TESTING SUBSTITUTION BIAS OF THE SOLOW-RESIDUAL MEASURE OF TOTAL FACTOR PRODUCTIVITY USING CES-CLASS PRODUCTION FUNCTIONS*

(Incomplete Draft)

Baoline Chen
Bureau of Economic Analysis
1441 L Street, NW
Washington, DC 20230
e-mail: baoline.chen@bea.gov

and

Peter A. Zadrozny
Bureau of Labor Statistics
2 Massachusetts Ave., NE
Washington, DC 20212
e-mail: zadrozny_p@bls.gov

January 26, 2005

Additional key words: Taylor-series approximation, model selection, numerical solution, tiered CES production function

JEL codes: C32, C43, C53, C63

---

*Material in this paper was presented at the Conference on Index Number Theory and Measurement of Productivity in Vancouver, Canada, on July 3, 2004. The paper's analysis and conclusions represent the authors' views and do not necessarily represent any positions of the Bureau of Economic Analysis or the Bureau of Labor Statistics.

ABSTRACT

For period t, let $q_t = f(v_t) + \tau_t$, where q_t denotes measured output quantity, $f(\cdot)$ denotes a production function, $v_t = (v_{1t}, ..., v_{nt})^T$ denotes a vector of n input quantities, τ_t denotes total factor productivity (TFP), and all variables are in natural-log form. Then, $f(v_t) = \sum_{i=1}^{n} \alpha_{it} v_{it}$, for $0 < \alpha_{it} < 1$ and $\sum_{i=1}^{n} \alpha_{it} = 1$, is a Cobb-Douglas first-order log-form approximation of a production function (denoted CD). If $f(\cdot)$ is approximated as a CD function, the share parameters, α_{it}, are set to successive two-period-averaged cost shares, and the observed input quantities are considered optimal or input-cost minimizing, then, $\tau_t = q_t - \sum_{i=1}^{n} \alpha_{it} v_{it}$ is the log-form Solow-residual measure of TFP (Solow, 1957). Solow-residual TFP could be subject to input-substitution bias for two reasons. First, the CD production function restricts all input substitutions to one. Second, observed inputs generally differ from optimal inputs, so that inputs observed in a sample tend to move not just due to substitution effects but for other reasons as well. In this paper, we test the possible input-substitution bias of the Solow-residual measure of TFP in capital, labor, energy, materials, and services (KLEMS) inputs data for U.S. manufacturing from 1949 to 2001. (1) Based on maximum likelihood estimation, we determine a best 4th-order approximation of a CES-class production function. The CES class includes not only the standard constant elasticity of input substitution production functions (denoted CES) but also includes so called tiered CES production functions (denoted TCES), in which prespecified groups of inputs can have their own input-substitution elasticities and input-cost shares are parameterized (i) tightly as constants, (ii) moderately as smooth functions, and (iii) loosely as successive averages. (2) Based on the best estimated production function, we compute the implied best TFP as $\tau_t = q_t - f(\hat{v}_t)$, where $f(\hat{v}_t)$ denotes the best estimated production function evaluated at the computed optimal inputs, $\hat{v}_t$. (3) For the data, we compute Solow-residual TFP and compare it with the best TFP. We conclude that for this data, the Solow-residual TFP is on average .1% lower, with a .6% standard error, than the best TFP and, hence, is very slightly downward biased, although the sampling-error uncertainty dominates this conclusion. In further work, we shall attempt to reduce this uncertainty with further testing based on more general CES-class production functions and more finely estimated parameters.

1. Introduction.

The paper is specifically motivated as discussed in the preceeding abstract, but is also more generally motivated by the desire to accurately compute price indexes based on explicit forms of the functions being maximized. There are two main, mathematically identical, but economically different applications: computing price indexes of production inputs based on maximizing output of a production function for given input costs, as here, and computing price indexes of consumer goods based on maximizing utility of consumed goods for given expenditures, as in Zadrozny and Chen (2004). Here, we consider standard constant elasticity of input substitution production functions (denoted CES), with one input-substitution elasticity for all inputs, and more general tiered CES production functions, with a different input-substitution elasticity for each group of inputs (denoted TCES).

We are also interested in using even more general production functions, which we call generalized CES production functions (denoted GCES), in which each input or good can have its own price elasticity parameter, but, for brevity, limit the present applications to CES and TCES production functions. CES and TCES production functions imply analytical solutions of their optimization problems. GCES production functions generally do not imply analytical solutions except in special homothetic cases, such as the CES and TCES cases. Generally, optimization problems based on GCES production functions can be solved only numerically. In Zadrozny and Chen (2004), we describe the multi-step perturbation (MSP) method as a quick and accurate method for numerically solving the corresponding utility maximization problem.

Here, we could have used analytical CES and TCES solutions, but, for two reasons, use numerical solutions produced by the MSP method. First, we use the MSP method in order to test its accuracy in solving the static optimization problems. In all cases, we obtained nearly double-precision or about 14-decimal-digit accuracy when we checked the numerical MSP solutions against the analytical solutions, which encourages us to work in the future with purely numerical solutions of GCES production functions. Second, we are intersted in studying TFP bias by generalizing the CD production function by adding higher-order log-form Taylor-series terms up to a specified order. However, to do this tractably we must restrict the number of estimated parameters and we do this by parameterizing in terms of these CES-class production functions.

We proceed here entirely in log form for four reasons: (i) TFP and related price and quantity indexes are usually considered in log form; (ii) log-form variables are unit free, scaled equivalently, and, hence, lie mostly within or close to a unit sphere, which promotes numerical accuracy; (iii) log-form derivatives of the CES-class production functions are easier to derive, program, and compute with; and, (iv) comparisons with benchmark Solow residuals are easier in log form.

As noted, q denotes the log of the quantity of observed goods and services, $f(\cdot)$ denotes the log of output produced by the production function, and, $\tau = q - \hat{q}$ denotes the log of the level of technology or TFP of $f(\cdot)$, where $\hat{q} = f(\hat{v})$ denotes the log of optimal output produced by optimal log-form inputs, $\hat{v}$. To distinguish between q and $f(\cdot)$, we, respectively, refer to them as "goods and services" and "output." Let $p = (p_1, ..., p_n)^T$ denote an $n\times1$ vector of logs of observed or computed input prices (superscript T denotes vector or matrix transposition) and let $v = (v_1, ..., v_n)^T$ denote an $n\times1$ vector of logs of observed or computed input quantities. The context of whether inputs are observed or optimal-computed will be spelled out in each case. Whether prices are in nominal or real (deflated) units makes no difference, so long as real prices in a period are obtained by deflating each nominal price by the same value.

We assume $f(\cdot)$ is analytical, hence, for a sufficiently large k, $f(\cdot)$ is arbitrarily well approximated by a k+1-order Taylor series. Let $e(x) = (\exp(x_1), ..., \exp(x_n))^T$ for any $n\times1$ vector $x = (x_1, ..., x_n)^T$. We write the input-cost line as $e(p)^Te(\hat{v}) = e(p)^Te(v)$, where p and v are given, so that $e(p)^Te(v)$ denotes observed expenditures on inputs and optimal $\hat{v}$ is computed. We consider the following output maximization problem: for given $f(\cdot)$, p, and v, maximize $f(\hat{v})$ with respect to $\hat{v}$, subject to $e(p)^Te(\hat{v}) = e(p)^Te(v)$. Because τ is absent from the statement of the problem, it plays no role in its solution. Like Solow, we compute τ residually: first $\hat{v}$, then τ. The difference with Solow is that $\hat{v}$ is computed as optimal and is not equated with observed v.

