

Productivity Dynamics with Technology Choice: An Application to Automobile Assembly

JOHANNES VAN BIESEBROECK
University of Toronto

First version received August 2001; final version accepted August 2002 (Eds.)

During the 1980's, all Japanese automobile producers opened assembly plants in North America. Industry analysts and previous research claim that these transplants are more productive than incumbent plants and that they produce with a substantially different production process. I compare the production processes by estimating a model that allows for heterogeneity in technology and productivity, both of which are intrinsically unobservable. The model is estimated on a panel of assembly plants, controlling for capacity utilization and price effects.

The results indicate that the more recent technology uses capital more intensively and it has a higher elasticity of substitution between labour and capital. Hicks-neutral productivity growth is estimated to be lower, while capital-biased (labour-saving) productivity growth is higher for the new technology. Using the estimation results, I decompose industry-wide productivity growth and find plant-level changes in lean plants to be the most important contributor. Plant-level productivity growth is further decomposed to reveal the importance of capital-biased productivity growth, increases in the capital–labour ratio, and returns to scale.

1. INTRODUCTION

Measuring technological change, which is defined as the shift of a production function over time, becomes harder if plants can choose which production technology to adopt. In the automobile assembly industry there is evidence that plants choose to operate with mass or lean technology. I estimate technical change for the industry using a rich plant-level data set for the U.S. Firms choose which technology to employ, but their choice is not observable to me. The shape of both technologies and the rate and factor-bias of their shift over time is estimated jointly with the technology decision. I use the model to decompose aggregate productivity growth and find that firm-level change in lean production plants is the most important contributor to aggregate productivity growth for the industry, but it only appears after controlling for the technology choice. In contrast with previous results in the literature, relocation of resources between plants is only the second most important effect. Labour productivity growth at lean plants is driven by increased capital use and capital-biased technical change. In mass production plants, on the other hand, almost all labour productivity advance is the result of Hicks-neutral technological growth.

Traditional productivity measures suggest that the automobile industry enjoyed a considerable productivity improvement since the 1980's, the period when the first Japanese producers entered the industry in the U.S. The number of vehicles produced per worker, an often quoted statistic, is on average higher in Japanese transplants than in plants owned by American producers. Researchers have concluded that entry of more productive plants caused industry-wide productivity to increase. Compositional changes at the intensive margin—relocation of shares between active plants—and the extensive margin—plant turnover—are thought to be the principal sources of aggregate productivity change.

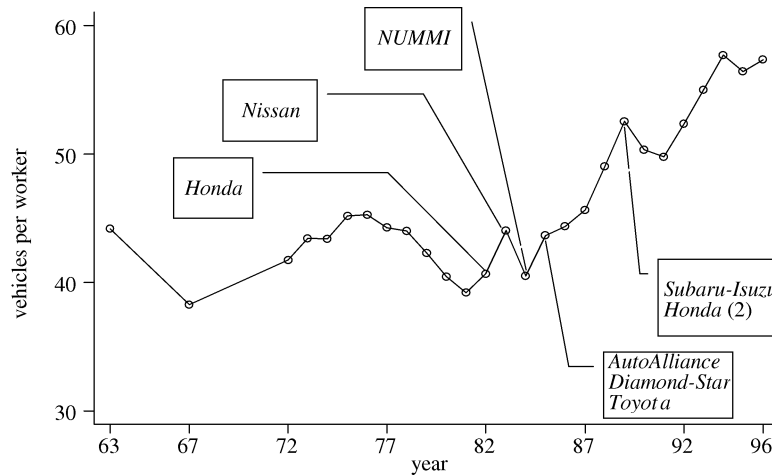


FIGURE 1

Two stylized facts: (1) average vehicles produced per worker for the industry increased significantly between 1980 and 1996 and (2) many Japanese plants entered right before the acceleration in (labour) productivity growth

At the same time, a largely separate literature claims that the entrants produce with a different technology, dubbed lean, or modern, manufacturing. Characteristics of lean production are team work, less standardization, flexible equipment, decentralization of decisions, less emphasis on scale economies, and increased flow in the production process.¹ Heterogeneity in technology is explained by the existence of two *systems* of production, lean and mass, each experiencing technological change at different rates and possibly different factor-bias. In the estimation, I allow for both technologies to coexist in the sample, but it would be overly restrictive to assume that the technology of a plant is exogenous. Firms decide what technology to employ at each plant, but the choice is not directly observable to me. I use observable features of the production process to predict the technology each plant is operating with.

Both stylized facts—aggregate productivity growth and entry by Japanese plants—are depicted in Figure 1. A popular measure of labour productivity, vehicles produced per worker, is calculated using all firms in the data set. Some researchers have concluded from Figure 1 that entrants are more productive and that productivity growth is driven by compositional effects. This is premature, because other effects could explain the increase in labour productivity as well, for example biased technological change in existing plants that use the old technology or technology switching that provides a one-off productivity increase for the plants involved. If I simply calculate labour productivity separately for plants built before and after 1982, as in Figure 3, the productivity increase is mostly situated with old plants. This casts doubt on explanations that focus solely on the entrants and compositional effects. The model I propose allows me to explicitly address a number of questions. Do Japanese-owned plants experience higher rates of productive growth than American-owned plants? Are the production technologies really different and in what way? What is driving aggregate productivity change in the industry?

A myriad of factors influence the number of vehicles produced per worker for the industry as a whole. The emergence of lean technology caused a major shift in production paradigm.

1. Many articles from the international motor vehicle program (IMVP) have described in great detail how Japanese plants differ from their American and European competitors along several dimensions. The program has generated more than 328 working papers since 1986. The book by Womack, Jones and Roos (1990) presents several findings of the program.

Some researchers argued that imitation of the technology used in transplants by incumbent plants led to a one-time productivity gain. Because firm-level changes have to be aggregated, changes in weight play a role as well. If resources are relocated from below to above average productive plants, the industry can advance without any productivity growth at the plant-level. Labour productivity can also be misleading. Relative factor prices changed significantly from 1963 to 1996, which can lead firms to choose a different combination of inputs, even without any change in technology. Estimating technological change at the firm-level as a shift in the production function takes input substitution into account. The rise or decline of one technology can also have compositional effects, because technologies might differ in rate and factor-bias of technical change. Disentangling the underlying effects is necessary to make robust conclusions about the most important determinants and evolution of productivity in the industry.

The task is complicated by the basic unobservability of technology. Unobserved heterogeneity in productivity levels introduces a potential simultaneity bias, because plants choose inputs conditional on their own level of productivity. I use a structural model of technology and input choice to control for both unobservables. I exploit the sequential nature of input choice together with distributional assumptions to control for unobserved productivity. A plant is assumed to make an explicit choice between both technologies before production starts. In the consequent operation, the input tradeoffs a plant faces and technological change it experiences are determined by the technology chosen before. Knowledge about the assembly process is used to construct a production function that can capture the most important differences between both technologies. Extra data are collected to control explicitly for other important effects, such as capacity utilization and price-setting.

I constructed a new and comprehensive data set of automobile (and light truck) assembly plants in the U.S. Input measures are obtained from the longitudinal research data set (LRD), from the Bureau of the Census. It provides reliable input statistics, because all plants are legally required to report the information. Variations in capacity utilization distort the relationship between measured input levels and the actual services a plant derives from them. Explicitly modelling and observing the number of shifts a plant is operated can control for it. An advantage of the automobile industry is the existence of a well-defined unit of output, a vehicle. I collected actual production volumes to avoid using deflated sales or value added as output measure, which contain the effects of price-setting. Data on the type of vehicle assembled, ownership, and the timing of changeovers,² are used to proxy the attractiveness of each technology or the cost of a technology change.

The two technologies I estimate differ significantly and in a meaningful way. The mass technology exhibits constant returns to scale and has a higher capital and labour share, but lower costs related to operating a shift. The technology identified as lean production, experiences decreasing returns to scale, but increasing returns to shifts. The elasticity of substitution between capital and labour and the own-price elasticities are significantly larger for the lean technology. These findings coincide with the notion that lean manufacturing is more flexible and relies less on standardization. The rate and factor-bias of technical change for each technology are identified separately. Lean production experiences a relatively lower rate of Hicks-neutral productivity growth and a significantly higher rate of capital-biased, labour-saving, productivity growth. Labour productivity growth for mass producers is mainly driven by Hicks-neutral productivity growth. For lean producers, capital-biased productivity growth and an increase in capital are the two most important contributors to labour productivity growth.

2. A plant undergoes a changeover before a substantially different vehicle can be assembled at a plant. These changeovers are accompanied by changes in the assembly line and the replacement of a substantial portion of the capital stock.

The lean technology has also become more popular over time. It is the result of new plants being more likely to enter with the lean technology and existing plants switching technologies, which is more likely in changeover years, when switching costs are lowered. Given the higher returns to scale associated with mass production, it is optimal for some plants to stick with the mass technology, especially if they produce only trucks and if model changes are rare. Decomposing industry-wide labour productivity growth illustrates the prime importance of plant-level growth in lean plants, although relocation of resources from mass to lean producers also accounts for a significant portion of the aggregate growth rate.

The next section provides an overview of the production process in the automobile assembly industry, which motivates the production function I adopt. The empirical strategy to identify both technologies and to control for unobserved productivity differences is described in Section 3. The timing of decisions and productivity shocks in the model are also described there. Section 4 contains a description of the plant-level data set and motivates the existence of two technologies. The likelihood function is derived in Section 5, which also contains the estimation results. In Section 6, I derive some conclusions about the evolution of productivity growth in the industry.

2. PRODUCTION AND PRODUCTIVITY IN AUTOMOBILE ASSEMBLY

The automobile assembly industry has not been a stranger to productivity growth or changes in technology. Henry Ford revolutionized the industry when he introduced the moving assembly line in 1913. He exploited returns to scale by ever-increasing specialization throughout the entire production process and achieved enormous cost reductions. Standardization of components and tasks and a limitation of model variety led the industry to operate branch assembly plants close to population centres, supplied by large component plants located in the Detroit area. Each assembly line produced only one model and inventories at dealerships were used to smooth production, allowing plants to operate as close to optimal scale as possible. Within the assembly plant, each worker or machine performed a single task. Little employee training was required and the focus was on standardization at all levels.³

After the Second World War, Japanese producers started to produce automobiles commercially.⁴ The specific domestic situation led them to develop a production process that was markedly different from the American model. The closed labour market, labour laws introduced during the American occupation, and a major strike after mass layoffs at Toyota in 1949, led to the establishment of quasi-lifetime employment. Lack of skilled workers made it necessary for companies to train employees extensively, which was only incentive compatible with low worker mobility. Lack of capital motivated manufacturers to shift part of the development of new components to suppliers. A more integrated supply chain was a natural complement, evolving into just-in-time delivery of components and lower inventories. The scarcity of capital also forced producers to adapt equipment to perform multiple tasks, instead of standardizing operations and tasks performed by employees and machines. As a result, workers became less interchangeable and work was organized in teams. Gradually, the need to perform multiple tasks made machines more sophisticated, with more scope for capital-biased technological progress. The domestic market was small and different groups of customers demanded vastly different vehicles. Small scale production combined with flexibility in assembly and design were features of lean production from the start. Assembly plants often produced more than one vehicle and changeover times were significantly shorter than in mass production plants.

3. Rubenstein (1992) provides an overview of the development of the industry in the U.S.

4. The description of lean production draws on the book by Womack *et al.* (1990).

Although the actual operation of an assembly plant and the organization of the supply chain and vehicle development differs significantly between mass and lean production, both systems share some important characteristics. Eiji Toyoda, who oversaw the development of the Toyota Production System,⁵ visited the Ford Rouge assembly plant in Detroit in 1950. Ford engineers made the reverse journey in 1981, visiting Mazda's main production complex in Hiroshima, which was modelled after Toyota city. Both systems operate with a moving assembly line, organize work in 8 hour shifts, install new capital and change models over the summer, and outsource the majority of components. A widespread misconception is that mass production plants make a vehicle from scratch, while lean plants simply bolt together imported components. Both production systems rely to a similar extent on outsourced components and the degree of outsourcing only increased after the sample period, with the recent trend towards modularization.⁶ American *firms* differ from their Japanese competitors by a greater extent of vertical integration. As a result, American assembly plants receive more components from in-house suppliers and they are more integrated and outsource less at the firm-level. There has been a trend in the industry to rely more on outside firms for components, but this had little impact on the activities carried out at the assembly plant.

