Gracia, Eduard

Working Paper

Uncertainty and capacity constraints: reconsidering the aggregate production function

Economics Discussion Papers, No. 2011-9

Provided in Cooperation with:
Kiel Institute for the World Economy – Leibniz Center for Research on Global Economic Challenges

Suggested Citation: Gracia, Eduard (2011) : Uncertainty and capacity constraints: reconsidering the aggregate production function, Economics Discussion Papers, No. 2011-9, Kiel Institute for the World Economy (IfW), Kiel

This Version is available at:
https://hdl.handle.net/10419/45492

Economics
The Open-Access, Open-Assessment E-Journal

Discussion Paper
No. 2011-9 | May 2, 2011 | http://www.economics-ejournal.org/economics/discussionpapers/2011-9

Uncertainty and Capacity Constraints: Reconsidering the Aggregate Production Function

Eduard Gracia
University of Barcelona

Abstract The Cobb-Douglas function is today one of the most widely-adopted assumptions in economic modeling, yet both its theoretical and empirical basis have long been under question. The purpose of this paper is to build an alternative production function on neoclassical microfoundations to address these issues, and then test it empirically.
An analysis of annual U.S. data from 1949 to 2008 suggest the model explains nearly 85 percent of GDP fluctuations, and a nonnested model comparison test concludes that it is empirically more robust than the Cobb-Douglas. Furthermore, both contemporary and lagged aggregate capital are rejected as explanatory variables. This lends support to the old "Cambridge Critique", which sustained that using valueweighted capital aggregates to explain production simply made no sense, and also strengthens the model in this paper for, unlike the Cobb-Douglas, it does not model installed capacity as aggregate capital, but as a sunk cost generating economic rents.
Taken at face value, these results not only pose a question on any macroeconomic model assuming a Cobb-Douglas function but also point towards an alternative interpretation of phenomena such as the way monetary policy impacts productivity.

JEL E22, E23

Keywords Production function; Cobb-Douglas; capital controversy

Correspondence Eduard Gracia, 39 Palace Court, London W2 4LS, United Kingdom; e-mail: eduard.gracia@booz.com

UNCERTAINTY AND CAPACITY CONSTRAINTS:

RECONSIDERING THE AGGREGATE PRODUCTION FUNCTION [1]

Eduard Gracia

eduard.gracia@booz.com

1. INTRODUCTION

It would be difficult to find a more widely adopted theoretical device than the Cobb-Douglas production function. Virtually all of today's mainstream Dynamic Stochastic General Equilibrium (DSGE) models assume their production functions to follow a variant of the equation introduced by Cobb & Douglas (1928), to the point of being often referred to as "the" neoclassical production function, as if it were a fundamental postulate without which a model could not be deemed to be "neoclassical".

Yet, in truth, not only is the Cobb-Douglas just one of many possible functions compatible with the neoclassical canon, but its very acceptability as "neoclassical" was subject to intense controversy from the start. As Paul Douglas recalled nearly forty years thereafter (Douglas 1967), the seminal 1928 paper met with "the most caustic criticism" from the leading neoclassical econometricians of the time (e.g. Ragnar Frisch). Nevertheless, in the decades following World War II the Cobb-Douglas function gradually gained acceptance in the neoclassical mainstream – even if only, as Solow (1966) would later put it, as "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 else better comes along".

[1] A very special thank you to Bernarda Zamora for providing numerous and extremely helpful comments to this paper

The debate did not end up there. Starting with Robinson (1953-54), it morphed into the so-called "Cambridge Capital Theory Controversy" which pitted the Cambridge, England school (Joan Robinson, Piero Sraffa…) against the Cambridge, Massachusetts one (Paul Samuelson, Robert Solow…) until the 1970s – see Cohen & Harcourt (2003) for an overview. At the risk of grossly simplifying a very complex topic, we could say the Cambridge (England) school claimed that, if we define capital as a price-weighted aggregate of heterogeneous goods, it makes no sense to model its return as a marginal product for, in production processes requiring time, the value of capital is determined by that rate of return acting as a discount rate – which makes the reasoning circular. The theoretical debate could actually be said to have settled with a Cambridge (England) victory when Samuelson (1966) acknowledged the validity of their objections and their impact on the economic interpretation of the role of capital.

This conclusion was reinforced by the so-called "aggregation" literature which developed at about the same time around the question of whether, and under what conditions, it would be possible to aggregate the individual functions of a set of profit-maximizing producers into a collective one dependent on total input quantitites (see Felipe & Fisher 2003 for an overview). Once again at the risk of oversimplification, it is probably fair to state that these authors (e.g. Fisher 1965, 1968 & 1969 or Gorman 1968) followed the principle, first proposed by May (1947), that an economic production function represents not just an aggregation of individual technical production functions but of their optimal production frontiers i.e. is subject to the additional constraint of optimal resource allocation by the individual producers. Unsurprisingly, none of the authors found any problem with the aggregation if only one, homogeneous input was considered; yet they also found that, with multiple

inputs including fixed as well as variable factors, only under quite restrictive, potentially unrealistic assumptions could an aggregate function be built.

Despite all these theoretical objections, and after falling somewhat out of favor in the 1970s and early 80s, the Cobb-Douglas function experienced a strong revival as the so-called "new classical" approach to macroeconomics became mainstream. One powerful reason for this is probably that, as Solow (1957) would put it, "as long as we insist on practicing macro-economics, we shall need aggregate relationships" – and, in truth, the opponents of the Cobb-Douglas function never managed to put forward a similarly-usable function that would overcome their own objections. The other, perhaps even more important reason, is the widespread perception that the Cobb-Douglas function, whatever its theoretical downsides, is empirically robust.

But, is it, really? True, if one estimates U.S. GDP growth as a function of capital and labor growth rates in an OLS regression without a time trend (thus following the approach in Cobb and Douglas' 1928 paper[2]), the results tends to support the original finding that the regression parameters of the two input variables approximate their shares of output. Yet we know since Abramovitz (1956) and Solow (1957) that a time trend is even more important a component of output growth than input volumes themselves – and, when one includes such a trend in the mix (as in Lucas 1970, Romer 1987, Klette & Griliches 1996, Griliches & Mairesse 1998 or Felipe & Adams 2005), the regression coefficient associated to aggregate capital usually turns out to be *negative* (albeit with low statistical significance), which is counterintuitive at best. DGSE models usually sidestep this issue by calibrating *a priori*, instead of estimating empirically, the coefficients of capital and labor; yet, while this may be a valid

[2] Cobb & Douglas (1928) performed a regression on annual values, as opposed to annual increments, of GDP against labour and capital. Hence, the independent regression parameter captured a constant GDP level but not a constant growth trend, as would be the case working on annual increments

approach for other purposes, it does not help us to validate the production function, since it assumes away what it should actually test. This not only suggests there may be a significant issue in the Cobb-Douglas specification, but also lends support to the objections of both the Aggregation authors and the old Cambridge (England) school[3].

Furthermore, the fact that the regression without a time trend displays such good results should not be taken as strongly positive evidence because (as first pointed out by Phelps Brown 1957 and then by Simon & Levy 1963, Fisher 1971, Shaikh 1974 & 1980, Samuelson 1979, Felipe & Fisher 2003 or Felipe & McCombie 2005 & 2009) such a specification is formally identical to the Fundamental Growth Accounting Identity. Thus, as long as the factor shares of output remain relatively stable over time (as is indeed the case), the empirical fit is bound to be good, simply because it turns into a test of the accounting identity itself, not of the underlying production function.

All this suggests that there is a case for re-examining the problem of the aggregate production function. As the Aggregation literature made clear, only under certain restrictive conditions is it possible to devise a general expression for the macroeconomic production function. If there are concerns regarding the empirical fit (and thus also the underlying assumptions) of the Cobb-Douglas function, it is fair to ask whether there is any other set of assumptions which might still produce a usable functional form but would perform better against empirical evidence.

This paper puts forward the hypothesis that observed GDP fluctuations are better modeled by regarding capital, both physical and human, as a "sunk cost" (whose returns are thus economic rents) than as a variable input whose reward is its marginal

[3] One cannot help thinking that it is just too bad the advocates of the Cambridge (England) school presented their case nearly always on theoretical grounds, when a simple empirical regression test would probably have been more effective to discredit their opponents' arguments

product, as the basic Cobb-Douglas function assumes. The rationale is that capital investments require a lead time to put into production and so, in an uncertain world, the first response to random changes in the composition of demand will be subject to existing capacity constraints. This is by no means a new idea: the concept was already central to the theory of capital put forward by the so-called Austrian school (Böhm-Bawerk, Hayek, Mises, Schumpeter) in the early 20th Century, and has subsequently been repeatedly invoked in mainstream literature as a major explanatory variable for productivity fluctuations. Thus, for example, Jorgenson & Griliches (1967) put forward capacity utilization as a major explanatory factor for Solow's Residual; Kydland & Prescott (1982) resorted to the "time to build" constraint as a central propagation mechanism in their model of real business cycles; Basu & Fernald (2000) highlighted variable input utilization as a likely major cause of productivity fluctuations; and most recently Hansen & Prescott (2005) developed a DSGE model where capacity constraints allow to mimic the asymmetry of actual business cycles.

Yet each one of these models has been conceived as a refinement of the basic Cobb-Douglas function and is therefore subject to the same methodological objections. This paper will, conversely, start from the model's micro-foundations and make it explicit where the underlying assumptions diverge from the Cobb-Douglas framework. Importantly, the explicit focus on the short term simplifies the aggregation conditions dramatically, for now only inputs that are variable in the short run need to be included in the production function. Since many of the inputs that might appear as variable for individual companies are actually fixed at an aggregate level (e.g. the price of electricity is the rent generated by fixed costs and assets devoted to generation and distribution activities, the education premium is the rent generated by the time and money an individual invested in education, etc.), one could argue that, in a closed

economy, only one major input would truly be variable in the short run: labor time. Under these conditions, value-weighted factor aggregates are no longer necessary, and we are left with a single, homogeneous input, which allows to build an aggregate production function that is both compliant with the Aggregation literature conditions and exempt from the Cambridge Critique.

This, we should highlight again, does by no means negate the role of installed capital in the production process. At the time the investment was made, of course, the investor expected it to yield at least the market return; yet in an uncertain world these expectations may not be fulfilled, so that past costs become irrelevant after every new random shock, and thus all that matter are the capacity constraints they impose on new production. Hence, the return on fixed capital (namely, the interest rate) should not be modeled as a marginal product on a historical investment, but as an economic rent... For economic rents can indeed act as an indicator of capacity utilization: the higher the spare capacity, the lower the ability of asset owners to charge a rent.

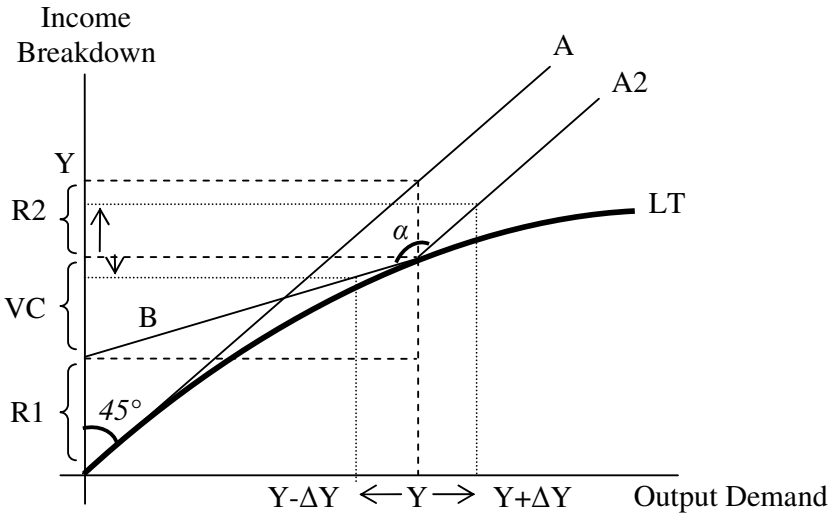
In sum, the purpose of this paper is to *build a model of the short-term aggregate production function, as determined by labor time and capital rents as explanatory variables*, and then *test it empirically to check how well it fits the data, as well as whether it might also be a better specification than the Cobb-Douglas function*.

The structure of this paper is as follows. Section 2 provides an intuitive rationale for the alternative model, which is then developed analytically in Section 3 – readers interested only in the analytical reasoning may want to jump directly to Section 3. Section 4 describes the empirical strategy and presents the statistical test results for both the proposed model itself and for the standard Cobb-Douglas function. Finally, Section 5 ends with a summary of findings and conclusions.