We consider only interior solutions which satisfy the usual first-order conditions (2.1) and (2.2). As functional forms, we consider CD production functions, standard CES production functions, and more general TCES production functions, which are multi-level generalizations of two-level CES functions (Sato, 1967; Burnside, Eichenbaum, and Rebelo, 1995), that allow different input groups to have different substitution elasticities. For each production function, we solve for optimal inputs using the MSP method. In the CD, CES, and TCES cases, we use analytical solutions to check the MSP method's accuracy and in the future, given the successful application of MSP with the CD, CES, and TCES production functions demonstrated here, we shall consider the more general GCES production functions which do not imply analytical solutions.

By a model we mean (i) a multiple-times-differentiable production function, $f(\cdot)$, (ii) a parameterization of $f(\cdot)$ over a data sample, and (iii) values of constant structural parameters which determine $f(\cdot)$ in the sample. We now consider three parameterizations in more detail: (a) unrestricted time-varying reduced-form parameters set every period to different values of structural parameters; (b) time-varying reduced-form parameters restricted by a smooth function of constant structural parameters; and, (c) constant reduced-form parameters equal to constant structural parameters.

For example, $f(v_t) = \sum_{i=1}^{n} \alpha_{it} v_{it}$ denotes a period-t log-form CD production function for mean-adjusted data, whose reduced-form parameters, α_{it}, depend on constant structural parameters in the vector θ. In the typical case (a) of a data-producing agency, reduced-form parameters are unrestricted, are set period-by-period to relative input costs, and are statistically unreliable (have infinite estimated standard errors), because the number of estimated structural parameters, $\dim(\theta)$, equals the number of observations, nT: $\alpha_{it} = \theta_{it}$, for i = 1, ..., n and t = 1, ..., T, so that $\dim(\theta) = nT$. In the typical academic case (c) of an econometric analysis, the reduced-form parameters are fixed over a sample in terms of the structural parameters and are statistically reliable because there are fewer estimated structural parameters than observations: $\alpha_{it} = \theta_i$, so that $\dim(\theta) = n < nT$. In the application in section 3, we consider the in-between case (b), in which nT reduced-form parameters vary smoothly according to an integrated moving-average (IMA) process (Gardner, 1985), such that $\dim(\theta) < nT$.

What difference does the extra generality of going beyond the CD production function make? Normally, empirical validity is measured by residual size. In this case, we have output residuals, $q-\hat{q}$, and input residuals, $v-\hat{v}$. However, because TPF and output residuals are identical, judging TFP's empirical validity using sizes of output residuals makes no sense. For example, statistically ideal zero output residuals imply zero log-TFP. Thus, instead, we propose judging TFP's empirical validity using an information criterion (IC) based on input residuals. The many IC which have been proposed differ in their propensities for choosing models with particular numbers of parameters. For example, Akaike's IC (1973) often picks less parsimonious models (i.e., with more parameters), while Schwarz's IC (1978) often picks more parsimonious models.

As usual, for a given data sample, we consider a model's parameter estimates and derived quantities like TFP as statistically reliable when the parameter estimates and derived quantities have finite standard errors. This occurs if and only if the degrees of freedom of the parameter estimates are positive. Among the models being considered, the one which minimizes a chosen IC is considered the best or empirically-most-valid model. An IC test based on input residuals for choosing the best model for computing TFP has several advantages. First, the test's justification does not depend on the method for estimating parameters. Second, the test can compare nonnested models. Third, the test does not require data on produced goods and services, q, although, of course, these data are ultimately needed to compute TFP.

By setting input-share parameters period-by-period to relative input costs, a Solow-residual analysis treats observed inputs as optimal, so that input residuals are exactly zero, degrees of freedom of estimated parameters are exhausted, and, strictly, the estimated parameters and implied TFP have no statistical reliability. By contrast, by testing with an IC based on input residuals, we can select the empirically-most-valid model among CD, CES, TCES, and possibly other models, compute the best model's implied TFP, and compare it with benchmark Solow residuals. Along the way, we can check the MSP method's accuracy by comparing analytical and MSP-numerical solutions in the CD, CES, and TCES cases which imply analytical solutions. We illustrate these ideas using annual data on capital, labor, energy, materials, and services (KLEMS) inputs in manufacturing industries, from 1949 to 2001, obtained from the Bureau of Labor Statistics. Thus, we provide a method for computing the "empirically-most-valid" TFP, potentially more valid than the Solow residual. In other words, we check the robustness of Solow residuals to deviations of the production function from the CD approximation which underlies the Solow residuals.

The remainder of the paper is organized as follows. Section 2 discusses using the MSP method to compute optimal inputs. Section 3 discusses the econometric design. Section 4 does three things: (1) it applies the MSP method to the KLEMS data to compute input residuals for the CD, CES, and TCES models being considered; (2) it selects a best model which minimizes the information criteria of Akaike (1973), Schwarz (1978), and Hurvich and Tsai (1989); and, (3) it computes TFP implied by the best model, computes the benchmark Solow residual, and compares the two TFP computations. Section 5 contains concluding remarks.

2. Using MSP to Compute Optimal Inputs.

We exploit the property for simplifying computations that maximizing a function in a constraint set results in a solution which is equivalent to the solution obtained by maximizing a monotonic transformation of the function in the same constraint set. In original units of measurement, the output maximization problem is: for given $F(\cdot)$, P, and V, maximize $F(\hat{V})$ with respect to $\hat{V}$, subject to $P^T\hat{V} = P^TV$, where $F(\cdot)$, P, V, and $\hat{V}$ denote antilogs of $f(\cdot)$ and the elements of p, v, and $\hat{v}$. Although the original-unit and log-unit formulations of the problem lead to slightly different first-order conditions, they have equivalent solutions, namely, $\hat{V} = \exp(\hat{v})$. As noted before, proceeding in log form has several advantages.

We want to compute optimal and residual inputs, for each period, in a sample of input prices and quantities, for CD, CES, and TCES production functions. Let $\{p_t, v_t\}_{t=1}^{T}$ denote a given sample of observed input prices and quantities. Then, for given $f(\cdot)$, p_t, and v_t in period t, the vector of optimal inputs, $\hat{v}_t$, which solves the output maximization problem, implies the vector of input residuals, $v_t - \hat{v}_t$.

Figure 1 illustrates MSP computation of $\hat{v}_t$ in terms of two inputs, in the movement from points A to B. Points A and B denote start and end points of an MSP computation. Straight lines AA and BB, through A and B, denote start and end input-cost lines. Curved lines f_A and f_B, tangent at A to AA and tangent at B to BB, denote start and end isoquants. Observed input prices and quantities are $p = (p_1, p_2)^T$ and $v = (v_1, v_2)^T$. Observed v is at A and BB denotes the "observed" cost line defined by observed p and v, $e(p)^Te(\hat{v}) = e(p)^Te(v)$. The objective is to compute $\hat{v}$, the optimal combination of inputs on BB. The implied negative input residual, $\hat{v} - v$, is depicted by the vector difference B - A.