To decide on the functional form of the production function one should bear in mind that it represents a technological relationship between physical output and physical inputs. In the data set, I only observe the value of material input, which displays two trends. The (value of) material input per vehicle ($\frac{P^m M}{Q}$) increases significantly over the sample period. At the same time, the material-sales ratio ($\frac{P^m M}{P^q Q}$) remained virtually constant. Quality upgrading of vehicles, through higher quality components, provides one explanation consistent with both observations. Newer cars use better fabric, better quality paint, more powerful brakes, etc. The amount of intermediate inputs per vehicle remained constant, but the quality and (real) price increased over time. This is consistent with the observed increase in price for the final product. The real price per vehicle increased from \$6000 to \$13,000 (in 1987 \$) over the sample period. An evolution in outsourcing activities provides an alternative explanation for the two stylized facts. The use of more material inputs in the assembly process can explain the upward trend in the material-vehicle ratio. To reconcile this with the constant material-sales ratio, the price increase for vehicles has to outpace the price increase for materials. We do not find evidence for such a trend at all.⁷

I adopt the first interpretation, quality upgrading with a constant amount of material inputs per vehicle. The volatility of the material-sales ratio is small, compared with the volatility of capital or labour input, providing additional support for the assumption of constant material input per vehicle.⁸ The production function is Leontief, allowing no substitution between materials and other inputs, but the proportion of material input is allowed to vary across technologies. Most components or subsystems, such as the engine, transmission, brake system, or electrical

5. Credit for devising the Toyota Production System, the prototypical lean production model, generally goes to Taiichi Ohno, the chief production engineer when Eiji Toyoda was president of the Toyota Motor Company.

6. Only in recent years do we observe the exact activities carried out at each plant, see Harbour and Associates (1999). Surprisingly, Japanese-owned assembly plants are marginally less outsourced.

7. This explanation faces an additional problem if the capital-labour combination to assemble different sub-components of equal value differs. In that case, the production function depends on exactly which activities are outsourced and no stable material aggregate exists. Without observing the actual material inputs, it would be impossible to remedy this problem.

8. I use the following formulas to evaluate the volatility of the material-sales ratio over time and across plants: $\frac{1}{J} \sum_{j=1}^J \frac{\frac{1}{T} \sum_t \left(\frac{M}{S} - \frac{1}{T} \sum_t \frac{M}{S} \right)^2}{\frac{1}{T} \sum_t \frac{M}{S}}$ and $\frac{1}{T} \sum_{t=1}^T \frac{\frac{1}{J} \sum_j \left(\frac{M}{S} - \frac{1}{J} \sum_j \frac{M}{S} \right)^2}{\frac{1}{J} \sum_j \frac{M}{S}}$, and similarly for labour and capital (plants are indexed by $j = 1 \dots J$, time by $t = 1 \dots T$). These measures reveal that capital-sales is six times more volatile than material-sales and labour-sales is four times more volatile. In addition, the difference between material volatility across plants hardly exceeds the volatility over time for a given plant.

components, are outsourced to in-house suppliers or outside firms, and simply installed at the assembly plant. Production for plant j producing at time t with technology i is governed by the following relationship:⁹

$$Q_{jt} = \min \left[\alpha_i^0 \left(\frac{M_{jt}}{P_{ijt}} \right)^{\alpha_i^m}, \bar{Q}_{jt} e^{\epsilon_{ijt}^q} \right]. \quad (1)$$

The first part of equation (1) relates output to the volume of material inputs, obtained by dividing the observable value of materials by a price index,

$$P_{ijt} = e^{p_i^0 + \theta_i^m t + \epsilon_{ijt}^m}. \quad (2)$$

The index captures quality upgrading of components over time (θ_i^m), which can vary across technologies. For example, over the sample period most manufacturers substituted cheaper drum brakes for more expensive, but higher quality, disc brakes. The component cost per vehicle increased and it is not captured by price deflation, because both brakes are different goods. At the same time, the amount of material input per vehicle remained constant (four brakes per vehicle) and assembly time also did not change. ϵ_{ijt}^m is a stochastic component that is unobservable to the econometrician.¹⁰ The coefficient on materials and the price index are estimated, but the main interest is in the other coefficients of the production function.

The second part of the production function links observed output (Q) to planned output (\bar{Q}). Firms in the industry operate in a decentralized fashion. The headquarters of the firm decides on a production volume for the plant, based on market demand predictions. The plant manager tries to satisfy the production requirement at minimal cost. *Ex post*, the actual production volume will differ from planned output, which is used to decide input levels. I model this as the realization of an independent shock to production (ϵ_{ijt}^q). The fixed-coefficient technology for materials makes it reasonable to assume that materials are determined after the realization of the production shock.¹¹

Planned output is a function of capital, labour, shifts, and productivity, which are not all decided simultaneously. In practice, a plant operates an 8 hour shift or stands idle. Whenever it operates, the entire assembly line or all of the capital stock is used. New investment is generally installed in the summer, when the assembly line is retooled for the new model and the plant does not operate for a couple of weeks. During the year, the plant manager decides how many shifts to operate the assembly line and at what speed, which determines the labour input required. In effect, the plant manager has two degrees of freedom, number of shifts and line-speed, and can satisfy the production requirement using different combinations of shifts (capital use) and labour input. The extent to which more workers are needed to increase line-speed, is determined by the elasticity of substitution between labour and capital. Workers are not laid off when the line is idled, but I observe actual hours worked and, therefore, variations in actual labour input. Because the assembly line has a minimum speed to operate efficiently, it is often advantageous to idle the plant for a number of shifts. Capacity utilization at the extensive margin varies significantly between plants and across time periods.

9. Benkard (2000) also uses a Leontief technology to model the production function in aircraft production, motivated by the impossibility to substitute between engines and other inputs.

10. It does not matter whether a plant observes ϵ_{ijt}^m or not, because it has no control over the amount of materials to use. Output is determined exogenously (at the firm-level) and material input is linked to output through a constant-coefficients technology.

11. It is straightforward to adjust the model to make material input choice precede the production shock. The only change to the estimation would be the introduction of a positive correlation between the errors in the production and material equation.

Total output is given by an output per shift function multiplied by productivity and the number of shifts the plant is operated,

$$\bar{Q}_{jt} = S_{jt}^{\alpha_i^s} f_i \left(\frac{L_{jt}}{S_{jt}}, \hat{K}_{ijt} \right) e^{\theta_i^n t + \omega_{ij}^n}. \quad (3)$$

The scale factor for the number of shifts has a factor α_i^s , to capture returns to shifts. This coefficient can be smaller than one if, for example, running more shifts reduces maintenance time, leading to more machine breakdown and lower production per shift. It can be larger than one if there are positive spillover effects between shifts, for example shared overhead or reductions in start-up time. Output per shift is a function of labour per shift and efficiency units of capital, and is multiplied with a plant-specific productivity factor. The shape of $f_i(\cdot)$ determines the technological substitution possibility between capital and labour for technology i .

I adopt the translog specification for the per shift production function. It is a flexible functional form which allows identification of Hicks-neutral and capital-biased technological change and it also does not restrict the elasticity of substitution to be the same for different technologies:

$$\log f_i = \alpha_i + \alpha_i^l l + \alpha_i^k \hat{k}_i + \frac{1}{2} \beta_i^l l^2 + \frac{1}{2} \beta_i^k \hat{k}_i^2 + \beta_i^{lk} l \hat{k}_i,$$

where l is the logarithm of labour per shift and \hat{k}_i are efficiency units of capital, also in logarithms ($\hat{k}_i = k + \theta_i^k t + \omega_i^k$).

The technology a plant produces with—as captured by the production function—is a relationship that determines the maximum output that can be obtained from a bundle of physical inputs. Some plants produce more output with the same amount of inputs. These plants can still be considered to operate with the same technology, although with a different level of productivity. Input substitution possibilities are identical, but the production frontier is shifted radially. Hicks-neutral differences between plants are captured by ω_{ij}^n and neutral productivity growth, θ_i^n , affects all plants equally.

In this industry, it is often assumed that a significant part of technological progress comes through improved machinery and equipment. Over time, it changes the input tradeoff embodied in the technology. In mass production, machines perform a single task, while lean technology relies on flexible equipment. Both potentially experience different rates of technological change. Capital-biased productivity growth, affecting all plants that produce with technology i similarly, is captured by θ_i^k . In addition, I allow for a plant-specific shock to productivity, ω_{ijt}^k , that is proportional to its capital stock. A plant observes all four productivity terms and chooses (variable) inputs accordingly.¹²

In equilibrium, both parts of the production function (1) will hold with equality. Substituting the price index for materials (2), planned output (3), and efficiency units of capital, taking logarithms and rearranging gives two estimating equations,

$$(IIIa) \quad q_{jt} = \alpha_i^s s_{jt} + \log f_i \left(\frac{L_{jt}}{S_{jt}}, K_{ijt} e^{\theta_i^k t + \omega_{ijt}^k} \right) + \theta_i^n t + \omega_{ij}^n + \epsilon_{ijt}^q$$

$$(IIIb) \quad m_{jt} = \alpha_i^0 + \frac{1}{\alpha_i^m} q_{jt} + \theta_i^m t + \epsilon_{ijt}^m.$$

All coefficients vary by technology, which is indicated by the index i . A question remains how many technologies to allow. At one extreme, one could postulate that all plants in the sample produce with the same technology. All observed heterogeneity would then be attributed to

12. A plant chooses capital knowing three of the productivity terms, but not ω_{ijt}^k . After the new investment is installed, it makes a production run and learns ω_{ijt}^k . Variable inputs and shifts are only chosen afterwards and a plant will choose them conditional on all four productivity terms.

measurement or sampling error. At the other extreme, one could assume a different technology for each plant and use random coefficients. This approach has not yielded satisfactory results, see Mairesse and Griliches (1990). Diamond, McFadden and Rodriguez (1978) show that it is impossible to identify the bias in productivity growth from the elasticity of substitution using only time-series variation. I take an intermediate position by allowing two—but only two—technologies. The trade press takes this stance by drawing a sharp distinction between lean and mass production. Milgrom and Roberts (1990) provide theoretical support for such restriction. They describe modern manufacturing as a set of activities that exhibit complementarities. The marginal product of adopting the new technology for one activity is increasing in adoption on other dimensions. It makes intermediary systems that are composed of elements from the old and new systems unstable.¹³

Because the technology a plant produces with is unobserved, it is not possible to rely on the theory of index numbers to obtain productivity measures. I estimate the production functions econometrically, which is flexible enough to allow for variable returns to scale and variations in capacity utilization. Two sets of coefficients, one for each technology, and an equation governing the technology choice are estimated jointly. In the next section, I describe how to control for unobserved productivity differences, ω^n and ω^k , and how the technology choice is modelled.

For the automobile industry, Friedlaender, Winston and Wang (1983) were the first to estimate a firm-level production function. They assumed one homogeneous technology, although one of the four American producers was excluded because its production technology was thought to differ substantially. Krafcik (1988) used detailed plant-level data, also enforcing the one-technology assumption. In the discussion of his results, as in many other IMVP studies, Japanese-owned plants are treated as a separate class that have noticeably higher productivity. One study, Griffith (1999), addresses productivity differences between domestic and foreign plants in the U.K. explicitly, but the difference is restricted to affect only the constant term in the production function. Finally, Fuss and Waverman (1992) estimate an industry-level production function for four countries, paying particular attention to capacity utilization.

3. EMPIRICAL STRATEGY

3.1. *Timing*

To estimate both equations generated by the production function, derived in the previous section, it is necessary to control for unobserved productivity differences and the technology choice. I exploit the timing and level (plant vs. firm) of decision making in the industry, using observed decisions to infer unobserved variables. It does entail a number of restrictions on how firms operate and it suggests a maximum likelihood estimation procedure. In Section 3.2, I discuss how to control for ω_{ij}^n using plant-fixed effects and for ω_{ijt}^k by inverting the first-order conditions for variable input choice. In Section 3.3, an explicit model of technology choice is developed, which generates a framework for estimation. An outline of the model, with the timing of decisions and when errors are realized, is provided in Figure 2.

When a plant is built, the firm makes a technology choice, mass or lean ($i = \mathcal{M}$ or \mathcal{L}). The constant component of productivity (ω_{ij}^n), known to the plant, but not to outsiders, is realized next, before any production decision is made. In the following years, mass technology plants have the option to adopt the lean technology, at which point a new plant-level productivity term gets

13. The complementarities were apparent in the preceding discussion of lean technology. Providing extensive training is only optimal if worker mobility is low. It also makes it easier to use flexible equipment and organize work in teams. An integrated supply chain is easier to achieve if suppliers are involved in R&D, and only then is a just-in-time inventory system feasible.

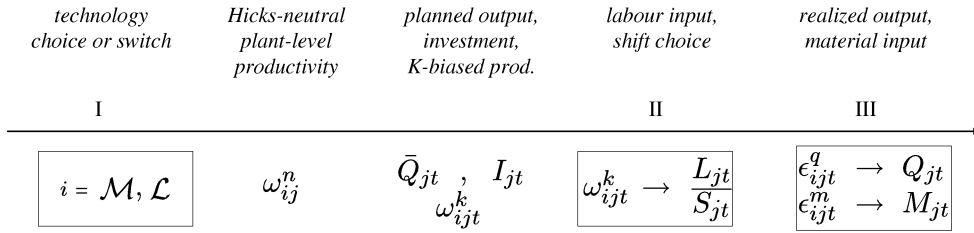


FIGURE 2
Timing of decisions and errors

realized. ω_{ij}^n does not vary over time. Planned output (\bar{Q}) is determined at the next stage. Plants get the production requirement handed down from headquarters and minimize costs subject to the production constraint. The first decision made at the plant-level is investment (I). The annual choices of output and investment are not modelled explicitly. Before production starts, a plant tests the assembly line and learns the productivity of the capital stock (ω_{ijt}^k), which determines the efficiency units of capital that enter the production function. The optimal choice of labour per shift is a function of capital productivity and it generates the first estimating equation. The final stage is actual production. The technology generates two estimating equations, already derived in Section 2. Actual output (Q) will differ from planned output because of an *ex post* shock to production (ϵ^q). Finally, material input (M) is proportional to the actual output produced and is influenced by a price shock (ϵ^m).