2. MODEL RATIONALE

Imagine an economy composed of many production units, each one devoted to transforming inputs into a given set of outputs. For every given output volume, there are multiple productive processes or "techniques" available, each one requiring a given fixed investment in plant capacity in addition to a variable cost per unit produced. We assume each one of these producers is rationally aiming to select the output volume and productive process that maximizes real profit (i.e. the difference between output value and input costs, measured in output units). We also assume that the optimal technology curve that results from their selecting the most profitable technique at each production level displays economies of scale i.e. that, given an increase in production volume, there is always a technique that would allow to reduce the overall cost per unit (always expressed in terms of output units) and therefore increase the real profit. This is represented graphically by curve 'LT' in Figure 1.

Figure 1: Long-term vs. short-term production functions



In this diagram, both the horizontal and vertical axes represent real output, whereas every line in the quadrant represents the cost structure of a given productive process

or technique. Therefore, if we select an output demand level on the horizontal axis, its projection on the vertical axis according to the curve representing a given technique indicates how much of its output value would correspond to fixed cost ('R1') vs. variable cost ('VC') vs. "pure" profit ('R2'). For example, curve 'LT´ represents the optimal long-term production cost curve i.e. the lowest production cost possible for every given output level, regardless of how long it would take to deploy the associated production process. Every point along this curve (say, point 'Y') is associated to a given optimal production technique (for example 'B') requiring a fixed upfront cost ('R1') plus a certain variable cost per unit (represented by the slope of 'B'), up to a plant capacity equal to 'Y'. Conversely, the straight line 'A' represents the cost profile of a production technique with constant variable costs per unit, no fixed costs and no barriers of entry, which would of course result in the unit price equating the marginal (i.e. variable) cost (hence it forms a 45° angle respective to the axes). Evidently, technique 'A' has a steeper slope than 'B' because its variable costs are higher, and is more inefficient for a given level of production 'Y' because the overall cost per unit for technique 'B', including both fixed and variable costs, is lower than that of technique 'A' – the difference being of course the "pure" profit R2.

In economics, prices are conventionally broken down into marginal costs (i.e. the incremental cost of producing the last output unit) and economic rents (that is, the difference between marginal cost and actual price). Evidently no rational, profit-maximizing producer would be willing to sell for less than the marginal cost (i.e. in the case of a production level 'Y', for less than the length of segment 'VC'). Conversely, from a short-term perspective, fixed cost investments constitute a "sunk cost" whose historical size is simply irrelevant for the maximization of future profit,

and whose return therefore constitutes an economic rent (i.e. in this case the sum of segments ‘R1’ and ‘R2’) since the concept of “marginal cost” does not apply to it.

Sure enough, if a production unit were considering whether to commit an upfront investment or not, it would require the present value of the expected return to equal or exceed the upfront cost. Hence, in a deterministic world it would make sense to model those future rents as equating marginal costs with marginal returns for that historical investment amount; yet, in an uncertain world, as soon as the conditions change so do those rents, after which no relationship may exist between them and the sunk costs.

The assumption behind the long-term curve ‘LT’ is that, when planning for the long run, producers can jump from one technique to the next as their output volumes change, choosing for every level of production the technique with the lowest cost. Conversely, when unexpected shocks hit demand, it is not possible to do this in the short run, for the upfront investments to expand capacity and deploy a more efficient production process cannot be deployed instantly, nor can installed capacity be easily divested, even if such a thing is possible at all. Hence, if demand, for example, drops unexpectedly by a magnitude ‘ΔY’ (i.e. down to ‘$Y-\Delta Y$’), the producer will use the same technique to produce at less than full capacity, whereas, if demand increases instead by ‘ΔY’ (i.e. up to ‘$Y+\Delta Y$’), the producer will have to resort to the less efficient technique A to produce the supplementary units required. This means that, although the long-term cost function is concave (i.e. has positive returns to scale), the short-term one is convex (that is, displays diseconomies of scale), for, given a planned output level ‘Y’, actual production follows segment B when it falls below ‘Y’ but segment ‘A2’ (parallel to ‘A’) when it raises above ‘Y’.

Let's imagine now that we have many industries in an economy, each producing a different set of goods and services but all subject to a cost function with the same characteristics. Then, if the structure of aggregate demand changes unexpectedly, so that demand for one product increases at the expense of another while the total consumer budget stays the same, the costs of those industries whose demand dropped will go down comparatively less than the costs of the industries with higher demand will go up. This will therefore result in an aggregate loss of productive efficiency:

1. The larger the variability of demand (i.e. 'ΔY' in the diagram) and/or
2. The smaller the angle 'α' between segments 'B' and 'A2'

In turn, since any increment in the share of economic rents ('R1+R2') over the total revenue 'Y' results in squeezing segment 'VC' and hence flattening segment 'B' and closing angle 'α', we may say that, under demand uncertainty, the higher the ratio $\frac{R1+R2}{Y}$, the lower the overall productive efficiency in the short run.

As a result, under uncertainty, the link between economic rents and productivity in the short run is exactly opposite to the long run. In the long run what matters is curve 'LT' and , since segment 'B' represents the tangent to this curve, the larger the ratio $\frac{R1+R2}{Y}$, the flatter (i.e. the more concave) the curve will be at that point – i.e. the higher its economies of scale. Conversely, in the short run, the larger this ratio (i.e. the more concave the 'LT' curve), the more convex the angle 'α' will be. If the world were deterministic, only the long-term function would matter, since rational agents would plan only at one point in the beginning of time with a view in the long run, and never have to revise their expectations again. Yet, in a stochastic world, circumstances

change continuously, and as the agents adapt every time to the new conditions it is the short-term function that determines the initial response.

In sum, the model predicts that the short-term production is a function of *variable* inputs only (as opposed to both fixed and variable), combined with the percentage of economic rents over total output (which determines the angle 'α' and therefore the degree of convexity) as well as the variability of demand composition (which in essence represents the average shock 'ΔY'). Since most apparently variable inputs translate at an aggregate level into fixed capital plus labor (e.g. electricity costs may seem variable to the consumer, yet they correspond to generation and distribution fixed capital investments plus some labor input), and even labor costs incorporate a certain portion of economic rents as a return for investments made at an earlier point in time (e.g. the returns for an investment in education), it could be argued that, at a macroeconomic level the only truly "variable" input is labor time.

Importantly, since the shape of the curve is directly dependent on the ratio $\frac{R1+R2}{Y}$, any policy aimed at changing the weight of economic rents over production would actually have the power to change the shape of the production function, and thus, by implication, to manipulate the rate of productivity growth. The implications are substantial, because the weight of economic rents can actually be modified through public intervention, be it by the central bank (e.g. through interest rates, which transfer income from borrowers to lenders) or by the government (e.g. through taxes and subsidies, or through legal monopolies and other constraints on competition). This means, for example, that, even if the market were perfectly rational and efficient, so that money supply were absolutely neutral from a demand perspective, the central bank's discretionary control on monetary supply would have the power to modify the

shape of the production function (and hence the rate of productivity growth), accelerating it with low real interest rates and slowing it down with higher ones. An important prediction of this model is, therefore, that there is a direct, positive correlation between cheap, abundant credit and observed productivity.

3. ANALYTICAL FRAMEWORK

3.1. The individual production function

Consider an economic system with $i = 1...n$ types of goods and services and $j = 1...m$ production units, each one of which transforms, at any given point in time *t*, a certain set of input quantities of goods and services $\{x_{1,j,t}...x_{n,j,t}\}$ into another set of output quantities $\{y_{1,j,t}...y_{n,j,t}\}$. We represent this transformation of one set of goods into another as a *technical* production function $f_{j,t}\{..\}$ such that:

$$f_{j,t}\{x_{1,j,t}...x_{n,j,t}\} \equiv \{y_{1,j,t}...y_{n,j,t}\} \quad (1)$$

Note that the output of this function $f_{j,t}\{..\}$ is expressed as a set of quantities of different products $\{y_{1,j,t}...y_{n,j,t}\}$, as opposed to a single scalar, since the various products may not always be totally independent from each other (e.g. some outputs may be by-products or co-products of others).

We now define the real output flow ('$Y_{j,t}$') of a production unit *j* at a given point in time *t* as the sum of all the component output quantities $\{y_{1,j,t}...y_{n,j,t}\}$ produced multiplied by their market prices $\{p_{1,t}...p_{n,t}\}$ and expressed in terms of a given basket

of consumable goods and services selected as *numéraire* (and whose unit price we conventionally represent as 'P_t') i.e.:

$$Y_{j,t} \equiv \frac{1}{P_t} \sum_{i=1}^{n} p_{i,t} y_{i,j,t} \tag{2}$$

On this basis, we define the *economic* production function $F_{j,t}\{x_{1,j,t}...x_{n,j,t}\}$ as the transformation leading from the input quantities to the real output flow i.e.:

$$Y_{j,t} \equiv F_{j,t}\{x_{1,j,t}...x_{n,j,t}\} \equiv \frac{1}{P_t} \sum_{i=1}^{n} p_{i,t} y_{i,j,t} \tag{3}$$

Note that this *economic* production function $F_{j,t}\{...\}$ differs from the *technical* one $f_{j,t}\{...\}$ in its using output prices to weight the various (and otherwise heterogeneous) output quantities into a scalar metric, the real output flow. These functions (economic and technical) would be the same, of course, if the production unit had a single good as output and, in addition, this good were chosen as the *numéraire*; otherwise, however, they can be very different. Since the objective of this exercise is to develop an expression linking a set of input units to a measure of GDP (i.e. our variable '$Y_{j,t}$'), it is the *economic* production function that this paper will focus on.

We now define the profit ('$\Pi_{j,t}$') of the production unit *j* as the difference between its output flow at current prices and its total production cost ('$C_{j,t}$') i.e. $\Pi_{j,t} \equiv P_t Y_{j,t} - C_{j,t}$, where $C_{j,t}$ is a function of the input quantities $\{x_{1,j,t}...x_{n,j,t}\}$. When this profit is expressed in terms of the basket of goods and services we have selected as the real output *numéraire*, we refer to it as "real" profit ('$\Pi^*_{j,t}$') i.e.:

$$\Pi^*_{j,t} \equiv Y_{j,t} - C^*_{j,t} \qquad (4)$$

Where obviously $C^*_{j,t} \equiv \frac{C_{j,t}}{P_{j,t}}$.

At this point we introduce four assumptions that will be central to this section:

1. *Profit maximization:*

   Producers select their output quantities $Y_{j,t}$ as well as their corresponding demand for inputs to maximize their real profit flow $\Pi^*_{j,t} \equiv Y_{j,t} - C^*_{j,t}$

2. *Existence of an optimum:*

   There is at least one finite maximum point for every real profit function $\Pi^*_{j,t}$

3. *Continuity and differentiability:*

   The real profit function is continuous and partially-differentiable respective to all input and output quantities

4. *Homogeneous production function:*

   The economic production function $Y_{j,t} \equiv F_{j,t}\left\{x_{1,j,t}....x_{n,j,t}\right\}$ is such that, given a scalar $a \in \Re$, the expression $F_{j,t}\left\{ax_{1,j,t}....ax_{n,t}\right\} = a^{h_{j,t}} F_{t,j}\left\{x_{1,j,t}....x_{n,j,t}\right\}$ holds (where the degree of homogeneity is independent of $\left\{x_{1,j,t}....x_{n,j,t}\right\}$)

For convenience, we also introduce the following two definitions:

- We define the *combined input* (‘ $X_{j,t}$ ’) as:

$$X_{j,t} \equiv \frac{1}{c^*_{1,j,t}} \sum_{i=1}^{n} c^*_{i,j,t} x_{i,j,t} \tag{5}$$

Where $\{c^*_{1,j,t} ... c^*_{n,j,t}\}$ represent the real marginal costs i.e. $c^*_{i,t} \equiv \frac{\partial C^*_t}{\partial x_{i,t}}$, and where $c^*_{1,j,t}$ represents the marginal cost of an input '1' selected as aggregation unit (and which, incidentally, does not need to coincide necessarily with the basket of goods we selected as output *numéraire*).

