Returns to scale: concept, estimation and analysis of Japan's turbulent 1964–88 economy

W. Erwin Diewert *University of British Columbia*
Takanobu Nakajima *Keio University*
Alice Nakamura *University of Alberta*
Emi Nakamura *Columbia University*
Masao Nakamura *University of British Columbia*

Abstract. There is policy interest in factoring productivity growth into technical progress and returns to scale components. Our approach uses exact index number methods to reduce the parameters that must be estimated, and allows us to exploit the cross-sectional dimension of plant-level panel data. We show that the same equation can also be used to estimate 'Harberger' scale economies and technical progress indicators that require fewer assumptions. Estimates of the elasticity of scale for Japanese establishments in three major industries over 1964–88 are presented. Our study spans the high growth era of the 1960s, two oil shocks, and other exogenous shocks. JEL classification: C43, D24

Rendements à l'échelle: concept, estimés, et analyse de l'économie turbulente (1964–1988) du Japon. Il y a intérêt en politique publique à identifier les composantes de la croissance de la productivité attribuables au progrès technique et aux rendements à l'échelle. L'approche utilise les méthodes des nombres indices exacts pour réduire les paramètres qui doivent êtres estimés, et pouvoir exploiter la dimension transversale des données de panel au niveau de l'établissement. On montre que la même équation peut être utilisée pour estimer les indicateurs d'économies d'échelle et de progrès technique à la Harberger (lesquels nécessitent un plus petit nombre de postulats). On présente des évaluations de l'élasticité d'échelle pour des établissements japonais dans trois industries importantes entre 1964 et 1988.L'étude couvre la période de forte croissance des années 1960, celle des deux chocs pétroliers et d'autres chocs exogènes.

The authors thank Yusuke Yamamoto for outstanding research assistance, the participants in an Industry Canada workshop, a Canadian Economics Association session, and participants in a session at the *World Congress on National Accounts and Economic Performance Measures for Nations*, including especially helpful verbal and written comments from Michelle Alexopoulos, Susanto Basu, Kevin Fox, Jon Steinsson, and Hiroki Tsurumi. All errors and interpretations are our sole responsibility. This research was supported by research grants from the Social Science and Humanities Research Council of Canada (SSHRC). Email: alice.nakamura@ualberta.ca; diewert@econ.ubc.ca; nakajima@fbc.keio.ac.jp; enakamura@columbia.edu; masao.nakamura@sauder.ubc.ca

Canadian Journal of Economics / Revue canadienne d'Economique, Vol. 44, No. 2
May / mai 2011. Printed in Canada / Imprimé au Canada
0008-4085 / 11 / 451–485 / © Canadian Economics Association

1. Introduction

Productivity growth is a widely used performance metric for nations, industries, firms and establishments (see Diewert and Nakamura 1993, 1999, 2003, 2007; Alexopoulos and Cohen 2011; Alexopoulos and Tombe 2009; Freeman et al. 2011; Federal Reserve Bank of Boston). In its simplest form, productivity growth is measured as the change in total output relative to the change in input usage. If outputs grow relatively more quickly, there is a kind of welfare improvement, at the aggregate level at least. However, observing productivity growth by itself provides no information about how this growth was achieved or can be improved upon.

From a policy perspective, there is interest in factoring productivity growth into technical progress and returns to scale components. These are components that many government and corporate decision makers view as subject to policy influence. For example, government tax concessions for firms that perform research and development (R&D) are often promoted as a way of boosting the technical progress component of productivity growth, and free trade agreements are viewed as a way of opening up larger potential markets and thus of enabling businesses to reap productivity gains via increasing returns to scale. However, the empirical support for these views is mixed.

Indeed, some economists argue that increasing returns to scale are not important as more than a theoretical possibility, for U.S. industries at least. Burnside, Eichenbaum, and Rebelo (1995) conclude for the United States that 'there is virtually no evidence to suggest that there are important deviations from constant returns to scale in the manufacturing industry.' Growth accounting, as advocated by Jorgenson and his collabourators and as carried out for EU-KLEMS, incorporates the explicit assumption of constant returns to scale (see Jorgenson, Ho, and Strioh 2004; Timmer, O'Mahony, and van Ark 2007). The division of views among economists regarding the significance of increasing returns to scale is mirrored in statistical agency practices. Some national statistics agencies carefully avoid making use of assumptions regarding scale economies in their data production procedures.[1] Others have built the assumption of constant returns to scale into their procedures for compiling the national accounts data and their national productivity statistics.[2]

1 See Balk (2010a, b) regarding the situation in the Netherlands.

2 For example, in explaining how the Canadian productivity accounts are compiled, Baldwin and Gu (2007) write: 'The advantage of using the method that employs endogenous rates is that it is provides a fully integrated set of accounts. The surplus is taken directly from the National Accounts that provides the underlying data for the productivity accounts. Capital is directly estimated from the investment flows that are also part of the System of National Accounts. In Canada, investment flows are integrated with the input-output tables and are thus consistent with output at the industry level. Equally important, the assumptions that are required to make use of the surplus in estimating capital services are fully compatible with the assumptions that underlie the nonparametric productivity estimates – that of a fully competitive economy with a production process subject to constant returns to scale.' They note that this is contrary to the

In this paper, we begin by reviewing alternative possible definitions for the elasticity of scale. We then derive an estimating equation with only two coefficients to be determined even though a flexible functional form is adopted for the producer behavioural relationship. The proposed estimation approach is of practical use particularly when there are large numbers of outputs and inputs, a situation that often results in problems of inadequate degrees of freedom using standard econometric methods and aggregate annual data and in multicollinearity problems even when firm or establishment data are used.

Estimates of the elasticity of scale for Japanese establishments in three major industries over the period 1964–88 are presented and discussed. The period spanned by our data includes the high-growth era of the 1960s, the two oil shocks, and the slow growth years of the 1980s. Finally, we propose alternative indicators of scale economies and technical progress referred to as 'Harberger indicators,' since they are inspired by observations of Harberger (1997, 1998).

2. Characterizing a production scenario

In this section, we introduce the main terminology and definitions used in this paper. In particular, we introduce the definitions of the quantity and price indexes, and show how these relate to revenue, cost and total factor productivity. Importantly, we wish to analyze a situation in which there are multiple cross-sectional units that we observe over time. We will therefore use indexes to compare production units both over time and in cross-sectional dimensions. We order the units according to size for the purpose of defining indexes in the cross-sectional dimension.

For a general N-input, M-output production process, the period t input and output price vectors are denoted by $w^t \equiv [w_1^t, \ldots, w_N^t]$ and $p^t \equiv [p_1^t, p_2^t, \ldots, p_M^t]$, while $x^t \equiv [x_1^t, \ldots, x_N^t]$ and $y^t \equiv [y_1^t, \ldots, y_M^t]$ denote the period t input and output quantity vectors. Nominal total cost C and revenue R are defined for production scenarios t and s as

$$C^t \equiv \sum_{n=1}^{N} w_n^t x_n^t, R^t \equiv \sum_{m=1}^{M} p_m^t y_m^t \tag{1}$$

$$C^s \equiv \sum_{n=1}^{N} w_n^s x_n^s \quad \text{and} \quad R^s \equiv \sum_{m=1}^{M} p_m^s y_m^s.$$

more recent recommendations of Schreyer (2004) at the OECD. Moreover other departments of the Government of Canada have been pursuing initiatives aimed at helping Canadian businesses reap the benefits of increasing returns to scale (see, e,g., Government of Canada 2007).

We also define four hypothetical quantity aggregates.[3] The first two are

$$\sum_{n=1}^{N} w_n^s x_n^t \quad \text{and} \quad \sum_{m=1}^{M} p_m^s y_m^t. \tag{2}$$

These aggregates are what the cost and revenue would have been if the period t inputs had been purchased and the period t outputs had been sold at period s prices. In contrast, the third and fourth hypothetical quantity aggregates are sums of period s quantities evaluated using period t prices:

$$\sum_{n=1}^{N} w_n^t x_n^s \quad \text{and} \quad \sum_{m=1}^{M} p_m^t y_m^s. \tag{3}$$

These are what the cost and revenue would have been if the period s inputs had been purchased and the period s outputs had been sold at period t prices.

The eight aggregates given in (1) through (3) are all that are needed to define the Paasche (P), Laspeyres (L), and Fisher (F) quantity, price, and TFPG indexes:

$$Q_P^{s,t} \equiv \sum_{i=1}^{M} p_i^t y_i^t \Big/ \sum_{j=1}^{M} p_j^t y_j^s, \quad Q_L^{s,t} \equiv \sum_{i=1}^{M} p_i^s y_i^t \Big/ \sum_{j=1}^{M} p_j^s y_j^s, \quad Q_F^{s,t} \equiv (Q_P Q_L)^{(1/2)}. \tag{4}$$

Similarly, the Paasche, Laspeyres, and Fisher input quantity indexes can be defined as

$$Q_P^{*s,t} \equiv \sum_{i=1}^{N} w_i^t x_i^t \Big/ \sum_{j=1}^{N} w_j^t x_j^s, \quad Q_L^{*s,t} \equiv \sum_{i=1}^{N} w_i^s x_i^t \Big/ \sum_{j=1}^{N} w_j^s x_j^s, \quad Q_F^{*s,t} \equiv (Q_P^* Q_L^*)^{(1/2)}. \tag{5}$$

The output and input price indexes are

$$P_P \equiv \sum_{i=1}^{M} p_i^t y_i^t \Big/ \sum_{j=1}^{M} p_j^s y_j^t \quad \text{and} \quad P_P^* \equiv \sum_{i=1}^{N} w_i^t x_i^t \Big/ \sum_{j=1}^{N} w_j^s x_j^t \tag{6}$$

$$P_L \equiv \sum_{i=1}^{M} p_i^t y_i^s \Big/ \sum_{j=1}^{M} p_j^s y_j^s \quad \text{and} \quad P_L^* \equiv \sum_{i=1}^{N} w_i^t x_i^s \Big/ \sum_{j=1}^{N} w_j^s x_j^s \tag{7}$$

3 Formally, the first two of these can be shown to result from deflating the period t nominal cost and revenue by a Paasche price index. The second two result from deflating the period t nominal cost and revenue by a Laspeyres price index. See Horngren and Foster (1987, chap. 24, pt 1) or Kaplan and Atkinson (1989, chap. 9) for examples of this accounting practice of controlling for price level change without explicit use of price indexes.

$$P_F \equiv (P_P P_L)^{(1/2)} \quad \text{and} \quad P_F^* \equiv (P_P^* P_L^*)^{(1/2)}. \tag{8}$$

A price index is the implicit counterpart of a quantity index (and vice versa) if the product of the quantity and price indexes equals the total cost ratio for input side indexes or the total revenue ratio for output side indexes.[4]

A TFPG index can be defined as a ratio of output and input quantity indexes:

$$TFPG^{s,t} \equiv Q^{s,t}/Q^{*s,t}. \tag{9}$$

All real production processes make use of multiple inputs and most yield multiple outputs. Nevertheless, it is easier to understand the various indexes used in this paper and how they relate to each other and to revenues and costs in the simplified context of a production process with one input and one output (a 1–1 process).