There are two sources of unobserved productivity differences in the production function. I control for both to avoid simultaneity bias. The first component (ω_{ij}^n) is constant over time and captured by plant-fixed effects. The second component (ω_{ijt}^k) varies over time and represents a shock to capital productivity. Variable inputs are only chosen after the actual productivity of the capital stock is observed. Inverting the first-order conditions gives an expression for ω_{ijt}^k in terms of observable variables. The following table summarizes the primitives of the empirical model, i indexes technology (\mathcal{M} or \mathcal{L}) and the j and t subscripts for plants and years are suppressed. β_i represents the technology-specific parameters to estimate.

decision	equation	explanation
(I)	$p = \Pr(i = \mathcal{M}; \beta_i)$	→ technology choice probability
(II)	$g_i\left(\frac{L}{S}\right) = g^l(w_i^n, K, t; \beta_i) + \omega_i^k$	→ f.o.c. for variable inputs
(IIIa)	$Q = g^q(S, L, K, t, \omega_i^k, \omega_i^n; \beta_i) + \epsilon_i^q$	→ ω_i^k from (II), ω_i^n is plant-fixed effect ϵ_i^q independent of inputs, ω_i^k , and ω_i^n
(IIIb)	$M = g^m(Q, t; \beta_i) + \epsilon_i^m$	→ ϵ_i^m realized at the very end

The technology decision (I) is used to derive the probability a plant produces with either technology in each year. The labour per shift decision (II) and the production function (III) provide three estimating equations. The equations imply a distribution for each endogenous variable conditional on technology. The likelihood function for the unconditional joint distribution is obtained by multiplying the probability for each technology by the joint conditional distribution for the three endogenous variables. I assume that the three error terms— ϵ_i^k , ϵ_i^q , ϵ_i^m —are independently distributed and exogenous to the R.H.S. variables. I derive the first-order condition (II) and the probability of technology choice (I) in the remainder of this section. The exact functional forms for equations (IIIa) and (IIIb) were derived in Section 2.

3.2. *Unobserved productivity differences*

The level of productivity is only defined relative to a particular production function. Only the productivity levels of plants that produce with the same technology can be compared meaningfully. Productivity growth, on the other hand, is always well defined. It measures the shift of a production frontier over time. I will focus on comparing θ_i^n and θ_i^k across technologies, controlling for differences in ω_i^n and ω_i^k in the estimation. In the long run, the technology with the highest productivity growth rate will dominate.

A problem that studies of the automobile industry have not adequately addressed is the endogeneity of unobserved productivity differences. Plants choose inputs with knowledge of their own productivity level generating simultaneity bias. Least squares estimation of the production relation is inconsistent and instrumental variables are the traditional solution. In this industry, the dependency of plant decisions on productivity can work through shifts or hours worked if headquarters systematically allocate higher production volumes to plants with higher productivity, even conditional on the installed capital stock. If higher productivity plants are better able to use the entire range of assembly line-speed, there can be another effect of productivity on labour per shift. Exit, on the other hand, is unlikely to be correlated with productivity. Almost all plants that closed during the sample produced compact cars. The (exogenous) evolution of demand, more than low plant-level productivity, will determine which plants are closed.

Blundell and Bond (2000) demonstrate that in the context of the production model it is hard to come up with powerful instruments. The stochastic frontier literature uses distributional assumptions to integrate out unobserved productivity, but this is rather restrictive. Another solution is to use a behavioural equation to obtain an expression for the unobserved productivity in terms of observable variables. For example, Olley and Pakes (1996) invert the investment function non-parametrically to substitute the productivity term from the production function. Levinsohn and Petrin (1999) follow a similar approach, but they use material inputs instead of investment. My solution is similar, but by inverting the labour per shift equation, I obtain an exact expression for ω_i^k , avoiding the non-parametric approximation.

Controlling for the Hicks-neutral plant-level productivity ω_{ij}^n is straightforward because it is assumed to be constant over time. I include plant dummies in the production function, that serve a dual purpose. They control for the endogeneity of productivity and they allow to pool plants, using the number of vehicles produced as output measure. Controlling for ω_i^k , which varies over time, requires more work. Inverting the first-order conditions for the variable input choice will give me an expression for unobserved capital productivity. Most productivity studies using a translog production or cost function rely on Shepard's lemma to derive the factor-share equations, to aid in identification. The introduction of shifts as choice variable and the fact that plants choose labour and capital at different times, makes it necessary to solve the cost minimization problem explicitly.

As already mentioned in Section 2, every period a production plan is handed down from headquarters. Capital, output and technology are not decision variables at this point. The plant chooses labour and shifts to satisfy the production requirement and to minimize variable costs, which are twofold. First, the wage bill is proportional to the number of shifts and the average number of hours worked on a shift. Second, there are fixed and variable costs of operating a shift. The existence of labour unions complicates the tradeoff between labour and shifts the plant makes, because unionized plants save only a fraction of wages when they reduce labour input. Labour contracts negotiated in this industry specify that a percentage of the normal wage is paid even when plants are idled. Plants only take the variable portion of the wage into account and I multiply the observed wage by a factor that lies between zero and one (δ). For non-unionized plants, this fraction is normalized to be one.

Costs associated with operating a shift are not observed directly, but I estimate a fixed component (μ_i) and a component proportional to capital (ρ_i).¹⁴ Both components are equal across plants, but potentially differ across technologies. Many plants idle the capital stock for part of the year, which indicates a positive marginal cost of operating the capital stock. Complementary inputs (labour and energy), depreciation in-use, or maintenance cost are likely to influence the cost of operating a shift.

A plant minimizes costs, while satisfying the production requirement in expectation.¹⁵

$$\begin{aligned} \min_{\{L, S\}} & S \times \left(w\delta \frac{L}{S} + \rho_i K + \mu_i \right) & (4) \\ \text{s.t.} & Q = S^{\alpha_i^s} f_i \left(\frac{L}{S}, K e^{\theta_i^k t + \omega_i^k} \right) e^{\theta_i^n t + \omega_i^n + \epsilon_i^q} \\ & E(Q) \geq \bar{Q} \\ & \delta = 1 \quad \text{if plant is not unionized} \\ & \delta \in [0, 1] \quad \text{if plant is unionized} \\ & K \leq \bar{K}. \end{aligned}$$

In the objective function, w is the observed wage rate and δ is the fraction of the wage that is not paid when a unionized plant is idled. $\frac{L}{S}$ is the average hours worked per shift by all employees. K , the capital stock, is fixed at this point. S is the number of shifts the plant is operated over the entire year. Only the relevant part of the production function is repeated.

The model allows for the large under-utilization of capital often noted in the automobile assembly industry. If capital has a positive operating cost ($\rho_i > 0$) and the elasticity of substitution between capital and labour is finite, it can be advantageous to idle the plant for some shifts. A manufacturer can produce the same amount by employing more workers on each shift and running fewer shifts to save on capital depreciation, maintenance, and energy. This can only be accomplished by running the assembly line at a higher jobs-per-minute (JPM) rate, made possible by the increased labour input. Since reported capacity numbers for the industry are calculated as potential output, assuming 10 shifts per week and the initially reported JPM rate, this substitution behaviour will show up as lowered capacity utilization. The tradeoff between labour and capital is determined by the operating cost per shift (ρ_i), the returns to shifts (α_i^s), and the elasticity of substitution between capital and labour. All three factors are identified separately in the model.

To obtain an estimating equation, I solve the minimization problem and rework the first-order conditions. The Lagrangian to minimize is

$$\mathcal{L}\mathcal{G}(L, S) = w\delta L + (\rho_i \bar{K} + \mu_i)S + \lambda \left[\bar{Q} - S^{\alpha_i^s} f_i \left(\frac{L}{S}, K e^{\theta_i^k t + \omega_i^k} \right) e^{\theta_i^n t + \omega_i^n} \right],$$

where L and S are decision variables. The fixed cost per shift and scarcity of resources will guarantee that $K = \bar{K}$ and $E(Q) = \bar{Q}$.¹⁶ The first-order conditions are:

14. There is no unique solution for the minimization problem if returns to scale for the per shift production function are equal to α_i^s . With excess capacity, a plant can produce the same output in t shifts, using L hours per shift, or using all tL hours in one shift and employing t times as much capital. Introducing a fixed cost per shift guarantees that a plant will always use the entire capital stock if it operates a shift.

15. While this cost minimization problem may not be profit maximizing for the firm, it reflects the constraints and incentives the plant manager faces.

16. The only variable in the problem a plant does not observe at this point is the *ex post* shock to production, ϵ_i^q . The error term is assumed to be i.i.d. normally distributed and is orthogonal to the labour and input choice.

$$w\delta = \lambda e^{\theta_i^n t + \omega_i^n} S^{\alpha_i^s - 1} \frac{\partial f_i}{\partial L/S}$$

$$\rho_i \bar{K} + \mu_i = \lambda e^{\theta_i^n t + \omega_i^n} S^{\alpha_i^s - 1} \left(\alpha_i^s f_i - \frac{\partial f_i}{\partial L/S} \frac{L}{S} \right).$$

Dividing both equations and rearranging gives

$$\underbrace{\frac{w\delta L}{w\delta L + S(\rho_i \bar{K} + \mu_i)}}_{\lambda_i} = \frac{1}{\alpha_i^s} \frac{\partial \log f_i(\frac{L}{S}, \hat{K})}{\partial \log(\frac{L}{S})}. \quad (5)$$

The expression on the left is the labour share in variable costs. It is not directly observable because the parameters ρ_i , μ_i and δ have to be estimated. The optimal labour–share ratio in equation (5) does not depend on the plant-specific neutral productivity term or the shock to production.

Using the translog specification for f_i I find

$$(II) \quad \frac{w\delta \frac{L}{S}}{w\delta \frac{L}{S} + \rho_i \bar{K} + \mu_i} - \beta_i^l \log \frac{L}{S} = \alpha_i^l + \beta_i^{lk} (k + \theta_i^k t + \omega_i^k).$$

The dependent variable, $\frac{L}{S}$, is a nonlinear function of the disturbance, ω_i^k . I use this expression to substitute ω_i^k in the production function, allowing for consistent estimation of equation (IIIa). In addition, I estimate equation (II) directly, to aid in identification.

3.3. Unobserved technology choice

To compare productivity growth for the lean and mass technology, I need to obtain consistent estimates for the parameters in both production functions. All three previously derived estimating equations are conditional on the technology choice. Simply regressing output on inputs cannot produce consistent estimates, because the technology choice is not observed directly. Only in recent years are the exact activities carried out at each plant observed. In addition, much of the difference between technologies is in the organization of work or the product flow, which are hard to measure.

The easiest solution would be to separate plants in two groups using an observable characteristic. A Chow test for structural breaks can be used to test whether the production technology is different for plants in both groups. The most straightforward criterion is to equate lean technology with plants that were built most recently. Figure 3, however, illustrates that most of the productivity growth is situated in the older plants. The newer, mostly Japanese, plants enter with a higher level of labour-productivity, but they increase productivity at a slower pace.

Performing similar exercises using time, nationality of owner, or other variables supposedly correlated with the technology choice, *e.g.* level of outsourcing or inventory levels, to separate plants in two groups, the null hypothesis of identical technologies can always be rejected. Unfortunately, the coefficient estimates for parameters in the production function differ for different criteria to split the sample. There is evidence that some mass technology plants adopt the lean technology during the sample period. It is also not certain that the newer, lean technology strictly dominates mass technology. In particular, the mass technology is thought to be better at exploiting scale economies, making adoption of the new technology not guaranteed. A more systematic method to separate plants, allowing for technology switching, is needed.

One possibility is to use an endogenous switching model as in Dubin and McFadden (1984), using a selection equation for technology that is a linear combination of observable variables.

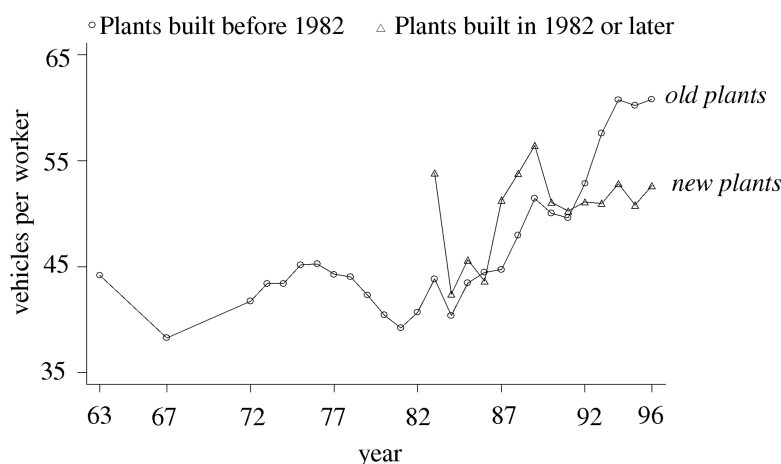


FIGURE 3
Evolution of vehicles produced per worker for new and old plants

I exploit the fact that the initial technology choice happens only once and that switching is rare. The probability that a plant operates with one technology depends to a large extent on the technology employed in the previous period. I estimate the initial technology choice and the probability for technology switching, instead of estimating the probability for the technology in each year directly. Beard, Caudill and Gropper (1991) also model technology transitions, but they fix the switching probability to a constant. I allow the probability of switching to differ across plants and over time as a function of observable variables and estimate the transition probabilities jointly with the production function, derived earlier.