- We also define the *rent ratio* or *rent coefficient* ('$\rho_{j,t}$') as the output value premium over real marginal costs per unit of input i.e.:

$$\rho_{j,t} \equiv \frac{Y_{j,t}}{\sum_{i=1}^{n} c^*_{i,j,t} x_{i,j,t}} - 1 \equiv \frac{Y_{j,t}}{c^*_{1,j,t} X_{j,t}} - 1 \tag{6}$$

Following Euler's homogeneous functions theorem, Assumption 4 implies that:

$$h_{j,t} = \sum_{i=1}^{n} \frac{\partial Y_{j,t}}{\partial x_{i,j,t}} \frac{x_{i,j,t}}{Y_{j,t}} \tag{7}$$

Since producers fine-tune their demand for inputs in order to maximize their real profit (Assumption 1), if a finite maximum profit point exists (Assumption 2), and if the real output and cost functions are continuous and partially-differentiable (Assumption 3), then the real profit maximum must be such that, for every input $x_{i,j,t} \in \{x_{1,j,t} ... x_{n,j,t}\}$, the following expression holds:

$$\frac{\partial \Pi^*_{j,t}}{\partial x_{i,j,t}} = \frac{\partial Y_{j,t}}{\partial x_{i,j,t}} - \frac{\partial C^*_{j,t}}{\partial x_{i,j,t}} = 0$$

$$\frac{\partial Y_{j,t}}{\partial x_{i,j,t}} = \frac{\partial C^*_{j,t}}{\partial x_{i,j,t}} = c^*_{i,j,t} \tag{8}$$

Which, replacing into the definition of rent ratio, yields:

$$\rho_{j,t} \equiv \frac{Y_{j,t}}{c^*_{1,j,t} X_{j,t}} - 1$$

$$X_{j,t} \equiv \frac{1}{c^*_{1,j,t}} \sum_{i=1}^{n} c^*_{i,j,t} x_{i,j,t}$$

$$\frac{1}{1+\rho_{j,t}} \equiv c^*_{1,j,t} \frac{X_{j,t}}{Y_{j,t}} \equiv \sum_{i=1}^{n} c^*_{i,j,t} \frac{x_{i,j,t}}{Y_{j,t}} = \sum_{i=1}^{n} \frac{\partial Y_{j,t}}{\partial x_{i,j,t}} \frac{x_{i,j,t}}{Y_{j,t}} \tag{9}$$

Thus, by comparing expressions (7) and (9), we can conclude that:

$$h_{j,t} = \frac{1}{1+\rho_{j,t}} \tag{10}$$

This implies that, since $h_{j,t}$ is independent of the input, so is $\rho_{j,t}$. Hence:

$$\frac{1}{1+\rho_{j,t}} \equiv c^*_{1,j,t} \frac{X_{j,t}}{Y_{j,t}} = \frac{\partial Y_{j,t}}{\partial x_{1,j,t}} \frac{X_{j,t}}{Y_{j,t}} = \frac{\partial Y_{j,t}}{\partial X_{j,t}} \frac{X_{j,t}}{Y_{j,t}}$$

$$\frac{\partial Y_{j,t}}{Y_{j,t}} = \frac{1}{1+\rho_{j,t}} \frac{\partial X_{j,t}}{X_{j,t}}$$

And, integrating both sides, we obtain:

$$Y_{j,t} = A_{j,t} X_{j,t}^{\frac{1}{1+\rho_{j,t}}} \tag{11}$$

Where both the integration coefficient $A_{j,t}$ (which we will refer to as "productivity coefficient") and the rent ratio $\rho_{j,t}$, are functions independent of $X_{j,t}$ and $Y_{j,t}$.

This is an important expression, for it represents the analytical equivalent of the broken line B-A2 in Figure 1 (only, now under the homogeneity condition imposed in Assumption 4, which the broken line B-A2 would obviously not fulfill). As expected, the curve will be convex (i.e. display diseconomies of scale) if rents are positive, whereas if they are negative (which is not inconceivable in a stochastic world where actual returns may be lower than expected[4]) the shape will be concave[5].

3.2. The path of output growth

The purpose of this subsection is to determine the path of output growth over time under the conditions established in Subsection 3.1.

Note that, contrary to a substantial portion of the literature, the model developed here operates under continuous, not discrete, time. Philosophically, one might argue that continuous time is a more realistic assumption, as real time is after all a continuous variable, and it is in real time, not in annual stints, that economic agents make their decisions. More pragmatically, though, the advantage of this assumption for us resides in its simplicity: although discrete time may seem more intuitive, as soon as the development gets a bit complex the discrete-time form just becomes analytically unmanageable. In this paper, for example, it is easy to see that the development

[4] To take but a simple example: when inflation is higher than expected, the rents generated by fixed income assets such as rent apartments or bank deposits may well become negative

[5] In some of the initial working paper discussions about this model, this conclusion caused some discomfort, as it was perceived as contradictory with the standard development as it appears in a number of papers (e.g. Basu & Fernald 2000). As shown in Appendix 1, however, there is no contradiction and, given the same assumptions, one can easily be derived from the other

leading to expression (14) in Subsection 3.2. below (which is already complex enough to be relegated to Appendix 2) would simply be impossible if, instead of applying Itô's lemma (which is no more than the stochastic version of the chain rule) we had to resort to Taylor's rule, which represents its equivalent in discrete analysis.

We now introduce the following three additional assumptions:

5. *Neutral price breakdown:*

   The way prices are broken down between marginal costs and economic rents (as represented by the rent ratio $\rho_{j,t}$) has no impact, other things being equal, on output demand i.e. $\forall \rho_{j,t} \in \Re \quad ; \quad \frac{\partial Y_{j,t}}{\partial \rho_{j,t}} = 0$

6. *Random walk perturbation on technology shocks:*

   Technology shocks follow an Itô stochastic diffusion process subject to a Wiener perturbation (that is, a linear, continuous, normally-distributed random-walk process otherwise known as "Brownian motion") i.e.:

$$\frac{dA_{j,t}}{A_{j,t}} = \gamma_{j,t} dt + s_{j,t} dW_{j,t} \qquad (12)$$

   Where $A_{j,t}$ represents the productivity coefficient in $Y_{j,t} = A_{j,t} X_{j,t}^{\frac{1}{1+\rho_{j,t}}}$, the symbols $\gamma_{j,t}$ and $s_{j,t}$ represent functions whose value is known at time *t*, and W_t is a Wiener process i.e. a stochastic process such that $W_0 \equiv 0$ and

$dW_{j,t} \equiv \omega_{j,t}\sqrt{dt}$, where $\omega_{j,t}$ is a serially-uncorrelated, normally-distributed standardized white noise (i.e. analytically $\omega_{j,t}$ ~ N$[0,1]$).

7. *Random walk perturbation on output demand:*

The growth rate of output demand also follows an Itô stochastic process subject to a Wiener perturbation, i.e.:

$$\frac{dY_{j,t}}{Y_{j,t}} = \mathrm{E}_t\left[\frac{dY_{j,t}}{Y_{j,t}}\right] + s_{j,t}dW_{j,t} + \sigma_{j,t}dV_{j,t} \qquad (13)$$

Where $\sigma_{j,t}$ is a function whose value is known at time *t*, the operator $\mathrm{E}_t[\circ]$ indicates the expected value according to the information available at instant *t*, and $V_{j,t}$ is a Wiener process i.e. a serially-uncorrelated, normally-distributed standardized white noise just as $W_{j,t}$. For simplicity, we will also assume that $V_{j,t}$ and $W_{j,t}$ are independent random processes so that $dV_{j,t}dW_{j,t} = 0$.

Assumption 5 is fairly intuitive: demand may be impacted by changes in output prices, but what portion of that price is devoted to paying for marginal costs as opposed to economic rents has no relevance for the consumers, and should therefore have no impact on their demand – which, it should be noted, we are implicitly assuming to always equal supply. Note that the condition $\forall\rho_{j,t} \in \Re$ implies that price composition is neutral throughout the whole real domain i.e. that $\frac{\partial Y_{j,t}}{\partial \rho_{j,t}} = 0$ holds for

every possible value of $\rho_{j,t}$, and thus means that $\frac{\partial^2 Y_{j,t}}{\partial \rho_{j,t}^2} = 0$ and, in general, $\frac{\partial^2 Y_{j,t}}{\partial \rho_{j,t} \partial V_{j,t}} = 0$ (where $V_{j,t}$ represents any other variable in the model).

The other two assumptions introduce a stochastic element into the model. The assumption that technology progress follow a random walk (Assumption 6) is consistent with the standard literature (e.g. in DSGE models), except perhaps for its being expressed in continuous instead of discrete time. Similarly, Assumption 7, which postulates that output growth perturbations are the sum of technology and non-technology shocks, should also be fairly intuitive (except perhaps for the fact that, for the sake of analytical simplicity, we assume these two sources of uncertainty to be independent from each other), although the second source of uncertainty is sometimes assumed to be nil in standard DGSE models.

Under these conditions, if we differentiate expression (11) (see the detailed derivation in Appendix 2) we obtain that:

$$\frac{dY_{j,t}}{Y_{j,t}} = \gamma_{j,t} dt + s_{j,t} dW_{j,t} - \frac{\sigma_{j,t}^2}{2} \rho_{j,t} dt + \frac{1}{1+\rho_{j,t}} \frac{dX_{j,t}}{X_{j,t}} \qquad (14)$$

There are two key takeaways from this expression (14). One is obvious: the output fluctuations that are not matched by technology change (as represented by $dW_{j,t}$) must be reflected on input change (as represented by $dX_{j,t}$). The other, however, is more important: the higher the non-technology variance ('$\sigma_{j,t}^2$') and/or the rent ratio ('$\rho_{j,t}$'), the slower will output growth be, even under the same levels of input growth and technology progress. This was exactly the conclusion anticipated in Section 2.

3.3. Aggregation

To extend expression (14) to a macroeconomic scale, we still need to aggregate output across all the production units $j = 1...m$.

We define the aggregate output ('$\overline{Y}_t$') as the sum of all individual outputs (i.e. $\overline{Y}_t \equiv \sum_{j=1}^{m} Y_{j,t}$); the combined input ('$\overline{X}_t$') as the sum of inputs weighted by their marginal costs (i.e. $\overline{X}_t \equiv \sum_{j=1}^{m} X_{j,t}$); the average rent coefficient ('$\overline{\rho}_t$') as the ratio $\overline{\rho}_t \equiv \frac{\overline{Y}_t}{c^*_{1,t}\overline{X}_t} - 1$; the relative output share ('$\alpha_{j,t}$') of a production unit *j* as the percentage of its production over the total (i.e. $\alpha_{j,t} \equiv \frac{Y_{j,t}}{\overline{Y}_t}$); and, finally, the combined input share ('$\beta_{j,t}$') of a production unit *j* as the percentage of its combined input consumption over the total (i.e. $\beta_{j,t} \equiv \frac{X_{j,t}}{\overline{X}_t}$).

In this context, we introduce one additional assumption:

8. Single marginal cost for the reference input:

> The marginal cost $c^*_{1,\varepsilon,t}$ of the input '1' selected as aggregation unit is the same for all producers i.e. $c^*_{1,1,t} = ... = c^*_{1,\mu,t} = \overline{c}^*_{1,t}$

This is actually quite a weak assumption, as it only requires to find *one* commodity that displays the same marginal cost for all producers (for example, a currency trading in a highly liquid, competitive market), and then take it as the accounting unit of $X_{i,t}$.

