An Analysis of U.S. Productivity: 1960-2017

Nathan Fernig
Vancouver School of Economics

August 17, 2018

Contents

1 Introduction

"Productivity isn't everything, but in the long run it is almost everything." - Paul Krugman

Productivity is defined as the ratio of outputs to inputs used in production. An economy's ability to raise its per capita standard of living over time depends almost entirely on its ability to more efficiently use its inputs into production. Thus there is no problem more fundamental to economic analysis than understanding productivity growth, and this paper will therefore be devoted to measuring it in the US postwar period.

Since the end of World War II, the US economy has experienced a prolonged period of economic growth and increase in living standards that is unprecedented in human history. During the period 1960-2017, we will see that real output grew 5.5 fold and per capita real incomes grew almost three fold, cementing America's place as the worlds pre-eminent powerhouse economy. This robust growth occurred largely on the back of strong productivity gains over the period, with the overall level of productivity increasing over 1.6 fold.

The story of US productivity is not characterized by smooth trends but by four distinct periods with varying growth rates. Indeed, from 1960-73 the US experienced rapid productivity growth, almost 1.3% per year, due in part to low energy prices, human capital (G.I. Bill) and infrastructure improvements, as well as technological advances particularly in manufacturing automation. This changed with the oil shocks and high inflation of the 70s, and despite massive advances in computing in the 80s, productivity growth was then subpar through the early 90s, leading to the computer productivity paradox as described by Solow (1987). Then in the mid 90s there was an explosion in productivity growth that lasted through the mid 00s, largely spurred on by information and communication technologies (ICT) finally being put to practical use in the economy. Even before the great recession, productivity growth began to slow again, as gains from the first ICT revolution were thought to be exhausted. We will see the current decade could very well be the worst for productivity growth of the postwar period, leading to a *modern* productivity paradox; brynjolfsson (2017). To underscore the importance of these eras, we note that if productivity growth had been able to be maintained at either the 1961-73 or 1994-04 trends throughout the entire period, per capita incomes would be nearly 50% higher today! If this empirical result is not enough to convince one of the importance of productivity, we will also present an analytical result, the key finding of the Solow Model, being that in the long run, the per capita rate of consumption growth in an economy is determined solely by the rate of productivity growth.

As raising productivity growth should therefore be one of the basic goals of economic policy, to infer how such policy and changing economic circumstances effect productivity, we must first be able to rigorously and accurately measure it. To do so, we will employ three broad techniques: index numbers, econometrics, and non-parametric estimates of efficiency. Index number methods will be primarily concerned with aggregating outputs and inputs and measuring their relative growth. Such methods receive strong justification from microeconomics and will allow us to decompose real income growth into explanatory multiplicative factors, following Diewert and Morrison (1986). Econometric estimates involve statistically estimating various profit functions where we will concern ourselves particularly with *flexible functional forms*. It will be shown that such functional forms, which include the normalized quadratic, do not arbitrarily restrict elasticities, and have enough free parameters for us to impose curvature conditions on them so our estimated equations are consistent with microeconomic theory. These estimated equations can therefore be justified from and are consistent with first principles, in a way that other standard workhorse utility functions of macroeconomics like the CES are not. Our non-parametric estimates will follow in the work of Farrell (1957), exploiting the geometry of the situation and employing linear programming techniques to measure efficiency as distance from observed quantity data to the production possibilities frontier.

As a vindication of productivity measurement as a science, it will be shown that these three disparate methods all produce near identical estimates on average. As being able to approach a problem from different angles and arrive at the same conclusion is a hallmark of scientific rigor, we will be confident in claiming we have estimated the true level of US productivity growth over the sample period. The relative merits of each approach will be vigorously discussed, but this "all roads lead to Rome" result should be kept in mind throughout the paper.

A roadmap of the paper is as follows. Chapter 2 will describe the process of constructing and aggregating our data that is used in all subsequent estimation. We will briefly introduce index number theory, as well as the key findings of the Solow Model which will emphasize the importance of the rest of our work. With the Solow Residual we will then obtain our first approximate measure of productivity growth. Chapter 3 describes index number methods and our decomposition of real income growth into explanatory factors. We also briefly wade into the current intellectual debate between the likes of Gordon, Jorgenson, Brynjolfsson and others on the future of American productivity growth. Chapter 4 will be devoted to econometric estimates of flexible functional forms, with various results from basic microeconomic theory intertwined along the way. Lastly Chapter 5 presents the theory of linear programming and its implementation in our nonparametric estimation. The analytically minded reader is invited to read sections 2.3, 2.6, 4.1, and 5.1 at their leisure, which are all self-contained. The historically minded reader is invited to read sections 3.2 and 4.4.2, which delve into the history (and future) of American productivity growth and consumer trends in more detail.

2 Data construction

Here we describe the process of constructing our data which will be used in our index number, econometric, and non-parametric estimates of total factor productivity, as well as our detailed consumer model. We will juxtapose the theory of our data construction alongside exposition of the broad trends or interesting facts in the data we initially observe as this will provide a sanity check on our methodology and may also later help explain some seemingly anomalous calculations. The main sources used will be the Organization for Economic Co-operation and Development online repository (OECD.stat) where national accounts statistics initially constructed by the Bureau of Economic Analysis (BEA) are compiled, pre-1970 data from OECD Nation Accounts: Main Aggregates (1999), and the Federal Reserve at St. Louis Economic Data Base (FRED), which will primarily be used to extrapolate our national account series from 2016 to 2017. The main series we will concern ourselves with constructing is GDP from the expenditure approach ($Y = C + I + G + X - M$) and GDP from the income approach (by constructing price and quantity series for capital and labour income). Along the way we will compile population statistics, impute effective tax rates, and construct series for depreciation, inflation, and interest rates. Lastly, we present the theory of the Solow model, which yields the conclusion that long run per capita rates of consumption are determined solely by the rate of technical progress. This result therefore provides immense justification for the importance of TFP measurement. We will then compute the Solow Residual, which will be seen to be a good approximation to our other more rigorous methods of TFP measurement.

2.1 Output Series

We begin with the expenditure approach to GDP. Quantity series are read in at both constant dollar (OECD 2010 base year) and current year prices. A prefix Q denotes a constant dollar or real variable, and a prefix V denotes a current price or nominal variable. Pre 1970 series come from OECD Nation Accounts: Main Aggregates (1999), and our 2017 observation is extrapolated by taking the growth rate of the corresponding variable from FRED, and applying it to the 2016 OECD.stat observation. The justification for doing so is that although the FRED series differ slightly from the OECD series, their growth rates are almost identical, so this allows us to extend all series through 2017. Once we have real and nominal quantity series, we can deflate nominal variables by real ones to get their corresponding prices.

- QC, VC: Household Consumption Expenditure
- QG, VG: Government Consumption Expenditure
- QI1, VI1: Gross Fixed Capital Formation
- VI2: Changes in Inventory (only nominal)

- QX, VX: Exports of Goods and Services
- QM, VM: Imports of Goods and Services

and the corresponding prices are constructed as:

$$PC = VC/QC \qquad PG = VG/QG \qquad PI = VI1/QI1$$
$$PX = VX/QX \qquad PM = VM/QM$$

with a price for investment we could then deflate our changes in inventories to get:

$$QI2 = VI2/PI$$

and then add real gross fixed capital formation to real changes in inventories to get real investment:

$$QI = QI1 + QI2$$

Next we document how the hard copy pre 1970 data was linked to the OECD.stat post 1970 data. We will show the construction just for consumption as all other variables followed the same procedure. OECD Nation Accounts: Main Aggregates (1999) contains volume indices and price indices. We first construct real quantities as:

$$QC_i = \frac{CVOL_i}{CVOL_{1970}} * QC_{1970} \qquad \text{i=1960,...,1970}$$

where CVOL is our final consumption expenditure volume index. This method of linking has the obvious benefits of preserving growth rates and making sure our 1970 observation remains unchanged. We then generate the value index:

$$CVAL_i = CVOL_i * PIC_i \qquad \text{i=1960,...,1970}$$

where PIC_i is our price index for final consumption expenditure from the hard copy, and now we can link our nominal series in the same fashion that we did with the real series above:

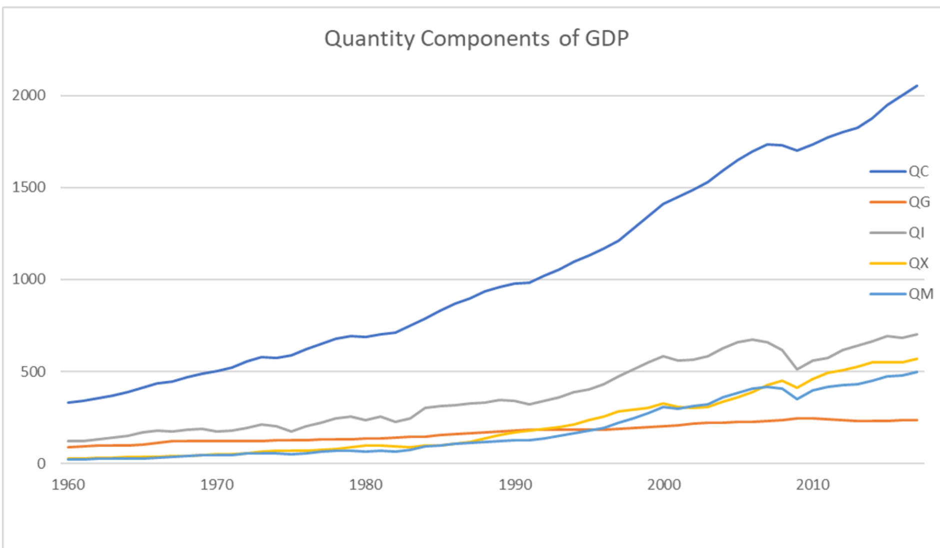
$$VC_i = \frac{CVAL_i}{CVAL_{1970}} * VC_{1970} \qquad \text{i=1960,...,1970}$$

and lastly deflate our nominal our nominal variable by its real quantity to get prices:

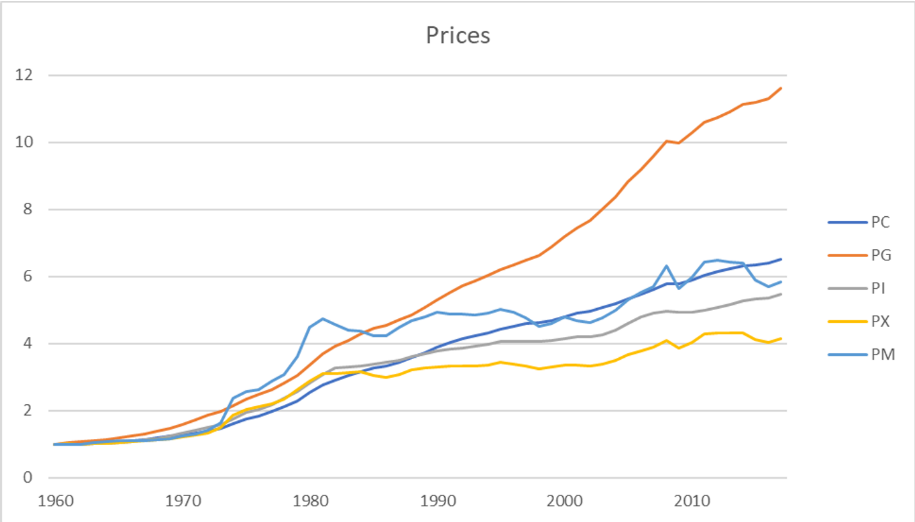
$$PC_i = VC_i/QC_i \qquad \text{i=1960,...,1970}$$

The same methodology was applied to link our series for government, investment, exports and imports. For inventory changes the hard copy data had nominal values, $VI2$ which were then used to construct real values by taking $QI2 = VI2/PI$ and finally we let $QI = QI1 + QI2$ as before.

Next we plot the real quantity components of GDP and their corresponding prices. The full data series these plots represent can be found in the appendix. Note due to a normalization the scale for our quantities figure is in millions USD. It is seen that real government output in the US is quite low. This is because the US federal government is quite limited in scope in comparison to other advanced economies. Paul Krugman and other economists often refer to the US federal government as "an insurance corporation with an army", meaning they are responsible for Social Security (pensions), Medicare (healthcare for the elderly), and the military, and not much else. The social safety net is relatively modest, which can be seen by government spending (automatic stabilizers) not picking up significantly in 2008. We also see that investment is the most volatile component of GDP as we would expect, with sharp dips occurring in 1973, 1990, 2008, and it is also interesting to note the surge in investment in the early 80's after inflation was finally tamed and private enterprise faced a more certain business environment.



For the price components of GDP, we see that price of government has risen by far the most, almost 12 fold over the sample period. This is because we do not observe government output, so we need to value it at input, and it's input is comprised mostly of wages which have experienced significant inflation over the period, as we will soon see. Also in our prices figure, we see the oil shocks of the 70s show up very saliently in our import prices series. In 1973 the Arab nations of OPEC enacted an embargo against western nations including the US who had supported Israel during the Yom Kippur War, causing the price of oil more than quadruple in a time span of less than a year. As oil was such a large proportion of US imports at the time, this caused the price of imports to surge 50% in one year. The second oil shock showing up prominently in the data happened in 1979-80 due to the Iranian Revolution and then the Iran-Iraq war. It is also seen that the price of imports dropped substantially in 2008, when oil fell from almost \$150 to under \$50 a barrel, so oil prices still figure prominently into USA import prices, although in a more muted manner than that of the 70s and 80s due to the USA consuming a broader array of energy sources and imported goods in general, as well as greatly increased domestic oil production over the sample period.



Next, we construct the terms of trade as simply

$$TOFT = \frac{PX}{PM}$$

There are two general facts about the terms of trade that are worth mentioning here. The first is that a country with favourable TOFT (defined as $TOFT > 1$) may be able to enjoy a higher standard of living (consume more) than its productive capacity would otherwise indicate. The reason for this is simply because they are selling to the rest of the world at a higher price than they are buying. This phenomenon is exemplified by certain Middle Eastern oil exporting countries who have faced incredibly favorable terms of trade due to high oil prices, coupled with the fact that exports and imports make up an outsized portion of their economies. From our calculation we see that USA terms of trade declined dramatically in the 70s due to the two oils shocks, and have remained poor ever since. As a result we can say that USA GDP has been adversely effected by unfavourable terms of trade over the sample period. Lastly we note that for the USA $TOFT$ is actually an endogenous variable. That is changes in USA consumption or production are so substantial that they will move global prices and thus alter the terms of trade it faces, which is not the case for smaller countries or developing nations where the terms of trade is treated as exogenous.

2.2 Input Series

Here we describe the construction of our series for inputs, capital and labour, over the sample period. A failing of our model is that we will not include land or natural capital as an input, because either the data is not readily available, or constructing prices for these series in an accurate and reliable manner is incredibly complex and still an ongoing problem in measurement economics.

Depreciation estimates from 1970 onwards were obtained from OECD.stat as Consumption of Fixed Capital. We assume that depreciated capital can be replaced 1 for 1 with new capital at the same cost as new investment and thus set the price of depreciation equal to the price of investment,

$$PD = PI$$

For pre-1970 depreciation we needed to read in a nominal value for Consumption of Fixed Capital from the hard copy OECD Nation Accounts: Main Aggregates (1999), and we linked it to the newer data by setting

$$VD_i = \frac{VDN_i}{VDN_{1970}} * VD_{1970} \qquad \text{i=1960,...,1970}$$

where again this linking procedure has the benefits of preserving growth rates and leaving our 1970 observation unaltered. We can simply then get real depreciation by deflating our nominal series by prices,

$$QD = \frac{VD}{PD}$$

Before discussing our depreciation series and the capital-output ratio that it will help generate, we need to aggregate our quantity series to GDP, for which we will use *index numbers.*

2.3 An Interlude on Index Number Theory

Now that we have constructed price and quantity components of GDP, it is natural to ask how these can be aggregated to GDP. For this task (and others) we will use index numbers. The main purpose of index numbers is to summarize large amounts of information on prices and quantities in an efficient and accurate manner. In this light index numbers can be thought of as a descriptive statistic. Index numbers are pervasive in economics, specifically at national statistical agencies and central banks, as their main purposes are to decompose value flows in the national accounts into price and quantity components, and as general measures of price inflation in the economy, although there are countless other uses of index numbers, such as in productivity measurement as we will soon see. Here we will very briefly develop the theory and highlight a few main results required to read the rest of this paper, relying heavily on and borrowing freely from the estimable lectures notes of Diewert (2018).

As a warm up we begin by considering the levels version of the index number problem, and show that it quickly leads to a fruitless end, which will then set the stage for the economic approach to bilateral index numbers. Indeed, define $p^t = (p_1^t, ..., p_N^t)$ as the period t vector of strictly positive prices over N commodities, and $q^t = (q_1^t, ..., q_N^t)$ as the period t quantity vector. We can then form the value aggregate

$$V^t = p^t \cdot q^t = \sum_{i=1}^{N} p_n^t q_n^t \tag{1}$$

and our index number problem is thus to decompose this value aggregate into an aggregate period t price level P^t and an aggregate period t quantity level Q^t such that

$$V^t = P^t Q^t \tag{2}$$

We assume that the aggregate period t price level is solely a function of period t prices, so we can write $P^t = c(p^t)$, and that Q^t is solely a function of period t quantities, $Q^t = f(q^t)$ substituting these functions into our formula for the value aggregate, and then equating this with (1) we get

$$V^t = f(q^t)c(p^t) = \sum_{i=1}^{N} p_n^t q_n^t \tag{3}$$

As our value aggregate is a positive number by definition, it seems reasonable to assume that $f(q) > 0$ and $c(p) > 0$ for all $q >> 0_N$ and $p >> 0_N$ where we have now dropped time superscripts. If we let $\mathbf{1}_N$ denote the vector of 1s in $\mathbb{R}^N$, then we must necessarily have $c(\mathbf{1}_N) = a$ for some $a \in \mathbb{R}$ and $f(\mathbf{1}_N) = b$ for some $b \in \mathbb{R}$. Although these assumptions seem quite minimalist, we have already imposed too much structure on the problem for it to be tractable, and this is Eichorn's (1978) Impossibility Theorem.

Theorem: If $N > 1$, there cannot exist positive functions c and f that satisfy (3).

Proof: We begin by substituting $p = \mathbf{1}_N$ into the right-hand equality of (3) to get

$$f(q)c(\mathbf{1}_N) = \sum_{i=1}^{N} (1) q_n \tag{4}$$

$$f(q) = \frac{1}{a} \sum_{i=1}^{N} q_n \tag{5}$$

where again we have assumed that $c(\mathbf{1}_N) = a$, and note that the above equation (5) must hold *for all* $q >> 0_N$. Similarly, we can substitute $q = \mathbf{1}_N$ into the right hand equality of (3) to get:

$$c(p) = \frac{1}{b}\sum_{i=1}^{N} p_n \tag{6}$$

Where again we note that (6) must hold *for all* $p >> 0_N$. Now recalling that $c(p)f(q)$ equals our value aggregate $\sum_{i=1}^{N} p_n q_n$, combining (5) and (6) yield

$$\left(\sum_{i=1}^{N} \frac{q_n}{a}\right)\left(\sum_{i=1}^{N} \frac{p_n}{b}\right) = \sum_{i=1}^{N} p_n q_n \tag{7}$$

But now the absurdity is at hand, as clearly (7) cannot hold *for all* positive price and quantitiy, $p >> 0_N, q >> 0_N$ vectors when $N > 1$, contradiction.

Thus we cannot decompose our value aggregate into a price aggregate which is purely a function of prices and a quantity aggregate which is purely a function of quantities. Although this is a purely mathematical result, it confirms our economic intuition that we should not be able to specify prices independently of quantities, and that they should be linked by some economic model of optimizing behavior, as prices and quantities are determined jointly in equilibrium. This provides justification for an economic approach to index number theory. Thus, we will now restrict ourselves to the bilateral framework in which we want to decompose a *value ratio* across the two periods into a price change component P times a quantity change component Q, that is we need two functions of $4N$ variables such that

$$\frac{p^1 \cdot q^1}{p^0 \cdot q^0} = P(p^0, p^1, q^0, q^1)Q(p^0, p^1, q^0, q^1)$$

and note that once we have determined a functional form for P, this uniquely determines the functional form for Q. Now we will briefly describe some of the common formula for $P(p^0, p^1, q^0, q^1)$ that will appear in this paper, how they are arrived at and some important facts pertaining to them.

One of the earliest and most common methods for determining relative price change across two periods is to take the fixed basket approach, where we fix our basket of commodities q,and then compare the cost of purchasing the basket at period 1 prices versus period 0 prices, and this provides us with an estimate of the increase in the price level across the two periods. For an arbitrary q this defines the Lowe Index:

$$P_L(p^0, p^1, q) = \frac{p^1 \cdot q}{p^0 \cdot q} \tag{8}$$

As our q was arbitrary, we could clearly take it to be the period 0 quantity vector or the period 1 quantity vector. These define the Laspeyres and Paasche price indices.

$$P_L(p^0, p^1, q^0, q^1) = \frac{p^1 \cdot q^0}{p^0 \cdot q^0} \tag{9}$$

$$P_P(p^0, p^1, q^0, q^1) = \frac{p^1 \cdot q^1}{p^0 \cdot q^1} \tag{10}$$

Now define the period t expenditure share to be $s_n^t = \frac{p_n^t q_n^t}{p^t \cdot q^t}$ Writing the numerator of (9) in summation formula and and then multiplying and dividing this through by p_n^0 yields:

$$\begin{aligned} P_L &= \frac{\sum_{n=1}^N p_n^1 q_n^0}{p^0 \cdot q^0} \\ &= \frac{\sum_{n=1}^N (\frac{p_n^1}{p_n^0}) p_n^0 q_n^0}{p^0 \cdot q^0} \\ &= \sum_{n=1}^N s_n^0 (\frac{p_n^1}{p_n^0}) \end{aligned}$$

Thus we have expressed the Laspeyres price index as a share weighted *arithmetic mean* of the price relatives across the two periods. In a similar calculation, we can invert the fraction in (10), then multiply and divide the numerator through by p_n^1, yielding:

$$\begin{aligned} P_P &= \left(\frac{p^0 \cdot q^1}{p^1 \cdot q^1}\right)^{-1} \\ &= \left(\frac{\sum_{n=1}^N p_n^0 q_n^1}{p^1 \cdot q^1}\right)^{-1} \\ &= \left(\frac{\sum_{n=1}^N (\frac{p_n^0}{p_n^1}) p_n^1 q_n^1}{p^1 \cdot q^1}\right)^{-1} \\ &= \left(\frac{\sum_{n=1}^N (\frac{p_n^1}{p_n^0})^{-1} p_n^1 q_n^1}{p^1 \cdot q^1}\right)^{-1} \\ &= \left(\sum_{n=1}^N s_n^1 (\frac{p_n^1}{p_n^0})^{-1}\right)^{-1} \end{aligned}$$

Thus we have expressed the Paasche price index as a share weighted *harmonic mean* of the price relatives across the two periods. Now, recall the definition of the weighted mean of order r:

$$M_r(x) = \left(\sum_{n=1}^N \alpha_n x_n\right)^{1/r}$$

where the α_i are positive and sum to 1. Schlömilch's inequality then says that the mean of order r is an increasing function of r, so an application of this is the harmonic mean ($r = -1$) is less than the arithmetic mean ($r = 1$). Lastly noting that as expenditure shares clearly sum to 1, $\sum_n p_n^t q_n^t / p^t \cdot q^t = 1$, we can apply this result to our share weighted averages for the Paasche and Laspeyres above. Thus if $s_n^0 = s_n^1$ (expenditure shares are equal across the two periods), Schlömilch's inequality says:

$$P_L(p^0, p^1, q^0, q^1) > P_P(p^0, p^1, q^0, q^1) \tag{11}$$

if the prices are not constant across the two periods. More generally, as shown in Balk(2008) if $s_n^0 \neq s_n^1$ but the covariance of relative price changes and relative quantity changes is negative, then we also get the same inequality above. Although we have derived this result without any appeal to economics, our economic intuition confirms it: if we fix our commodity basket to q^0, we will overstate the pure price increase across the two periods because we are not accounting for the fact that people will substitute away from the more expensive goods into cheaper goods in q^1. Conversely, if we restrict ourselves to q^1, then we are not accounting for the fact that people have substituted into cheaper goods in period 1, and so we are probably understating the true level of price increase. This logic suggests that the true pure level of price increase lies between P_L and P_P. It is then natural to take a symmetric average of the Laspeyres and Paasche, which if we choose this average to be the geometric mean, this defines the Fisher Ideal Index:

$$P_F = \left[\left(\frac{p^1 \cdot q^0}{p^0 \cdot q^0} \right) \left(\frac{p^1 \cdot q^1}{p^0 \cdot q^1} \right) \right]^{1/2} \tag{12}$$

The list of reasons for using Fisher Ideal indexes is expansive and impressive, so we will mention some of these briefly here.

In the economic approach to index number theory the starting point is often assuming a representative consumer solves the following cost minimization problem:

$$C(u^t, p^t) = \min_q \{p^t \cdot q \mid f(q) = u^t\}$$

where $f(q)$ is a continuous, increasing, quasi concave utility function. For a general utility level u and price vector p these define the expenditure function $C(u, p)$. It is then natural to define the *family* of Konus true cost of living indices

$$P_K(p^0, p^1, q) = \frac{C(f(q), p^1)}{C(f(q), p^0)}$$

which is the ratio of minimum costs of achieving the same level of utility u = f(q) across the two periods. It is then easy to show that $P_K(p^0, p^1, q^0) \leq$

$P_L(p^0, p^1, q^0, q^1)$ and $P_K(p^0, p^1, q^1) \geq P_P(p^0, p^1, q^0, q^1)$. Indeed,

$$\begin{aligned} P_K(p^0, p^1, q^0) &= \frac{C(f(q^0), p^1)}{C(f(q^0), p^0)} \\ &= \frac{C(f(q^0), p^1)}{p^0 \cdot q^0} \\ &\leq \frac{p^1 \cdot q^0}{p^0 \cdot q^0} \\ &= P_L(p^0, p^1, q^0, q^1) \end{aligned}$$

where the inequality above follows becuase q^0 is clearly feasible for the optimization problem $C(f(q^0), p^1)$, but not necessarily optimal, implying $C(f(q^0), p^1) \leq p^1 \cdot q^0$. Similarly,

$$\begin{aligned} P_K(p^0, p^1, q^1) &= \frac{C(f(q^1), p^1)}{C(f(q^1), p^0)} \\ &= \frac{p^1 \cdot q^1}{C(f(q^1), p^0)} \\ &\geq \frac{p^1 \cdot q^1}{p^0 \cdot q^1} \\ &= P_P(p^0, p^1, q^0, q^1) \end{aligned}$$

Where again the inequality follows because q^1 is clearly feasible for the problem $C(f(q^1), p^0)$, but not necessarily optimal. Konus (1939) then proved the following remarkable result, which we will prove here because its proof technique has recently re-emerged in the literature, see Feenstra, Inklaar and Timmer (2016).

Theorem: There exists a q^* which is a convex combination of q^0 and q^1, $q^* = \lambda^* q^0 + (1 - \lambda^*) q^1$, $\lambda^* \in [0, 1]$ such that the unobservable true cost of living index $P_K(p^0, p^1, q^*)$ lies between $P_P(p^0, p^1, q^0, q^1)$ and $P_L(p^0, p^1, q^0, q^1)$.

Proof: Let $\lambda \in [0, 1]$. Define the auxillary function

$$g(\lambda) = P_K(p^0, p^1, (1 - \lambda) q^0 + \lambda q^1)$$

We note a few facts about $g(\lambda)$. The first is that it is *continuous* on $[0, 1]$. This follows since it is a ratio of cost functions, and cost functions are (minimum) value functions which are continuous by the Theorem of the Maximum under some mild regularity conditions which hold in an economic context. Next, from our results above we note that

$$\begin{aligned} g(1) &\geq P_P \\ g(0) &\leq P_L \end{aligned}$$

Now assume $P_L \geq P_P$. Combining this inequality with the above two, combinatorially we see that one of the following must hold:

$$g(0) \leq P_P \leq P_L \leq g(1)$$
$$g(0) \leq P_P \leq g(1) \leq P_L$$
$$P_P \leq g(0) \leq P_L \leq g(1)$$
$$P_P \leq g(1) \leq g(0) \leq P_L$$
$$P_P \leq g(0) \leq g(1) \leq P_L$$

Since g is continuous as mentioned above, it assumes all intermediate values on $[0,1]$. Thus, the above inequalities all show by the Intermediate Value Theorem, we *must have* $\lambda^* \in [0,1]$ such that $P_p \leq g(\lambda^*) \leq P_L$. Now suppose $P_L \leq P_P$. Given our constraints on $g(1)$ and $g(0)$ then it must be the case that

$$g(0) \leq P_L \leq P_P \leq g(1)$$

But again, $g(\lambda)$ is continuous on $[0,1]$ so it achieves all intermediate values, and thus by the intermediate value theorem there exists λ^* such such $P_L \leq g(\lambda^*) \leq P_P$

To see how this result justifies the use of the Fisher Ideal index, note that if the true cost of living index lies *between* P_P and P_L, then taking a symmetric average of the two, which necessarily lies between them, is probably a good approximation to the true cost of living index, and the natural symmetric average to take of P_L and P_P is the geometric mean, which gives us the Fisher Ideal.

We now show that if we are to explicitly assume the consumers cost function is a certain *flexible functional form*, then the Fisher Ideal is equal to *exactly* the ratio of costs across the two periods. Indeed suppose that the consumer's unit cost function is a homogeneous quadratic, that is

$$c(p) = [p^T Bp]^{1/2}$$

where $B = B^T$. Differentiating c(p) yields:

$$\nabla c(p) = (1/2)(p^T Bp)^{-1/2}(2Bp) = \frac{Bp}{c(p)}$$

where we have used the symmetry condition $B = B^T$. Lastly recall *Shepherd's Lemma* which says (when f(q) is linearly homogenous):

$$\frac{q^t}{p^t \cdot q^t} = \frac{\nabla c(p^t)}{c(p^t)}$$

so applying this formula to our unit cost function we see

$$\frac{q^t}{p^t \cdot q^t} = \frac{Bp}{c(p^t)^2} \tag{13}$$

Substituting these facts into our Fisher Ideal price index we see:

$$\begin{aligned}
P_F &= \left[\left(\frac{p^1 \cdot q^0}{p^0 \cdot q^0} \right) \left(\frac{p^1 \cdot q^1}{p^0 \cdot q^1} \right) \right]^{1/2} \\
&= [p^1 \cdot (q^0/p^0 \cdot q^0)/p^0(q^1/p^1 \cdot q^1)]^{1/2} \\
&= [(p^1 \cdot (Bp^0/c(p^0)^2))/(p^0 \cdot (Bp^1/c(p^1)^2)]^{1/2} \qquad \text{applying Eq(13)} \\
&= \left(\frac{c(p^1)}{c(p^0)} \right) \left(\frac{p^{1T}Bp^0}{p^{0T}Bp^1} \right)
\end{aligned}$$

and lastly note that since B is a symmetric matrix and $p^{1T}Bp^0$ is a scalar we have $p^{1T}Bp^0 = (p^{1T}Bp^0)^T = p^{0T}B^Tp^1 = p^{0T}Bp^1$ so $\left(\frac{p^{1T}Bp^0}{p^{0T}Bp^1} \right) = 1$ and thus

$$P_F(p^0, p^1, q^0, q^1) = \frac{c(p^1)}{c(p^0)} \tag{14}$$

Thus, the Fisher ideal price index P_F is *exactly* equal to the true price index $c(p^1)/c(p^0)$! Hence we have shown the Fisher Ideal to be a Superlative Index Number, a term famously coined by Diewert (1976).

By now it should be obvious that the Fisher Ideal Index receives incredibly strong justification from the economic approach. It is also worth mentioning that in a separate approach to index number theory (the test or axiomatic approach) the Fisher Ideal also performs marvelously. This approach involves examining whether our bilateral index number formula $P(p^0, p^1, q^0, q^1)$ satisfies certain reasonable properties we would like a price index to possess. Some examples of early axioms are whether $P(p^0, p^1, q^0, q^1)$ is a strictly positive function, if its continuous in all $4N$ variables, if it is equal to unity when prices are constant across the two periods, that is whether $P(p, p, q^0, q^1) = 1$, and is it linearly homogeneous in period 1 prices, so that $P(p^0, \lambda p^1, q^0, q^1) = \lambda P(p^0, p^1, q^0, q^1)$ for all $\lambda > 0$. A very important axiom which is a stumbling block for many commonly used indexes is the *time reversal test*, which can be stated as the property:

$$P(p^0, p^1, q^0, q^1) = \frac{1}{P(p^1, p^0, q^1, q^0)}$$

This says that if we swap all period 0 and 1 data, the price index that is formed with it is the reciprocal of the original index. To see the fundamental nature of this test, consider the one good case where our price index from period 0 to 1 will just be p_1/p_0. If we swap these prices across periods, we would now like our index to be p_0/p_1, which is exactly the reciprocal of the aforementioned index. There are two other reversal tests, the quantity reversal test:

$$P(p^0, p^1, q^0, q^1) = P(p^0, p^1, q^1, q^0)$$

and the price reversal test:

$$\left((p^1 \cdot q^1)/(p^0 \cdot q^0)\right) \Big/ P(p^0, p^1, q^0, q^1) = \left((p^0 \cdot q^1)/(p^1 \cdot q^0)\right) \Big/ P(p^1, p^0, q^0, q^1)$$

where it is seen that if we use the implicit quantity index the above involves simply checking $Q(p^0, p^1, q^0, q^1) = Q(p^1, p^0, q^0, q^1)$, that is if we swap period 0 and 1 prices our quantity index remains unperturbed.

