusion of New Technology," *Journal of Political Economy,* Vol. 99, no. 6, 1991.

Cooley, Thomas, F., Jeremy Greenwood, and Mehmet Yorukoglu, "The Replacement Problem," University of Rochester, July 1995.

Davis, Steven and John Haltiwanger, "Gross Job Creation, Gross Job Destruction and Employment," *The Quarterly Journal of Economics,* August 1992.

Dhrymes, Phoebus, "The Structure of Production Technology: Productivity and Aggregation Effects," Columbia University, Department of Economics Discussion Paper Series No. 551, June 1991.

Dhrymes, Phoebus, *Introductory Econometrics*, 1978, Springer Verlag, New York.

Dixit, Avinash, "Competitive Industry Equilibrium with Firm-Specific Uncertainty," Princeton University, 1992.

Dwyer, Douglas W., "Disaggregate Production Functions," Columbia University, 1994.

Dwyer, Douglas W., "Technology Locks, Creative Destruction and Non-Convergence in Productivity Levels" Center for Economic Studies Discussion Paper, CES 95-6, 1995a (available via ftp.census.gov/pub/ces/wp/95-6) a more recent version is available from the author upon request.

Dwyer, Douglas W., "Whittling Away At Productivity Dispersion," Center for Economic Studies Discussion Paper, Center for Economic Studies Discussion Paper, CES 95-5, March 1995b (available via ftp.census.gov/pub/ces/wp/95-5) a more recent version is available from the author upon request.

Ericson, Richard, and Ariel Pakes, "Markov Perfect Industry Dynamics: A Framework For Empirical Work," *Review of Economic Studies,* Vol. 62, no. 1, pages 53-84, January, 1995.

Gort, Michael, Byong H. Bahk, and Richard A. Wall, "Decomposing Technical Change," *Southern Economic Journal*, Vol. 62, pages 220-234, July, 1993.

Gray, Wayne B., "Productivity Database," Clark University, 1989 (This database and its documentation is available on line; see http://nber.harvard.edu for information and ftp://nber.harvard.edu:/pub/productivity for the data.).

Green, William H., *Econometric Analysis*, second edition, 1993, Prentice Hall, New Jersey.

Griliches, Zvi, and Jacques Mairesse, "Production Functions: The Search for Identification," NBER working paper no. 5067, 1995.

Griliches, Zvi, "Economic Data Issues," Zvi Griliches and M.D. Intriligator, eds. *Handbook of Econometrics, Volume III*, Elsevieur Science Publishers, 1986.

Hamilton, James, *Time Series Analysis*, Princeton University Press, 1994.

Ichniowski, Casey and Kathryn Shaw, "Old Dogs and New Tricks: Determinants of the Adoption of Productivity-Enhancing Work Practices," *Brookings Papers: Microeconomics*, pages 1-65, 1995.

Irwin, Douglas and Pete Klenow, "High Tech and R&D Subsidies: Estimating the Effects of Sematech," NBER working paper no. 4974, 1994.

Mairesse, Jacques and Zvi Griliches, "Heterogeneity in Panel Data: Are There Stable Production Functions" in P. Champsauer et. al. eds., *Essays in Honor of Edmond Malinvaud, Volume 3: Empirical Economics*, MIT Press, 1990, pages 192-231.

Jovanovic, Boyan, "Selection and the Evolution of Industry," *Econometrica,* Vol. 50, no. 3, pages 649-670, May 1982.

Jovanovic, Boyan and Glenn M. MacDonald, "Competitive Diffusion," *Journal of Political Economy*, Vol. 102, no. 1, pages 24-52.

Jones, Derek C. and Takao Kato "The Productivity Effects of Employee Stock-Ownership Plans and Bonuses: Evidence from Japanese Panel Data," *The American Economic Review*, Vol. 85, no. 3, pages 391-415, June, 1995.