A crucial assumption is that the choice of technology is made at the firm-level, before production starts. It makes the technology choice orthogonal to the plant-level productivity factors and error terms. The firm makes a net present value comparison between both technologies and chooses the technology that gives it the highest discounted profits, taking expectations and switching costs into account. Building a structural model to explicitly model the comparison is too complicated. A myriad of effects enter this decision: strategic considerations, pricing of a durable good, the joint decision for many plants, etc. Specific assumptions about the expectations of all exogenous variables and the optimal response functions for all choice variables would be needed.

Instead, I take a reduced-form approach to control for the technology decision and assume the probability for a technology is a function of observable variables. The probability that plant j enters the sample with the older, mass technology is¹⁷

$$(I') \quad \psi_{jt} = \frac{1}{1 + \exp(W_{jt}\eta)}$$

17. Equation (I') can be rationalized by a selection problem in the first period:

$$i_{(jt)} = \begin{cases} \mathcal{M} & \text{if } W_{jt}\eta < \epsilon^1 \\ \mathcal{L} & \text{if } W_{jt}\eta \geq \epsilon^1 \end{cases}$$

where ϵ^1 follows the extreme variable distribution and t is the entry year for plant j . The $W_{jt}\eta$ aggregate captures the difference in net present value for the firm from operating plant j with the lean instead of the mass technology.

TABLE 1
Transition probabilities between technology

		Time t	
		\mathcal{M}	\mathcal{L}
(I)	\mathcal{M}	$p_{jt} = \frac{1}{1 + \exp(Z_{jt}\gamma)}$	$1 - p_{jt} = \frac{\exp(Z_{jt}\gamma)}{1 + \exp(Z_{jt}\gamma)}$
	\mathcal{L}	0	1

Variables in W capture the relative profitability of the lean vs. mass technology. The evolution of both technologies, experience the firm has operating each technology, and the market segment for the model a plant produces, are likely to be important determinants. For example, it is argued that lean production is better suited if demand is volatile or for production at a lower scale. The trend towards more cars per household has led to a larger demand for speciality or niche vehicles, making the lean technology more popular over time.

A firm makes the initial technology choice before a plant enters the sample. In addition, at the start of each year, it has the option of switching technology. The data spans 27 years, generating 2^{27} possible paths for technology. It is possible to estimate a model that allows plants to shift freely using the EM algorithm, as illustrated in Hamilton (1989). It necessitates fixing the transition probabilities between technologies to be constant across plants and over time. I decided to be less flexible on the direction of change and more flexible on the transition probabilities, making them vary over time and across plants. To estimate the model, I assume that lean production is an absorbing state, which reduces the number of possible paths for technology to 28. Plants built before the lean technology was available can end up with the “wrong” technology for their characteristics. Only these plants will opt to switch and incur switching costs. The transition probabilities are a function of observable variables, just like the initial probability choice in (I'). The likelihood will be constructed from the entire time path of the endogenous variables, instead of year by year.

The transition probabilities are illustrated in Table 1.¹⁸ Variables in Z determine the probability a firm finds it more beneficial to produce with the lean technology in plant j , rather than sticking with the mass technology. The same demand variables as in W influence this transition probability. In addition, I incorporate a dummy for changeover years to capture switching costs. When a substantially modified model is introduced, a plant has to adjust a large part of the assembly line. This is likely to be a good moment to make the technology switch as well, because much of the capital stock has to be replaced anyway.

4. DATA ON ASSEMBLY PLANTS

4.1. Observed variables

From the preceding discussion it is clear that three sources of data are needed: variables characterizing the technology choice, output quantities and input levels, adjusted for intensity of use.

18. Underlying the probabilities in (I) is a different selection problem:

$$i_{(jt)} = \begin{cases} \mathcal{M} & \text{if } [i_{(jt-1)} = \mathcal{M} \ \& \ Z_{jt}\gamma < \epsilon^2] \\ \mathcal{L} & \text{if } [i_{(jt-1)} = \mathcal{M} \ \& \ Z_{jt}\gamma \geq \epsilon^2] \text{ or } i_{(jt-1)} = \mathcal{L}, \end{cases}$$

where ϵ^2 also follows the extreme variable distribution and plant j entered before year t .

The principal data source is the LRD constructed by the Center for Economic Studies at the Bureau of the Census. The data are collected from plant responses to the yearly Annual Survey of Manufactures and the 5-yearly Census of Manufactures. Observations are plant-years and plants are linked over time. Coverage includes all plants with SIC code 3711 (motor vehicles and car bodies) for their main product. The data spans the years 1963, 1967 and 1972–1996.

Industry publications are used to supplement the LRD. These cover a smaller number of companies and plants. Only statistics for plants owned and operated by one of the large automobile or truck companies are available. Omitted plants are owned by smaller firms and specialize in converting cars to limousines, trucks to campers, or they only make car or truck bodies. In addition, the LRD contains some engine or component plants that produce a large number of bodies or completed vehicles only sporadically.¹⁹ Seventy-four percent of the observations in the LRD with SIC code 3711 could be matched to the data from other sources. These plants represent 94% of the employment in the industry.

This data set provides reliable and complete input statistics. The labour input measure we use is total hours worked at the plant. Hours worked by non-production workers are imputed using their relative wage.²⁰ Volatility in production combined with a high degree of unionization have resulted in companies having a number of temporarily unemployed workers on their payroll. Only actual hours worked are counted in labour input, although the amount paid to temporarily unemployed workers is included in the labour costs. The input choice model explicitly recognizes the different incentives unionized plants face and provides a method to control for unionization in the estimation as well. Capital input is constructed from book values. An alternative measure, using the perpetual inventory method, yielded almost identical results. It is deflated using the capital goods deflator for the industry from the NBER productivity data set. Material input includes raw materials and intermediate products, fuels and electricity. All are scaled by the appropriate deflator from the NBER data set.²¹ In principal, energy could be included separately in the production function. It was not done because it is smaller than 1% of costs for almost all plants and already included in raw materials for some. Table 2 contains summary statistics for the relevant variables.

The second piece of information needed is output. Most productivity studies use deflated sales or value added as output measure, because actual production volumes are not generally available. If a firm has price-setting power, price changes will erroneously be interpreted as productivity changes. For example, if a firm produces subject to an inelastic demand, it can increase sales by raising the price. Deflation by an industry-wide price index does not capture the individual price movements. Because output and inputs do not change the increase in “output” will be interpreted as a productivity gain.²² For a solution using specific assumptions about competition and demand, see Klette and Griliches (1996). In a concentrated industry like automobile assembly, price-setting is likely to be important. One example is the sale of identical vehicles by different firms. All American producers have joint-ventures with Japanese partners, assembling vehicles jointly and selling identical products under different nameplates.

19. Plants are classified according to the industry category of their main product. Some engine or component plants assemble a limited amount of vehicles as well, which in some years can make up the majority of sales. This can result in these plants being classified in SIC industry 3711 in some years.

20. Hours worked is directly observed for production workers, but only the total number of non-production workers is known. Assuming that wages for non-production workers vary in proportion to hours worked, we can estimate the hours they worked from the evolution of the ratio of wage per worker for production and non-production workers.

21. The NBER productivity data set is constructed by Bartelsman, Becker, and Gray, and is available on the Internet at <http://www.nber.org>.

22. Even if demand is not perfectly inelastic, sales will go down less than output after a price increase. If inputs vary more or less in proportion to output, a price decrease will still be interpreted as a productivity gain.

TABLE 2
Summary statistics for the automobile assembly industry
 (1963, 1967, 1972–1996)

Variable	Mean	Standard deviation
Inputs		
Total hours worked per shift	22,386	17,562
Total employment	4332	3133
Production workers (% of total)	0.86	0.05
Book value of capital (% of sales)	0.334	0.293
Materials–sales ratio	0.75	0.10
Energy–sales ratio	0.006	0.004
Output		
Cars produced	181,780	92,076
Light trucks produced	130,547	81,637
Total vehicles produced	196,902	100,177
Other variables		
Only-cars dummy	0.57	0.49
Only-trucks dummy	0.23	0.42
Cars and trucks dummy	0.20	0.40
Japanese ownership dummy	0.05	0.22
Changeover dummy	0.04	0.18
Annual capacity	237,095	83,904
Shifts operated (per year)	424	114
Union dummy	0.97	0.18
Number of observations	1358	
Number of plants	78	
Number of firms	17	

Sources: LRD, Bureau of the Census, 2000; *Ward's Automotive*; and *Automotive News* weekly magazine (various years).

The Japanese model invariably fetches a higher price.²³ I collected information on the number of cars and light trucks produced by each plant. For the years 1985–1996 these data were obtained directly from *Ward's Automotive*. For the preceding years two data sources in *Ward's Automotive Yearbook* are matched.²⁴

The use of vehicles as output measure is not entirely without problems. The comparison of plants that produce different vehicles is straightforward, as I include plant dummies in the production function to control for constant productivity differences. It does take more resources to build a luxury car than it does to build a pickup truck, but as long as the substitution possibilities between inputs and productivity growth are similar, the parameters in the production functions will be estimated consistently. The comparison over time is potentially more troublesome, given the large improvements in quality cars and trucks have experienced over the last decades. In so far as the quality improvement comes from purchased components, it is controlled for in the model. Other sources of quality improvements are likely to bias Hicks-neutral productivity growth downwards, assuming that they do not change the input tradeoff a manufacturer faces.²⁵ Unobserved quality improvements that are associated with one input

23. This price difference can be substantial. NUMMI, the joint-venture between GM and Toyota in Fremont, CA, produces the Chevrolet Prizm and Toyota Corolla. Both models are identical, assembled from the same components on the same assembly line. On average, the Prizm is sold for \$3000 less than the Corolla, which is about 20% of the average retail price. Some of this represents lower profits for Chevrolet dealers, but most of it is a factory rebate. Using deflated sales or value added as output measure makes NUMMI look much more productive assembling Corollas.

24. Details about the calculations are available upon request. As a robustness check the calculated production for 1985 is compared with the information obtained directly. The correlation was a reassuring 0.99.

25. To get a sense of the magnitude of this effect, note that labour productivity growth averages only 1.3% per year over the sample period using vehicles as output measure, while it is 2.5% using deflated sales.

factor or only affect one technology have the potential to bias the production function estimates and affect the parameters in the technology choice specification, but by their very nature there is nothing I can do about it.²⁶

The input and output statistics are augmented by variables that allow us to distinguish both technologies. These include dummies for the type of vehicle produced and a dummy for Japanese ownership. A dummy for changeover years will be used as a proxy for switching costs.²⁷

Information on shifts is collected to account for capacity utilization. A major concern for productivity measurement in this industry is the volatility in capacity utilization. The Harbour report (Harbour and Associates, 1999) calculates assembly plant utilization rates for 1998, a record production year, between 34 and 148%.²⁸ Even though most of the capital cost is sunk after it is installed, many plants choose to remain idle for part of the year. The tradeoff a plant faces is to run few shifts, with many workers on each, or run a lot of them, with fewer workers. The number of shifts to operate is an explicit choice the plant makes, resulting in endogenous variation in capacity utilization. From *Automotive News* weekly magazine we obtained the number of weeks a plant ran overtime, worked on Saturdays, and the number of weeks a plant closed for vacation, for inventory adjustment, or for retooling and model changeovers. Using these five measures, the total number of shifts the plants operated each year can be computed.

4.2. Identification of two technologies

Both the trade press and theoretical arguments support the existence of *two* technologies. While the existence of two technologies is an untestable hypothesis of the model, a comparison of the characteristics of the estimated two technologies with prior expectations on the differences between lean and mass technology provides a validity check. Some descriptive evidence consistent with the two technologies assumption is also presented.

I already mentioned that a Chow test for a structural break between two different samples rejects homogeneity of technology, using a number of different criteria to sort plants. If the technology choice were observable, both production functions could be estimated separately. In practice, different criteria yield different sample splits and different production function estimates. The intrinsic unobservability of the technology choice raises an identification problem.

