Why do Aggregate Production Functions Work? Fisher's simulations, Shaikh's identity and some new results

JESUS FELIPE & CARSTEN A. HOLZ

ABSTRACT *The literature on aggregation has shown that the conditions for successful aggregation of micro production functions into an aggregate production function are far too stringent to be believable (Fisher 1969, 1971). Despite this, aggregate production functions continue being used. The reason is that they seem to 'work'. This happens, however, because underlying every aggregate production function is the income accounting identity that links input and output, i.e. output equals wages plus profits. A simple algebraic transformation of this identity yields a form that resembles a production function (Shaikh, 1974, 1980). This paper uses Monte Carlo simulations to study two questions. First, how much spuriousness can help explain the relatively good fits of the Cobb–Douglas production function? The simulations show that the contribution of spuriousness to a high R^2 is minor once we properly account for the fact that input and output data used in production function estimations are linked through the income accounting identity. It is mostly the link through this identity that explains the results. Secondly, we study how much factor shares have to vary in an economy so as to render the Cobb–Douglas production function with a time trend a bad choice for modelling and estimation purposes. We conclude that the Cobb–Douglas form is robust to relatively large variations in the factor shares. What makes this form often fail are the variations in the growth rates of the wage and profit rates.*

1. Introduction

Over 30 years ago, Franklin Fisher (1969, 1993) showed, in a series of seminal papers, that aggregate production functions *should not* work in empirical applications. The reason is that the conditions for successful aggregation are so stringent that one cannot expect these conditions to hold for real economies. The most general conditions state that the existence of an aggregate capital stock requires that the production functions of individual firms differ at most by capital-augmenting technical differences; the existence of a labour aggregate requires that every firm hire the same proportions of each type of labour; finally, the existence of an aggregate of output requires that every firm produce the same market basket of outputs.

Jesus Felipe, Georgia Institute of Technology, Atlanta, GA 30332–0610, USA. E-mail: jesus.felipe@inta.gatech.edu. Carsten A. Holz, Social Science Division, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.

ISSN 0269-2171 print/ISSN 1465-3486 online/01/030261-25
DOI: 10.1080/02692170110052338

Despite Fisher's results, economists have continued using the aggregate production function in both theoretical and applied works. It is difficult to know why. Our conjecture is that *sometimes* they *appear* to work when estimated econometrically (tested?). By this we mean, following Lucas (1970) and Fisher (1971), that the fit obtained is high, the elasticities are close to the factor shares in output, and the marginal productivity of labour explains the wage rate well. This observation led Fisher (1971) to study, using simulations, the empirical conditions under which the aggregate Cobb–Douglas function would yield good results. In his pioneering simulation work, Fisher concluded that the Cobb–Douglas production function would work well whenever factor shares were constant. This is not a trivial observation, for Fisher had generated aggregate series violating the aggregation conditions. Thus, he was surprised to see that the aggregate Cobb–Douglas function works under the circumstances explained above.[1] The reality, however, is that when one estimates aggregate production functions, results tend to be rather poor, in particular in time series analyses. Very seldom are results consistent with the neoclassical theory. This was recently pointed out by Sylos Labini (1995).

Shaikh (1974, 1980) explained why what Fisher had discovered was due to something rather surprising and discomforting. Shaikh showed that underlying Fisher's (1971) simulations was the accounting identity that links output to the sum of the wage bill plus overall profits. It can be shown that a simple algebraic transformation of this identity yields a form that *resembles* a production function. Shaikh (1980) showed that when Fisher generated the series of output and inputs for the simulations, he had linked them through the income identity. No wonder the aggregate production function has to work econometrically! More recently, McCombie (1987), McCombie & Dixon (1991), Felipe (2000), and Felipe & McCombie (2001a) have taken up the issue again and provided empirical evidence that all aggregate production functions do is to track the income accounting identity.[2]

This paper revisits the problem. The reason is that today economists still define the aggregate production function as a summary of the aggregate technology (Mankiw, 1997); estimate econometrically aggregate production functions (believing that the good fits reinforce their findings);[3] and derive implications from such results. Examples of this practice are the literature on growth, externalities and increasing returns (Romer, 1987; Hall, 1990; Caballero & Lyons, 1992; Mankiw *et al.*, 1992; Basu & Fernald, 1995, 1997);[4] or the literature on the measurement on infrastructure productivity (Aschauer, 1989; Munnell, 1992).

We use Monte Carlo simulations and a Cobb–Douglas production function to answer two questions. First, how much does spuriousness help explain the high fit of the production function? This matters because economists today tend to estimate the production function in first differences in order to eliminate the possible spuriousness inherent in the estimation in levels, assuming output and input are difference stationary processes (or an error correction transformation if they find the series to be cointegrated). However if, as indicated above, the aggregate production function is essentially a transformation of an identity, the problem of spuriousness should not matter, and it should not make any difference whether the equation is estimated in levels or in first differences.

The second question is: how much do factor shares have to vary so that the Cobb–Douglas production function yields poor results? Fisher (1971) concluded quite correctly that the Cobb–Douglas function works well when factor shares are constant. In addition, by implication, it does not work when factor shares are not constant. But how much variation do we need to observe in the factor shares? An interesting aspect, well known by practitioners, is that if one fits a Cobb–Douglas production function with a time trend to a data set, most likely the results will be 'bad' despite the fact that factor shares are (sufficiently) constant. This has often led economists to open Pandora's box and search for all sorts of reasons. This paper offers a parsimonious explanation for these findings.

In doing this, the present paper generalizes Fisher's experiments in two obvious ways. First, there is an explicit use of the accounting identity in the generation of the series, thus acknowledging Shaikh's explanation. And second, we take care of the statistical properties of the data (i.e. the possibility of unit roots in some of the series), something not well understood in the early 1970s. The rest of the paper is structured as follows. In Section 2 we summarize Shaikh's arguments. It will be shown that any production function can be derived as a transformation of the said income accounting identity. Sections 3 and 4 discuss the above two questions and the simulation results. To keep matters simple we use the standard Cobb–Douglas production function with a time trend. Once the arguments in Section 2 are understood, it will be clear that the choice of this form does not affect the substance of the argument. The same could be done with the CES or translog production functions, but at the cost of increasing the complexity of the simulations. Section 5 provides empirical evidence. Section 6 concludes.

2. Production Functions and the Accounting Identity

To begin with, and to motivate the questions posed, we provide suggestive empirical evidence. Table 1 shows the OLS estimates of a Cobb–Douglas production function for two US manufacturing branches, in growth rates (equations A), and in levels (equations B). The poor results displayed here are 'standard'; in particular, the finding of an insignificant or a negative elasticity of capital (Klette & Griliches, 1996; Griliches & Mairesse, 1998). It is also worth mentioning the substantial differences in R^2 and Durbin–Watson (DW) statistics between the regressions in growth rates and in levels; in the latter regressions R^2 is higher, and the DW is lower. The most likely argument is that this is due to the effect of spuriousness. As is well known, in regressions of integrated variables (as economic variables in levels often are) the R^2 statistic tends to one in probability, while the DW tends to zero also in probability. Likewise, the estimated parameters are very different. This paper asks whether there is a parsimonious interpretation for these results, and whether what is at stake is an econometric problem, or something different. We shall return to these results in Section 5.

To understand the claims made in the introduction, let us show how any production function can be derived as an algebraic transformation of the income accounting identity. This identity can be written as

$$Q_t \equiv w_t L_t + r_t K_t \tag{1}$$

where Q, w, r, L and K denote real value added, the average real wage rate, the average profit rate (% per annum), employment, and value of the stock of capital, respectively. Expression (1) holds as an identity for every period. It is not Euler's Theorem. It states that income equals the sum of the wage bill plus all types of

Table 1. Cobb–Douglas production function

SIC 22: Textile and mill products

(A) $q_t = \varphi = \gamma_3 l_t + \gamma_4 k_t$

	φ	γ_1	γ_2
	0.048 (5.71)	1.45 (7.37)	−0.617 (−2.01)
		R^2 = 0.653; DW = 1.89	

(B) $\ln Q_t = C + \varphi t + \gamma_1 \ln L_t + \gamma_2 \ln K_t$

Constant	φ	γ_1	γ_2
3.22 (5.43)	0.05 (12.79)	1.80 (10.45)	−0.62 (−4.75)
		R^2 = 0.976; DW = 1.24	

SIC 36: Electric and electronic equipment

(A) $q_t = \varphi + \gamma_3 l_t + \gamma_4 k_t$

	φ	γ_1	γ_2
	0.075 (4.46)	1.165 (11.00)	−0.825 (−2.57)
		R^2 = 0.806; DW = 1.845	

(B) $\ln Q_t = C + \varphi t + \gamma_1 \ln L_t + \gamma_2 \ln K_t$

Constant	φ	γ_1	γ_2
3.20 (1.72)	0.03 (2.93)	0.91 (10.21)	0.05 (0.26)
		R^2 = 0.992; DW = 0.75	

Data for the U.S. for 1959–91.

profits. It does not assume constant returns to scale, or perfect competition. It holds in every type of market. The operating surplus (i.e. all profits on all types of capital goods) is written as the ex-post average profit rate (not the rental price of capital) times the value of the stock of capital. In the words of Samuelson: 'No one can stop us from labeling this last vector [residually computed profit returns to "property" or to the non-labour factor] as (RC_j), as J.B. Clark's model would permit – even though we have no warrant for believing that noncompetitive industries have a common profit rate R and use leets capital (C_j) in proportion to the $(P_j q_j - W_j L_j)$ elements!' (Samuelson, 1979, p. 932).[5]

Now let us rewrite Equation (1) in growth rates (lowercase letters denote the growth rates of output, labour and capital, and ^ denotes the growth rates of the wage and profit rates) as

$$q_t = a_t\hat{w}_t + (1 - a_t)\hat{r}_t + a_t l_t + (1 - a_t)k_t = \varphi_t + a_t l_t + (1 - a_t)k_t \quad (2)$$

where $a_t = w_t L_t / Q_t$ and $1 - a_t = r_t K_t / Q_t$ are the labour and capital shares, respectively, in total output, and $a_t\hat{w}_t + (1 - a_t)\hat{r}_t = \varphi_t$ is the weighted average of the growth rates of the wage and profit rates (his expression will play an important part in the arguments).