The MSP method starts at A but generally works correctly only if the starting point is optimal. Generally, A is not optimal on the observed cost line BB, because isoquant f_A, which passes through A, is not tangent to BB at A. However, A is optimal on AA, because AA is constructed to be tangent to f_A at A. Accordingly, AA is defined by $e(\hat{p})^Te(\hat{v}) = e(\hat{p})^Te(v)$, where $\hat{p}$ satisfies the first-order conditions (2.1) and (2.2) below, for given $f(\cdot)$ and v. Thus, $\hat{p}$ and AA are "optimal" at A. The MSP method computes the change in optimal inputs as they move from A to B in response to the counterclockwise rotation of the input-cost line at the initial point A, as the price vector flattens from $\hat{p}$ in AA to p in BB.

As before, for given assumed $f(\cdot)$ and given observed p and v, the objective is to compute optimal $\hat{v}$. For these given quantities, the log-form output-maximization problem is: maximize $f(\hat{v})$ with respect to $\hat{v}$, subject to $e(p)^T e(\hat{v}) = e(p)^T e(v)$. The Lagrangian function of the problem is $\ell = f(\hat{v}) + \hat{\lambda}(e(p)^T(e(v) - e(p)^T e(\hat{v}))$, where $\hat{\lambda}$ denotes the Lagrange multiplier. We obtain the first-order conditions of the maximization problem by differentiating ℓ with respect to $\hat{v}$ and $\hat{\lambda}$ and setting the results to zero,

(2.1) $$\nabla f(\hat{v}) = \hat{\lambda} e(p + \hat{v})^T,$$

(2.2) $$e(p)^T e(\hat{v}) = e(p)^T e(v),$$

where $\nabla f(\hat{v}) = [\partial f(\hat{v})/\partial v_1, \ldots, \partial f(\hat{v})/\partial v_n]$ denotes the $1\times n$ gradient row vector of first-partial derivatives of $f(\hat{v})$. For given $f(\cdot)$, p and v, equations (2.1) and (2.2) can be solved for unique values of $\hat{v}$ and $\hat{\lambda}$, at least locally and numerically, if second-order conditions hold.

As discussed before, we start the MSP method at observed inputs and need to treat them as optimal. Because observed inputs, v, are generally not optimal at observed prices, p, we first need to compute the "optimal" price vector, $\hat{p}$, at which v is optimal. We do this by considering the first-order conditions (2.1) and (2.2) as $\nabla f(v) = \hat{\lambda} e(\hat{p}+v)^T$ and $e(\hat{p})^T e(v) = e(p)^T e(v)$, for given assumed $f(\cdot)$ and given observed p and v, and solving for $\hat{\lambda}$ and $\hat{p}$. Let $E(x) = \text{diag}(e(x))$ denote the $n\times n$ diagonal matrix with $n\times 1$ vector e(x) on the principal diagonal; because all original units of observed inputs are positive, E(v) has finite and nonzero diagonal elements and, hence, is nonsingular; $E(v)^{-1}e(v) = u$, where $u = (1, \ldots, 1)^T$ denotes the $n\times 1$ unit vector of ones; $e(\hat{p})^T e(v) = e(p)^T e(v)$ when computing $\hat{p}$, because the computed input-cost line defined by $\hat{p}$ and the observed input-cost line defined by p both pass through the observed inputs, v. The solution values of $\hat{\lambda}$ and $\hat{p}$ are

(2.3) $$\hat{\lambda} = \nabla f(v)u/e(p)^T e(v),$$

$$e(\hat{p}) = E(v)^{-1}\nabla f(v)^T/\hat{\lambda}.$$

At this point, having computed $\hat{p}$ according to equations (2.3), we now consider $\hat{p}$ as observed and given, and relabel it as p. Thus, we now consider as given the same $f(\cdot)$ and v as before and the computed $\hat{p}$ relabelled as p. For these given quantities, we now differentiate first-order conditions (2.1) and (2.2) with respect to $\hat{v}$, $\hat{\lambda}$, and p and write the result as

(2.4) $$F(x)d\hat{y} = G(x)dp,$$

or

$$\begin{bmatrix} \nabla^2 f(\hat{v}) - \hat{\lambda}E(p+\hat{v}) & -e(p+\hat{v}) \\ -e(p+\hat{v})^T & 0_{1x1} \end{bmatrix} \begin{bmatrix} d\hat{v} \\ d\hat{\lambda} \end{bmatrix} = \begin{bmatrix} \hat{\lambda}E(p+\hat{v}) \\ e(p+\hat{v})^T - e(p+v)^T \end{bmatrix} dp,$$

where $\nabla^2 f(\cdot)$ denotes the $n\times n$ Hessian matrix of second-partial derivatives of $f(\cdot)$, $F(x)$ is an $(n+1)\times(n+1)$ matrix function, $G(x)$ is an $(n+1)\times n$ matrix function, $x = (\hat{y}^T, p^T)^T$ contains all 2n+1 variables, $\hat{y} = (\hat{v}^T, \hat{\lambda})^T$ contains the n+1 "endogenous" variables to be determined, and p contains the n given or "exogenous" prices. Although all of x is computed recursively and hats emphasize computed values, for simplicity, we omit them from x because, unlike in v or y, we do not need to distinguish between hatted and unhatted x. If $f(\cdot)$ is differentiable k+1 times, then, F(x) is differentiable k-1 times; G(x) is always differentiable any number of times.

The elements of x are all known, because they are either observed or computed. For given x, equation (2.4) implies the unique value $d\hat{y} = H(x)dp$, where $H(x) = F(x)^{-1}G(x)$, if and only if $|F(x)| \neq 0$, where $|\cdot|$ denotes the determinant of a square matrix. This condition holds because the second-order conditions of the problem imply that

(2.5) $$(-1)^{n+1}|F(x)| > 0$$

(Mann, 1943). Thus, when x maximizes output and satisfies second-order condition (2.5), equation (2.4) has the unique solution

(2.6) $$d\hat{y} = H(x)dp,$$

where $H(x) = F(x)^{-1}G(x)$ is an $(n+1)\times n$ matrix function of x. Although equation (2.6) derives from the true y process, we write its left side as $d\hat{y}$ to emphasize that the true dy is approximated using this equation.

3. Econometric Design.

We now discuss the econometric design of the tests. First, we explain computation of an information criterion (IC). We have a sample of observations on input prices and quantities, $\{p_t,v_t\}_{t=1}^{T}$, for periods t = 1, ..., T. Input residuals in period t are observed input quantities minus computed optimal input quantities, denoted $\xi_t = v_t - \hat{v}_t$. Suppose the residuals are distributed normally, identically, independently, with zero means, and the covariance matrix Σ_ξ or $\xi_t \sim NIID(0,\Sigma_\xi)$. Let $LL(\theta)$ denote $-(2/T)\times$log-likelihood function, except for terms independent of parameters, where θ denotes the model's parameters. Then, $LL(\theta) = \ln|\hat{\Sigma}_\xi|$, where $\ln|\cdot|$ denotes the natural logarithm of a determinant and $\hat{\Sigma}_\xi = (1/T)\sum_{t=1}^{T}\xi_t\xi_t^T$, where the residuals, ξ_t, are evaluated at the particular value of θ. Finally, an IC = LL + $P(\theta)$, where $P(\cdot)$ is a penalty term which depends on the number of estimated parameters. For example, in the Akaike information criterion, $P(\theta) = (2/T)\cdot\#(\theta)$, where $\#(\theta)$ denotes the number of estimated parameters.