The discussion of switching regressions in Reiss and Wolak (2002) illustrates that one needs parametric assumptions on the unobservables to identify unobservable regimes. It is the non-normality of the reduced form errors that determines to what extent different regimes can be identified. Omitting the time dimension, the underlying structure of the economic model tells me that in any year the empirical distribution of output given inputs can be factored into two components:

$$f(q_{jt}|X_{jt}) = \Pr(i = \mathcal{M})h(q_{jt}|X_{jt}, i = \mathcal{M}) + \Pr(i = \mathcal{L})h(q_{jt}|X_{jt}, i = \mathcal{L}). \quad (6)$$

If two unobserved technologies exist, the observed, empirical distribution will be the weighted sum of two unobservable conditional distributions—conditional on technology choice—each

26. It does not imply that the introduction of machinery and equipment that improves accuracy in the assembly process—and therefore the quality of output—necessarily leads to biased estimates, because all plants in the industry tend to adopt better machinery when it becomes available. Plants differ to a greater extent in the variety of functions machines and workers are expected to perform and in the way workers interact with machinery, both of which are unlikely to affect quality directly.

27. Every 4–8 years a substantially improved or completely new model is introduced. In such changeover year a plant's downtime is substantially larger because of extensive retooling of the assembly line. A changeover year is likely to be a good time to switch technologies as well, because a significant part of the capital stock already has to be replaced.

28. Capacity utilization can exceed 100% since plants can run more than two shifts a day, 5 days a week, while capacity is calculated for regular operation.

weighted by the probability a plant operates with that technology. Functional form assumptions on the probability of employing each technology, $\Pr(i = \mathcal{M})$, will allow the untangling of the different terms in (6).²⁹ Even though it is possible to characterize the empirical distribution of output given the inputs non-parametrically, it is unlikely to be of much economic interest. In order to know how output responds to changes in exogenous variables one needs to know the partial derivatives of each term in (6). Similarly, if we want to estimate the input substitution possibilities or returns to scale that plants face, we need to distinguish the two conditional density functions.

While it is possible to identify the conditional densities solely using functional form assumptions, as in Porter (1983), exclusion restrictions aid in identification. If we observe variables that are correlated with the probability a plant operates either technology and if the correlation between output and inputs differs depending on whether one conditions on those variables or not, those variables will help identify the two conditional densities in (6). The variables in Z and W in equations (I) and (I') that are excluded from the production function play exactly that role.

A validity check on the model will be whether the estimated characteristics of the two technologies correspond to the characteristics that are generally attributed to lean and mass production. For example, the results will show that the technology choice is correlated with ownership. In particular, Japanese-owned plants are more likely to adopt the lean technology, which is plausible. Consequently, if the correlation between output and the number of shifts a plant operates differs between Japanese and domestically owned plants, it will be possible to identify different conditional densities $h(\cdot|X, \mathcal{M})$ and $h(\cdot|X, \mathcal{L})$. In particular, lean plants are thought to operate at a higher rate of capacity utilization, which will be interpreted in the production function (3) as higher returns to shifts ($\alpha_{\mathcal{L}}^s > \alpha_{\mathcal{M}}^s$), which is borne out by the estimation results.³⁰ At the same time, returns to scale are expected to be lower for the lean technology and we do find that $\hat{\alpha}_{\mathcal{L}}^l + \hat{\alpha}_{\mathcal{L}}^k < \hat{\alpha}_{\mathcal{M}}^l + \hat{\alpha}_{\mathcal{M}}^k$. Combined with the higher return to shifts, it will be optimal for lean producers to operate relatively more shifts and produce relatively less output on each, requiring a lower line-speed and less labour per shift. Given that the amount of capital to construct an assembly line varies proportionally less, the capital-labour ratio for lean producers will be high. Mass producers, on the other hand, will run fewer shifts, but produce a lot of output on each. To achieve this they will man the assembly lines with more workers, resulting in a low capital-labour ratio per shift.

More support for the two technologies assumption comes from Figure 4, which plots the distribution of an observable statistic: the capital-labour ratio per shift. It suggests, or is at least consistent with, the existence of two technologies and the possibility that plants switch between the two.

The left panels plot the non-parametric density for the ratio of capital to labour per shift in the first 5 years of the sample.³¹ The right panels plot the distribution for the same graphs in the last 5 years of the sample. The top panels contain the ratio for all plants and the bottom graphs are limited to plants that remained in the sample continuously from 1963 to 1996. The ratio has a bimodal distribution in both time periods and the popularity of each mode changed over time. In the early years, most plants choose the low capital-labour technology, the left mode. In later years, most plants prefer the technology with a higher capital-labour ratio, leading to an increase in the right mode.

29. The likelihood function in the estimated model is more complicated than (6), because the entire time-path of the endogenous variables is modelled, but the underlying idea is the same.

30. The coefficient estimates that are referred to in this section are in Table 3.

31. These graphs can be interpreted as smoothed histograms. An observation is the average capital-labour ratio for a plant over the relevant 5 year period. Confidentiality considerations preclude me from reporting the underlying statistics directly.

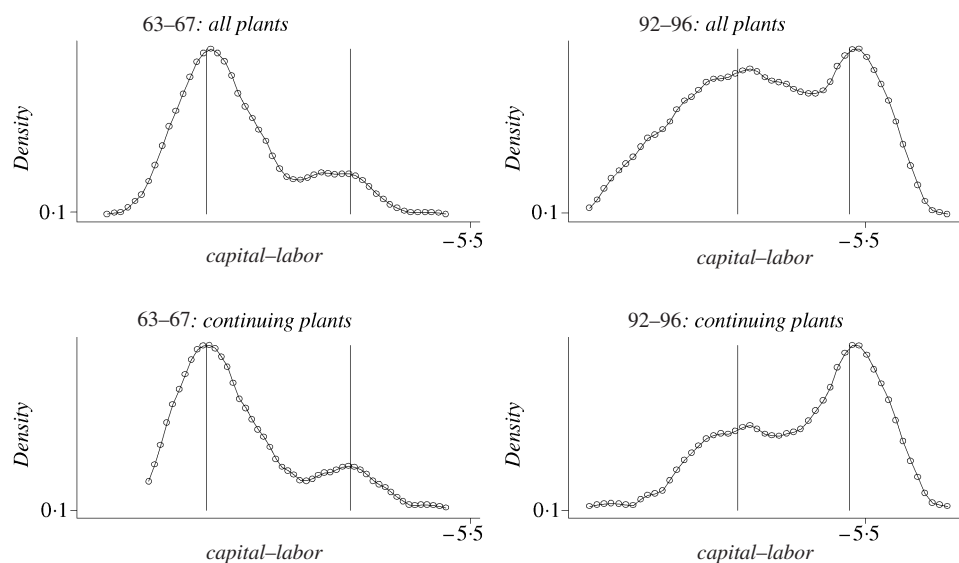


FIGURE 4

Non-parametric distribution for the average capital-labour (per shift) ratio for different groups of plants

This increase in popularity for the capital-intensive technology in the top panels is consistent with technology switching by existing plants, as well as with the entry of new plants that predominantly prefer the capital-intensive technology. However, the pattern holds up in the bottom panels. The same increase in the capital-intensive mode appears if the sample is limited to plants that operated throughout the entire sample period. This is consistent with plants changing their technology, which the model has to allow for.

Changes over time will help identify two sets of production function parameters. Comparing the left and right panels in Figure 4 reveals that in the top and bottom panels both modes have shifted to the right. The increased cost of labour has led all plants, regardless of the technology they employ, to substitute workers for machinery. It is also true that the right modes, capturing plants producing with a more capital-intensive technology and also the mode that gained in popularity over time, has shifted further to the right than the left mode. While the probability for either technology is identified by functional form assumptions and exclusion restrictions, the larger rightward shift in the right mode will show up in the production function estimates. The estimation results (later) will indicate that the new technology has a higher elasticity of substitution. It allows lean plants to respond more to the same change in relative factor prices, leading to a larger increase in capital-labour ratio for the new technology. The estimation results will also reveal that the new technology experiences a higher capital-biased technological change ($\hat{\theta}_{\mathcal{L}}^k > \hat{\theta}_{\mathcal{M}}^k$). Both characteristics of the estimated technologies are consistent with the patterns observed in Figure 4.³²

An example clearly illustrates that failure to control for the existence of different technologies will bias productivity growth estimates. Figure 5 plots simplified unit isoquants for both types of technology in input space. The producers with the lean technology have a

32. The translog specification of the production function combined with the functional form assumptions on technology switching determines to what extent differences in growth of the capital-labour ratio between mass and lean plants will be attributed to differences in capital-bias technology change or elasticity of substitution.

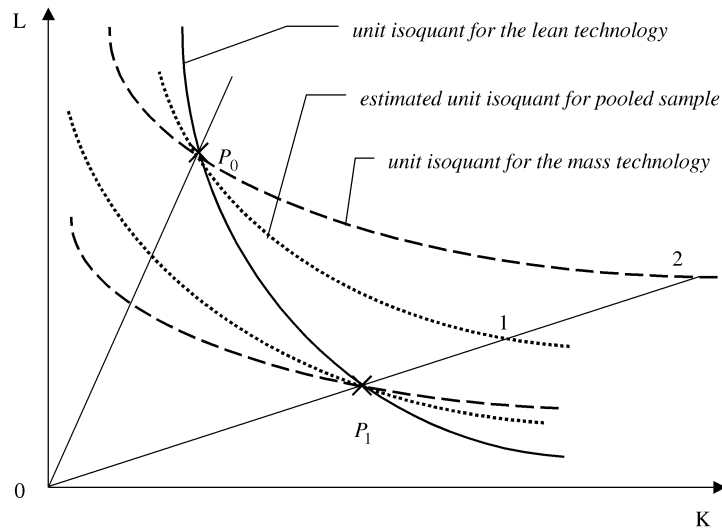


FIGURE 5

Neglecting the existence of two technologies biases productivity growth estimates

steeper isoquant (the solid line), choosing a higher capital–labour ratio at the same factor prices. Producers with the mass technology face a different tradeoff between labour and capital, captured by the dashed line. If both types of plants are pooled, the estimated technology will lie between the two existing ones and have the shape of the dotted line. Take a plant with an initial production plan at P_0 and a production plan P_1 in some later year. Without knowledge of the input tradeoff the technology allows, it is impossible to know the growth in productivity the plant experienced. We have to separate movement along the isoquant from the shift in the function. If we estimate only one production function for the pooled sample, productivity growth will be understated for plants with the mass technology. Actual productivity growth is $2P_1/02$, although it is estimated to be $1P_1/01$. For producers with the new technology, productivity growth is overstated. It is actually zero, although it is estimated to be $1P_1/01$ as well. If we believe heterogeneity in technology to exist, we have to control for it to measure productivity growth correctly.

5. ESTIMATION

5.1. Likelihood function

Equations (II), (IIIa) and (IIIb) describe the distribution of the endogenous variables—materials, output, and labour per shift—conditional on technology. The probability a plant produces with each technology was derived separately and is given by (I) and (I'). The vector $y = [m, q, \frac{L}{S}]'$ contains the endogenous variables of the system. It depends on two vectors of disturbances (one for each technology, $i = \mathcal{M}$ or \mathcal{L}), $\epsilon_i = [\epsilon_i^m, \epsilon_i^q, \omega_i^k]'$, two parameter vectors β_i , and the matrix X of exogenous variables (S, L, K, t, W, Z). A word of explanation is warranted about the exogeneity. The capital stock depends on the Hicks-neutral productivity term, ω_i^n , but not on any of the variables in ϵ_i , because it is realized earlier. S and L will depend on ω_i^k , but not on the two other error terms in ϵ_i . The dependence of $\frac{L}{S}$ on ω_i^k is explicitly modelled and used to substitute the capital-productivity from the production function. As a result, both S and L are exogenous variables at the production stage.

I assume that the three errors are independent and normally distributed. Given the nonlinear relation between the labour-shift ration and ω_i^k in (II), the Jacobian for the transformation of variables has to be derived,

$$\frac{\partial \omega_i^k}{\partial L/S} = \left(\frac{L}{S}\right)^{-1} (\lambda_i + \lambda_i^2 - \beta_i^l). \quad (7)$$

The conditions for convexity of the production function also guarantee that the relationship between ω_i^k and $\frac{L}{S}$ is monotone. λ_i represents the share of labour in the variable cost, derived in Section 3.2. It is a function of observed variables and technology-specific parameters.