It is an incredible fact that not only does the Fisher ideal pass all 20 generally agreed upon axioms (see Diewert 2009 for the full list), but they actually uniquely *characterize* the Fisher, in so far as the Fisher is the *only* index number formula to pass all tests. The proof of this involves simply re-arranging the reversal tests above, and showing that a bilateral index number formula which passes all 3 must indeed be the Fisher. This results synthesizes our intuition from the economic approach, and the more abstractly mathematical axiomatic approach, that the Fisher Ideal is the "correct" way to measure price increase across two periods.

The other index number formula which will be a major player in the aggregation done in this paper is the Törnqvist-Thiel. It is expressed as

$$\log P_T(p^0, p^1, q^0, q^1) = \sum_{n=1}^{N} \frac{1}{2}(s_n^0 + s_n^1) \log\left(\frac{p_n^1}{p_n^0}\right)$$

or by taking anti-logs as:

$$P_T(p^0, p^1, q^0, q^1) = \prod_{n=1}^{N} \left(\frac{p_n^1}{p_n^0}\right)^{\frac{1}{2}s_n^0 + \frac{1}{2}s_n^1}$$

where the above formula for $\log P_T(p^0, p^1, q^0, q^1)$ is often justified through the so called stochastic approach. That is, if we were to draw price relatives at random in such a way that each dollar of base period expenditure share has an equal chance of being selected, then the overall mean logarithmic price chance is $\sum_{n=1}^{N} s_n^0 \log(\frac{p_n^1}{p_n^0})$. Note, this is a true mean since our s_n^0 are positive and sum to 1, they are indeed a *discrete probability measure.* We could also weight log price relatives by period 1 expenditure share, giving us mean price change of $\sum_{n=1}^{N} s_n^1 \log(\frac{p_n^1}{p_n^0})$ As each of these overall measures of log price change are equally valid, we then take their arithmetic average which defines the Törnqvist-Thiel above.

Aside from being derived from and the obvious conclusion of the stochastic approach, the Törnqvist-Thiel possesses numerous desirable properties. Indeed, it passes the time reversal test:

$$\begin{aligned} P_T(p^1, p^0, q^1, q^0) &= \prod_{n=1}^{N} \left(\frac{p_n^0}{p_n^1} \right)^{\frac{1}{2}s_n^0 + \frac{1}{2}s_n^1} \\ &= \frac{1}{\prod_{n=1}^{N} \left(\frac{p_n^1}{p_n^0} \right)^{\frac{1}{2}s_n^0 + \frac{1}{2}s_n^1}} \\ &= \frac{1}{P_T(p^0, p^1, q^0, q^1)} \end{aligned}$$

and is linearly homogeneous in period 1 prices:

$$P_T(p^0, \lambda p^1, q^0, q^1) = \lambda P_T(p^0, p^1, q^0, q^1)$$

It also turns out that the Törnqvist-Thiel is a *superlative index number formula*, as defined above. Indeed, it is exact for the *translog function*, a result which we will hold off on until chapter 3. It is also worth mentioning that the Törnqvist-Thiel is the discrete time approximation to the continuous time Divisia index, although we will not delve into continuous time indexes here. Lastly, it is worth mentioning that the Fisher and Törnqvist-Thiel approximate each other to the 2^{nd} order around an equal price and quantity point.

Theorem: The Paasche, Laspeyres, Fisher and Törnqvist-Thiel all approximate each other to the 1^{st} order around an equal price and quantity point. That is,

$$\nabla P_P(p, p, q, q) = \nabla P_L(p, p, q, q) = \nabla P_F(p, p, q, q) = \nabla P_T(p, p, q, q)$$

$$= \begin{pmatrix} -q/p^T q \\ q/p^T q \\ 0_N \\ 0_N \end{pmatrix}$$

Theorem: The Fisher and Törnqvist-Thiel approximate each other to the 2^{nd} order around an equal price and quantity point. That is,

$$\nabla^2 P_F(p, p, q, q) = \nabla^2 P_T(p, p, q, q)$$

$$= \begin{pmatrix} \frac{2qq^T}{(p^T q)^2} & \frac{-qq^T}{(p^T q)^2} & -\frac{1}{2}\frac{I_N}{p^T q} + \frac{1}{2}\frac{qp^T}{(p^T q)^2} & \frac{1}{2}\frac{I_N}{p^T q} - \frac{1}{2}\frac{qp^T}{(p^T q)^2} \\ \frac{-qq^T}{(p^T q)^2} & 0_{NxN} & \frac{1}{2}\frac{I_N}{p^T q} - \frac{1}{2}\frac{qp^T}{(p^T q)^2} & \frac{1}{2}\frac{I_N}{p^T q} - \frac{1}{2}\frac{qp^T}{(p^T q)^2} \\ \frac{1}{2}\frac{qp^T}{(p^T q)^2} - \frac{I_N}{p^T q} & \frac{1}{2}\frac{I_N}{p^T q} - \frac{1}{2}\frac{qp^T}{(p^T q)^2} & 0_{NxN} & 0_{NxN} \\ \frac{1}{2}\frac{I_N}{p^T q} - \frac{1}{2}\frac{qp^T}{(p^T q)^2} & \frac{1}{2}\frac{I_N}{p^T q} - \frac{1}{2}\frac{qp^T}{(p^T q)^2} & 0_{NxN} & 0_{NxN} \end{pmatrix}$$

The above theorem is of enormous importance and we will see it time and time again throughout this paper. It states that using relatively smooth time series data, we should expect the difference between the Fisher and Törnqvist-Thiel to be negligible, thus for all intensive purposes we can compute whichever index we think best suits the situation, and know that the other will be nearby. As another consequence of this theorem, briefly recall the axiomatic approach described above. It can be shown that the Törnqvist-Thiel fails 9 out of 20 standard tests, but because of its second order approximation property to the Fisher, which passes all 20, we can conclude that the Törnqvist-Thiel must not fail any of these tests catastrophically. Indeed, we have now stated a lengthy list of reasons to use Törnqvist-Thiel index numbers, which will play a prominent role in this paper. As one final justification for using this index on the basis of an appeal to authority, we note that the US CPI is computed by a Törnqvist-Thiel index by the BLS.

We conclude our brief exposition on index number theory with a discussion of chain weighting. As it may be difficult to obtain current period quantity (or expenditure) information, it is then natural to calculate a price index such as the Laspeyres which fixes quantities at their period 0 level. The problem with such a calculation is that over long periods of time, the period t quantity basket may be wildly different from the period 0 quantity basket, due to substitution, changes in consumer trends, and various other factors. The result is then that the spread between $P_L(p^0, p^1, q^0, q^1)$ and $P_P(p^0, p^1, q^0, q^1)$ may be very large. Even if we are to assume that the true price index lies somewhere between P_L and P_P, if this difference is sufficiently large then taking a symmetric average of the Paasche and Laspeyres may still not be enough to bring us "close" to this true price index. A solution to this problem is via chain-weighting, where one period rates of change (chainlinks) are cumulated to yield relative price levels.

Definition: If our fixed based indexes for time t=0,1,2,... are:

$$1,\ P(p^0, p^1, q^0, q^1),\ P(p^0, p^2, q^0, q^2), ...$$

then the *chained indexes* for $t = 0, 1, 2, ...$ are:

$$1,\ P(p^0, p^1, q^0, q^1),\ P(p^0, p^1, q^0, q^1)P(p^1, p^2, q^1, q^2), ...$$

which by denoting our period t chain link as $P^t_{CHL} = P(p^{t-1}, p^t, q^{t-1}, q^t)$ can be defined succinctly as

$$P^t = P^t_{CHL}P^{t-1}$$

By cumulating these one period rates of change, we will reduce the spread between the Paasche and Laspeyres, bringing us closer to the "truth". The last question to answer then is when is it appropriate to chain.

Rule: One should chain if the prices and quantities pertaining to adjacent periods are *more similar* than the prices and quantities of distant periods.

This is the case if there are systematic trends in prices and quantities over long periods of time, such as national accounts price or quantity data at high levels of aggregation. It is not appropriate to chain when prices “bounce” or oscillate, which may be due to seasonality, or sales at lower levels of aggregation. As most of the data in this paper is from long run time series with smooth trends, we will be justified in habitually applying the chain principle.

2.4 Aggregation to GDP and Input Series Continued

We are now in a position to construct price and quantity series for GDP. Indeed, first we change the sign of our import quantity series for all observations, so they are now negative. We then form the price and quantity vectors

$$p^t = (PC^t, PG^t, PI^t, PX^t, PM^t)$$
$$q^t = (QC^t, QG^t, QI^t, QX^t, QM^T)$$

We first construct a *chained* Törnqvist-Thiel price index for GDP as:

$$PY_T = P_T(p^0, p^1, q^0, q^1)P_T(p^1, p^2, q^1, q^2)...P_T(p^{t-1}, p^t, q^{t-1}, q^t)$$

which implicity defines our quantity series for GDP:

$$QY_T = \left(\frac{p^t \cdot q^t}{p^0 \cdot q^0}\right)\bigg/ PY_T$$

We can also construct a chained Fisher index for our price series as

$$PY_F = P_F(p^0, p^1, q^0, q^1)P_F(p^1, p^2, q^1, q^2)...P_F(p^{t-1}, p^t, q^{t-1}, q^t)$$

and proceeding in the same manner construct chained Laspeyres and Paasche price indexes for GDP over this sample period. It will be illustrative of the theory presented in section 2.3 to examine the resultant (2017) values of these *chained* series. Indeed, consider the following table:

Chained Index	PY_T	PY_F	PY_P	PY_L
2017	6.783	6.783	6.786	6.780

Here we see our second order approximation property at work: The chained Fisher and chained Törnqvist-Thiel are equal to three decimal places! It is also worth noting that the “spread” between the Paasche and Laspeyres is extremely small, .006, and that the Fisher falls exactly between them. This example nicely illustrates the efficacy of chaining smooth trending long run time series data.

We can also calculate the Laspeyres, Paasche, Fisher and Törnqvist-Thiel without chaining (fixed base), simply applying the formulas:

$$PY_L(p^0, p^t, q^0, q^t) = \frac{p^t \cdot q^0}{p^0 \cdot q^0} \qquad PY_P(p^0, p^t, q^0, q^t) = \frac{p^t \cdot q^t}{p^0 \cdot q^t}$$

$$PY_F(p^0, p^t, q^0, q^t) = (PY_L PY_P)^{\frac{1}{2}} \quad PY_T(p^0, p^1, q^0, q^1) = \prod_{n=1}^{N} \left(\frac{p_n^1}{p_n^0}\right)^{\frac{1}{2}s_n^0 + \frac{1}{2}s_n^1}$$

Calculating these from our price and quantity series yields:

Index	PY_T	PY_F	PY_P	PY_L
2017	6.783	6.673	6.340	7.023

Our empirical software (Shazam) automatically chains the Törnqvist-Thiel so it is the same as above. Otherwise note that without chaining the Laspeyres (period 0 quantities) now has a significant upward bias as we would expect, and so we have the inequality $PY_L(p^0, p^t, q^0, q^t) \geq PY_P(p^0, p^t, q^0, q^t)$. Taking a symmetric average of PY_P and PY_L (the Fisher) brings us closer to our true level of price increase over the sample period (which we assume to be $PY_T =$ chained PY_F), but it is still "off" by .11. In comparing the two tables, we note that chaining has reduced the spread of the Paasche and Laspeyres from .683 to .003, an impressive feat!

The price and quantity series for GDP used throughout this paper will be our chained Törnqvist-Thiel, which for all intensive purposes is the same as using the chained Fisher ($PY_T = PY_F = 6.683$ in 2017, and $QY_T = 2862$ in 2017 whereas $QY_F = 2863$ in 2017, all chained). Proceeding with the Törnqvist-Thiel we can now construct nominal GDP for all years simply as

$$VY = PY_T \cdot QY_T$$

and going forward we will drop the T subscripts on PY_T and QY_T.

We will treat depreciation in the *geometric* framework. Our equation of motion for the capital stock is thus

$$QK_t = QK_{t-1} + QI_{t-1} - QD_{t-1} \tag{15}$$

Since this definition is recursive in K_t and we have depreciation and investment for all observations, we only need a beginning of period capital stock (1960) to construct capital stocks for all other years. Recall that one of Kaldor's Facts (1957) says the capital-output ratio should be stationary over time. A good estimate for this ratio is usually around 2.5, which we will take to be our starting capital output ratio, and then check that it remains roughly stationary over time. We can then construct our starting period capital stock by simply noting

$$QK_t = (2.5)QY_t \qquad \text{t=1960}$$

and then all subsequent K_t can be constructed by equation (15). We now set the price of beginning of year capital stock equal to the price of investment for the previous year, and for 1960 we set the price of QK at the beginning of 1960 equal to the price of investment

$$PK_t = PI_t \qquad \text{t=1960}$$
$$PK_t = PI_{t-1} \qquad \text{t=1961,...,2017}$$

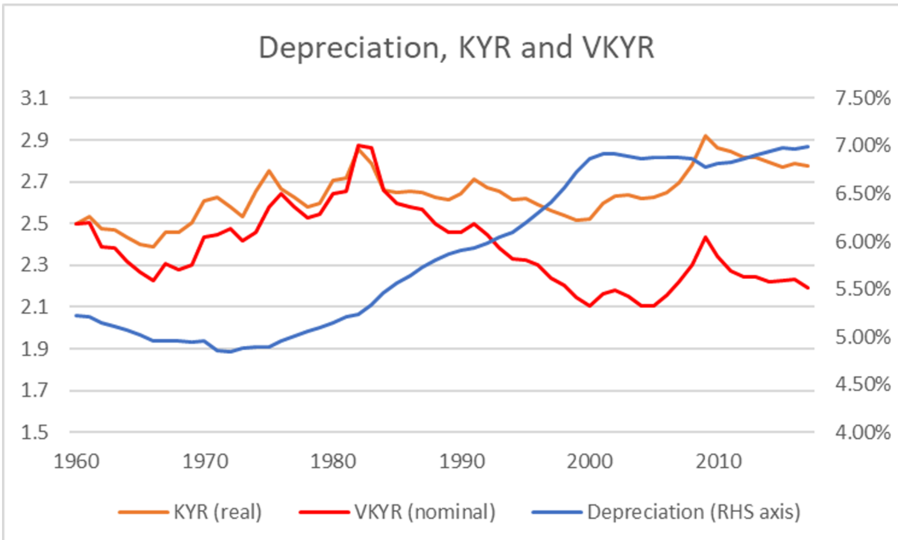
We take $VK = PK \cdot QK$ for all observations as our value of the capital stock. Lastly, we take our depreciation rate to be

$$D = \frac{QD}{QK}$$

and generate our real and nominal capital output ratios as

$$KYR = \frac{QK}{QY} \qquad \text{(real)}$$

$$VKYR = \frac{VK}{VY} \qquad \text{(nominal)}$$



for which we include the plot of these series as well as depreciation. We note our real capital output ratio, KYR, is roughly stationary around 2.5, with a minimum of 2.3875 and a maximum of 2.9191, thus we feel relatively justified in calibrating our intital capital output ratio to 2.5. If we run the simple linear regression

$$KYR_t = \beta t + \epsilon_t$$

we get $\beta = .0048$, so on average the real capital output ratio increased by .48% per year which reflects capital deepening. From the graph above, it is clear that most of this capital deepening happened from the mid 1960s to early 1980s, a fact we shall return to later. We note an upward trend in depreciation rates over the sample period, with its minimum observation being $D = 0.04841$ in 1972, and then steadily climbing over the sample period, to a maximum value in our final observation of $D = 0.06999$ in 2017. If we run the simple linear regression

$$D_t = \beta t + \epsilon_t$$

we see that $\beta = .000457$, or .045% per year. This upward trend in depreciation rates is due to increased investment in high depreciation machinery and equipment, and decreased investment in low depreciation structures over the sample period.

We now briefly turn our attention to labour input series. The relevant definitions we will work with are:

$$CE = WE + UP \tag{16}$$

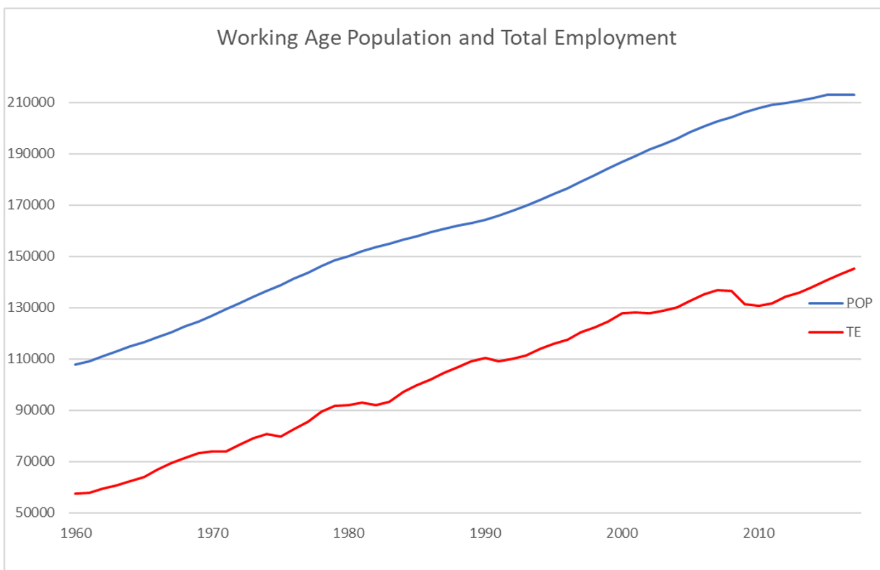
$$TE = WE + AF \tag{17}$$

$$LF = AF + U \tag{18}$$

CE	Civilian Employment
WE	Wage Earners
UP	Unpaid Family Workers
TE	Total Employment
AF	Armed Forces
LF	Labour Force
U	Unemployed

For constructing the series for Working Age Population (15-64) OECD.stat was missing the observations for 2015,2016 and 2017. The Federal Reserve has this data, but observations differed enormously from the OECD.stat data, by a magnitude of almost 10million in the year 2014. To confound matters the Federal Reserve lists the OECD.stat series as their source. My suspicion is that the larger series is making more of an effort to include undocumented workers who may not be picked up by the labour force surveys. In most countries measuring working age population should be incredibly easy, but in the USA there is a very large population of undocumented people ($\tilde{1}$5million) who largely live in the shadows. This presents an enormous problem for economic analysis for two reasons. The first being obviously that any per capita estimate is going to be incredibly sensitive to adding or subtracting 15million people. The second being that it builds in an enormous amount of heterogeneity to our model. Undocumented peoples are forced to work in lower wage jobs in sectors such as agriculture and construction, so their wage profiles will be very different from a typical American worker. Also in our tax imputations that we will soon calculate, many of these workers by definition will not be accounted for in payroll tax or federal income tax. Problems of economic analysis aside, it is my opinion that every effort should be made to include these workers in our statistics. The reason being is that having access to millions of people who are notoriously hard working and willing to work low wage service sector jobs has obviously been a massive boon for the USA economy, one that is too impactful to ignore, especially if the primary goal of this paper is to produce accurate productivity estimates. Lastly I will note the census claims to include all people living in the USA, and there is every incentive to do so as cities and municipalities are often awarded funding by number of residents (not necessarily citizens) but I am still somewhat skeptical of this data. This issue seems to be largely ignored by academic economists which is somewhat understandable given how politically charged it has become.

In the end I used the larger of the two series (OECD), and for the 2015, 2016, and 2017 observations I took growth rates from the FRED data and applied them to my series. These growth rates were very small due to large numbers of retiring baby boomers as well as lower population growth. The other imputations made in the labour force data include filling in armed forces data for 2014-2017. I assumed the armed forces to be constant over this period, which may be a tenuous assumption due to the 2013 Budget Sequestration in which the USA military faced large automatic cuts. The US military is large enough that this may not be an insubstantial error to make, although other errors we make will surely be larger. For the Self-Employed and Unpaid Family workers series I was missing 2017 observations, for which I found somewhat analogous series in FRED, so I applied their 2016-2017 growth rates to the OECD data to get the 2017 observation.



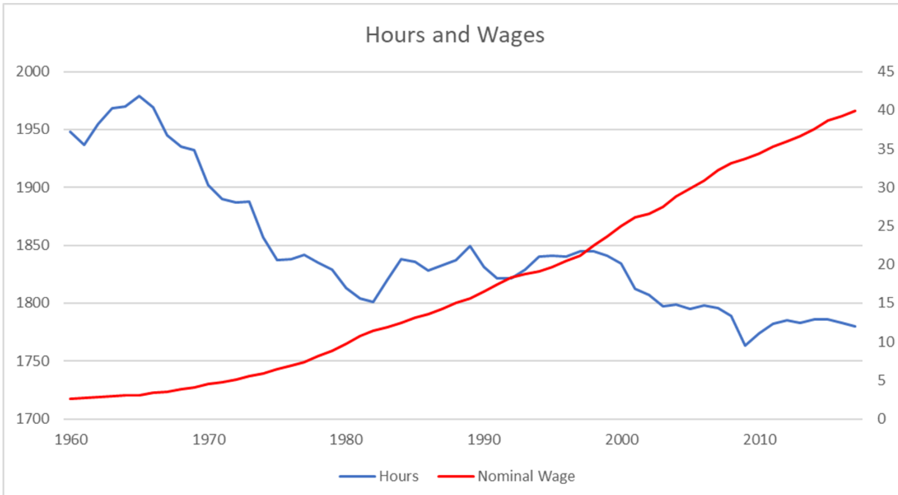
Here we include a plot of working age population and total employment. We see working age population has increased dramatically over the sample period, almost exactly doubling from 107million in 1960 to 213million in 2017. Although upward sloping there is a somewhat sinusoidal trend built in to this data due to the natural population cycle. In the USA this trend has been a massive explosion in population following World War II, the so-called Baby Boom, who in our data we can see entering working age population from the mid 60s onwards. Next there was Generation X, who were early children of baby boomers or late children of the great generation, which we can see entering working age population from the early 80s onwards. Lastly we have millenials who can be seen entering working age population from the mid 00's onwards. It is also worth noting that from 2015 onwards our POP series begins to flat line. This is primarily due to there not being enough millenials entering the labour force and immigration to replace the massive amount of retiring baby boomers. It is understood and well documented that demographic problems (an aging population) will be one of the great challenges for the US economy going forward, see Furman (2017). We will later show that labour input growth has been instrumental to GDP growth over our sample period, which supports this claim. It also illustrates how naive various recent bills by the US Senate and Congress to greatly restrict immigration, such as Tom Cotton's RAISE Act are. This bill if passed will reduce legal immigration into the USA by almost 50%, which as we will see even from analysis as cursory as carried out in this project, would have a devastating impact on the US economy. President Trump has also proposed equally naive, shortsighted and tribalist policies on immigration. US economic performance has been very strong over this sample period, but this may change in the future if the US working age population begins to flatline or recede due to poor policy choices. The bright red line illustrates total employment and moves 1 for 1 with the business cycle over the sample period. It's most severe dip occurs in 2008 in which over 5million jobs were lost.

Next we construct our series on average annual hours worked. We can then multiply our Total Employees (TE) series by hours to get a quantity series for employee labour. We also generate a quantity series for the self employed and unpaid family workers while we are here:

$$QE = (TE)(Hours)/1000000$$
$$QSE = (SE)(Hours)/1000000$$
$$QUP = (UP)(Hours)/1000000$$

and note that QE is thus measured in billions of hours. We can then construct an hourly wage rate series using our earlier Value of Earnings (VE) series simply as

$$PE = \frac{VE}{QE}$$



We include a plot of average annual hours worked and this nominal wage rate. Note from 1965 to 1980 hours worked underwent a dramatic decline, falling from 1975 hours to 1800. Since 1980 average annual hours worked has been more stable, but it did dip substantially in 2008, part of which may be attributed to the well documented trend in more businesses trying to shift employees from full time to part time since the Great Recession of 2008. The hour series then recovered and began to fall again after 2010, which may be due to bizzare tax incentives created by "Obamacare" in which private employers began to shift employees onto part time employment so they were not obligated to pay for their health insurance under the Obamacare Mandate. We see that nominal wages increased from \$2.6 per hour in 1960 to \$40 per hour in 2017, although much of this is obviously inflation and real wage growth has been much more muted over the sample period as we will soon see.

Lastly, we need to form a labour input aggregate which will be used in subsequent computations. As we have constructed quantity series for wage earners (QE), self-employed (QSE), and unpaid family workers (QUP) and a price series for wage earners (PE), we only need a price series for the self-employed and unpaid family workers. As the data for value of earnings for these types of workers does not exist, we will need to make a few assumptions. In particular, we will assume that the self-employed earn 2/3 as much as wage earners, and unpaid family workers earn 1/3 as much as wage earners. This is somewhat of an arbitrary assumption, but the quantity series for unpaid family workers and self employed are very small relative to the wage earners series (<10%) so it will not cause any major distortions even if incorrect. Thus we set

$$PSE = \frac{2}{3}PE$$
$$PUP = \frac{1}{3}PE$$

For each period t we can now form the 3x1 price and quantity vectors as $p^t = (PE^t, PSE^t, PUP^t)$ and $q^t = (QE^t, QSE^t, QUP^t)$ and as the data is annual and near observations are more alike than distant observations, we will use a chained Fisher to aggregate, where the period t index is defined as

$$P_F(p^{t-1}, p^t, q^{t-1}, q^t) = \left[\left(\frac{p^t \cdot q^{t-1}}{p^{t-1} \cdot q^{t-1}}\right)\left(\frac{p^t \cdot q^t}{p^{t-1} \cdot q^t}\right)\right]^{1/2}$$

and the chained index is thus then

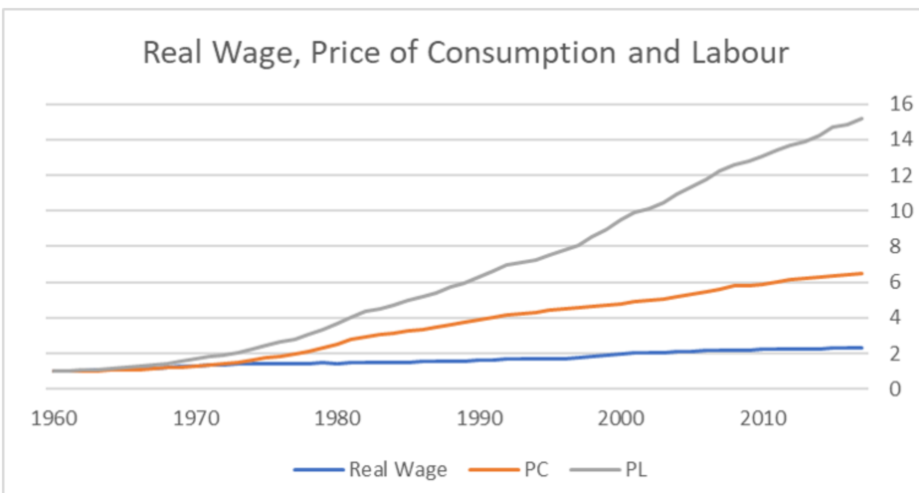
$$P_{FCH}(p^0, p^t, q^0, q^t) = P_F(p^0, p^1, q^0, q^1)P_F(p^1, p^2, q^1, q^2)...P_F(q^{t-1}, q^t, p^{t-1}, p^t)$$

which then implicity defines the quantity index and finally we can form the labour value aggregate as

$$VL = PL \cdot QL$$

We also note in our data (see appendix) that the wage rate for employees (PE), is equal to a constant (c=2.63315) times the wage rate for all types of workers, PL , which is an application of Hick's Aggregation Theorem, since we have made the wages of the self-employed and unpaid family workers proportional to that of the employed wage earners. With this price series for labour, we can now calculate the real wage simply as

$$\text{Real Wage} = \frac{PL}{PC}$$



We note that according to this calculation, real wages only grew 2.32 fold over the sample period. Part of this relatively muted real wage growth may be attributed to the large growth in labour force (QL grew 2.15 fold) which can suppress wages. If we form the corresponding *growth rate* series for real wages, we see real wages grew on average at 1.5% per year over the sample period and had a downward trend of -.02% per year, that is the growth rate of real wages remained almost constant.

2.5 Taxes, Inflation, and Interest Rates

Here we describe the procedure use to impute the tax rates and subsidy rates that will be used throughout this paper. Broadly, we will concern ourselves with taxes on consumption, capital, exports, imports and labour. The specific series from OECD.stat we used were:

- T1000: Taxes on income, profits and capital gains
- T1100: Taxes on income, profits and capital gains of individuals
- T2000: Social Security Contributions
- T3000: Taxes on Payroll and Workforce
- T4000: Taxes on Property
- T5000: Taxes on goods and Services
- T5123: Customs and Import Duties
- T5124: Taxes on Exports
- T6000: Other Taxes

We will assume that most income taxes T1100 fall on labour, and the difference of T1000 and T1100 falls on capital. We take social security contributions to be a tax on labour, as well as payroll taxes. We also take T4000 (property) and T6000 (other) to be taxes that fall on capital. Thus we define:

- TC=T5000-T5123-T5124 (taxes on consumption)
- TX=T5124 (taxes on exports)
- TM=T5123 (taxes on imports)
- TL=T1100+T2000+T3000 (taxes on labour)
- TK=T1000-T1100+T4000+T6000 (taxes on capital)

and note that in this construction we are assuming no taxes fall on government or investment. We next compute tax rates:

$$\begin{array}{lll} TRC = TC/VC & TRX = TX/VX & TRM = TM/VM \\ TRL = TL/VL & TRK = TK/VK & \end{array}$$

As our data was missing observations for the years 1961-1964, we used a linear interpolation, that is we computed:

$$TRC_{1960+n} = \left(\frac{5-n}{5}\right) TRC_{1960} + \left(\frac{n}{5}\right) TRC_{1965} \qquad \text{n=1,...,5}$$

$$TRX_{1960+n} = \left(\frac{5-n}{5}\right)TRX_{1960} + \left(\frac{n}{5}\right)TRX_{1965} \qquad \text{n=1,...,5}$$

$$TRM_{1960+n} = \left(\frac{5-n}{5}\right)TRM_{1960} + \left(\frac{n}{5}\right)TRM_{1965} \qquad \text{n=1,...,5}$$

$$TRK_{1960+n} = \left(\frac{5-n}{5}\right)TRK_{1960} + \left(\frac{n}{5}\right)TRK_{1965} \qquad \text{n=1,...,5}$$

$$TRL_{1960+n} = \left(\frac{5-n}{5}\right)TRL_{1960} + \left(\frac{n}{5}\right)TRL_{1965} \qquad \text{n=1,...,5}$$

Next, as we did not have any tax data for 2017, we set 2017 tax rates equal to those of 2016, that is $TRC_{2017} = TRC_{2016}$ and the same for all other series. This is a reasonable assumption to make as there were no major tax policy changes at the federal level enacted effective in 2017. Obviously President Trump's TCJA was passed in 2017, but it is not law until fiscal year 2018, so our constant tax assumptions for 2017 seems reasonable. For subsidies, we use the data series on subsidies on products, which act as an increased selling price of the subsidized output by the subsidized industry. The OECD.stat series has all observations 1960-2016, and the 2017 observation was taken from the same series on FRED. We then compute the subsidy rate simply as

$$SRC = \frac{VS}{VC}$$

As we now have all series on taxes and subsidies, we can construct gross profits (gross return to capital) at producer prices:

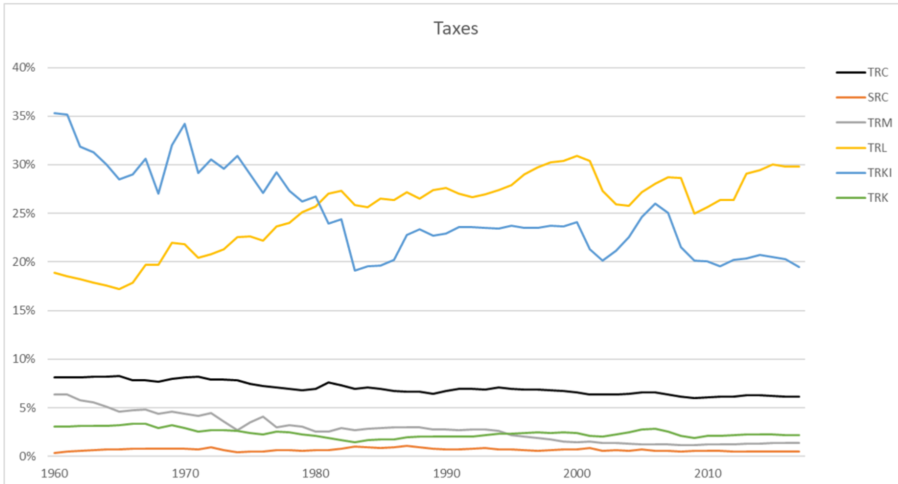
$$\begin{aligned} GPROF = {} & PC(1 - TRC + SRC)QC + PI(QI) + PG(QG) \\ & + PX(1 - TRX)QX - PM(1 + TRM)QM - PL(QL) \end{aligned}$$

to go from gross profits to net profits we just need to subract the value of depreciation:

$$NPROF = GPROF - VD$$

which we can then use to define a profits (capital income) taxation rate:

$$TRKI = \frac{TK}{NPROF}$$



Here we include a plot of all imputed tax rates over the sample period, for which there is a lot to comment on. The first is that some of the wild year over year swings in tax rates should not be read to much in to. These jumps do not actually represent changes in taxation policy, but our graph is picking up recessions, as a relic of our methodology. For example in 2008 it appears the tax rate on capital income (TRKI) declines dramatically, but this is just tax receipts falling by more than profits. Reasons for this most likely include firms claiming capital losses, various deductions, as well as cascading tax rates. This being said, as Dr. Diewert pointed out in a presentation, what we have done is actually calculate *effective* tax rates. A *statutory* tax rate is the tax rate that is on the books (legislated), for example in 2017 the US Corporate Tax Rate was 35%. The effective tax rate is what the firm or individual actually ends up paying after taking advantage of various loopholes and deductions in the tax code, which we see for US corporations in 2017 turned out to be about 20%. An often parroted talking point among certain American politicians is that the USA has the highest corporate tax rate of any advanced economy. We have just shown this is not so! This leads us into our next point, which from our graph we see that for an advanced economy, the USA is a relatively low taxation country. The USA has no federal sales tax, and many individual states have either no state sales tax or no state income tax. We see from our plot that consumption tax rates have fallen steadily from over 8% in 1960 to 6% in 2017. As of 2017, the highest individual income tax bracket was 39% (down to 37% in 2018 after the TCJA) which is low compared to Canada and most European nations.