For each time period the quantity of the one input used is given by x_1^t, its unit price is w_1^t, the quantity of the one output produced in period t is y_1^t, and its unit price is p_1^t. TFP can be defined conceptually as the rate of transformation of total input into total output:[5]

$$TFP^t \equiv \left(y_1^t/x_1^t\right). \tag{10}$$

TFPG usually stands for total factor productivity growth over time. However, we are also interested in relative productivity – that is, in the total factor productivity gradient from one productive unit to the next for the sequential pairs in an ordered cross-section. TFPG, as used in either a time series or cross-sectional context, can be defined in several ways, three of which are considered here. The first is the rate of growth over time for *TFP*. This concept of TFPG, denoted here by *TFPG*(1), can be measured as [6]

$$TFPG(1)^{s,t} \equiv \left(\frac{y_1^t}{x_1^t}\right) \Big/ \left(\frac{y_1^s}{x_1^s}\right) = TFP^t/TFP^s. \tag{11}$$

Secondly, TFPG can be defined as the ratio of output and input quantity indexes. For a 1–1 case we have

$$TFPG(2)^{s,t} \equiv \left(\frac{y_1^t}{y_1^s}\right) \Big/ \left(\frac{x_1^t}{x_1^s}\right) = Q^{s,t}/Q^{*s,t}. \tag{12}$$

4 The implicit price (quantity) index corresponding to a given quantity (price) index can always be derived by imposing the product test and solving for the price (quantity) index that satisfies this rule.

5 Some authors also use TFP to refer to total factor productivity growth. In line with Bernstein (1999), we use TFPG rather than TFP for total factor productivity growth so as to avoid the inevitable confusion that otherwise results.

6 Here we refer to t and s as time periods. However, the 'period s' comparison situation could be for some other unit of production in the same time period.

Thirdly, TFPG can be defined as the real revenue to cost ratio for t versus s:

$$TFPG(3)^{s,t} \equiv \left[\frac{R^t/R^s}{p_1^t/p_1^s}\right] \Big/ \left[\frac{C^t/C^s}{w_1^t/w_1^s}\right] = \left[\frac{R^t/C^t}{R^s/C^s}\right]\left[\frac{p_1^s/w_1^s}{p_1^t/w_1^t}\right], \tag{13}$$

where it can easily be seen that $(R^t/R^s)/(p^t/p^s) = y_1^t/y_1^s$ and $(C^t/C^s)/(w^t/w^s) = x_1^t/x_1^s$.

Comparing the expressions in (11), (12), and (13), we can readily see that for the 1–1 case all three concepts of TFPG lead to the same measure. This result carries over exactly to the general multiple input, multiple output case for the second and third concepts and carries over approximately for the first concept (see Diewert and Nakamura 2007).

In general also, an output quantity index can always be specified as the revenue growth rate corrected for output side price change using an appropriate output price index:

$$(R^t/R^s)/P^{s,t} = Q^{s,t}. \tag{14}$$

Similarly, an input quantity index equals cost growth corrected for input side price change:

$$(C^t/C^s)/P^{*s,t} = Q^{*s,t}. \tag{15}$$

3. Defining the elasticity of scale

In the *New Palgrave: A Dictionary of Economics*, J. Eatwell (1987) gives the conventional definition of the elasticity of scale: 'If all inputs are multiplied by a positive scalar, t, and the consequent output represents as $t^\gamma y$, then the value of γ may be said to indicate the magnitude of returns to scale. If $\gamma = 1$, then there are constant returns to scale: any proportional change in all inputs results in an equiproportionate change in output. If $\gamma > 1$, there are increasing returns to scale. If $\gamma < 1$ (though not less than 0, given the possibility of free disposal), then there are decreasing returns.' This definition is widely accepted. For example, Quah (2003) repeats it in verbal form: 'A production technology shows increasing returns to scale or simply increasing returns when an equiproportional increase in factor inputs results in a greater than proportional increase in output . . . Under constant returns to scale, an equiproportional increase in factor inputs results in an exactly proportional increase in output; under decreasing returns to scale, an equiproportional increase in factor inputs results in less than a proportional increase in output.' This definition is also built into important bodies of economic theory (see Diewert and Fox 2010).

The conventional approach to estimating the elasticity of scale is to derive an expression for this parameter in the context of a production, revenue, or cost function representation of the technology menu.

Suppose that $Y^t = f^t(x^t)$, where f^t is a production function, Y^t is the quantity for a single output good, and $x^t \equiv [x_1^t, ..., x_N^t]$ is a period t vector of the quantities for the N input goods. Suppose, moreover, that the selected production function can be factored into a pure output augmenting technical progress component – specified here, as in many studies, as a simple multiplicative term, $e^{\theta t}$ – and an atemporal production function, f.[7] That is, suppose the period t production function can be written as[8]

$$f^t(x^t) \equiv e^{\theta(t-s)} f(x^t) \quad \text{for all } t. \tag{16}$$

Within this framework, the elasticity of scale can be defined locally as the percentage change in total output due to a 1% increase in quantity for each of the N inputs:

$$\begin{aligned} \gamma_P(x^t) &\equiv [f^t(x^t)]^{-1} df^t(\lambda x^t)/d\lambda \,|_{\lambda=1} = \sum_{n=1}^{N} f_n(x^t) x_n^t / f(x^t) \\ &= \sum_{n=1}^{N} \partial \ln f_n(x_n^t) / \partial \ln x_n^t. \end{aligned} \tag{17}$$

A production function framework for defining the elasticity of returns to scale is appropriate when there is only one output (i.e., when $M = 1$). It is also appropriate with multiple outputs provided the output mix is approximately fixed.[9] In that case, some fixed weighted aggregate of the output quantities, such as total sales evaluated at constant prices, can be used to represent the total output. Hall (1988, 1990), Yoshioka, Nakajima, and M. Nakamura (1994), Nakajima, M. Nakamura, and Yoshioka (1998, 2001), Basu and Fernald (1997), and Diewert and Lawrence (2005) all implicitly make the latter assumption by the act of treating sales as the sole output. However, this approach is unsuitable in cases where there are large changes in the output mix.

When there are multiple outputs as well as multiple inputs, a revenue function can be used to characterize a production unit's technology. The revenue function,

7 A variety of more general treatments of disembodied, Hicks-neutral technical progress are possible here, and in the following sections, with minimal complication to the derivations of the semi-exact estimators we present for the returns to scale. Substantial adjustments are required, however, to incorporate more relaxed formulations of time related technical progress that do not presume Hicks neutrality such as are found in Diewert (1980).

8 Solow in his classic 1957 paper states that the variable t 'for time' appears in the production function F 'to allow for technical change.' Having introduced t in this way, he goes on to observe that 'slowdowns, speed-ups, improvements in the education of the labour force, and all sorts of things will appear as 'technical change.'

9 This is the condition needed to justify using Leontief's (1936, 54–7) Aggregation Theorem.

$R^t(p^t, x^t)$, conditional on the availability of the vector of inputs and evaluated at the period t output price vector $p^t = [p_1^t, ..., p_M^t]$, is

$$R^t(p^t, x^t) \equiv \max_y \{p^t y : (y, x^t) \text{ belongs to } S^t\}, \tag{18}$$

where $p^t y^t \equiv \sum_{m=1}^{M} p_m^t y_m^t$ denotes the inner product between the vectors p^t and y^t, and S^t denotes the feasible set of inputs and outputs. The period t revenue function, $R^t(p^t, x^t)$, is re-specified using a temporally invariant aggregator, R, for the term, $R(p^t, x^t)$, and a time dependent term, $e^{\theta t}$; that is,

$$R^t(p^t, x^t) \equiv e^{\theta(t-s)} R(p^t, x^t). \tag{19}$$

With this setup, a measure of the elasticity of returns to scale can be defined as the percentage change in revenue due to a 1% increase in quantity for each of the N inputs, controlling for output side changes in the price level from s to t:

$$\gamma_R(p^t, x^t) \equiv [R^t(p^t, x^t)]^{-1} dR^t(p^t, \lambda x^t)/d\lambda\,|_{\lambda=1} = \sum_{n=1}^{N} R_n(p^t, x^t) x_n^t / R(p^t, x^t)$$
$$= \sum_{n=1}^{N} \partial \ln R(p^t, x^t)/\partial \ln x_n^t. \tag{20}$$

The cost function provides a dual alternative to the primal production and revenue function characterizations of the technology of a production unit that have been introduced above. Like the revenue function case, the cost function framework is suitable when there are multiple outputs as well as inputs. The firm's period t cost function, $C^t(w, y)$, conditional on target levels for a set of outputs and given a vector of input prices $w \equiv [w_1^t, \ldots, w_N^t]$ is

$$C^t(w, y) \equiv \min_x \{w^t \cdot x^t : (y^t, x) \text{ belongs to } S^t\}, \tag{21}$$

where $w^t \cdot x^t \equiv \sum_{n=1}^{N} w_n^t x_n^t$ and where S^t is a feasible set of inputs and outputs, as above. It is assumed that the period t cost function, C^t, can be related to an atemporal function, C, as follows:[10]

$$C^t(w^t, y^t) = e^{\theta(t-s)} C(w^t, y^t). \tag{22}$$

With this setup, a reciprocal cost function based measure of the elasticity of scale is defined as the percentage change in cost due to a 1% increase in each of the

10 See Diewert and Fox (2008) for more on this approach. They also deal explicitly with market structure issues that we do not go into here.

output quantities (see Panzar 1989, 8, def. 3):

$$[1/\gamma_C(w^t, y^t)] \equiv [C^t(w^t, y^t)]^{-1} dC^t(w^t, \lambda y^t)/d\lambda \,|_{\lambda=1} = \sum_{m=1}^{M} C_m(w^t, y^t) y_m^t / C(w^t, y^t)$$
$$= \sum_{m=1}^{M} \partial \ln C(w^t, y^t)/\partial \ln y_m^t. \qquad (23)$$

Intuitively, we might expect primal and dual definitions of the elasticity of scale to differ. However, for a given production situation that satisfies the usual conditions, the primal and dual definitions are equal, so it does not matter which is used. Given the importance of this result, in appendix B we provide proofs for the 1-N and M-N cases.

The component that is a function of time has been specified in the simplest possible way in (16), (19), and (22). Little would change if a more complex time specification were utilized, provided the separability of the time component is maintained. One way of allowing for more complex time-related effects is to use rolling two-period panels over the interval of time spanned by the available data, as is done in the empirical part of this study. This approach is possible, of course, only if establishment or firm panel data are available. The time component will pick up factors that affect production units of all sizes in a given time period, including business cycle changes that have capacity utilization effects.[11]

TFPG can then be decomposed into technical progress (TP) and returns to scale components.