Jorgenson, Dale, Frank Gollop and Barbara Fraumeni, *Productivity and U.S. Economic Growth,* Harvard University Press, Cambridge, Massachusetts, 1987.

McGuckin, Robert H. and George A. Pascoe, "The Longitudinal Research Database (LRD): Status and Research Possibilities," *Survey of Current Business*, 1988.

MIT Commission on Industrial Productivity, "The Working Papers of the MIT Commission of Industrial Productivity," Volume 2, 1989, The MIT Press, Cambridge, Massachusetts.

Mueller, Dennis C., *Profits in the Long Run*, 1986, Cambridge University Press, Cambridge.

Olley, Steven, G., and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," Center for Economic Studies Discussion Paper, CES 92-2, February 1992.

Pakes, Ariel, "Mueller's Profits in the Long Run," *Rand Economic Journal*, Vol. 18, no. 2, pages 319-332, Summer 1987.

Pakes, Ariel and Richard Ericson, "Empirical Implications of Alternative Models of Firm Dynamics," 1995, *Journal of Economic Theory* (forthcoming).

Parente, Stephen L., "Technology Adoption, Learning-by-Doing and Economic Growth," *Journal of Economic Theory*, Vol 63, pages 346-369, 1994.

Roberts, Mark J., and Dylan Supina, "The Magnitude and Persistence of Output Price Dispersion For U.S. Manufactured Products," the Pennsylvania State University, December 1994.

Rumuelt, Richard P. "How Much Does Industry Matter?" *Strategic Management Journal,* Vol. 12, pages 167-185, 1991.

Solow, Robert M., "Technical Change and the Aggregate Production Function," *The Review of Economics and Statistics,* Vol. 39, pages 312-320, August 1957.

Upton, David, "What Really Makes Factories Flexible," *Harvard Business Review*, pages 74-87, July-August 1995.

Wang, Duan, "Irreversibility and Corporate Investment," Columbia University, 1994.

Appendix I: Analysis of Variance on a Balanced Panel with Serial Correlation

In panel data, as the degree of serial correlation approaches a unit root the data will look like it has fixed effects, even in the absence of fixed effects. Here, I develop unbiased estimates of the variances of the fixed effects and the error term, assuming the that error term follows an AR1 process, which is known, for a balanced panel.

Let tfp_{it} be NxT observations (N and T are the number of plants and years, respectively.) generated by:

$$tfp_{it} = v_t + a_i + \epsilon_{it};\ \epsilon_{it} = \rho\epsilon_{it}-1 + \mu_{it},$$

where v_t are deterministic time trends, a_i are non-stochastic with a mean of zero and a of variance of σ_a^2[26], and μ_{it} are independently drawn from a distribution whose mean is zero and variance is σ_μ^2. Define

$$\sigma_\epsilon^2 = E(\epsilon_{it}^2) = \frac{1}{(1-\rho^2)}\sigma_\mu^2.$$

Let $x_{\cdot t}$ and $x_{i\cdot}$ denote the mean of x in a given year and across time, respectively.

[26]Alternatively, a_i could be independently drawn from a distribution with a mean of zero and a variance of σ_a^2.

Let

$$ESS1=E\left(\sum_t \sum_i \left(tfp_{it}-tfp._t\right)^2\right)$$
$$=E\left(\sum_t \sum_i tfp_{it}^2 - \sum_t \sum_i tfp._t^2\right)$$
$$=N\sum_t v_t^2+TN(\sigma_a^2+\sigma_e^2)-N\sum_t v_t^2 - \frac{1}{N^2} E\sum_t \sum_i ((\sum_i a_i)^2+(\sum_i e_{it})^2)$$
$$= T(N-1)(\sigma_a^2+\sigma_e^2).$$

Let

$$x_{it} = tfp_{it} - tfp._t.$$

Define

$$ESS2 = E\left(\sum_i \sum_t \left(x_{it} - x_i.\right)^2\right)$$
$$=E\left(\sum_i \sum_t x^2 - 2\sum_i \sum_t x_{it}\, x_i. + \sum_i \sum_t x_i.^2\right).$$