4. Application to KLEMS Data.

We now discuss the application to annual data for U.S. manufacturing from 1949 to 2001, from the Bureau of Labor Statistics (2002). The data are prices and quantities of capital (K), labor (L), energy (E), materials (M), and services (S) used by U.S. manufacturing firms to produce output. The raw data are indexes of input quantities (with 1996 values being 100), expenditures on inputs in billions of current dollars, and the value of output in billions of current dollars. Prices of inputs are computed as expenditures divided by input quantity indexes. As noted before, it makes no difference whether the prices are in current or constant dollars.

The Solow-residual is based on a first-order CD approximation of any differentiable production function. Here, a production function parameterized in a certain way is a model. We consider CES and TCES models of the five KLEMS inputs. The parameters are input-cost shares, denoted α_1, ..., α_5, and input substitution elasticities, denoted σ in the CD and CES models and σ_1 and σ_2, for $\sigma_1 > \sigma_2$, in the TCES models. For the 53 annual periods, we consider "constant" α's estimated as sample means, "IMA" α's equal to one-period ahead forecasts of estimated IMA(1,1) models of the cost shares, and "Törnqvist" α's set to .5 $\times$ period t's observed input-cost shares + .5 $\times$ period t-1's observed input-cost shares. We estimate the IMA parameters by applying maximum likelihood estimation (MLE) to the raw cost-share observations. In each case, because the cost shares must sum to one, we set the α's of the four largest LMKS-cost shares as noted and set the remaining E-cost shares residually, as one minus the sum of the other α's. For CES models, we consider σ = .1, .5, 1., 2., 10. Thus, we do a kind of MLE over a coarse grid of σ's, conditional on estimated α's. In TCES models, we consider two σ's over a similar grid, such that the "outer" one is always larger than the "inner" one. We do not consider joint estimation of parameters, such as MLE, because often this results in implausible α's. For example, until he introduces utilization rates (an extension which is beyond the scope of this paper), Tatom (1980) obtains MLE $\alpha_L > 1$ and $\alpha_K < 0$, which contradicts $0 \leq \alpha \leq 1$.

We evaluate estimated models in terms of information criteria (IC). We consider the basic Akaike IC or AIC, the bias corrected AIC or BCAIC (Hurvich and Tsai, 1989), and the Schwarz (1978) Bayesian IC or BIC. We are especially concerned about degrees of freedom (DF) of estimated parameters and, for a chosen IC, consider as "best" the model which minimizes that IC. We are concerned with DF because a model with zero DF implies that the model's estimated parameters and any derived quantities, such as TFP, have infinite variances and, hence, have no statistical reliability. To varying extents, the ICs considered here account for DF by adding penalty terms to -(2/T) $\times$ log-likelihood function. Among the ICs considered here, in tables 1 to 3, BCAIC most effectively accounts for DF, because it is the only IC which approaches $+\infty$ as DF approach zero from above. Thus, we set BCAIC = $+\infty$ when DF are exhausted. An IC is parsimonious if it selects as "best" the models with the fewest parameters. ICs in Table 1 are ordered in increasing parsimony as AIC, BCAIC, and BIC.

4.1. Results from CD and CES Models

We considered 15 CES models. The top panel of Table 1 reports ICs of the best CES production function for five KLEMS input residuals; constant, IMA, and Törnqvist cost-share processes; and five values of the input elasticity of substitution. Of course, when the elasticity of substitution is one, the CES function reduces to the CD function.

The DF in for the CES models are obtained as follows. Each model has five KLEMS inputs. Because the cost shares sum to one, there are four free cost shares in each of the 53 sample periods. Each model also has an elasticity parameter. Thus, constant-cost-share models have 4 + 1 = 5 estimated parameters, hence, have DF = max[53-5,0] = 48. Each IMA(1,1)-cost-share model has two estimated parameters, a moving-average coefficient and a white-noise disturbance variance. Thus, IMA-cost-share models have $4 \times 2 + 1 = 9$ estimated parameters, hence, have DF = max[53-9,0] = 44. Finally, Törnqvist-cost-share models have $53 \times 4 + 1 = 213$ estimated parameters, hence, have DF = max[53-213,0] = 0. Figure 2 depicts the largest cost-share inputs, L, M, K, and S. That is, the smallest cost shares of E are not graphed. In figure 2, each panel contains time plots of constant, IMA, and Törnqvist cost shares for each of the LMKS inputs. Strictly each panel has three cases, but practically each panel has two cases, because the IMA and Törnqvist graphs are nearly identical. Thus, the IMA and Törnqvist models differ significantly only in their DF.

In the constant-cost share case, $\sigma = .5$ yields the lowest IC values with positive DF. Because the IMA- and Törnqvist-cost share graphs are nearly identical but IMA DF = 44, whereas Törnqvist DF = 0, we consider the IMA models as better, regardless of IC values. Thus, even if we chose to follow AIC and disregard the other ICs, we would consider IMA model 2 better than Törnqvist model 3, because model 2 has positive DF, even though model 3's AIC is lower.

4.2. Results from TCES Models

We also considered 24 TCES models. Even if we limit the TCES model search to two-tiered models, this still implies more models than we could evaluate in practice, because there are 16 possible groupings with one to five KLEMS inputs. Thus, we look at figures 3 and 4 to obtain guidance about which input groups to form.

Figure 3 depicts the 10 pairwise scatter plots of the KLEMS inputs in log form. In the figure, all pairwise plots except those involving L follow clear, noiseless, mostly upward, straight or curved lines. Plots involving L are quite noisy. Thus, figure 3 suggests that all non-L inputs move in close to fixed proportions and have low substitutability. That is, figure 3 suggests a two-tiered TCES model with an outer group of L and KEMS, with high substitution σ_1, and an inner group of K, E, M, and S, with low substitution σ_2. The L-KEMS two-tiered CES model, $TCES_1$, takes the form

$$(4.1) \qquad Q = [\alpha_1 L^\rho + \alpha_2(\beta_1 K^\gamma + \beta_2 E^\gamma + \beta_3 M^\gamma + \beta_4 S^\gamma)^{\rho/\gamma}]^{1/\rho},$$

where α_i, $\beta_i > 0$, $\alpha_1 + \alpha_2 = \beta_1 + ... + \beta_4 = 1$, and ρ, $\gamma < 1$. The outer group, L and KEMS, has $\sigma_1 = |1-\rho|^{-1}$; and the inner group, K, E, M, and S, has $\sigma_2 = |1-\gamma|^{-1}$.