The joint density of the y_{jt} -vector, conditional on technology i , is

$$h(y_{jt}|i) = \frac{1}{s_i^m} \phi\left(\frac{m_{jt} - g^m(X_{jt}; \beta_i)}{s_i^m}\right) \times \frac{1}{s_i^q} \phi\left(\frac{q_{jt} - g^q(X_{jt}; \beta_i)}{s_i^q}\right) \\ \times \left| \left(\frac{L_{jt}}{S_{jt}}\right)^{-1} (\lambda_{ijt} + \lambda_{ijt}^2 - \beta_i^l) \right| \frac{1}{s_i^k} \phi\left(\frac{g_i\left(\frac{L_{jt}}{S_{jt}}\right) - g^l(X_{jt}; \beta_i)}{s_i^k}\right), \quad i = \mathcal{M}, \mathcal{L}$$

where $\phi(\cdot)$ is the standard normal density and s_i^x is the variance of ϵ_i^x . The parametric form for the functions g^m , g^q , g^l , $g_{\mathcal{M}}$, and $g_{\mathcal{L}}$ are given by equations (II), (IIIa), and (IIIb). I assume further that, conditional on the technology choice, errors are uncorrelated over time. This is the natural assumption, even for ω_i^k , given the interpretation of the error terms. The shock to capital-productivity only contains the difference between the realized productivity and the plant's expectation. It is the higher or lower productivity of the capital stock that was not anticipated and optimized against. If plants do not make systematic errors, the realizations of ω_i^k will not be serially correlated. The density of y_j for the entire sample becomes³³

$$H(y_{j1}^i \dots y_{jT}^i) = h(y_{j1}|i) \dots h(y_{jT}|i).$$

These densities cannot be estimated directly if the technology choice is not observed. For each sequence $(y_{j1}, y_{j2}, \dots, y_{jT})$ we sum over all possible technology paths it can represent. For example, the probability of the sequence $(y_{j1}^{\mathcal{M}}, y_{j2}^{\mathcal{M}}, y_{j3}^{\mathcal{L}}, \dots, y_{jT}^{\mathcal{L}})$ occurring is $\psi_{j1} p_{j2} (1 - p_{j2})$. The plant started with the mass production technology, stuck with it after the first year, and switched to the new one after the second year. The probability for a sequence y_j —for a plant present in the sample from time 1 to time T —is

$$\begin{aligned} \text{LH}(y_{j1}, y_{j2}, \dots, y_{jT}) = & (1 - \psi_{j1}) H(y_{j1}^{\mathcal{L}}, y_{j2}^{\mathcal{L}}, \dots, y_{jT}^{\mathcal{L}}) \\ & + \psi_{j1}(1 - p_{j2}) H(y_{j1}^{\mathcal{M}}, y_{j2}^{\mathcal{L}}, \dots, y_{jT}^{\mathcal{L}}) \\ & + \psi_{j1} p_{j2}(1 - p_{j3}) H(y_{j1}^{\mathcal{M}}, y_{j2}^{\mathcal{M}}, y_{j3}^{\mathcal{L}}, \dots, y_{jT}^{\mathcal{L}}) \\ & + \dots \\ & + \psi_{j1} p_{j2} \dots (1 - p_{jT}) H(y_{j1}^{\mathcal{M}}, y_{j2}^{\mathcal{M}}, \dots, y_{jT-1}^{\mathcal{M}}, y_{jT}^{\mathcal{L}}) \\ & + \psi_{j1} p_{j2} \dots p_{jT} H(y_{j1}^{\mathcal{M}}, y_{j2}^{\mathcal{M}}, \dots, y_{jT-1}^{\mathcal{M}}, y_{jT}^{\mathcal{M}}). \end{aligned}$$

ψ_{j1} is the probability a plant enters the sample with the mass technology and p_{jt} is the probability a plant remains with the mass technology in period t , given that it had this technology in period $t-1$. The starting probability is a function of variables W and the transition probabilities depend on variables Z . Substituting the previous expressions for $H(\cdot)$ and $h(\cdot|i)$ and multiplying over all plants generates the full likelihood function.

33. I write y_{jt}^i to denote the distribution of y_{jt} conditional on producing with technology i in year t .

5.2. Results

We estimate equations (II), (IIIa), (IIIb) and expressions (I) and (I') using the likelihood function in (8). This controls for unobserved heterogeneity in technology by jointly estimating the starting and switching probability for each technology with the parameters in both production functions. Endogeneity of productivity is accounted for by including plant-fixed effects and substituting ω_i^k in the production function using the labour per shift equation. All coefficients are estimated jointly and I impose cross-equation restrictions. Estimation results are in Table 3.

The set of coefficients in the first column represents the mass technology, with standard errors in the second column. The coefficients in the third column represents the difference between the parameters for both technologies. These differences are estimated directly and their standard errors are reported in the fourth column. The coefficients for the lean—absorbing—technology are at the far right, in the fifth column. They are not estimated directly, but obtained by summing the coefficients in the first and third columns.

The top panel contains the linear and quadratic input coefficients in the production function, the parameter for returns to shifts, Hicks-neutral and capital-biased productivity growth, and the estimated standard deviation for the shock to production. Most of the standard errors for the difference terms (fourth column) are surprisingly small. The two technologies are estimated to be significantly different. The shape of the estimated production functions is interpreted in the next section. The mass technology is associated with a higher rate of Hicks-neutral productivity growth. The lean technology, on the other hand, experiences a very high rate of capital-biased productivity growth, but Hicks-neutral productivity growth is slightly negative. In Section 6 we discuss the impact of these estimates on industry-wide productivity growth.

The second panel contains the coefficients in the materials equation (IIIb). Although there are no cross-equation restrictions, we estimate it jointly with the other equations, because it helps to predict technology choice if the amount of materials per vehicle differs by technology. Quality upgrading through improved components is estimated to be slightly lower for the lean technology. The parameter on output is indistinguishable between technologies, indicating both are to a similar degree outsourced. An increase in output is associated with a less than proportional increase in material input. The larger standard deviation for the error terms in the material and output equations indicates that the endogenous variables are less well predicted for the lean technology. The lower emphasis on standardization in this technology leads to larger variations between plants.

The cost-parameters and standard deviation for the labour per shift equation are in the third panel. The parameters determining the marginal product of labour were reported earlier, in the first panel. The dummy for labour unions is assumed constant across technologies. It is estimated at 0.97, meaning that non-unionized plants save an extra 3% of the wage rate, compared to unionized plants, when they idle the assembly lines, but the difference is not significant. It indicates that the tradeoff between labour and shifts does not vary by unionization status. Both the fixed and variable costs related to shifts are estimated to be lower for the lean technology. The variable component is not significantly different from zero for lean producers. It will lead lean plants to operate more shifts, reinforcing the effect of higher returns to shifts, already reported earlier.

In the bottom panel are the coefficients governing the technology choice. For the starting probability, the results indicate that plants are gradually becoming more likely to start out with the lean technology. The coefficient on time is estimated at 0.09, but not significantly different from zero. The sign is positive, as expected. Because the probability a plant operates with the mass technology is modelled with the logit formula, the marginal effect is $\frac{\partial \Pr(i=\mathcal{M})}{\partial \text{year}} = -\eta_y \Pr(1 - \Pr)$. If a new plant is equally likely to operate either technology in 1 year, the

TABLE 3
Estimation results for the three equations and technology choice

		Mass technology		Difference		Lean technology
		$\hat{\beta}_{\mathcal{M}}$	(S.E.)	$\hat{\beta}_{\mathcal{L}} - \hat{\beta}_{\mathcal{M}}$	(S.E.)	$\hat{\beta}_{\mathcal{L}}$
Production function:						
Shifts	α_i^S	0.9684	(0.004)	0.1334	(0.034)	1.1017
Labour	α_i^L	0.8222	(0.023)	0.0224	(0.023)	0.8447
Capital	α_i^K	0.1364	(0.039)	-0.0325	(0.021)	0.1039
Labour squared	β_i^L	-0.0148	(0.037)	0.0931	(0.037)	0.0783
Capital squared	β_i^K	0.0268	(0.012)	-0.0204	(0.012)	0.0065
Labour \times capital	β_i^{Lk}	0.1483	(0.019)	-0.1262	(0.019)	0.0221
Capital-biased PG	θ_i^K	0.0706	(0.019)	0.2533	(0.036)	0.3239
Hicks-neutral PG	θ_i^N	0.0176	(0.005)	-0.0269	(0.007)	-0.0093
Standard deviation	s_i^q	0.2353	(0.009)	0.3132	(0.097)	0.5485
Materials equation:						
Constant term	α_i^0	3.6582	(0.270)	0.0003	(0.313)	3.6586
Output	$\frac{1}{\alpha_i^m}$	0.8571	(0.022)	0.0099	(0.026)	0.8670
Quality upgrading	θ_i^m	0.0210	(0.002)	-0.0075	(0.003)	0.0135
Standard deviation	s_i^m	0.2231	(0.008)	0.2257	(0.008)	0.4488
Labour-shifts equation:						
Capital cost	ρ_i	2.9E-6	(5.5E-6)	-2.9E-6	(3.5E-4)	1.4E-11
Fixed cost per shift	μ_i	67.483	(11.4)	-10.021	(11.3)	57.462
Union dummy	δ	0.9745	(0.123)	X	X	0.9745
Standard deviation	s_i^k	0.0644	(0.008)	-0.0340	(0.005)	0.0304
Technology choice:						
Starting probability: η						
Constant term		-2.0345	(0.613)			
Dummy for cars		0.6853	(0.255)			
Dummy for trucks		1.1867	(0.427)			
Japanese dummy		1.0404	(0.281)			
Time		0.0911	(0.046)			
Transition probability: γ						
Constant term		-1.9412	(0.755)			
Dummy for cars		-0.1328	(0.204)			
Dummy for trucks		-0.7739	(0.723)			
Changeover dummy		0.4080	(0.303)			
Time		-0.0383	(0.057)			
Number of observations		1358				
Log-likelihood		1890.8				

probability a plant starting the next year operates with the lean technology is 2.25% higher. The increase in probability for the lean technology slows down when it becomes higher, but Figure 6 illustrates that the probability new plants employ the mass technology decreases fast. Plants that produce only trucks, and to a lesser degree plants producing only cars, prefer the lean technology, although the coefficients are not estimated very precisely. Surprisingly, the new technology seems

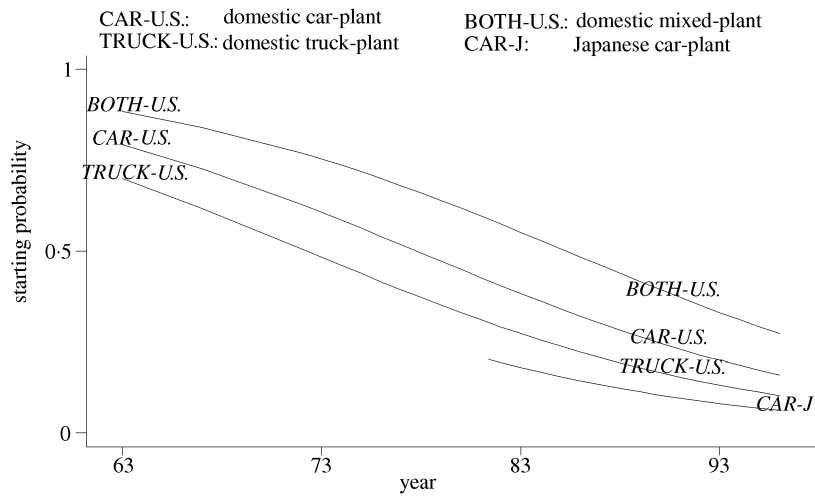


FIGURE 6

Probability that a new plant starts out with the mass technology

to be less favoured by plants that produce both trucks and automobiles. Japanese producers are more likely to choose the lean technology. The average starting probability for the mass technology is 0.81 at the beginning of the sample period and 0.23 at the end.

Figure 6 plots the starting probability for different types of plants. The three dummy variables in Z —only-cars, only-trucks, Japanese ownership—define six types of plants, all of which face a different probability for the mass technology in each year. For readability, only the schedule for four of the six types are plotted in Figure 6. The positive coefficient on time makes the probability for the mass technology decline over the years. The positive coefficient estimate on Japanese ownership shifts the schedule for each vehicle category down. The first Japanese plants entered the sample in 1982.³⁴

For the transition probability, the negative coefficient on time indicates that the probability for a mass plant switching to the lean technology declines over the sample period. During a year in which a plant has a major changeover, it is more likely to adopt the new technology as well. I reported earlier that plants producing both cars and trucks had the highest probability to start out with the mass technology. The results also indicate that they have the highest probability to make the transition to lean. Figure 7 traces the inverse of the transition probability, the probability that a mass production plant remains with the old technology, in regular and changeover years for the different types of plants. The average switching probability over the entire sample period is 0.07. This ranges from a low of 0.02 for a truck producer if 1996 was a regular year, to a high of 0.18 for a car and truck producer if 1963 was a changeover year. These statistics can be interpreted as hazard rates. The average transition probability suggests that over a 10 year period half of the mass producers would switch to the lean technology. In 1963 it would only take three and a half changeover year to achieve the same percentage of mixed-plants switching. A small increase in the transition probability has large effects.

34. The three joint ventures between American and Japanese producers, NUMMI, AutoAlliance and Diamond-Star, are considered Japanese plants.