4. Semi-exact estimators for the elasticity of scale and technical progress

In this section we describe the estimation approach employed in the empirical application in this paper. A key feature of our empirical approach is that we are able to take advantage of the panel structure of our establishment-level data to identify the returns to scale parameter. This panel data approach was developed by Yoshioka, Nakajima, and M. Nakamura (1994) and Nakajima, M. Nakamura, and Yoshioka (1998, 2001) and is also used in Nakajima et al. (2007).[12]

11 Some analysts also distinguish capacity utilization as a separate component and others, explicitly or implicitly, treat capacity utilization as a component of returns to scale. The latter accords with business world commentary. Managers make multitudes of choices that combine scale and capacity utilization issues as joint choices.

12 This stream of work can be viewed as a generalization and extension of the basic theoretical results of Diewert (1976, lemma 2.2, eq. (2.11) and theorem 2.16), the material on non-competitive approaches in Diewert (1978), and additional results in Diewert (1981, including the s. 7 results on the markup change). Diewert and Fox (2008) extended this approach to allow for multiple outputs, but with the assumption of competitive output markets

The semi-exact estimation approach employs the following five steps: (1) Choose a producer behavioural equation to use as the analysis framework. (2) Write the specified equation in terms of a separable term for time varying phenomena and an atemporal producer behavioural relationship. (3) Define the elasticity of scale in the context of the choices made for the first two steps. (4) Choose a flexible functional form for the atemporal producer behavioural equation. (5) Use exact index number methods[13] to derive a parsimonious estimating equation that isolates the elasticity of scale as an unknown coefficient that can be estimated from available data while also controlling for the separable time varying factors. We now apply these steps to derive a semi-exact estimating equation that is used in our empirical application.

With pure output augmenting technical progress, a period t production function can be written as

$$f^t(x^t) \equiv e^{\theta(t-s)} f(x^t), \tag{24}$$

where θ is a conventional technical progress parameter. A homogeneous translog function with no constraint on the degree of homogeneity, k, is used for the time invariant f, since this allows for possible increasing or decreasing returns to scale and places no restrictions on the input substitution elasticities.[14]

In appendix C, we show the derivation for the following semi-exact equation of the form derived by Yoshioka, Nakajima, and Nakamura (1994) and Nakajima, Nakamura, and Yoshioka (1998, 2001):

$$\ln Y = \ln f(x) \equiv \beta_0 + \sum_{n=1}^{N} \beta_n \ln x_n + (1/2) \sum_{n=1}^{N} \sum_{j=1}^{N} \zeta_{nj} \ln x_n \ln x_j. \tag{25}$$

The parameters on the right-hand side of (25) satisfy the following restrictions:

$$\sum_{n=1}^{N} \beta_n = k > 0 \tag{26}$$

$$\sum_{j=1}^{N} \zeta_{nj} = 0 \quad \text{for} \quad n = 1, \ldots, N \tag{27}$$

and price taking behaviour in these markets. Diewert and Fox (2008) further generalized the semi-exact approach to a cost function framework that allows for limited types of imperfect competition and markups in output markets, but they implement the approach with aggregate time series data for industries.

13 Diewert (1976, 1981, 2002) pioneered the exact index number approach and also the related concept of superlative index numbers. Both the Fisher and the Törnqvist indexes are superlative.

14 The basic translog functional form was introduced by Christensen, Jorgenson and Lau (1971, 1973).

$$\zeta_{nj} = \zeta_{jn} \quad \text{for all } 1 \leq n < j \leq N. \tag{28}$$

The parameter k in (13) is the degree of homogeneity. For a homogeneous production function, the degree of homogeneity equals the local elasticity of scale everywhere, so $\gamma(x) = \gamma = k$. The elasticity of scale, γ as defined in (17), can potentially be estimated by assuming a stochastic specification, estimating all the unknown parameters of (24) and (25), and then summing the estimates of $\beta_1, \ldots, \beta_N$ as in (26). This is the conventional approach. However, the number of parameters can be overly large.[15]

Adopting instead the semi-exact approach, where $Q_T^{s,t}$ and $Q_T^{*s,t}$ are Törnqvist output and input quantity indexes[16] and u^t is an error term, the estimating equation is

$$ln\left[Q_T^{s,t}\right] = \theta(t-s) + \gamma \ln Q_T^{*s,t} + u^t. \tag{29}$$

In estimating equation (29), we use grouped panel data on establishments.[17] In each time period, we ordered the establishments by size, from smallest to largest. We then constructed chained quantity indexes, with the links of the chain being bilateral comparisons for the successive establishment group pairs, moving from smaller to larger across the first year in each two-year panel, and then from the first to the second year via a chain link comparison over time for the smallest establishment group, and then across the successive pairs for the second year of the two-year panel.

If $\gamma > 1$, then we say there is evidence of increasing returns to scale. The fact that our estimating approach allows us to use variation in both the cross-sectional and the time dimensions generates a great deal of variation in both the output and input indexes (the left- and right-hand-side variables of equation (29)), allowing for improved identification of the returns to scale parameter.

5. Harberger indicators of scale economics and technical progress

In the previous sections, we developed the semi-exact estimation approach for which results are shown in section 7. This approach was derived under particular assumptions about producer behaviour. In this section we present an alternative interpretation of the same estimating equation. This interpretation draws on the

15 For example, in estimating scale economies and technical change using aggregate time series data, Berndt and Khaled (1979) and also Chan and Mountain (1983) had to estimate 22 unknown parameters using 25 observations.

16 The definition for a Törnqvist quantity index is given in appendix B. Our theoretical analysis in section 2 focused on the Laspeyres and Paasche indexes instead of the Törnqvist index. However, this difference is probably not important, since the Törnqvist index is closely approximated by the Fisher index, the geometric average of the Laspeyres and Paasche indexes. Further details of the derivation are given in appendix C.

17 This is the form in which the data were made available.

analysis of productivity of Harberger (1998). Harberger writes: 'Economies of scale may take on a new aspect when one considers that negative TFP experiences can arise from firms being driven back up their short run average cost curves as competitive forces cause output to fall well below designed capacity.' He draws our attention to factors affecting production costs, and TFPG, via ongoing interactions between size choices and economic circumstances.

We define the presence of 'Harberger' returns to scale as a situation where, at a given point in time, larger firms have higher productivity than smaller firms. Notice that it is exactly in the situation in which an equation of form (29) will yield a coefficient on the input index greater than one – that is, evidence of returns to scale according to the standard definition analyzed in the previous sections of this paper. Given the interpretation of the Harberger scale economies indicator, then the residual differences in productivity over time that will be captured by the constant term in the estimating equation are naturally classified as technical progress: what we term the Harberger technical progress indicator.

6. Returns to scale results for three Japanese industries

Before moving on to our own empirical work, we first note some problems that have been flagged as possible sources of erroneous estimates of returns to scale in other studies. For one, Burnside (1996) argues that the imposition of cross-industry restrictions (e.g., Hall 1990) may lead to upwardly biased estimated of returns to scale. A remedy is to not pool data over different industries. Secondly, Basu and Fernald (e.g., 1995) argue that, when value added output data are utilized in the presence of markups (e.g., Hall 1990), some of the contribution of intermediate products are likely to be incorrectly attributed, leading also to upwardly biased estimates of returns to scale. A remedy is to not use value added output data in studies of this sort. Thirdly, Basu and Fernald (1997) call attention to additional aggregation issues. They argue for the use of establishment level data as the remedy, but use aggregate data in their own study because of not having access to appropriate establishment data.

6.1. Our data and estimating equation

The establishment-level data we use in our empirical application spans the turbulent years of 1964–88 for the Japanese economy. Our data were compiled by what was formerly the Japanese Ministry of International Trade and Industry (MITI) and is now the Ministry of Economy, Trade and Industry (METI). Each year, the Census of Manufacturing by Industry is carried out for establishments (plants and other places of business). The establishments are classified by size measured by the number of employees: (1) 30–49, (2) 50–99, (3) 100–99, (4) 200–99, (5) 300–499, (6) 500–999, and (7) 1,000 and more.[18] Our data consist of average figures, by industry, for the establishments in each of the designated size groups. We use these data for 1964–88 in the form of 24 industry-specific rolling two-year panels. We do not pool over industries. Moreover, the panels we create also allow us to examine how the estimated scale economies change over time. Thus, we can check how the estimates of the scale economies accord with information in MITI documents and from former MITI employees. The first and second years for each panel are denoted as t and $t + 1$.

Output is measured as establishment gross sales (not value added) plus the net increases in final product inventories evaluated at current period prices. The production input attributes included in our study are the number of workers, the fixed assets at the beginning of each year (as a proxy for the available services of these assets for that year), and intermediate and raw materials, all measured per establishment in the original survey and available by establishment size group.[19]

For each year of each two-year industry panel, the establishment data are ordered as explained above.[20] Our estimating equation is[21]

$$\ln \tilde{Q}_{Ti} = \theta D_i + \gamma \ln Q^*_{Ti} + e_i \tag{32}$$

for the pooled observations consisting of the information for each year of the two years for each industry panel, where the different size groups are ordered from smallest to largest and where

$$D_i = 0 \quad \text{for observations in the first year}\,(t-1)\,\text{of each two - year panel}$$
$$= 1 \quad \text{for observations in the second year}\,(t)\,\text{of each two - year panel.}$$

We treat the error term, e, as randomly distributed in the cross sectional dimension for each two-year panel with zero mean and constant variance and as autocorrelated over the two years. We estimated (19) using generalized least

18 The number of these groups and hence the definitions of size groups have varied somewhat over time.

19 The input cost price deflators for labour and capital are based on the average annual cash earnings per worker and the depreciation rate for fixed assets plus the average interest rate for a one-year term-deposit for capital. In computing the capital stock, new investment in fixed assets is deflated using the industry-specific investment goods deflators published by the Economic Planning Agency. The investment goods deflator is also used to adjust the input price of capital. The Bank of Japan input price deflator is used to deflate the materials input.

20 The rationale for ordering the establishment observations in this way is the same as for ordinary chaining.

21 For the left-hand variable to be the log of a true implicit Törnqvist output quantity index, a Törnqvist output price index must be used to deflate the total average sales figures. Similarly, for the quantity variable on the right-hand side to be a true Törnqvist input quantity index, we would require implicit Törnqvist input price indexes for the inputs. In fact, we had no choice but to use the price information and deflators available to use from the Government of Japan. The resulting errors of approximation should be minimized by our use of two-year, industry-specific panels.

squares. The error variance and autoregressive parameter can vary freely over panels.[22]

We present our scale indicator estimation results in table 1. Values statistically different from 1 using a critical region of 0.05 are starred. Coefficient values that are also significantly greater than 1 and hence that indicate significantly increasing returns to scale are in **boldface type**. Coefficient values that are significantly less than 1 and hence indicate significantly decreasing returns to scale are in *italics*. In columns 2, 4, and 6 of table 2, we show for each industry and two-year panel whether the estimated progress indicator is significantly positive, significantly negative, or insignificantly different from zero. These results can be compared with the corresponding summary results for the scale indicator in columns 1, 3, and 5, where a > sign denotes significant increasing scale economies, a < sign denotes significant decreasing scale economies, and an = sign indicates that the estimated coefficient (shown in table 1) is insignificantly different from 1, indicating constant returns to scale.