Figure 4 suggests a two-tiered TCES model with so-called L-E-KMS input groups. The top panel of figure 4 depicts the following broad input-price movements: all input prices except E prices follow the same upward trend, exhibit relatively minor differences about the trend, and E prices are relatively constant during 1949-1972 and 1982-2001 and rise sharply during 1973-1981. The bottom panel of figure 4 depicts the following broad input-quantity movements: L is relatively constant, K, M, and S follow each other very closely along an upward trend, and E rises sharply until 1973 and thereafter grows slowly. In particular, the bottom panel of figure 4 suggests a two-tiered TCES model: an "outer" group of L, E, and KMS, with high substitution σ_1, and an "inner" group of K, M, and S, with low substitution σ_2. Because the bottom panel of figure 4 shows that K, M, and S move in close to fixed proportions, we expect σ_2 to be small. The relative constancy of L in figure 4 could also be interpreted as indicating nonneutral L-saving technical change, but we limit the analysis to homothetic production functions, hence, limit it to the neutral technical change of the Solow residual. The L-E-KMS two-tiered CES model, $TCES_2$, takes the form

$$(4.2) \qquad Q = [\alpha_1 L^\rho + \alpha_2 E^\rho + \alpha_3(\beta_1 K^\gamma + \beta_2 M^\gamma + \beta_3 S^\gamma)^{\rho/\gamma}]^{1/\rho},$$

where α_i, $\beta_i > 0$, $\alpha_1 + \alpha_2 + \alpha_3 = \beta_1 + \beta_2 + \beta_3 = 1$, and ρ, $\gamma < 1$. The outer group, L, E, and KMS, has $\sigma_1 = |1-\rho|^{-1}$; the inner group, K, M, and S has $\sigma_2 = |1-\gamma|^{-1}$.

The middle and bottom panels of Table 1 contain ICs from the $TCES_1$ and $TCES_2$ models. Because there are outer and inner elasticities of substitution in the TCES models, DF is equal to 47 in the constant-cost share models, is 43 in the IMA-cost share models, and remains 0 for the Törnqvist-cost share models. In the TCES models, the outer elasticity of substitution is $\sigma_1 \in \{.5, 1\}$ and the inner elasticity of substitution is $\sigma_2 \in \{.1, .17, .5, .67\}$. In the $TCES_1$ models, the IMA-cost-share model 5, with $\sigma_1 = 1$ and $\sigma_2 = .67$, has the lowest ICs and positive DF. Similarly, in the $TCES_2$ models, the best IMA-cost-share model 8, with $\sigma_1 = 1$ and $\sigma_2 = .67$, has the lowest ICs and positive DF.

Among the 15 CES and 24 TCES models, the IMA-cost-share model 5 has the lowest ICs, and therefore, is the best model. The results reject a single elasticity of substitution for all the KLEMS inputs and suggest that TFP computed from the IMA-cost-share L-KEMS TCES model 5 is more appropriate than TFP computed from a Törnqvist-cost-share CD model.

The MSP method worked accurately for all models and sample periods. The accuracy of MSP computations is measured as the largest absolute residual of the computed first-order conditions (FOC). For the KLEMS inputs, there are six scalar FOC, five marginal productivity conditions and a cost line. Each scalar FOC may be written as a scalar expression equal to zero. The computed value of each scalar expression is an FOC residual, which we want to be as close to zero as possible. For each sample period, the MSP method computes the input residuals in many steps. For each step, the method computes six absolute FOC residuals. For each case, the optimal inputs, hence, the residual inputs, were computed so that the FOC were satisfied with approximately double-precision or 10^{-14} accuracy. Because economic data usually have no more than 5-6 decimal digits, 14 decimal digit accuracy significantly exceeds the accuracy of usual economic data.

4.3. Best TFP versus Solow-Residual TFP.

We use the best production-function model, IMA-cost-share $TCES_1$ model 5, to compute $\%\Delta TFP_t^* = \%\Delta Q_t - \%\Delta F(\hat{v}_t)$, namely, period-to-period percentage change in optimal TFP, where $\%\Delta Q_t$ denotes period-to-period percentage change in observed output, $\%\Delta F(\hat{v}_t)$ denotes period-to-period percentage change in computed optimal output, for the best production function, $F(\cdot)$, at optimal inputs, $\hat{v}_t$. Similarly, let $\%\Delta TFP_t^{SR} = \%\Delta Q_t - c_{kt}\cdot\%\Delta K_t - \ldots - c_{St}\cdot\%\Delta S_t$, denote percentage change in Solow-residual TFP, where c_{kt}, ..., c_{st} denote Törnqvist input-cost shares and $\%\Delta K_t$, ..., $\%\Delta S_t$ denote percentage changes in the KLEMS inputs. To compare the percentage changes in optimal and Solow-residual TFP, we graph their difference in figure 5.

For the period 1949 to 2001, figure 5 and table 2 show a slightly positive mean, a slightly upward trend if one abstracts from outlying fluctuations, and a significantly declining variance in the difference between percentage growth in optimal and Solow-residual TFP or $\%\Delta TFP_t^* - \%\Delta TFP_t^{SR}$. In particular, table 2 shows a mean of $\mu = .10\%$ and a standard deviation of $\sigma = .60\%$. Although these numbers might seem small, they become much more significant when translated to levels at the end of the 53-year period, as follows.

For example, suppose the levels of the two TFP measures are both one in 1949. A measure which starts at one in 1949 and grows at $\gamma\%$ per year for 53 years equals $e^{53\gamma}$ in 2001. Thus, the one-standard-deviation bounds $\mu - \sigma = -.005 \le \%\Delta TFP_t^* - \%\Delta TFP_t^{SR} \le \mu + \sigma = .007$ on the differences in the growth rates from 1949 to 2001 imply the one-standard-deviation bounds $.767 \le TFP_{2001}^* - TFP_{2001}^{SR} \le 1.449$ on the differences in the levels of TFP in 2001. Thus, if optimal and Solow-residual TFP are both one in 1949, optimal TFP could be 45% higher or 23% lower than Solow-residual TFP by 2001. In other words, apparently small average differences and uncertainties in growth rates over 53 years translate into large differences and uncertainties in levels at the end of 53 years. We mention this example illustratively. Further investigation is needed to determine more conclusively whether there is a significant discrepancy between optimal and Solow-residual TFP.

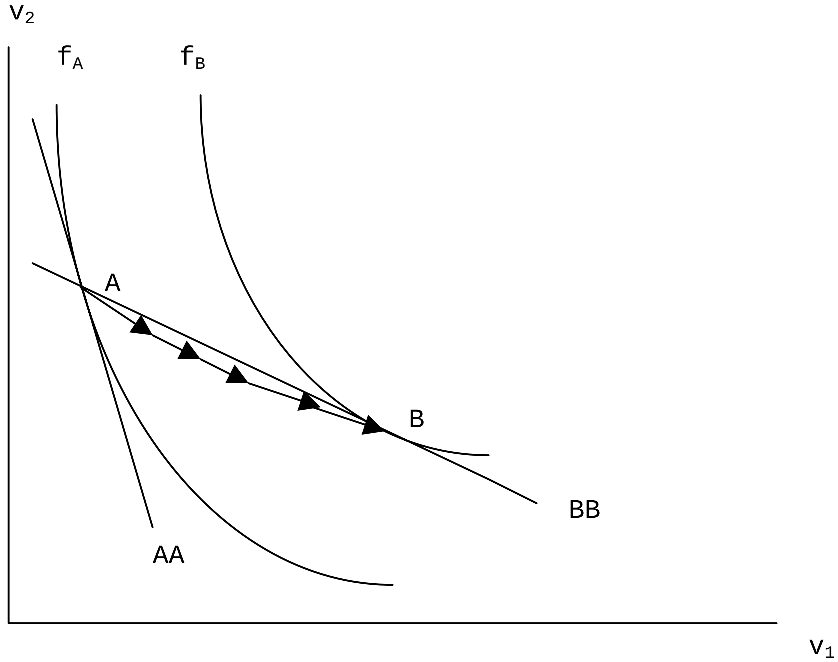
5. Conclusion.