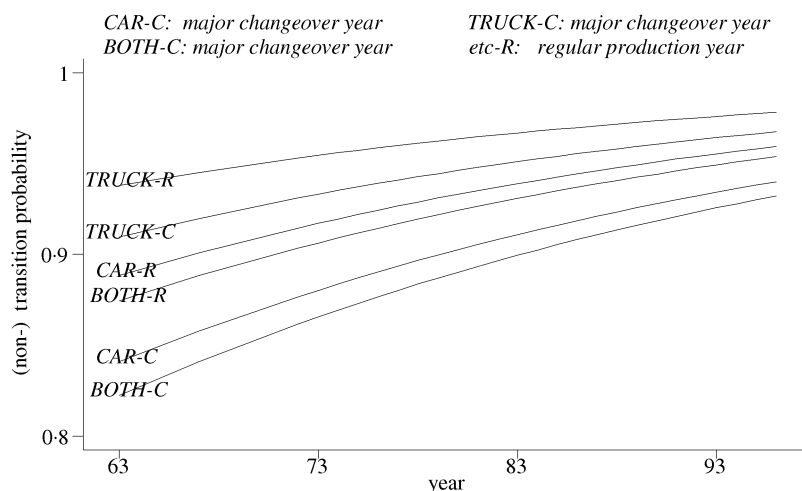


FIGURE 7

Probability a mass production plant does not switch to the lean technology

5.3. Interpretation

I draw two conclusions about technology from the estimation results. First, the two technologies are estimated to be rather distinct. The characteristics of both production functions that I estimate are consistent with an interpretation of the old technology (the one plants switch out of) as mass production and the new technology (the one modelled as an absorbing state) as lean production. Second, the proportion of plants producing with the mass technology declines significantly over the sample period. This is caused by a combination of entry by new plants, which are more likely to be lean, and technology switching by existing plants.

I do find evidence for two distinct technologies in the sample. The nature of their difference corresponds largely to the mass–lean distinction many industry observers have made. Interpretation of the parameters in the production function is complicated by the quadratic terms. In Table 4, I evaluate several characteristics of the technology, evaluated at the sample mean which is calculated separately for each technology. Plants are weighted by the imputed probability they operate with each technology. Table 4 contains estimates for factor shares, returns to scale, productivity growth, and elasticities of substitution. The first two columns contain the results for the two estimated technologies, mass and lean. For comparison, the far right column presents the same statistics for a translog production function, estimated on the full sample. The translog estimation uses the same output and input measures and includes plant-fixed effects as well.³⁵

The first two rows contain the labour and capital share in costs per shift (not including the cost of operating the shift), evaluated at the mean for each technology. The capital share does not contain the variable costs of operating a shift, which are proportional to the capital stock. It makes comparison between my results and the regular translog estimation less straightforward. The mass technology puts more weight on both inputs and produces with constant returns to scale. The lean technology has a relatively higher weight on capital than labour, as could be expected from the higher capital–labour ratio in Figure 4. Its capital share is 7.3 times higher than the labour share, while for mass technology it is only 5.9 times higher. Returns to scale are noticeable decreasing for the lean technology. This can be the result of less standardization.

35. Failure to correct for capacity utilization led invariably to a negative capital share. I approximated the capital services a plant derives from the observed capital stock by multiplying capital with an index of capacity utilization.

TABLE 4
Comparing both estimated technologies with regular translog results

	Mass technology	Lean technology	Translog estimation for the entire sample
Output per shift:			
Capital share in cost per shift	0.144	0.104	0.349
Labour share in cost per shift	0.854	0.764	0.796
Returns to scale (L & K)	0.998	0.868	1.145
Total factor shares:			
Labour share in variable cost	0.882	0.695	X
Shift share in variable cost	0.151	0.178	X
Returns to shifts	0.968	1.102	X
Productivity growth:			
Hicks-neutral	0.018	-0.009	0.004
Capital-biased	0.071	0.324	0.007
Labour-biased	X	X	-0.006
Elasticity of substitution (L-K)	1.038	1.156	0.334
Demand elasticities:			
ϵ_{LL}	-0.154	-0.884	-0.117
ϵ_{KK}	-0.861	-0.977	-0.266
ϵ_{LK}	0.149	0.120	0.117
ϵ_{KL}	0.886	0.883	0.265

Alternatively it can be caused by the practice of lean plants to assemble more than one model on the same assembly line. At the model-level returns to scale can be constant, while they are decreasing for the entire line.

The labour and shift shares in variable cost have an additional correction for the returns to shifts. The correct formula and interpretation is given by equation (5). Variable costs include fixed and variable costs associated with operating shifts. For the lean technology, returns to shifts are estimated to be increasing. As a result, inputs have to increase proportionally less to increase output and the factor shares sum to less than one. The lower variable cost of operating a shift for the lean technology ($\rho_{\mathcal{L}}$ was estimated to be very low) and the increasing returns to shifts both cause lean producers to operate at a higher level of capacity utilization. The higher number of shifts operated means that the share of variable costs that goes towards shifts is not that different between both technologies. The mass producers will spend relatively more on workers, 5.8 times as much as on shifts compared with only 3.9 times as much for lean producers, confirming the previously found low capital-labour ratio for the mass technology.

Hicks-neutral and capital-biased productivity estimates differ considerably between the two technologies, while labour-biased productivity growth is restricted to zero. The high rate of capital-biased productivity growth for lean producers is remarkable, but not surprising, given that lean production is associated with flexible machinery. In recent years, advanced equipment has allowed some Japanese plants to produce rather distinct models on the same assembly line. It also causes changeover times for the assembly lines between models to be much lower in Japanese plants. As could be predicted from Figure 5, Hicks-neutral productivity growth for the regular translog estimation, pooling all plants, is estimated to lay between productivity growth for lean and mass producers. One should use caution to compare the factor-biased productivity growth rates in my model with the results for the translog estimation. The latter does not have the same structural interpretation, because interaction between inputs will affect the productivity estimates as well.³⁶

36. As a rule of thumb, one can divide the factor-biased growth rates by the factor share, to be comparable with the results based on the translog estimates.

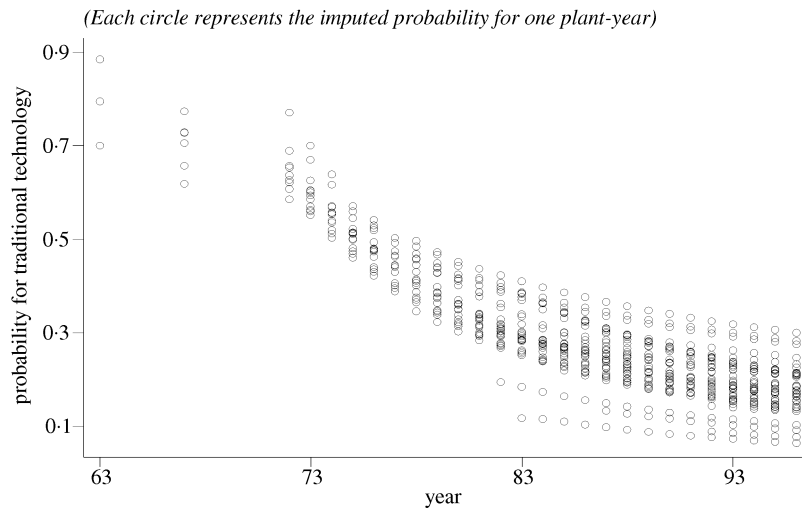


FIGURE 8
Probability for mass technology in the sample

The bottom panel in Table 4 contains the estimates for the elasticity of substitution between capital and labour. The lean technology displays a slightly higher elasticity of substitution, exceeding one for both technologies. Comparing these results with the translog estimation indicates that the failure to model capacity utilization explicitly results in an implausibly low elasticity estimate. The same finding holds for the factor-demand elasticities. The lean technology is estimated to be more flexible than the mass technology. Labour demand especially, is estimated to be significantly more elastic. Both technologies are estimated to be more price responsive than the results for the simple translog estimation indicate.

The second conclusion I draw from the analysis is that the industry as a whole has almost completed the transition from mass to lean production. I calculate the probability for the mass technology in each year for each of the 78 plants in the sample by updating the starting probability with the relevant transition probabilities for each plant. Figure 8 shows that the proportion of plants in the sample that are likely to produce with the mass technology declines over time, but in any given year there is considerable variation. The probability for the lean technology was very small for all plants in 1963, but the reverse is true in 1996.

The decline for mass technology is caused by two trends. The starting probabilities in Figure 6 already indicated that new plants were gradually more likely to use the lean technology. At the end of the sample period, every existing plant has a lower than 0.4 probability of using the mass technology, with the probability significantly lower for some types of plants. Of the 49 plants in the sample in 1963 only half remain in 1996. The average entry year for the other plants present at the end of the sample is 1983. In that year the average starting probability for the mass technology was already as low as 0.42.

The second trend leading to the disappearance of mass production is technology switching by existing plants. For example, car-plants faced a switching probability of around 0.1 in the first part of the sample, even in regular production years. This translates into half of the plants making the technology switch in less than 7 years. In addition, switching is more prevalent in changeover years, which were more common earlier in the sample when switching probabilities were higher to start with. The average year for all observations in the data set is 1983, while the average year for a significant model changeover is 1980. The results confirm the conclusion

from the bottom two panels of Figure 4 that technology switching is important. Figure 7 indicated that the non-transition probability tends to one towards the end of the sample. Gradually, fewer plants with the “wrong” technology, that choose to make a costly technology switch, remain. It makes the probability for the mass technology for any given plant asymptote to a value above zero, although Figure 8 indicates that this value is rather low for many plants. For some plants, especially American-owned truck producers that had few changeover years and entered the sample early on, the mass technology is the optimal production technology, even at the end of the sample.

6. IMPLICATIONS FOR AGGREGATE PRODUCTIVITY GROWTH

Finally, I investigate what the model teaches us about productivity growth in the industry. The estimated productivity growth rates are shifts of the production functions, with the function for each technology shifting independently. The mass technology experiences a higher rate of Hicks-neutral productivity growth, 1.8 vs. -0.9% for the lean technology. The lean technology, on the other hand, has a higher capital-biased productivity growth rate, 32.4 vs. 7.1%. The capital-biased results are not directly comparable to the usual estimates obtained from a translog cost or production function. The growth rates I estimate only affect total output or costs in proportion to the share of capital and through the interaction between capital and labour. The interpretation is the same though: productivity growth is labour-saving. Over time it will lead firms to substitute capital for labour. The strong upward trend of the real wage has a similar effect. Both trends help explain the observed increase in vehicles per worker for the industry in Figure 1. Given the evolution of wages, the higher rate of labour-saving growth reinforces the benefits of lean technology, which explains its increased popularity over time.

The estimated technologies are used to analyse the evolution of productivity in the industry. Figure 1 suggested a break in the trend growth rate for labour productivity in the early 1980s, when productivity growth accelerated strongly. Instead of modelling this as an exogenous shift, I calculate the impact of trends in fundamentals on the observed industry-wide productivity growth. A first decomposition divides aggregate labour productivity growth for the industry into contributions of lean producers, mass producers, and a switching effect. Relocation of resources between plants and entry and exit play an important role as well. A second decomposition divides the growth in labour productivity for each technology into the contribution of several effects that are estimated in the model.

Following Baily, Hulten and Campbell (1992), I decompose aggregate labour productivity growth for the industry into a term measuring plant-level productivity change, a term capturing the relocation of inputs between plants and two terms capturing the effect of entry and exit. The decomposition was improved for unbalanced panels by Haltiwanger (1997), which introduced an extra covariance term:

$$\begin{aligned} \Delta \overline{LP}_t &= LP_t - LP_{t-1} = \sum_{n=j,k}^{\text{stay, enter}} \vartheta_{nt} LP_{nt} - \sum_{n=j,l}^{\text{stay, exit}} \vartheta_{nt-1} LP_{nt-1} \\ &= \sum_j^{\text{stay}} \vartheta_{jt-1} \Delta LP_{jt} + \sum_j^{\text{stay}} \Delta \vartheta_{jt} (LP_{jt-1} - LP_{t-1}) + \sum_j^{\text{stay}} \Delta \vartheta_{jt} \Delta LP_{jt} \\ &\quad + \sum_k^{\text{enter}} \vartheta_{kt} (LP_{kt} - LP_{t-1}) - \sum_l^{\text{exit}} \vartheta_{lt-1} (LP_{lt-1} - LP_{t-1}) \end{aligned} \quad (8)$$

$LP = \log\left(\frac{Q}{L}\right)$ is the logarithm of the level of labour productivity. The entire expression measures the average labour productivity growth, weighing each plant by its share in industry employment.

TABLE 5

Decomposition of industry-wide labour productivity growth

	Entire sample average: 1963–1996			After Japanese entry: 1982–1996		
	Lean plants	Mass plants	Switching	Lean plants	Mass plants	Switching
$\Delta \bar{LP}$ (aggregate)	(1.06%)	1.28% (0.20)	(0.01)	(2.25%)	2.51% (0.14)	(0.11)
Plant-level ΔLP	0.80%	0.55%	0.03%	1.85%	0.88%	0.09%
Relocation	0.44%	0.17%		0.90%	–0.24%	
Covariance	–0.60%	0.01%	–0.03%	–1.04%	0.21%	0.02%
Net entry	0.42%	–0.53%		0.53%	–0.71%	

I use these weights to investigate the reallocation of labour over time.³⁷ The first term measures the contribution of labour productivity growth at the plant-level, only calculated for plants that stay in the sample from year $t - 1$ to year t . The second term measures the relocation effect, again only for plants that stayed in the sample. If inputs are relocated to plants with above average labour productivity, this term will be positive. The third term is more difficult to interpret as it measures the correlation between productivity growth and input growth. It will be positive if productivity improvements are more likely to be found in expanding plants or if plants that downsize are more prone to experience negative productivity growth. The fourth and fifth terms measure the contribution of plants entering and exiting the sample. If new plants are relatively more productive than exiting plants, the sum of the last two terms will be positive.