The general trend is thus moving away from taxes on capital and moving towards taxes on labour. There are three main events in USA taxation policy over the sample period that are worth mentioning. The first is the Reagan tax cuts. In The Economic Recovery Tax Act of 1981, passed in response to the Volcker recession, all marginal tax rates were cut substantially, most notably with the top income bracket falling from 70% all the way to 50%. Next in The

Tax Reform Act of 1986, Reagan again cut income tax rates substantially, with the top rate then falling from 50% to 38.5%, and consolidated many brackets. As this bill was true "tax reform", it also enacted the Alternative Minimum Tax on corporations, and closed various loopholes, with the result being a bill that is revenue neutral. The Tax Reform Act of 1986 remains quite popular with economists for these reasons. Next there was the Bush Tax Cuts of 2001 and 2003. These cuts lowered all individual income taxes, lowered the capital gains tax, and lowered the inheritance tax, with there being a one year window (2010) where the estate tax was entirely repealed. Any bill which is non-revenue neutral for more than 10 years must pass the Byrd Rule in the US Senate, which means it requires 60 votes out of 100 to be passed into law. Democrat politicians would not support the Bush tax cuts, specifically as they viewed the changes to the inheritance tax and capital gains tax as regressive, so these bills were unable to pass the Byrd Rule, and were thus sunsetted (set to expire) in 10 years. This is what caused the famous "Fiscal Cliff" for then President Obama. With a then Democrat president, and a Republican congress, many of which were hardline libertarians from the 2010 Tea Party movement, there was no political consensus on how to make permanent the popular parts of the Bush tax cuts (rate cuts for lower earners), but let expire the unpopular parts (cuts in the inheritance tax). Thus it turned into an incredibly high stakes political standoff in which Republicans would only extend all tax cuts or nothing at all, but Democrats controlling the presidency and the bully pulpit wouldn't go along with this. If one accepts the premise that cutting marginal rates corrects incentives in a way that is beneficial for economic growth and output, then surely all marginal rates increasing over night would be disastrous for the economy. There were other complex fiscal issues intertwined in the fiscal cliff, but the expiring Bush tax cuts formed the basis for it. Crisis was averted at the last minute, but it easily could have plunged the US economy back into recession at a time when it was already quite fragile and the post 2008 recovery was still sluggish. This leads us into the President Trump's Tax Cuts and Jobs Act of 2017, which he initially wanted to call the Cuts Cuts Cuts Act. The hallmarks of this legislation were lowering the corporate tax rate from 35% to 21%, a doubling of the standard deduction, and eliminating various deductions for state and local taxes. This legislation does not show up in my data set, but its effects on the US economy and the US government's fiscal position will certainly be interesting to track going forward. Due to President Trump being deeply unpopular, this bill was not able to receive bipartisan support (pass the Byrd Rule) so it will be sunsetted in 10 years, which will almost assuredly create another political and economic crisis. My own editorial comment on these matters is I think it is ironic that sunsetted tax legislation is being passed to create a more business friendly and pro growth environment, when the ticking time bomb expiration dates on these rate cuts does the exact opposite. Republicans may argue that that they would make these cuts permanent if they could, but if this is the case then they should make a greater effort to reach across the aisle and pass bipartisan tax reform. Lastly, we note that the first Reagan tax cut, the Bush tax cut, and now the TCJA are all deficit financed and have contributed to the increasingly tenuous fiscal position of the US federal government, although that is beyond the scope of this paper.

We can now calculate an overall real rate of return that the US economy earned each year. Recall our formula for the end of period user cost of capital was

$$PU = PK\Bigg((1+R) - (1+IK)(1-D)\Bigg) \tag{19}$$

where PK is the beginning of period of year price of capital, R is the production unit's cost of financial capital, and IK is the actual or anticipated asset inflation rate. We will go with Jorgenson's preferred methodology and and let IK be the actual or ex-post asset inflation rate.

$$IK_t = \frac{PK_{t+1}}{PK_t} - 1 \qquad\qquad \text{t=1960,...,2016}$$

We now know that the beginning of the period value of capital stock is simply VK, and its end of period value is VK adjusted positively for asset inflation and negatively for depreciation, that is $(1-D)(1+IK)VK$, so the change in the value of the capital stock over the period is simply

$$DVK = VK - \Bigg((1-D)(1+IK)\Bigg)VK$$

and thus we can calculate our ex-post nominal rate of return on the value of assets R as

$$i_t = \frac{GPROF_t - DVK_t}{VK_t}$$

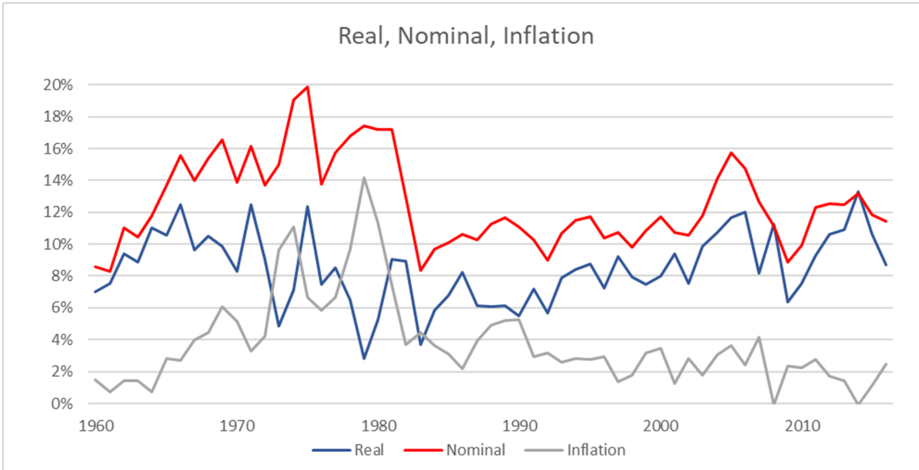
Now that we have a nominal rate of return, we need only inflation to calculate the real rate of return. For this we use CPI data from the BLS, where for each year we set the CPI to be equal to its January observation. We then normalize the series to be equal to 1 in 1960 (divide all entries through by the January 1960 CPI value), and then we calculate inflation as

$$\pi_t = \frac{CPI_{t+1}}{CPI_t} - 1$$

With this we can then calculate the real rate of return from the exact form of the Fisher equation, that is

$$r_t = \frac{1+i_t}{1+\pi_t} - 1$$

and we include a plot for the real interest rate, nominal interest rate and inflation over the sample period.

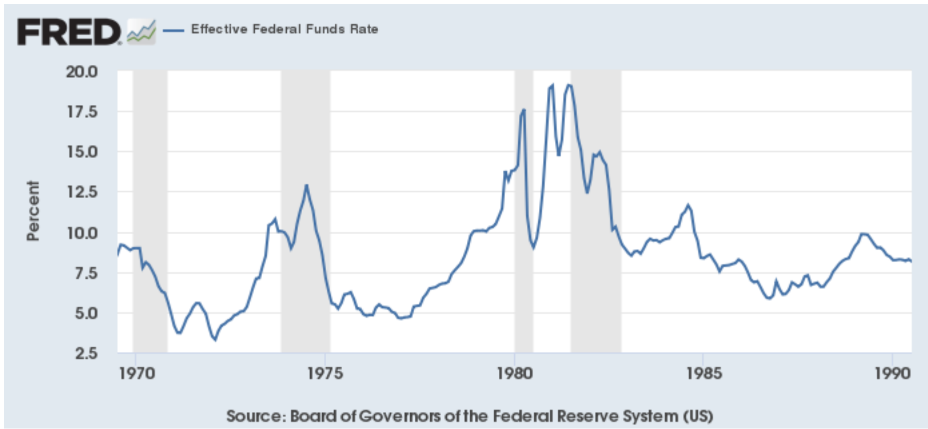


This graph nicely illustrates some of the broad macroeconomic trends in the USA over the sample period. We begin by examining the inflation series. In December 1969 the US economy entered a mild recession which ended as of November 1970, although unemployment remained high thereafter. Then President Nixon understood that high unemployment coupled with sluggish GDP growth would be perilous for his re-election campaign in 1972. Nixon began to exert extreme pressure on Burns, both directly and through Office of Management and Budget Director George Shultz, to engage in expansionary monetary policy leading up to his election. In a fascinating paper, Abrams (2006) documents all relevant private oval office and phone conversations between Nixon and Burns. Nixon if nothing else understood the short run Phillips curve, and wanted to increase inflation and reduce unemployment at all costs. Although Burns was a sycophantic Republican, he was also a well respected and reputable academic economist who initially was resistant to Nixon plans. Indeed, even by the time the money supply was sky-rocketing in 1972 and Burns was fully on board with Nixon's plan, Nixon said "This is the last time I want to see him... [garbled] or get the hell out of here. War is going to be declared if he doesn't come around some." Nixon notes further that Burns is "talking to the Jewish press." Other select quotes from Nixon in private conversation in 1972 are "The economy has to be good, strong expanding economy this year. So much at stake on that", "Keep the money supply going up!"" and "Great. Great, You can lead 'em. You can lead 'em. You always have, now. Just kick 'em in the rump a little." Where the last quote is in conversation with Burns a day before a Federal Open Market Commitee meeting in which Nixon has convinced Burns to ask the FOMC to increase the money supply yet again. Indeed, Nixon was such a shrewd political operator he exerted this pressure not only through direct conversation and Shultz, but also through actions such as a steady stream of anonymous leaks to pressure Burns, including floating one proposal to expand the size of the Federal Reserve (so that Nixon could appoint a majority of the new members) and another proposal to give the White House more control over the Fed, while planting a false story that Burns was requesting a large pay raise, when in fact Burns had suggested taking a pay cut.

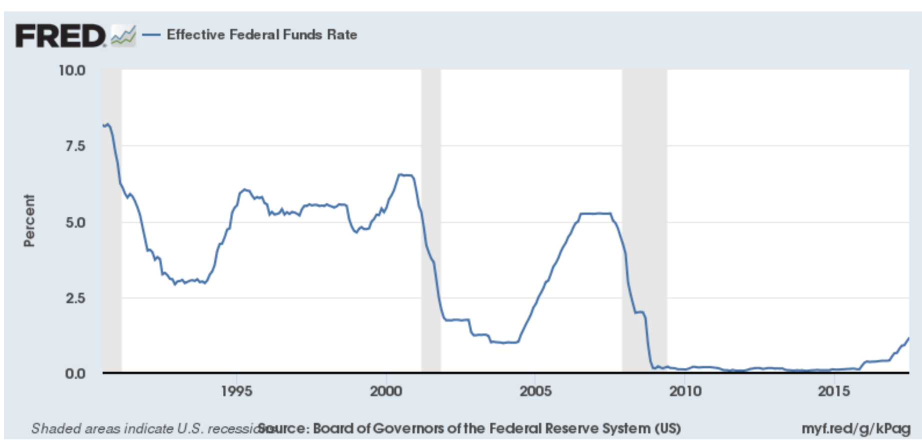
Needless to say, this relentless pressure by Nixon was a success, with the money supply (M2) in particular exploding from 1970-1972 and GDP growth and unemployment being restored in time for the 1972 election, which Nixon then won in a landslide. The following table is adapted from Abrams (2006) where M1 and M2 are from FRED and the real GDP growth rate is from our GDP quantity series QY_T.

	M1 Growth	M2 Growth	RGDP Growth
1970	4.51%	7.36%	2.80%
1971	6.77%	13.68%	5.20%
1972	7.56%	11.65%	5.60%

This high inflation of the early 70's was then a prelude to the extreme inflation of the mid and late 70's. Indeed, inflation peaked as high as 11.1% in 1974 during the OPEC oil embargo when the price of oil nearly quadrupled, and then as high as 14.2% in 1979 when the price of oil skyrocketed again due to the Iranian revolution and the Iran-Iraq war. We see that the wild inflation of the 70's was then tamed in the early 80's by Fed Chairman Paul Volcker, leading to the so-called "Volcker Recession". This was when the Fed finally credibly commited to controlling inflation. To do so, the money supply needed to be tightened considerably which involved raising interest rates to unheard of levels. Indeed, the following Fred graph shows the effective federal funds rate at 9% in the early 1980s, which then almost *doubled* in the course of a year to 18%, plunging the US economy into a second recession. Although this monetary tightening had painful short term consequences, it was successful in vanquishing the persistently high inflation of the 70s. Paul Volcker has been considered a very successful Fed Chairman thanks to this feat, although he was fired in 1986 by Ronald Reagan because he was not doing enough to advance the deregulation agenda of the administration. His successor was then Alan Greenspan. Lastly in the following graph, one should also note Arthur Burns cutting the effective federal funds rate from nearly 10% at the start of 1970 to 4% by Richard Nixon's election in 1972.



From the mid 80's until 2008 inflation remained very low and stable. This era is often referred to as "the great moderation" by economists, a term popularized by Ben Bernanke, as almost all macroeconomic variables saw a marked decrease in their volatility, coupled with the sustained booms of the mid to late 80s, 90s, and post tech bubble 00s. This era was brought to an end by the financial crisis of 2008. Finally, we note that our inflation series actually became *negative* in both 2008 and 2013, which was due to the impotency of monetary policy at the zero lower bound for nominal interest rates. Indeed, we also provide a plot of the effective Fed Funds Rate over the rest of the sample period, and see that it has essentially been 0 since 2010.



We conclude with descriptive statistics for real and nominal rates and various measures of inflation. Note the last two columns are average *levels* over the sample period and the other columns are average yearly growth rates. Indeed, we note our average real rate of return $\mu_r = .08486$ is much too high. This is primarily due to our neglect of land and natural resources in our asset base. A comparable measure, return on equity, may be the inflation adjusted S&P 500 which has averaged roughly 7% per year over the sample period. We also note the real rate of return is an important measure of the efficiency of the economy, and it has shown no sign of a slowdown over the sample period, a sign of the vigor of the economy, despite a significant dip in 2008. Inflation averaged only 3.8% per year over the sample period, although there was substantial volatility about this mean as previously discussed. Lastly we note our average CPI level and average price of consumption level differ somewhat substantially. As the CPI is calculated for *urban consumers* and PC is over *all* consumption, I would guess if anything that PC may have a relative upward bias to the CPI but this does not seem to be the case. This difference may be due to the treatment of owner occupied or revisions to the CPI over the sample period.

	r_t	i_t	π_t	IK_t	CPI	PC
Mean	0.08486	0.1259	0.03813	0.0303	4.0834	3.5016

2.6 The Solow Model, Growth Accounting and the Solow Residual

Here we briefly present the theory of the Solow (1957) model. We do this for a few reasons. First, the Solow model exemplifies the importance of technological progress. We will see it is an equilibrium growth model in which all per capita variables grow at the rate of technology in steady state. As the rest of this paper is dedicated to measuring technological progress in a more sophisticated fashion, this model makes it crystal clear why we should care about TFP growth. Secondly, the Solow model yields a "rough and ready" easy to apply formula for measuring technical progress, called the Solow Residual. We will show that this Solow Residual gives us an estimate of TFP growth which is *strikingly* close to our index number, econometric, and non-parametric estimates. As it is the easiest to implement of all these methods, the Solow Residual is then an excellent first approximation to measuring total factor productivity growth.

We begin by assuming a production function of the form

$$Y = F(K, AL)$$

Here K is the capital stock, L is labour input, and A is the level of technology, which enters multiplicatively through A, so we say technology is labour-augmenting or Harrod-Neutral. We will refer to the quantity AL as units of *effective labour*. Next we assume this production function is linearly homogeneous in inputs, so it is subject to constant returns to scale, that is

$$F(cK, cAL) = cF(K, AL)$$

This assumption of constant returns to scale allows us to work with our production function in *intensive form*, that is

$$\begin{aligned}
\frac{Y}{AL} &= \frac{1}{AL}F(K, AL) \\
y &= F(\frac{K}{AL}, 1) \\
y &\equiv f(k)
\end{aligned}$$

where $y = Y/AL$ is output per units of effective labour and $k = K/AL$ is capital per units of effective labour. To get convergence to a steady state, we will need to make some assumptions about to intensive form of the production function. Indeed, we let $f(0) = 0$, $f'(k) > 0$, and $f''(k) < 0$. We note that the marginal product of capital is just

$$\begin{aligned}
\frac{\partial F(K, AL)}{\partial K} &= \frac{\partial}{\partial K} ALf\left(\frac{K}{AL}\right) \\
&= AL \cdot \frac{1}{AL} f'(k) \\
&= f'(k)
\end{aligned}$$

Thus the marginal product of capital is positive and subject to diminishing returns per unit of effective labour. Lastly we will need to assume that

$$\begin{aligned}
\lim_{k\to\infty} f'(k) &= 0 \\
\lim_{k\to 0} f'(k) &= \infty
\end{aligned}$$

where the above limits are often referred to as the Inada Conditions.

We assume that both labour and knowledge grow at constant rates, so

$$\frac{\dot{L(t)}}{L(t)} = n \qquad\qquad \frac{\dot{A(t)}}{A(t)} = g$$

where our goal is to show that in the steady state equilibrium of this economy, per capita output and consumption both grow at g. The last assumption we need to make is regarding the evolution of the aggregate capital stock. We assume that output which is not consumed is re-invested, so our equation of motion for capital is

$$\dot{K(t)} = sY(t) - \delta K(t)$$

where s is the *savings rate* which is taken to be exogenous and constant, and δ is the depreciation rate which is also taken to be constant.

We now derive the main equation of the Solow model. As 2 of our 3 inputs, labour and technology, have growth rates that are exogeneous, we now want to examine the dynamics of capital per unit of effective labour. Indeed,

$$\begin{aligned}
\dot{k(t)} &= \frac{\partial}{\partial t}\left(\frac{K(t)}{A(t)L(t)}\right) \\
&= \frac{\dot{K(t)}}{A(t)L(t)} - \frac{K(t)}{(A(t)L(t))^2}\left(A(t)\dot{L(t)} + L(t)\dot{A(t)}\right) \\
&= \frac{\dot{K(t)}}{A(t)L(t)} - \frac{K(t)}{A(t)L(t)}\frac{\dot{L(t)}}{L(t)} - \frac{K(t)}{A(t)L(t)}\frac{\dot{A(t)}}{A(t)} \\
&= \frac{sY(t) - \delta K(t)}{A(t)L(t)} - k(t)n - k(t)g \\
&= sf(k(t)) - \delta k(t) - nk(t) - gk(t) \\
\dot{k(t)} &= sf(k(t)) - (n + g + \delta)k(t)
\end{aligned}$$

where the second equality is the product rule, the third equality follows from multiplying out the parentheses, and the fourth equality comes from substituting in our equations of motion for the aggregate capital stock, as well as our growth rates for labour and technological progress.

We now show that this economy as we have described it will converge to a steady state level of capital per unit of effective labour k^*, and that output and consumption grow at the rate of g in the steady state.

Definition: A steady state for the above dynamical system is a k^* such that

$$\dot{k(t)}\Big|_{k=k^*} = 0$$

Indeed we can easily solve for the steady state of our model simply as

$$\begin{aligned}
0 &= sf(k^*) - (n + g + \delta)k^* \\
sf(k^*) &= (n + g + \delta)k^*
\end{aligned}$$

The above equation has a nice economic interpretation. The left hand term, $sf(k^*)$ is actual investment per unit of effective labour. The right hand term, $(n + g + \delta)k^*$, is *break-even investment*, that is it is the amount of investment that must be done to keep k at its existing level. Since k is capital per unit of *effective labour*, this break even term includes not only depreciation, but also the growth rates of labour and technology, since AL is growing at rate $n + g$. Thus, we have found our steady state k^* to be simply where actual investment is equated with break even investment.

We now show that this economy as described (with our assumptions on the production function f) will converge to the unique steady state k^*.

Theorem: The steady state k^* of the 1-D dynamical system $\dot{k} = sf(k) - (n+g+\delta)k$ is globally stable. That is, for any intial capital stock $k_0 > 0$, we will have $k_0 \to k^*$, at which then $\dot{k^*} = 0$

Proof: First recall we assumed our intensive form production function f to be globally concave, that is $f''(k) < 0$ for all k. Using an alternative characterization of concavity, we note that

$$\frac{f(k) - f(0)}{k - 0} > f'(k)$$

and letting $f(0) = 0$ we see $f(k) > kf'(k)$ and dividing through by k and multiplying both sides by s yields

$$\frac{sf(k)}{k} > sf'(k)$$

Now, the condition for a steady state, k^*, to be a stable equilibrium of this dynamical system is

$$\frac{\partial \dot{k(t)}}{\partial k}\Big|_{k=k^*} < 0$$

and we note that

$$\begin{aligned}
\frac{\partial \dot{k(t)}}{\partial k}\Big|_{k=k^*} &= sf'(k^*) - (n+g+\delta) \\
&< \frac{sf(k^*)}{k^*} - (n+g+\delta) \\
&= \frac{(n+g+\delta)k^*}{k^*} - (n+g+\delta) \\
&= 0
\end{aligned}$$

Thus for any strictly positive k_0 we have $k_0 \to k^*$.

Now that we have shown our economy will converge to a steady state, we can describe the evolution of aggregate and per capita variables at this steady state. Indeed, as $K = kAL$ and k is constant at k^* with AL growing at rate $n+g$, we see aggregate capital stock grows at rate $n+g$. Now that aggregate capital and effective labour are both growing at rate $n+g$, the constant returns to scale property of our production function $Y = F(K, AL)$ implies Y is growing at rate $n+g$. Lastly, as $Y/L = Af(k^*)$ in steady state, this implies output per worker grows at rate g, and as aggregate consumption is $C = (1-s)Y$, consumption per capita is then $C/L = (1-s)Y/L$ which then grows at rate g.

Result: On the balanced growth path (steady state) of this Solow economy, output per capita and consumption per capita both grow at the rate of technical progress g.

This result, combined with the fact that the US economy is described reasonably well by the balanced growth path of the Solow model (we witnessed a relatively stable capital output ratio, as well as rising output per worker and capital per worker in our data construction) makes the rest of this paper worth reading. It states that in the long run, technical progress is king. If we are to increase per capita living standards, this must ultimately happen through g. Before we can even dream of finding policies to increase g, we must be able to accurately and consistently measure it, which is the work we devote the rest of this paper to.

We finish our exposition on the Solow model with a few general remarks. The first is that it can be shown that an increase in the savings rate s will have a level (temporary) positive effect on output, but not a permanent one. In the long run we cannot continually increase per capita output by simply saving and investing more. Our next remark is that this savings rate is taken to be *exogeneous*, an undesirable feature of the model. Its natural extension, the Ramsey-Cass-Koopmans model, rectifies this where a savings rate is then determined by utility maximization problems of representative agents and is thus endogenous to the economy. Lastly, even if we accept the conclusions of the Solow model that ultimately all increases in per capita living standards must come from technical progress, the fact that our measure of such technical progress g is taken to be exogenous (coming from heaven) is very unsatisfying. Models of endogenous growth will incorporate elements such as human capital, research and development spending and learning by doing, see Romer (1986, 1990). As this paper is fundamentally concerned with the measurement of technical progress, not prescribing policies to amplify it, we will not comment any further on endogeneous growth.

We now calculate the Solow Residual and show that it generates TFP growth estimates that are almost identical to our index number estimates that will be presented shortly. The basic idea of the Solow Residual is that if we assume a neoclassical production function as in this section, then we can observe output Y, and inputs K and L, and the amount by which observed output exceeds observed input (the residual) must be due to technical progress in our production function. Indeed, we now go to discrete time (as our data is yearly) and assume a Cobb-Douglas production function,

$$Y_t = A_t K_t^{\alpha} L_t^{1-\alpha}$$

where note now we have assumed technical progress is *Hicks Neutral*, and $1-\alpha$ is the labour income share. Output at time $t+1$ is then:

$$Y_{t+1} = A_{t+1} K_{t+1}^{\alpha} L_{t+1}^{1-\alpha}$$

and taking log differences yields:

$$\log Y_{t+1} - \log Y_t = \log A_{t+1} - \log A_t + \alpha(\log K_{t+1} - \log K_t) + (1-\alpha)(\log L_{t+1} - \log L_t)$$
$$\log\left(\frac{Y_{t+1}}{Y_t}\right) = \log\left(\frac{A_{t+1}}{A_t}\right) + \alpha \log\left(\frac{K_{t+1}}{K_t}\right) + (1-\alpha)\log\left(\frac{L_{t+1}}{L_t}\right)$$
$$\log(1+g_y) = \log(1+g_g) + \alpha \log(1+g_k) + (1-\alpha)\log(1+g_l)$$

Where note we have slightly changed notation, g_y is the growth of aggregate output, g_g is the growth rate of technical progress, g_k is the growth rate of our aggregate capital stock, and g_l is the growth in labour input. We also clearly could have derived this same formula by taking a time derivative of the logged production function, but as our estimation is done in discrete time this is more illustrative. We now apply the approximation

$$\log(1+x) \approx x \qquad\qquad \text{for small } x$$

so the above reduces to

$$g_y = g_g + \alpha g_k + (1-\alpha)g_l$$

or

$$g_g = g_y - \alpha g_k - (1-\alpha)g_l$$

This formula simply says productivity growth is just the difference between output growth and observable input growth, weighted by their income shares. We can then easily calculate this residual from our series on aggregate output, capital stock, labour input, and labour shares.

The following table presents descriptive statistics of our Solow Residual, g_g, as well as our index number measurements of TFP growth which will be presented in Section 3. It is shown that these two methods produce very similar results. On the one hand maybe this is not surprising, as both are simply measuring growth of output against growth of labour and capital inputs. On the other hand it is re-assuring that they produce near identical estimates.

	Mean	Min.	Max.	St. Dev.	Correlation
Solow Residual	0.0093	-0.0245	0.0317	0.0122	0.9993
TFPG	0.0089	-0.0259	0.0308	0.0121	0.9993

Reasons for using this Solow Residual as a TFP estimate may be that it is our most easily derived formula, as well as the fastest to apply to the data. This makes it a good first approximation to the problem. Reasons against using it are that it is clearly unrealistic to assume a Cobb-Douglas production function for the entire US economy, as this functional form arbitrarily restricts elasticities. Our index number estimates will be preferred as they avoid these undesirable assumptions. The conclusions of the Solow model, that the long run per capita growth rate of output and consumption are determined solely by the rate of technical progress, frames the immense importance of rigorously measuring TFP growth, which we now proceed to do in more preferred settings.

3 Index Number Estimates of Total Factor Productivity

In this section we will compute traditional index number measures of total factor productivity in both the gross and net real income framework. We will also follow the Diewert and Morrison (1986) methodology and do a Kohli (1991) type decomposition of real income growth into explanatory multiplicative factors, which will also give us the all important measure of per capita real income growth over the sample period. Lastly we will discuss the issue of the declining labor share of income.

3.1 Gross Real Income Framework

We begin by constructing chained Törnqvist-Thiel input and output aggregates. For our input aggregate, define the 2x1 price and quantity vectors $p^t = (PL^t, PKS^t)$ and $q^t = (QL^t, QKS^t)$ where L is labour and KS capital services, whose price is previously defined by equation (19). The Törnqvist-Thiel price index for period t is then

$$\log P_T(p^{t-1}, p^t, q^{t-1}, q^t) = \sum_{i=1}^{2} \frac{1}{2}(s_n^t + s_n^{t-1}) \log\left(\frac{p_n^t}{p_n^{t-1}}\right)$$

where expenditure shares s_n^t are defined in the usual way. We can then define our chained index as

$$PZ_T(p^0, p^t, q^0, q^t) = P_T(p^0, p^1, q^0, q^1) P_T(p^1, p^2, q^1, q^2) ... P_T(p^{t-1}, p^t, q^{t-1}, q^t)$$

which then implicity defines our quantity index

$$QZ_T(p^0, p^t, q^0, q^t) = \left(\frac{p^t \cdot q^t}{p^0 \cdot q^0}\right) \Big/ PZ_T(p^0, p^t, q^0, q^t) \tag{20}$$

For our output aggregate, we made the following adjustments using our tax data.

$$PPC = PC(1 - TRC + SRC) \quad PPG = PG \quad PPI = PI$$
$$PPX = PX(1 - TRX) \quad PPM = PM(1 - TRM)$$

where recall we are treating subsidies as an increase in the selling price of the subsidized output of subsidized industries, and also assuming no taxes fall on investment and government. We then form the the 5x1 price and quantity vectors $p^t = (PPC^t, PPG^t, PPI^t, PPX^t, PPM^t)$ and $q^t = (QC^t, QG^t, QI^t, QX^t, -QM^t)$ noting that we have made QM negative. We then aggregate using chained Törnqvist-Thiel indexes in the exact same manner we did above with our input series. Our Törnqvist-Thiel index for output prices is now

$$\log P_T(p^{t-1}, p^t, q^{t-1}, q^t) = \sum_{i=1}^{5} \frac{1}{2}(s_n^t + s_n^{t-1}) \log\left(\frac{p_n^t}{p_n^{t-1}}\right)$$

and the chained output price index then cumulates these one period rates of change simply as

$$PY_T(p^0, p^t, q^0, q^t) = P_T(p^0, p^1, q^0, q^1)P_T(p^1, p^2, q^1, q^2)...P_T(p^{t-1}, p^t, q^{t-1}, q^t)$$

Indeed, letting $PY_T(p^0, p^t, q^0, q^t)$ be our chained output price index, we then have the quantity index implicity defined as

$$QY_T(p^0, p^t, q^0, q^t) = \left(\frac{p^t \cdot q^t}{p^0 \cdot q^0}\right) \Big/ PY_T(p^0, p^t, q^0, q^t) \tag{21}$$

The reader will recall this is almost identical to our aggregation to GDP in section 2.4, except now we have incorporated our tax data. Indeed, the index number estimates for our tax adjusted output price series are:

Chained Index	PY_T	PY_F	PY_P	PY_L
2017	6.917	6.917	6.923	6.912

Where we see our 2^{nd} order approximation property at work, PY_T and PY_F are equal to 3 decimal places, chaining has nicely reduced the spread between the Paasche and Laspeyres to .011, and the Fisher falls exactly in between them. We note that these prices are *slightly* higher than our GDP series index number estimate. The reason for this is that we are now adjusting for taxes, (primarily consumption), and consumption taxes declined over the sample period, so our early period price of consumption is relatively lower to end of period consumption price, making our price index slightly larger (higher inflation). As a result our quantity series are slightly lower than before:

Chained Index	QY_T	QY_F	QY_P	QY_L
2017	2691	2691	2688	2693

Thus we can define total factor productivity simply as the ratio of our output and input quantity indexes:

$$PROD = \frac{QY_T}{QZ_T} \tag{22}$$

and the growth rate of productivity can be defined by setting

$$YG = \frac{QY_{Tt}}{QY_{T(t-1)}} \qquad ZG = \frac{QZ_{Tt}}{QZ_{T(t-1)}} \tag{23}$$

and thus computing

$$PRODG = \frac{YG}{ZG}$$

The detail oriented reader will note that we have used Törnqvist-Thiel output and input quantity indexes to form our productivity estimate, but we could be equally justified in using the Fisher to this end as well. It turns out that due to our second order approximation property there is no difference! In our empirical work we computed index number estimates of productivity using *both* ratios of Fisher output and input quantity indexes, and ratios of Törnqvist-Thiel output and input quantity indexes, and observed they were identical to 5 decimal places! This is more decimal places than we need report, so although what follows is technically our Törnqvist-Thiel index number measure of productivity, this is identical to the Fisher measure.

3.2 A Historical Retrospective on US Productivity Growth

It will be illuminating to present our productivity estimates over varying time frames to help us observe and understand the broad trends in American productivity over the sample period. Indeed, we first present summary statistics for productivity over the entire sample period. We see the average productivity growth rate from 1960-2017 was 0.89% per year. This is an *arithmetic average*. We would be equally justified in taking the *level* of productivity in 2017, which was 1.65098, and raising it to the $\frac{1}{57}$, which gives us the *average geometric growth rate* of 0.88% per year. We note our maximum observation was productivity growth of 3.08% in the year of 1984. This extremely high number appears at a time when there was a malaise in productivity growth, as we will soon see, so it is somewhat of a mystery. A potential explanation involves going back to our investment quantity series and noting that it surged in 1984, which is what is driving the large output increase in that year. This surge in investment may have been driven by a perfect storm of factors: inflation had finally been brought under control, creating a much more desirable business environment to invest in, as well as the Reagan administration had cut taxes and was slashing regulation, also spurring investment. This confluence of events can explain the large increase in investment which greatly increased output and thus productivity in 1984.