We have used the multi-step perturbation (MSP) method to compute input residuals for CD, CES, and TCES production functions of KLEMS inputs, for inelastic, unit elastic, and elastic input substitution, using KLEMS data from the Bureau of Labor Statistics, representing aggregate U.S. manufacturing from 1949 to 2001. We then used the input residuals to compute various ICs. We focus on ICs because, like log-likelihood functions, they provide a scalar measure of the empirical fit of a multiple equation model, in this case the five, numerically computed, demand functions of the KLEMS inputs. By extending $-(2/T) \times$ log-likelihood function with positive penalty terms, the ICs acknowledge that adding parameters is statistically costly because degrees of freedom (DF) are used up. Adding too many parameters reduces DF to zero so that estimated parameters and derived quantities, such as TFP, based on them, strictly have infinite variances, hence, have no statistical reliability.

For the results in tables 1, $TCES_1$ model 5, with IMA-cost shares, is the best model with the lowest ICs and positive DF. According to AIC, CD model 3 has a slightly lower AIC and, in this respect, is better but we dismiss Törnqvist models with DF = 0 as statistically unreliable. Figure 2 indicates that the IMA-cost shares and the Törnqvist-cost shares follow each other very closely. Thus, the Solow-residual TFP implied by model 3 should be close to TFP of model 5 computed as $\tau_t = q_t - f(\hat{v}_t)$.

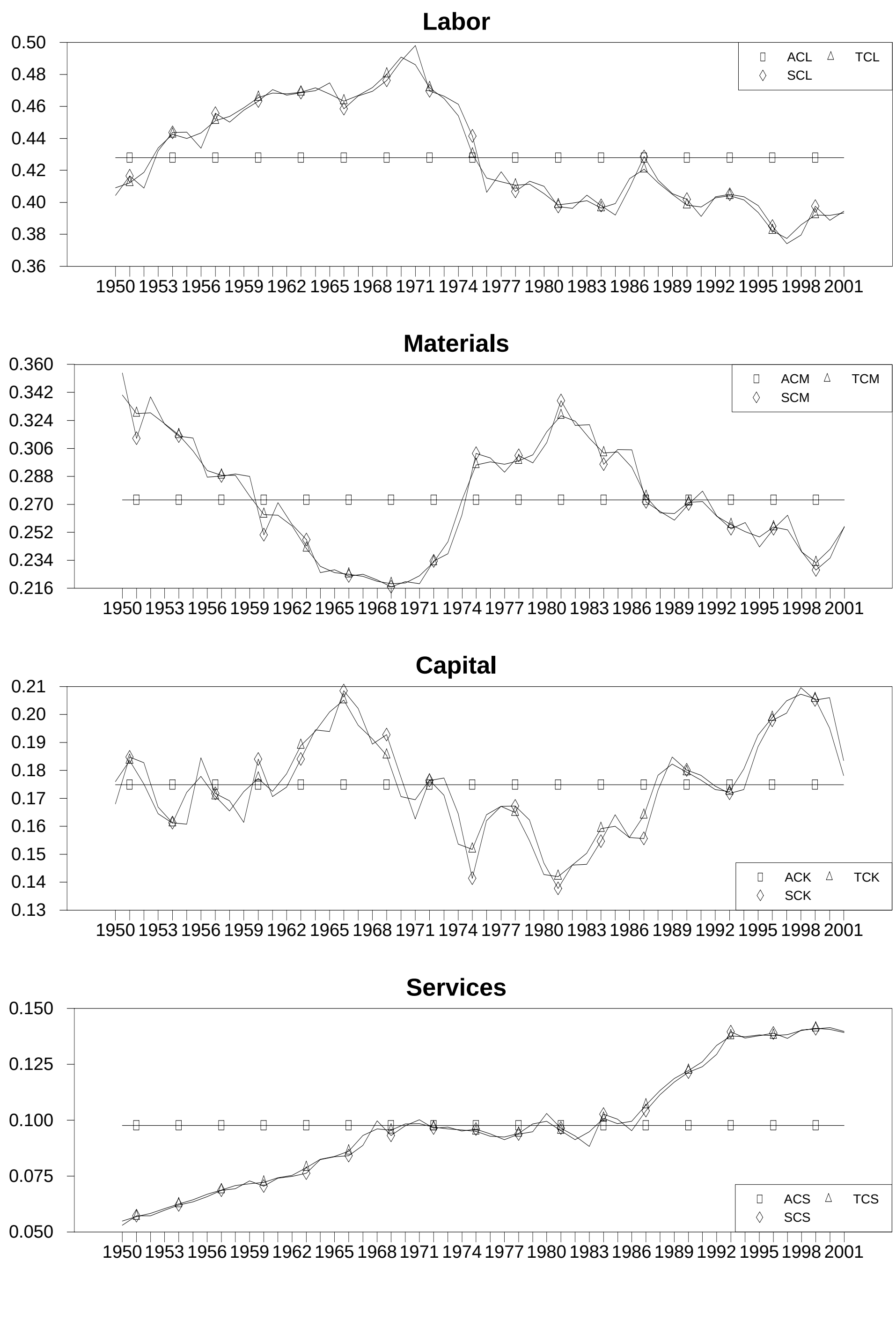
We have chosen a best $TCES_1$ model, within the CES and TCES classes of models, for computing optimal TFP. For a given IC and a given model, conditional on estimates of the cost-share parameters, α_i, we chose the elasticity parameter, σ, over a coarse grid of values, so as to minimize the IC. In the future, we shall consider estimating σ using maximum likelihood, conditional on estimates of the α_i's. Unless the production function includes a measure of capacity, estimating σ jointly with the α_i's may result in implausible estimates (Tatom, 1980). Also, we shall consider using more general production functions, such as the log-form GCES function, $f(v) = (1/\gamma)\cdot\ln(\sum_{i=1}^{n}\alpha_i e^{\rho_i v_i})$, where $\rho_i < 1$ is an input-specific elasticity parameter and setting $\gamma = \sum_{i=1}^{n}\alpha_i\rho_i$ implies local constant returns to scale. If the ρ_i's are unequal, then, the GCES function is globally nonhomothetic and first-order conditions (2.1) and (2.2) have no analytical solution. Because, for the CES and TCES applications in this paper, the MSP method produced very accurate solutions, with almost double precision ($\cong 10^{-14}$) accuracy, we expect the method to produce similarly accurate solutions for GCES applications.

Figure 1: Illustration of Multi-Step Perturbation.



Input-cost lines AA and BB are, respectively, defined by $e(\hat{p})^T e(\hat{v}) = e(\hat{p})^T e(v)$ and $e(p)^T e(\hat{v}) = e(p)^T e(v)$, for given precomputed "optimal" $\hat{p}$ and given observed p and v.

Figure 2: Constant, IMA, and Törnqvist LMKS Input Cost Shares, 1949-2001.



In ACL, SCL, and TCL, "A" denotes average or constant cost shares, "S" denotes "smooth" or IMA cost shares, "T" denotes Törnqvist cost shares, "L" denotes labor, and similarly in ACM, ..., TCS.

Figure 3: Scatter Plots of Pairwise Log of KLEMS Input Quantities.