Combining Assumption 8 with the identity $Y_{j,t} \equiv (1+\rho_{j,t}) c^*_{1,j,t} X_{j,t}$ we obtain that:

$$\frac{Y_{j,t}}{(1+\rho_{j,t})X_{j,t}} = \bar{c}^*_{1,t} = \frac{\bar{Y}_t}{(1+\bar{\rho}_t)\bar{X}_t} \tag{15}$$

And also that:

$$1+\rho_{j,t} = (1+\bar{\rho}_t)\frac{Y_{j,t}}{\bar{Y}_t}\frac{\bar{X}_t}{X_{j,t}} = (1+\bar{\rho}_t)\frac{\alpha_{j,t}}{\beta_{j,t}} \tag{16}$$

Hence, if we now replace $\rho_{j,t}$ into (14) we obtain:

$$\frac{dY_{j,t}}{Y_{j,t}} = \left(\gamma_{j,t} + s_{j,t}\frac{dW_{j,t}}{dt} - \frac{\sigma^2_{j,t}}{2}\left((1+\bar{\rho}_t)\frac{\alpha_{j,t}}{\beta_{j,t}} - 1\right)\right)dt + \frac{\bar{Y}_t}{1+\bar{\rho}_t}\frac{dX_{j,t}}{\bar{X}_t Y_{j,t}} \tag{17}$$

If we now multiply both sides of the equation by $\alpha_{j,t}$ and aggregate for all $j = 1...m$:

$$\sum_{j=1}^{m}\frac{dY_{j,t}}{\bar{Y}_t} = \sum_{j=1}^{m}\alpha_{j,t}\left(\gamma_{j,t} + s_{j,t}\frac{dW_{j,t}}{dt} - \frac{\sigma^2_{j,t}}{2}\left((1+\bar{\rho}_t)\frac{\alpha_{j,t}}{\beta_{j,t}} - 1\right)\right)dt + \sum_{j=1}^{m}\frac{1}{1+\bar{\rho}_t}\frac{dX_{j,t}}{\bar{X}_t}$$

$$\frac{d\bar{Y}_t}{\bar{Y}_t} = \sum_{j=1}^{m}\alpha_{j,t}\left[\gamma_{j,t} + s_{j,t}\frac{dW_{j,t}}{dt} - \frac{\sigma^2_{j,t}}{2}\left((1+\bar{\rho}_t)\frac{\alpha_{j,t}}{\beta_{j,t}} - 1\right)\right]dt + \frac{1}{1+\bar{\rho}_t}\frac{d\bar{X}_t}{\bar{X}_t} \tag{18}$$

For convenience, we introduce now the following definitions:

- The coefficient $\delta_{j,t} \equiv \sigma^2_{j,t}\frac{\alpha_{j,t}}{\beta_{j,t}}$ and its aggregate $\bar{\delta}_t \equiv \sum_{j=1}^{m}\alpha_{j,t}\delta_{j,t}$
- The aggregate rate $\bar{\gamma}_t \equiv \sum_{j=1}^{m}\alpha_{j,t}\gamma_{j,t}$

- The composite rate $\Gamma_{j,t} \equiv \gamma_{j,t} + \frac{\sigma_{j,t}^2 - \delta_{j,t}}{2}$ and its aggregate $\overline{\Gamma}_t \equiv \sum_{j=1}^{m} \alpha_{j,t}\Gamma_{j,t}$

- The aggregate technology perturbation $\bar{s}_t d\overline{W}_t \equiv \sum_{j=1}^{m} \alpha_{j,t} s_{j,t} dW_{j,t}$, which obviously follows a Gaussian distribution i.e. $\bar{s}_t d\overline{W}_t \sim \mathrm{N}\left(0, \bar{s}_t^2\right)$ and therefore $d\overline{W}_t \sim \mathrm{N}(0,1)$

- The overall non-technology variance $\bar{\sigma}_t^2 dt \equiv \left(\frac{1}{1+\bar{\rho}_t}\frac{d\bar{X}_t}{\bar{X}_t}\right)^2$ i.e. the equivalent of expression (2.7) in Appendix 2, only calculated here at a macroeconomic level

Applying these definitions, expression (18) can be rewritten as follows:

$$\frac{d\bar{Y}_t}{\bar{Y}_t} = \overline{\Gamma}_t dt + \bar{s}_t d\overline{W} - \frac{\bar{\delta}_t}{2}\bar{\rho}_t dt + \frac{1}{1+\bar{\rho}_t}\frac{d\bar{X}_t}{\bar{X}_t} \tag{19}$$

Which, in turn, can be integrated through Itô's lemma to obtain:

$$\bar{Y}_t = \bar{A}_t \bar{X}_t^{\frac{1}{1+\bar{\rho}_t}} \tag{20}$$

Where the aggregate productivity coefficient $\bar{A}_t$ is a martingale such that:

$$\frac{d\bar{A}_t}{\bar{A}_t} = \left(\overline{\Gamma}_t + \frac{\bar{\sigma}_t^2 - \bar{\delta}_t}{2}\bar{\rho}_t\right)dt + \bar{s}_t d\overline{W}_t \tag{21}$$

Importantly, notice that expression (21) is *NOT* just a weighted aggregate of the individual productivity growth expression (12), as one might expect. This is because the aggregate productivity coefficient $\bar{A}_t$ reflects not just the impact of technology changes but also that of changes in the share of output of different production units whose individual production functions, as depicted graphically in Section 2 and

analytically in Subsection 3.1, are not necessarily linear. It is indeed easy to see that (21) would turn into a simple aggregation of expression (12) if all the non-technology perturbations where zero (so that $\sigma_{1,t} = ... = \sigma_{m,t} = 0$) and/or if all the production functions were linear (i.e. $\rho_{1,t} = ... = \rho_{m,t} = 0$ and thus also $\frac{\alpha_{i,t}}{\beta_{i,t}} = \frac{\alpha_{j,t}}{\beta_{j,t}}$).

The question remaining now, however, is how to measure $\overline{X}_t$: after all, this metric is defined as the sum of the individual producers' input quantities weighted by their own marginal costs, which may be different from one produce to the next. This is, at a simplified level, why Fisher and Gorman asserted that only under very special circumstances could this be transformed into an aggregation of inputs without having to give different weights to the same variable inputs as used by different producers. The definition of "variable input", however, depends on the time horizon of analysis: the shorter it is, the fewer inputs can be regarded as variable. Hence, there must be a period short enough for only one input to be genuinely regarded as variable.

Which one? For an individual production unit, variable cost can be any input it buys externally; but in a closed economy, at an aggregate level, the intermediate products the various units buy and sell from each other cancel out, leaving only a combination of labor time and a stock of capital resulting from investments made (and sunk) at some earlier point in time (be it as physical fixed assets, as human capital or any other form)[6]. Therefore, if the "short term" is defined as short enough, we may assume:

[6] For example, at a microeconomic level utilities constitute variable costs whose consumption is decided literally at the flip of a switch; but at an aggregate level generation capacity depends on fixed investments which allow to produce electricity at virtually zero marginal cost. Similarly, labour time of permanent employees could be regarded as fixed costs, since the company will be paying the same salaries regardless of the actual work donw; but from a macroeconomic viewpoint these rents between producers of labour (i.e. workers) and its consumers (i.e. companies) cancel out and the fact remains that work hours have a personal variable cost in terms of sacrificed leisure time.

9. Labor hours as only variable input within the target time horizon:

The aggregate system is a closed economy where non-weighted labor time constitutes the only variable input within the relevant decision time horizon

According to this assumption, if we represent the total sum of labor hours as $\overline{H}_t$, then obviously expression (20) can be rewritten as:

$$\overline{Y}_t = \overline{A}_t \overline{H}_t^{\frac{1}{1+\overline{\rho}_t}} \qquad (22)$$

Where $\overline{H}_t$ represents a non-weighted sum of work hours of all persons (as opposed to the usual Tornquist labor input aggregate, where input time is weighted by the share of output of each type of worker).

In short: *this model's main prediction is that GDP growth is better explained by taking non-weighted labor time and an estimate of the rent ratio as explanatory variables than by a Cobb-Douglas function depending on value-weighted capital and labor aggregates*. This is the hypothesis that will be tested in Section 4.

3.4. Comparison with the Cobb-Douglas function

For reference, it may now make sense to establish what sort of assumptions would derive a Cobb-Douglas production function within the same analytical framework. This requires introducing two additional assumptions. The first is the following:

10. Market prices equal marginal costs:

The marginal costs of all inputs equal their market prices i.e. $c_{i,j,t} = p_{i,t}$

This assumption implies that:

- The rent ratios are always zero (i.e. $\rho_{i,t} = 0$), as there are no economic rents
- Assumption 9 becomes indefensible, as now all inputs are instantly variable

Under this assumption, the aggregate production function takes the form:

$$\overline{Y}_t = \overline{A}_t \overline{X}_t \tag{23}$$

Where the aggregate productivity coefficient $\overline{A}_t$ is a martingale such that:

$$\frac{d\overline{A}_t}{\overline{A}_t} = \overline{\gamma}_t dt + \overline{s}_t d\overline{W}_t \tag{24}$$

And where $\overline{X}_t$ can now be calculated as a sum of inputs multiplied by their market prices i.e. $\overline{X}_t \equiv \frac{1}{p_{1,t}} \sum_{i=1}^{n} \sum_{j=1}^{m} p_{i,t} x_{i,j,t}$.

The impact of Assumption 10 is therefore substantial: not only does the function become linear (i.e. exhibits constant returns of scale), but also, as a result, the growth rate of the productivity coefficient $\overline{A}_t$ becomes a direct aggregation of the individual expression (12), which, per Assumption 6, is only determined by technology factors.

The second key assumption is related to aggregate output elasticity. Specifically, given the usual Tornquist aggregates for labor ('$\overline{L}_t$') and capital ('$\overline{K}_t$'), it assumes:

11. Constant elasticity of aggregate capital vs. output:

The elasticity $\frac{\partial \overline{Y}_t}{\partial \overline{K}_t} \frac{\overline{K}_t}{\overline{Y}_t} = \alpha_K$ is a constant.

Since, according to expression (23), this is a homogeneous function of degree one (because $\bar{\rho}_t = 0$), we know that $\bar{Y}_t = \alpha_K \frac{\partial \bar{Y}_t}{\partial \bar{K}_t} \bar{K}_t + (1-\alpha_K) \frac{\partial \bar{Y}_t}{\partial \bar{L}_t} \bar{L}_t$, which means that the production function can be rewritten as:

$$\bar{Y}_t = \bar{A}_t \bar{K}_t^{\alpha_K} \bar{L}_t^{\alpha_L} \qquad \text{(where } \alpha_L = 1-\alpha_K\text{)} \qquad (25)$$

Which is of course the constant-returns Cobb-Douglas function.

Note that (as shown in Appendix 1) by dropping Assumption 1 (i.e. profit maximization) this result can be extended to non-constant returns (i.e. $\alpha_K + \alpha_L \neq 1$).

4. EMPIRICAL STRATEGY AND TESTING

4.1. Empirical model

Following the traditional approach to testing macroeconomic production functions empirically, this paper resorts primarily to an OLS regression against a GDP growth series. This requires rearranging expression (22) in logarithmic form:

$$\ln \bar{Y}_t = \ln \bar{A}_t + \frac{1}{1+\bar{\rho}_t} \ln \bar{H}_t \qquad (26)$$

Where, since we know from expression (21) that $\bar{A}_t$ follows an Itô diffusion process:

$$d \ln \bar{A}_t = \frac{d\bar{A}_t}{\bar{A}_t} - \frac{1}{2}\left(\frac{d\bar{A}_t}{\bar{A}_t}\right)^2 = \frac{d\bar{A}_t}{\bar{A}_t} - \frac{\bar{s}_t^2}{2} dt \qquad (27)$$

And therefore we can rewrite expression (26) as:

$$\ln \overline{Y}_t = \int \left(\overline{\Gamma}_t + \frac{\overline{\sigma}_t^2 - \overline{\delta}_t}{2} \overline{\rho}_t - \frac{\overline{s}_t^2}{2} \right) dt + \int \overline{s}_t d\overline{W}_t + \frac{\ln \overline{H}_t}{1 + \overline{\rho}_t} \tag{28}$$

Which, taking discrete time increments, becomes:

$$\Delta \ln \overline{Y}_t = \Delta \int \left(\overline{\Gamma}_t + \frac{\overline{\sigma}_t^2 - \overline{\delta}_t}{2} \overline{\rho}_t - \frac{\overline{s}_t^2}{2} \right) dt + \Delta \int \overline{s}_t d\overline{W}_t + \Delta \frac{\ln \overline{H}_t}{1 + \overline{\rho}_t} \tag{29}$$

To formulate the model in linear form we introduce two ancillary assumptions:

a) The coefficients $\overline{\Gamma}$, $\overline{s}$, $\overline{\sigma}$ and $\overline{\delta}$ are constant

b) Time increments are short enough to allow the approximation

$$\Delta \int \overline{\rho}_t dt \approx \overline{\rho}_t \Delta t \qquad \text{and} \qquad \Delta \frac{\ln \overline{H}_t}{1 + \overline{\rho}_t} \approx \frac{\Delta \ln \overline{H}_t}{1 + \overline{\rho}_t}$$

Under these assumptions we can rewrite expression (29) as:

$$\Delta \ln \overline{Y}_t = \left(\overline{\Gamma} - \frac{\overline{s}^2}{2} \right) \Delta t + \frac{\overline{\sigma}^2 - \overline{\delta}}{2} \overline{\rho}_t \Delta t + \frac{\Delta \ln \overline{H}_t}{1 + \overline{\rho}_t} + \overline{s} \Delta \overline{W}_t \tag{30}$$

Which, conventionally taking the discrete time increments in the series as unity (i.e. $\Delta t \equiv 1$), can be estimated as the regression:

$$\Delta \ln \overline{Y}_t = \alpha + \beta_1 \overline{\rho}_t + \beta_2 \frac{\Delta \ln \overline{H}_t}{1 + \overline{\rho}_t} + u_t \tag{31}$$

Where α, β_1 and β_2 are regression parameters whose correspondences with the model are the following:

$$\alpha \equiv \bar{\Gamma} - \frac{\bar{s}^2}{2}$$

$$\beta_1 \equiv \frac{\bar{\sigma}^2 - \bar{\delta}}{2}$$

$$\beta_2 \equiv 1$$

$$u_t \equiv \bar{s}\Delta\bar{W}_t$$

Note that the definition of β_2 imposes an additional test of consistency, because only the value $\beta_2 = 1$ is compatible with the theoretical model, and therefore the regression results must be such that this hypothesis cannot be rejected.