Assume, without affecting the substance of the argument, that factor shares are constant over time, i.e. $a_t = a$, and that wages and profit rates grow at a constant rate, i.e. $\hat{r}_t = \hat{r}$ and $\hat{w}_t = \hat{w}$, where $r_t = r_0 e^{\hat{r}t}$ and $w_t = w_0 e^{\hat{w}t}$ (departures from these assumptions will be discussed below). Under these assumptions, equation (2) becomes

$$\hat{Q}_t = a\,\hat{w} + (1 - a)\,\hat{r} + a\,\hat{l}_t + (1 - a)\,\hat{k}_t = \varphi + a\,\hat{l}_t + (1 - a)\,\hat{k}_t \quad (3)$$

where $\varphi = a\hat{w} + (1 - a)\hat{r}$ (4)

Integrating Equation (3) leads to

$$\ln(Q_t) = \varphi t + a\ln(L_t) + (1 - a)\ln(K_t) + C \quad (5)$$

where C is the constant of integration. And taking antilogarithms yields

$$Q_t = Ae^{(a\hat{w} + (1-a)\hat{r})t} L^a_t K^{1-a}_t = Ae^{\varphi t} L^a_t K^{1-a}_t \quad (6)$$

where A is also a constant. What is Equation (6)? Given our derivation, it must be the accounting identity, Equation (1), rewritten under the assumptions of constant factor shares and constant growth rates of the wage and profit rates. The interesting aspect is, of course, that it is identical to the Cobb–Douglas production function with constant returns to scale and a 'neutral' time effect, the coefficient of which has been traditionally interpreted as the rate of total factor productivity (TFP) growth. The above derivation shows that if the assumptions of constant factor shares and constant growth rates of the wage and profit rates *are correct*, the econometric estimation of the rewritten accounting identity in logarithms, that is, of

$$\ln(Q_t) = a + \varphi t + \gamma_1 \ln(L_t) + \gamma_2 \ln(K_t) + u_t \quad (7)$$

where u_t is the error term, must yield a perfect fit (i.e. $R^2 = 1$ and $u_t = 0$) because it is an identity. The income identity is, of course, compatible with any aggregate production technology, or lack of it, but will give a perfect fit to a putative Cobb–Douglas production function as long as factor shares are constant. Factor shares can be constant for many reasons, such as a constant mark-up on unit labour costs or the Kaldorian theory of distribution (Kaldor, 1956), both of which do not depend on an underlying Cobb–Douglas production function. Should this be the case, the data will indicate that a Cobb–Douglas form is the correct one even though the true underlying technology is, for example, fixed coefficients.

By implication, if factor shares are not (sufficiently) constant, and/or wage and profit rates do not grow at constant rates, estimation of (7) will yield poor results.

This way, Shaikh (1980) was able to show that since Fisher (1971) had used the identity (1) to generate the series (and imposed constant factor shares), his simulations with the Cobb–Douglas function had to yield high fits, and the estimated parameters had to be very close to the factor shares, even though the capital series had been generated violating the aggregation conditions. Fisher's observation that the aggregate Cobb–Douglas function works when factor shares are constant despite violating the aggregation conditions was easily explained.[6]

A number of further implications are the following. (i) The coefficient of the time trend, φ (the proxy for the rate of technological progress), equals the weighted average of the constant growth rates of the wage and profit rates, i.e. Equation (4). This matters for explaining the simulation results, in particular those in Section 4. (ii) The estimates of γ_1 and γ_2 in Equation (7) must equal the constant shares of labour and capital, i.e. a and $(1-a)$. This implies that the values of the factor shares determine the values of the output elasticities in a statistical sense, rather than the other way around for economic reasons (e.g. factor shares are paid their marginal products under competitive conditions). (iii) γ_1 plus γ_2 must add up to 1, thus showing putative 'constant returns to scale'. However, this is due to the underlying accounting identity, and does not imply that, in actual production, returns to scale are constant. They might or might not be. (iv) Regressors' endogeneity as well as the statistical properties of the series (i.e. the underlying trend and the possibility of cointegration) are issues of secondary nature (if important at all) since we are estimating an accounting identity (this is further studied in Section 3).[7] (v) If, in practice, we do not obtain the identity it must be because *one or both* assumptions regarding factor shares and the growth rates of the wage and profit rates do not hold. If factor shares in the economy under consideration vary, or if the growth rates of the wage and profit rates are not constant (i.e. φ_t is not constant) then we need to hypothesize different paths (probably more complex), which will lead to other 'production functions' (see Felipe, 2000; Felipe & McCombie, 2001a, for the translog and CES cases). However, the general argument remains: all we are doing is rewriting the accounting identity. Suppose, for example, that in the economy under consideration wage and profit rates are constant, i.e. $\hat{w}_t = 0$ and $\hat{r}_t = 0$ (instead of growing at a constant rate, as assumed before). Substitution into Equation (2) implies $\varphi = 0$. Then the income accounting identity can be rewritten as a Cobb–Douglas function without the time trend. Or suppose factor shares, instead of being constant, follow a path $a_t = f(\ln L_t, \ln K_t)$ such as $a_t = \alpha_1 + \alpha_2 \ln K_t + \alpha_3 \ln L_t$ (and similarly for capital's share). Then, substitution of this path into Equation (2) and integrating will yield a translog production function. We must stress the context of these arguments: most likely an aggregate production function cannot be derived theoretically, and yet we fit one and it works. Why? The reason is that all aggregate production functions are different approximations (depending on the paths of the factor shares, wage and profit rates) to the income accounting identity.[8]

3. How Far can Spuriousness go to Explain the Fit of the Cobb–Douglas Production Function

This section studies the effects of the spurious regression phenomenon in the case of production function estimations. Nelson & Kang (1984) and Durlauf & Phillips (1988) discussed the econometric implications of including a linear time trend as one of the right-hand side variables in the regression of one difference stationary process (e.g. output) on one or more unrelated difference stationary processes (e.g., labour and capital). Using simulation analysis, Nelson & Kang (1984) showed that high R^2 values and significant coefficient t-values are due to spurious detrending. From the economic point of view, the trend is included in the production function as a measure of technological progress. From the econometric point of view, however, if output and inputs are difference stationary processes (DSP), the regression is spurious and should therefore be run in first differences. Only if all time series are trend stationary processes (TSP) is inclusion of a time trend econometrically acceptable.

Nelson & Kang made a reference to the case under study, namely, the production function. They ran the regression $Y_t = \alpha + \beta t + \upsilon X_t + u_t$ '. . . where $\{Y_t\}$ is a nonstationary variable such as output, $\{X_t\}$ is a nonstationary independent variable (or set of such variables) such as a production input, and $\{u_t\}$ is a sequence of disturbances. The role of time is to account for growth in Y not attributable to X, for example, the impact of technological change on output' (Nelson & Kang, 1984, p. 78). In order to obtain a lower bound for the R^2 value, Nelson & Kang created Y and X as independent zero-drift random walks with unit variance Normal distributions, and then regressed Y on a constant, time and X. Across 1000 runs they obtained a mean R^2 value of 0.501, the time coefficient $\hat{\beta}$ was significant at a nominal 5% level in 83% of the runs, and the coefficient of X, $\hat{\gamma}$, was significant at a nominal 5% (1%) level in 64% (55%) of the runs.

We extend Nelson & Kang's (1984) simulations to take into account that the output and inputs series used to estimate the production function $Q_t = A_t F(K_t, L_t)$ are linked through the accounting identity (1). We report the results of the simulation experiment which analyses how much of the R^2 value observed in the estimation of the Cobb–Douglas function with an exponential time trend can be explained by spuriousness. The results are shown in Table 2.

The first set of simulations simply extends Nelson & Kang's (1984) analysis for the case of three unrelated random walks.[9] (In this and the next set, results are shown for estimations in levels and first differences). In the second set the three random walks are related through the accounting-identity-link Equation (1) and the definition of the labour share, i.e. $a_t = w_t L_t / Q_t$. The theoretical derivation in Section 2 was based on certain paths for the variables (e.g. constant growth rates of the wage and profit rates). It would be meaningless, however, to run the simulations with those exact paths, since we know that we would be estimating an accounting identity, and thus R^2 must be 1, independently of the statistical properties of the data. Besides, if the data exactly satisfied the assumptions about the factor shares, wage and profit rates, and we tried to estimate a Cobb–Douglas production function with a time trend, we would have the problem of perfect multicollinearity (this can be easily checked by substituting the variables in the identity into the production function). Therefore, we explore what occurs when the data follow other paths (e.g. wage and profit rates are random walks). This is not incompatible with the analysis in the previous section, for the accounting identity continues holding. The third set extends the results to the case where all variables are TSP (results in levels). All statistics cover 1000 runs, where each run has 100 observations. It must be emphasized that the precise values of the variables in the simulations are of no consequence. All that matters is that the identity is imposed on the series (one of the five series must be used to 'close' the identity. We chose K). It is also important to emphasize that, unlike Fisher (1971), there is no explicit aggregation process.

Table 2. How far can spuriousness go to explain the fit of the Cobb–Douglas production function?