Interestingly, our minimum observation of -2.59% productivity growth occurred just 2 years earlier in 1982. This abysmal productivity regress can be explained by examining both the input and output quantities that determine our productivity estimate. First, on the input side, the price of capital services remained quite high relative to other recessions. For example, price of capital services declined by less than 1% in this recession, whereas it declined by more than 3% in the great recession. This is unique to 1982 (recall it was the Volcker recession), because nominal rates were sky high due to the monetary tightening the economy was undergoing, and a high nominal rate means a high user cost of capital so our input quantity index (denominator) was "too large" given the weak nature of the economy. On the output side, obviously the economy had entered a recession so real output declined. Thus we can explain the terrible productivity performance of 1982 by the unique and highly unusual circumstances created by the Volcker recession.

	Mean	σ	Min.	Max.
PRODG	1.0089	0.0121	0.9741	1.0308
YG	1.0295	0.0206	0.9717	1.0731
ZG	1.0204	0.0118	0.9770	1.041

We next examine the arithmetic average of productivity growth rates decade by decade.

Decade	1961-69	1970-79	1980-89	1990-99	2000-09	2010-17
PRODG	1.0133	1.0078	1.0074	1.0132	1.0045	1.0073

We note that the worst decade for productivity growth was 2000-2009, with an average growth rate of only .45% per year. Although the main driving force of this was the great recession, it is worth noting that productivity growth actually slowed down *before* 2008. Indeed, productivity growth rates were negative over the entire 2006-2009 period for the US economy. We will soon follow the work of growth theorists and consider this decade as two distinct periods, one in which there was a productivity *surge* through 2005, and then the decline following it. Indeed, these decade boundaries are somewhat arbitrary for if we were to include 2010 (which was the most robust year of the millenia so far) alongside the 2000s, then the current decade, 2011-2017 would be the worst for productivity performance. Either way, American productivity has continued to be sluggish in the current decade, growing only .073% per year from 2010-2017. This figure is arguably more worrying than that of the previous decade, because the US economy has not been in recession at any point over this period.

This paper only intends to measure productivity, not to explain it, but we will offer a few remarks regarding this slowdown. It is clear that the technology of the 2010's have not yet worked their way into productivity statistics. These technologies include smart phones and their related apps (Uber, AirBNB etc.) 3D printing, self-driving cars, and a leap forward in business analytics through machine learning and artificial intelligence. It then remains to ask is there a gestation period for which these technologies will work their way into productivity statistics, such as was the case with the internet and information technology, or are these inventions simply not as impactful for productivity as previous inventions such as the internet and the integrated circuit. It seems our new technologies are simply "built on top of" old ones, so a strong case could be made for diminishing marginal returns.

Another potential explanation could be that of *spatial misallocation* as proposed by Hsieh and Moretti (2018). In this paper, the authors examine productivity at the regional instead of national level. They find that certain regions and in particular, certain cities, such as San Fransisco and New York, are incredibly productive, even by historical standards. More productivity means higher wages, so this begs the question as why aren't greater proportions of the population moving to these highly productive areas to increase the productivity of the USA as a whole. The answer compellingly provided in this paper is that workers cannot afford to due to exorbitant housing prices and rents. The authors claim that this misallocation, workers not being in the right place at the right time, lowered overall *growth* by 36% over the period 1964-2009. With housing prices and rents currently being at all time highs in most of these US metropolitan areas, it certainly seems plausible that this is a major headwind against US productivity growth, and accounts for some of the observed slowdown we are measuring here.

A third potential answer which seems somewhat less interesting than the previous two (our gestation period for new technologies and spatial misallocation problems) but is certainly not without merit is the increasing complexity of the US tax code and regulatory environment. Indeed, although tax rates have been trending down in general over this period as was previously discussed, this has been coupled with the tax code becoming vastly more complex. Indeed, the US tax code is now over *75,000 pages* and the US code of federal regulations is over *25,000 pages*. Clearly such a burdensome regulatory environment and complex tax code will be a drag on productivity growth. This is exemplified by a quote by Ronald Reagan in which he said "The US Federal Governments views on the economy is as follows: If it moves, tax it. If it keeps moving, regulate it. And if it stops moving, subsidize it."

There is a vigorous debate among economists both over the causes of this productivity slowdown, as well as trying to forecast it into the future. We note that the views of Gordon (2015), are particularly morose. Citing an enormous range of factors from demographics to human capital attainment to fiscal problems, Gordon claims there is no feasible path to US productivity ever recovering to previous levels. Furthermore he claims he can forecast the efficacy of current technological innovations 50-100 years into the future, and that current innovations such as machine learning and artificial intelligence will not be nearly as impactful as "techno-optimists" imagine. Next we note that other growth theorists such as Jorgenson, Ho and Samuels (2014) and Cowen (2011) forecast continued weak productivity growth for the US economy, but their outlooks are not nearly as dire as that of Gordon.

Finally, there is a class of "techno-optimists" such as Brynjolfsson, Rock and Syverson (2017), who believe the USA is on the verge of restored robust productivity growth. They believe there is a time lag for new technologies to diffuse and be put into widespread use, with the most important technology being machine learning. Citing a wide range of examples these researchers document how artificial intelligence developed by companies such as Google, Microsoft and Amazon has undoubtedly surpassed human intelligence along many dimensions, and once these "smart machines " are put to practical use, they will have a significant impact on US productivity. As additional evidence of this hypothesis, they point to stock market valuations of these high tech firms. As financial market valuations provide among our bests guess of future growth, it is clear the astronomical P/E ratios of these firms means the market has already priced in enormous future growth. That is, Brynjolfsson, Rock and Syverson are not alone in their technological optimism, most Wall Street analysts appear to agree with them.

For what it is worth I agree with the views of Brynjolfsson. It is ridiculous to try to forecast the efficacy of nascent technologies as far into the future as Gordon does, especially when these technologies are primarily proprietary machine learning algorithms, which are distinctly different and will interact with the economy in a fundamentally different way than past physical inventions. Indeed, smart algorithms will *scale* and are more multifaceted than technological advances of the past, facts Gordon will not admit. Jorgenson is too bogged down in lamenting the end of Moore's Law. The fact that this rule of thumb for measuring technological progress (the number of transistors on an integrated chip will double every 18 months) is coming to an end, as Jorgenson often cites, is entirely irrelevant. At this point, we have enough computing power to do everything we could hope to ever do. What matters is not the underlying hardware, but the *software* built on top of it, whose sophistication now was unimaginable even a decade ago. Indeed, advances in hardware are ultimately limited by laws of physics and will eventually exhibit diminishing marginal returns as all other physical assets do. This is not necessarily the case for a self-learning algorithm. Certainly Gordon and Jorgenson have made more accurate predictions than Brynjolfsson of productivity growth so far this decade, but as we saw with the late 90's and early 00's surge, things can change very quickly.

Lastly, it will be illuminating to divide the sample period into four broad eras of varying productivity trends, as growth theorists usually do.

Era	1961-73	1974-93	1994-2004	2005-17
ProdG	1.0129	1.0070	1.0136	1.0038

Productivity growth was very strong from 1961-1973. This is due to a multitude of factors, for which we will list a few which surely contributed. First, energy prices were low and stable. Indeed a barrel of crude oil was just \$2.91 in 1960 and \$3.39 in 1970. At a time when oil was such a major input into production, cheap stable energy prices were clearly a boon for the US economy. Next, there were massive investments in infrastructure. For example, the Highway Act of 1956 was a \$26 *billion* public works project, which is large even by modern standards, that built over 66,000km of highways and roads in the decade that followed. This would allow for huge efficiency gains in trucking and transportation. Next, there were substantial improvements in manufacturing technologies and automation, through key inventions such as feedback controllers starting to be used in production processes. Finally, there were also substantial improvements in human capital. The G.I. bill following WWII granted over 8million servicemen access to college education or retraining in a technical field.

This period of robust productivity growth ended with the oil shock of 1973. Indeed, in the era that followed, which we have defined as 1974-1993, productivity growth was a dismal .70% per year on average, only slightly better than *half* of what it had been in our previously defined era. Apart from the obvious detrimental effects of high and unstable energy prices, other explanations for this slowdown include a decline in *labour quality* and a depletion of investment opportunities. The decline in labour quality is attributed to large proportions of young woman entering the labour force in the 70s, who earned lower wages and were seen to be less productive at the time than their older male counterparts, as documented in Jorgenson, Gollop, and Fraumeni (1987). It is documented in Nordhaus (1982) that there was a substantial decline in patent applications over this period, lending credence to the idea that maybe all the good ideas had been "used up" in the 60's, and now finding new inventions and technologies to boost productivity is substantially more difficult. Other researches such as Griliches (1988) were unconvinced by this explanation and believed the rate of technological progress had not slowed, but just that it was not showing up in the productivity statistics.

This provides an appropriate segue into our next era. By the late 80s and early 90s, it had become a mystery as to why the most important invention of the 20th century, the Personal Computer, as well as the associated technology of the internet, had not yet made a substantial impact of productivity statistics. This is often referred to as the "computer productivity paradox," famously summarized by Robert Solow (1987) as "you can see the computer age everywhere but in the productivity statistics." Indeed, this all began to change in the mid 90s when productivity began to surge due to rapid growth in ICT (information and communications technology) industries. Jorgenson, Ho and Stiroh (2014) divide this era into two further periods, one pre 2000 and one post 2000. They attribute the pre 2000 productivity gains due to growth in ICT industries themselves and massive investment in these industries, and the post 2000 gains to the related industries which were able to successfully implement these ICT advances. Businesses finally making use of advances in computing is generally accepted as fact as the main driver of the observed productivity growth of this period, although other theories have been advanced. These include the idea that a greatly tightening labour market meant job openings were not necessarily able to be filled, so businesses were forced to *squeeze* more productivity out of there existing workforce, as well as that the strong US dollar of the time made US companies less competitive internationally, again forcing them to become more productive to make up the difference. Both these arguments do not seem to hold water, as currently the USA has a very tight labour market, as well as a strong currency, and productivity growth remains sluggish. We accept the hypothesis that the late 90s and early 00s productivity surge was due to the ICT revolution. As we have already discussed the most current trends in productivity growth, this concludes our historical analysis of the sample period. I believe one of the most important takeaways from this discussion is an understanding of how quickly productivity trends can change and begin to deviate from previous long run growth rates. Indeed, it appears in both 1972 and 2004, no one was predicting imminent slowdowns or that previous productivity gains had been exhausted, and in the early 90s it was not the mainstream orthodox view that there was about to be a massive productivity revival. Despite all the difficulties associated with measuring productivity growth, it is clear that forecasting it is much harder.

3.3 Net Real Income Framework

We will now generate TFP estimates in the net income framework. We begin by generating a new net investment aggregate that is equal to our old gross investment aggregate less depreciation. Let PKE_t be the end of period price of capital, which we take to be $PKE_t = PK_{t+1}$. The end of period quantity of capital stock is just the beginning of period less depreciation, that is $QKE_t = QK_t(1-D)$, so letting VKE be the end of period value of capital stock, and VK be the beginning of period value of capital stock, we get a value of time series depreciation:

$$VDTS = VK - VKE$$

and a value of cross sectional depreciation:

$$VDCR = D \cdot PK \cdot QK$$

and importantly note that we will have VDCR $\geq$ VDTS, which is because our cross-sectional measure of depreciation does not account for asset price inflation. We thus take our new net product to be

$$\text{VNY} = \text{PY} \cdot QY - PK \cdot QK + PKE \cdot QKE$$

We then set the price of capital waiting services to be $PWS = R \cdot PK$ and $QWS = QK$. For our input aggregate, we define the price and quantity vectors $p^t = (PL^t, PWS^t)$ and $q^t = (QL^t, QWS^t)$ and we used a chained Tornqvist-Theil index to do the aggregation as above, which gives us real net domestic product from the income side, QZ. For our output aggregate, we form the price and quantity vectors $p^t = (PPC^t, PPG^t, PPI^t, PPX^t, PPM^t, PK^t, PKE^t)$ and $q^t = (QC^t, QG^t, QI^t, QX^t, -QM^t, QK^t, QKE^t)$ which we again aggregate as above using a chained Tornqvist-Thiel to get our real ouput aggregate QZ. We can then calculate the level of net domestic product productivity as simply the ratio of our real input and output aggregates,

$$PRODN = \frac{QY}{QZ}$$

and if we take input and output growth rates as before

$$YG = \frac{QY_t}{QY_{t-1}} \qquad ZG = \frac{QZ_t}{QZ_{t-1}} \tag{24}$$

then we can compute the net domestic TFP growth rate

$$\text{PRODG} = \frac{YG}{ZG} \tag{25}$$

	Mean	σ	Min.	Max.
ProdG Gross	1.0089	0.0121	0.9741	1.0308
ProdG Net	1.0091	0.0125	0.9751	1.0314

We see the arithmetic average rate of TFP growth in the net income framework is .91% per year, which is .02% higher per year than in the gross framework. Thus going from a gross product framework to a net product framework has magnified the effects of TFP growth.

3.4 The Diewert-Morrison Methodology

Here we will briefly review the theory of decomposing growth in real income over time into technical progress, growth in real output prices and growth of primary inputs, following the work of Diewert and Morrison (1986). Assume the market sector of the economy produces M net outputs, $y = (y_1, ..., y_M)$ which are sold at positive prices $P = (P_1, ..., P_M)$, and that the market sector of the economy uses positive quantities of inputs $x = (x_1, ...x_N)$ purchased at positive input prices $W = (W_1, ...W_N)$. In our work elements of y will be components of GDP and elements of x will be capital waiting services and labour services. Given a vector of output prices P and a vector of primary input x, we can define the period t market sector Net Product function, $g^t(P, x)$ as:

$$g^t(P, x) = \max_y \{P \cdot y : (y, x) \in S^t\} \tag{26}$$

where S^t is the period t production possibilities set that is assumed to be closed and convex. Next, under the constant returns to scale assumption on the technology set S^t,

$$g^t(P^t, x^t) = P^t \cdot y^t = W^t \cdot x^t$$

we can deflate this by P_C^t ,the *period t consumption expenditures deflator*, so we will concern ourselves with $\rho^t = g^t(p^t, x^t)$, where we have defined ρ^t to be real income generated by the market sector in period t, p^t is a real output price and now w^t is a real input price. We now define the factors which we would like to compose ρ^t into. The first is our period t productivity growth factors, where we hold output prices and input quantities constant, and only allow the g function to vary:

$$\tau(p, x, t) = g^t(p, x)/g^{t-1}(p, x)$$

there are two obvious measures of τ to consider, a Laspeyres type measure τ_L^t that uses period $t-1$ prices and inputs, and a Paasche type measure τ_P^t that uses period t prices and inputs.

$$\tau_L^t = g^t(p^{t-1}, x^{t-1})/g^{t-1}(p^{t-1}, x^{t-1}) \tag{27}$$

$$\tau_P^t = g^t(p^t, x^t)/g^{t-1}(p^t, x^t) \tag{28}$$

There is no good reason to use one of these measures over the other, so if we want to take a symmetric average that also satisfies the time reversal test from index number theory, we take their geometric mean to get

$$\tau^t = (\tau_L^t \tau_P^t)^{1/2} \tag{29}$$

We now define the family of indexes for the effects of real output price growth on real income

$$\alpha(p^{t-1}, p^t, x, s) = g^s(p^t, x)/g^s(p^{t-1}, x) \tag{30}$$

where we have held our g function and input vector x constant and only allowed p to vary. Again we consider both the Laspeyres and Paasche type measures

$$\alpha_L^t = g^{t-1}(p^t, x^{t-1})/g^{t-1}(p^{t-1}, x^{t-1}) \tag{31}$$

$$\alpha_P^t = g^t(p^t, x^t)/g^t(p^{t-1}, x^t) \tag{32}$$

and proceed to take their geometric mean as

$$\alpha^t = (\alpha_L^t \alpha_P^t)^{1/2}$$

Lastly, we consider the family of indexes for input quantity growth

$$\beta(x^{t-1}, x^t, p, x) = g^s(p, x^t)/g^s(p, x^{t-1}) \tag{33}$$

which holding g and p fixed at $t-1$ or t will give us Paasche and Laspeyres type indexes:

$$\beta_L^t = g^{t-1}(p^{t-1}, x^t)/g^{t-1}(p^{t-1}, x^{t-1}) \tag{34}$$

$$\beta_P^t = g^t(p^t, x^t)/g^t(p^t, x^{t-1}) \tag{35}$$

and their geometric mean is

$$\beta^t = (\beta_L^t \beta_P^t)^{1/2}$$

If we let γ^t be the period t chain rate of growth factor for real income,

$$\gamma^t = \rho^t/\rho^{t-1} \tag{36}$$

we can obtain a decomposition involving τ^t, β^t and α^t, indeed, note that:

$$\begin{aligned}
\gamma^t &= \rho^t/\rho^{t-1} && (37)\\
&= g^t(p^t, x^t)/g^{t-1}(p^{t-1}, x^{t-1}) && (38)\\
&= \left(\frac{g^t(p^t, x^t)}{g^{t-1}(p^t, x^t)}\right)\left(\frac{g^{t-1}(p^t, x^t)}{g^{t-1}(p^{t-1}, x^t)}\right)\left(\frac{g^{t-1}(p^{t-1}, x^t)}{g^{t-1}(p^{t-1}, x^t)}\right) && (39)\\
&= \tau_P^t \alpha(p^{t-1}, p^t, x^t, t-1)\beta_L^t && (40)
\end{aligned}$$

where in the third equality we multipled and divided through by both $g^{t-1}(p^t, x^t)$ and $g^{t-1}(p^{t-1}, x^t)$ and the fourth inuality comes from applying definitions (28),

(30) and (34) in that order. Similarly,

$$\gamma^t = \rho^t/\rho^{t-1} \tag{41}$$
$$= g^t(p^t, x^t)/g^{t-1}(p^{t-1}, x^{t-1}) \tag{42}$$
$$= \left(\frac{g^t(p^{t-1}, x^{t-1})}{g^{t-1}(p^{t-1}, x^{t-1})}\right)\left(\frac{g^{t-1}(p^t, x^{t-1})}{g^{t-1}(p^{t-1}, x^{t-1})}\right)\left(\frac{g^t(p^t, x^t)}{g^t(p^{t-1}, x^{t-1})}\right) \tag{43}$$
$$= \tau_L^t \alpha(p^{t-1}, p^t, x^{t-1}, t)\beta_P^t \tag{44}$$

where we obtained the third equality by multiplying and dividing through by both of $g^t(p^{t-1}, x^{t-1})$ and $g^{t-1}(p^{t-1}, x^{t-1})$, and then (44) comes from applying definitions (27), (30) and (35) in that order. Multiplying (40) and (44) together and then taking square roots of both sides yields:

$$\gamma^t = \left(\tau_P \tau_L\right)^{1/2}\left(\alpha(p^{t-1}, p^t, x^t, t-1)\alpha(p^{t-1}, p^t, x^{t-1}, t)\right)^{1/2}\left(\beta_L^t \beta_P^t\right)^{1/2} \tag{45}$$
$$= \tau^t(\alpha(p^{t-1}, p^t, x^t, t-1)\alpha(p^{t-1}, p^t, x^{t-1}, t))\beta^t \tag{46}$$

where the second equality comes from our definitions of β and τ. Our intuition tells us that

$$(\alpha(p^{t-1}, p^t, x^t, t-1)\alpha(p^{t-1}, p^t, x^{t-1}, t))^{1/2} \approx \alpha^t \tag{47}$$

so applying this in (46) we get the approximation

$$\gamma^t \approx \tau^t \alpha^t \beta^t \tag{48}$$

which we will soon prove holds exactly if g is a translog functional form. Now that we have growth rates for technical progress, real output prices and real input quantities, we can express these in *levels* as suggested by Kohli (1990). Indeed, let T be the level of TFP, A the level of real output prices, and B the level of primary input quantities. That is, set:

$$T^0 \equiv 1; T^t \equiv T^{t-1}\tau^t; t = 1, 2, ... \tag{49}$$
$$A^0 \equiv 1; A^t \equiv A^{t-1}\alpha^t; t = 1, 2, ... \tag{50}$$
$$B^0 \equiv 1; B^t \equiv B^{t-1}\beta^t; t = 1, 2, ... \tag{51}$$

and thus note

$$\frac{\rho^t}{\rho^0} = \frac{\rho^t}{\rho^{t-1}}\frac{\rho^{t-1}}{\rho^{t-2}}...\frac{\rho^1}{\rho^0} \tag{52}$$
$$\approx (\tau^t\alpha^t\beta^t)(\tau^{t-1}\alpha^{t-1}\beta^{t-1})...(\tau^1\alpha^1\beta^1) \tag{53}$$
$$\approx (\tau^t\tau^{t-1}...\tau^1)(\alpha^t\alpha^{t-1}...\alpha^1)(\beta^t\beta^{t-1}...\beta^1) \tag{54}$$
$$\approx T^t A^t B^t \tag{55}$$

where the first equality follows from telescoping the fractions, the second equality comes from applying our approximation (48) to each ratio in (52), and the last equality comes from recursively applying definitions (49), (50) and (51). We will now show that the approximation (48) holds exactly if our NDP function $g^t(p,x)$ has the translog functional form

$$\log g^t(p,x) \equiv a_0^t + \sum_{m=1}^{M} a_m^t \log p_m + (1/2)\sum_{m=1}^{M}\sum_{k=1}^{M} a_{mk}\log p_m \log p_k$$
$$+ \sum_{n=1}^{N} b_n^t \log x_n + (1/2)\sum_{n=1}^{N}\sum_{j=1}^{N} b_{nj}\log x_n \log x_j + \sum_{m=1}^{M}\sum_{n=1}^{N} c_{mn}\log p_m \log x_n$$

where the coefficients satisfy a variety of restrictions for g^t to be linearly homogeneous, the details of which can be found in Diewert (1974). It can now be shown that

$$\gamma^t = \tau^t \alpha^t \beta^t$$

and in fact Diewert and Morrison (1986) showed that $\tau^t, \alpha^t, \beta^t$ can be calculated using empirically observable price and quantity data for periods t-1 and t:

$$\log \alpha^t = \sum_{m=1}^{M}(1/2)\left(\frac{p_m^{t-1}y_m^{t-1}}{p^{t-1}\cdot y^{t-1}} + \frac{p_m^t y_m^t}{p^t \cdot y^t}\right)\log(\frac{p_m^t}{p_m^{t-1}}) \tag{56}$$
$$= \log P_T(p^{t-1}, p^t, y^{t-1}, y^t) \tag{57}$$
$$\log \beta^t = \sum_{n=1}^{N}(1/2)\left(\frac{w_n^{t-1}x_n^{t-1}}{w^{t-1}\cdot x^{t-1}} + \frac{w_n^t x_n^t}{w^t \cdot n^t}\right)\log(\frac{p_m^t}{p_m^{t-1}}) \tag{58}$$
$$= \log Q_T(w^{t-1}, w^t, x^{t-1}, x^t) \tag{59}$$
$$\tau^t = \gamma^t / \alpha^t \beta^t \tag{60}$$

I will prove just (56) as the rest is analogous and straightforward once one knows to use Diewert's (1976) quadratic approximation lemma. Indeed, recalling definitions (31) and (32) we have that

$$\log \alpha^t = \frac{1}{2}\bigg(\log \alpha_L^t + \log \alpha_P^t\bigg)$$
$$= \frac{1}{2}\bigg(\log g^{t-1}(p^t, x^{t-1}) - \log g^{t-1}(p^{t-1}, x^{t-1}) + \log g^t(p^t, x^t) - \log g^t(p^{t-1}, x^t)\bigg)$$

now g is quadratic in $\log p$ so we can apply the generalization of Diewert's quadratic identity to get

$$\log \alpha^t = \frac{1}{2}\bigg(\nabla_{\log p}\log g^{t-1}(p^{t-1}, x^{t-1}) + \nabla_{\log p}\log g^t(p^t, x^t)\bigg)^T \bigg(\log p^t - \log p^{t-1}\bigg) \tag{61}$$

the t-th element of $\nabla_{\log p} \log g^{t-1}(p^{t-1}, x^{t-1})$ is

$$\frac{\partial \log g^{t-1}(p^{t-1}, x^{t-1})}{\partial \log p_i^{t-1}} = \frac{\partial g^{t-1}(p^{t-1} x^{t-1})}{\partial p_i^{t-1}} \frac{p_i^{t-1}}{g^{t-1}(p^{t-1}, x^{t-1})} \tag{62}$$

$$= \frac{p_i^{t-1} y_i^{t-1}}{p^{t-1} \cdot y^{t-1}} \tag{63}$$

where in the second equality we have applied both Hotelling's Lemma, $\frac{\partial g^{t-1}(p^{t-1} x^{t-1})}{\partial p_i^{t-1}} = y_i^{t-1}$ and our perfect competition assumption $p^{t-1} \cdot y^{t-1} = g^{t-1}(p^{t-1}, x^{t-1})$ and a similar calculation gives us

$$\frac{\partial \log g^t(p^t, x^t)}{\partial \log p_i^t} = \frac{p_i^t y_i^t}{p^t \cdot y^t} \tag{64}$$

Plugging our equations (63) and (64) into the dot product in (61) we get

$$\begin{aligned} \log \alpha^t &= \sum_{m=1}^{M} (1/2) \left(\frac{p_m^{t-1} y_m^{t-1}}{p^{t-1} \cdot y^{t-1}} + \frac{p_m^t y_m^t}{p^t \cdot y^t} \right) \log\left(\frac{p_m^t}{p_m^{t-1}}\right) \\ &= \log P_T(p^{t-1}, p^t, y^{t-1}, y^t) \end{aligned}$$

as desired.

3.5 Implementing the Diewert-Morrison Methodology

Here we describe how we implemented the above equations in our emprirical work to decompose real income growth into explanatory factors in the *gross* framework. We first define our real income aggregate as

$$RI = PC \cdot QC + PG \cdot QG + PI \cdot QI + PX \cdot QX + PM \cdot QM$$

and let

$$RLINK^t = \frac{RI^t}{RI^{t-1}}$$

Recalling that if we assume a translog functional form we have $\gamma^t = \tau^t \alpha^t \beta^t$ where τ is technical progress, α is growth in real output prices and β is real input quantity growth, we can apply the equations derived (56), (58), (60) above in our 5 output two input model to get:

$$\log \alpha^t = \sum_{m=1}^{5} (1/2) \left(\frac{p_m^{t-1} y_m^{t-1}}{p^{t-1} \cdot y^{t-1}} + \frac{p_m^t y_m^t}{p^t \cdot y^t} \right) \log\left(\frac{p_m^t}{p_m^{t-1}}\right)$$

$$\log \beta^t = \sum_{n=1}^{2} (1/2) \left(\frac{w_n^{t-1} x_n^{t-1}}{w^{t-1} \cdot x^{t-1}} + \frac{w_n^t x_n^t}{w^t \cdot n^t} \right) \log\left(\frac{p_m^t}{p_m^{t-1}}\right)$$

and then

$$\tau^t = \frac{\gamma^t}{\alpha^t \beta^t}$$

so now we can define our put price link factors for consumption, investment, government, exports and imports:

$$PCLINK = exp\left(\frac{1}{2}\left(\frac{PC^t \cdot QC^t}{RI^t} + \frac{PC^{t-1} \cdot QC^{t-1}}{RI^{t-1}}\right) \cdot \log(\frac{PC^t}{PC^{t-1}})\right) \tag{65}$$

$$PGLINK = exp\left(\frac{1}{2}\left(\frac{PG^t \cdot QG^t}{RI^t} + \frac{PG^{t-1} \cdot QG^{t-1}}{RI^{t-1}}\right) \cdot \log(\frac{PG^t}{PG^{t-1}})\right) \tag{66}$$

$$PILINK = exp\left(\frac{1}{2}\left(\frac{PI^t \cdot QI^t}{RI^t} + \frac{PI^{t-1} \cdot QI^{t-1}}{RI^{t-1}}\right) \cdot \log(\frac{PI^t}{PI^{t-1}})\right) \tag{67}$$

$$PXLINK = exp\left(\frac{1}{2}\left(\frac{PX^t \cdot QX^t}{RI^t} + \frac{PX^{t-1} \cdot QX^{t-1}}{RI^{t-1}}\right) \cdot \log(\frac{PX^t}{PX^{t-1}})\right) \tag{68}$$

$$PMLINK = exp\left(\frac{1}{2}\left(\frac{PM^t \cdot QM^t}{RI^t} + \frac{PM^{t-1} \cdot QM^{t-1}}{RI^{t-1}}\right) \cdot \log(\frac{PM^t}{PM^{t-1}})\right) \tag{69}$$

and our input growth effects for labour and capital:

$$QLLINK^t = exp\left(\frac{1}{2}\left(\frac{PL^t \cdot QL^t}{RI^t} + \frac{PL^{t-1} \cdot QL^{t-1}}{RI^{t-1}}\right) \cdot \log(\frac{QL^t}{QL^{t-1}})\right) \tag{70}$$

$$QKLINK^t = exp\left(\frac{1}{2}\left(\frac{PKS^t \cdot QKS^t}{RI^t} + \frac{PKS^{t-1} \cdot QKS^{t-1}}{RI^{t-1}}\right) \cdot \log(\frac{QKS^t}{QKS^{t-1}})\right) \tag{71}$$

and then our productivity link factor is:

$$TLINK^t = \frac{RLINK^t}{(PCLINK)^t(PGLINK)^t(PILINK)^t(PXLINK)^t(PMLINK)^t(QLLINK)^t QKLINK^t}$$

To go from growth rates to levels which we can plot we do a Kohli decomposition, recalling equations (49), (50) and (51) our levels are:

$$\begin{aligned}
RIL^t &= RIL^{t-1} \cdot RLINK^t \\
TT^t &= TT^{t-1} \cdot TLINK^t \\
CC^t &= CC^{t-1} \cdot PCLINK^t \\
GG^t &= GG^{t-1} \cdot PGINK^t \\
II^t &= II^{t-1} \cdot PILINK^t \\
XX^t &= XX^{t-1} \cdot PXLINK^t \\
MM^t &= MM^{t-1} \cdot PMLINK^t \\
TTT^t &= TTT^{t-1} \cdot PTLINK^t \\
LL^t &= LL^{t-1} \cdot QLLINK^t \\
KK^t &= KK^{t-1} QKLINK^t
\end{aligned}$$

where we have taken

$$PTLINK^t = PXLINK^t \cdot PMLINK^t$$

to combine the effects of exports and imports, and also recall that in our base year (1960) all levels are normalized to 1, that is

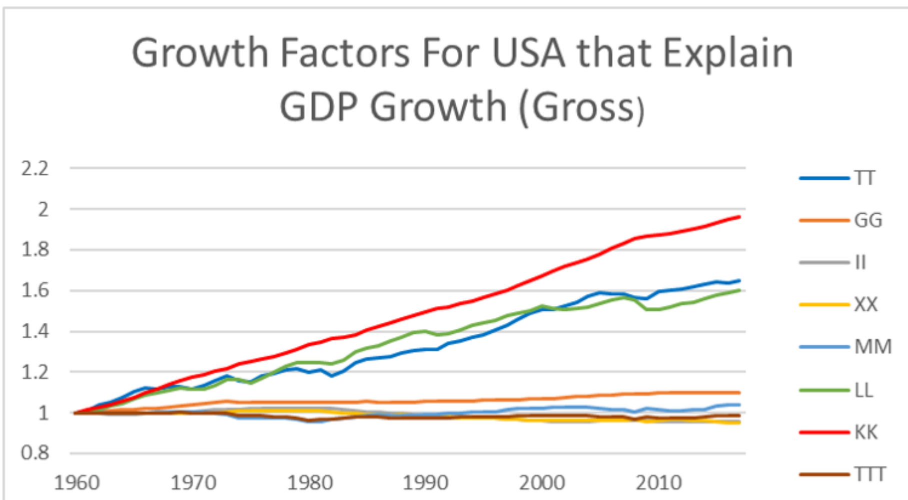
$$RIL = TT = DD = XX = MM = TTT = LL = KK = 1$$

in 1960.

3.6 Results

Here we plot our Kohli type decomposition for real income growth into explanatory factors in both the gross and net income frameworks. The following table lists all levels of explanatory factors as of 2017:

RIL	TT	GG	II	XX	MM	LL	KK	TTT
5.3898	1.6510	1.0987	0.9563	0.9518	1.0375	1.6018	1.9645	0.9874

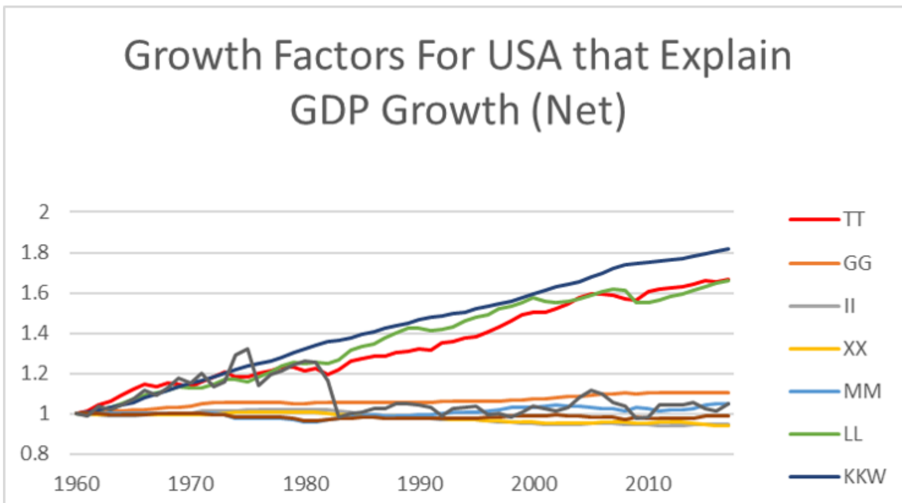


Over the sample period, real income in the USA grew 5.38 fold. If we scale this by population growth, we get per population gross real income grew 2.73 fold from 1960-2017. The factors that explain most of the growth in real gross income are growth of capital services (KK) (red line at the top), growth of labour input (LL) and technical progress (TT). The remaining growth factors are small but one can see that from the mid 90s onwards, MM (growth factor for imports) , increased from about .95 to 1.03, which we can probably attribute to a China effect, and this contributed significantly to gross real income growth over the latter part of the sample period.