It has been argued (e.g., Burnside 1996; Burnside et al. 1995) that studies such as Hall (1990) find evidence of increasing returns to scale in part at least because of a failure to allow for cyclical variation in capital utilization. Our measures of returns to scale incorporate the effects of varying capacity utilization. As a consequence, one possible source of decreasing returns to scale in some time periods is that larger firms were unable to make use of their installed capacity in periods of low demand.

6.2. *Discussion of findings*

Leading into the period for which we have data, the years of 1945 to 1960 are often referred to as the Reconstruction phase for Japan. This phase was followed by the Rapid Growth years through some time in the early to mid-1970s, also called the Golden Years (Komiya, Okuna, and Suzumura 1988; Johnson 1982). Many believe that the Government of Japan, acting primarily through the Japanese Ministry of International Trade and Industry (MITI), established in 1949, played a role in bringing about the Golden Years (Patrick 1986; Johnson 1982). According to Tsuruta (1988), MITI sought to develop industries that could survive international competition by raising the productivity of Japanese industries. The Government of Japan finalized a vision in 1963 with two criteria for an optimum industrial structure, one being the 'Productivity Increase Rate Criteria.' Increasing the scale of production facilities was the main

22 Correlation of the error term in (32) with the input index on the right-hand side is less likely than correlation of the error term for a standard producer behavioural equation such as a production function with right-hand side input quantities. Nevertheless, we experimented with using as instruments the average interest rate on a one-year term deposit that varies over time but not establishments and also the average annual cash earnings per worker and the depreciation rate for fixed assets both of which vary over establishments. The null hypothesis for the Hausman test is rejected for only 3 cases out of our 72 regressions. Thus, we show generalized least squares results in table 1.

TABLE 1
Estimates of the returns to scale parameter[a]

	Years	Textiles	Pulp/paper	Electrical machinery
1	64–65	*0.989** (34.6, 18)[a]	**1.018*** (13.7, 16)	**1.021*** (8.8, 18)
2	65–66	*0.990** (33.2, 18)	**1.016*** (10.7, 18)	**1.017*** (7.7, 18)
3	66–67	*0.991** (19.3, 18)	**1.013*** (11.3, 18)	**1.025*** (10.8, 18)
4	67–68	*0.992** (4.9, 16)	**1.027*** (85.7, 16)	**1.034*** (22.9, 16)
5	68–69	*0.974** (110.8, 16)	**1.022*** (53.1, 16)	**1.044*** (35.5, 16)
6	69–70	*0.982** (88.3, 16)	**1.017*** (31.0, 16)	**1.039*** (24.5, 16)
7	70–71	*0.989** (5.7, 16)	**1.020*** (33.4, 16)	**1.035*** (25.3, 16)
8	71–72	0.991 (4.1, 16)	**1.023*** (83.0, 16)	**1.029*** (20.3, 16)
9	72–73	1.006 (3.5, 16)	**1.022*** (32.6, 16)	**1.038*** (26.2, 16)
10	73–74	**1.008*** (11.0, 16)	**1.013*** (5.8, 16)	1.043 (39.5, 16)
11	74–75	1.004 (1.3, 16)	*0.987** (16.4, 16)	**1.039*** (31.6, 16)
12	75–76	0.985 (3.2, 14)	0.993 (3.2, 14)	**1.060*** (110.6, 14)
13	76–77	*0.978** (11.5, 14)	1.007 (0.9, 14)	**1.057*** (183.8, 14)
14	77–78	*0.986** (8.3, 14)	0.998 (0.2, 14)	1.051 (174.6, 14)
15	78–79	**1.018*** (8.5, 8)	0.994 (0.6, 14)	**1.064*** (146.6, 14)
16	79–80	**1.031*** (189.0, 8)	0.993 (0.9, 14)	**1.066*** (101.7, 14)
17	80–81	**1.023*** (10.9, 10)	0.995 (0.5, 14)	**1.051*** (112.4, 14)
18	81–82	1.003 (0.1, 10)	1.004 (0.6, 14)	**1.054*** (117.5, 14)
19	82–83	1.003 (0.1, 10)	0.996 (0.4, 14)	**1.057*** (110.5, 14)
20	83–84	**1.022*** (12.6, 10)	1.009 (2.7, 14)	**1.055*** (151.5, 14)
21	84–85	**1.018*** (20.0, 8)	0.994 (2.1, 14)	**1.061*** (149.6, 14)
22	85–86	**1.013*** (7.8, 8)	1.002 (0.5, 14)	1.047 (53.4, 12)
23	86–87	**1.040*** (11.0, 10)	**1.019*** (11.3, 14)	**1.036*** (22.4, 12)
24	87–88	**1.063*** (178.2, 8)	**1.021*** (6.2, 14)	**1.042*** (29.8, 12)

a The null hypothesis is H_0: RS $= 1$, where RS is our measure of returns to scale, γ. The first number in parentheses is the F statistic with degrees of freedom of 1 in the numerator and $n - 3$ in the denominator where n is the number of size groups of establishments. The second number is n. Values of the elasticity of scale that are significantly different from one using this F test with a critical region of 0.05 are starred.

TABLE 2
Returns to scale and progress indicator results

		Textiles		Pulp/paper		Electrical machinery	
	Years[c]	H-scale effect[a]	H-progress effect[b]	H-scale effect[a]	H-progress effect[b]	H-scale effect[a]	H-progress effect[b]
1	64–65	<	0	>	0	>	+
2	65–66	<	+	>	+	>	+
3	66–67	<	+	>	0	>	+
4	67–68	<	+	>	+	>	+
5	68–69	<	+	>	0	>	+
6	69–70	<	+	>	+	>	+
7	70–71	<	+	>	0	>	0
8	71–72	=	+	>	+	>	+
9	72–73	=	+	>	0	>	+
10	73–74	>	+	>	0	=	+
11	**74–75**	=	0	<	0	>	−
12	75–76	=	+	=	+	>	+
13	**76–77**	<	−	=	−	>	+
14	77–78	<	0	=	0	=	+
15	78–79	>	+	=	0	>	+
16	79–80	>	+	=	+	>	+
17	**80–81**	>	−	=	−	>	−
18	81–82	<	0	=	+	>	0
19	82–83	<	+	=	0	>	+
20	83–84	>	+	>	+	>	0
21	**84–85**	>	−	=	0	>	0
22	**85–86**	>	−	=	0	=	0
23	86–87	>	+	>	0	>	0
24	87–88	>	−	>	+	>	+

a Statistically significant increasing (decreasing) returns to scale are indicated by > (<), and = indicates that the H-scale indicator is not statistically different from 1 and hence that we accept the null hypothesis of constant H-returns to scale.
b Statistically significant positive (negative) technical progress is indicated by *a* + (−). A zero indicates that the H-progress indicator was not significantly different from zero, and hence that we accept the null hypothesis of no H-progress over that 2 year period.
c The years are in bold for any panel where technical progress was significantly negative for any one of the three industries.

strategy for raising the productivity of industries. The historical record reveals that government encouraged many large-scale mergers and that, where there were many suppliers, tried to foster systems of specialized producers. These rationalization plans were aimed at bringing about decreasing costs in industries by increasing the scale of production. Other government assistance was intended to minimize the financial, market and technological risks of investment by industry to modernize and expand capacity.

Roy (2005) argues that one reason these policies were successful, through the mid-1970s at least, is because infant industry type development measures were combined with export promotion and also the domestic market was sufficiently large that multiple strong domestic competitors could co-exist in most industries.

Moreover, the protection afforded to industries was reduced, in fact, as the industries became competitive. Industries that continued to have poor productivity performance were ultimately pushed by MITI to contract, MITI's objective being that the industry would end up leaner but better able to compete without assistance. Even competitive market advocate Michael Porter (1998) lauds this policy, noting explicitly that efforts were made to increase the scale of production facilities: 'This sort of government role was constructive. Competitive advantage depended on having modern, large scale facilities. Government's levers at this stage were powerful ones.'

In addition to managing industries that were believed to be promising for the future growth of the nation because of relatively high productivity and what MITI interpreted to be returns to scale (what we have now labelled 'Harberger scale economies'), MITI also managed declining industries. Harberger (1997, 1998) recommends that attention be paid to businesses doing poorly as part of efforts to raise national productivity. For example, the Industry Stabilization Law of 1978 aimed at suspension or scrapping of capacity in depressed industries. Laid-off workers were covered by insurance and firms were encouraged to submit re-employment assistance plans. In addition, distress loans were given to smaller businesses trying to adjust. Adjustment assistance policies facilitated employment switchovers by workers, resource transfers of many sorts, mergers to cut capacity, and modernization of equipment. By these measures, MITI may have helped make it possible for Japan to proceed, with relatively little social disruption, with liberalizations that were necessary to open the doors of international markets for Japan's successful industries.

Before turning to our scale economies estimation results for specific industries, we note also that external economic shocks seem to explain the significantly negative values for our Harberger progress indicator shown in table 2. The estimated values of this indicator (not shown) are small in all cases. The years for the panels with a significantly negative progress indicator for any one of the industries are in bold in table 2. The progress indicator controls for conditions affecting establishments of all sizes, including general technical progress and also shocks to the economy of the nation. Burbidge and Harrison (1984) examine aggregate developments in five major industrial countries, including Japan, over much of the period spanned by our study. They conclude that the oil shocks in the early 1970s had significant negative effects on the economies of all five countries. They find also that whereas the effects of the 1979–80 oil shocks were minimal for four of the countries examined, Japan's economy was hard hit.

In this study, we chose to focus on three industries that underwent major changes in size and structure over the economically eventful period of 1964–88. One was a leading industry heading into this period that then fell on hard times: the textile industry. A second – pulp and paper – was small as of 1964, but was viewed as promising over the next several years and then had to be scaled down beginning in the later 1970s, owing to a shortage of a crucial raw material input. The third was small in 1964, but was already seen as a potential leader for the

TABLE 3
Establishments and workers in textile mills, 1965–84

	Establishments		Workers	
	Number	Value for the given year as a percentage of 1965 value	Number	Value for the given year as a percentage of 1965 value
1960	38,773		1,163,253	
1965	100,157	100.0	1,326,872	100.0
1970	112,754	112.6	1,264,228	95.3
1975	114,111	113.9	995,669	75.0
1980	39,741	39.7	691,018	52.1
1984	36,269	36.2	626,567	47.2

SOURCE: Japan Statistical Association (1988), *Historical Statistics of Japan*, 2: 283

economy and this hope was born out over the coming decades: the electrical machinery industry. These are three industries where a great deal is known about what MITI officials and others believed, year to year, were the challenges and possible means of doing better.

Textiles was a leading industry for Japan prior to World War II. As rebuilding began during the Occupation (1945–52), the textile industry was designated as a key sector to lead the nation's economic recovery. One reason for this was that the U.S. government realized that Japanese textile companies could provide a market for U.S. surplus supplies of raw cotton. The Korean War (1950–3) gave a kick-start to the textile industry. Japan's textile industry supplied United Nations forces.