Statistic	Mean	Standard deviation	Number of runs (out of 1000, 5% signif. level): With sign. t-val.	Number of runs (out of 1000, 5% signif. level): Rejecting CRS	Mean	Standard deviation	Number of runs (out of 1000, 5% signif. level): With sign. t-val.	Number of runs (out of 1000, 5% signif. level): Rejecting CRS
	All variables are DSP created as $X_t = \mu + X_{t-1} + \epsilon_t$; $\epsilon_t \sim N(0,1)$							
	Spurious regression				Regression in first differences			
Sets 1–2	$\ln(Q_t) = \alpha + \beta_t + \gamma_1 \ln(L_t) + \gamma_2 \ln(K_t) + u_t$				$q_t = \beta + \gamma_1 l_t + \gamma_2 k_t + v_t$			
Set 1	Q, L and K are RW with $Q_0 = K_0 = L_0 = 1000$, $\mu = 0$.							
R^2	0.575357	0.242049		902	0.020858	0.020051		1000
α	6.882530	3.993195	901					
S.e.(α)	1.004910	0.425451			0.000002	0.000007		
β	−0.000008	0.000120	771		0.000002	0.000101		
S.e.(β)	0.000016	0.000008			0.000100	0.000487		
γ_1	0.014979	0.396637	610		0.000754	0.107413	64	
S.e.(γ_1)	0.103367	0.041911			0.101153	0.010787		
γ_2	−0.011447	0.415785	627		0.002655	0.099560	46	
S.e.(γ_2)	0.101244	0.039940			0.100753	0.010426		
DW	0.327604	0.141562			1.990242	0.203015		
Set 2	w, r and L are RW with $w_0 = r_0 = L_0 = 1000$, $\mu = 0$. $Q_t = w_t L_t / a_t$, and $K_t = (Q_t - w_t L_t)/r_t$, where $a_t = a = 0.6$.							
R^2	0.873451	0.115429		623	0.752246	0.043172		55
α	7.654263	1.964840	997					
S.e.(α)	0.498156	0.198476						
β	−0.000003	0.000085	819		0.000001	0.000070	47	
S.e.(β)	0.000012	0.000006			0.000071	0.000005		
γ_1	0.500462	0.350018	857		0.500805	0.089568	999	
S.e.(γ_1)	0.087817	0.035798			0.087681	0.008813		
γ_2	0.494506	0.202831	964		0.498748	0.053252	1000	
S.e.(γ_2)	0.051409	0.021679			0.050449	0.005016		
DW	0.322063	0.140261			1.999682	0.198818		

Table 2. (Continued)

Statistic	Mean	Standard deviation	Number of runs (out of 1000, 5% signif.t level) With sign. t-val.	Rejecting CRS	Mean	Standard deviation	Number of runs (out of 1000, 5% signif.t level) With sign. t-val.	Rejecting CRS
Set 3	All TSP variables are created as $X_t = 1000\exp(0.05t + z_t \sim N(0,1)$ Regression: $\ln(Q_t) = \alpha + \beta t + \gamma_1\ln(L_t) + \gamma_2\ln(K_t) + u_t$							
	Without accounting identity link; Q, L and K are TSP				With accounting identity link; w, r and L are TSP $Q_1 = w_tL_t/0.6$, $K_t = (Q_t - w_tL_t)/r_t$			
R^2	0.684621	1.044977		1000	0.952968	0.007881		60
α	7.823030	1.209966	1000		8.610850	0.603917	1000	
S.e.(α)	1.157776	0.119029			0.586601	0.061528		
β	0.049195	0.008299	1000		0.049726	0.004558	1000	
S.e.(β)	0.008011	0.000824			0.004375	0.000444		
γ_1	0.004402	0.105725	52		0.502638	0.090402	1000	
S.e.(γ_1)	0.102048	0.010554			0.088039	0.009167		
γ_2	0.008073	0.102499	52		0.499985	0.051964	1000	
S.e.(γ_2)	0.101539	0.010273			0.051274	0.005434		
DW	2.019948	0.196317			2.019756	0.193111		

Mean denotes the mean value across 1000 runs. Standard deviation denotes the sample standard deviation of the 1000 mean values. Mean S.e. of a parameter denotes the mean of the 1000 standard errors. Standard deviation of the S.e. of a parameter denotes the sample standard deviation of the 1000 standard errors. $v_t = u_t - u_{t-1}$.

However, this is inconsequential. The reason is that since the data are linked through the income accounting identity, we know that the latter holds at all levels of aggregation.

The first set of results is a simple extension of Nelson & Kang's experiments. It constitutes the 'lower bound' case for three unrelated random walks (two explanatory variables). Compared with Nelson & Kang's regression with one explanatory random walk, the R^2 value is up slightly at 0.5754 (from 0.501). All other results resemble those in Nelson & Kang's regression: parameter estimates are not close to any realistic factor share; there are clearly no constant returns to scale; and differencing leads to a negligible R^2 value.[10]

The second set of results is based on random walks generated according to the accounting identity (1) and the definition of the labour share. In this set and in the following we arbitrarily choose a value for the labour share of 0.6. Since the growth rates of the wage and profit rates in this set are not constant (w_t and r_t are random walks, and thus Equation (6) is not the correct functional form that corresponds to the identity expression (1)), the R^2 value is not equal to one and there is still some scope for spuriousness to improve the R^2 value (by 0.12) when going from the non-spurious regression in first differences (R^2 value of 0.7522) to the spurious regression in levels (R^2 value of 0.8735).[11]

Regarding the estimated parameters, we observe the following. First and foremost, the parameters can be interpreted in terms of factor shares. Second, given the existence of the accounting identity link, taking first differences neither changes significantly the value of the parameters nor reduces the significance of these parameters, unlike in Set 1 (this occurs because the labour share is maintained perfectly constant). Finally, the number of runs with significant t-values (at the 5% significance level) for the parameters of capital and labour is up compared with the previous set, and close to 1000.

Since the growth rates of the wage and profit rates are not constant, constant returns to scale are rejected in 62.3% of all runs in levels; this percentage is lower than in the previous set where the accounting identity did not hold. The mean sum of the parameters of labour and capital across 1000 runs is 0.985042 with a standard deviation of 0.276800, a maximum value of 2.097430 and a minimum value of 0.008175. After differencing, the mean sum of the parameters of labour and capital is 0.999553 with a standard deviation of 0.069087, a maximum value of 1.234804 and a minimum value of 0.763048. First differencing reduces this number to a more appropriate 5.5%. The effects of spuriousness are further visible in the regression in levels in two respects. First, the Durbin–Watson statistic is very low at 0.3221. Second, the parameter of time, β, traditionally interpreted as the rate of technical progress, is significantly different from zero at the 5% level in 819 runs, despite the fact that the mean is zero (the weighted average of the growth rates of the wage and profit rates must be zero because w_t and r_t are random walks). First differencing then yields perfect results with a Durbin–Watson of 2, while β is significantly different from zero in 47 runs only.

For the third set of results variables are created as TSP, subject to the same size of (distinct) shocks and the same growth rate of an arbitrarily chosen 5%.[12] Since all variables now are TSP, the time trend is correctly included in the regression, and, unlike in the previous sets, the two regressions are not subject to the problems of spurious detrending. We report the results for the case that all variables are unrelated versus the case that the variables are linked in accordance with the accounting identity; the effects of imposing the accounting identity link are similar to those obtained for DSP variables. Once the accounting identity link has been instituted, R^2 improves from 0.684 (i.e. a 'lower bound' for the non-spurious regression) to 0.9530, constant returns to scale can no longer be rejected, and, again crucial, coefficient estimates turn highly significant and 'credible' in terms of a factor share interpretation. The parameter of time is correctly estimated. It equals $0.6{\times}5\% + 0.4{\times}5\% = 0.05$ (see Equation (4)) where 5% is the assumed value of the growth rate of the wage rate as well as that of the profit rate.

Finally, to take into account the possibility of cointegration among the series we fitted a dynamic reparameterization in error correction model (ECM) form of Set 2 with two lags in output and inputs (results available upon request).[13] Following the arguments in Section 2, it can easily be shown why this form is subject to the same issues discussed there, and why it can be interpreted in terms of the income identity. The results show a slight increase in the R^2 value (to 0.774) compared to the regression in first differences, but no change in the (long-run) parameters of capital and labour.

The conclusion of these Monte Carlo simulations is that if the profit rate, wage rate, and labour are difference stationary processes, then spuriousness may explain a small part of the good fit of a production function estimation, but obscure some issues such as constant returns to scale. The accounting identity link in the simulations turns out to be the major force leading to a high R^2 value, independent of the issue of spuriousness, and only the accounting identity link can explain the proximity of the estimated parameters to the factor shares, and thus the emergence of constant returns to scale. The importance of this last point must be stressed. Only the link of the variables through the identity gives rise to parameter values close to the factor shares. And this is why the generation of 'independent' random walks (as in Nelson & Kang, 1984, or in set 1 in Table 2) or trend stationary variables (as in set 3 in Table 2 here) yields parameter values that cannot be interpreted in terms of a Cobb–Douglas production function. Paradoxically, Fisher (1971) did not report the summary of the estimates he obtained.[14] There is no doubt, however, that, overall, the estimates he obtained had to be close to the shares. The reason, as discussed in Section 2, is that Fisher linked the series using the identity. Our simulations show that without this link, there is no way that the estimates will be anywhere close to the shares.

4. When will the Cobb–Douglas Function not Work?

We have seen that if factor shares and the growth rates of the wage and profit rates are perfectly constant, the Cobb–Douglas form is observationally equivalent to the accounting identity. This explains the results of Fisher (1971). In his seminal paper, Fisher concluded that the aggregate Cobb–Douglas production function would yield good results when factor shares were relatively constant, even in case such aggregate production function did not exist. The arguments in Section 2 explain why this had to be the case: if factor shares happen to be constant, the underlying accounting identity can be rewritten as a form that looks like the Cobb–Douglas form.

In this section we ask the following question: how much do the two assumptions of constant factor shares and constant growth rates of the wage and profit rates have to be relaxed for the regression estimation results to no longer be 'good'? (As indicated above, if factor shares were perfectly constant, and wage and profit rates grew at perfectly constant rates, one would not be able to obtain the OLS estimates of a Cobb–Douglas with a time trend due to perfect multicollinearity.) To answer this question, the two assumptions are now relaxed step by step with the economy being simulated again using 1000 runs, each with 100 observations. It is important to remember that, in Section 2, we showed that the coefficient of the time trend is definitionally the weighted average of the growth rates of the wage and profit rates (Equation (4)). Likewise, the fact that factor shares now vary does not imply that the accounting identity does not hold (Equation (2) remains). All it means is that the assumptions used to derive Equations (3)–(6) do not hold *exactly*. Therefore, if one fits Equation (6), we should not expect a perfect fit, and elasticities will not equal the factor shares. In this case, we need another path to track the factor shares. This will lead, as we saw above, to another production function (i.e. to another form of the identity). What we ask here is what occurs when factor shares vary, but one nevertheless fits the Cobb–Douglas form. Will it work?