We also performed a Kohli type decomposition of real income growth into multiplicative effects, following the the procedure again of section (3.4) above. Note here KKB is increase in real price of beginning of period capital stock and KKE is increase in real price of end of period capital stock, and in our plot we only include the net effect KN. This is because KKB and KKE largely cancel each other out because an increase in the beginning of period price of capital stock will decrease income since it is an input at that point, but an increase in end of period price will increase income since its an output at that point. The following table lists all levels of explanatory factors as of 2017:

RIL	TT	GG	II	XX	MM
5.4871	1.6681	1.1076	0.9451	0.9441	1.051

KKB	KKE	LL	KKW	TTT	KN
1.9388	0.5409	1.6632	1.8157	0.9922	1.0488



We see that real income grew 5.49 fold over the sample period. If we normalize this by population we get per capita net real income grew 2.78 fold over the sample period. The three most important explanatory growth factors in the net income framework are the same as was the case for the gross income framework: the growth factor for waiting services (KKW) ends up at 1.81 (the navy line), the growth factor for labour input (LL) ends up at 1.66 (the green line) and the technical progress level (TT) ends up at 1.66 (the red line). Note that the growth factors for real net income growth are generally more volatile than the corresponding growth factors for real gross income growth.

3.7 Labour and Capital Shares

Lastly we will comment briefly on the labour and capital shares of income we generated in this section. In the gross framework, recalling that our real input quantity aggregate was QZ and our real input price aggregate was PZ, we set

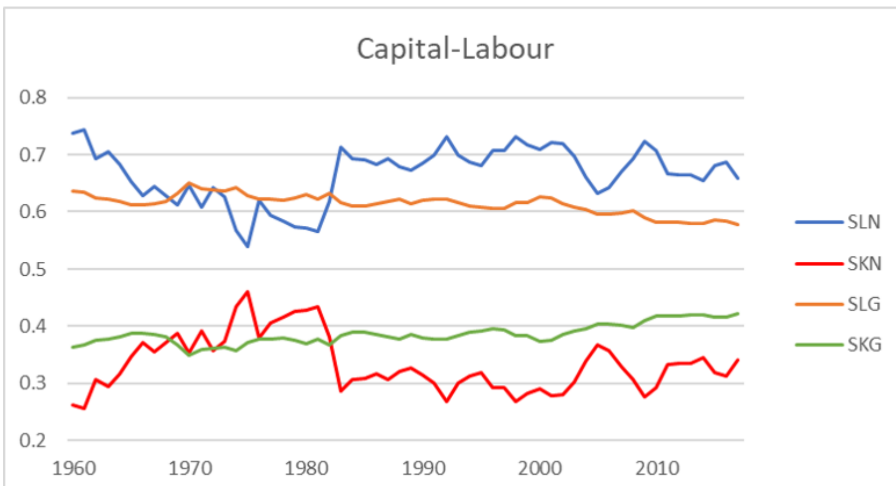
$$VZ = PZ \cdot QZ$$

and calculate the share of labour and capital services as a fraction of GDP as

$$SLG = \frac{PL \cdot QL}{VZ} \qquad\qquad SKG = \frac{PKS \cdot QKS}{VZ} \tag{72}$$

we carry out the same computation in the net framework, and provide descriptive statistics and plots of these shares:

	Mean	σ	Min.	Max.
SLN	0.6658	0.0486	0.5395	0.7434
SKN	0.3342	0.0486	0.2566	0.4605
SLG	0.6137	0.0175	0.5784	0.6509
SKG	0.3863	0.0175	0.3491	0.4216



It is immediately obvious that in the gross framework the labour income share (orange line) has been in steady decline since the early 80's. Recently this decline has become quite dramatic, with our minimum observation of only .578 occurring in the most recent year, 2017. Although this has clearly been going on for a while, it seems to be only recently that the economics profession is taking notice. Indeed, it was also one of the famously aforementioned Kaldor (1957) stylized facts that the labour and capital share of income are stationary over time. We have just shown this is not the case, although still almost all macroeconomic models are calibrated taking a labour share of income as 2/3, which is clearly problematic.

A recent paper by Karabarbounis and Nieman (2014) documents the fall in the labour share of income globally, and attributes about half of this decline to the decrease in the relative price of investment goods to labour, which they attribute to advances in information technology and the computers, inducing firms to shift away from labor and toward capital. This fact has broad implications for macroeconomic dynamics and also income inequality. If we assume that capital holding are very concentrated among high income individuals, increasing the capital share of income holding everything else equal widens the gap between wealthy capital holders and poorer workers. Thus it seems very plausible that the decline in labor share is one of the driving forces of the increase in observed income inequality.

This relationship seems quite related to productivity measurement, as if the productivity increases we get are in part due to automation of routine tasks, this will decrease the labor share and increase the capital share, and if capital is highly concentrated among wealthy individuals, this will drive inequality. We also note that for studying inequality, the net income framework seems to be more revelant, as we cannot consume depreciation, that is not all of GDP is income.

4 Econometric Estimation

In this section we will estimate systems of consumer and producer demand functions and producer supply functions. With the end goal being to produce estimates of technological progress and elasticities, we will attempt to find functional forms that are consistent with economic theory but do not arbitrarily restrict elasticities. To this end, we will make ample use of *flexible functional forms* and duality theory. Specifically, we will concern ourselves with generalized leontief, translog, and normalized quadratic functional forms. Most of our estimation will involve the normalized quadratic, as we will see that convexity or concavity can be imposed on this functional form in a parsimonious way without destroying its flexibility. Lastly, to achieve greater flexibility and reduce autocorrelation, we will implement linear splines in our estimating equations.

4.1 Flexible Functional Forms

We begin briefly by giving the definition of and the motivation for using flexible functional forms.

Definition: A *flexible functional form* f is a functional form that has enough parameters to approximate an arbitrary $f^* \in \mathscr{C}^2$ to the 2^{nd} order at an arbitrary point x^* in the domain of definition of both f and f^*.

Indeed, this implies f has enough parameters to satisfy the following $1 + N + N^2$ equations:

$$
\begin{aligned}
f(x^*) &= f^*(x^*) && \text{1 Equation} && (73)\\
\nabla f(x^*) &= \nabla f^*(x^*) && \text{N Equations} && (74)\\
\nabla^2 f(x^*) &= \nabla^2 f^*(x^*) && \text{N(N+1)/2 Equations} && (75)
\end{aligned}
$$

where the number of equations in (75) follows from the fact that since $f, f^* \in \mathscr{C}^2$ we can apply Young's Theorem which says $\partial^2 f(x^*)/\partial x_i \partial x_j = \partial^2 f(x^*)/\partial x_j \partial x_i$ and $\partial^2 f^*(x^*)/\partial x_i \partial x_j = \partial^2 f^*(x^*)/\partial x_j \partial x_i$ so our matrices $\nabla^2 f$ and $\nabla^2 f^*$ are symmetric, leaving only N(N+1)/2 parameters to be determined.

Example: Let $f(x) \equiv a_0 + a^T x + (1/2)x^T Ax$ where a is an Nx1 vector of parameters and $A = (a_{ij})$ is symmetric NxN matrix. Thus f has 1+N+N(N+1)/2 parameters. Note that $\nabla f(x) = a + Ax$ and $\nabla^2 f(x) = A$. Consider an arbitrary $f^* \in \mathscr{C}^2$, looking at (73)-(75), the equations we must satisfy are:

$$a_0 + a^T x^* + (1/2)x^{*T} Ax* = f^*(x^*) \tag{76}$$
$$a + Ax^* = \nabla f^*(x^*) \tag{77}$$
$$A = \nabla^2 f^*(x^*) \tag{78}$$

Thus if we take $A = \nabla^2 f^*(x^*)$ we satisfy (78), and then from (77) we see $a = \nabla f^*(x^*) - Ax^*$, and lastly $a_0 = f^*(x^*) - a^T x^* - (1/2)x^{*T} Ax^*$ is uniquely determined by (76). Thus we have shown f can approximate an arbitrary $f^* \in \mathscr{C}^2$

hence it is a flexible functional form.

One may note that the f in our example above is not linearly homogeneous, (a function is linearly homogeneous if $f(\lambda x) = \lambda f(x)$ for all $\lambda > 0$ and all x in its domain of definition), a property which we frequently desire of functions in economic analysis. Determining the number of parameters that a *linearly homogeneous* flexible functional form must involves applying Euler's Theorems on Differentiable Homogeneous Functions, which we will briefly state and prove.

Theorem (Euler): Let $f(x)$ be a linearly homogeneous function of N variables. If the first order partial derivatives of f exist, then they must satisfy:

$$f(x) = x^T \nabla f(x) \tag{79}$$

If the second order partial derivatives of f exist, then they must satisfy:

$$\nabla^2 f(x) x = 0_N \tag{80}$$

Proof: For $\lambda > 0$ we have $f(\lambda x_1, \lambda x_2, ..., \lambda x_N) = \lambda f(x_1, ... x_N)$ Differentiating both sides with respect to λ yields:

$$\begin{aligned} f(x) &= \sum_{n=1}^{N} \frac{\partial f(\lambda x_1, ..., \lambda x_N)}{\partial(\lambda x_n)} \frac{\partial(\lambda x_n)}{\partial \lambda} && \text{(Chain Rule)} && (81) \\ &= \sum_{n=1}^{N} \frac{\partial f(\lambda x_1, ..., \lambda x_N)}{\partial(\lambda x_n)} x_n && && (82) \\ &= x^T \nabla f(x) && \text{(Setting } \lambda = 1 \text{ in (82))} && (83) \end{aligned}$$

So we have proven (79). Again let $f(\lambda x_1, \lambda x_2, ..., \lambda x_N) = \lambda f(x_1, ... x_N)$ and now differentiate with respect to x_n:

$$\begin{aligned} f_n(\lambda x_1, ..., \lambda x_N) \partial(\lambda x_n)/\partial x_n &= \lambda f_n(x_1, ..., x_N) && \text{(Chain Rule)} && (84) \\ f_n(\lambda x_1, ..., \lambda x_N) \lambda &= \lambda f_n(x_1, ..., x_N) && && (85) \\ f_n(\lambda x_1, ..., \lambda x_N) &= f_n(x_1, ..., x_N) && && (86) \end{aligned}$$

where we have adopted the notation $f_n = \partial f(x_1, ... x_N)/\partial x_n$ and note (84)-(86) hold for $n = 1, ... N$. We can now differentiate (86) with respect to λ to get

$$\begin{aligned} 0 &= \sum_{j=1}^{N} \frac{\partial f_n(\lambda x_1, ..., \lambda x_N)}{\partial(\lambda x_j)} \frac{\partial(\lambda x_j)}{\partial \lambda} && \text{for n = 1,...,N} && (87) \\ 0 &= \sum_{j=1}^{N} \frac{\partial f_n(\lambda x_1, ..., \lambda x_N)}{\partial(\lambda x_j)} x_j && && (88) \\ 0_N &= \nabla^2 f(x) x && \text{(Setting } \lambda = 1 \text{ in (88))} && (89) \end{aligned}$$

and we are done.

We are now equipped to comment on the number of free parameters that a flexible linearly homogeneous functional form must have. Suppose f satisfies $\nabla f(x^*) = \nabla f^*(x^*)$. Then taking a dot product with x^* on both sides yields $x^{*T}\nabla f(x^*) = x^{*T}\nabla f^*(x^*)$. But by Euler's first theorem above this reduces to $f(x^*) = f^*(x^*)$, so we have eliminated 1 Equation, (73). Similarly, if we have $\nabla^2 f(x^*) = \nabla^2 f^*(x^*)$, by Euler's second theorem above we have $\nabla^2 f(x^*)x = \nabla^2 f^*(x^*)x = 0_N$. Thus if we have $\partial f(x^*)/\partial x_i\partial x_j = \partial f^*(x^*)/\partial x_i\partial x_j$ for all $i \neq j$ then this necessarily determines the diagonal elements, that is we have $\partial f(x^*)/\partial x_j\partial x_j = \partial f^*(x^*)/\partial x_j\partial x_j$ as well for $j = 1, ...N$, so we have eliminated another N equations. Noting that from (73)-(75) we started with $1+N+N(N+1)/2$ equations and we just eliminated $1+N$ equations by applying Euler's Theorems on homogeneous functions, we are left with $N(N+1)/2$ equations to satisfy. Thus, in order for $f(x)$ to be a flexible linearly homogeneous functional form, it must have at least $N(N+1)/2$ free parameters. If f has exactly this many free parameters we henceforth refer to it as a a *parsimonious flexible functional form*

Lastly, we note the importance of flexible functional forms. Let $C(y,p)$ be the producers cost function. By Shepherd's Lemma, the producer's system of cost minimizing input demand functions is $x(y,p) = \nabla_p C(y,p)$ Thus Recalling that the cross price elasticity of demand of good n with respect to price k is $e_{nk} = \bigg((p_k/x_n)\partial x_n(y,p)/\partial p_k\bigg) = \bigg((p_k/x_n)\partial C(y,p)/\partial p_n\partial p_k\bigg)$, we see that if C is not flexible, price elasticities of input demand will be a priori restricted in an arbitrary way. Thus to get accurate estimates of elasticities consistent with economic theory, we will concern ourselves with estimating flexible functional forms.

4.2 Producer Models

Here we estimate a a producer model with 7 goods: consumption, government output, gross investment, exports, imports, labour and capital services. We define the variable profit function as

$$V(p,k) = \max_y \{p^T y : k = F(y)\} \tag{90}$$

where F(y) is a capital requirements function, defined as the minimum amount of capital k that is required to produce the vector of net outputs y. Note if commodity i is an output, then $y_i > 0$ and if commodity i is an input then $y_i < 0$ We briefly note two properties of the variable profit function (90).

Theorem: The variable profit function defined by (90) is convex in p for fixed k

Proof: Let $V(p^1,k) = p^{1T}y^1$ and let $V(p^2,k) = p^{2T}y^1$. Consider a convex combinition of p^1 and p^2, $p^* = \lambda p^1 + (1-\lambda)p^2$, and let $V(p^*,k) = p^{*T}y^*$ As $F(y^*) = k$, we have that y^* is feasible but not necessarily optimal for the problem $V(p^1,k)$, so

$$p^{1T}y^1 \geq p^{1T}y^* \tag{91}$$

As $F(y^*) = k$, we have that y^* is feasible but not necessarily optimal for the problem $V(p^2,k)$, so

$$p^{2T}y^2 \geq p^{2T}y^* \tag{92}$$

Multiplying (91) by λ and (92) by $(1-\lambda)$ and adding the two inequalities together we get

$$\lambda p^{1T}y^1 + (1-\lambda)p^{2T}y^2 \geq \lambda p^{1T}y^* + (1-\lambda)p^{2T}y^* \tag{93}$$

$$\lambda V(p^1,k) + (1-\lambda)V(p^2,k) \geq (\lambda p^1 + (1-\lambda)p^2)^T y^* \tag{94}$$

$$\lambda V(p^1,k) + (1-\lambda)V(p^2,k) \geq V(p^*,k) \tag{95}$$

so $V(p,k)$ is convex for fixed k, as desired.

Using an alternative characterization of convexity for $V(p,k)$, this implies that $\nabla^2_{pp}V(p,k)$ is a positive semi-definite matrix. We will also make use of the following fact:

Theorem: If F(y) is linearly homogeneous (so that the technology exhibits constant returns to scale), then $V(p,k)$ decomposes into $v(p)k$ where $v(p)$ is the unit profit function.

Proof: Suppose F(y) is linearly homogeneous. Starting with our definition of $V(p,k)$:

$$\begin{aligned}
V(p,k) &= \max_y \{p^T y : k = F(y)\} \\
&= \max_y \{p^T y : 1 = F(\frac{y}{k})\} && \text{Since F is linearly homogeneous} \\
&= \max_y \{k\left(p^T(\frac{y}{k})\right) : 1 = F(\frac{y}{k})\} \\
&= k \max_y \{p^T(\frac{y}{k}) : 1 = F(\frac{y}{k})\} && \text{as } k > 0 \\
&= k \max_z \{p^T z : 1 = F(z)\} && \text{letting z=y/k} \\
&= kv(p)
\end{aligned}$$

as desired.

4.2.1 Translog Estimation

We begin by estimating the translog variable profit function with a time trend to try and capture the effects of technical progress. Define:

$$\log v(p,t) = \alpha_0 + \beta_0 t + \sum_{i=1}^{6} \alpha_i \log p_i + (1/2) \sum_{i=1}^{6} \sum_{j=1}^{6} \gamma_{ij} \log p_i \log p_j + \sum_{i=1}^{6} \beta_i t \log p_i \tag{96}$$

where we impose the following restrictions to ensure symmetry and linear homogeneity

$$\begin{aligned}
\gamma_{ij} &= \gamma_{ji}; & 1 \le i < j \le 6 & \qquad \text{(15 Restrictions)} \\
\sum_{i=1}^{N} \alpha_i &= 1; & & \qquad \text{(1 Restriction)} \\
\sum_{j=1}^{6} \gamma_{ij} &= 0 & i = 1, ..., 6 & \qquad \text{(6 Restrictions)} \\
\sum_{i=1}^{N} \beta_i &= 0 & & \qquad \text{(1 Restriction)}
\end{aligned}$$

We have period t variable profit as

$$V^t = p^{tT} y^t = \sum_{i=1}^{6} p_i^t y_i^t \tag{97}$$

and letting $V(k,p,t) = kv(p,t)$ we can thus write down our estimating equations for (96) for t=1,...58 as:

$$\log V^t = \log k^t + \log v^t + e_0^t \tag{98}$$

$$= \log k^t + \alpha_0 + \beta_0 t + \sum_{i=1}^{6} \alpha_i \log p_i + (1/2) \sum_{i=1}^{6} \sum_{j=1}^{6} \gamma_{ij} \log p_i \log p_j + \sum_{i=1}^{6} \beta_i t \log p_i + e_0^t \tag{99}$$

$$\log\left(\frac{V^t}{k^t}\right) = \alpha_0 + \beta_0 t + \sum_{i=1}^{6} \alpha_i \log p_i + (1/2) \sum_{i=1}^{6} \sum_{j=1}^{6} \gamma_{ij} \log p_i \log p_j + \sum_{i=1}^{6} \beta_i t \log p_i + e_0^t \tag{100}$$

To derive estimating equations for the output shares $s_i^t(k,p)$ for $i = 1, ..., 6$ and $t = 1, ..., 58$ we will require both Hotelling's Lemma and our unit profit decomposition shown above, indeed

$$s_i^t(k,p) = \frac{p_i^t y_i^t}{V(p,k,t)} \tag{101}$$

$$= p_i^t \frac{k\partial v(p,t)}{\partial p_i} \frac{1}{V(p,k,t)} \qquad \text{Hotelling's Lemma to kv(p)} \tag{102}$$

$$= p_i^t \frac{k\partial v(p,t)}{\partial p_i} \frac{1}{kv(p,t)} \qquad \text{unit profit decomposition} \tag{103}$$

$$= \frac{\frac{\partial v(p,t)}{v(p,t)}}{\frac{\partial p_i}{p_i^t}} \tag{104}$$

$$= \frac{\partial \log v(p,t)}{\partial \log p_i} \tag{105}$$

Equation (105) states that to get estimating equations for shares, we merely need to differentiate our translog unit profit function, (96), with respect to log prices. Indeed, differentiating (96) with respect to log prices yields:

$$s_i^t(k,p) = \alpha_i + \sum_{j=1}^{5} \gamma_{ij} \log p_j + \beta_i^t + e_i^t \tag{106}$$

As shares necessarily sum to 1, that is $\sum_{i=1}^{6} s_i^t = \frac{\sum_{i=1}^{5} p_i^t y_i^t}{V^t} = 1$ they have a linear dependence relationship and we will need to drop one share from our estimating equations.

We now explain how to substitute our restrictions into our estimating equations (100). We first use the condition $\sum_{i=1}^{6} \alpha_i = 1$ to eliminate α_6. We then use $\sum_{i=1}^{6} \beta_i = 1$ to eliminate β_6. Next we use the symmetry conditions $\gamma_{ij} = \gamma_{ji}$ and $\sum_{j=1}^{6} \gamma_{ij} = 0; i = 1, ...N$ to eliminate γ_{i6} and γ_{6i} for $i = 1, ...6$. After carrying out this procedure, for t=1,...,58, our estimating equations become:

$$\log \frac{V^t}{p_6^t k^t} = \alpha_0 + \beta_0 t + \sum_{i=1}^{5} \alpha_i \log \frac{p_i^t}{p_6^t} + \sum_{i=1}^{5} \beta_i t \log \frac{p_i^t}{p_6^t} + \frac{1}{2} \sum_{i=1}^{5} \gamma_{ij} (\log \left(\frac{p_i^t}{p_6^t}\right))^2 \tag{107}$$

$$+ \sum_{\substack{i=1 \\ i<j}}^{5} \sum_{j=1}^{5} \gamma_{ij} \log \left(\frac{p_i^t}{p_6^t}\right) \log \left(\frac{p_j^t}{p_6^t}\right) + e_0^t$$

and

$$s_i^t(k,p) = \alpha_i + \sum_{j=1}^{5} \gamma_{ij} \log p_j + \beta_i^t + e_i^t \qquad \text{i=1,...,5} \tag{108}$$

Thus the parameters to identify are $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \gamma_{11}, \gamma_{22}, \gamma_{33}, \gamma_{44}, \gamma_{55}, \gamma_{12}, \gamma_{13}, \gamma_{14}, \gamma_{15}, \gamma_{23}, \gamma_{24}, \gamma_{25}, \gamma_{34}, \gamma_{35}, \gamma_{45}$ which is a total of 27 parameters and can be identified from the 6 equations that compose (107) and (108) with a nonlinear regression technique.

Once we have identified the $\gamma_{ij}, \alpha_i, \beta_i$, we can calculate our fitted shares using the formula above. Indeed, letting s_i^{t*} denote fitted values from our nonlinear regression, we can then apply the following formulas to calculate cross price elasticities of net supply and own price elasticities of net supply:

$$e_{ij}^t = \frac{\partial \log y_i(k^t, p^t)}{\partial \log p_j} = (s_i^{t*})^{-1}\gamma_{ij}^* + s_j^{t*} \qquad i \neq j \tag{109}$$

$$e_{ii}^t = \frac{\partial \log y_i(k^t, p^t)}{\partial \log p_j} = (s_i^{t*})^{-1}\gamma_{ii}^* + s_i^{t*} - 1 \qquad i = 1, .., 6 \tag{110}$$

Indeed, we will desire *own price elasticities of net supply* for consumption, government, investment, and exports $(e_{11}, e_{22}, e_{33}, e_{44})$ to be positive, and own price elasticities of net supply for imports and labor (e_{55}, e_{66}) to be negative.

Next, we would like to obtain estimates for

$$\nabla_{pp}^2 V(k^t, p^t)$$

which if our estimated profit equation is to be convex in prices consistent with economic theory, we would like this matrix to be positive semi-definite. First, with our fitted period t shares s_i^{t*} and fitted period t profits V^{t*}, we can estimate our fitted period t net supplies y_i^t* as

$$y_i^{t*} = \frac{V^{t*} s_i^{t*}}{p_i^t} \tag{111}$$

Next, recalling the definition of our price elasticities of net supply we note that

$$e_{ij} = \frac{p_j}{y_i} \frac{\partial y_i(k, p)}{\partial p_j} \tag{112}$$

$$= \frac{p_j}{y_i} \frac{\partial V(p, k)}{\partial p_j \partial p_i} \qquad \text{Hotelling's Lemma} \tag{113}$$

and thus re-arranging we see that

$$\left(\nabla_{pp}^2 V^{t*}\right)_{ij} = \frac{e_{ij}^t y_i^{t*}}{p_j^t} \tag{114}$$

In our empirical work, we will test the definiteness of this matrix in two ways. The first is the necessary condition that all diagonal elements of a positive semi-definite matrix are non-negative, and the second is the sufficient condition that all its eigenvalues are positive.

Theorem: All diagonal elements of a positive semi-definite matrix V are non-negative.

Proof: Since V is positive semi-definite we have $a^T V a \geq 0$ for all $a \in \mathbb{R}^N$. Indeed, letting $a = e_i$ where e_i is the i-th standard normal basis vector for $\mathbb{R}^N$ we see $e_i^T V e_i = V_{ii} \geq 0$ so we have shown the arbitrary i-th diagonal element of V to be non-negative.

Theorem: A real symmetric matrix V is positive semi-definite iff all its eigenvalues are non-negative.

Proof: First let all the eigenvalues of V be non-negative. Since our V is real symmetric, it is *orthogonally diagonalizable*, that is $V = Q\Lambda Q^T$, where Λ is the diagonal matrix of the (non-negative) eigenvalues of V and the columns of Q form an orthonormal basis for the eigenspace so we necessarily have $Q^T Q = I_N$. Indeed, for an arbitrary $a \in \mathbb{R}^N$ we have

$$\begin{aligned} a^T V a &= a^T Q \Lambda Q^T a \\ &= v^T \Lambda v && v = Q^T a \\ &= \sum_{i=1}^{N} v_i^2 \lambda_i \\ &\geq 0 \end{aligned}$$

where the inequality follows since we assumed all λ_i are positive, and squares are necessarily positive. Now to go the other way let V be positive semi-definite and let λ_i be an eigenvalue, whose eigenvector is q. Then $Vq = \lambda_i q$. But V is positive semi-definite so $q^T V q = \lambda_i q^T q = \lambda_i \sum_{i=1}^N q_i^2 \geq 0$ which can only be true if $\lambda_i \geq 0$ so we are done.

Lastly, as we are principally interested in measuring technical progress, we show how we can measure the effects of TFP from our estimated equations. From equation (96) we define $V(k, p, t) = kv(p, t)$ and differentiate *with respect to time* t and evaluate the expression with our period t estimates:

$$\frac{\partial \log V(k, p, t)}{\partial t} = \beta_0^* + \sum_{i=1}^{N} \beta_i^* \log p_i^t \equiv T^t \tag{115}$$

Recalling that the *time derivative* of a logged variable gives a percentage increase, we see (115) is the percentage increase in variable profits due to the improvements in technology going from period $t-1$ to t. For comparison purposes, the problem with our above estimate is that our above estimate for technical progress T^t will be about 3 times the size of our index number estimates for TFP growth. This is because payments to capital are roughly (1/3) of payments to labour. The TFP growth rate is the growth rate of outputs divided by the growth rate of inputs, where inputs are labour and capital services. Thus the denominator in this estimate of technical progress is roughly three times as large as the implicit denominator in T^t whose sole input is capital. To adjust for this we then calculate

$$T^t = \left(\beta_0^* + \sum_{i=1}^{N} \beta_i^* \log p_i^t\right)\left(\frac{V^{t*}}{V^{t*} - p_6^t y_6^t}\right) \tag{116}$$

which we will compare to our index number estimates.

We now implement linear splines into our translog estimating equations to allow for varying time trends for technical progress. To do so we look at plots of fitted residuals from our previous 6 estimating equations, which exhibitted a high degree of autocorrelation. We then identify "break-points", in which linear trends reverse (where a "zig" turns into a "zag"). By identifying 6 break-points, we allow for roughly 1 time trend per decade to model technical progress over the sample period. Indeed, the time trend for our previous estimating equation will now be modeled as the piecewise linear function $d_i(t)$, defined as follows:

$$\beta_{i1}t \qquad 1 \le t \le t_1$$

$$\beta_{i1}t_1 + \beta_{i2}(t - t_1) \qquad t_1 \le t \le t_2$$

$$\beta_{i1}t_1 + \beta_{i2}(t_2 - t_1) + \beta_{i3}(t - t_2) \qquad t_2 \le t \le t_3$$

$$\beta_{i1}t_1 + \beta_{i2}(t_2 - t_1) + \beta_{i3}(t_3 - t_2) + \beta_{i4}(t - t_3) \qquad t_3 \le t \le t_4$$

$$\beta_{i1}t_1 + \beta_{i2}(t_2 - t_1) + \beta_{i3}(t_3 - t_2) + \beta_{i4}(t_4 - t_3) + \beta_{i5}(t - t_4) \qquad t_4 \le t \le t_5$$

$$\beta_{i1}t_1 + \beta_{i2}(t_2 - t_1) + \beta_{i3}(t_3 - t_2) + \beta_{i4}(t_4 - t_3) + \beta_{i5}(t_5 - t_4) + \beta_{i6}(t - t_5) \quad t_5 \le t \le t_6$$

And thus our translog function variable profit function with time trends $d_i(t)$ is

$$\log v(p,t) = \alpha_0 + d_0 t + \sum_{i=1}^{6} \alpha_i \log p_i + (1/2)\sum_{i=1}^{6}\sum_{j=1}^{6} \gamma_{ij} \log p_i \log p_j + \sum_{i=1}^{6} d_i t \log p_i \tag{117}$$

We can again use our constraints $\sum_{i=1}^{6} \alpha_i = 1$ to eliminate α_6. We then next eliminate d_6. Finally we use the symmetry conditions $\gamma_{ij} = \gamma_{ji}$ and $\sum_{j=1}^{6} \gamma_{ij} = 0; i = 1, ...6$ to eliminate γ_{i6} and γ_{6i} for $i = 1, ...6$. After carrying out this procedure, for t=1,...,58, our estimating equations become:

$$\log \frac{V^t}{p_6^t k^t} = \alpha_0 + d_0 t + \sum_{i=1}^{5} \alpha_i \log \frac{p_i^t}{p_6^t} + \sum_{i=1}^{5} d_i t \log \frac{p_i^t}{p_6^t} + \frac{1}{2} \sum_{i=1}^{5} \gamma_{ij} (\log \left(\frac{p_i^t}{p_6^t} \right))^2$$

$$+ \sum_{\substack{i=1 \\ i<j}}^{5} \sum_{j=1}^{5} \gamma_{ij} \log \left(\frac{p_i^t}{p_6^t} \right) \log \left(\frac{p_j^t}{p_6^t} \right) + e_0^t \tag{118}$$

and

$$s_i^t(k,p) = \alpha_i + \sum_{j=1}^{5} \gamma_{ij} \log p_j + d_i^t + e_i^t \qquad \text{i=1,...,5} \tag{119}$$

As we have added 6 time trends into 6 equations, we note that we have added an addition $6 * 6 = 36$ parameters. Thus we must identify the 27 variables above, plus the 36 linear time trend co-efficients, giving us a total of 63 parameters to identify from the above 6 equations, which can then be estimated using non-linear regression techniques.

Once these equations have been estimated, fitted shares can be calculated and used to calculate cross price elasticities of net supply and own price elasticities of net supply using formulas (109) and (110) listed above. We can also use equation (111) to calculate fitted period t net supplies y_i^{t*} and then we have all the ingredients to again calculate $\nabla^2_{pp} V(k^t, p^t)$ using equation (114) which we desire to be positive semi-definite. Lastly, to get our estimate of technological progress, we again differentiate our translog proit function with respect to time, and we multiply through by $\frac{V^{t*}}{V^{t*} - p_6^t y_6^t}$ to make this estimate of technological progress directly comparable with our index number estimates. Indeed, we note that

$$\begin{aligned} T^t &= \frac{\partial \log V(k,p,t)}{\partial t} \\ &= \frac{\partial d_0^*(t)}{\partial t} + \sum_{i=1}^{6} \frac{\partial d_i^*(t)}{\partial t} \log p_i^t \\ &= \left(\frac{\partial d_0^*(t)}{\partial t} + \sum_{i=1}^{6} \frac{\partial d_i^*(t)}{\partial t} \log p_i^t \right) \left(\frac{V^{t*}}{V^{t*} - p_6^t y_6^t} \right) \quad \text{to make comparisons} \end{aligned}$$

4.2.2 Leontief No Substitution Estimation

We now turn our attention to estimating a Leontief no-substitution profit function. We will estimate this function both with and without imposing constant returns to scale. With non constant returns to scale (allowing for fixed costs), our variable profit function is

$$V(p,k,t) = a^T p + b^T pk + c^T pkt + d^T pt \tag{120}$$

where from this definition it should be clear why this function correpsonds to no substitution, since no terms will survive taking 2^{nd} derivatives. Recalling Hotelling's Lemma which states

$$y(p,k) = \nabla_p V(p,k)$$

we can derive estimating equations for (120) simply as

$$y_i = \alpha_i + \beta_i k + c_i tk + d_i t + \epsilon_i \qquad \text{i=1,...6} \tag{121}$$

which gives us 6 equations and 24 parameters to identify (we must identify α_i, β_i c_i and d_i for $i = 1, ..., 6$) for $t = 1, ..., 58$ which can be estimated using non-linear regression techniques. We can then calculate returns to scale as

$$RS = \frac{b^{*T} pk + c^{*T} pkt}{V^{*t}}$$

where b^* and c^* are our fitted co-efficients and V^{*t} is fitted period t variable profits.