However, the end of the Korean War brought a sluggish domestic market and led to the first post-war MITI 'recommended curtailment' of operations in March 1952. Higher economic growth for Japan as a whole soon improved textile industry performance. The number of employees in the industry rose from 1.16 million in 1960 to 1.26 million in 1970. However, the 1970 figure represents a decrease from the 1965 employment figure (see table 3). The Continuous Automated Spinning (CAS) system was an important factor in the post-1965 employment decline. MITI was also worried about the increasing complaints of U.S. producers and potential emerging overcapacity. By the mid-1960s, MITI officials were rationing the facilities that could be built for production of synthetic fibres. This is despite the fact that Japanese manufacturers continued to be successful in selling to the United States. Also, a five-year Textile Industry Reorganisation Programme begun in 1967 called for modernizing equipment.

The Japanese Textile Federation and MITI tried to diffuse growing U.S. political opposition by adopting voluntary curbs on textile exports beginning in 1971. However, President Nixon and the U.S. textile industry denounced this initiative as inadequate. Moreover, in 1971, Nixon announced that U.S. dollars could no longer be converted to gold. A floating system for major currencies was adopted, and the yen appreciated versus the U.S. dollar. The new currency arrangements also pushed the major oil-exporting countries (OPEC) to raise oil prices. In 1974, the first of a series of OPEC price hikes increased Japanese synthetic textile production costs. Still, Japanese production of synthetics, which had been increasing since 1957, continued to rise into 1975, the output mostly being exported to the United States, where domestic producers were complaining. Also, producers in lower-wage Asian countries took advantage of the tariff reductions Japan had to make and began marketing their textile products in domestic Japanese markets. In October 1977, MITI intervened with production curtailments.

Already during the 1970s, large-scale textile companies such as Mitsubishi Rayon[23] were reinventing their businesses by actions such as shifting their portfolios to include more synthetic textile production and diversifying into non-textile lines of business as well as increasing their overseas investments (particularly Southeast Asian). Smaller firms found it harder to adjust to the changing circumstances. For the next five years, from 1978, the textile industry was designated a 'depressed industry' under the terms of the Temporary Measures Law for Stabilization of Specific Depressed Industries. In 1979, MITI moved even more firmly to discourage further expansion of the synthetic fibre industry and subsequently required manufacturers to dispose of a massive 18% of their facilities.

The Japanese textile industry was restructured in major ways over the 1964–88 years. Even as output grew under government encouragement, the number of regular workers was reduced and the number of establishments diminished, while retooling of the continuing establishments enabled higher volumes of output per establishment. The ongoing labour reallocation as a result of structural change in the textile industry was far greater than might appear from observation of only aggregate industry employment, because huge shifts were also taking place among the main sorts of textile production (see table 4).

In our view, the end of decreasing estimated scale economies in 1970–1 followed by increasing returns to scale in 1973–4 and then once more in 1978–9 and the 1980s probably reflect the measurable success of MITI efforts to restructure the Japanese textile industry.[24]

Pulp and paper was a relatively small industry for the first years spanned by our data. However, the high-growth years of the 1960s for Japan brought a rapid expansion in the demand for pulp and paper. Raw materials, initially in the form of logs and then wood chips and dry pulp, became an expanding part of world trade (Schreuder and Anderson 1988, 174). Japan became a dominant player first in the world wood chip market and subsequently in the dried pulp market (Pappens 1994, 24). In 1964, Japan constructed the first ocean-going vessel

23 http://www.fundinguniverse.com/company-histories/MITSUBISHI-RAYON-CO-LTD-Company-History.html

24 For more on these measures, see Nakamura and Vertinsky (1994). It should be noted that some government programs to help this industry were still in effect as of 1987–88. In particular, the Temporary Measures Law for the Structural Adjustment for Specific Industries lasted for five years from 1983.

TABLE 4
Value of textile exports, 1965–80 (¥m.)

	Cotton Yarn	Raw silk	Cotton textiles	Woolen textiles	Silk textiles	Synthetic yarn and fibres
1960	18,861	18,162	126,507	19,628	18,779	97,833
1965	7,353	4,867	108,944	31,266	13,049	328,179
	(1.5)	(1.0)	(22.0)	(6.3)	(2.6)	(66.5)
1970	5,314	464	67,541	27,190	5,279	510,729
	(0.9)	(0.0)	(11.0)	(4.4)	(0.8)	(82.8)
1975	6,676	1	77,107	13,060	4,432	654,681
	(0.9)	(0.0)	(10.2)	(1.7)	(0.6)	(86.6)
1980	11,730		110,037	11,252	10,360	625,530
	(1.5)		(14.3)	(1.5)	(1.3)	(81.3)

SOURCE: Japan Statistical Association (1988), *Historical Statistics*, 2: 39

designed specifically for the transportation of wood chips and began importing sawmill residues and wood chips from old-growth forests in Canada and the United States (Shimokawa 1977, 27; Schreuder and Anderson 1988, 169). Product demand conditions plus new technologies for paper production that the Japanese industry quickly took advantage of created conditions of establishment-level increasing returns to scale.[25]

However, by the mid-1970s, shortages of pulp wood supplies had become a serious problem for Japanese pulp and paper producers. Then, in 1979, U.S. interest rates shot up, causing a slump in the U.S. housing market. As house-building activity fell, this caused a reduction in production for both U.S. and Canadian lumber mills. The reduction in lumber mill residues led to a shortage of the wood chips and dry pulp needed by the Japanese pulp and paper industry. Demand for chips was still high in the United States, too, so the price of chips increased dramatically. Weyerhauser for example increased the price of Douglas fir wood chips by 138% in a six-month period. This hike in prices became known in the Japanese pulp and paper industry as 'chipshock' (Schreuder and Anderson 1988, 176–7).

New technologies were developed and instituted for utilizing pulp from hardwoods. These technologies made it possible to use chips and pulp from the tropics, and Japan had the ships needed to handle the transportation. Thus, new technologies that allowed for input substitution, new sources of chips and pulp that could be tapped because Japan also had the needed transportation capabilities, and rising demand for paper products for use with computers and photocopiers all are believed to have helped bring the Japanese pulp and paper industry back

25 Increasing returns to scale for the pulp and paper industry were reported for other nations too. See, for example Hailu and Veeman (2000) and Mohnen, Jacques, and Gallant (1996) for Canadian studies that find evidence of increasing returns to scale for Canada that persisted for more years than for Japan. This makes sense, since Canada continued to have access to sufficient supplies of chips to keep large pulp and paper mills operating closer to full capacity.

TABLE 5
Effective rate of protection (in percentage terms) in Japan by industry and year

Industry	1963	1968	1973	1975	1978
Textiles	54.3	28.2	18.6	38.6	38.3
Paper	9.7	18.0	11.0	17.3	9.4
Electrical machinery	30.9	16.5	5.4	10.2	7.4

SOURCE: Komiya, Okuno, and Suzumura (1988) quoting Shouda (1982)

into increasing returns to scale conditions by the late 1980s. In summary, the optimistic reports for this industry in the earlier part of the period spanned by our data are reflected in our empirical results by significantly increasing returns to scale for 1964–5 through the 1973–4 panels (table 1, col. 2). The severe chip shortage conditions and other problems this industry faced show up in our estimation results as decreasing or constant returns to scale from 1974–5 through 1985–6, after which the pulp and paper industry is again found to exhibit increasing returns to scale.

Electrical machinery was an expanding industry already by the decade prior to the period spanned by our data. Tariff barriers were one means that MITI used to try to enable domestic industries to grow and achieve the scale economies. In the years leading into the time period spanned by our data, the Japanese electrical machinery industry enjoyed relatively high effective tariff protection (see table 5). Beason and Weinstein (1996) report that the electrical machinery industry received less than the average industry in terms of government subsidies. However, MITI reportedly acted at many points to try to help raise the scale of production in this industry. As Okimoto (1989) explains, in Japan, the electrical machinery industry has had a high percentage of smaller firms. The Government of Japan also sponsored and helped to coordinate research of critical importance for the Electrical Machinery industry.

Over the years of 1955–97, electrical machinery has the highest growth rate of the industries considered by Porter and Sakakibara (2004, 41, table 3), based on empirical research of Beason and Weinstein (1996). Beason and Weinstein report that for the four industries where their estimate of the elasticity of scale was greater than 1 (fabricated metal, general machinery, transportation equipment, and electrical machinery) they were unable to reject the hypothesis of constant returns to scale. Porter and Sakakibara use that result in their analysis. However, Beason and Weinstein obtain their returns to scale estimates using a translog production function and following the setup of Chan and Mountain (1983). We suspect that the large number of parameters that this approach necessitates estimating and the collinear nature of many of the explanatory variables are the reasons that Beason and Weinstein did not find evidence of significant increasing returns to scale.

7. Concluding remarks

In this paper, we derive a very simple estimating equation for returns to scale and technical progress that avoids the problems of multicollinearity and inadequate degrees of freedom that typically arise in conventional approaches. We use what we term semi-exact estimators that surmount the conventional estimation problems. In this method, the exact index number approach is used to greatly reduce the number of auxilary parameters that must be estimated along with the elasticity of scale while not losing the flexibility of the chosen production framework. In addition, building on the insights of Yoshioka, Nakajima, and M. Nakamura (1994), Nakajima, M. Nakamura, and Yoshioka (1998, 2001), our approach allows us to exploit the cross-sectional dimension of grouped plant-level panel data to obtain improved estimates of returns to scale parameters.

In the empirical portion of this study, we analyze Japanese establishments in three major industries over the turbulent 1964–88 period using two-year rolling panels of establishment-level data. The period spanned by our data includes the high-growth era of the 1960s, the two oil shocks, and the slow growth years of the 1980s. For these industries, we have compiled a narrative record of events relating to demand fluctuations. Our results on time variation in returns to scale in the Japanese economy are intuitive in light of the narrative evidence on economic occurrences and government policies over this period.

Appendix A: Törnqvist index number measures of TFPG

The natural logarithm of a Törnqvist (1936) output quantity index is[26]

$$\ln Q_T = (1/2)\sum_{m=1}^{M}\left[\left(p_m^s y_m^s \Big/ \sum_{i=1}^{M} p_i^s y_i^s\right) + \left(p_m^t y_m^t \Big/ \sum_{j=1}^{M} p_j^t y_j^t\right)\right] \ln\left(y_m^t / y_m^s\right). \tag{A1}$$

The Törnqvist input quantity index Q_T^* is defined analogously as

$$\ln Q_T^* = (1/2)\sum_{n=1}^{N}\left[\left(w_n^s x_n^s \Big/ \sum_{i=1}^{N} w_i^s x_i^s\right) + \left(w_n^t x_n^t \Big/ \sum_{j=1}^{N} w_j^t x_j^t\right)\right] \ln\left(x_n^t / x_n^s\right). \tag{A2}$$

26 Törnqvist (1936) indexes are also known as translog indexes following Jorgenson and Nishimizu (1978), who introduced this terminology because Diewert (1976, 120) related the indexes to a translog production function.