Table 3. Overview of the simulations

Regression	$\ln(Q_t) = \alpha + \varphi t + \gamma_1 \ln(L_t) + \gamma_2 \ln(K_t) + u_t$	
Accounting identity link	$Q_t = \dfrac{w_t L_t}{a_t}$	$K_t = \dfrac{Q_t - w_t L_t}{r_t}$
Labor share process	$a_t = 0.06 + 0.9a_{t-1} + z_{1t}, z_{1t} \sim N(0,\mathrm{sd}_1^2)$, $\mathrm{sd}_1 = \{0.001, 0.01, 0.1\}$, and $a_0 = 0.6$	
All variables are DSP	All variables are TSP	Mixed case
$r_t = r_{t-1} + z_{2t}$ $z_{2t} \sim N(0,\mathrm{sd}_2^2), r_0 = 1$ $w_t = w_{t-1} + z_{3t}$ $z_{3t} \sim N(0,\mathrm{sd}_3^2), w_0 = 1$	$r_t = 1e^{0.01t + z_{2t}}$ $z_{2t} \sim N(0,\mathrm{sd}_2^2), r_0 = 1$ $w_t = 1e^{0.05t + z_{3t}}$ $z_{3t} \sim N(0,\mathrm{sd}_3^2), w_0 = 1$	$r_t = r_{t-1} + z_{2t}$ $z_{2t} \sim N(0,\mathrm{sd}_2^2), r_0 = 1$ $w_t = 1e^{0.05t + z_{3t}}$ $z_{3t} \sim N(0,\mathrm{sd}_3^2), w_0 = 1$
$L_t = L_{t-1} + z_{4t}$ $z_{4t} \sim N(0,\mathrm{sd}_4^2)$ $\mathrm{sd}_4 = 0.01, L_0 = 1$	$L_t = 1e^{0.025t + z_{4t}}$ $z_{4t} \sim N(0,\mathrm{sd}_4^2)$ $\mathrm{sd}_4 = 0.01, L_0 = 1$	$L_t = 1e^{0.025t + z_{4t}}$ $z_{4t} \sim N(0,\mathrm{sd}_4^2)$ $\mathrm{sd}_4 = 0.01, L_0 = 1$

Table 3 presents an overview over the three types of simulations conducted. We fit a Cobb–Douglas production function in levels including a time trend. In the first type of simulation, labour, profit rate, and wage rate are difference stationary processes (DSP); in the second type these variables are trend stationary processes (TSP); and in the third type, labour and wage rate are TSP, while the profit rate is DSP. Since the results of these three simulations do not differ systematically we focus in the following on the third one only. As before, once the accounting identity is imposed, the precise values of the variables are inconsequential.

Table 4 translates various shock sizes into characteristics of the time series of the labour share, profit rate growth, wage growth, and labour growth.[15] The most important results are reported in the graphs in the Appendix for a standard deviation of the shock to the labour share of $\mathrm{sd}_1 = 0.01$. We report 12 graphs: coefficient of time and its standard deviation; coefficient of labour and its standard deviation; coefficient of capital and its standard deviation; number of runs with significant time, labour, and capital coefficients; R^2; Durbin–Watson; and number of runs rejecting constant returns to scale.

For each case, four time series are created, namely, employment, labour share, profit rate and wage rate. Output and capital stocks are derived from these four time series through the accounting identities. Throughout all simulations the standard deviation of the shock to employment (sd_4) is held at 0.01, with an imposed mean growth rate of 2.5% and an initial value $L_0 = 1$, yielding a mean of mean labour growth rates of 2.55% and a mean standard deviation of labour growth rates of 0.0209 (bottom of Table 4). The choice of initial value has no influence on growth rates and their variation.

Table 4. Characteristics of the simulated variables in terms of shock sizes

		Standard deviation of shock	Mean of means	Mean standard deviation	Redraws
Labour share	$a_t = 0.06 + 0.9\, a_{t-1} + z_{1t},\ a_0 = 0.6,\ z_{1t} \sim N(0,\mathrm{sd}_1^2)$				
		$\mathrm{sd}_1 = 0$	0.6000000	0.000000	0
		$\mathrm{sd}_1 = 0.001$	0.6000013	0.001998	0
		$\mathrm{sd}_1 = 0.005$	0.6000045	0.010003	0
		$\mathrm{sd}_1 = 0.01$	0.599955	0.019940	0
		$\mathrm{sd}_1 = 0.05$	0.598241	0.100183	0
		$\mathrm{sd}_1 = 0.01$	0.582391	0.182977	1808
		$\mathrm{sd}_1 = 1$	0.502730	0.283634	169530
Growth rate of profit rate	Where the profit rate is created as: $r_t = r_{t-1} + z_{2t},\ r_0 = 1,\ z_{2t} \sim N(0\mathrm{sd}._2{}^2),\ \mathrm{sd}_2 = e^{k_2}$				
	$k_2 = -11$	$\mathrm{sd}_2 = 1.670{*}10^{-5}$	0.000000	0.000017	0
	$k_2 = -10$	$\mathrm{sd}_2 = 4.540{*}10^{-5}$	0.000000	0.000045	0
	$k_2 = -9$	$\mathrm{sd}_2 = 1.234{*}10^{-4}$	0.000000	0.000123	0
	$k_2 = -8$	$\mathrm{sd}_2 = 3.355{*}10^{-4}$	0.000001	0.000334	0
	$k_2 = -7$	$\mathrm{sd}_2 = 9.119{*}10^{-4}$	0.000006	0.000910	0
	$k_2 = -6$	$\mathrm{sd}_2 = 0.002479$	0.000008	0.002484	0
	$\boldsymbol{k_2 = -5}$	$\mathbf{sd_2 = 0.006738}$	**0.000021**	**0.006816**	**0**
	$\boldsymbol{k_2 = -4}$	$\mathbf{sd_2 = 0.018316}$	**0.002658**	**0.039559**	121
	$k_2 = -3$	$\mathrm{sd}_2 = 0.049787$	0.069520	0.573574	3741
	$k_2 = -2$	$\mathrm{sd}_2 = 0.135335$	0.189191	1.629150	9015
Growth rate of wage rate	Where the profit rate is created as: $w_t = 1e^{0.05t + z_{3t}},\ z_{3t} \sim N(0\mathrm{sd}_3{}^2),\ \mathrm{sd}_3 = e^{k_3},\ w_0 = 1$				
	$k_3 = -11$	$\mathrm{sd}_3 = 1.670{*}10^{-5}$	0.051271	0.000025	
	$k_3 = -10$	$\mathrm{sd}_3 = 4.540{*}10^{-5}$	0.051271	0.000067	
	$k_3 = -9$	$\mathrm{sd}_3 = 1.234{*}10^{-4}$	0.051271	0.000183	
	$k_3 = -8$	$\mathrm{sd}_3 = 3.355{*}10^{-4}$	0.051271	0.000498	
	$k_3 = -7$	$\mathrm{sd}_3 = 9.119{*}10^{-4}$	0.051271	0.001361	
	$k_3 = -6$	$\mathrm{sd}_3 = 0.002479$	0.051278	0.003683	
	$\boldsymbol{k_3 = -5}$	$\mathbf{sd_3 = 0.006738}$	**0.051317**	**0.009994**	
	$\boldsymbol{k_3 = -4}$	$\mathbf{sd_3 = 0.018316}$	**0.051629**	**0.027306**	
	$k_3 = -3$	$\mathrm{sd}_3 = 0.049787$	0.053865	0.074314	
	$k_3 = -2$	$\mathrm{sd}_3 = 0.135335$	0.071079	0.208469	
Growth rate of labour	Where labour is created as: $L_t = 1e^{0.025t + z_{4t}},\ z_{4t} \sim N(0\mathrm{sd}_4{}^2),\ L_0 = 1$				
		$\mathbf{sd_4 = 0.01}$	**0.025535**	**0.020939**	

'Redraws' denotes the number of times an observation had to be redrawn to avoid its value at any one point of time exceeding or falling below a certain limit. The labour share was forced to fall within the]0,1[interval. With an imposed mean labour share of 0.6, the upper limit is more likely to be reached than the lower limit, truncation is more likely to occur at the upper limit, and for large shocks the mean of means is therefore biased downward. Due to the truncation the mean standard deviation of labour share series for large shocks is also biased downward. The profit rate was forced to be positive. For large shocks the mean of means is therefore biased upward, while the mean standard deviation is biased downward. Large shocks to the growth rate of a trend stationary process (wage rate and labour) lead to mean growth rates of the series exceeding the imposed growth rate due to the asymmetrical nature of the exponential function.

The size of the other three shocks, namely the shock to the labour share, sd_1, to the profit rate, sd_2, and to the wage rate, sd_3, is then varied independently to assume a wide range of values. The DGP of the labour share is an AR(1) with drift 0.06 and first order autoregressive parameter 0.9. (This process generates a labour share series that looks credible and consistent with actual series. Graph available upon request.) The profit rate follows a DSP with initial value of unity and zero drift, while the wage rate follows a trend stationary process with initial value unity and an imposed mean growth rate of 5%.[16]

How is Table 4 to be interpreted? The table shows (see rows in bold), for example, that standard deviations of the shock to the labour share of $sd_1 = 0.01$ and $sd_1 = 0.005$ translate into standard deviations of the labour share of 0.01994 and 0.010003, respectively. A shock to the labour share of $sd_1 = 0.01$ implies, therefore, that if the last period's value of the labour share was 0.6, this period's value will fall within the interval 0.56 to 0.64 with a probability of 95% (i.e. 0.6 $\pm$ 2 $\times$ standard deviation). If $sd_1 = 0.005$, this period's labour share will be between 0.58 and 0.62 with the same probability. If, on the other hand, $sd_1 = 0.1$, the standard deviation of the labour share is 0.182977, which is too high to be realistic (i.e. this period's labour share would fall between 0.234 and 0.965 in 95% of all cases).

In order to cover a wide range of values systematically, the standard deviations of the shocks to the profit rate (sd_2) and wage rate (sd_3) are varied following an exponential processes (see Table 4). For example, values of $k_2 = -5$ and $k_2 = -4$ imply a standard deviation of the normally distributed shocks to the profit rate of $sd_2 = 0.0068$ and $sd_2 = 0.0183$, and a mean (across 1000 runs) standard deviation of the growth rate of the profit rate of almost 0.7% and 4.0% (corresponding to a $\pm$0.7 and $\pm$4.0 percentage point change in the growth rate between two periods). These values give rise to a growth rate of the profit rate that lies between –1.4% to 1.4% (0.000021 $\pm$ (2$\times$0.006816)) for $k_2 = -5$ with a probability of 95% (mean $\pm$ two standard deviations). And for $k_2 = -4$, the corresponding growth rate interval ranges from –8% and 8%.