As we are primarily concerned with measuring technical progress, we may differentiate (120) with respect to t to get a measure of TFP, which to make comparable to our index number estimates we then adjust by multiplying through by the ratio of fitted gross profits to fitted gross profits less fitted value of labour input, indeed:

$$T^t = \frac{\partial V(p,k,t)}{\partial t}\left(\frac{V^{t*}}{V^{t*} - p_6^t y_6^{t*}}\right) \tag{122}$$

$$= (c^{*T} pk + d^{*T} p)\left(\frac{V^{t*}}{V^{t*} - p_6^t y_6^{t*}}\right) \tag{123}$$

It will be seen in the results section that this model is not successful as k and t are too collinear so it cannot distinguish between technical progress and returns to scale, so we now impose constant returns to scale. Indeed, with constant returns to scale and a single time trend in each input or output equation to capture technical progress our variable profit function is now

$$V(p,k,t) = b^T pk + c^T ptk \tag{124}$$

a straightforward application of Hotelling's Lemma, gives us the estimating equations for t=1,...,58

$$y_i = b_i k + c_i tk + \epsilon_i \qquad i = 1, ..., 6 \tag{125}$$

So we now have 6 equations and 12 parameters to identify (all α_i and β_i) which can be estimated using non-linear regression techniques. Once we have these estimates, we then note that our measure of technical progress is

$$T^t = \frac{\partial V(p,k,t)}{\partial t}\left(\frac{V^{t*}}{V^{t*} - p_6^t y_6^{t*}}\right) \tag{126}$$

$$= c^T pk\left(\frac{V^{t*}}{V^{t*} - p_6^t y_6^{t*}}\right) \tag{127}$$

4.2.3 Normalized Quadratic Estimation

We now begin the estimation of what turns out to be our preferred functional form, the normalized quadratic (unit) profit function. Recall our example from section (4.1) where we saw that a quadratic form is a *flexible functional form*, meaning it has the 2^{nd} order approximation property for all $\mathscr{C}^2$ functions, so we are not arbitrarily restricting elasticities. Furthermore, we will see that part of the power of the normalized quadratic is that we can *impose* curvature conditions onto it (convexity), so we will now get estimated equations that are consistent with economic theory (recall a profit function should be convex in prices), whereas the translog function could not do this, and we considered such a major failing of the estimation (see the results section for more details). We define the normalized quadratic profit function as:

$$V(k,p,t) = a^T pk + b^T ptk + (1/2)\frac{p^T Bp}{\alpha^T p}k \tag{128}$$

and using our usual decomposition we note that $V(k,p,t) = v(p)k$ where

$$v(p,t) = a^T p + b^T pt + (1/2)\frac{p^T Bp}{\alpha^T p} \tag{129}$$

and we will concern ourselves with estimating (129). We note that α is a 6x1 vector of pre-determined parameters for scaling, defined as $\alpha_i = \frac{\mu_i}{\sum \mu_i}$ where μ_i is the mean of y_i. We note that a necessary and sufficient condition for (128) to be convex in prices is for the B matrix to be positive semi-definite, which we will get by construction.

In successive regressions, we will build up the B matrix to see what rank substitution matrix the data can support. We begin by letting $B = cc^T$ where c is a $6x1$ vector of parameters. Thus (129) becomes

$$v(p,t) = a^T p + b^T pt + (1/2)\frac{p^T cc^T p}{\alpha^T p} \tag{130}$$

and

$$\nabla_p v(p,t) = a^T + b^T t + \frac{Bp}{\alpha^T p} - (1/2)\frac{p^T Bp}{(\alpha^T p)^2} \tag{131}$$

with $B = cc^T$ equations (131) will be our estimating equations for $i = 1, ...6$ For identification we need the normalization $c^T 1_N = \sum_{i=1}^{6} c_i = 0$ so we set $c_1 = -\sum_{i=2}^{6} c_i$. Thus we have to estimate $a_1, a_2, a_3, a_4, a_5, a_6, b_1, b_1, b_2, b_3, b_4, b_5, b_6, c_2, c_3, c_4, c_5, c_6$ which is 17 parameters and (131) defines 6 estimating equations, which can be estimated with a non-linear regression technique.

We now continue to build up the rank of our B matrix. Clearly $B = cc^T$ was rank one, and now we let $B = cc^T + dd^T$ which defines a rank 2 matrix where d is a $6x1$ vector of parameters. Differentiation gives us the following 6 estimating equations, $i = 1, ..., 6$

$$\nabla_p v(p,t) = a^T + b^T t + \frac{(cc^T + dd^T)p}{\alpha^T p} - (1/2)\frac{p^T(cc^T + dd^T)p}{(\alpha^T p)^2} \tag{132}$$

for identification we must set $d_1 = 0$ and $d_2 = -\sum d_i$ so that $\sum d_i = 0$, thus we have to identify all the previous 17 parameters, plus the 4 additional parameters d_3, d_4, d_5, d_6 thus (132) defines 6 equations and we must estimate these 21 parameters, which can be done with non-linear regression techniques.

Next we considered a rank 3 substitution matrix so that $B = cc^T + dd^T + ff^T$ where f is a $6x1$ vector of parameters and our estimating equations become $i = 1, ...6$ were

$$\nabla_p v(p,t) = a^T + b^T t + \frac{(cc^T + dd^T + ff^T)p}{\alpha^T p} - (1/2)\frac{p^T(cc^T + dd^T + ff^T)p}{(\alpha^T p)^2} \tag{133}$$

For identification we must set $f_1 = f_2 = 0$ and $f_3 = -\sum f_i$ so that $\sum f_i = 0$, thus we must identify all previous 21 parameters plus 3 additional parameters f_4, f_5 and f_6. Thus (133) defines 6 estimating equations for our 24 parameters, which can be estimated using non-linear regression techniques.

Lastly we will consider a rank 4 substitution matrix so that $B = cc^T + dd^T + ff^T + hh^T$ where h is a $6x1$ vector of parameters. Our estimating equations for $i = 1, ..., 6$ are then

$$\nabla_p v(p,t) = a^T + b^T t + \frac{(cc^T + dd^T + ff^T + hh^T)p}{\alpha^T p} - (1/2)\frac{p^T(cc^T + dd^T + ff^T + hh^T)p}{(\alpha^T p)^2} \tag{134}$$

and for identification we must set $h_1 = h_2 = h_3 = 0$ and $h_4 = -\sum h_i$ so that $\sum h_i = 0$. Thus we must identify the previous 24 parameters plus h_5 and h_6 so that we have 26 parameters to identify in total, from estimating equations (134).

To see that our B matrix is positive semi-definite which was the necessary and sufficient for (128) to be convex in prices, simply note that for an arbitrary $z \in \mathbb{R}^N$ we have $z^T B z = z^T(cc^T + dd^T + ff^T + hh^T)z = z^T cc^T z + z^T dd^T z +$

$z^T ff^T z + z^T hh^T z = \sum r_i^2 + \sum s_i^2 + \sum u_i^2 + \sum v_i^2 \geq 0$ where $r_i = c_i z_i, s_i = d_i z_i, u_i = f_i z_i, v_i = h_i z_i$, so B is positive semi-definite as desired.

To generate our measure of technical progress, we simply differentiate (128) with respect to time, and then to make it comparable to our index number estimates we again multiply through through by the ratio of fitted gross profits to fitted gross profits less fitted value of labour input, indeed:

$$T^t = \frac{\partial V(p,k,t)}{\partial t}\left(\frac{V^{t*}}{V^{t*} - p_6^t y_6^{t*}}\right) \tag{135}$$

$$= (b^{*T} pk)\left(\frac{V^{t*}}{V^{t*} - p_6^t y_6^{t*}}\right) \tag{136}$$

For this functional form we have that the cross price elasticity of net supply is:

$$e_{mn}(p,t) = (p_n/y_m)\partial y_m(k,p,t)/\partial p_n$$

$$= b_{mn}(p_n/\alpha^T p)(k/y_m) - (p_n/y_m)\sum_{j=1}^{N} b_{mj}\alpha_n p_j k/(\alpha^T p)^2$$

$$-(p_n/y_m)\sum_{j=1}^{N} b_{nj}\alpha_m p_j k/(\alpha^T p)^2 + (p_n/y_m)(1/2)\alpha_m\alpha_n p^T Bpk/(\alpha^T p)^3$$

Given that $Bp^* = 0_N$, empirically around our normalized prices the last 3 terms in the above expression will be small in magnitude. Thus, what will drive our elasticity e_{mn} is the first term, $b_{mn}(p_n/\alpha^T p)(k/y_m)$. Although b_{mn} is a constant, all other terms, including price p_n, net output y_m, and $\alpha^T p$, the fixed basket price index of all variable input and outputs, will all have substantial trends over the sample periods. Thus we will see trending elasticities in our empirical estimates, which is built in by the functional form we have specified. To deal with this, we will now allow for time trends in our *B matrix*. The methodology we will employ is we will set B equal to a weighted average of two matrices, one which characterizes substitution possibilities at the beginning of the period, and one which characterizes substitution possibilities at the end of the sample period. Rather than do this all at once, we will build up the time trends in our rank 4 substitution matrix one vector at a time, beginning with just the c vector, then the c and d vector, and so on and so forth.

We begin as described above with a single time trend in the c vector of our rank 4 substitution matrix, where we take a weighted average of beginning of period substitution possibilities and end of period substitution possibilities.

Indeed, our variable profit function is now

$$V(p,k,t) = a^T pk + b^T ptk + (1/2)\frac{p^T((1-t/T)cc^T + (t/T)c'c'^T + dd^T + hh^T + ff^T)p}{\alpha^T p}k \tag{137}$$

and our estimating equations from the *unit* profit function will be

$$\begin{aligned}\nabla_p v(p,t) = a^T + b^T t + {} & \frac{((1-t/T)cc^T + (t/T)c'c'^T + dd^T + ff^T + hh^T)p}{\alpha^T p} \\ -(1/2) & \frac{p^T((1-t/T)cc^T + (t/T)c'c'^T + dd^T + ff^T + hh^T)p}{(\alpha^T p)^2}\end{aligned}$$

where c' is a $6x1$ vector of parameters, and for identification we set $c'_1 = -\sum c'_i$ so we have 5 additional parameters to identify. We now have 31 parameters to identify and 6 equations, which can be estimated using non-linear regression techniques.

In successive regressions, we build up a time trend in all of the c, d, f, h vectors of our substitution matrix, so our final B is essentially a weighted average of two matrices, one which characterizes substitution possibilities at the beginning of the sample period and one which characterizes substitution possibilities at the end of the sample period. Our variable profit function is now

$$\begin{aligned}V(p,k,t) = {} & a^T pk + b^T ptk \\ & + (1/2)\frac{p^T((1-t/T)(cc^T + dd^T + hh^T + ff^T) + (t/T)(c'c'^T + dd'^T + hh'^T + ff'^T))p}{\alpha^T p}k\end{aligned} \tag{138}$$

and our estimating equations for the unit profit function are:

$$\begin{aligned}\nabla_p v(p,t) = a^T + b^T t + {} & \frac{((1-t/T)(cc^T + dd^T + hh^T + ff^T) + (t/T)(c'c'^T + dd'^T + hh'^T + ff'^T))p}{\alpha^T p} \\ -(1/2) & \frac{p^T((1-t/T)(cc^T + dd^T + hh^T + ff^T) + (t/T)(c'c'^T + dd'^T + hh'^T + ff'^T))p}{(\alpha^T p)^2}\end{aligned}$$

For identification, we must have have $d'_1 = 0$ and $d'_2 = -\sum d'_i$, as well as $h'_1 = 0, h'_2 = 0$ and $h'_3 = -\sum h'_i$ and finally $f'_1 = 0, f'_2 = 0, f'_3 = 0$, and $f'_4 = -\sum h'_4$ which gives us 40 parameters to estimate in total. With fitted values we can again estimate elasticities simply as

$$e^*_{mn} = \frac{\nabla^2 V^*_{mn} p_n}{y^*_m}k$$

where $\nabla^2 V^*_{mn}$ is our estimated m, n eliminate of the 2nd derivative matrix, which came from simply differentiating the estimated unit profit function twice with respect to prices. Lastly, we would like to be able to generate a measure of technical progress, which is the partial derivative of variable profits with respect to time divided by the fitted value of labour and capital input. As our substitution matrix is now a *function of time*, this calculation is slightly more complex and we will have a new term in our measure of technical progress. Indeed, let C denote our final substitution matrix. We can now difference this matrix term by term, so we let

$$\partial C = C_t - C_{t-1}$$

whose ij element is just the ij element of C at time t less the ij element of C at time t-1. We now proceed as before, differentiating our profit function with respect to time. We first note that the bias term, call it β_0 created by the time trend in the substitution matrix is

$$\beta_0 = \frac{1}{2} p^T \partial C p \frac{k}{\alpha^T p}$$

and thus we see

$$\frac{\partial V(p,k,t)}{\partial t} = \beta_0 + b^{*T} pk \tag{139}$$

which we then make the standard adjustment to get

$$T^t = \left(\frac{1}{2} p^T \partial C p \frac{k}{\alpha^T p} + b^{*T} pk \right) \left(\frac{1}{V^{t*} - p_6^t y_6^t} \right) \tag{140}$$

In the results section it will be seen that although the performance of this last model was satisfactory, the elasticities still displayed substantial trends and our fitted residuals exhibited a very high degree of auto-correlation.

To combat this high degree of auto-correlation of fitted results, as well as to improve the fit of our equations, we adapt the previous model, a normalized quadratic profit function with rank 4 substitution matrix, but now we implement splines to allow for linear time trends in technical progress. To do so we look at plots of fitted residuals from our previous 6 estimating equations, which exhibitted a high degree of autocorrelation. We then identify "break-points", in which linear trends reverse (where a "zig" turns into a "zag"). By identifying 6 break-points, we allow for roughly 1 time trend per decade to model technical progress over the sample period. Indeed, the time trend for our previous estimating equation will now be modeled as the piecewise linear function $d_i(t)$, defined as follows:

$$\beta_{i1} t \qquad 1 \leq t \leq t_1$$

$$\beta_{i1} t_1 + \beta_{i2}(t - t_1) \qquad t_1 \leq t \leq t_2$$

$$\beta_{i1} t_1 + \beta_{i2}(t_2 - t_1) + \beta_{i3}(t - t_2) \qquad t_2 \leq t \leq t_3$$

$$\beta_{i1}t_1 + \beta_{i2}(t_2 - t_1) + \beta_{i3}(t_3 - t_2) + \beta_{i4}(t - t_3) \qquad t_3 \leq t \leq t_4$$

$$\beta_{i1}t_1 + \beta_{i2}(t_2 - t_1) + \beta_{i3}(t_3 - t_2) + \beta_{i4}(t_4 - t_3) + \beta_{i5}(t - t_4) \qquad t_4 \leq t \leq t_5$$

$$\beta_{i1}t_1 + \beta_{i2}(t_2 - t_1) + \beta_{i3}(t_3 - t_2) + \beta_{i4}(t_4 - t_3) + \beta_{i5}(t_5 - t_4) + \beta_{i6}(t - t_5) \quad t_5 \leq t \leq t_6$$

and our variable profit function is now

$$V(p,k,t) = a^T pk + \sum d_i(t)p_i k + (1/2)\frac{p^T((1-t/T)(cc^T + dd^T + hh^T + ff^T) + (t/T)(c'c'^T + d'd'^T + h'h'^T + f'f'^T))p}{\alpha^T p}k$$

Thus by adding 6 time trends in 6 equations, we have added 36 parameters, giving us a total of 76 parameters and 6 equations, which can be estimated with non-linear regression techniques. Our measure of technical progress will be similar to equation (140), except now when we differentiate our variable profit function with respect to time, we pick up co-efficients of our linear time trend $d_i(t)$

4.2.4 Empirical Results

Here we present the findings from our econometric estimation of the various aforementioned producer models. First we note the details of how the data is presented. As each model had 6 estimating equations, rather than report individual R^2, we take an *arithmetic* average of these R^2 and report just that as an overall estimate of goodness of fit. Next, the curvature row refers to whether our estimated equations were convex in prices for fixed k. This involved examining the definiteness of the matrix of 2^{nd} order partials as described above. Obviously the log likelihood measure is useful for comparisons of goodness of fit *within* a functional form; Ex. compare the log-likelihood for a normalized quadratic with a rank 1 substitution matrix to a normalized quadratic with a rank 4 substitution matrix, but should not be used to compare goodness of fit for different functional forms (translog to leontief for example) as this is arbitrary. The E_{ii} rows are the *own price elasticities of net supply* for consumption, government, investment, exports, imports and labour. To be consistent with economic theory, we desire the first four of these elasticities to be positive, and the latter two to be negative. Lastly, we comment on our "ideal" level of technical progress. Recall that both our Fisher and Tornqvist-Thiel conventional index number estimates of total factor productivity growth had an arithmetic average of .89% per year. This is consistent with index number theory as we expect these two indexes to approximate each other to the 2nd order around an equal price and quantity point, as proved earlier in our index number section. Rather than report this .89% figure for comparisons, instead we took the *level* of TFP in 2017 from our Diewert-Morrison decomposition (a Tornqvist-Thiel measure of TFP essentially), where this level was 1.65101. We then proceeded to take a *geometric average* of the level in 2017, which gave us TFP growth of .883% per year. The only difference between this and the .89% figure is that one is an arithmetic average and one is a geometric average, and the geometric average is less, a consequence of *Schlömilch's inequality.*

We now comment on our Translog functional form estimation. The major failing of this estimation was that our estimated translog function did not satisfy the curvature conditions. We checked both necessary (diagonal elements being non-negative) and sufficient (all eigenvalues being non-negative) conditions for this function to be convex in prices, and they all failed. As we started with an incredibly minimalist set of assumptions on the profit function to derive its convexity in prices, see Eq. (90) and the theorem preceeding it, (essentially we just assume optimizing behavior and convex technology sets), the fact that our estimated equation cannot satisfy this should be seen as a very damning indictment of it. Indeed, our estimated equation is not even consistent with basic microeconomic theory, which should make one wary of such estimation going forward. We also note that linear splines (adding a time trend for technical progress, roughly 1 per decade) was very successful. We see that for the additional 36 parameters we gained 217 log likelihood points and increased our average R^2 from .687 to .893, as well as substantially reduced autocorrelation of residuals. This success of linear splines to model technological progress bodes well for their implementation in our subsequent preferred functional forms as well. Failings aside, it is worth noting that our estimate of the growth rate of technology over the sample period from the Translog function with linear splines is strikingly close to our average index number estimate (.885% vs. .883%)!

	Ideal	PMOD1	PMOD2	PMOD3	PMOD3
Function		Translog	Translog	Leontief	Leontief
Returns to Scale		CRS	CRS	1.9532	CRS
Splines		None	Imposed	None	None
Curvature Conditions		Failed	Failed	Imposed	Imposed
Log-Likelihood		867.049	1084.005	-1290.795	-1422.008
Average R^2		.687	.893	.98225	.9767
E11	(+)	0.4778	0.2936		
E22	(+)	0.0423	-0.2084		
E33	(+)	2.1586	1.5408		
E44	(+)	1.4559	2.2788		
E55	(-)	-0.5294	-1.0526		
E66	(-)	-1.7552	-0.6639		
Technology	.883	.870	.885	-.002	.800

It is seen from the results that the Leontief no substitution estimation was unsatisfactory. Besides from the model's inability to generate non-zero elasticities consistent with economic theory, we also note that our estimation failed catastrophically when we did not impose constant returns to scale. Indeed, without imposing constant returns to scale we get a massive estimate for returns to scale and a *negative* estimate for technological progress. This is nonsensical, and stemmed from the fact that because K and t are too collinear, the model is unable to distinguish between tehcnical progress and returns to scale. This led us to impose constant returns to scale in all subsequent estimation.

In building up the substitution matrix B, it is seen that that we were able to increase log likelihood and R^2 by adding the c, d and f vectors to the substitution matrix, so it was rank 3. At this point, adding the h vector (2 additional parameters to identify) to the substitution matrix did not increase log likelihood or R^2, so it is seen that a rank 3 substitution matrix is all the USA data will support over the sample period. For PMOD 4-7, we also note that our curvature conditions follow by construction of the functional form, and our average R^2 is quite good. All elasticities are the correct sign consistent with our theory, and our measure of technical progress is again extremely close to our index number estimate, (.9074% vs. .883%) Trending elasticities and autocorrelation of residuals aside, this estimation should be considered satisfactory.

	Ideal	PMOD4	PMOD5	PMOD6	PMOD7
Function		NQ	NQ	NQ	NQ
Rank		1	2	3	4
Splines		None	None	None	None
Curvature Conditions		Imposed	Imposed	Imposed	Imposed
Log-Likelihood		-1337.382	-1296.805	-1291.529	-1291.529
Average R^2		.965	0.968	0.969	0.969
E11	(+)	0.2372	0.4030	0.3928	0.3928
E22	(+)	0.1287	0.8625	0.8758	0.8758
E33	(+)	1.3612	1.4761	1.5061	1.5061
E44	(+)	0.0073	0.0540	1.1051	1.1051
E55	(-)	-0.0714	-0.3483	-0.3572	-0.3572
E66	(-)	-0.8650	-1.3202	-1.2953	-1.2953
Technology	.883	0.9123	0.9052	0.9074	0.9074

Pmod11 reports the results of our estimation after building up time trends in the c, d, f, and h vectors of our substitution matrix, respectively. These time trends allow B to be a weighted average of two matrices, one which characterizes beginning of period substitution possibilites and one which characterizes end of period substitution possibilities. Again the reason for employing this methodology was to deal with the strong trends we observed in our elasticities. Going from Pmod7 to Pmod11, we see that we gained 27 log likelihood points for the additional 14 time trend parameters in our substitution matrix. All our elasticities are the correct sign, although $E_{44} = 3.9052$ is particularly large in magnitude. Again our estimate of technical progress is incredibly close to our index number average estimate, (.875% vs. .883%). For comparison purposes we also report descriptive statistics for the elasticities from Pmod11 estimation:

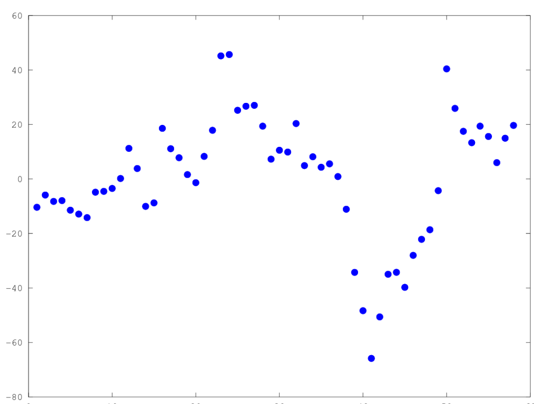
Pmod11	N	Mean	St. Dev	Variance	Min	Max
E11	58	0.45119	4.98E-02	2.48E-03	0.37945	0.53356
E22	58	0.62153	0.19731	3.89E-02	0.14132	0.84865
E33	58	1.4831	0.27607	7.62E-02	1.0114	1.9535
E44	58	3.9052	4.173	17.414	0.61413	19.287
E55	58	-0.63721	0.57462	0.33019	-2.2471	-4.67E-02
E66	58	-1.1509	0.16397	2.69E-02	-1.41	-0.83746

Lastly, Pmod12 reports the results of our estimation after implementing linear splines to allow for a time trend for technical progress. It is seen that we gained 219 log likelihood points for the additional 36 time trend parameters, and that our average R^2 is now a lofty .9957. Clearly linear splines have been a success for our estimation. Our elasticities are all the correct sign, and on average they are slightly smaller in magnitude than our elasticities from Pmod11. The problem of our outlier estimate for E_{44} in early years is also remedied, giving us a more reasonable average of the own price elasticity of net supply for exports. Pmod12 elasticities also have smaller standard deviations on average as they are trending less. We also note that the measure of technical progress we generated should still be considered very close to our index number estimate, (.823% vs. .883%). For comparison purposes, we report the descriptive statistics of our elasticities below.

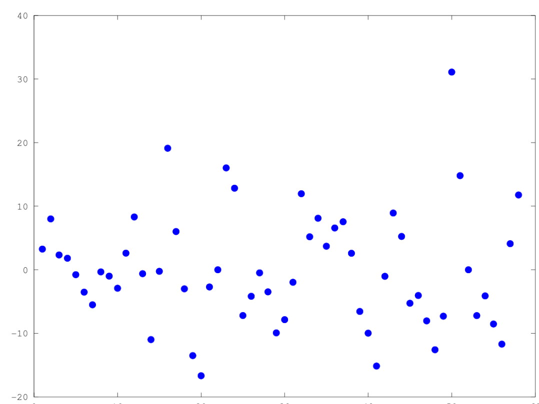
Pmod12	N	Mean	St. Dev	Variance	Min	Max
E11	58	0.51042	8.55E-02	7.30E-03	0.32612	0.64429
E22	58	0.35052	0.21288	4.53E-02	6.91E-02	0.82207
E33	58	0.94909	0.35619	0.12687	0.23619	1.5117
E44	58	1.3824	0.35583	0.12662	0.94873	2.2103
E55	58	-0.32426	0.19445	3.78E-02	-0.63485	-7.60E-02
E66	58	-0.71304	0.17421	3.04E-02	-1.009	-0.42826

	Ideal	PMOD11	PMOD12
Function		NQ	NQ
Rank		4	4
Splines		None	Imposed
Curvature Conditions		Imposed	Imposed
Log-Likelihood		-1264.341	-1018.015
Average R^2		0.9694	0.9957
E11	(+)	0.4512	0.5104
E22	(+)	0.6215	0.3505
E33	(+)	1.4831	0.9490
E44	(+)	3.9052	1.3824
E55	(-)	-0.6372	-0.32426
E66	(-)	-1.1509	-0.7130
Technology	.883	.875	.823

As one final justification for our use of linear splines, consider the following two plots. The first plots fitted residuals for the 6th estimating equation of Pmod11. These resdiuals exhibit a high degree of autocorrelation. (I randomly selected the 6th estimating equation to illustrate this point, all other 5 equations have a similarily high degree of autocorellation of fitted residuals). The next plot is the fitted residuals of Pmod12, which are now a cloud, independent and normally distributed as we would hope for.

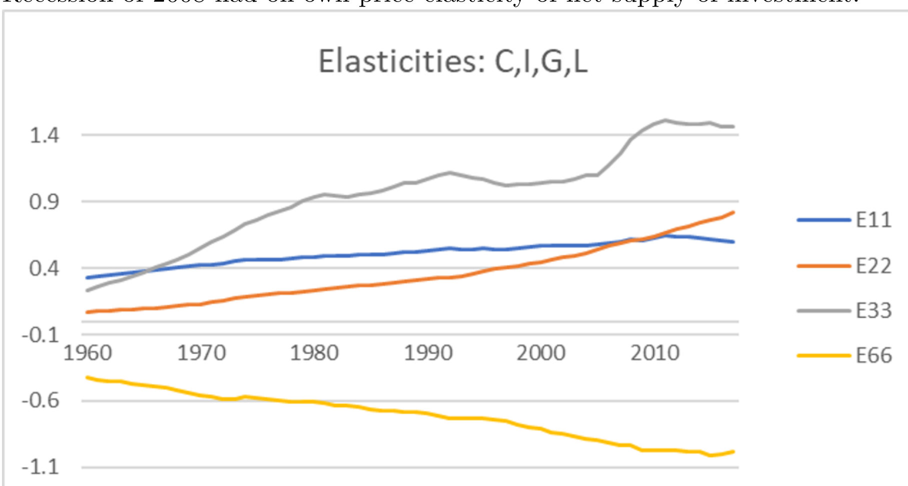


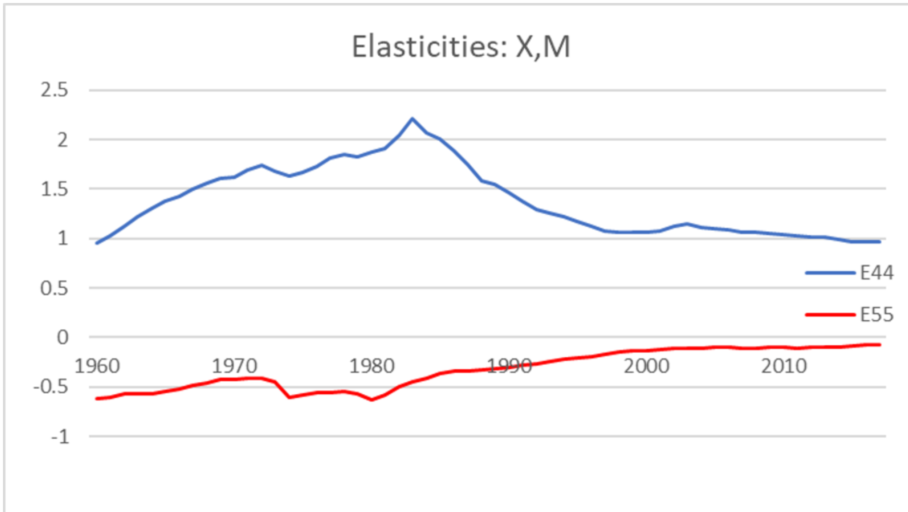
$\rightarrow$



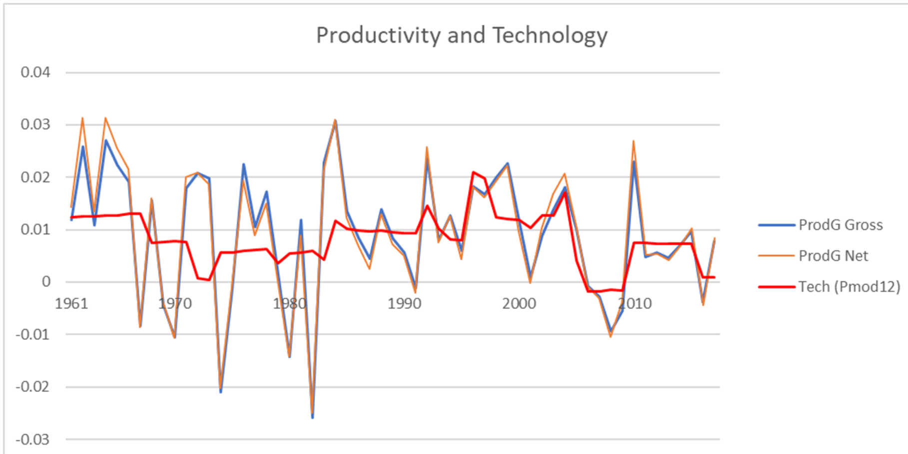
Indeed, Dr. Diewert stated that "Autocorellation is not a stochastic specification problem, but a problem of specifying the correct functional form for the structural part of your estimating equations" and the above plots provide very compelling evidence for this bold assertion.

Next we plot the own price elasticities of net supply for consumption, government, investment, exports, imports and labour generated by Pmod12, the normalized quadratic with a rank 4 substitution matrix and linear time trends for technical progress. Even if this was not our preferred estimate of technical progress, these are surely our best estimates of elasticities, as they are reasonable in magnitude, all the correct sign, do not exhibit as strong trends as our other estimates, and we largely eliminated autocorrelation from our estimating equations. We note that the own price elasticity of net supply is increasing in magnitude for all of consumption, government, investment and labour. Increasing in magnitude consumption supply and labour demand elasticities was a common trend in all Pmod estimation. This may be taken as evidence that the economy is becoming more flexible, but this is hard to reconcile with our own price elasticities of net supply of export and demand for imports, which have been steadily decreasing in magnitude since the early 1980's. It is plausible that these decreasing in magnitude trade elasticities are due to the increase in cross country supply chains. Finally, we note the strong effect that the Great Recession of 2008 had on own price elasticity of net supply of investment.





Lastly, we provide a few general comments about our econometric estimates of TFP growth. In this section, we have seen that various flexible functional forms can provide us with estimates of TFP that are incredibly close to our average index number estimates. The relative merits of index number methods versus econometric methods could be debated, but it appears that on average, "all roads lead to Rome". Instead of declaring one method superior to another, this should be seen as a vindication of productivity measurement as a science, in so far as we have measured productivity in a multitude of different ways and arrived at the same answer. This being said, I do prefer our index number estimates, for the reasons that they impose less structure on the problem, and I believe they are actually more justifiable from an economic theory context than all of our econometric estimates, save maybe the normalized quadratic. As we see with the following graph in which we plot our index number estimates of TFP growth in the gross and net income framework, as well as our econometric growth rates from Pmod12, a shortcoming of our index number estimates is that they oscillate substantially around their average, making it harder to read into year over year changes. Clearly a relative merit of our econometric estimate is that it does a nice job of smoothing out our estimates of technical progress into something more plausible, so we can gain valuable information by looking at the TFP growth rate in any given year, rather than being restricted to just looking at long run averages as we must do with our index number estimates. This discussion of comparing and contrasting methods of productivity measurement is not yet complete, and we will return to it after our non-parametric estimates of total factor productivity growth in section 5.



4.3 Consumer Models

We now move on from the producer side of the economy to the consumer side of the economy. Here we will estimate a simple two good consumer regression model of demand for consumption and leisure, with the end goal being to produce realistic estimates of Hicksian price elasticities of demand and income elasticities of demand. To do the estimation we will employ our favourite flexible functional form, the normalized quadratic cost function. As described in detail in section (4.1), the beauty of this functional form is that it does not arbitrarily restrict elasticities, and that it has enough free parameters for us to impose curvature conditions (concavity here) on it so our estimated equations are always consistent with economic theory. We will describe the problem created by the fact that utility is not directly observable, which will first be dealt with by using indirect utility estimates, and then tackled by using our exact index number formula for utility for the normalized quadratic functional form. Finally we will again make use of linear splines to achieve greater flexibility. Lastly note that this is a model of consumption and leisure demand, and series for leisure were not constructed in our data work, section (2). To deal with this we therefore convert per population labour supply into leisure demand. The methodology for doing so was simply to assume effective hours as twice the *average* number of hours supplied, and then leisure is simply effective hours less hours worked. A shortcoming of this methodology is clearly that maximum number of hours worked is somewhat arbitrary, but the leisure series we construct seems reasonable and displays no strong trends.