Reversing the role of the prices and quantities in (A1) yields the Törnqvist output price index, P_T, defined by

$$\ln P_T = (1/2)\sum_{m=1}^{M}\left[\left(p_m^s y_m^s \Big/ \sum_{i=1}^{M} p_i^s y_i^s\right) + \left(p_m^t y_m^t \Big/ \sum_{j=1}^{M} p_j^t y_j^t\right)\right] \ln\left(p_m^t / p_m^s\right). \tag{A3}$$

The input price index P_T^* is defined in a similar manner. As Diewert and Nakamura (2007) explain, a price index is the implicit counterpart of a volume index if the product rule (also called the product test or axiom) is satisfied.[27] This rule requires that the product of the volume and price indexes must equal the cost ratio for input indexes or the revenue ratio for output indexes. The implicit Törnqvist output quantity index,
$\tilde{Q}_T$, is defined implicitly by

$$(R^t/R^s)/P_T \equiv \tilde{Q}_T \tag{A4}$$

(see Diewert 1992, 181), and the implicit Törnqvist input quantity index, $\tilde{Q}_T^*$, is defined analogously using the cost ratio and P_T^*(see Diewert 2005, 39; Diewert and Nakamura 2007, s. 3.5). The implicit Törnqvist output price index, $\tilde{P}_T$, is given by

$$(R^t/R^s)/Q_T \equiv \tilde{P}_T, \tag{A5}$$

and the implicit Törnqvist input price index, $\tilde{P}_T^*$, is defined analogously.

Appendix B: The equivalence of the primal and dual definitions of the elasticity of scale

We briefly summarize the results on this equivalence below for the one output, many inputs $(1 - N)$ case, and then for a more realistic, many outputs, many inputs $(M\text{-}N)$ production situation.

B.1. The 1 − N case

If we let $y = f(x)$ denote a production function, the measure of returns to scale, $\gamma(x)$, is defined here in the same manner as in (8) (where $x >> 0_N$ and $f(x) > 0$).

Ohta (1974) shows that the direct production function measure and the cost function based elasticity of scale are equal for the $1 - N$ case. Here we re-derive this result using a new approach that facilitates understanding the methodological interrelationships and derivation of the $M\text{-}N$ case (below) that Ohta does not provide.

27 For more on the properties of direct versus implicit indexes, see Allen and Diewert (1981).

PROPOSITION 1. *Given $y > 0$ and $w >> 0_N$, suppose $x^* >> 0_N$ solves the cost minimization problem*:

$$\min_x\{w \cdot x : y = f(x)\} = C(y, w). \tag{B3}$$

Suppose further that f is once continuously differentiable at the point x^* and the gradient vector of the production function is weakly positive at this point so that $\nabla f(x^*) > 0_N$. Then the dual cost function measure of returns to scale is equal to the primal production function measure of returns to scale; that is, we have the following equality:

$$\beta(y, w) = \gamma(x^*). \tag{B4}$$

Proof. Form the Lagrangian for the cost minimization problem (B3):

$$L(x, \mu) \equiv w \cdot x - \mu[f(x) - y]. \tag{B5}$$

Using the differentiability of the production function and the assumption that x^* solves (B3), the following first-order necessary conditions for x^* to solve (B3) must be satisfied for some μ^*:[28]

$$\nabla_x L(x^*, \mu^*) = w - \mu^* \nabla f(x^*) = 0_N \tag{B6}$$

$$\nabla_\mu L(x^*, \mu^*) = -[f(x^*) - y] = 0. \tag{B7}$$

Take the inner product of equations (B6) with x^*. The resulting equations imply that

$$\mu^* x^* \cdot \nabla f(x^*) = w \cdot x^* = C(y, w), \tag{B8}$$

where the last equation follows, since x^* is a solution to the cost minimization problem (B3). Thus, we obtain

$$\mu^* = C(y, w)/x^* \cdot \nabla f(x^*) > 0, \tag{B9}$$

where the inequality follows from our assumptions that $x^* >> 0_N$, $w >> 0_N$ and $\nabla f(x^*) > 0_N$. Finally, Samuelson's (1947, 34) Envelope Theorem implies that

$$\partial C(y, p)/\partial y = \partial L(x^*, \mu^*, w, y)/\partial y = \mu^*, \tag{B10}$$

28 We also require that the classical constraint qualification condition (that $\nabla f(x^*)$ be a non-zero vector) hold in order to ensure that conditions (B6) and (B7) hold. That $\nabla f(x^*) \neq 0_N$ follows from assuming $\nabla f(x^*) > 0_N$.

where the last equality follows by differentiating $L(x, \mu, w, y)$ with respect to y where L is given by

$$L(x, \mu, w, y) \equiv w \cdot x - \mu[f(x) - y]. \tag{B11}$$

According to the Envelope Theorem, the impact of a small parameter change on maximum revenue or minimum cost will be the same with, and without, full adjustment of all decision variables to the new parameter value (Samuelson 1947; 1983, 34). Indirect effects on the value of the decision maker's objective by way of adjustments of the decision variables do not matter. Thus marginal cost, $\partial C(y, p)/\partial y$, is equal to the optimal Lagrange multiplier, μ^*. Substituting (B10) into (B9) gives us the following expression for $x^* \nabla f(x^*)$:

$$x^* \nabla f(x^*) = C(y, w)/\partial C(y, p)/\partial y. \tag{B12}$$

Thus, from (B1), we have

$$\gamma(x^*) \equiv x^* \nabla f(x^*)/f(x^*) = \beta(y, w) \tag{B13}$$

using (B10) and $y = f(x^*)$. □

B.2. The M-N Case

For the many outputs and many inputs case, Panzar and Willig (1977, 488) defined the returns to scale using the cost function as follows:

$$\beta(y, w) \equiv C(y, w) \Bigg/ \sum_{m=1}^{M} [\partial C(y, w)/\partial y_m] y_m = C(y, w)/y \cdot \nabla_y C(y, w)/C(y, w). \tag{B14}$$

To show the equivalence between the dual cost function and the multi-output counterpart of the primal definition of returns to scale, the latter must be defined. We could represent the technology (locally) using a production function:

$$y_1 = f(y_2, \dots, y_M, x), \tag{B15}$$

where the right-hand expression is the maximum that can be produced for output 1 given that $y_2, \dots, y_M$ must be produced and that the input vector x is available to the producer, and where f is non-increasing in $y_2, \dots, y_M$ and non-decreasing in the components of x. Or, we could represent the technology using an input requirements function:

$$x_1 = g(y, x_2, \dots, x_N), \tag{B16}$$

where $g(y, x_2, \ldots, x_N)$ is the minimum amount of input 1 that is required to produce the vector of outputs y, given that amounts $x_2, \ldots, x_N$ of inputs 2 to N are available to the producer, and where g is non-decreasing in the components of y and non-increasing in $x_2, \ldots, x_N$. Or, following Caves, Christensen, and Diewert (1982, 1402), we could restate both of the above representations of the technology using a Hicksian transformation function, $t(y, x)$. Thus, we can represent both (B15) and (B16) by means of the following constraint, where we assume that t is non-decreasing in the components of x and non-increasing in the components of y:[29]

$$t(y, x) = 0. \tag{B17}$$

We can now consider how to locally define returns to scale in a many output, many input setting. Following Caves, Christensen, and Diewert (1982), suppose we increase all inputs by λ. Suppose, also, that we let $u(\lambda, y, x)$ be the factor of proportionality by which all outputs must be increased, so that the inflated input and output vectors are on the production surface. Thus, $u(\lambda, y, x)$ is defined implicitly[30] by the following equation:

$$t(u(\lambda, y, x)y, \lambda x) = 0. \tag{B18}$$

$\gamma(y, x)$ is now defined as the rate of change of u with respect to a change in λ, evaluated at $\lambda = 1$:

$$\gamma(y, x) \equiv \partial u(\lambda, y, x)/\partial\lambda_{\lambda=1}. \tag{B19}$$

In order to determine the derivative on the right-hand side of (B19), differentiate both sides of (B18) with respect to λ and evaluate the resulting derivatives at $\lambda = 1$. We obtain the following equation:

$$y \cdot \nabla_y t(y, x)\gamma(y, x) + x \cdot \nabla_x t(y, x) = 0. \tag{B20}$$

If $y \cdot \nabla_y t(y, x)$ is not zero, we obtain the following formula for the primal measure in the M-N case:

$$\gamma(y, x) \equiv -x \cdot \nabla_x t(y, x)/y \cdot \nabla_y t(y, x). \tag{B21}$$

Note that since $t(y, x)$ is non-increasing in the components of y and non-decreasing in the components of x, $\gamma(y, x)$ must be nonnegative. Also, if we are

29 Note that in order to convert the representation of the technology given by (B15) into the representation given by (B17), we need only define $t(y, x) \equiv -y_1 + f(y_2, \ldots, y_M, x)$. In order to convert the representation of the technology given by (B16) into the representation given by (B17), we need only define $t(y, x) \equiv x_1 - g(y, x_2, \ldots, x_N)$.

30 In order to ensure the existence of the implicit function $u(\lambda, y, x)$ in a neighborhood of $u(1, y, x^*)$, we need to assume that $y \cdot \nabla t(y, x^*) \neq 0$. We will assume that $\nabla t(y, x^*) < 0_M$ which will imply $y \cdot \nabla t(y, x^*) \neq 0$.

in the single output case and $t(y_1, x) = -y_1 + f(x)$, where f is the usual 1-N case production function, then (B21) becomes the usual 1-N case measure of returns to scale:

$$\gamma(y_1, x) \equiv x \cdot \nabla_x f(x)/y_1 = x \cdot \nabla_x f(x)/f(x). \tag{B22}$$

The equivalence between the primal and the dual cost-function based approaches was established for the many output-many input case by Panzar and Willig (1977, 486–90) and Caves, Christensen, and Swanson (1981, 995).[31] We now prove the result under weaker regularity conditions that the cost function measure of returns to scale, $\beta(y, w)$ defined by (B14), is equal to the transformation function measure of returns to scale, $\gamma(y, x)$ defined by (B21); that is, we prove the following equality: $\beta(y, w) = \gamma(y, x^*)$. Therefore, either approach can be used (and is equivalent) for estimating returns to scale.