For the growth rate of the wage rate, values of $k_3 = -5$ and $k_3 = -4$ (a standard deviation in the shock of $sd_3 = 0.0068$ and $sd_3 = 0.0183$) imply mean standard deviations of 1% and 2.7%, respectively. These values give rise to a growth rate of the wage rate that lies between 3.13% and 7.13% for $k_3 = -5$ with a probability of 95%, while for $k_3 = -4$, the corresponding growth rate interval ranges from –0.30% and 10.56%, respectively.

The graphs in the Appendix report the effects of a certain variation in the growth rate of the profit rate, and a certain variation in the growth rate of the wage rate, on the regression results for a given shock to the labour share. We show the graphs for the case of $sd_1 = 0.01$ (We also used $sd_1 = 0.001$ and $sd_1 = 0.1$. The graphs corresponding to these two shocks are available upon request.) They show how, for very small values of the standard deviations of the shock to the profit rate and of the shock to the wage rate (sd_2 and sd_3, respectively), the coefficients of time, labour, and capital are very accurately estimated with minimal standard errors and significant coefficients (the expected coefficient of time equals the labour share times the mean growth rate of the wage rate, 0.05, plus the capital share times the mean growth rate of the profit rate, i.e. $0.6\times0.05 + 0.4\times0 = 0.03$).

Three observations as to what happens to the coefficients when the variation in the growth rate of the profit rate and/or in the growth rate of the wage rate increases, are the following. First, once sd_3 exceeds a threshold level (i.e. the value of the shock for which the coefficient estimates begin deteriorating) of $sd_3 = e^{-4}$ (i.e. a standard deviation of the growth rate of the wage rate of 2.73 percentage points) the coefficients of time and labour move towards zero rapidly as long as sd_2 remains smaller than e^{-4} (i.e., a standard deviation of the growth rate of the profit rate of 3.95 percentage points). The coefficient of capital, on the other hand, increases. Second, once sd_2 exceeds a threshold level of $sd_2 = e^{-5}$ (i.e. a standard deviation of the growth rate of the profit rate of 0.68 percentage points), the coefficients of time and labour rise above their actual values as long as sd_3 remains smaller than e^{-3} (i.e. a standard deviation of the growth rate of the wage rate of 7.43 percentage points).

The coefficient of capital, on the other hand, decreases. Thirdly, if sd_3 and sd_2 both exceed e^{-5} (i.e. standard deviations of 0.99% and 0.68%, respectively) all three coefficients on a very narrow path of shock combinations *may* remain very accurate, although the standard errors increase very rapidly (i.e. a high variation in the growth rate of the profit rate leads to high coefficients of time and labour, while a high variation in the growth rate of the wage rate leads to low coefficients of time and labour, and vice-versa for capital; the result in case of a mix of these two influences is a narrow range where the coefficients remain accurate).

Throughout, the number of runs with significant coefficients of time and labour (in 1000 runs) drops off rapidly from 1000 once sd_3 exceeds e^{-4} independent of the size of sd_2. The coefficient of capital exhibits similar behaviour, except that the dependence on sd_2 and sd_3 is reversed.[17] The R^2 value remains very high across all values of sd_2 and sd_3. It starts falling minimally once sd_2 or sd_3 exceeds e^{-5}; it experiences its biggest fall, to a minimum value of 0.9980, once $sd_2 = sd_3 = e^{-2}$ (i.e. a standard deviation of 162.91% for the growth rate of the profit rate, and of 20.84% for the growth rate of the wage rate). The Durbin–Watson statistic appears to vary almost independently of sd_2 and reaches values close to two around $sd_3 = e^{-5}$. The null hypothesis of constant returns to scale can practically never be rejected: rejections hover between 35 and 65 out of 1000 runs.

Comparing these results with the cases of $sd_1 = 0.001$ and $sd_1 = 0.1$ allows a generalization of the above observations and some additional conclusions.

(1) The size of the shocks to the profit and wage rates relative to the size of the shock to the labour share matters, i.e. the individual threshold levels of sd_2 and sd_3 depend on sd_1. The larger sd_1, the larger may sd_2 and sd_3 be before parameter estimates deteriorate. In the case of $sd_1 = 0.1$, the coefficients exhibit small unsystematic fluctuations around the actual values; however, the standard errors of the parameter estimates increase with sd_1.

(2) The size of the shock to the profit rate relative to the size of the shock to the wage rate matters for all levels of the shock to the labour share, sd_1. If only one of the first two shocks exceeds a certain threshold level determined by sd_1, then parameter estimates drop off or rise sharply, and standard errors increase; the number of runs with significant parameter estimates drops off depending only on the size of the shock corresponding to this time series (where the coefficient of time depends on sd_3), unless sd_1 is relatively large, in which case the number of runs with significant estimates of time and labour coefficients are low throughout. If sd_2 and sd_3 jointly exceed a certain threshold level determined by sd_1, the parameter estimates across a narrow range of sd_2 and sd_3 combinations may remain close to the actual values but their standard errors increase rapidly and the R^2 value drops.

(3) Even when sd_2 and sd_3 exceed their threshold levels, the R^2 value still remains very high and close to unity across all levels of sd_1.
(4) The Durbin–Watson statistic deteriorates quickly even in the case of the smallest shock to the labour share, i.e. $sd_1 = 0.001$. It remains acceptable only for a very narrow range of shocks sd_2 and sd_3.
(5) The number of runs rejecting constant returns to scale is independent of the variation in the labour share, and independent of the variation in the growth rate of the profit rate and in the growth rate of the wage rate; it stays, correctly, around 50 runs out of 1000 runs under all shock combinations.

Overall, the most remarkable feature is that already small variations in the growth rates of the wage and profit rates make the Cobb–Douglas production function with a time trend yield poor results; coefficient estimates are far from the factor shares (but the R^2 value remains high due the presence of the identity). Only rather low variation in the growth rates of the wage and profit rates, as well as very specific combinations of higher variation in these growth rates, lead to good Cobb–Douglas production function estimation results.

5. Revisiting the Empirical Evidence

The simulations in the previous section indicate an important result, namely, that (small) variations in the growth rates of the wage and profit rates are responsible for failures of the Cobb–Douglas production function with a time trend to perform well. In fact, the variations in the simulations for which the Cobb–Douglas form deteriorates are, in general, smaller than those displayed by actual series. On the other hand, this functional form is very robust to variations in the factor shares. To provide evidence that this is the case, we first fit equation (2) to time series for the US for SIC 22 and SIC 36 for 1959–91, but *only* under the assumption of constant factor shares ($a_t = a$). This yields $q_t = a\hat{w}_t + (1 - a)\hat{r}_t + al_t + (1 - a)k_t$, and $\ln Q_t = c + a\ln w_t + (1 - a)\ln r_t + a\ln L_t + (1 - a)\ln K_t$, in growth rates and in levels, respectively. If it is true that the parameters are (sufficiently) constant, all four coefficients estimated unrestricted (denoted γ_i for $i = 1,2,3,4$) will be precisely estimated, should be approximately equal to the factor shares ($\gamma_1 = \gamma_3 = a$; $\gamma_2 = \gamma_4 = 1-a$), and the regression should display a very high fit. Results are shown in Table 5. This is confirmed for both cases and with the regressions in growth rates (equation (2)) and in levels. The extremely high fits and high t-statistics, as well as the proximity of the estimated coefficients (highly significant) to the factor shares in both regressions can only be explained in terms of the accounting identity. Note the small difference in R^2 between the regressions in levels and in first differences (despite the difference in Durbin–Watson). What drives the high fit is clearly the underlying accounting identity.

We can now return to the results obtained in Table 1 and compare them with those here. The equations estimated were interpreted as Cobb–Douglas production functions. However, in the light of our arguments, they can be interpreted as the identity under the assumptions of constant factor shares and constant growth rates of the wage and profit rates (i.e. equations (3) and (5)). The first regression is in growth rates and the second in levels. In the first case, the weighted average of the growth rates of the wage and profit rates becomes the constant; and in the second case, the regression in levels, it is the parameter of the time trend. Recall that we argued that these expressions are identical with the Cobb–Douglas production

Table 5. Value added accounting identity

SIC 22: Textile and mill products

(A) $q_t = \gamma_1 \hat{w}_t + \gamma_2 \hat{r}_t + \gamma_3 l_t + \gamma_4 k_t$

γ_1	γ_2	γ_3	γ_3	$\bar{a}$; σ	Max; Min
0.522 (27.00)	0.476 (46.99)	0.519 (22.90)	0.468 (19.93)	0.508; 0.045	0.579; 0.423

$R^2 = 0.998$; DW = 1.36

(B) $\ln Q_t = C + \gamma_1 \ln w_t + \gamma_2 \ln r_t m + \gamma_3 \ln L_t + \gamma_4 \ln K_t$

Constant	γ_1	γ_2	γ_3	γ_4
1.02 (16.78)	0.526 (24.33)	0.48 (41.21)	0.495 (36.10)	0.463 (33.45)

$R^2 = 0.999$; DW = 0.69

SIC 36: Electric and electronic equipment

(A) $q_t = \gamma_1 \hat{w}_t + \gamma_2 \hat{r}_t + \gamma_3 l_t + \gamma_4 k_t$

γ_1	γ_2	γ_3	γ_3	$\bar{a}$; σ	Max; Min
0.507 (22.78)	0.502 (62.75)	0.503 (43.11)	0.502 (34.22)	0.48; 0.048	0.564; 0.404

$R^2 = 0.998$; DW = 1.80

(B) $\ln Q_t = C + \gamma_1 \ln w_t + \gamma_2 \ln r_t m + \gamma_3 \ln L_t + \gamma_4 \ln K_t$

Constant	γ_1	γ_2	γ_3	γ_4
0.82 (17.60)	0.53 (33.96)	0.51 (54.69)	0.47 (50.57)	0.50 (62.20)

$R^2 = 0.999$; DW = 0.81

t-values in parenthesis. Data for the US for 1959–91. $\bar{a}$ is the average labour share; σ is the standard deviation; Max. and Min. are maximum and minimum values of the labour share, respectively.

function, but need not be such. The estimated factor elasticities were far from the factor shares, and the fit in the regressions in growth rates is substantially lower than in the regressions in Table 5; while R^2 in the regressions in levels was still very high. Given our previous arguments and the results in Table 5, we know now that the reason behind the results in Table 1 *must be* the variability in the growth rates of the factor prices, namely, that the weighted average of the growth rates of the wage and profit rates is not well approximated with a constant (φ). Since the weighted average of the factor prices is not well proxied by a constant, spuriousness matters a bit, as

Table 6. Variability in factor prices

SIC 22: Textile mill products

	Mean	Max.	Min.	Stan. Deviation
$\hat{w}$	0.024	0.108	−0.077	0.035
$\hat{r}$	0.023	0.224	−0.144	0.087

SIC 36: Electric and electronic equipment

	Mean	Max.	Min.	Stan. Deviation
$\hat{w}$	0.026	0.049	−0.044	0.0195
$\hat{r}$	0.0057	0.173	−0.217	0.086

discussed in Section 3 (the series are I(1)). Unlike in the regressions in Table 1, now there is virtually no difference in fit as well as in the size of the estimated parameters between the regressions in levels and in differences. However, there is still a substantial difference in the Durbin–Watson. This simply 'proves' that spuriousness is a secondary issue.