In addition, if the regression is performed against the data of a closed economy (or one large enough for the foreign sector to be proportionally small e.g. the USA), one should also expect the value of β_1 to be lower than zero. The reason is that, although in theory the definition of $\beta_1 \equiv \frac{\bar{\sigma}^2 - \bar{\delta}}{2}$ could also be compatible with a positive sign, in a closed economic system, where the total number of labor hours available is a given, increases in the demand of hours by some producers can only be fulfilled though reallocation from other producers. Under these conditions, the overall system variance $\bar{\sigma}^2 dt = \left(\frac{d\bar{H}_t}{(1+\bar{\rho}_t)\bar{H}_t} \right)^2$ will logically be zero, whereas this does not apply to the aggregation of individual variances $\bar{\delta}$ (because, unlike $\bar{\sigma}^2$, this metric excludes the covariances across units, which in this case are negative). Hence, although a positive value for β_1 is in principle possible in an open system, given that imports represent a relatively small portion of the U.S. economy, the result should be close to that of a closed system, and thus a result where $\beta_1 \geq 0$ should raise a question mark.

4.2. Data and Approach

The variables of the regression are obtained on the basis of the following dataset:

- As estimators of **aggregate output** ('$\overline{Y}_t$'), **total labor time** ('$\overline{H}_t$'), **aggregate labor input** ('$\overline{L}_t$'), and **capital input** ('$\overline{K}_t$') we use *aggregate U.S. private nonfarm growth data for, respectively, real GDP, hours of all persons, labor input and capital services from 1949 to 2008* according to the data provided by the *U.S. Bureau of Labor Statistics*[7].

  The difference between '$\overline{H}_t$' and '$\overline{L}_t$' is that the first is a non-weighted sum of all workers' time, whereas the second is the standard Tornquist aggregate that is used for macroeconomic Cobb-Douglas estimations. In truth, both variables are so highly correlated (R^2=97.77% between the two in this particular sample) that it would matter little which one we used in a model – but the distinction is important from a conceptual viewpoint.

- As for the **overall rent ratio** $\overline{\rho}_t$, there is unfortunately no single, all-encompassing indicator of the portion of economic rents embedded in the overall price of productive factors. Nevertheless, since, in this model, all capital investments are regarded as "sunk" costs, any return it generates must be a rent: hence, a good measure of it could be the risk-free interest rate. Indeed, in this interpretation the money market is but a process of wealth transfer between lenders (i.e. savers) and borrowers (i.e. producers), in which credit represents funds moving from the former to the latter and interests (and credit repayments) from the latter to the former. The balance of interests

[7] The source data can be found in http://stats.bls.gov/mfp/ and have been reproduced in Appendix 3

generated less increments of credit provided is thus equivalent to a rent paid by producers for their use of capital invested.

In addition, since expression (30) takes as an input its value at instant t instead of its increment between instants t and $t + \Delta t$, we need to take the value of this indicator at the start of the period – which, as we deal with annual data, means at the start of the year. Of course this is a fairly crude measure, as there is nothing inherently "special" in the rent ratio in the month of January, or in the calendar year as a time horizon, but this should help us to test whether the rent ratio today does indeed help us to predict the rate of output growth tomorrow.

As a result of all this, we take as an estimator of the rent ratio $\overline{\rho}_t$, the *U.S. nominal risk-free interest rate (specifically, the Federal Reserve Prime Rate) at the start of every year from 1949 to 2008, less the annualized rate of growth of USD money supply (measured as M2) during the month of January of that year*, always according to the *Federal Reserve Bank of St. Louis*[8].

Under these conditions, the model is subjected to four empirical tests:

- ***TEST 1:*** First the proposed model is estimated through an OLS regression in order to analyze its empirical fit and its consistency with the theory.
- ***TEST 2:*** Next, a Cobb-Douglas production function is estimated under the same conditions, and its empirical fit compared with the previous one[9].

---

[8] The source data are in http://research.stlouisfed.org/fred2/ and http://research.stlouisfed.org/aggreg/ and have also been reproduced in Appendix 3

[9] The values of the aggregate labour and capital services variables, which are required for this regression analysis, are provided by the *U.S. Bureau of Labor Statistics*. The source data can be found in http://stats.bls.gov/mfp/ and have been reproduced in Appendix 3

- ***TEST 3:*** The two models are then compared through a Davidson-MacKinnon test (following Davidson & MacKinnon 1981) to assess whether one should be preferred to the other, or if they should be regarded as complementary.
- ***TEST 4:*** Finally, the regression in test 1 is repeated, only now adding capital input changes (both contemporary and lagged) as an explanatory variable to test whether aggregate capital is better or worse a measure of installed capacity than financial rents, or if the two are complementary.

4.3. Empirical Results

4.3.1. TEST 1

Here we estimate the parameters of the model as formulated in expression (31) by means of an OLS regression[10]. The results appear on Table A:

Table A
Results of an OLS regression on the model proposed in this paper

Analytical Expression: $\Delta \ln \overline{Y}_t = \alpha + \beta_1 \overline{\rho}_t + \beta_2 \dfrac{\Delta \ln \overline{H}_t}{1+\overline{\rho}_t} + u_t$

Independent Variables:

Constant	2.41653 ***	2.42891 ***
	(0.184043)	(0.176064)
$\overline{\rho}_t$	-0.160799 ***	-0.160809 ***
	(0.0288181)	(0.0287530)
$\dfrac{\Delta \ln \overline{H}_t}{1+\overline{\rho}_t}$	0.993380 ***	0.981038 ***
	(0.0648514)	(0.0593733)

[10] In this and the following tables, the asterisks beside the estimated values indicate their significance according to Student's *t*-test: one asterisk indicates significance at 90% confidence or more; two asterisks, significance at 95% confidence or more; three asterisks, significance at 99% confidence or more. For further clarity, standard deviations are always represented between brackets.

Statistical Fit Metrics:	
Degrees of Freedom	57
R^2	82.9808 %
F-statistic	138.9580
Akaike Information Criterion	201.8077
Bayesian (a.k.a. Schwarz) Information Criterion	208.0907
Hannan-Quinn Information Criterion	204.2653
Durbin-Watson Test[11] *(critical values 1.51442 and 1.65184)*	2.045814
Breusch-Godfrey Test (with a lag of order one)	0.0886961
Breusch-Godfrey P-value for P(F(1,56) > 0.0886961)	76.6943 %

The R^2 value suggests that this expression can explain nearly 85 percent of the total variability of the series. The results are also consistent with the model's predictions regarding parameters β_1 and β_2. The *t*-test suggests both are different to zero at over 99% confidence, the estimated value of β_1 is negative (as expected) and the *t*-ratio for the null hypothesis $\beta_2 = 1$ is equal to $\frac{0.993380 - 1}{0.0648514} = -0.10208$, which does not allow to reject the hypothesis[12]. Moreover, the robustness of this result is confirmed by both the Durbin-Watson and the Breusch-Godfrey tests: in both cases, the result indicates that there is no first-order autocorrelation in the residuals.

4.3.2. TEST 2

Next we follow the same procedure to test a Cobb-Douglas function as expressed in (25) i.e. $\overline{Y}_t = \overline{A}_t \overline{K}_t^{\beta_1} \overline{L}_t^{\beta_2}$ (only, for the sake of generality, allowing $\beta_1 + \beta_2 \neq 1$) and where we assume the magnitudes $\overline{\gamma}_t$ and $\overline{s}_t$ in (24) are constant. The results are in Table B:

[11] Durbin-Watson critical values are calculated for 5% significance

[12] In fact this ratio corresponds to a *t*-distribution probability of 46% of being larger than unity (whereas obviously that of being larger than the estimated value is, by definition, 50%). In other words, the estimated value is so close to unity that the *t*-likelihood of the two differs in just 4%

Table B
Results of an OLS regression on a Cobb-Douglas function

Analytical Expression: $\Delta \ln \bar{Y}_t = \alpha + \beta_1 \Delta \ln \bar{K}_t + \beta_2 \Delta \ln \bar{L}_t + u_t$

Independent Variables:

Constant	2.57462 ***	2.42891 ***
	(0.814243)	(0.176064)
K_t	-0.150151	--0.160809
	(0.195456)	(0.195456)
L_t	1.06501 ***	1.06501 ***
	(0.0952961)	(0.0952961)

Statistical Fit Metrics:

Degrees of Freedom	57
R^2	69.9311 %
F-statistic	66.28242
Akaike Information Criterion	235.9566
Bayesian (a.k.a. Schwarz) Information Criterion	242.2396
Hannan-Quinn Information Criterion	238.4142
Durbin-Watson Test (critical values 1.51442 and 1.65184)	1.566992
Breusch-Godfrey Test (with a lag of order one)	2.831592
Breusch-Godfrey P-value for P(F(1,56) > 2.831592)	9.8 %

The statistical fit is clearly much worse than in Table A. Not only is the R^2 lower in Table B, but the *F*-statistic is also smaller, and all the information criteria (Akaike, Schwarz and Hannan-Quinn), which signal higher likelihood the smaller they are, take much larger values. Furthermore, unlike in Table A, the Durbin-Watson test takes a value which suggests that there might be an autocorrelation issue: indeed, the value of this test falls right between the Durbin-Watson lower and upper critical values for 5% significance (1.51442 and 1.65184, respectively), and therefore its result is inconclusive. A more powerful tool, the Breusch-Godfrey test, allows to reject the autocorrelation hypothesis at 5% significance – but only just, because at 10% it would

not. These inconclusive results suggests a possibility that *t*-ratio values might be somewhat overestimated, thus reducing even further their explanatory value.

Moreover, the regression results simply do not support the Cobb-Douglas specification. In particular, the estimated regression coefficient associated to capital input ('β_1') is *negative* (i.e. that capital is a factor of value *destruction*) which is contradictory to the model, although with such a large standard deviation it is not even possible to reject the hypothesis that the actual coefficient be $\beta_1 = 0$ (which would still contradict the Cobb-Douglas framework). Nevertheless, the estimated value is so far away from the theoretical one (i.e. from capital's share of output, which is approximately $\beta_1 = 0.32$) that we can reject it as a null hypothesis, for its *t*-ratio would be $\frac{-0.150151 - 0.32}{0.195456} = -2.40541$, which allows to reject it with a 95% confidence margin. The conclusion for the parameter ('β_2') associated to labor is similar to that of capital (although, in this case, its regression coefficient is statistically different from zero): its *t*-ratio respective to the null hypothesis that it be equal to labor's share of output (i.e. approximately $\beta_2 = 0.68$) is $\frac{1.06501 - 0.68}{0.0952961} = 4.04014$, which allows to reject the hypothesis with over 99% confidence.

As discussed in the introduction, this is by no means a novel finding. If it is not highlighted more often is probably because most mainstream models are designed to work around it, be it by performing the regression without a constant time trend factor (i.e., in terms of Table B, by imposing *a priori* that $\alpha = 0$) and then estimating parameters '$\hat{\beta}_1$' and '$\hat{\beta}_2$', be it by calibrating the theoretical weights of capital and labor (i.e. by defining *a priori* the values of 'β_1' and 'β_2') and then estimating only

'α' as a residual growth rate. These modeling methods are useful for other purposes, but do not help us to estimate the production function, for they assume away what should actually be proven. In this sense, the results in Table B are clear: neither can parameter 'α' be equated to zero (because the *t*-statistic rejects such hypothesis with a confidence margin of over 99%) nor, once this time trend element is introduced, can the hypothesis of parameters 'β_1' and 'β_2' representing factor shares be sustained.