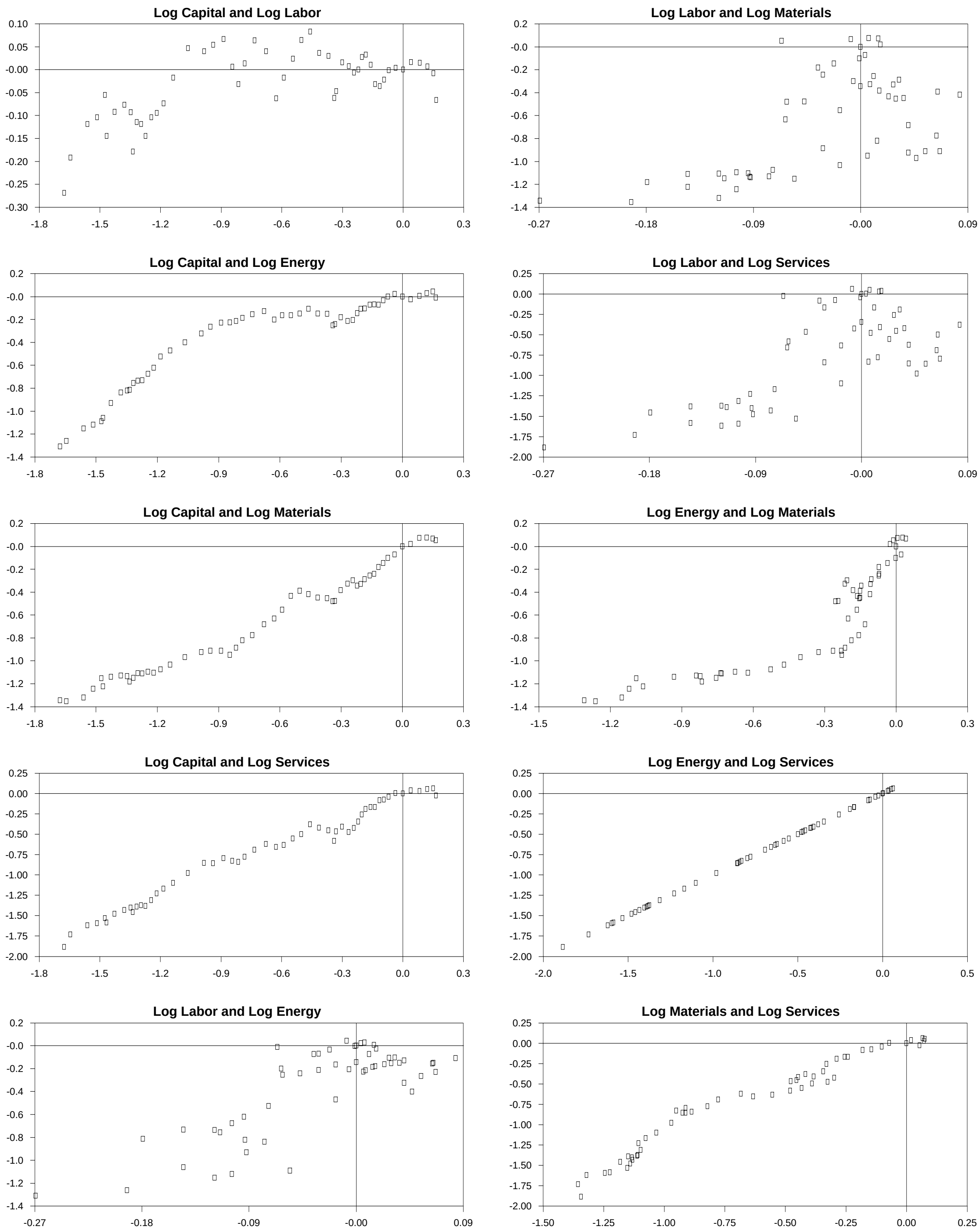
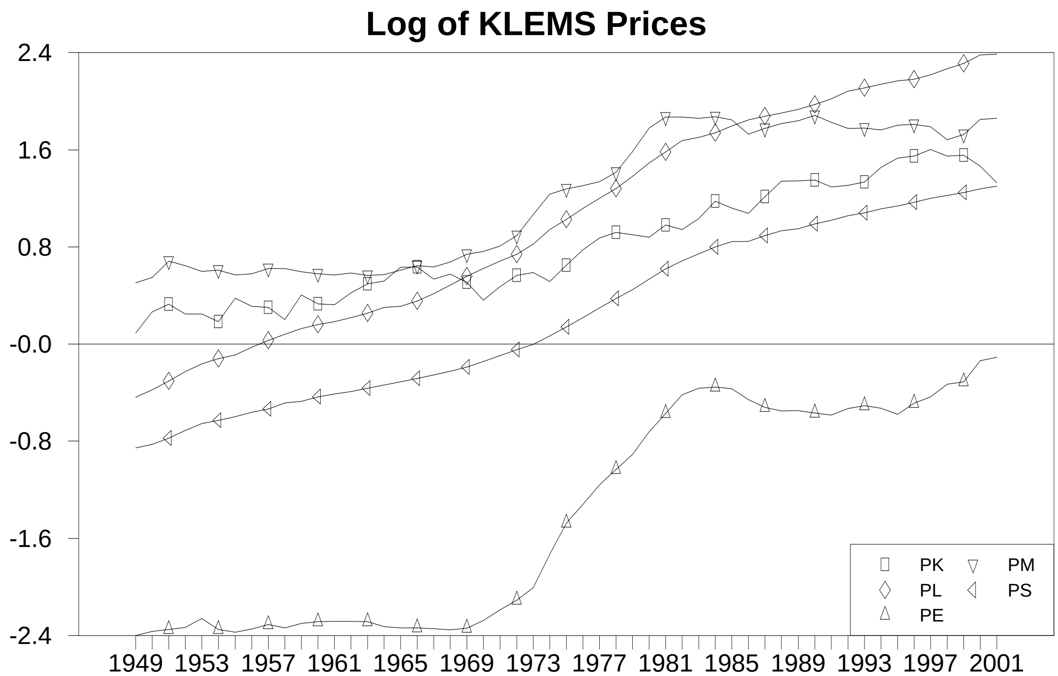


Figure 4: Log of KLEMS Input Prices and Quantities, 1949-2001.