Table 6 shows that the standard deviations of the growth rates of the wage and profit rates are well into the region where the parameters deteriorate, in particular the standard deviation of the growth rate of the profit rates, which corresponds to a shock value exceeding $k_2 = -4$ (see Table 4) in both examples.

Figures 1, 2 and 3 plot the growth rates of the wage and profit rates for the two sectors considered (SIC 22, SIC 36), and the weighted average of the growth rates of the wage and profit rates. The first two graphs show that the variation in the growth rate of the profit rate is substantially higher than that in the growth rate of the wage rate. The last one indicates that the weighted average cannot be well approximated by a constant. As Shaikh (1980) showed, one would have to resort to a complex time trend with sines and cosines.

How can one get such complex approximation? There is no textbook method to obtain it. It is simply a matter of trial and error. We have already seen that factor shares are sufficiently stable for regression purposes. Thus, we can construct an approximation to the path of the weighted average of the wage and profit rates from the identity under such an assumption. For example, for SIC22, and using the value of the average factor shares, i.e. $\bar{a} = 0.508$ and $(1 - \bar{a}) = 0.492$, such a path is $(0.508 \times \ln w_t + 0.492 \times \ln r_t)$. this is shown in Figure 4. Now all we have to find is the appropriate mathematical form to track it. Through trial and error we constructed $A(t) = [T - \mathrm{Sin}(T) - \mathrm{Cos}(T) - \mathrm{Sin}(T^2) - \mathrm{Cos}(T^2) + \mathrm{Cos}(T^3) + \mathrm{Sin}(T^4) + \mathrm{Cos}(T^6)]$, where T is a time trend, 'Sin' is the sine function, and 'Cos' is the cosine function. Estimation results are (t-values in parenthesis):

$$\ln Q_t = \underset{(3.00)}{0.0185} \times A(t) + \underset{(2.19)}{0.577} \times \ln L_t + \underset{(3.03)}{0.552} \times \ln K_t$$

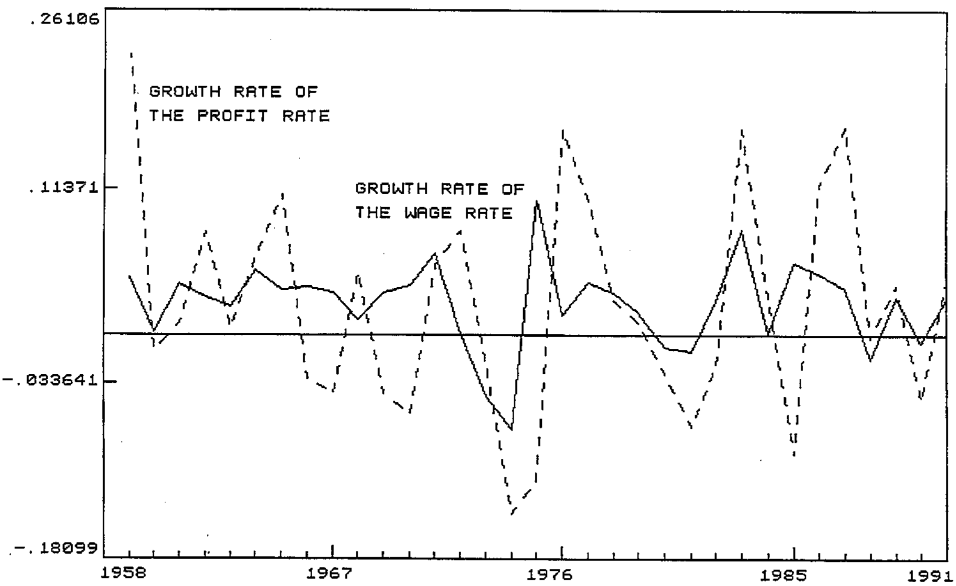


Fig. 1. Growth rates of the wage and profit rates. SIC 22.

Certainly this form is not perfect yet (the correlation between the path ($0.508 \times \ln w_t$ + $0.492 \times \ln r_t$) and $A(t)$ is 0.923), but a comparison with the results displayed in Table 1, when the linear time trend was used, is indicative of the great improvement, in particular in terms of the proximity of the estimates to the factor shares.[18]

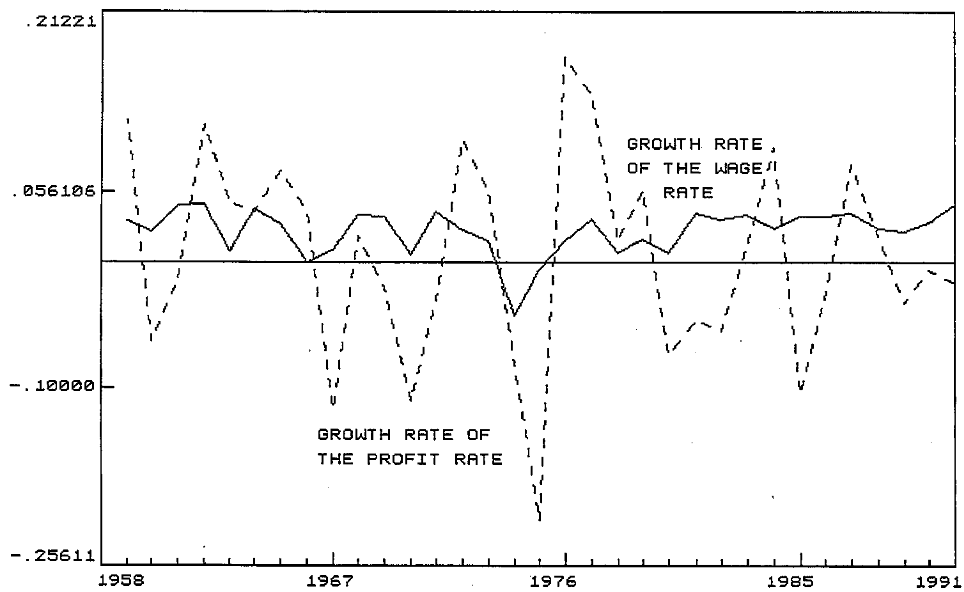


Fig. 2. Growth rates of the wage and profit rates. SIC 36.

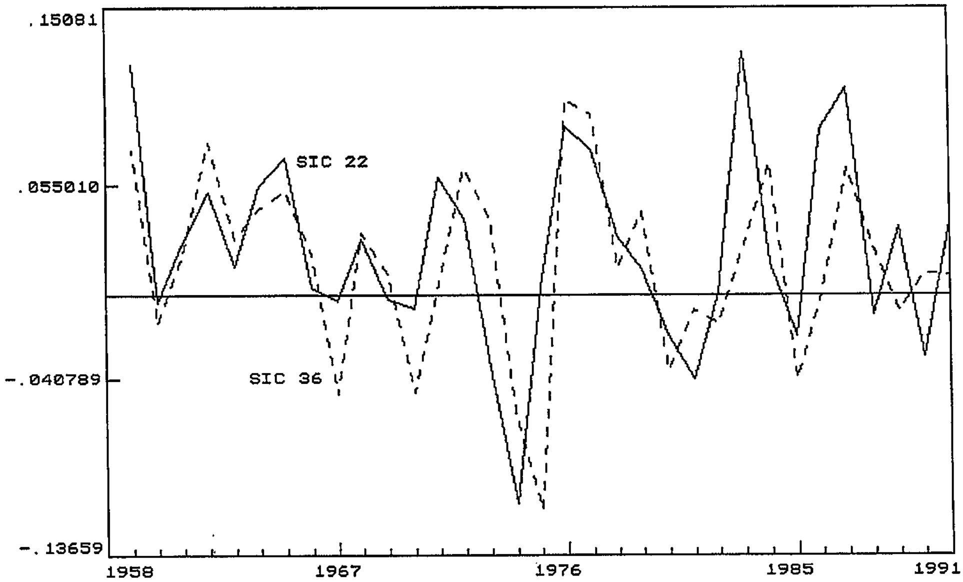


Fig. 3. Weighted average of the growth rates of the wage and profit rates.

6. Conclusions

This paper has analysed why aggregate production functions appear to fit (at least sometimes) in the light of Fisher's work on aggregation and Shaikh's work with the accounting identity. It is rather unfortunate that despite this important body of work, economists still refer to the aggregate production function as a summary of

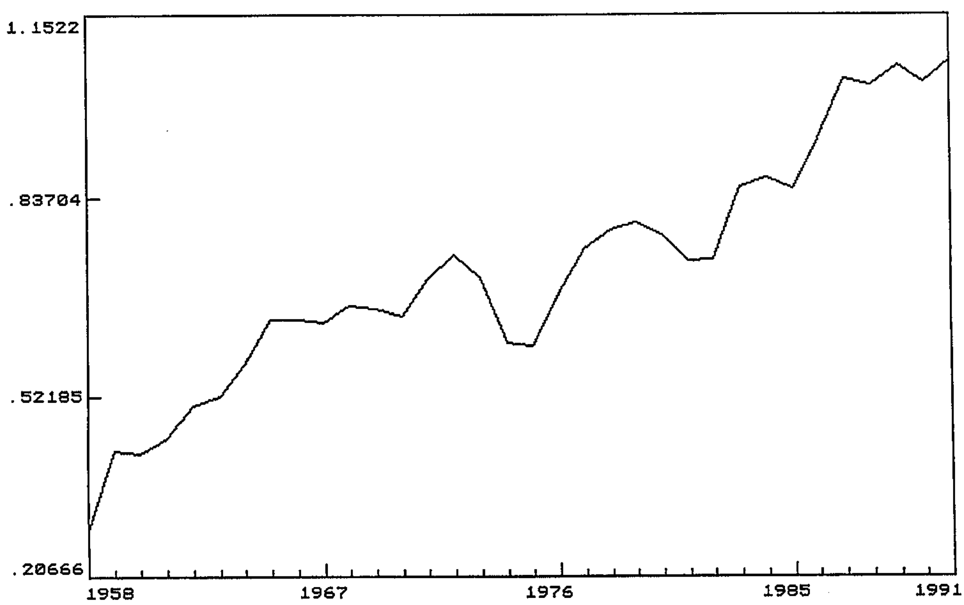


Fig. 4. Weighted average of the wage and profit rates.

the aggregate technology. Fisher (1971) used simulation analysis to study the empirical conditions under which the Cobb–Douglas function would work well. This was that factor shares had to be constant. However, Shaikh (1980) showed that Fisher's observation was simply the result of the fact that underlying all aggregate production functions there is the income identity that relates output to the sum of the wage bill plus total profits.