4.3.3. TEST 3

Although all the criteria we have reviewed so far support the regression in Table A (i.e. the model postulated in this paper) against the one in Table B (the standard Cobb-Douglas function), this comparison alone does not suffice to reject the latter, for there is still a possibility that both models be complementary and a "hybrid" one combining the explanatory variables of the two offer an even better result.

To distinguish between these two possibilities we resort to a Davidson/MacKinnon test (following Davidson & MacKinnon 1981). This test requires to re-run each one of the two original regression models but now including as explanatory variables**,** in addition to the original ones, the fitted values of the alternative model. A rejection of one of the two models will take place when its fitted values come out as not statistically significant (on the basis of a *t*-test) when included in the alternative model and at the same time the fitted values of the alternative model come out as significant when included as an independent variable in the first one. As a result, this test does not "force" the identification of a single winner: it may happen that none of the models is rejected (in which case we would probably need to consider a "hybrid" model), or that both are (which might point towards a more fundamental specification

issue). Hence, if we *do* end up finding a clear winner, this does arguably provide a very strong piece of evidence in favor of adopting it at the expense of the other model.

As shown in Tables C and D, the outcome of this test is clear. Table C shows the result of performing the same regression as in Table A but including now the fitted values from the one in Table B as an explanatory variable (represented by '$\hat{\hat{B}}_t$'):

Table C
Davidson/MacKinnon Test 1

Analytical Expression: $\Delta \ln \overline{Y}_t = \alpha + \beta_1 \overline{\rho}_t + \beta_2 \frac{\Delta \ln \overline{H}_t}{1+\overline{\rho}_t} + \beta_3 \hat{\hat{B}}_t + u_t$

Independent Variables:

Constant	*Coefficient ('$\hat{\alpha}$')*	2.05381 **
	Standard Deviation	(0.896827)
$\overline{\rho}_t$	*Coefficient ('$\hat{\beta}_1$')*	-0.163373 ***
	Standard Deviation	(0.0296904)
$\frac{\Delta \ln \overline{H}_t}{1+\overline{\rho}_t}$	*Coefficient ('$\hat{\beta}_2$')*	0.845029 **
	Standard Deviation	(0.364774)
$\hat{\hat{B}}_t$	*Coefficient ('$\hat{\beta}_3$')*	0.151979
	Standard Deviation	(0.367651)

Statistical Fit Metrics:

Degrees of Freedom	56
R^2	83.0326 %
F-statistic	91.34813

This result proves that, under these conditions, the fitted values from the regression in Table B (i.e. from the Cobb-Douglas model) *lose* their statistical significance, whereas the other variables remain significant.

Conversely, Table D displays the results of the opposite exercise, namely, performing the same regression as in Table B but including the fitted values from the one in Table A (represented by variable '$\hat{\hat{A}}_t$') as an input variable, and proves that, under these conditions, *only* the fitted values from that regression Table A retain their statistical significance as an explanatory variable.

Table D
Davidson/MacKinnon Test 2

Analytical Expression: $\Delta \ln \overline{Y}_t = \alpha + \beta_1 \Delta \ln \overline{K}_t + \beta_2 \Delta \ln \overline{L}_t + \beta_3 \hat{\hat{A}}_t + u_t$

Independent Variables:

Constant	*Coefficient ('$\hat{\alpha}$')*	0.0780094
	Standard Deviation	(0.725731)
K_t	*Coefficient ('$\hat{\beta}_1$')*	-0.00931168
	Standard Deviation	(0.149858)
L_t	*Coefficient ('$\hat{\beta}_2$')*	0.0319070
	Standard Deviation	(0.173350)
$\hat{\hat{A}}_t$	*Coefficient ('$\hat{\beta}_3$')*	0.975332 ***
	Standard Deviation	(0.148739)

Statistical Fit Metrics:

Degrees of Freedom	56
R^2	82.9912 %
F-statistic	91.08028

In short, these results reinforce the conclusions from Tests 1 and 2, suggesting that *the model put forward in this paper constitutes a better-fitting specification than the traditional Cobb-Douglas production function* and, furthermore, that *a hybrid combining the explanatory variables of both models would not provide any additional explanatory power.*

4.3.4. TEST 4

Although the results of Tests 1, 2 and 3 are robust and clear-cut, one could still wonder whether they should simply be interpreted as rejection of the Cobb-Douglas specification or, more fundamentally, of the significance of the variables it is built upon. The question is almost a moot point in the case of Tornquist aggregate labor ('$\overline{L}_t$'), since it is so highly correlated to the non-weighted sum of work hours ('$\overline{H}_t$'), although the latter is more closely correlated to GDP (as is well known – see for example Barro 2008), just as the model in this paper would predict. In the case of aggregate capital, conversely, the question of whether it is a better or worse measure of installed capacity than economic rents is relevant for two reasons:

a) The production function proposed in this paper is defined as a "short-term" function where capacity is a constraint, but lacks a mechanism to explain how this capacity evolves in the long run. To the extent capacity is built up through investment, one would expect today's net investments (i.e. changes in aggregate capital) to determine future capacity expansions.
b) On the flip side, if (as the Cambridge Critique would predict) it makes no sense to measure capital input as a value-weighted capital aggregate, then one would expect to find that net investment plays no role in capacity expansion.

To test these alternative hypotheses, we perform a regression on an encompassing model (following the method proposed by Mizon & Richard 1986) including both the production function in (31) and the change in aggregate capital, lagged *n* periods. To cover a wide range of possibilities, the regression is repeated for aggregate capital lags ranging from zero (i.e. contemporary) to nine years, as shown in Table E.

Table E
Results of OLS regressions on a comprehensive model including both rent ratio and lagged net investment

Analytical Expression: $$\Delta \ln \overline{Y}_t = \alpha + \beta_1 \overline{\rho}_t + \beta_2 \frac{\Delta \ln \overline{H}_t}{1+\overline{\rho}_t} + \beta_3 \Delta \ln \overline{K}_{t-n} + u_t$$

Lag (n)		$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	Degrees of Freedom	R^2	F
0	Coefficient	2.42695 ***	-0.160649 ***	0.993808 ***	-0.00263317	56	82.9809%	91.01400
	St. Dev.	(0.647701)	(0.0304137)	(0.0702052)	(0.156788)			
1	Coefficient	2.50365 ***	-0.160649 ***	0.993808 ***	-0.00263317	55	82.6823%	87.53122
	St. Dev.	(0.666359)	(0.0304137)	(0.0702052)	(0.156788)			
2	Coefficient	1.76982 ***	-0.149895 ***	1.00602 ***	0.127561	54	84.5758%	98.69962
	St. Dev.	(0.611189)	(0.0267011)	(0.0640098)	(0.134288)			
3	Coefficient	2.18125 ***	-0.150875 ***	0.988664 ***	0.0330864	53	83.8601%	91.79266
	St. Dev.	(0.606440)	(0.0271605)	(0.0649257)	(0.134746)			
4	Coefficient	2.14379 ***	-0.153336 ***	0.985094 ***	0.0471138	52	84.0698%	91.47464
	St. Dev.	(0.599357)	(0.0272911)	(0.0648228)	(0.134932)			
5	Coefficient	3.04774 ***	-0.154033 ***	0.985779 ***	-0.163779	51	84.4047%	92.00753
	St. Dev.	(0.616606)	(0.0272029)	(0.0647449)	(0.137464)			
6	Coefficient	2.99013 ***	-0.154498 ***	0.970040 ***	-0.141506	50	83.4339%	83.94044
	St. Dev.	(0.642998)	(0.0275939)	(0.0682558)	(0.141171)			
7	Coefficient	3.26846 ***	-0.146056 ***	0.951912 ***	-0.206147	49	83.1418%	80.55323
	St. Dev.	(0.653130)	(0.0278211)	(0.0687830)	(0.142637)			
8	Coefficient	3.62372 ***	-0.137721 ***	0.968504 ***	-0.282509 **	48	85.7281%	96.10870
	St. Dev.	(0.605256)	(0.0262653)	(0.0634209)	(0.134347)			
9	Coefficient	3.88621 ***	-0.138796 ***	0.984766 ***	-0.352919 **	47	86.2609%	98.36325
	St. Dev.	(0.603281)	(0.0256572)	(0.0630930)	(0.137118)			

The outcome is unambiguous. Both the rent ratio and the labor input variables retain their significance in every single case, and none of the regressions with lagged net investment provides any clear improvement in explanatory power (as measured by the R^2 and F statistics). Conversely, for net investments with up to seven years' lag not only is the coefficient negative in most of the cases, but in none of them is it possible to reject the null hypothesis of it being zero… and for lags of eight or nine years, bizarrely, the test indicates statistical significance but for a ***negative*** coefficient: in other words, it suggests that investments made eight or nine years ago have over 95% probability of resulting in a ***reduced*** productive capacity today (!).

Two straightforward interpretations spring to mind: either there really is a mechanism by which net investment is not only irrelevant but even detrimental to production (which seems contrary to common sense) or using aggregate capital as a measure of non-labor input simply makes no sense, as Robinson and Sraffa defended so long ago. Faced with this choice, one must let Occam's Razor rule and conclude with the Cambridge (England) school that *value-weighted aggregate capital metrics play no measurable role in the macroeconomic production function*.

5. SUMMARY AND CONCLUSIONS

Fisher and Gorman proved long ago that a macroeconomic production function cannot be built on the basis of just any set of individual production functions, even if behind them are rational, profit-maximizing agents. One can always add up individual production functions into an aggregate, to be sure, but the aggregate input results from adding up individual input quantities weighted by their marginal costs, and these may have different productivities depending on which producer uses them. Additional assumptions are therefore required to develop a usable function.

One such assumption is perfect competition, i.e. that marginal costs always equal market prices. Arguably, this postulate is central to the Cobb-Douglas function, and can be justified by supposing that, in the long run, most inputs are variable and so economic rents may be discounted as a mere form of short-term friction. As explained above, however, this function poses a number of theoretical and empirical issues, which suggests there is room for exploring an alternative. This paper proposes to base this alternative on the exact opposite assumption: that it is the short term that matters most to understand output fluctuations, and that, within this time horizon, only one major aggregate input (labor time) may be regarded as variable. From a theoretical perspective, this approach presents a number of advantages, among them not being subject to the Cambridge Critique; nevertheless, one could equally think of some strong theoretical points to support the Cobb-Douglas function, so this alone cannot drive the choice: to select a model above the other we need to resort to empirical data.

We perform this empirical analysis in Section 4, with the following results:

1. The model proposed in this paper is empirically robust, and can explain nearly 85 percent of the U.S. GDP fluctuation from 1949 to 2008.
2. Against the same data set, the Cobb-Douglas function proves to be a lot less robust, with substantially lower fit and likelihood metrics as well as significant evidence of residual autocorrelation, and the fundamental hypothesis that the parameters associated to capital and labor are equal to their shares of GDP can be rejected with over a 95% confidence margin.
3. A non-nested model comparison (specifically, a Davidson/MacKinnon test) concludes that the model put forward in this paper represents a better

specification than the Cobb-Douglas function and, furthermore, that a hybrid of the two would not add any meaningful explanatory power either.

4. Last but not least, the estimated impact of capital, both contemporary and lagged, on production is nearly always negative (although the variance around it does not allow to exclude its simply being nil), which suggests that, as the Cambridge Critique postulated long ago, aggregate capital makes no sense as a measure of non-labor productive inputs.

Why does it matter which model we select? It matters because their implications are very different. A Cobb-Douglas world where output depends on the quantities of capital and labor and their marginal productivities, which also happen to be their prices, is very different from one where every income above the marginal product of basic unqualified labor constitutes a rent, and where these rents in turn determine how overall productivity behaves. In this latter world, for example, central bank policy would not require any sort of irrationality or monetary illusion to be effective as long as interest rates and credit flows act as economic rents and therefore drive the behavior of the aggregate production function. Conversely, in such a world it might not be licit to model the choice between investment and consumption as most DSGE papers do, for aggregate capital (i.e. accumulated net investment) could not be represented as a productive input… It would go well beyond the scope of this paper to analyze each one of these implications; but it is probably fair to say here that the predictions from models based on the production function put forward in this paper would probably diverge substantially from the standard ones.

****

APPENDICES

Appendix 1

The purpose of this appendix is to show the development of a widely-used derivation of a production function's degree of homogeneity (see for example Varian 1986 for a textbook example, or Basu & Fernald 1997 and 2000 for empirical applications), and so to highlight where its assumptions differ from those in Subsection 3.1.