Here, the first term is just ESS1. In order to evaluate the next, two terms observe that:

$$x_i = \left(tfp_{it} - tfp._t\right)$$
$$=v_t + a_i + e_{it} - \left(\frac{1}{N}\right)\sum_i \left(v_t+a_i+e_{it}\right)$$
$$=a_i + e_{it} - \left(\frac{1}{N}\right)\sum_i \left(a_i+e_{it}\right).$$

Therefore, the second two terms in the ESS2 expression can be evaluated as follows:

$$\sum_i \sum_t (x_{it}\, x_{i\cdot} - x_{i\cdot}^{2})\Bigg) = E \sum_i \sum_t (x_{i\cdot})^2$$

$$= \frac{N}{T} E\left(\sum_t \left(a_i - \sum_i \frac{a_i}{N} + \epsilon_{it} - \sum_i \frac{\epsilon_{it}}{N} \right)\right)^2$$

$$= \frac{N}{T} E\left((\sum_t a_i)^2 + \left(\sum_t \sum_i \frac{a_i}{N}\right)^2 + (\sum_t \epsilon_i)^2 + \left(\sum_t \sum_i \frac{\epsilon_{it}}{N}\right.\right.$$

$$-2\frac{N}{T} E\left(\sum_t a_i \sum_t \sum_i \frac{a_i}{N} - \sum_t \epsilon_i \sum_t \sum_i \frac{\epsilon_{it}}{N} \right)$$

$$= TN\left(1 - \frac{1}{N}\right)\sigma_a^2 + \frac{N}{T}\left(1-\frac{1}{N}\right) E\,(\sum_t \epsilon_{it})^2.$$

Therefore,

$$ESS2 = T(N-1)\sigma_\epsilon^2 - \left(\frac{N-1}{T}\right) E\left(\left(\sum_t \epsilon_{it}\right)^2\right).$$

Finally,

$$E((\sum_t \epsilon_{it})^2) = T\sigma_\epsilon^2 + 2\sum_{k=1}^{T-1}\sum_{j=1}^{k} \rho^j \sigma_\epsilon^2$$

$$= T\sigma_\epsilon^2 + 2\sum_{k=1}^{T-1}\left(\frac{\rho - \rho^{k+1}}{1-\rho}\right)\sigma_\epsilon^2$$

$$= T\sigma_\epsilon^2 + 2(T-1)\frac{\rho}{1-\rho} - 2\left(\frac{\rho^2 - \rho^{T+1}}{(1-\rho)^2}\right).$$

Therefore,

$$ESS2 = \left(TN - T - N + 1 - \left(\frac{N-1}{T}\right)\left(\frac{2(T-1)\rho}{1-\rho} - \left(\frac{2\rho^2 - \rho^{T+1}}{(1-\rho)^2}\right)\right)\right)\sigma_\epsilon^2.$$

Now solving for F_a^2 and $F_{,}^2$ is just a matter of solving to equations and two unknowns.

Appendix II: Computing Asymptotic Standard Errors for the Simple Method of Moments Estimates.

First, in order to make the linear approximation, I need to differentiate 2 with respect to MV. This derivative can be obtained implicitly. Because 2̂ maximizes G at an interior of the parameter space (its an open space), $DG*_{2=\hat{2}}$ = (0,0,0,0)', where DG is the derivative of G with respect to 2. Noting that G is a function of both 2 and MV yields:

$$\frac{d(DG_i)}{dMV_j} = \frac{\partial DG_i}{\partial MV_j} + \frac{\partial DG_i}{\partial \hat{\theta}_i}\frac{\partial \hat{\theta}_i}{\partial MV_j} = 0,$$

for all i and j. Therefore, 2̂ can be approximated by:

$$\hat{\theta} \approx \theta_0 + H \cdot (MV - PMV),$$

where H is 4x((T+1)T/2) and

$$h_{ij} = \frac{\partial \hat{\theta}_i}{\partial MV_j} = -\frac{\dfrac{\partial DG_i}{\partial MV_j}}{\dfrac{\partial DG_i}{\partial \hat{\theta}_i}}.$$

This linear approximation provides a means for determining the limiting distribution, because it becomes arbitrarily accurate as 2̂ approaches 2 (provided 2̂ is continuous in MV).