This paper has elaborated on the issue, and has related Fisher's simulations to Shaikh's arguments by explicitly linking the variables in the production function and the income identity. Likewise, the statistical properties of the data (e.g. the presence of unit roots) have been considered. We have used Monte Carlo simulations to answer two questions. First, how much spuriousness can help explain the relatively good fits of the Cobb–Douglas production function? The simulations show that the contribution of spuriousness to a high R^2 is minor once we properly account for the fact that input and output data used in production function estimations are linked through the income accounts, i.e. output equals wages plus profits, in value terms. It is mostly the link through this identity that explains the results. Secondly, we have studied how much factor shares have to vary in an economy so as to render the Cobb–Douglas production function with a time trend a bad choice for modelling and estimation purposes.

We conclude that the Cobb–Douglas form is robust to relatively large variations in the factor shares. However, what makes this form quite often fail are the variations in the growth rates of the wage and profit rates. The weighted average of these two growth rates has been shown to be the coefficient of the time trend. This implies that, in most applied work, a Cobb–Douglas form (i.e. approximation to the income accounting identity) should work. We just have to find *which* Cobb–Douglas form with a dose of patience in front of the computer.

Notes

We thank Colin Hargreaves, Lutz Kilian, John McCombie, Miguel Ramirez, the participants at the Western Economic Association International meetings in San Francisco, 28 June – 2 July 1996, and Vancouver, 29 June – 3 July 2000, and the participants of the Economics Seminars at the Hong Kong University of Science and Technology, the University of Hong Kong, Universidad Carlos III de Madrid, and the Federal Reserve Bank of Atlanta for their comments on earlier versions. Likewise, we are grateful to two anonymous referees for their suggestions, which led to a substantial improvement in the argument. A previous version of this paper was circulated earlier under the title 'On Production Functions, Technical Progress and Time Trends.' Jesus Felipe acknowledges financial support from the Center for International Business Education and Research (CIBER) at the Georgia Institute of Technology. The usual disclaimer applies.

1. Solow, being aware of the aggregation problems, once claimed: 'I have never thought of the macroeconomic production function as a rigorously justifiable concept. In my mind it is either an *illuminating parable*, or else a mere device for handling data, to be used so long as it gives good empirical results, and to be abandoned as soon as it doesn't, or as soon as something better comes along (Solow, 1966, 1259–1260; italics added).
2. In an unfortunately unknown paper, the Nobel Prize Laureate Herbert Simon (1979) had argued in a similar way.
3. For example, recently Mankiw commented: '[. . .] I have always found the high R^2 reassuring when I teach the Solow growth model. Surely, a low R^2 in this regression would have shaken my faith . . .' Mankiw (1997, 104).
4. Interestingly, this literature provides a theoretical justification for why the elasticities may differ from the factor shares, contrary to Fisher's initial motivation (i.e. that usually elasticities equal the factor shares). Romer (1987) even talked about a 'suggestive puzzle.' In fact, in empirical work with the Cobb–Douglas, most often the elasticities do not equal the factor shares.
5. LEETS is simply the word STEEL spelled backwards. This is how Joan Robinson did it, in honour of J. E. Meade.

6. Fisher (1971) used the identity to generate total profits residually; and used the marginal productivity condition of labour to allocate the latter efficiently. But for the Cobb–Douglas this condition is that $w_t = a(Q_t/L_t)$. However, it can be seen that this expression also follows from the definition of the labour share $a_t = (w_t L_t)/Q_t$, under the assumption that the labour share is constant. See Shaikh (1980).
7. For a recent use of cointegration analysis in production functions see Otto & Voss (1996).
8. One could argue that most studies today do not fit the production function directly, as discussed here (e.g. Caballero & Lyons, 1992; Mankiw *et al.*, 1992; Basu & Fernald, 1995). These other cases are analysed in Felipe (2001). It is shown that they are special cases of the one considered here.
9. The regressions here—in contrast to those of Nelson & Kang—are run in logarithms. Taking the logarithm of a negative value is avoided by choosing a sufficiently large positive first-period value of 1000 in combination with Nelson & Kang's $N(0,1)$ shock. Running a regression in logarithms of one random walk on an unrelated second random walk with initial values 1000 we fully reproduce Nelson & Kang's results.
10. If all DSP processes include a drift of 0.5 the R^2 value reaches 0.9743 (dropping to 0.0201 once the regression is run in first differences), but otherwise there is no systematic difference to the version without drift.
11. We could have chosen a smaller shock or a larger initial value than the ones used here ($\varepsilon_t \sim N(0,1)$, $X_0 = 1000$). This is not a problem. All that matters here is to show the effect of the accounting identity link.
12. The initial value of each variable does not matter. We leave it at 1000. We have introduced a shock z_t to the growth rates in order to avoid the perfect multicollinearity that appears in the case of a perfectly constant labour share and perfectly constant growth rates of the wages and profit rates.
13. The model estimated was:

$$q_t = \beta + \delta_1 q_{t-1} + \delta_2 l_t + \delta_3 l_{t-1} + \delta_4 k_t + \delta_5 k_{t-1} + \lambda_1 Q_{t-1} + \gamma_2 L_{t-1} + \lambda_3 K_{t-1} + u_t,$$

where the long-run elasticities are given by $\eta_1 = -(\lambda_2 / \lambda_1)$ and $\eta_2 = -(\lambda_3 / \lambda_3)$.
14. Fisher (1971, p. 312) indicates that in some cases he obtained 'ridiculous' coefficients.
15. Throughout all estimations that follow below we impose restrictions in our simulations. The labour share cannot exceed unity or fall below zero; the profit rate, wages and employment cannot fall below zero. If any of them does, its value is redrawn.
16. Introduction of a drift in the profit rate process would yield a growth rate different from zero, but since the growth rate would continuously decrease across 1000 observations, the objective of relaxing the assumption of constant growth rates by applying a shock of potentially equal size in each period would not be achievable.
17. The only exception is the standard errors, which increase rapidly as either sd_2 or sd_3 exceeds e^{-6} but then drop off again once sd_2 or sd_3 exceeds e^{-3}—possibly due to the large technical impact of truncation on the mean growth rate of the profit rate as well as its variation (see Table 4).
18. Felipe & Adams (2001) and Felipe & McCombie (2001b) have also approximated the weighted average of the growth rates of the wage and profit rates through complex functions with two other different data sets. In both cases the approximations provide quasi-perfect fits. It is simply a matter of trial and error, and a lot of patience in front of the computer.

References

Aschauer, D. A. (1989) Is public expenditure productive? *Journal of Monetary Economics*, 23(2), pp. 177–200.

Basu, S. & Fernald, J. F. (1995) Are apparent spillovers a figment of specification error?, *Journal of Monetary Economics*, 36, pp. 165–188.

Basu, S. & Fernald, J. F. (1997) Returns to scale in U.S. production: estimates and implications, *Journal of Political Economy*, 105(2), pp. 249–283.

Caballero, R. J. & Lyons, R. K. (1992) External effects in U.S. procyclical productivity, *Journal of Monetary Economics*, 29, pp. 209–225.

Durlauf, S. N. & Phillips, P. C. B. (1988) Trends versus random walks in time series analysis, *Econometrica*, 56, pp. 1333–1354.

Felipe, J. (2000) On the myth and mystery of Singapore's 'zero TFP', *Asian Economic Journal*, 14(2), pp. 187–209.

Felipe, J. (2001) Endogenous growth, increasing returns, and externalities: an alternative interpretation of the evidence, *Metroeconomica*, forthcoming.

Felipe, J. & Adams, F. G. (2001) 'A Theory of Production.' The estimation of the Cobb–Douglas function: a retrospective view, Georgia Institute of Technology and Northeastern University, Manuscript.

Felipe, J. & McCombie, J. S. L. (2001a) The CES production function, the accounting identity, and Occam's Razor, *Applied Economics*, forthcoming.

Felipe, J. & McCombie, J. S. L. (2001b) Some methodological problems with recent neoclassical analyses of the East Asian miracle, Georgia Institute of Technology and Downing College (Cambridge), manuscript.

Fisher, F. M. (1969) The existence of aggregate production functions, *Econometrica*, 37(4), pp. 553–577.

Fisher, F. M. (1971) Aggregate production functions and the explanation of wages: a simulation experiment, *The Review of Economics and Statistics*, 53(4), pp. 305–325.

Fisher, F. M. (1993) *Aggregation. Aggregate Production Functions and Related Topics* John Monz (Ed.) (Cambridge, MA, The MIT Press).

Griliches, Z. & Mairesse, J. (1998) Production functions: the search for identification, in: Steinar Strøm (Ed.) *Econometrics and Economic Theory in the 20th Century* (Cambridge University Press).

Hall, R. E. (1990) Invariance properties of Solow's productivity residual, in: Peter Diamond (Ed.) *Growth/Productivity/Employment* (Cambridge, The MIT Press).

Kaldor, N. (1956) Alternative theories of distribution, *Review of Economic Studies*, 23, pp. 83–100.

Klette, T. J. & Griliches, Z. (1996) The inconsistency of common scale estimators when output prices are unobserved and endogenous, *Journal of Applied Econometrics*, 11, pp. 343–361.

Lucas, R. (1970) Capacity, overtime, and empirical production functions, *American Economic Review*, 60(2), pp. 23–27.