We define the marginal cost respective to the output quantity $C'_{j,t} \equiv \frac{\partial C_{j,t}}{\partial Y_{j,t}}$. Combining this with the definition of marginal cost respect to input $c_{i,j,t} \equiv \frac{\partial C_{j,t}}{\partial x_{i,j,t}}$, if we adopt Assumption 3 (i.e. continuity and differentiability), then the relationship between these two variables may be expressed as $c_{i,j,t} = C'_{j,t} \frac{\partial Y_{j,t}}{\partial x_{i,j,t}}$.

Thus, if we also adopt Assumption 4 (i.e. homogeneity) in terms of expression (5) and combine it with the above we obtain:

$$h_{j,t} = \sum_{i=1}^{n} \frac{c_{i,j,t}}{C'_{j,t}} \frac{x_{i,j,t}}{Y_{j,t}} = \frac{1}{C'_{j,t} Y_{j,t}} \sum_{i=1}^{n} c_{i,j,t} x_{i,j,t} \qquad (1.1)$$

It is common at this point to introduce an assumption to the effect that the cost function $C_{j,t}$ is homogeneous of degree one respective to the inputs $\{x_{1,j,t} \ldots x_{n,j,t}\}$ so that $C_{j,t} = \sum_{i=1}^{n} c_{i,j,t} x_{i,j,t}$, and then to establish the following definitions:

- Mark-up ratio $m_{j,t} \equiv \frac{P_t}{C'_{j,t}}$

- Profit ratio as $\pi_{j,t} \equiv \frac{P_t Y_{j,t} - C_{j,t}}{P_t Y_{j,t}}$

Which allow to rewrite expression (1.1) in the more familiar form:

$$h_{j,t} = \frac{1}{C'_{j,t} Y_{j,t}} \underbrace{\sum_{i=1}^{n} c_{i,j,t} x_{i,j,t}}_{=C_{j,t}} = \frac{C_{j,t}}{C'_{j,t} Y_{j,t}} = \frac{C_{j,t}}{C'_{j,t} Y_{j,t}} \frac{P_t}{P_t} = \underbrace{\frac{P_t}{C'_{j,t}}}_{m_{j,t}} \underbrace{\frac{C_{j,t}}{P_t Y_{j,t}}}_{1-\pi_{j,t}}$$

$$h_{j,t} = m_{j,t}\left(1 - \pi_{j,t}\right) \qquad (1.2)$$

Note, however, that, up to this point, we have resorted to Assumptions 3 and 4 (and also implicitly to Assumption 2, at least to the extent that, if no optimal point existed, the whole exercise would be pointless) – yet Assumption 1 (i.e. maximization of the real profit function $\Pi^*_{j,t}$) has played no role so far. If we now formally introduce this assumption, then evidently:

$$\frac{\partial \Pi^*_{j,t}}{\partial Y_{j,t}} \equiv \frac{\partial}{\partial Y_{j,t}}\left(Y_{j,t} - C^*_{j,t}\right) \equiv 1 - C'^*_{j,t} = 0 \qquad (1.3)$$

Which means that $m_{j,t} \equiv \frac{P_t}{C'_{j,t}} \equiv \frac{1}{C'^*_{j,t}} = 1$. On the other hand, since applying definition (6) to the profit ratio leads to $\pi_{j,t} = \frac{\rho_{j,t}}{1+\rho_{j,t}}$, the degree of homogeneity must be $h_{j,t} = 1 - \pi_{j,t} = \frac{1}{1+\rho_{j,t}}$. This is exactly the conclusion we reached in Subsection 3.1.

Q.E.D.

Appendix 2

The purpose of this appendix is to show the analytical derivation of expression (14) in Subsection 3.2.

As a first step, we need to differentiate the production function under the form introduced in expression (11), that is, $Y_{j,t} = A_{j,t} X_{j,t}^{\frac{1}{1+\rho_{j,t}}}$. According to Itô's lemma (i.e. the stochastic version of the chain rule) the differential of this expression is as follows (note that, due to its length, this expression spans over two lines of text):

$$dY_{j,t} = \frac{\partial Y_{j,t}}{\partial t} dt + \frac{\partial Y_{j,t}}{\partial W_{j,t}} dW_{j,t} + \frac{\partial Y_{j,t}}{\partial X_{j,t}} dX_{j,t} + \frac{\partial Y_{j,t}}{\partial \rho_{j,t}} d\rho_{j,t} + \frac{1}{2} \frac{\partial^2 Y_{j,t}}{\partial W_{j,t}^2} \left(dW_{j,t}\right)^2 + \frac{1}{2} \frac{\partial^2 Y_{j,t}}{\partial X_{j,t}^2} \left(dX_{j,t}\right)^2 + \ldots$$

$$\ldots + \frac{1}{2} \frac{\partial^2 Y_{j,t}}{\partial \rho_{j,t}^2} \left(d\rho_{j,t}\right)^2 + \frac{\partial^2 Y_{j,t}}{\partial W_{j,t} \partial X_{j,t}} dW_{j,t} dX_{j,t} + \frac{\partial^2 Y_{j,t}}{\partial W_{j,t} \partial \rho_{j,t}} dW_{j,t} d\rho_{j,t} + \frac{\partial^2 Y_{j,t}}{\partial X_{j,t} \partial \rho_{j,t}} dX_{j,t} d\rho_{j,t}$$

Which, as we know from Assumption 5 that all the differentials respective to ρ_t are zero, can be simplified into the following expression (still expanding over two lines):

$$dY_{j,t} = \frac{\partial Y_{j,t}}{\partial t} dt + \frac{\partial Y_{j,t}}{\partial W_{j,t}} dW_{j,t} + \frac{\partial Y_{j,t}}{\partial X_{j,t}} dX_{j,t} + \frac{1}{2} \frac{\partial^2 Y_{j,t}}{\partial W_{j,t}^2} \left(dW_{j,t}\right)^2 + \frac{1}{2} \frac{\partial^2 Y_{j,t}}{\partial X_{j,t}^2} \left(dX_{j,t}\right)^2 + \ldots$$

$$\ldots + \frac{\partial^2 Y_{j,t}}{\partial W_{j,t} \partial X_{j,t}} dW_{j,t} dX_{j,t} \qquad (2.1)$$

If we now divide this formula by $Y_{j,t}$ and replace everything with its equivalent according to expression $Y_{j,t} = A_{j,t} X_{j,t}^{\frac{1}{1+\rho_{j,t}}}$, then (2.1) turns into the following (still requiring two lines of text):

$$\frac{dY_{j,t}}{Y_{j,t}} = \frac{1}{A_{j,t}}\left(\frac{\partial A_{j,t}}{\partial t}dt + \frac{\partial A_{j,t}}{\partial W_{j,t}}dW_t + \frac{1}{2}\frac{\partial^2 A_{j,t}}{\partial W_{j,t}^2}\left(dW_{j,t}\right)^2\right) + \frac{1}{1+\rho_{j,t}}\frac{dX_{j,t}}{X_{j,t}} + \ldots$$

$$\ldots + \frac{1}{1+\rho_{j,t}}\frac{\partial A_{j,t}}{\partial W_{j,t}}\frac{dW_{j,t}dX_{j,t}}{A_{j,t}X_{j,t}} - \frac{\rho_{j,t}}{2\left(1+\rho_{j,t}\right)^2}\left(\frac{dX_{j,t}}{X_{j,t}}\right)^2 \qquad (2.2)$$

Now, since we know from Assumption 5 that the perturbations on the increments of A_t are distributed according to a Wiener process, then we can also use Itô's lemma to decompose it as follows:

$$\frac{dA_{j,t}}{A_{j,t}} = \frac{1}{A_{j,t}}\left(\frac{\partial A_{j,t}}{\partial t}dt + \frac{\partial A_{j,t}}{\partial W_{j,t}}dW_{j,t} + \frac{1}{2}\frac{\partial^2 A_{j,t}}{\partial W_{j,t}^2}\left(dW_{j,t}\right)^2\right) = \gamma_{j,t}dt + s_{j,t}dW_{j,t} \qquad (2.3)$$

Which, applied to expression (2.2), leads to:

$$\frac{dY_{j,t}}{Y_{j,t}} = \gamma_{j,t}dt + s_{j,t}dW_{j,t} + \frac{1}{1+\rho_{j,t}}\left(\frac{dX_{j,t}}{X_{j,t}} - \frac{\rho_{j,t}}{2\left(1+\rho_{j,t}\right)}\left(\frac{dX_{j,t}}{X_{j,t}}\right)^2 + s_{j,t}\frac{dW_{j,t}dX_{j,t}}{X_{j,t}}\right) \qquad (2.4)$$

Or, rearranging terms:

$$\frac{dX_{j,t}}{\left(1+\rho_{j,t}\right)X_{j,t}} = \frac{dY_{j,t}}{Y_{j,t}} - \gamma_{j,t}dt - s_{j,t}dW_{j,t} - s_{j,t}\frac{dW_{j,t}dX_{j,t}}{\left(1+\rho_{j,t}\right)X_{j,t}} + \frac{\rho_{j,t}}{2\left(1+\rho_{j,t}\right)^2}\left(\frac{dX_{j,t}}{X_{j,t}}\right)^2 \qquad (2.5)$$

If we replace $\frac{dY_t}{Y_t}$ according to Assumption 7, this expression becomes:

$$\frac{dX_{j,t}}{(1+\rho_{j,t})X_{j,t}} = \mathrm{E}_t\left[\frac{dY_{j,t}}{Y_{j,t}}\right] + \sigma_{j,t}dV_{j,t} - \gamma_{j,t}dt - s_{j,t}\frac{dW_{j,t}dX_{j,t}}{(1+\rho_{j,t})X_{j,t}} + \frac{\rho_{j,t}}{2(1+\rho_{j,t})^2}\left(\frac{dX_{j,t}}{X_{j,t}}\right)^2 \quad (2.6)$$

If we now square both sides of this expression and, as usual in Wiener processes, eliminate all terms $dtdW_{j,t}$, $dtdV_{j,t}$ or $dtdX_{j,t}$ as well as those under the form $(dt)^n$ or $(\mathrm{E}_t[dY_{j,t}])^n$ or $(dX_{j,t})^{n+1}$ where $n>1$, and then replace $(dV_{j,t})^2 = dt$, we obtain:

$$\left(\frac{dX_{j,t}}{(1+\rho_{j,t})X_{j,t}}\right)^2 = \sigma_{j,t}^2 dt \quad (2.7)$$

Thus, replacing this back into expression (2.6) we obtain:

$$\frac{1}{1+\rho_{j,t}}\frac{dX_{j,t}}{X_{j,t}} = \mathrm{E}_t\left[\frac{dY_{j,t}}{Y_{j,t}}\right] + \sigma_{j,t}dV_{j,t} - \gamma_{j,t}dt - s_{j,t}\frac{dW_{j,t}dX_{j,t}}{(1+\rho_{j,t})X_{j,t}} + \frac{\sigma_{j,t}^2}{2}\rho_{j,t}dt \quad (2.8)$$

Given that the stochastic component of $\frac{dX_{j,t}}{(1+\rho_{j,t})X_{j,t}}$ is $\sigma_{j,t}dV_{j,t}$, since Assumption 7 postulated that this variable would be independent from technology progress i.e. $dV_{j,t}dW_{j,t} = 0$, we can further simplify expression (2.8) into:

$$\frac{1}{1+\rho_{j,t}}\frac{dX_{j,t}}{X_{j,t}} = \mathrm{E}_t\left[\frac{dY_{j,t}}{Y_{j,t}}\right] + \sigma_{j,t}dV_{j,t} - \gamma_{j,t}dt + \frac{\sigma_{j,t}^2}{2}\rho_{j,t}dt$$

$$\frac{dY_{j,t}}{Y_{j,t}} = \gamma_{j,t}dt + s_{j,t}dW_{j,t} - \frac{\sigma_{j,t}^2}{2}\rho_{j,t}dt + \frac{1}{1+\rho_{j,t}}\frac{dX_{j,t}}{X_{j,t}} \quad (2.9)$$

Which is equal to expression (14) in Subsection 3.2.

Q.E.D.

Appendix 3

The purpose of this appendix is to display the dataset supporting the empirical tests in Section 4.