At first, it seems that our production function framework previously developed can be adapted to estimate consumer preferences. Examining Equation (90), we simply replace output y by utility u, our production function is replaced by a utility function, and we re-interpret the input vector x as commodity demands and reinterpret the vector p as a vector of commodity prices. Thus the consumer problem of minimizing cost of attaining a given utility level u is simply:

$$C(u,p) = \min_x \{p^T x | f(x) \geq u\} \tag{141}$$

Theorem (Shephard): If the cost function defined by (141) is differentiable with respect to prices, then the consumers system of commodity demand functions for a given u and p is

$$x(u,p) = \nabla_p C(u,p) \tag{142}$$

Proof: The Lagrangian associated with the problem defined by (141) is

$$\mathscr{L} = p^T x - \lambda(f(x) - u)$$

A straightforward application of the *Envelope Theorem* tells us that differentiating our *optimized* value function with respect to a parameter (prices) is the same as differentiating the Lagrangian with respect to the same parameter. Thus

$$\frac{\partial C(u,p)}{\partial p} = \frac{\partial \mathscr{L}}{\partial p} = x$$

While we are here we prove one more basic result which will be fundamental to the analysis that follows:

Theorem: $C(u,p)$ as defined by (141) is concave in prices for fixed u.

Proof: Let x^1 be the solution to $C(u,p^1)$ so that $C(u,p^1) = p^{1T}x^1$ and let x^2 be the solution to $C(u,p^2)$ so that $C(u,p^2) = p^{2T}x^2$. Note we necessarily have $f(x^1) = f(x^2) = u$. Consider the convex combination of prices $p^* = \lambda p^1 + (1-\lambda)p^2$ for $\lambda \in (0,1)$ and let x^* be the solution to $C(u,p^*)$ so that $C(u,p^*) = p^{*T}x^*$ Note since $f(x^*) = u$, x^* is feasible but not necessarily optimal for $C(u,p^1)$, so

$$x^{1T}p^1 \leq x^{*T}p^1 \tag{143}$$

and applying the same logic at prices p^2 we note

$$x^{2T}p^2 \leq x^{*T}p^2 \tag{144}$$

and then simply multiplying (143) by λ and (144) by $(1-\lambda)$ and summing the inequalities gives us the desired result.

$$\lambda x^{1T}p^1 + (1-\lambda)x^{2T}p^2 \leq \lambda x^{*T}p^1 + (1-\lambda)x^{*T}p^2 \tag{145}$$

$$\lambda C(u,p^1) + (1-\lambda)C(u,p^2) \leq C(u,p^*) \tag{146}$$

The systems of demand functions defined by (142) are referred to as *Hicksian* demand functions. The principle problem associated with using (142) as the basis of our estimating equations is that the utility level u which enters the equations as an explanatory variable is not actually observable. This problem can be dealt with by equating the cost function $C(u, p)$ with observable expenditure Y, so that $Y = C(u, p)$. For well behaved (invertible) expenditure functions, we can then solve for u as a function of Y and p, so that

$$u = g(Y, p) \tag{147}$$

and we refer to equation (147) as the *consumer's indirect utility function*. One may note that if preferences are *homothetic* (meaning the utility function is linearly homogeneous) then we have the decomposition $C(u, p) = uc(p)$. In this case the inversion process is very straightforward and we have $u = Y/c(p)$ or $g(Y, p) = Y/c(p)$, but in this section we will be principally concerned with *non-homothetic preferences*. Either way, we can replace the u in the system of Hicksian demand functions (142) by $g(Y, p)$ to obtain the consumer's system of (observable) market demand functions:

$$x = \nabla_p C(g(Y, p), p) \tag{148}$$

and equations such as (144) will form the basis of our estimating equations.

4.3.1 Normalized Quadratic Estimation

We now make explicit assumptions about the functional form of our cost function defined by equation (141). Indeed, we assume $C(u, p)$ is a normalized quadratic flexible functional form modeling non-homothetic preferences, that is:

$$C(u, p) = a^T p + \left(b^T p + \frac{\frac{1}{2} p^T C p}{\alpha^T p} \right) u \tag{149}$$

where a and b are 2x1 vectors of parameters and C is a 2x2 symmetric matrix, and α is a pre-determined 2x1 vector of parameters defined as

$$\alpha^T = \left(\frac{\mu_{x_1}}{\mu_{x_1} + \mu_{x_2}}, \frac{\mu_{x_2}}{\mu_{x_1} + \mu_{x_2}} \right)$$

where μ_{x_1} and μ_{x_2} are the sample means of per capita consumption and leisure demand, respectively. In our base year (1960) we have $p^* = (1, 1)$ and we impose $a^T p^* = 0$, $b^T p^* = 1$ and $Cp = 0_2$ This means a_2 is uniquely determined by a_1, b_2 is uniquely determined by b_1, and c_{12} is uniquely determined by c_{11}, and by symmetry c_{12} determines c_{21} which determines c_{22}, thus we only need to estimate 3 parameters, a_1, b_1 and c_{11}. The rational for these restrictions is shown in Diewert (1980) where he considers an arbitrary twice continuously differentiable cost function $C^*(u, p)$ that satisfies money metric scaling at the positive reference price vector p^*, so that $C^*(u, p^*) = u$. It is shown then that $C(u,p) = \alpha^T p + uc(p)$ can approximate C^* to the second order at (u^*, p^*) where $c(p)$ is a flexible unit cost function. Thus our restrictions above, $a^T p^* = 0, b^T p^* = 1$ and $Cp^* = 0_2$ on (149) have imposed "expenditure = utility in base year", $C(u, p^*) = u$ in (149), so with these restrictions we have successfully cardinalized utility

It is a fact that that our normalized quadratic cost function is concave in prices iff the C matrix is *negative semi-definite.* Using our restrictions above,

$$
\begin{aligned}
C &= \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} && (150) \\
&= \begin{pmatrix} c_{11} & -c_{11} \\ -c_{11} & c_{11} \end{pmatrix} && (151)
\end{aligned}
$$

which for arbitrary $z \in \mathbb{R}^2$ has quadratic form

$$
\begin{aligned}
z^T C z &= c_{11}(z_1^2 - 2z_1 z_2 + z_2^2) && (152) \\
&= c_{11}(z_1 - z_2)^2 && (153) \\
&\leq 0 \qquad \text{iff } c_{11} \leq 0 && (154)
\end{aligned}
$$

Thus we have shown that the concavity of $C(u,p)$ is determined entirely by whether c_{11} is negative. We will discuss our estimates of this paramter in the results section, but it should also be noted that equation (149) is a *flexible functional form*, so we can always impose curvature on it by alternative methods if c_{11} fails to be negative, as we will show how to do shortly.

To derive estimating equations from (149), we apply Shepherd's Lemma to get Hicksian demands, and then we evalute these at our indirect utility function $g(Y,p)$ Indeed,

$$
x = \nabla_p C(u,p) = a + (b + \frac{Cp}{\alpha^T p} - \frac{(1/2)p^T C p \alpha}{(\alpha^T p)^2})u \tag{155}
$$

and letting $Y = C(u,p)$ we note

$$
\begin{aligned}
Y &= a^T p + \left(b^T p + \frac{\frac{1}{2} p^T C p}{\alpha^T p}\right) u && (156) \\
u &= \frac{Y - a^T p}{b^T p + \frac{\frac{1}{2} p^T C p}{\alpha^T p}} && (157)
\end{aligned}
$$

which defines our indirect utility function and thus can be substituted into (155), yielding:

$$
x = a + \left(b + \frac{Cp}{\alpha^T p} - \frac{(1/2)p^T C p \alpha}{(\alpha^T p)^2}\right)\left(\frac{Y - a^T p}{b^T p + \frac{\frac{1}{2} p^T C p}{\alpha^T p}}\right) \tag{158}
$$

Lastly, we multiply the furthest right hand side parentheses through (top and bottom) by $\alpha^T p$ and then we let

$$v = \frac{p}{\alpha^T p} \tag{159}$$

making these substitutions into (158) gives us:

$$x = a + \left(b + Cv - \frac{1}{2}v^T Cv\right)\left(\frac{\frac{Y}{\alpha^T p} - \alpha^T v}{b^T v + \frac{1}{2}v^T Cv}\alpha\right) \tag{160}$$

Now recall that the i-th expenditure share is

$$s_i = \frac{p_i x_i}{Y} \tag{161}$$

and as there are only two goods and expenditure shares necessarily sum to 1, $s_1 + s_2 = 1$, we have a linear dependence relationship between them so we only concern ourselves with s_1 in our estimating equations. Thus adding errors and time superscripts, $t = 1, ...58$ we get that

$$s^t = \frac{p^t}{Y^t}\left(a + \left(b + Cv^t - \frac{1}{2}v^{tT} Cv^t\right)\left(\frac{\frac{Y^t}{\alpha^T p^t} - \alpha^T v^t}{b^T v^t + \frac{1}{2}v^{tT} Cv^t}\right)\right) + \epsilon^t \tag{162}$$

and the first equation of (162), s_1^t will be our sole estimating equation used to estimate a_1, b_1 and c_{11} which can be done with numerical non-linear regression techniques. Letting a^*, b^* and C^* be our estimates of vectors a, b and matrix C respectively from this regression, we can then calculate indirect utility u^* from equation (157), and in turn use this to calculate fitted expenditure Y^* from (156). With these values as well as our fitted demands, x_1^* and x_2^*, we can then calculate our Hicksian elasticities of demand with respect to real income or utility, as

$$EC_i = \frac{\partial x_i}{\partial u}\frac{u^*}{x_i^*} \tag{163}$$

$$= \left(b_i^* + (C^* v)_i - \frac{1}{2}v^T C^* v\right)\frac{u^*}{x_i^*} \qquad i = 1, 2 \tag{164}$$

Similarly we can also calculate the Hicksian price elasticity of demand of good i with respect to price j as

$$EC_{ij} = \frac{\partial x_i}{\partial p_j}\frac{p_j}{x_i^*} \tag{165}$$

$$= \frac{\partial^2 C(u,p)}{\partial p_j \partial p_i}\frac{p_j}{x_i^*} \tag{166}$$

$$= \left(\nabla_{pp}^2 C(u,p)\right)_{ij}\frac{p_j}{x_i^*} \tag{167}$$

where the second equality above follows from *Shephard's Lemma* and the matrix in the third equality is our estimate of the 2^{nd} order partials of $C(u,p)$ which we see by differentiating (156) with respect to prices is of the form

$$\nabla^2_{pp}C(u,p) = \Bigg(\frac{C}{\alpha^T p} - \frac{Cp\alpha^T}{(\alpha^T p)^2} - \frac{p\alpha^T C}{(\alpha^T p)^2} + \frac{(p^T C p)(\alpha\alpha^T)}{(\alpha^T p)^3}\Bigg)u \tag{168}$$

$$= \Bigg(C - Cv\alpha^T - v\alpha^T C + v^T C v(\alpha\alpha^T)\Bigg)(\frac{u}{\alpha^T p}) \tag{169}$$

where we have made use of our normalization $v = \frac{p}{\alpha^T p}$ and of course in our empirical work the above is computed using our estimated C matrix C^*

It will be seen in the results section that our estimated C matrix C^* was not negative semi-definite, meaning our normalized quadratic cost function defined by (149) was not concave in prices. Thankfully, as previously explained the normalized quadratic is a flexible functional form, meaning it has enough free parameters for us to *impose* concavity upon it. Indeed to do so, in our next estimation we now let

$$C = -AA^T \tag{170}$$

where A is lower triangular. To see that C is now necessarily negative semi-definite, simply note for $z \in \mathbb{R}^2$

$$\begin{aligned} z^T C z &= z^T(-AA^T)z \\ &= -z^T AA^T z \\ &= -\omega^T\omega \qquad\qquad \omega = A^T z \\ &= -\sum_{i=1}^{2}\omega_i^2 \\ &\leq 0 \end{aligned}$$

and note that if we require $Ap^* = 0_2$ then we necessarily have $Cp^* = 0_2$ so our construction of C still satisfies the utility cardinalization process described above, and also with this restriction and using the standard rule for matrix multiplication we note we must only identify c_{11} where $c_{11} = -a_{11}a_{11}$ The problem with this methodology is it turns out by impose $C = -AA^T$, all elements of our estimated price second derivative matrix, $\nabla^2_{pp}C(u,p)$, defined by equation (169) turn out to be 0. This implies that our Hicksian price elasticities of demand,

$$EC_{ij} = \Bigg(\nabla^2_{pp}C(u,p)\Bigg)_{ij}\frac{p_j}{x_i^*}$$

will all turn out to be 0.

Due to the limitations of the above model and its lackluster performance, we now take a different approach to trying to estimate the preferences and elasticities associated with the normalized quadratic cost function. Indeed, now rather than calculating utility from the indirect utility function defined by (147), we will calculate utility using the *exact* index number formula for the normalized quadratic functional form. Recall an index number formula is exact if it decomposes (measures exactly) the ratios of utility across the two periods u^1/u^0 using just observable price and quantity data. We now briefly sketch the derivation of this index number formula. Consider our normalized quadratic cost function defined by (149), at times $t = 0$ and $t = 1$ By Shephard's lemma, we can differentiate this with respect to prices at times 0 and 1 to get our Hicksian demands:

$$x^0 = a + \left(b + \frac{Cp^0}{\alpha^T p^0} - \frac{(1/2)p^{0T}Cp^0\alpha}{(\alpha^T p^0)^2}\right)u_0 \tag{171}$$

$$x^1 = a + \left(b + \frac{Cp^1}{\alpha^T p^1} - \frac{(1/2)p^{1T}Cp^1\alpha}{(\alpha^T p^1)^2}\right)u_1 \tag{172}$$

we then pre-multiply both sides of (171) and (172) by the transpose of $(\alpha^T p^0)p^1 + (\alpha^T p^1)p^0$, which then after simplifying down the quadratic terms gives

$$((\alpha^T p^0)p^1 + (\alpha^T p^1)p^0)^T x^0 = \left((\alpha^T p^0)p^1 + (\alpha^T p^1)p^0)^T(a+b) + p^{1T}Cp^0\right)u_0 \tag{173}$$

$$((\alpha^T p^0)p^1 + (\alpha^T p^1)p^0)^T x^1 = \left((\alpha^T p^0)p^1 + (\alpha^T p^1)p^0)^T(a+b) + p^{0T}Cp^1\right)u_1 \tag{174}$$

and we note that since our C is symmetric we have $p^{1T}Cp^0 = p^{0T}Cp^1$ so we can divide (174) by (173) to get

$$\frac{u_1}{u_0} = \frac{((\alpha^T p^0)p^1 + (\alpha^T p^1)p^0)^T x^1}{((\alpha^T p^0)p^1 + (\alpha^T p^1)p^0)^T x^0} \tag{175}$$

which is our *exact* index number formula. Lastly, it is illuminating to write the formula defined by (175) in terms of our normalized prices $v = p/(\alpha^T p)$ which can be done if we multiply and divide both numerator and denominator of the above by $(\alpha^T p^0)(\alpha^T p^1)$ and then simplify to:

$$Q_{NQ} = \frac{(v^1 + v^0)^T x^1}{(v^1 + v^0)^T x^0} \tag{176}$$

The reader may note that if one was to multiply through all v_0 and v_1 by (1/2) the above Q_{NQ} may be interpret as a weighted arithmetic average of real prices pertaining to the two periods and quantity vectors, thus equation (176) defines an *additive superlative index number formula*. As near observations are more alike than distant observations, we will chain the above index number formula in the usual manner, defining

$$NQ_{LINK}(p^{t-1}, p^t, q^{t-1}, q^t; \alpha) = \frac{(v^t + v^{t-1})^T x^t}{(v^t + v^{t-1})^T x^{t-1}}$$

and then our chained index is defined simply as

$$UNQ_{CH}^{t} = UNQ_{CH}^{t-1} \cdot NQ_{LINK}(p^{t-1}, p^{t}, q^{t-1}, q^{t}; \alpha) \tag{177}$$

Going forward we simply let $u = UNQ_{CH}^{t}$ in our estimating equations, and we perform the normalization of letting $u = Y$ in our base year 1960, so we have that expenditure equals utility and thus we have successfully cardinalized utility. Note with this normalization we now longer require the restrictions $a^T p^* = 0$ and $b^T p^* = 1$, thus we must now estimate all of a_1, a_2, b_1, b_2 Again applying Shephard's Lemma to our normalized quadratic cost function we see

$$x = a + \left(b + \frac{Cp}{\alpha^T p} - \frac{(1/2)p^T Cp\alpha}{(\alpha^T p)^2}\right)u \tag{178}$$

$$= a + (b + Cv - \frac{1}{2}v^T Cv\alpha)u \tag{179}$$

where we will still impose concavity on our expenditure function so $C = -AA^T$ for a lower-triangular matrix A. Adding errors and time superscript for t=1,...,58, we see the explicit estimating equations defined by (179) are:

$$x_i^t = a_i + \left(b_i - a_{11}^2(v_1 - v_2) + a_{11}^2(v_1 - v_2)\right)u + \epsilon_i^t \qquad i = 1, 2 \tag{180}$$

Thus we have two equations and 5 parameters (the symmetry conditions on the C matrix and the fact that we still have $Cp^* = 0$ imply that we only need estimate a_{11} which uniquely determines c_{11} by the logic above) which can be estimated using non-linear regression techniques. Letting a^*, b^*, C^* be our estimated parameters, we can then calculate indirect utility simply as

$$UI = \frac{Y - a^{*T}p}{b^{*T}p + \frac{1}{2}(v^T C^* v)(\alpha^T p)} \tag{181}$$

which in our results section we will discuss descriptive statistics of UI and u. Lastly we can apply our previously stated formulas (163) and (165) to estimate Hicksian price elasticities of demand and income elasticities of demand. It will be seen that our data was not able to support estimating the two equations defined by (180) simultaneously (the 2^{nd} estimated equation will have 0 explanatory power), so we will spline utility to achieve greater flexibility and improve the fit of our equations.

Indeed, let u_t be our exogenous utility series calculated from our exact index number formula for the normalized quadratic functional form. By examining the plots of the fitted residuals of our estimated equations defined by (180), we choose 6 break points where a “zig” turns to a “zag”. Using these break points, (which occur at $t = 10, 20, 23, 30, 38, 50$, recalling that $t = 0$ corresponds to 1960) we then spline utility as:

$$
\begin{aligned}
u_1^* &= u_t && 0 \leq t \leq 10 \\
u_2^* &= u_t - u_{10} && 11 \leq t \leq 20 \\
u_3^* &= u_t - u_{20} && 21 \leq t \leq 23 \\
u_4^* &= u_t - u_{23} && 24 \leq t \leq 30 \\
u_5^* &= u_t - u_{30} && 31 \leq t \leq 38 \\
u_6^* &= u_t - u_{38} && 39 \leq t \leq 50 \\
u_7^* &= u_t - u_{50} && 51 \leq t \leq 58
\end{aligned}
$$

Thus with splined utility our normalized quadratic cost function is now:

$$
C(u,p) = a^T p + \sum_{i=1}^{7} b(i)^T p u_i^* + (\frac{\frac{1}{2} p^T C p}{\alpha^T p}) u \tag{182}
$$

$$
= a^T p + \sum_{i=1}^{7} b_1(i) p_1 u_i^* + \sum_{i=1}^{7} b_2(i) p_2 u_i^* + +(\frac{\frac{1}{2} p^T C p}{\alpha^T p}) u \tag{183}
$$

and a straightforward application of Shephard's Lemma as we have now done many times gives us our two estimating equations:

$$
x_1 = a_1 + b_{11}u_1^* + b_{12}u_2^* + b_{13}u_3^* + b_{14}u_4^* + b_{15}u_5^* + b_{16}u_6^* + b_{17}u_7^* - (a_{11}^2(v_1 - v_2))u + .5\alpha_1(v_1 - v_2)^2 u + \epsilon_1
$$

$$
x_2 = a_2 + b_{21}u_1^* + b_{22}u_2^* + b_{23}u_3^* + b_{24}u_4^* + b_{25}u_5^* + b_{26}u_6^* + b_{27}u_7^* + (a_{11}^2(v_1 - v_2))u + .5\alpha_1(v_1 - v_2)^2 u + \epsilon_2
$$

As we have added 12 parameters (6 break points in 2 equations), we now have a total of 12 parameters to be estimated by the two equations which can be done with non-linear regression techniques. Lastly we let $Y = C(u,p)$ and solve (183) for u to get our estimate of indirect utility, UI, which is the same as (181) only with u now entering in peicewise linear fashion, and we always apply equations (163) and (165) to get our estimates of Hicksian price elasticities of demand and income elasticities of demand.

4.3.2 Empirical Results

Here we present the findings from our econometric estimation of the various aforementioned consumer models. First a note on how the data is represented. In the table that follows, the log likelihood is indeed comparable across across all programs of estimation (cmod1-4). The curvature conditions row refers to whether our estimated equations were concave in prices for fixed u. EC_{11} and EC_{22} are Hicksian *own price* elasticities of demand, which we expect to be negative, and EC_{12} and EC_{21} are Hicksian *cross price* elasticities of demand. ECU_1 and ECU_2 are income elasticities of demand for consumption and leisure, where we say good i is *normal* if $ECU_i > 0$ and *inferior* if $ECU_i < 0$. Lastly the U, UI

row lists the correlation co-efficient between our indirect measure of utility UI and our u calculated using the exact index number formula for the normalized quadratic functional form.

Recall that for our normalized quadratic cost function to be concave in prices for fixed u, we needed the C matrix to be negative semi-definite. As our C was a 2x2 symmetric matrix with the restriction $Cp^* = 0_2$, I proved above that our estimated C was negative semi-definite iff $c_{11} \leq 0$. In Cmod1, our estimated c_{11} was *positive*, ($c_{11} = 0.23982$) meaning our C matrix was not even indefinite but *positive definite*. Thus our estimated equation cost function was certainly not concave in prices. As we derived concavity of the cost function under a very minimalist set of assumptions (esentially just optimizing behavior and convex choice sets) , the cost function estimated in cmod1 is not even consistent with basic microeconomic theory. Furthermore, the Hicksian own price elasticities of demand we generated are the incorrect sign (a consequence of our estimated cost function failing to be concave in prices). The income elasticities posit that consumption is a normal good and leisure is an inferior good. The estimation described by cmod1 is clearly unsatisfactory, and going forward we will impose concavity upon our cost function.

In cmod2 we imposed concavity on our cost function by construction (which we can do since the normalized quadratic is a *flexible functional form*). Recall we showed the concavity of $C(u, p)$ depended solely on the sign of the parameter c_{11} in the C matrix, and we then let $C = -AA^T$ so $c_{11} = -a_{11}^2$ which is necessarily negative. Doing so came at a substantial cost, as we see that our log likelihood fell by 20 points, and the explanatory power of the model, our R^2, fell considerably as well from .9374 to .8783. Furthermore, as a relic of this imposition of concavity upon our normalized quadratic cost function, all Hicksian elasticities of demand (both own and cross) are now 0!

	CMOD1	CMOD2	CMOD3	CMOD4
Function	NQ	NQ	NQ	NQ
Splines	None	None	None	Imposed
Curvature Conditions	Failed	Imposed	Imposed	Imposed
Log-Likelihood	178.422	158.2684	145.4900	254.9531
EQ1 R^2	0.9374	0.8783	0.9889	0.9981
EQ2 R^2			0.0001	0.8218
EC11	0.52944	0	-0.03802	0
EC12	-0.52944	0	0.03802	0
EC21	-0.78590	0	0.54447	0
EC22	0.78590	0	-0.54447	0
ECU1	1.8456	1.5417	1.4340	1.7297
ECU2	-0.72853	-0.10565	0.35163	-0.36626
U,UI			0.99988	0.96559

In cmod3 we then calculated utility using our exact index number formula for the normalized quadratic functional form. With an exogenous utility series, we were then able to derive two estimating equations, one for each good (consumption and leisure). Unfortunately, the data was unable to support such estimation. It is seen that the second equation has no explanatory power: its R^2 is zero. The Hicksian price elasticities of demand now have the correct sign, although EC_{11} and EC_{12} (own price elasticity of demand for consumption and elasticity of consumption with respect to the price of leisure) are too small in magnitude to be considered reasonable. Lastly, it is seen that our correlation co-efficient between our utility calculated by the exact index number formula for the normalized quadratic functional form, u, and our indirect utility estimate ui is incredibly high, .99988. A criticism of this calculation may be that we first calculated u, allowed it to enter our estimating equations for x_1 and x_2, and then used these estimating equations to calculate ui, so we should not necessarily be *too* surprised that u and ui are correlated. This criticism does not explain the incredibly high degree to which we observe their correlation, and that speaks to the efficacy of the exact index number formula for the normalized quadratic functional form.

In cmod4 we then splined our utility series (from the exact index number formula for the normalized quadratic functional form) to achieve greater flexibility. This was very successful from the standpoint of improving fit and explanatory power of the model. Indeed, for 12 additional parameters (6 breakpoints in 2 equations), bringing us to a total of 17 parameters, we increased our log likelihood function by 110 points. The data was now able to support estimating a second equation (which has an R^2 of .8218, and our estimating equation for consumption demand x_1 has a lofty R^2 of .9981. The problem with this estimation is then that as a relic of splining utility, all Hicksian price elasticities of demand are 0. Thus from the standpoint of generating price elasticities of demand this estimation is useless. The income elasticities of demand appear okay in magnitude and sign, showing that consumption is a normal good and leisure is an inferior good which seems reasonable, although these estimates are at odds with the results of cmod3. We also see that our correlation between u and ui is very high, which means our exogenous utility series u is most likely accurate.

In sum, this estimation is unsatisfying along almost all dimensions. At first our estimated equation failed to satisfy curvature conditions consistent with basic economic theory, and then imposing curvature conditions destroy our price elasticity estimates. With an exogenous utility series the data was unable to support estimating two equations, and then when we splined utility to improve explanatory power this again destroyed our elasticity estimates.

We note that in cmod3 our positive leisure income elasticity of demand means leisure is a normal good. In this case the *income effect* is dominating and as we get wealthier we work less. Conversely in cmod4 leisure income elasticity of demand is negative and as we get wealthier we work more. These inconsistencies speak to the need for a *dynamic* optimizing model of labour leisure tradeoff, one in which decisions of how much to work are also influenced by expectations of future income, something this model lacks as we are only solving a static optimization problem in each year. In such a model we would expect years with large temporary shocks to income to generate strong substitution effects, but longer lasting and more persistent shocks will generate a relatively stronger income effect. Our model smooths out these effects too much so we are not able to observe this phenomena from our estimated equations.

This estimation illustrates that a *static* optimization model should not be the basis for studying the labour-leisure tradeoff. Indeed, by only examining this data in the cross section we are unable to model agents expectations, limiting us from getting realistic estimates of income and substitution effects of income shocks to leisure from year to year. This estimation should be re-done in a *dynamic* general equilibrium model to correct for this.

4.4 Detailed Consumer Model

Now we turn to estimation of our 17 good consumer regression model. With our detailed consumer data, the end goal will be to produce reliable estimates of income elasticities of demand, own price elasticities of demand, and geometric growth rates of per capita consumption for each good type. Although our two good model was quite general, this estimation should have applications to many specific areas of economics. Indeed, consider the problem of "sector investing" in finance. Rather than choosing just individual companies to invest in, many investment firms will instead choose entire sectors to invest in (Eg. healthcare vs. communications) recognizing that at the individual firm level there may be too much variance of outcome, but still having faith that they can "beat the market" by picking more profitable sectors. Indeed, for such analysis a reliable estimate of income elasticities of demand for various sectors of the economy would be an incredibly useful metric to consider. If a firm is forecasting strong growth for the economy as a whole, then ceteris paribus sectors with high income elasticities of demand should be expected to be more profitable going forward. In the long run, these income elasticities of demand show us "where the economy is heading".

Aside from having private sector applications in finance and industrial organization, this model will also provide a good "check" on our methodology. Although on the surface a 17 good model may seem much more complex, we will show that all of our previously developed analytical tools can be brought to bare on it. Indeed, our primary mode of estimation will be our favourite flexible functional form, the normalized quadratic cost function. We will use the exact index number formula for the normalized quadratic functional form to calculate an exogenous utility series which can enter our estimating equations, and we will also spline utility to achieve greater flexibility and improve fit. A straightforward application of Shephard's Lemma will be used to derive estimating equations, and we will be able to impose concavity on our estimated equations as previously done, so they will be consistent with economic theory. Previously derived formulas for Hicksian price elasticities of demand and income elasticities of demand can also be applied. Lastly, we note that although estimation of our two good consumer regression model was unsatisfactory, it will be shown that the normalized quadratic can indeed handle this more robust, richer data set to estimate preferences. The fact that our normalized quadratic can fit this more complex model and produce reasonable estimates of elasticities should be seen as a vindication of our mode of estimation in spite of previous failings.

4.4.1 Leontief and Normalized Quadratic Estimation

We begin by estimating non-homothetic Leontief preferences where we do not impose money metric scaling. We first develop some preliminaries which will be used in all subsequent models. Let μ_{q_i} denote the sample mean of the i-th good over the period under consideration. We then define a vector of constants as $\alpha^T = (\alpha_1, ..., \alpha_{17})$ where $\alpha_i = \frac{\mu_{q_i}}{\sum_{j=1}^{17} \mu_{q_i}}$ and with this we take our normalized prices to be $v = \frac{p}{\alpha^T p}$ with this, we can now calculate our exact index number formula for the normalized quadratic functional form derived in section 4.3.1. Indeed,

$$\frac{u_1}{u_0} = Q_{NQ}(p^0, p^1, q^0, q^1; \alpha) = \frac{(v^1 + v^0)^T q^1}{(v^1 + v^0)^T q^0} \tag{184}$$

where (184) defines an *additive* superlative index number formula. As with this data set, near observations are more alike than distant observations, so we will apply the chain principle in the usual manner, defining

$$NQ_{LINK}(p^{t-1}, p^t, q^{t-1}, q^t; \alpha) = \frac{(v^t + v^{t-1})^T x^t}{(v^t + v^{t-1})^T x^{t-1}}$$

and then our chained index is defined simply as

$$UNQ^t_{CH} = UNQ^{t-1}_{CH} \cdot NQ_{LINK}(p^{t-1}, p^t, q^{t-1}, q^t; \alpha) \tag{185}$$

Going forward (until we spline utility) we simply let $u = UNQ^t_{CH}$ in our estimating equations, and we perform the normalization of letting $u = Y$ in our base year 1971, so we have that expenditure equals utility and thus we have successfully cardinalized utility. As a sanity check we also calculated chained Fisher quantity indexes over our 17 goods and compared it to (185) and saw that it equals UNQ to 3 decimal places, as predicted by index number theory since both are superlative index number formula which approximate each other to the 2^{nd} order.

We now define our cost function to represent non-homothetic Leontief no-substitution preferences, so

$$C(u,p) = a^T p + b^T pu \tag{186}$$

recalling Shephard's Lemma which says $q = \nabla_p C(u,p)$ adding errors to these Hicksian demand functions gives us our estimating equations:

$$q_i = a_i + b_i u + \epsilon_i \qquad\qquad i = 1, ..., 17 \tag{187}$$

which defines 17 estimating equations and 34 parameters to be estimated which can be done with linear regression. We can calculate our indirect measure of utility then simply as

$$IU = \frac{Y - a^{*T} p}{b^{*T} p} \tag{188}$$

which can be compared to our index number estimates. Lastly, for these preferences Hicksian price elasticities of demand are clearly 0 (no terms survive taking second derivatives) but we can calculate Hicksian income elasticities of demand as

$$EU_i = \frac{\partial q_i}{\partial u} \frac{u}{q_i^*} \tag{189}$$

$$= \frac{\partial^2 C(u,p)}{\partial p_i \partial u} \frac{u}{q_i^*} \tag{190}$$

$$= \frac{b_i^* u}{q_i^*} \tag{191}$$

We will now proceed to spline utility to achieve greater flexibility. To do so, we examined plots of fitted residuals of the above 17 estimating equations, and looked for the most common break points where trends in residuals underwent a reversal, that is where a “zig” turned into a “zag”. We then define our splined utility series as

$$\begin{aligned} u_1 &= u_t & \qquad 1 \le t \le 19 \\ u_2 &= u_t - u_{20} & \qquad 20 \le t \le 37 \\ u_3 &= u_t - u_{37} & \qquad 38 \le t \le 46 \end{aligned}$$

where it is of interest to note that our two breakpoints ended up being at the largest recession of the sample period (2008) and another recession (1990), which is surely not a coincidence. If time is scarce or there are too many autocorellation plots to examine, it seems reasonable to suggest that as a “first guess” of where splines should be placed is in recession years for preference estimation. Thus our Leontief no substitution cost function with splined utility is

$$C(u,p) = a^T p + b_1^T p u_1 + b_2^T p u_2 + b_3^T p u_3 \tag{192}$$

to which an application of Shephard's Lemma gives us our estimating equations

$$q_i = a_i + b_{1i}u_1 + b_{2i}u_2 + b_{3i}u_3 \qquad\qquad i = 1, ..., 17 \tag{193}$$

Thus with 2 breakpoints in 17 equations we have added 34 parameters to be estimated so (193) defines a system of 17 equations with 68 parameters to be estimated which can be done by linear regression. With our estimated parameters $a_1^*, b_1^*, b_2^*, b_3^*$ we can estimate Hicksian income elasticities of demand the exact same as above, and also generate our measure of indirect utility which is now slightly more involved because we have splined utility:

$$
\begin{aligned}
IU_t &= \frac{Y - a^{*T}p}{b_1^{*T}p} && 1 \le t \le 19 \\
IU_t &= IU_{19} + \frac{Y - a^{*T}p - b_1^{*T}pIU_{19}}{b_2^{*T}p} && 19 \le t \le 37 \\
IU_t &= IU_{37} + \frac{Y - a^{*T}p - b_1^{*T}pIU_{19} - b_2^{*T}(IU_{37} - IU_{19})}{b_3^{*T}p} && 38 \le t \le 46
\end{aligned}
$$

and as a check on our calculations we can perform descriptive statistics on IU_t and u_t and see that their correlation co-efficient is almost 1.