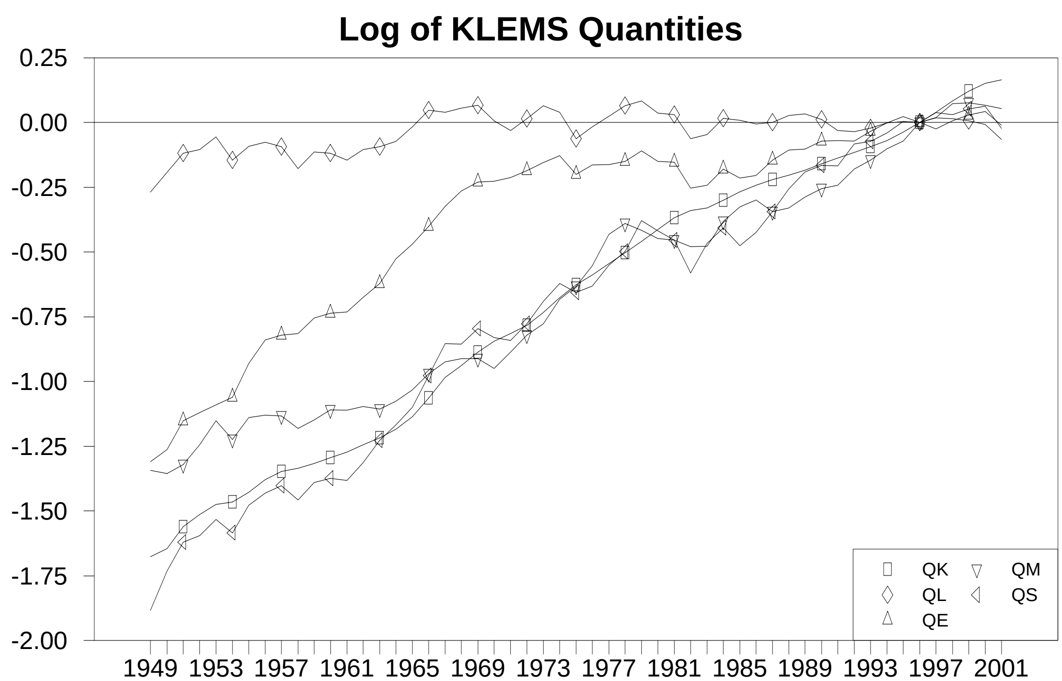


Figure 5: $\%\Delta TFP_t^* - \%\Delta TFP_t^{SR}$ from 1949 to 2001.

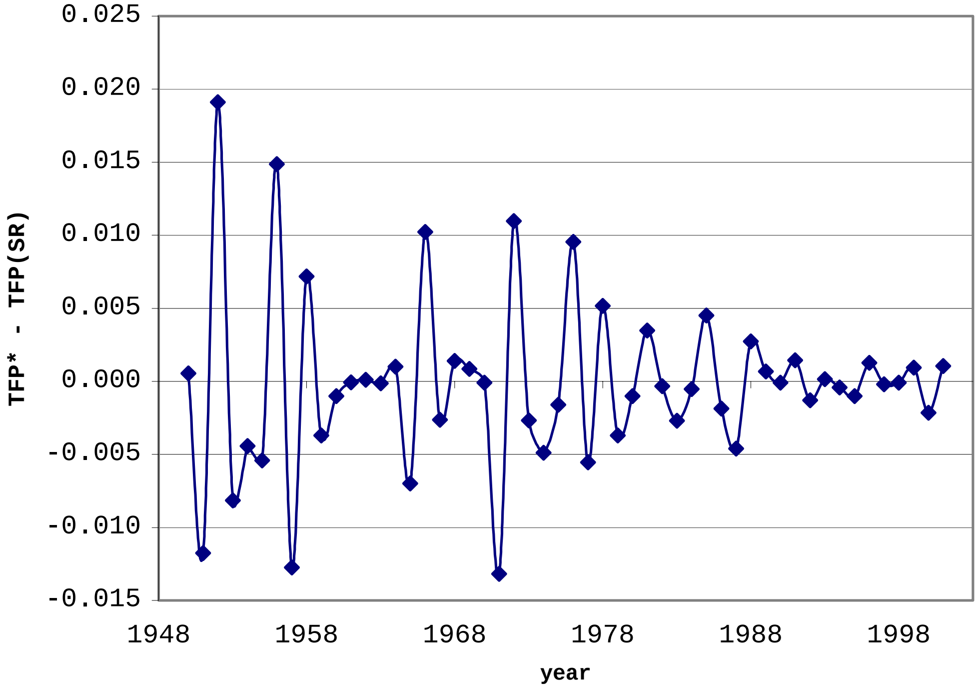


Table 1: Summary Statistics of the Best Estimated Models.

1	2	3	4	5	6	7	8	9
Model	α_{it}	σ_1	σ_2	**DF**	**-2LL/T**	**AIC**	**BCAIC**	**BIC**
				Best CES Models				
1	**Const**	.50	---	48	-26.66	-26.47	-26.45	-26.29
2	**IMA**	1.0	---	44	-32.62	-32.28	-32.21	-31.94
3	**Tornq**	1.0	---	0	-42.22	-34.18	$+\infty$	-26.26
				Best $TCES_1$ Models				
4	**Const**	.50	.17	47	-27.69	-27.50	-27.48	-27.32
5	**IMA**	1.0	.67	43	-33.54	-33.21	-33.13	-32.87
6	**Tornq**	1.0	.67	0	-37.26	-26.98	$+\infty$	-19.06
				Best $TCES_2$ Models				
7	**Const**	.50	.10	47	-24.09	-23.90	-23.88	-23.71
8	**IMA**	1.0	.67	43	-33.24	-32.90	-32.82	-32.56
9	**Tornq**	1.0	.67	0	-40.35	-32.32	$+\infty$	-24.40

Comment: The CES, $TCES_1$, and $TCES_2$ production functions are, respectively, $Q = (\alpha_1 K^\rho + \alpha_2 L^\rho + \alpha_3 E^\rho + \alpha_4 M^\rho + \alpha_5 S^\rho)^{1/\rho}$, $Q = [\alpha_1 L^\rho + \alpha_2(\beta_1 K^\gamma + \beta_2 E^\gamma + \beta_3 M^\gamma + \beta_4 S^\gamma)^{\rho/\gamma}]^{1/\rho}$, and $Q = [\alpha_1 L^\rho + \alpha_2 E^\rho + \alpha_3(\beta_1 K^\gamma + \beta_2 M^\gamma + \beta_3 S^\gamma)^{\rho/\gamma}]^{1/\rho}$.

Table 2: Summary Statistics of $\%\Delta TFP^{*} - \%\Delta TFP^{SR}$.

$\%\Delta TFP^{*} - \%\Delta TFP^{SR}$	
Min	-1.91%
Max	1.32%
Mean	.10%
Std dev	.59%

REFERENCES

Akaike, H. (1973), "Information Theory and Extension of the Maximum Likelihood Principle," pp. 267-281 in Second International Symposium on Information Theory, B.N. Petrov and F. Csaki (eds.), Budapest: Akademia Kiado.

Berndt, E.R. (1991), The Practice of Econometrics: Classic and Contemporary, Reading, MA: Addison-Wesley.

Bureau of Labor Statistics (2002), Multifactor Productivity web page, http://www.bls.gov/mfp/home.htm.

Burnside, C., M. Eichenbaum, and S. Rebelo (1995), "Capital Utilization and Returns to Scale," Working Paper No. 5125, National Bureau of Economic Research, Cambridge, MA.

Christensen, L.R., D.W. Jorgenson, and L.J. Lau (1973), "Transcendental Logarithmic Production Functions," Review of Economics and Statistics 55: 28-45.

Gardner, E.S. (1985), "Exponential Smoothing: The State of the Art," Journal of Forecasting 4: 1-28.

Hurvich, C.M. and C.L. Tsai (1989), "Regression and Time Series Model Selection in Small Samples," Biometrika 76: 297-307.

Sato, K. (1967), "A Two-Level Constant Elasticity-of-Substitution Production Function," Review of Economic Studies 34: 201-218.

Schwarz, G. (1978), "Estimating the Dimension of a Model," Annals of Statistics 8: 147-164.

Solow, R.M. (1957), "Technical Change and the Aggregate Production Function," Review of Economics and Statistics 39: 312-320.

Tatom, J.A. (1980), "The 'Problem' of Procyclical Productivity," Journal of Political Economy 88: 385-394.

Zadrozny, P.A. and B. Chen (2004), "Multi-Step Perturbation Method for Accurately Computing and Empirically Evaluating the Cost of Living," typescript, Bureau of Labor Statistics, Washington, DC.