	Source: U.S. Bureau of Labor Statistics				*Source: Federal Reserve Bank of St. Louis*			*Calculated Variables*	
	(a)	*(b)*	*(c)*	*(d)*	*(e)*	*(f)*	*(g)*	*(h)*	*(i)*
Year	Real Value-Added Output Growth Rate	Sum of Work Hours of All Persons, Growth Rate	Aggregate Labor Growth (Tornquist Aggregate)	Aggregate Capital Growth Rate	Federal Reserve Prime Rate on January 1	Money Supply (M2) on January 1	Money Supply (M2) on February 1	Rent Ratio[13]	Work Hours Growth Weighted by Rent Ratio[14]
t	$\Delta \ln \overline{Y}_t$	$\Delta \ln \overline{H}_t$	$\Delta \ln \overline{L}_t$	$\Delta \ln \overline{K}_t$	-	-	-	$\overline{\rho}_t$	$\frac{\Delta \ln \overline{H}_t}{1+\overline{\rho}_t}$
1949	-0.7	-4.3	-4.2	2.9	2.00	171.5	171.7	0.5	-4.3
1950	10.2	3.1	3.6	4.1	2.00	174.6	175.5	-4.4	3.2
1951	7.6	4.8	4.8	5.4	2.44	182.3	182.6	0.4	4.8
1952	2.9	0.9	1.7	3.4	3.00	193.7	194.8	-4.0	0.9
1953	4.9	2.5	3.3	3.3	3.00	205.9	206.6	-1.2	2.5
1954	-1.6	-3.6	-3.0	2.4	3.25	216.6	217.6	-2.4	-3.7
1955	8.5	4.0	4.2	3.7	3.00	230.8	232.4	-5.6	4.2
1956	1.7	2.5	2.6	4.1	3.50	241.7	242.3	1.0	2.5
1957	2.0	-0.7	-0.1	3.3	4.00	252.7	253.7	-0.8	-0.7
1958	-2.2	-4.4	-4.3	2.3	4.34	263.2	265.7	-7.5	-4.8

[13] Consistently with the approach laid out in Subsection 4.2, this variable is calculated as $(h) \equiv (e) - \left(\left(1+\ln(g)-\ln(f)\right)^{12}-1\right)*100$

[14] This variable is calculated as $(i) \equiv (b)/\left(1+(h)/100\right)$

1959	8.7	4.5	4.6	2.6	4.00	286.6	287.7	-0.7	4.5
1960	1.7	0.4	0.4	3.1	5.00	298.2	298.5	3.8	0.4
1961	2.0	-1.4	-0.4	2.6	4.50	314.1	316.5	-5.0	-1.5
1962	6.9	2.2	3.3	3.3	4.50	337.5	340.1	-5.1	2.3
1963	4.7	1.0	1.2	3.7	4.50	365.2	367.9	-4.7	1.0
1964	6.8	2.4	2.4	4.0	4.50	395.2	397.6	-3.0	2.5
1965	7.1	3.8	3.7	5.0	4.50	427.5	430.4	-3.9	4.0
1966	7.2	3.5	3.3	6.1	5.00	462.0	464.6	-1.9	3.6
1967	1.7	-0.1	-0.2	5.9	5.96	481.6	485.1	-3.1	-0.1
1968	5.3	2.1	1.8	4.6	6.00	527.4	530.4	-1.0	2.1
1969	3.0	2.9	3.3	5.2	6.95	569.3	571.9	1.3	2.9
1970	-0.1	-1.7	-1.3	4.9	8.50	589.6	586.3	15.0	-1.5
1971	3.8	-0.4	-0.7	4.2	6.29	633.0	641.0	-9.9	-0.4
1972	6.8	3.5	3.4	4.5	5.18	717.7	725.7	-9.0	3.8
1973	7.4	4.0	3.9	5.7	6.00	810.3	814.1	0.2	4.0
1974	-1.4	0.0	0.6	5.5	9.73	859.7	864.2	3.3	0.0
1975	-1.7	-4.4	-4.4	3.9	10.05	906.3	914.1	-0.7	-4.4
1976	7.1	3.4	3.1	3.2	7.00	1026.6	1040.3	-10.1	3.8
1977	5.7	4.1	4.2	3.9	6.25	1165.2	1177.6	-7.2	4.4
1978	6.6	5.2	5.2	4.6	7.93	1279.7	1285.5	2.4	5.1
1979	3.2	3.5	3.2	5.5	11.75	1371.6	1377.8	6.2	3.3
1980	-1.1	-0.9	-0.5	5.5	15.25	1482.7	1494.5	5.3	-0.9
1981	2.2	0.7	1.5	5.4	20.16	1606.9	1618.7	11.0	0.6
1982	-3.1	-2.3	-1.3	4.6	15.75	1770.4	1774.5	12.9	-2.0
1983	6.6	1.9	2.3	3.8	11.16	1962.4	1999.6	-13.9	2.2
1984	8.3	6.1	6.2	5.0	11.00	2140.4	2160.5	-0.8	6.1
1985	4.3	2.6	2.9	5.4	10.61	2334.9	2356.4	-1.0	2.6
1986	3.9	0.8	1.4	4.8	9.50	2504.8	2515.3	4.4	0.8
1987	3.6	3.1	3.2	4.1	7.50	2747.1	2750.8	5.9	2.9
1988	4.6	2.9	3.5	3.7	8.75	2851.8	2874.8	-1.3	2.9

1989	3.5	2.8	3.3	3.9	10.50	2997.2	2997.6	10.3	2.5
1990	1.4	-0.4	0.1	3.4	10.11	3172.4	3184.8	5.3	-0.4
1991	-0.9	-2.5	-1.1	2.9	9.52	3292.7	3309.5	3.2	-2.4
1992	3.8	-0.2	1.0	2.5	6.50	3385.7	3404.7	-0.4	-0.2
1993	3.6	2.9	3.3	3.2	6.00	3424.7	3420.4	7.5	2.7
1994	4.7	3.6	4.1	3.7	6.00	3482.0	3482.7	5.8	3.4
1995	3.3	2.8	2.7	4.5	8.50	3502.2	3499.5	9.4	2.6
1996	4.4	1.8	2.2	4.9	8.50	3659.0	3673.0	3.8	1.7
1997	5.1	3.6	4.0	5.5	8.25	3833.3	3844.6	4.7	3.4
1998	5.2	2.2	2.4	6.6	8.50	4055.9	4088.0	-1.4	2.2
1999	5.6	2.3	2.6	7.1	7.75	4398.1	4423.4	0.6	2.3
2000	4.5	1.0	1.2	6.8	8.50	4656.4	4673.6	4.0	1.0
2001	1.0	-2.1	-1.6	5.1	9.05	4964.3	4999.9	0.1	-2.1
2002	1.9	-2.6	-2.0	3.5	4.75	5446.7	5473.1	-1.2	-2.6
2003	3.1	-0.6	-0.4	2.9	4.25	5789.6	5826.2	-3.6	-0.6
2004	4.2	1.3	1.2	2.4	4.00	6058.9	6096.9	-3.8	1.4
2005	3.4	1.7	1.9	3.3	5.25	6402.7	6415.7	2.8	1.7
2006	3.2	2.2	2.5	3.4	7.26	6709.8	6729.3	3.7	2.1
2007	2.4	0.5	1.2	3.3	8.25	7067.1	7073.8	7.1	0.5
2008	-0.2	-2.2	-1.7	2.8	6.98	7461.7	7536.9	-5.7	-2.3

BIBLIOGRAPHY

Abramovitz, M. (1956) "Resource and Output Trends in the United States since 1870" *American Economic Review*, Vol. 46, 5-23

Barro, R. J. (2008) *Macroeconomics: A Modern Approach* Thompson South-Western

Basu, S. & Fernald, J. (1997) "Aggregate Productivity and Aggregate Technology" *Board of Governors of the Federal Reserve System International Finance Discussion Paper* No. 593, October

Basu, S. & Fernald, J. (2000) "Why Is Productivity Procyclical? Why Do We Care?" *NBER Working Paper* No. 7940, October

Cobb, C. W. and Douglas, P. H. (1928) "A Theory of Production" *American Economic Review*, December, 139-65

Cohen, A. J. and Harcourt G. C. (2003) "Whatever Happened to the Cambridge Capital Theory Controversies?" *Journal of Economic Perspectives*, Winter, 199-214

Davidson, R. and MacKinnon, J. G. (1981). "Several Tests for Model Specification in the Presence of Alternative Hypotheses" *Econometrica* 49, 781-793

Douglas, P. H. (1967) "Comments on the Cobb-Douglas Production Function" *The Theory and Empirical Analysis of Production: Studies in Income and Wealth*, edited by M. Brown, National Bureau of Economic Research, 15-22

Felipe, J. & Adams, F.G. (2005) "A Theory of Production'. The Estimation of the Cobb-Douglas Function: A Retrospective View" *Eastern Economic Journal*, vol.31, 427-445

Felipe, J. & Fisher, F. M. (2003) "Aggregation in Production Functions: What Applied Economists Should Know" *Metroeconomica*, May/September, 208-62

Felipe, J. & McCombie, J. S. L. (2005) "How Sound are the Foundations of the Aggregate Production Function?" *Eastern Economic Journal*, Summer, 467-488.

Felipe, J. & McCombie, J. S. L. (2009) "On Accounting Identities, Simulation Experiments and Aggregate Production Functions: A Cautionary Tale for (Neoclassical) Growth Theorists" *Handbook of Alternative Theories of Economic Growth*, edited by M. Setterfield, Cheltenham, Edward Elgar

Fisher, F. M. (1965) "Embodied technical change and the existence of an aggregate capital stock" *Review of Economic Studies* 32, 263–88

Fisher, F. M. (1968) "Embodied technology and the existence of labor and output aggregates" *Review of Economic Studies* 35, 391–412

Fisher, F. M. (1969) "The Existence of Aggregate Production Functions" *Econometrica*, October, 553-77

Fisher, F. M. (1971) "Aggregate Production Functions and the Explanation of Wages: A Simulation Experiment" *Review of Economics and Statistics* 53, 305-325

Gorman, W. M. (1968) "Measuring the Quantities of Fixed Factors", in *Value, Capital and Growth: Papers in Honour of Sir John Hicks*, Edinburgh University Press, 141–172

Griliches, Z. & Mairesse, J. (1998) "Production Functions: The Search for Identification" *Econometrics and Economic Theory in the 20th Century*, edited by S. Strom. Cambridge University Press, 169-203

Hansen, G. D. & Prescott, E. C. (2005) "Capacity Constraints, Asymmetries, and the Business Cycle" *Review of Economic Dynamics* vol. 8(4), 850-865

Jorgenson, D. W. & Griliches, Z. (1967) "The Explanation of Productivity Change" *Review of Economic Studies* 34 (July), 349–83

Klette, T. J. & Griliches, Z. (1996) "The Inconsistency of Common Scale Estimators when Output Prices are Unobserved and Endogenous" *Journal of Applied Econometrics*, July-September, 343-61

Kydland, F. E. & Prescott, E. C. (1982) "Time to Build and Aggregate Fluctuations" *Econometrica*, Vol. 50, No. 6, 1345-1370

Lucas, R. (1970) "Capacity, Overtime, and Empirical Production Functions" *American Economic Review*, May, 23-27

May K. (1947) "Technological Change and Aggregation" *Econometrica*, 15, 51–63

Mizon, G.E. and J.F. Richard, 1986, "The Encompassing Principle and its Application to Testing Non-nested Hypotheses" *Econometrica* 54, 657-678

Phelps Brown, E. H. (1957) "The Meaning of the Fitted Cobb-Douglas Function" *Quarterly Journal of Economics*, November, 546-60

Robinson, J. (1953-54) "The Production Function and the Theory of Capital" *Review of Economic Studies*, 21:2, 81-106.

Romer, P. (1987) "Crazy Explanations for the Productivity Slowdown" *NBER Macroeconomics Annual*, MIT Press, 163-202

Samuelson, P. A. (1966) "A Summing Up" *Quarterly Journal of Economics*. November, 568-83

Samuelson, P. A. (1979) "Paul Douglas's Measurement of Production Functions and Marginal Productivities" *Journal of Political Economy*, October, 923-39

Shaikh, A. (1974) "Laws of Production and Laws of Algebra: The Humbug Production Function" *The Review of Economics and Statistics*, February, 115-20

Shaikh, A. (1980) "Laws of Production and Laws of Algebra: Humbug II" in *Growth, Profits and Property, Essays in the Revival of Political Economy,* edited by E. J. Nell. Cambridge University Press, 80-95

Simon, H. and Levy, F. (1963) "A Note on the Cobb-Douglas Function" *Review of Economic Studies*, June 1963, 93-94

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

Solow, R. M. (1966) "Review of Capital and Growth" *American Economic Review*, vol. 56, 1257-60

Varian, H. R. (1986) *Microeconomic Analysis* W. W. Norton & Co.