In order to determine the variance covariance matrix of 2̂, we need the variance covariance matrix of MV. This can be obtained by Monte Carlo techniques for the balanced panel. I generated 100 industries according to the parameter estimates (I drew the underlying shocks from a normal distribution.); I constructed 100 MV; and I computed the variance covariance matrix of the these 100 random moment vectors. The variance-covariance matrix of 2 is HEH', where E is the estimated variance covariance matrix of MV. The standard errors are computed accordingly.

Appendix III: Data

My data set consists of the textile plants (SIC 2200-2299) in the Longitudinal Research Database (LRD), which is based on the Annual Survey of Manufactures (ASM) and the Census of Manufactures (CM).[27] The sample runs from 1972 until 1987.

The CM is carried out every five years (1967, 1972, 1977, 1982, and 1987) and each plant is, in principle, sampled with probability one. The ASM draws a sample of plants two years after the census, and then follows this sample for five years (these samples begin in 74, 79, and 84). It adds newly created plants to the sample every year. The sample probability is

[27]For a detailed description of this database see McGuckin and Pascoe (1988).

increasing in plant size.

My sample is a subset of a sample that includes all information available on every plant ever in the SIC codes 2200-2299 from 1967 to 1989. The sample is truncated to drop administrative record cases, which are small plants for which only a limited amount of information is collected, and drops pre-1972 and post-1987 observations. The pre-1972 observations were dropped in order to construct a complete time series and the post-1987 observations were dropped, because machine and capital retirements were not collected in 1988 or 1989. The regressions are ran separately for each four-digit SIC code, and therefore a plant was only included in the regression if it was in that textile industry. My unbalanced sample contains four years in which all firms are sampled with probability one (in theory), and three different samples in which large firms are sampled with a higher probability.

To resolve an apparent inconsistency in the classification of plants in census and non-census years the following aggregations are made: SIC 2258 includes DIND 2258 and 2292; SIC 2273 includes DIND 2271, 2272 and 2279; SIC 2283 includes DIND 2281, 2283 and 2284; SIC 2299 includes DIND 2291, 2293, 2294 and 2299 (DIND is the derived industry code). The relevant prices indices were computed as a Laspeyres price index with 1987 as a base year via Gray's productivity database with total value of shipments as the relevant weights (Gray, 1989).

Variable Construction:

RVA (Real value added)

Value added is computed as the total value of shipments plus changes in the value of inventories less the cost of materials (including materials, supplies, fuel, electric energy, cost of resales, and cost of contract work). Value added is deflated through Gray's shipments price index to generate RVA.

TE (Total employment)

Total employment is the sum of the average number of production workers and nonproduction workers.

BOOK (Gross book value of capital)

The only measure of assets that can be calculated consistently across small plants (which are intermittently sample) and large plants is book value. That is the book value of buildings and machinery at the end of the period plus the capitalized value of rental payments deflated by Gray's investment price index.

$$Assets_t = (BAE_t + MAE_t)/PINV_t + (BR_t+MR_t)/(r_tPINV_t).$$

Here BAE and MAE are the gross book value of assets and machinery at the end of the period; BR and MR are rents paid for buildings and machinery, and r is the user cost of capital (Wang, 1994).

Payroll and Average Wages

Payroll is the sum of total salaries and wages (SW) plus legally required supplemental labor costs (LE) and voluntary supplemental labor costs (VLC). Average wages are payroll divided by total employment (TE).