PROPOSITION 2. *Given $y >> 0_M$ and $w >> 0_N$, suppose $x^* >> 0_N$ solves the cost minimization problem:*

$$\min_x\{wx : t(y, x) = 0\} \equiv C(y, w). \tag{B23}$$

Suppose, further, that t is once continuously differentiable at the point (y, x^*) and the gradient vector of the transformation function is weakly positive with respect to x and weakly negative with respect to y at this point, so $\nabla_x t(y, x^*) > 0_N$ and $\nabla_y t(y, x^*) < 0_M$. Then the cost function measure, $\beta(y, w)$ defined by (B14), is equal to the transformation function measure, $\gamma(y, x)$ defined by (B23); that is, we have the following equality:

$$\beta(y, w) = \gamma(y, x^*). \tag{B24}$$

Proof. Form the Lagrangian for the cost minimization problem (B23):

$$L(x, \mu) \equiv wx - \mu[t(y, x)]. \tag{B25}$$

Given the differentiability of t and assumptions that x^* solves (B23) and $\nabla_x t(y, x^*) > 0_N$, so the constraint qualification holds, then the first-order necessary conditions for x^* to solve (B23) are satisfied for some μ^*:

$$\nabla_x L(x^*, \mu^*) = w - \mu^* \nabla_x t(y, x^*) = 0_N \tag{B26}$$

31 Caves, Christensen, and Swanson (1981, 995) used a similar framework but they did not spell out the mathematical details. Panzar and Willig (1977) used a different framework to define returns to scale in the primal, but in the end they did arrive at the primal formula [(C21) here] reported here.

$$\nabla_{\mu} L(x^*, \mu^*) = -t(y, x^*) = 0. \tag{B27}$$

Take the inner product of equations (B26) with respect to x^*. The resulting equations imply that

$$\mu^* x^* \cdot \nabla_x t(y, x^*) = w \cdot x^* = C(y, w), \tag{B28}$$

where the last equation follows since x^* is a solution to the cost minimization problem (B23). Thus, we obtain

$$\mu^* = C(y, w)/x^* \cdot \nabla_x t(y, x^*) > 0, \tag{B29}$$

where the inequality follows from assuming $x^* >> 0_N$, $w >> 0_N$ and $\nabla_x t(y, x^*) > 0_N$. Thus, from (B29), we have

$$x^* \cdot \nabla_x t(y, x^*) = C(y, x^*)/\mu^*. \tag{B30}$$

To show the relationship of (B14) and (B21), we need Samuelson's (1947, 34) Envelope Theorem; it implies

$$\nabla_y C(y, p)/\nabla_y L(x^*, \mu^*, w, y) = -\mu^* \nabla_y t(y, x^*), \tag{B31}$$

where this equality follows by differentiating $L(x, \mu, w, y)$ with respect to y, where L is defined more fully as

$$L(x, \mu, w, y) \equiv w \cdot x - \mu[t(y, x)]. \tag{B32}$$

Now, inner product both sides of (B31) with the vector y in order to obtain the following equation:

$$y \cdot \nabla_y t(y, x^*) = -y \cdot \nabla_y C(y, p)/\mu^*. \tag{B33}$$

From (B21) and using our assumption that $\nabla_y t(y, x) < 0_M$, so that $y \cdot \nabla_y t(y, x) < 0$, we have

$$\begin{aligned} \gamma(y, x) &\equiv -x \cdot \nabla_x t(y, x)/y \cdot \nabla_y t(y, x) \\ &= [C(y, x^*)/\mu^*]/y \cdot \nabla_y C(y, p)/\mu^* \quad \text{using (B30) and (B33)} \\ &= \beta(y, w) \quad \text{by definition where } \beta \text{ and } \gamma \text{ must be positive.} \end{aligned} \tag{B34}$$

So, the cost function based definition of returns to scale equals the direct transformation function definition. □

Appendix C: Details for the production function framework

Consider the cost minimization problem given by

$$\min_{x}\left\{\sum_{n=1}^{N} w_n x_n : Y = f^t(x)\right\}. \tag{C1}$$

Solving the minimization problem given in (C1) yields the following N first order necessary conditions.

$$w_n = \lambda[\beta_n + \zeta_{nj} \ln x_j^t]/x_n^t, n = 1, \ldots, N. \tag{C2}$$

From these and the period s translog production function given in (12), we can determine x and the Langrange multiplier λ. Multiplying (C2) through by x_n^t and summing over the N inputs yields

$$\begin{aligned}\sum_{n=1}^{N} w_n^t x_n^t &= \lambda \sum_{n=1}^{N}\left[\beta_n + \sum_{j=1}^{N} \zeta_{nj} \ln x_j^t\right] \\ &= \lambda \sum_{n=1}^{N} \beta_n = \lambda\gamma_P, \text{ using (14) and then (13).}\end{aligned} \tag{C3}$$

Solving (C3) for λ, substituting the resulting expression into (C2), and then multiplying through by x_n^t yields

$$\gamma_P c_n^t = f_n^s(x_n^t)x_n^t/f_n^s(x^t), \quad n = 1, \ldots, N; \quad t = 1, \ldots, T, \tag{C4}$$

where c_n^t denotes the period t share of total cost expended for input n. The right-hand side of the base period s production function specified in (12) is quadratic in the logarithms of the input quantities. The Diewert Quadratic Identity (Diewert 1976, lemma 2.2) can be used to obtain an expression relating the change in the logarithms of the input quantities going from s to t to the corresponding change in the logarithm of the output quantity:

$$\begin{aligned}\ln f^t(x^t) - \ln f^s(x^s) &= \theta(t-s) + \ln f^s(x^t) - \ln f^s(x^s) \\ &= \theta(t-s) + (1/2)\sum_{n=1}^{N}\{[f_n^s(x^s)x_n^s/f^s(x^s)] \\ &\quad + [f_n^s(x^t)x_n^t/f^s(x^t)]\}[\ln x_n^t - \ln x_n^s] \\ &\quad \text{since } \ln f^s(x) \text{ is quadratic in the variables } \ln x_n \\ &= \theta(t-s) + (1/2)\sum_{n=1}^{N}\{[\gamma_P c_n^s] + [\gamma_P c_n^t]\}[\ln x_n^t - \ln x_n^s] \\ &= \theta(t-s) + \gamma_P \ln Q_T^{*s,t}, \end{aligned} \tag{C5}$$

where $\ln Q_T^{*s,t}$ is the logarithm of the Törnqvist input quantity index given by

$$\ln Q_T^{*s,t} = (1/2)\sum_{n=1}^{N}\left[\left(w_n^s x_n^s \Big/ \sum_{i=1}^{N} w_i^s x_i^s\right) + \left(w_n^t x_n^t \Big/ \sum_{j=1}^{N} w_j^t x_j^t\right)\right]\ln\left(x_n^t/x_n^s\right). \tag{C6}$$

$f^t(x^t)$ and $f^s(x^s)$ on the left-hand side of (C5) can be replaced by observable quantities, Y^t and Y^s.

Appendix D: Examples of increasing returns to scale

We searched through both the current and classic literatures for examples of why returns to scale matter. What follows is a small sample of what we found. We organize the examples according to the type of input factor said to be economized on as the scale of production is increased. For each example, we maintain that the *different input mix* used for larger-scale production is a key source of the claimed reduction in unit output cost with greater scale. The input factors are taken up in the order of the KLEMS paradigm (K = capital, L = labour, E = energy, M = materials, S = savings).

D.1. Scale-related savings on plant and equipment and other capital inputs

In addition to laying the foundations of modern economics, Smith (1963, 7) gave a number of examples of increasing returns to scale that are still often quoted as relevant.[32] He notes, for instance, that equipment typically comes in discrete

32 Though the publication dates for available editions of his work are in the 1900s, Smith lived from 1723 to 1790.

sizes. Once durables are chosen, the expenses become fixed costs. Cost savings per unit of output can result from averaging fixed costs over higher output levels.

Larger producers can use higher-capacity equipment and other facilities, and physical laws result in potentially lower unit costs for larger facilities. For example, Kaldor (1972, 1253) observes: 'there is the important group of cases ... due to the three dimensional nature of space.' Lipsey (2000) elaborates: 'The geometrical relation governing any container typically makes the amount of material used, and hence its cost (given constant prices of the materials with which it is made), proportional to *one dimension less* than the service output, giving *increasing returns to scale* ... Blast furnaces, ships, and steam engines are a few examples of the myriad technologies that show such geometrical scale effects' (emphasis added).

Inventories are also part of business capital. Edgeworth (1888, 124) applied the Law of Large Numbers to the inventory-stocking problem and derived the rule that optimal inventory stocks are proportional to the square root of anticipated demands. Diewert (2005) notes that the square root inventory replenishment rule was soon widely adopted by classical industrial engineers (e.g., Green 1915; Harris 1915, 48–52) and has been incorporated, too, into the models of modern economists, including Allais (1947, 238–41), Baumol (1952), and Tobin (1956). Research and development is an intangible capital investment activity, where increasing returns are said to matter (Huang and Diewert 2010). The non-rival nature within a business of many sorts of intangible capital is another claimed source of savings with greater scale.[33] Arnold (2004, 13) explains: 'Consider a firm that produces $10 million using one plant, 10 workers, and a certain amount of technical knowledge. This firm could produce $20 million by building a second plant, hiring 10 more workers, and *using the same knowledge*. Since output has doubled *with a less-than-doubling of inputs* ... production is characterized by increasing returns to scale' (emphasis added).

Finally, Lipsey and Carlaw (2004) argue convincingly that capital-related scale economies are typically made invisible by the pervasive practice of measuring capital input by the capital service flow without attention to how that flow is produced or how the scale of that flow can be altered. If these arguments were to be taken seriously, much of the existing empirical productivity literature would need to be revised and returns to scale would surely prove more important.

D.2. Aggregate labour-saving scale effects

Adam Smith (1963, 7) also gave examples of how labour savings can arise with greater scale: 'This great increase of the quantity of work, which, in consequence of the division of labour, the same number of people are capable of performing, is owing to three things; first, to the increase of dexterity in every particular workman; secondly, to the saving of time which is commonly lost in passing

33 Business process knowledge and development is relevant in this regard (see Alexopoulos and Cohen 2010; Alexopoulos and Tombe 2009).

from one species of work to another; and lastly, to the invention of a great number of machines which facilitate and abridge labour, and enable one man to do the work of many.' Young (1928, 530) notes that labour costs may be saved, too, by the substitution of machines for workers: 'It would be wasteful to make a hammer to drive a single nail; it would be better to use whatever awkward implement lies at hand. It would be wasteful to furnish a factory with an elabourate equipment . . . to build a hundred automobiles; it would be better to rely mostly upon tools and machines of standard types, so as to make a relatively larger use of directly-applied and a relatively smaller use of indirectly-applied labour. Mr. Ford's methods would be absurdly uneconomical if his output were very small.' Babbage (1835, 175) acknowledges the contributions of Smith, and adds an insight of his own: 'That the master manufacturer, by dividing the work to be executed into different processes, each requiring different degrees of skill or force, can purchase exactly that precise quantity of both which is necessary for each process; whereas, if the whole work were executed by one workman, that person must possess sufficient skill to perform the most difficult, and sufficient strength to execute the most labourious, of the operations into which the art is divided.' Also, when the size of a piece of equipment is increased, often only one driver or operator is still needed.

D.3. Scale-related energy and materials savings

Energy and materials savings are attributed to greater scale too. Marshall (1920, 290) writes:[34] 'A ship's carrying power varies as the cube of her dimensions, while the resistance offered by the water increases only a little faster than the square of her dimensions; so that a large ship requires less coal in proportion to its tonnage than a small one.'Babbage (1935, 174) attributes materials savings with greater scale to reduced worker error rates: 'A certain quantity of material will, in all cases, be consumed unprofitably, or spoiled by every person who learns an art . . . if each person confine his attention to one process . . . the division of labour will diminish the price of production.'