McCombie J. S. L. (1987) Does the aggregate production function imply anything about the laws of production? A note on the Simon and Shaikh critiques, *Applied Economics*, 19, 1121–1136.

McCombie, J. S. L. & Dixon, R. (1991) Estimating technical change in aggregate production functions: a critique, *International Review of Applied Economics*, 24–46.

Mankiw, G. N. (1997) Comment, *NBER Macroeconomics Annual 1997*, pp.103–107.

Mankiw, G., Romer, D. & Weil, D. N. (1992) A contribution to the empirics of economic growth, *Quarterly Journal of Economics*, 107(2), pp. 407–437.

Munnell, A. (1992) Infrastructure investment and economic growth, *Journal of Economic Perspectives*, 6(4), pp. 189–198.

Nelson, C. & Kang, H. (1984) Pitfalls in the use of time as an explanatory variable in regressions, *Journal of Business and Economic Statistics*, 2, pp. 73–82.

Otto, G. & Voss, G. M. (1996) Public capital and private production in Australia, *Southern Economic Journal*, 62(3), pp. 723–738.

Romer, P. (1987) Crazy explanations for the productivity slowdown, *NBER Macroeconomics Annual 1987*, pp. 163–201.

Samuelson, P. (1979) Paul Douglas's measurement of production functions and marginal productivities, *Journal of Political Economy*, 87(5), pp. 923–939.

Shaikh, A. (1974) Laws of production and laws of algebra: the Humbug production function, *The Review of Economics and Statistics*, 56(61), pp. 115–120.

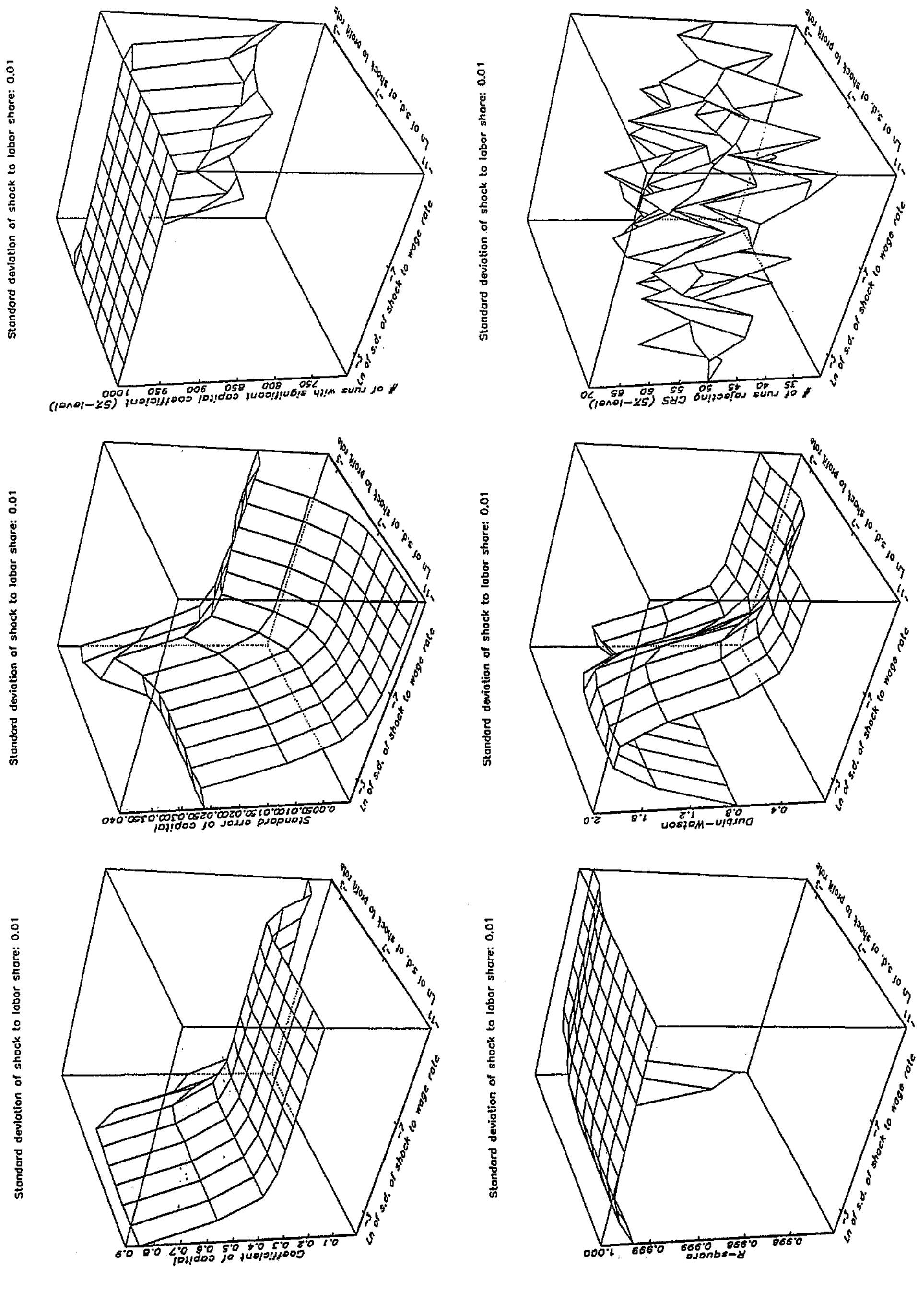
Shaikh, A. (1980) Laws of production and laws of algebra: Humbug II, in: Edward J. Nell (Ed.) *Growth, Profits and Property. Essays in the Revival of Political Economy* (Cambridge, MA, Cambridge University Press).

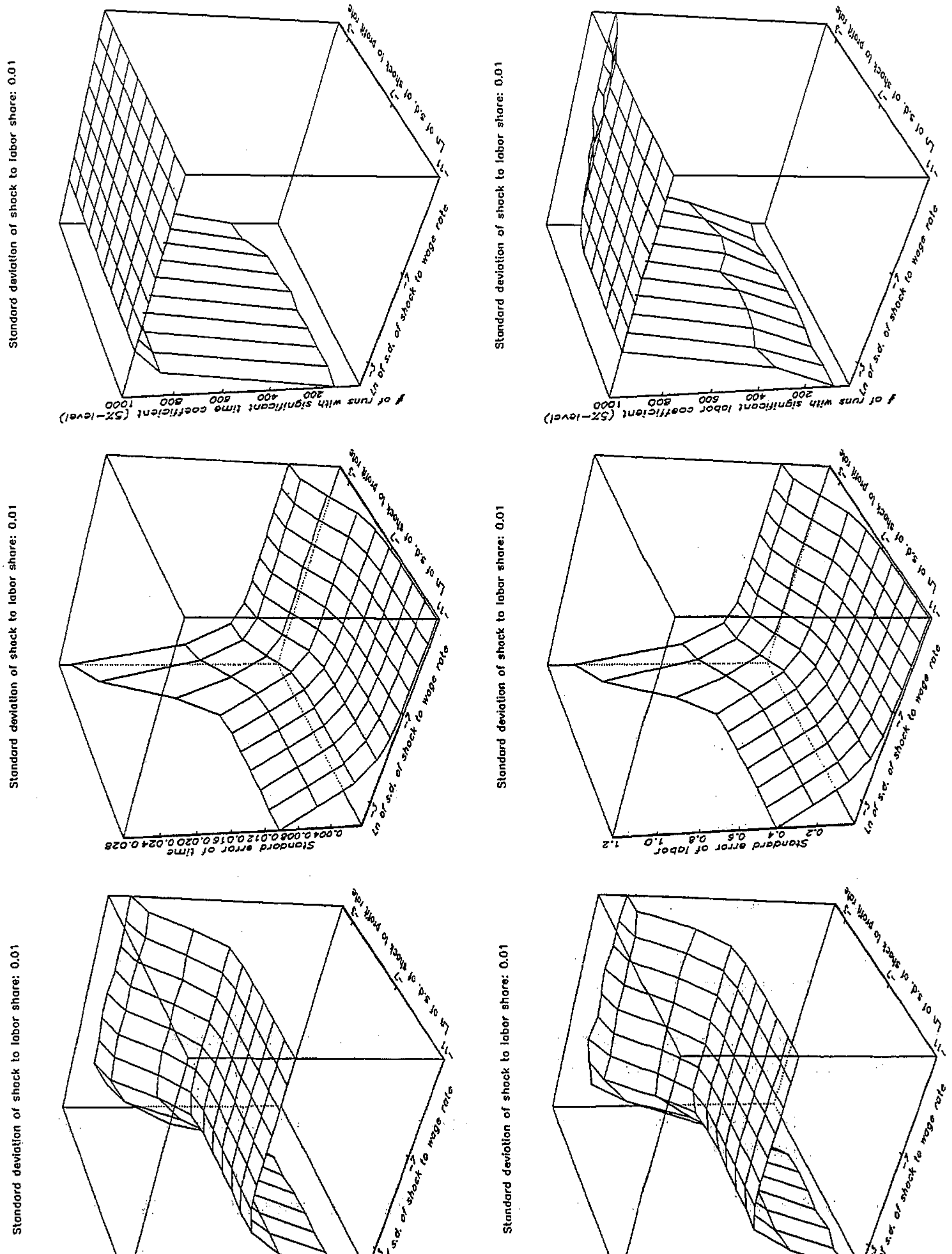
Simon, H. A. (1979) On parsimonious explanations of production relations, *Scandinavian Journal of Economics*, 81, pp. 459–474.

Solow, R. (1966) Review of capital and growth, *American Economic Review*, 56, pp.1257–1260.

Sylos Labini, P. (1995) Why the interpretation of the Cobb–Douglas production function must be radically changed, *Structural Change and Economic Dynamics*, 6, pp. 485–504.

Appendix


Standard deviation of shock to labor share: 0.01
# of runs with significant capital coefficient (5%-level)
Standard error of capital
Coefficient of capital
# of runs rejecting CRS (5%-level)
Durbin-Watson
R-square
Ln of s.d. of shock to wage rate
Ln of s.d. of shock to profit rate


Standard deviation of shock to labor share: 0.01
# of runs with significant time coefficient (5%-level)
Ln of s.d. of shock to wage rate
Ln of s.d. of shock to profit rate
# of runs with significant labor coefficient (5%-level)
Standard error of time
Standard error of labor
Coefficient of time
Coefficient of labor