For each plant I also decompose the level and growth of labour productivity into the contribution of the mass and lean technology:

$$\Delta LP_{jt} = \underbrace{\varphi_{jt} \Delta LP_{jt}}_{\text{mass}} + \underbrace{(1 - \varphi_{jt-1}) \Delta LP_{jt}}_{\text{lean}} + \underbrace{(\varphi_{jt-1} - \varphi_{jt}) \Delta LP_{jt}}_{\text{switching}} \quad (9)$$

$$LP_{nt} = \underbrace{\varphi_{nt} LP_{nt}}_{\text{mass}} + \underbrace{(1 - \varphi_{nt}) LP_{nt}}_{\text{lean}}, \quad n = j, k, l \quad (10)$$

where the weight, φ_{jt} , is the probability for the mass technology for plant j in year t . The first term in (9) captures the contribution to the mass technology, multiplying the labour productivity growth at plant j with the probability it is producing with the mass technology in year t . The second term measures the similar contribution to lean technology, multiplying labour productivity growth with the probability plant j was lean in both years. The third term multiplies the productivity growth with the probability a plant made the technology switch at the start of year t . Equation (10) similarly decomposes the level of labour productivity into a mass and lean component.

Substituting equations (9) and (10) in decomposition (8) gives 12 terms. I sum the contribution of entry and exit for each technology to calculate a net entry effect. Table 5 contains the results for all 10 terms, averaged over all years in the sample.

The lean technology generates the bulk of industry-wide productivity growth. Plants using this technology increase productivity faster and new plants are more likely to be lean and have higher productivity than exiting plants. The relocation of labour between plants that remained

37. Haltiwanger (1997) uses output shares as weights, but it is more intuitive to use denominator shares to aggregate plant-level productivity, because aggregate labour productivity, $LP' = \frac{\sum_i Y_i}{\sum_i L_i}$, can be written as $LP' = \sum_i \frac{L_i}{\sum_i L_i} \frac{Y_i}{L_i}$. Using logarithms destroys the exact relationship between aggregate and plant-level measures, but it is unavoidable to make the linear decomposition. Obviously, it has to hold that $\sum_j \vartheta_{jt} + \sum_k \vartheta_{kt} + \sum_l \vartheta_{lt} = 1$.

in the sample generates a positive contribution as well, indicating that workers move over time to plants with a higher productivity level. This effect is larger for lean plants than for the mass technology plants. The only negative contribution for lean plants comes from the covariance term. Productivity growth is more likely to occur in plants that decrease their workforce. The next decomposition, see Table 6, confirms that lean plants have exploited diseconomies of scale by reducing their size. If a mass technology plant switches technology, the size of the resulting lean technology plant is likely to exceed the optimal scale, given the lower scale economies estimated for the new technology.

The results on plant-level growth and relocation are similar, but smaller in magnitude, for mass technology plants. One different result is the negative contribution of net entry. It can be the result of entering mass plants being less productive than the sample average or exiting mass plants being relatively more productive. Another possibility is that highly productive mass plants diminish in importance from the moment of entry to the point of exit, caused by a reduced labour share or a reduction in the probability the plant operates with the mass technology.

The direct effect of switching, in the third column, is minor. Changes in probability for each technology in any given year are small for most plants. In addition, lack of experience is likely to make the newly configured plant operate with low productivity for some time immediately after a switch in technology.

The last three columns in Table 5 show the same results for the last 14 years of the sample, the period following the entry of Japanese firms. The results are very similar to the decomposition for the entire sample period, but the growth rates for lean plants are substantially larger and the difference between mass and lean technology is more pronounced. One new finding is the negative contribution of relocation for mass plants: labour is relocated from more to less productive plants. This inverse relocation effect is also present using capital as aggregation weight. The initial reaction of some manufacturers after the Japanese entry was to move extra resources, labour and capital, to plants with low productivity, relatively reallocating resources to below average productivity plants. This strategy proved not very effective and many plants, often recently refurbished ones, were closed permanently during the recession in the early 1990's. After the recession, new plants were built and the estimates suggest that the probability that those plants employed the mass technology was very small.

A second decomposition divides the growth rate of labour productivity for each technology into the contribution of underlying effects that were estimated. I decompose the effects by approximating the estimated production function by a Cobb–Douglas function,

$$Q = S^{\alpha^s} \left[\left(\frac{L}{S} \right)^{\alpha^l} (K e^{\theta^k t})^{\alpha^k} \right] e^{\theta^n t}.$$

Taking first differences of the logarithm of this function, deducting growth in labour input, and substituting the coefficient estimates, I can write labour productivity growth for each plant as

$$\begin{aligned} \Delta LP = (\dot{q} - \dot{l}) = & \underbrace{\hat{\theta}^n}_{\text{Hicks-neutral PG}} + \underbrace{\hat{\theta}^k \hat{\alpha}^k}_{\text{capital-biased PG}} + \underbrace{\hat{\alpha}^k (\dot{k} - \dot{l} + \dot{s})}_{\text{capital-labour growth}} \\ & + \underbrace{(\hat{\alpha}^l + \hat{\alpha}^k - 1)(\dot{l} - \dot{s})}_{\text{returns to scale}} + \underbrace{(\hat{\alpha}^s - 1)\dot{s}}_{\text{returns to shifts}} + \epsilon, \end{aligned} \quad (11)$$

where \dot{x} denotes the year on year growth rate for x . The last term captures the change in errors between years and the approximation error caused by the use of the Cobb–Douglas approximation.

I sum each of the five terms over all plants, using the probability for each technology as weight. All plant-years receive equal weight, as they did in the estimation. To make the

TABLE 6

Decomposition of labour productivity growth into fundamentals for each technology

	Lean technology		Mass technology	
	Growth (%)	Contribution (%)	Growth (%)	Contribution (%)
$\Delta \overline{LP}$	1.46	100	1.11	100
Growth in $\frac{K}{L/S}$	0.58	39	-0.01	-1
Hicks-neutral PG	-0.29	-20	0.69	62
Capital-biased PG	1.04	71	0.39	36
Returns to scale	0.08	6	0.05	4
Returns to shifts	0.05	4	-0.01	-1

decomposition add up (given the error term), I use the relative importance of the Hicks-neutral and capital-biased productivity growth estimates to divide the part of labour productivity growth not accounted for by the other three terms. Results for both technologies are in Table 6.

For the mass technology, almost all of the growth in labour productivity is Hicks-neutral and capital-biased productivity growth. For the lean technology most of the productivity growth comes from capital-biased productivity growth, unsurprising, given the high estimate for this coefficient. In contrast with the mass technology, there is also a sizable contribution from the increase in capital-labour ratio (for labour per shift) as well. Total factor productivity, which can be calculated by deducting the growth in capital-labour ratio from the labour productivity growth (deduct the second line from the first in Table 6), hardly differs between the two technologies. Returns to scale are decreasing for the lean technology, but it generates a positive contribution because the average scale of operation decreased. Many lean plants have become smaller over time, as a result of the proliferation of models produced. This trend in demand increased the popularity of the lean technology, since it is better suited to produce small production runs. Some plants now produce more than one model on the same assembly line, which again favours the more flexible, lean technology. The increase in capacity utilization, plants are operated more shifts, makes a small positive contribution to labour productivity growth for lean technology.

7. FINAL CONCLUSIONS

I estimated productivity growth using a structural model of production, accounting for unobserved heterogeneity in technology and productivity. The estimated technologies are consistent with the often made distinction between mass and lean production in this industry. The more recent (lean) technology is associated with higher capital-biased and lower Hicks-neutral productivity growth. I also find that the mass production technology is disappearing from the industry. It is caused by the entry of new plants, predominantly choosing the new technology, and technology switching by existing plants.

Using the estimation results, I investigate the trends underlying the large increase in aggregate labour productivity growth for the industry since the early 1980's. Plant-level growth, attributed mostly to lean producers, and the net entry of plants with the new technology are the two most important components of aggregate productivity growth. The plant-level labour productivity growth can be further decomposed. Capital-biased productivity growth and an increase in the capital-labour ratio are particularly important for the lean technology. Surprisingly, diseconomies of scale contributed positively as well, because plants decreased their scale of operation over the sample period.

Acknowledgements. I wish to acknowledge a large debt to Frank Wolak who went much beyond the call of duty as my thesis advisor providing comments and suggestions and improving all aspects of the research project. I also benefited from comments from Lanier Benkard, Peter Reiss, two referees, the editor, and numerous seminar participants at Stanford University and elsewhere. I am grateful to Mike Clune at the CCRDC for help preparing the data. Financial support from the Littlefield Fellowship, the GRO program at Stanford University, and the NSF under grant SES-9906403 is gratefully acknowledged. The research in this paper was conducted while the author was a Census Bureau research associate at the California Census Research Data Center. Research results and conclusions expressed are those of the author and do not necessarily indicate concurrence by the Bureau of the Census. This paper has been screened to insure that no confidential data are revealed.

REFERENCES

- BAILY, M. N., HULTEN, C. and CAMPBELL, D. (1992), "Productivity Dynamics in Manufacturing Plants", *Brookings Papers: Microeconomics*, **4** (1), 187–267.
- BEARD, R. T., CAUDILL, S. B. and GROPPER, D. M. (1991), "Finite Mixture Estimation of Multiproduct Cost Functions", *Review of Economics and Statistics*, **73** (4), November, 654–664.
- BENKARD, C. L. (2000), "Learning and Forgetting: The Dynamics of Aircraft Production", *American Economic Review*, **90** (4), September, 1034–1054.
- BLUNDELL, R. W. and BOND, S. R. (2000), "GMM Estimation with Persistent Panel Data: An Application to Production Functions", *Econometric Reviews*, **19** (3), 321–340.
- DIAMOND, P., McFADDEN, D. and RODRIGUEZ, M. (1978), "Measurement of the Elasticity of Factor Substitution and Bias of Technical Change", in M. Fuss and D. McFadden (eds.) *Production Economics: A Dual Approach to Theory and Applications*, Vol. 2 of *Contributions to Economic Analysis*, Vol. 111 (New York: North-Holland) 125–147.
- DUBIN, J. A. and McFADDEN, D. L. (1984), "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption", *Econometrica*, **52** (2), March, 345–362.
- FRIEDLAENDER, A. F., WINSTON, C. and WANG, K. (1983), "Costs, Technology, and Productivity in the U.S. Automobile Industry", *Bell Journal of Economics*, **14** (1), 1–20.
- FUSS, M. A. and WAVERMAN, L. (1992) *Cost and Productivity in Automobile Production: The Challenge of Japanese Efficiency* (Cambridge: Cambridge University Press).
- GRIFFITH, R. (1999), "Using the Ard Establishment Level Data to Look at Foreign Ownership and Productivity in the UK", *Economic Journal*, **109** (456), June, 416–442.
- HALTIWANGER, J. C. (1997), "Measuring and Analyzing Aggregate Fluctuations: The Importance of Building from Microeconomic Evidence", *Federal Reserve Bank St. Louis Review*, **79** (3), 55–77.
- HAMILTON, J. D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, **57** (2), March, 357–384.
- HARBOUR AND ASSOCIATES (1999) *The Harbour Report: Competitive Assessment of the North American Automotive Industry* (Rochester, MI: Harbour and Associates).
- KLETTE, T. J. and GRILICHES, Z. (1996), "The Inconsistency of Common Scale Estimators when Output Prices are Unobserved and Endogenous", *Journal of Applied Econometrics*, **11** (4), 343–361.
- KRAFCIK, J. F. (1988), "Comparative Analysis of Performance Indicators at World Auto Assembly Plants" (Master's Thesis, MIT, Cambridge).
- LEVINSOHN, J. and PETRIN, A. (1999), "When Industries Become more Productive, Do Firms?: Investigating Productivity Dynamics" (NBER Working Paper No. 6899).
- MAIRESSE, J. and GRILICHES, Z. (1990), "Heterogeneity in Panel Data: Are there Stable Production Functions", in P. Champsaur *et al.*, (eds.) *Essays in Honor of Edmond Malinvaud*, Vol. 3 (Cambridge: MIT Press) 125–147.
- MILGROM, P. and ROBERTS, J. (1990), "The Economics of Modern Manufacturing: Technology, Strategy, and Organization", *American Economic Review*, **80** (3), 511–528.
- OLLEY, G. S. and PAKES, A. (1996), "The Dynamics of Productivity in the Telecommunications Equipment Industry", *Econometrica*, **64** (6), 1263–1297.
- PORTER, R. H. (1983), "A Study of Cartel Stability: The Joint Executive Committee, 1880–1886", *Bell Journal of Economics*, **14** (2), Autumn, 301–314.
- REISS, P. C. and WOLAK, F. A. (2002), "Structural Econometric Modelling in Industrial Organization" (Stanford University).
- RUBENSTEIN, J. (1992) *The Changing U.S. Auto Industry: A Geographical Analysis* (New York: Routledge).
- WOMACK, J. P., JONES, D. T. and ROOS, D. (1990) *The Machine that Changed the World* (New York: Rawson Associates).