D.4. Scale-related savings on service inputs

The Law of Large Numbers can enable savings with greater scale for many business services, too. For example, a larger bank requires less in the way of cash reserves to meet random demands for savings withdrawals. Similarly, almost all sorts of insurance will usually be cheaper for larger firms.[35]

34 Although the publication dates of Marshall's work are typically more recent, he lived from 1842 to 1924.

35 Diewert (2005) explains that Edgeworth (1888, 122) and Whitin (1952, 506–11; 1957, 234–6) were among those who developed these applications of probability theory.

References

Alexopoulos, M., and J. Cohen (2011) 'Volumes of evidence: examining technical change in the last century through a new lens,' *Canadian Journal of Economics* 44, 413–50

Alexopoulos, M., and T. Tombe (2009) 'Management matters,' working paper, University of Toronto

Allais, M. (1947) *Economie et Intérêt* (Imprimerie Nationale)

Arnold, R.W. (2004) 'Modeling long-run economic growth,' Technical Paper Series, Paper 2003–4, Congressional Budget Office http://www.cbo.gov/ftpdocs/42xx/doc4284/2003-4.pdf

Babbage, C. (1835) *On the Economy of Machinery and Manufactures*, 4th ed. Reprint (New York: A.M. Kelley, 1965)

Basu, S., and J. F. Fernald (1995) 'Are apparent productive spillovers a figment of specification error?' *Journal of Monetary Economics* 36, 165–88

Basu, S. and J. F. Fernald (1997) 'Returns to Scale in U.S. Production: Estimates and Implications,' *Journal of Political Economy*, 105, 249–83

Baumol, W.J. (1952) 'The transactions demand for cash: an inventory theoretic approach,' *Quarterly Journal of Economics* 66, 545–56

Beason, R., and D.E. Weinstein (1996) 'Growth, economies of scale, and targeting in Japan (1955–1990),' *Review of Economics and Statistics* 78, 286–95

Burbidge, J. and A. Harrison (1984) 'Testing for the effects of oil-price rises using vector autoregressions,' *International Economic Review* 25, 459–84

Burnside, C. (1996) 'Production function regressions, returns to scale and externalities,' *Journal of Monetary Economics* 37, 177–201

Burnside, C., M. Eichenbaum, and S. Rebelo (1995) 'Capital utilization and returns to scale,' *NBER Macroeconomics Annual* 10, 67–110

Caves, D. W., L. Christensen, and W. E. Diewert (1982) 'The economic theory of index numbers and the measurement of input, output, and productivity,' *Econometrica* 50, 1393–414

Caves, D.W., L.R. Christensen, and J.A. Swanson (1981) 'Productivity growth, scale economies, and capacity utilization in U.S. railroads, 1955–74,' *American Economic Review* 71, 994–1002

Chan, M.W.L., and D.C. Mountain (1983) 'Economies of scale and the Törnqvist discrete measure of productivity growth,' *Review of Economics and Statistics* 65, 663–7

Diewert, W.E. (1976) 'Exact and superlative index numbers,' *Journal of Econometrics* 4, 221–56

— (1978) 'Hicks' Aggregation Theorem and the existence of a real value-added function,' in *Production Economics: A Dual Approach to Theory and Applications*, Vol. 2, ed. M. Fuss and D. McFadden (Amsterdam: North-Holland). Reprinted as chap. 15 in Diewert and Nakamura (1993, 435–70)

— (1981) 'The economic theory of index numbers: a survey,' in *Essays in the Theory and Measurement of Consumer Behaviour in Honour of Sir Richard Stone*, ed. A. Deaton (London: Cambridge University Press). Reprinted as chap. 7 in Diewert and Nakamura (1993, 177–221)

— (2002) 'The quadratic approximation lemma and decompositions of superlative indexes,' *Journal of Economic and Social Measurement* 28, 63–88

— (2005) 'The measurement of performance: productivity versus the real rate of return,' in *The Measurement of Business Capital, Income and Performance*, tutorial University Autonoma of Barcelona, Spain, 21–22 September. http://faculty.arts.ubc.ca/ediewert/barcelona.htm

Diewert, W.E., and K.J. Fox (2008) 'On the estimation of returns to scale, technical progress and monopolistic markups,' *Journal of Econometrics* 145, 174–93

— (2010) 'Malmquist and Törnqvist productivity indexes: returns to scale and technical progress with imperfect competition,' *Journal of Economics*, forthcoming

Diewert, W. E. and A. O. Nakamura (1999), 'Benchmarking and the measurement of best practice efficiency: an electricity generation application,' *Canadian Journal of Economics* 32, 570–88

— (2003), 'Index number concepts, measures and decompositions of productivity growth,' *Journal of Productivity Analysis* 19, 127–60

— (2007) 'The measurement of productivity for nations,' in *Handbook of Econometrics*, Vol. 6, ed. J.J. Heckman and E.E. Leamer (Elsevier)

Diewert, W. E., and A. O. Nakamura, eds (1993) *Essays in Index Number Theory*, Vol. 1 (Amsterdam: North-Holland)

Eatwell, J. (1987) 'Returns to scale,' *The New Palgrave: A Dictionary of Economics*, Vol. 4, 165–6

Edgeworth, F.Y. (1888) 'The mathematical theory of banking,' *Journal of the Royal Statistical Society* 51, 113–27

Federal Reserve Bank of Boston (2004) 'What is productivity?' http://www.bos.frb.org/education/ledger/ledger04/winter/whatisprod.pdf

Freeman, R.B., A.O. Nakamura, L.I. Nakamura, M. Prud'homme, and A. Pyman (2011) 'Walmart innovation and productivity: a viewpoint,' *Canadian Journal of Economics* 44, 486–508

Green, J.B. (1915) 'The perpetual inventory in practical stores operation,' *Engineering Magazine* 48, 879–888

Hall, R.E. (1990) 'Invariance properties of Solow's Productivity Residual,' in *Growth/Productivity/Employment*, ed. P. Diamond (Cambridge, MA: MIT Press)

Harberger, A.C. (1997) 'Studying the growth process: a primer,' prepared for the Conference on Capital Formation and Economic Growth, Hoover Institution, Stanford University

Harberger, A.C. (1998) 'A vision of the growth process,' *American Economic Review* 88, 1–32

Harris, F.W. (1915) *Operations and Cost* (A.W. Shaw)

Huang, N., and W.E. Diewert (2011) 'Estimation of R&D depreciation rates: a suggested methodology and preliminary application,' *Canadian Journal of Economics* 44, 387–412

Johnson, C. (1982) *MITI and the Japanese Miracle: the Growth of Industrial Policy, 1925–1975* (Stanford, CA: Stanford University Press)

Jorgenson, D.W., and M. Nishimizu (1978) 'U.S. and Japanese economic growth, 1952–1974: an international comparison,' *Economic Journal* 88, 707–26

Kaldor, N. (1972) 'The irrelevance of equilibrium economics,' *Economic Journal* 82, 1237–55

Komiya, R., M. Okuno, and K. Suzumura, eds (1988) *Industrial Policy of Japan*, trans. under the supervision of K. Sato (Academic Press)

Leontief, W. (1936) 'Composite commodities and the problem of index numbers,' *Econometrica* 4, 39–59

Lipsey, R.G., and K.I. Carlaw (2004) 'Total factor productivity and the measurement of technological change,' *Canadian Journal of Economics* 37, 1118–50

Mohnen, P., R. Jacques, and J.S. Gallant (1996) 'Productivity and research and development in two Canadian forest product industries,' *Forest Science* 42, 447–87

Marshall, A. (1898/1920) *Principles of Economics*, 4th ed. (London: Macmillan)

Nakajima, T., M. Nakamura, and K. Yoshioka (1998) 'An index number method for estimating scale economies and technical progress using time series of cross section data: sources of total factor productivity for Japanese manufacturing, 1964–1988,' *Japanese Economic Review* 49, 310–34

Ohta, M. (1974) 'A note on the duality between production and cost functions: rate of returns to scale and rate of technical progress,' *Economic Studies Quarterly* 25, 63–5
Okimoto, D.I. (1989) *Between MITI and the Market: Japanese Industrial Policy for High Technology*, (Stanford, CA: Stanford University Press)
Quah, D. (2003) 'Digital goods and the new economy,' in *New Economy Handbook*, ed. D. Jones (Academic Press / Elsevier Science)
Panzar, J.C. (1989) 'Technological determinants of firm and industry structure,' in *Handbook of Industrial Organization*, ed. R. Schmalensee and R.D. Willig, Vol. 1 (Elsevier Science)
Panzar, J.E., and R.D. Willig (1977) 'Economies of scale in multi-output production,' *Quarterly Journal of Economics* 91, 481–93
Pappens, R.A. (1994) 'Industry dusts off plans as future looks bright,' *Pulp and Paper International* 36(11), 23–6
Patrick, H., ed. (1986) *Japan's High Technology Industries* (University of Washington Press)
Porter, M.E. (1998) *The Competitive Advantage of Nations* (Free Press)
Porter, M.E., and M. Sakakibara (2004) 'Competition in Japan,' *Journal of Economic Perspectives* 18, 27–50
Roy, A. (2005) 'Industrial development measures: what India can learn from Japan,' CSIRD Discussion Paper 12/2005 http://www.csird.org.in/pdf/DP%2012.pdf
Samuelson, P.A. (1947) *Foundations of Economic Analysis* (Cambridge, MA: Harvard University Press)
Schreuder, G.F., and E.T. Anderson (1988) 'International wood chip trade: past developments and future trends with particular emphasis on Japan,' in *Global Issues and Outlook in Pulp and Paper* ed. G.F. Schreuder (University of Washington Press)
Schreyer, P. (2001) *Measuring Productivity: Measurement of Aggregate and Industry-Level Productivity Growth*, OECD manual (Paris: OECD)
Shimokawa, E. (1977) 'Japan's dependence upon wood chips for pulp,' *Unasylva* 29(115), 26–7.
Smith, A. (1963) *The Wealth of Nations*, Vol. 1 [1776] (Richard D. Irwin)
Solow, R.M. (1957) 'Technical change and the aggregate production function,' *Review of Economics and Statistics* 39, 312–20
Tobin, J. (1956) 'The interest elasticity of transactions demand for cash,' *Review of Economics and Statistics* 38, 241–7
Törnqvist, L. (1936) 'The Bank of Finland's Consumption Price Index,' *Bank of Finland Monthly Bulletin* 10, 1–8
Tsuruta, T. (1988) 'The rapid growth era,' chap. 3 in Komiya, Okunu, and Suzumura (1988, 49–88)
Walters, A.A. (1963) 'Production and cost functions: an econometric survey,' *Econometrica* 31, 1–66
Yoshioka, K., T. Nakajima, and M. Nakamura (1994) *Sources of Total Factor Productivity for Japanese Manufacturing Industries, 1964–1988: Issues in Scale Economies, Technical Progress, Industrial Policies and Measurement Methodologies*, Monograph No. 5, Keio Economic Observatory, Keio University
Young, A.A. (1928) 'Increasing returns and economic progress,' *Economic Journal* 38, 527–42