We now proceed to add a substitution matrix to our cost function. We first let our substitution matrix B be rank 1 of the form $B = cc^T$ where c is a 17x1 vector of parameters and we will set $c_1 = -\sum_{i=2}^{17} c_i$ for identification. Our cost function is thus a normalized quadratic of the form

$$C(u,p) = a^T p + b_1^T p u_1 + b_2^T p u_2 + b_3^T p u_3 - \left(\frac{\frac{1}{2}(p^T cc^T p)}{\alpha^T p} \right) u \tag{194}$$

It is a fact that the normalized quadratic cost function is concave in prices if the B matrix is negative semi-definite, and as we have $B = -cc^T$ this is clearly negative semi-definite by construction as shown earlier, so the cost function defined by (194) will always be concave in prices. Our estimating equations can again be derived by Shephard's Lemma and in vector form we see they are

$$q = a + b_1u_1 + b_2u_2 + b_3u_3 - cc^T vu + \frac{1}{2} v^T cc^T v \alpha u + \epsilon \tag{195}$$

where recall $v = \frac{p}{\alpha^T p}$ which defines a system of 17 equations with 84 parameters to be identified (recall c_1 is restricted by the other c_i for identification).

We will now "build up" the substitution matrix by continually adding vectors to it until it is rank 9, at which point we will decide that the parameters of the model do not have sufficient degrees of freedom to continue such estimation, even if we would be able to improve fit by adding more. As the model will become quite large, we note it is important to save estimated co-efficients from each regression and use them as starting co-efficients in the next regression, which will ensure that the gradient vector in our non-linear regression can actually converge in finite time. Thus if our normalized quadratic cost function is

$$C(u,p) = a^T p + b_1^T p u_1 + b_2^T p u_2 + b_3^T p u_3 - \left(\frac{\frac{1}{2}(p^T B p)}{\alpha^T p} \right) u$$

we then define our substitution matrix B in successive regressions as

$$\begin{aligned}
&B = cc^T && \text{Rank 1}\\
&B = cc^T + dd^T && \text{Rank 2}\\
&B = cc^T + dd^T + ee^T && \text{Rank 3}\\
&B = cc^T + dd^T + ee^T + ff^T && \text{Rank 4}\\
&B = cc^T + dd^T + ee^T + ff^T + hh^T && \text{Rank 5}\\
&B = cc^T + dd^T + ee^T + ff^T + hh^T + ii^T && \text{Rank 6}\\
&B = cc^T + dd^T + ee^T + ff^T + hh^T + ii^T + jj^T && \text{Rank 7}\\
&B = cc^T + dd^T + ee^T + ff^T + hh^T + ii^T + jj^T + kk^T && \text{Rank 8}\\
&B = cc^T + dd^T + ee^T + ff^T + hh^T + ii^T + jj^T + kk^T + mm^T && \text{Rank 9}
\end{aligned}$$

where $c, d, e, f, h, i, j, k, m$ are all 17x1 vectors of parameters. If we expand out the cost function we note it can be written as

$$\begin{aligned}
C(u,p) = a^T p + b_1^T p u_1 + b_2^T p u_2 + b_3^T p u_3 &- \left(\frac{\frac{1}{2}(p^T cc^T p)}{\alpha^T p} \right) u - \left(\frac{\frac{1}{2}(p^T dd^T p)}{\alpha^T p} \right) u - \left(\frac{\frac{1}{2}(p^T ee^T p)}{\alpha^T p} \right) u \\
&- \left(\frac{\frac{1}{2}(p^T ff^T p)}{\alpha^T p} \right) u - \left(\frac{\frac{1}{2}(p^T hh^T p)}{\alpha^T p} \right) u - \left(\frac{\frac{1}{2}(p^T ii^T p)}{\alpha^T p} \right) u \\
&- \left(\frac{\frac{1}{2}(p^T jj^T p)}{\alpha^T p} \right) u - \left(\frac{\frac{1}{2}(p^T kk^T p)}{\alpha^T p} \right) u - \left(\frac{\frac{1}{2}(p^T mm^T p)}{\alpha^T p} \right) u
\end{aligned}$$

and to derive estimating equations for our normalized quadratic with rank 9 substitution matrix we apply Shephard's Lemma, with in vector form gives us the estimating equations

$$\begin{aligned}
q = a + b_1 u_1 + b_2 u_2 + b_3 u_3 - cc^T v u + \frac{1}{2} v^T cc^T v \alpha u - dd^T v u + \frac{1}{2} v^T dd^T v \alpha u \\
- ee^T v u + \frac{1}{2} v^T ee^T v \alpha u - ff^T v u + \frac{1}{2} v^T ff^T v \alpha u - hh^T v u + \frac{1}{2} v^T hh^T v \alpha u \\
- ii^T v u + \frac{1}{2} v^T ii^T v \alpha u - jj^T v u + \frac{1}{2} v^T jj^T v \alpha u - kk^T v u + \frac{1}{2} v^T kk^T v \alpha u \\
- mm^T v u + \frac{1}{2} v^T mm^T v \alpha u
\end{aligned}$$

where the above defines a system of 17 equations with 176 parameters to be estimated, which follows from the restrictions on the vectors in the substitution matrix for identification, we had to set $d_1 = 0$, $d_2 = -\sum_{i=3}^{17} d_i$, $e_1, e_2 = 0$, $e_3 = -\sum_{i=4}^{17} e_i$ and so on and so forth which adds a total of $\sum_{n=1}^{9}(17-n) = 108$ parameters from the baseline model with no substitution matrix (68 parameters). We estimate the above with a non-linear regression and using the finishing estimated co-efficients from our rank 8 substitution model, the gradient vector of the log likelihood function will converge to 0, albeit slowly. We also note this will be our final model as we are now estimating 176 parameters with only 782 observations (17 goods across 46 years), so each parameter we are estimating has only ~ 4.5 degrees of freedom and further estimation is obviously tenuous. With our estimated equations, we can then calculate fitted values, Hicksian price elasticities of demand, Hicksian income elasticities of demand, our indirect utility estimate and the growth rate of per capita consumption over the sample period. Indeed, if $C^*(u,p)$ is our estimated cost function, then the Hicksian price elasticities of demand are

$$E_{ij} = (\nabla^2 C^*(u,p))_{ij} \frac{p_i}{q_j^*}$$

For calculating our income elasticities of demand, we need to take care that we have splined utility. So for $1 \leq t \leq 19$ we have:

$$\begin{aligned} EU_i = \Bigg(& b_{1i} - c_i^* c^{*T} v - d_i^* d^{*T} v - e_i^* e^{*T} v - f_i^* f^{*T} v - h_i^* h{*}^T v - i_i^* i^{*T} v - j_i^* j^{*T} v \\ & - k_i^* k^{*T} v - m_i^* m^{*T} v + .5\alpha_i \Big(v^T c^* c^{*T} v + v^T d^* d^{*T} v + v^T e^* e^{*T} v + v^T f^* f^{*T} v \\ & + v^T h^* h^{*T} v + v^T i^* i^{*T} v + v^T j^* j^{*T} v + v^T k^* k^{*T} v + v^T m^* m^{*T} v \Big) \Bigg) \frac{u}{q_i^*} \end{aligned}$$

for $20 \leq t \leq 37$ we have:

$$\begin{aligned} EU_i = \Bigg(& b_{2i} - c_i^* c^{*T} v - d_i^* d^{*T} v - e_i^* e^{*T} v - f_i^* f^{*T} v - h_i^* h{*}^T v - i_i^* i^{*T} v - j_i^* j^{*T} v \\ & - k_i^* k^{*T} v - m_i^* m^{*T} v + .5\alpha_i \Big(v^T c^* c^{*T} v + v^T d^* d^{*T} v + v^T e^* e^{*T} v + v^T f^* f^{*T} v \\ & + v^T h^* h^{*T} v + v^T i^* i^{*T} v + v^T j^* j^{*T} v + v^T k^* k^{*T} v + v^T m^* m^{*T} v \Big) \Bigg) \frac{u}{q_i^*} \end{aligned}$$

and for $38 \leq t \leq 46$ we have:

$$\begin{aligned} EU_i = \Bigg(& b_{3i} - c_i^* c^{*T} v - d_i^* d^{*T} v - e_i^* e^{*T} v - f_i^* f^{*T} v - h_i^* h{*}^T v - i_i^* i^{*T} v - j_i^* j^{*T} v \\ & - k_i^* k^{*T} v - m_i^* m^{*T} v + .5\alpha_i \Big(v^T c^* c^{*T} v + v^T d^* d^{*T} v + v^T e^* e^{*T} v + v^T f^* f^{*T} v \\ & + v^T h^* h^{*T} v + v^T i^* i^{*T} v + v^T j^* j^{*T} v + v^T k^* k^{*T} v + v^T m^* m^{*T} v \Big) \Bigg) \frac{u}{q_i^*} \end{aligned}$$

4.4.2 Empirical Results

We now present our results from the previously described estimation. We will be most interested in comparing the leontief estimation with and without splines (con1 and con2), and our initial normalized quadratic estimation (rank 1 and 2 substitution matrices) compared to our final normalized quadratic estimation (rank 9 substitution matrix) so our discussion and presentation will primarily be about these files. Indeed consider the following table which describes the fit of these models. Here our log likelihood measure is comparable across all modes of estimation and the R^2 is taken to be the arithmetic average across all 17 estimating equations. Going from con1 to con2 it is seen that splining utility with just 2 breakpoints (in 17 equations), to add a total of 34 parameters was a success. This allowed us to gain 180 log likelihood points and increase our average R^2 by .05. It is also seen that our exact index number for the normalized quadratic functional forms gives us an exogenous utility series which is highly correlated with our indirect utility estimates. We note that for all files con3 through con11 our estimated equations satisfied curvature conditions consistent with economic theory, (by construction), that is the cost function was concave in prices for fixed u as discussed extensively above. It is seen that by building up the substitution matrix from rank 1 to 9, we were able to substantially improve log likelihood, gaining 322 points. We could probably continue to improve log likelihood by building up the rank of the matrix, but at this point each parameter only had 4.5 degrees of freedom as previously discussed, so we need to be cognizant of not over fitting the model, which we may already be doing at this point. Indeed, if one is to examine the individual R^2 of con11, it is seen that some of the individual R^2 are still quite poor. Indeed, the 6th equation, electricity gas and other fuel, only has an R^2 of .11, down considerably from its estimation in con3 and con4 (0.3073 and 0.5752) respectively. Thus it appears the model has trouble simultaneously fitting all 17 equations at once, and will fit one equation at the detriment of another. This being said, overall the fit of the model is satisfactory given the complexity of the preferences we are trying to estimate across 17 goods.

	Con1	Con2	Con3	Con4	Con11
Function	Leontief	Leontief	NQ	NQ	NQ
Rank			1	2	9
Splines	None	Impose	Imposed	Imposed	Imposed
Curvature Conditions	Trivial	Trivial	Imposed	Imposed	Imposed
Log-Likelihood	2629.351	2812.866	2886.260	2899.611	3208.460
Average R^2	.8007	.8534	.8645	.83612	.8754
U,UI	.9951	.99609	.9993	.99991	1.0000

Before showing own price and income elasticities of demand we present the following table to remind the reader which index corresponds to which good.

1	**Food and non-alcoholic beverages**
2	Alcoholic beverages
3	Tobacco
4	Clothing and Footwear
5	Housing Aggregate
6	Electricity, gas and other fuels
7	Furnishings, households equipment and routine maintenance of the house
8	Health
9	Purchase of vehicles
10	Operation of personal transport equipment
11	Transport services
12	Communications
13	Recreation and culture
14	Education
15	Restaurants and hotels
16	Miscellaneous goods and services
17	Final consumption expenditure of resident households abroad

Below is our income elasticity of demand estimates for a leontief cost function with splined utility, a normalized quadratic with rank 1 substitution matrix, and a normalized quadratic with rank 9 substitution matrix. We begin by examining the con2 elasticities. We note that food and non-alcoholic beverages has a very small in magnitude income elasticity of demand, which makes sense. One can certainly spend more on food as they become wealthier, but there is clearly a baseline level of sustenance for consumers (achieved well before the sample period began) and after which we would expect spending on food to somewhat flatline. Alcohol income elasticities of demand are quite high given the success of various temperance movements over the sample period, as well as public health initiatives to advertise the health risks of excessive drinking. Tobacco has a negative income elasticity of demand in all estimation files. This makes sense as in 1964 (right before our sample period began) the surgeon general of the USA issued their first scientific report linking cigarette smoking to various maladies. A massive and successful public health initiative trying to get people to stop smoking followed shortly after. Thus this income elasticity of demand is somewhat confounded by the effect of this public health initiative over the sample period, not just people quitting smoking as they get wealthier. Indeed, on the one hand we would expect wealthier and more educated people to be the first to be made aware of the dangers of smoking, and thus more likely to quit. On the other hand, smoking is a very expensive habit, and historically has been a hobby of the wealthy, which we would then expect to see a positive income elasticity of demand, although this may be more a relic of European culture than American.

Next we note that income elasticity of demand for eletricity, fuel and gas is negative across all modes of estimation. This may be due to increased fuel efficiency of appliances, electronics and cars over the sample period. Indeed, average fuel efficiency of cars in the USA has increased 3 fold over this sample period, which is probably what is "driving" this elasticity. Next note we have positive and large in magnitude income elasticities of demand for household maintenance, health and purchase of vehicles, as we would hope which is a good check on our methodology. If any of these were to be negative it would be an immediate cause of concern but they are not. The healthcare income elasticity of demand was not quite as large in magnitude as I was expecting. Part of this may be due to insurance is aggregated in a different category (miscellaneous goods and services) and insurance and administrative costs make up a substantial portion of healthcare spending in the USA, and I am not sure if they are all accounted for in the healthcare category. We next note in con2 that our income elasticity of demand for communications is very large. Indeed, the splines caused a huge jump in the elasticity estimates at both breakpoints. We had multiple years with an income elasticity of demand of over 5 for communications, so although the average maybe seem somewhat reasonable the individual estimates are not and this series should be thrown out. Indeed our estimates of income elasticity of demand for our recreation and culture series had similar problems. We experienced very large jumps at breakpoints which generated elasticities which were too large to be considered reasonable. The average may be okay but the individual estimates are not. We note both the communications and recreation and culture series experienced relatively no inflation over the sample period, but very large real quantity growth, which may be what is driving this wonky estimates of income elasticities.

In con2 we had a negative income elasticity of demand for education which is nonsensical. It is well understood that as people become wealthier they will value education more, as well as the fact that over the sample period the relative importance of education grew, so we would like this elasticity to be positive ideally. Indeed, this equation also had a poor fit in con2 with an $R^2 = 0.3193$and part of the trouble in initially estimating this elasticity may be due to the fact that education experienced rapid inflation over the sample period. Lastly we note in con2 our income elasticities of demand for restaurants and hotels and consumption expenditure of residents abroad are both somewhat large in magnitude and positive, as we would hope.

It is seen that in con3 our elasticity estimates are very much similar to that of con2. The one noteworthy difference is that our income elasticity of demand is now positive. Indeed, con3 did a much better job of fitting this equation (its R^2 improved from 0.3193 to 0.7869. This elasticity is still smaller in magnitude that we would like, but atleast it is now the appropriate sign.

In con11 by the time we have built up our rank 9 substitution matrix there are some substantial differences in our income elasticity estimates. We first note our income elasticity estimate for food and beverages is now *negative* on average. In examining the series itself, it is seen that its entries oscillate between positive and negative quite often, and that there is also large jump discontinuities at our break points. Although a negative income elasticity of demand for food and beverages can certainly be rationalized in a very wealthy country, the erratic nature of the estimated series leads me to prefer our estimate from con2 or con3. In con11 it is seen that our alcohol income elasticity of demand is now quite large in magnitude, and our household expenditure income elasticity of demand is much smaller in magnitude. We note that our income elasticity of demand for transportation is now negative in con11. This is hard to rationalize as this series should include items such as airplane tickets which we would expect to have very large income elasticities of demand. Indeed, con11 had a worse fit for this series than our other estimation files, and the elasticities also exhibited large jump discontinuities in our breakpoint years. Thus for transportation services we prefer our earlier estimates of income elasticity of demand. Also in con11 it is now seen that our income elasticity of demand estimate for communications is far too large in magnitude. Indeed we had a hard time fitting this equation, in early years especially where our elasticity estimates were over 8. Lastly note that con11 gives us our most reasonable estimate of income elasticity of demand of education, which is now almost 1 in magnitude.

	Con2	Con3	Con11
EU1	0.0652	0.0026	-0.2569
EU2	0.6727	0.2889	1.0700
EU3	-1.5058	-1.4253	-0.3681
EU4	1.3591	1.0156	1.0386
EU5	0.7314	0.7737	1.3040
EU6	-0.1797	-0.2529	-1.2377
EU7	1.4654	1.5437	0.2596
EU8	1.1278	1.0812	1.3423
EU9	1.2071	1.2125	1.3516
EU10	0.5081	0.5394	0.8656
EU11	0.6910	0.7560	-0.2836
EU12	2.5759	2.7223	4.7264
EU13	2.6887	2.8154	0.6982
EU14	-0.1281	0.3156	0.9322
EU15	0.6458	0.6493	1.0762
EU16	1.1973	1.1651	1.6152
EU17	1.6522	2.0565	1.9885

It is difficult to make an assessment of which estimates of elasticities we ultimately prefer. All programs generated both elasticities we thought to be very realistic and elasticities that are almost impossible to rationalize. Con11 has very poor estimates for transportation and communications, but our best estimates of healthcare and education. Con2 has a very poor estimate of education but very reasonable estimates along most other dimensions, and so on and so forth. Unfortunately I do not consider it scientifically sound to *pick and choose* elasticity estimates among our different modes of estimation to form a coherent worldview. The reason for this is that it would be easy to dream up preferences that can fit and generate elasticities for one series while neglecting all others. Estimating preferences is ultimately about estimating *choices* and if we are splicing together elasticities from different estimated equations, we are neglecting this fundamental fact. If we were forced to choose, we would ultimately take con3 (the normalized quadratic cost function with rank 1 substitution matrix and splined utility) to be our most reasonable. Clearly much more in depth research is required to assess the veracity of these estimates than the informal discussion which appears in this paper, although we would argue that con3 and this work in general is probably quite a good 'first approximation' to the problem. In future estimation it is recommended that a different spline methodology for our exogenous utility series be used (maybe cubic) that won't generate such large jump discontinuities at break points, as well as simply repeating this estimation at a future date when we have more data so our parameters have more degrees of freedom may also be profitable.

We next examine our Hicksian own price elasticities of demand generated by the normalized quadratic with rank 1 substitution matrix (con3) and rank 9 substitution matrix (con11). First note that all these elasticities are negative as desired, because our estimated cost function $C(u, p)$ is negative semi-definite in prices, and as proved earlier negative semi-definite matrices have non-negative diagonal elements, so these elasticities are all of the form $\frac{\partial^2 C}{\partial p_i^2}\frac{p_i}{q_i^*}$ which is the product of a non-negative and strictly positive ratio, so it is negative. Here it is seen that our own price elasticity of demand estimates generated by the rank 9 substitution matrix are *much more* robust. Indeed, the own price elasticity of demand for many goods in con3 is 0 for all intensive purposes, which is clearly unrealistic. We find the price elasticities of demand generated by con11 to be quite satisfactory. Indeed, first note that two of our major discretionary spending categories (recreation and culture and spending abroad) both have very large own price elasticities of demand as we would hope, because discretionary spending should be more sensitive to prices than spending on necessities. To this end we see goods categories which people are unable to forgo (such as health care and electricity and gas) have very inelastic demand over the sample period. All in all, with the rank 9 substitution matrix this model seems to generate Hicksian own price elasticities of demand which are very realistic.

	Con3	Con11
E1P1	-0.0196	-0.6675
E2P2	-0.1114	-0.98513
E3P3	-0.0050	-0.5469
E4P4	-0.3844	-0.80844
E5P5	-0.0172	-0.51766
E6P6	-0.0092	-0.17277
E7P7	-0.0211	-0.99971
E8P8	-0.0210	-0.30068
E9P9	-0.0001	-0.2995
E10P10	-0.0037	-0.31429
E11P11	-0.0026	-0.33948
E12P12	-0.0313	-0.79535
E13P13	-0.0977	-2.3644
E14P14	-0.2438	-0.49152
E15P15	-0.0001	-0.33628
E16P16	-0.0082	-0.474
E17P17	-0.1104	-1.4605

Finally, we examine normalized growth rates for all 17 categories of goods over the sample period. To carry out this calculation, we simply divide every entry of a quantity series through by its base year (1971) value. Its normalized base year entry will then be 1. We then take a *geometric* average of its level in 2016 to get a average geometric growth rate of real quantities over the sample period. The result is the following table:

	Geometric Growth Rate
GR1	0.99982
GR2	1.0137
GR3	0.97557
GR4	1.02445
GR5	1.01497
GR6	0.99746
GR7	1.02453
GR8	1.0224
GR9	1.01858
GR10	1.00856
GR11	1.01289
GR12	1.04918
GR13	1.04867
GR14	0.99814
GR15	1.01168
GR16	1.02277
GR17	1.01905

Here growth rates which exceeded 2% per year are in green and quantities with negative growth rates over the sample period are in red. It is seen that the quantity series with the highest geometric growth rates over the sample period were communications and recreation and culture. The high growth of communications can be explained by the fact that virtually all communication formats in use in 2016 were not in existence in 1971 save for the fax and landlines. The high growth rate of recreation and culture can be explained by the high income growth over the period, and noting that in con2 and con3 we estimated the income elasticity of demand of recreation and culture to be very high. We also note quantities for health care and household furnishings had a very high geometric growth rate over the sample period as we would expect. Negative growth rates for food and beverages, tobacco, and fuel, electricity and gas over the sample period can all be rationalized. What is hard to explain is the negative growth rate for education. Upon examining the series, it is seen that if the sample period had ended as of 2003, we would have had a positive growth rate for our education quantity series. Indeed, the price of education experienced rapid inflation over the sample period, and at some point, probably in the early 2000s it became too prohibitively expensive, causing people to consume less of it. Indeed, the education quantity series also has been falling rapidly since 2010, with a final level of 0.9197 in 2016 relative to 1.0 in 1971. The implications of this going forward for the American economy and American society as a whole are clearly very bad.

5 Non-Parametric Techniques

Lastly we will discuss our non-parametric estimates of efficiency or technical progress. Here no explicit assumptions will be made on the functional form of our production function, in contrast to our index number estimates (in the Diewert-Morrison methodology we assumed a translog production function to get an exact decomposition of real income growth into observable factors), or our econometric estimates (translog, leontief, normalized quadratic). Instead, using observed input and output quantity data, following Farrell (1957) we will exploit the geometry of the situation and construct the convex free disposal hull from observed data, and then calculate various metrics of distance to the production possibility frontier, at first only accounting for technical efficiency, but then also considering prices (allocative efficiency) by imposing various restrictions such as cost minimizing or profit maximizing behavior. This method will identify years in which the economy is not on the production possibility frontier, because resources are being used inefficiently or productive capacity is left ideal. The latter corresponds to our standard understanding of a recession, in which the resources available to the economy (capital and labour in this case) are underutilized. Thus we will see this method will very quickly identify recession years as not being on the PPP as we would expect, although unfortunately it will not tell us *why* inputs are being used inefficiently, although it does offer a key hint (in my opinion), in so far as that when we begin to consider allocative efficiency (prices) we will begin to pick up many more inefficient points. This lends credence to the New-Keynesian thesis that (sticky) wages and prices may cause resources to be underutilized. More importantly, it will be seen that in our preferred framework for non-parametric measures of efficiency (variable profit maximization with convex constant returns to scale technology), when we normalize our index number estimates of productivity and compare it to this measure of efficiency, they are strikingly close. Going from econometric estimation assuming explicit functional forms, to non-parametric estimation and still arriving at very close to exactly the same answer (on average) will be seen as a major triumph of these methodologies as a science.

Lastly, we will replicate the results of Diewert and Fox (2016) by estimating TFP by decomposing the cost constrained value added function. This decomposition is incredibly useful and corrects a potential major failing of our analysis so far. In our index number and econometric estimates of total factor productivity, we see that productivity would fall during recession years. But if we want to think of TFP as primarily technology as one usually does, then there is no reason for this to be the case, as in 2008 we still had access to the exact same technology as we did in 2007. Thus this Diewert and Fox decomposition will decompose productivity into technology, value added efficiency and a mix input growth factor, and it will be shown that in recession years it is *value added efficiency* that causes a decline in productivity, while our technical progress measure will hold steady.

We begin by discussing the primary analytical tool of non-parametric estimates of efficiency, Linear Programming. In what follows I will borrow freely from class notes by Professors Erwin Diewert and Ozgur Yilmaz.

5.1 An Interlude on Linear Programming

Linear Programming involves optimizing linear objective functions subject to linear constraints. As these types of problems are ubiquitous in engineering, economics and statistics, linear programming has become a mainstay tool of these disciplines. The success of these techniques is surely due to the fact that it turns out that there is an efficient and easy to implement algorithm (the simplex method) to solve this class of optimization problems. In this section we will briefly describe the simplex method, give two examples of it and comment on its time complexity bounds. We will also state the dual problem and prove theorems of weak and strong duality.

Definition: The "Primal" problem for a linear program is:

$$\max_{x} \{c^T x : Ax \leq b\,;\; x \geq 0_N\} \tag{196}$$

where $c \in \mathbb{R}^N$ and $b \in \mathbb{R}^M$ are vectors of constants. A is an MxN matrix of constants and $x \in \mathbb{R}^N$ is our choice variable. We will often introduce "slack variables", one for each inequality constraint, so we transform the above problem into

$$\max_{x_0,x,s} \{x_0 : x_0 - c^T x = 0;\; Ax + I_M s = b;\; x \geq 0_N;\; s \geq 0_M\} \tag{197}$$

where $s \in \mathbb{R}^M$ and x_0 takes on the value of our objective. We will also make use of considering the problem with equality constraints, which gives us the linear program:

$$\max_{x_0,x} \{x_0 : x_0 + c^T x = 0;\; Ax = b;\; x \geq 0_N\} \tag{198}$$

Definition: An x^0 that satisfies $Ax^0 = b$ and $x^0 \geq 0_N$ is called a *feasible solution* for its associated Primal.

Definition: A feasible solution with at least N-M components equal to 0 is called a *basic feasible solution.*

Theorem: If $Ax = b$, $x \geq 0_N$ has a feasible solution, then it has a basic feasible solution.

Proof: Here we assume $M < N$ so A is singular. Assume a basic feasible solution x^* such that $Ax^* = b$ and $x^* > 0$. If x^* has has at most M non-zero components, we are done. Thus assume x^* has $k > M$ nonzero components. Without loss of generality, let the first k elements of x^* be non-zero. Now consider the matrix $A = (A_{\bullet 1}, A_{\bullet 2}, ..., A_{\bullet k})$ where we use the notation $A_{\bullet k}$ to denote the *kth column* of the matrix A. As we have assumed $k > M$, the matrix now defined by A is singular, thus there is a linear dependence relationship among its columns. Thus there exists a *non-zero* vector z such that $Az = 0_M$ As our first k elements of x are positive and atleast one element of z is positive, we can define the strictly positive number $\alpha = \max_j \{z_j / x_j^*\}$. Crucially note that this implies $\alpha x_j^* - z_j = 0$ for some $1 \leq j \leq k$ Thus,

$$\begin{aligned} b &= Ax^* \\ &= \sum_{k=1}^{K} A_{\bullet k} x_k^* \\ &= \sum_{k=1}^{K} A_{\bullet k} x_k^* - \frac{1}{\alpha} \sum_{k=1}^{K} A_{\bullet k} z_k \\ &= \frac{1}{\alpha} \sum_{k=1}^{K} A_{\bullet k} (\alpha x_k^* - z_k^*) \end{aligned}$$

Where above the second equality follows because all elements of x^* past k are 0, the third equality follows because we are subtracting something that is equal to 0 (see the linear dependence relationship above), and the fourth equality comes from multiplying and dividing the first term by α and then factoring. Thus, again noting that we have $\alpha x_j^* - z_j = 0$ for some $1 \leq j \leq k$, the last equality clearly shows we have constructed a feasible solution with at least 1 more zero component than our previous feasible solution x^*! As $k > M$ was arbitrary, we can continue this procedure until our x^* contains M zero components and we are done.

We are now in a position to state the simplex algorithm. If one has a feasible solution, then by the above theorem a basic feasible solution can be constructed to initiate the algorithm. Indeed, consider the following representation of (198):

$$\max_{x_0, x_1, \dots, x_N} \; \{x_0 : \; e_0 x_0 + \sum_{n=1}^{N} A_{\bullet n}^* x_n = b^*\} \tag{199}$$

where here we have defined $A_{\bullet n}^*$ by "stacking" the n-th element of c on top of A, that is

$$A_{\bullet n}^* = \begin{pmatrix} c_n \\ A_{\bullet n}^* \end{pmatrix} \qquad\qquad b^* = \begin{pmatrix} 0 \\ b \end{pmatrix} \qquad\qquad e_0 = \begin{pmatrix} 1 \\ 0_M \end{pmatrix}$$

so the first row of the constraint in (199) is just our objective function, and the next M rows are our usual constraints. Let x^0 be a basic feasible solution for

(199), whose elements 1 through M are non-zero and its next $N-M$ components are zero. This x^0 satisfies

$$e_0 x_0^0 + \sum_{m=1}^{M} A^*_{\bullet m} x_m^0 = b^*$$

where the sum above need only go up to M because all x_i past M are 0. For each non-zero element in x^0, we then take its associated index column from A^* to form the *basis matrix* :

$$B = \left(e_0, A^*_{\bullet 1}, A^*_{\bullet 2}, ..., A^*_{\bullet M}\right)$$

The above is an M+1 by M+1 matrix which for now we take to be nonsingular, so B^{-1} exists. We also note the terminology that will follow. I will say x_i is a *basic variable* if the associated i-th column of A^*, $A^*_{\bullet i} \in B$, and otherwise x_i is non-basic. Non-basic variables will be set to 0 shortly. With this definition of our basis matrix, we can thus write the constraint in (199) as

$$\left(B \mid A^*_{\bullet M+1}, ..., A^*_{\bullet N}\right) \begin{pmatrix} x_0 \\ \vdots \\ x_N \end{pmatrix} = b^* \tag{200}$$

now left multiply through by B^{-1} to get

$$B^{-1}\left(B \mid A^*_{\bullet M+1}, ..., A^*_{\bullet N}\right) \begin{pmatrix} x_0 \\ \vdots \\ x_N \end{pmatrix} = B^{-1} b^*$$

$$\left(I_{M+1} \mid B^{-1} A^*_{\bullet m+1}, ..., B^{-1} A^*_{\bullet N}\right) \begin{pmatrix} x_0 \\ \vdots \\ x_N \end{pmatrix} = B^{-1} b^*$$

which using our row column rule for multiplication we can write as:

$$e_0 x_0 + e_1 x_1 + ... + e_M x_M + B^{-1} A^*_{\bullet M+1} x_{M+1} + ... + B^{-1} A^*_{\bullet N} x_N = B^{-1} b^* \tag{201}$$

Setting all non-basic variables equal to 0, that is $x_{M+1}, ..., x_N = 0$, our initial basic feasible solution is then

$$x_m^0 = B^{-1}_{m\bullet} b^* \qquad\qquad m = 1, ..., M \tag{202}$$

We we now ask whether our x^0 is optimal. To answer this question we must examine equation (201). Suppose we try to increase a non-basic variable from 0, say x_{M+1}. Clearly, if $B^{-1}_{0\bullet} A^*_{\bullet M+1} < 0$ then we can increase x_{M+1} so $B^{-1}_{0\bullet} A^*_{\bullet M+1} x_{M+1} < 0$. To make sure we have still satisfied (201), we merely need to increase x_0 to offset this. Thus we have increased our objective so our x^0 was not optimal.

This is our optimality criterion for the simplex method:

Optimality Condition: If $B_{0\bullet}^{-1}A_{\bullet n}^{*} \geq 0$ for $n = M{+}1, ...N$ then the current basis matrix is optimal, and the solution to our linear program is given by (202).

Now suppose we have $B_{0\bullet}^{-1}A_{\bullet s}^{*} < 0$ for some s in our index of non-basic variables. Next suppose that $B_{m\bullet}^{-1}A_{\bullet s}^{*} \leq 0$ for $m = 1, ..., M$. The first inequality means we can increase x_s to increase our objective as described above, and the second inequality means we can increase $x_1, ...x_M$ to offset the corresponding decrease of elements $B_{m\bullet}^{-1}A_{\bullet s}^{*} \leq 0$ so we can continually satisfy equation (201) (stay feasible) while increasing x_s as much as we would like. Thus we have described the situation in which our optimal solution is unbounded. We will give an example of an unbounded LP shortly to make this more explicit.

Unbounded Solution: If $B_{0\bullet}^{-1}A_{\bullet s}^{*} < 0$ for some s in our index of non-basic variables and $B_{m\bullet}^{-1}A_{\bullet s}^{*} \leq 0$ for $m = 1, ..., M$ then we can increase x_s without bound while still satisfying (201), so our linear program has an unbounded solution.

Now, continue to suppose $B_{0\bullet}^{-1}A_{\bullet s}^{*} < 0$ but now let $B_{m\bullet}^{-1}A_{\bullet s}^{*} > 0$ for some $1 \leq m \leq M$. Thus we can increase our objective by increasing x_s, but to maintain feasibility, equation (201) shows us we must simultaneously decrease x_m^0. As $x_m^0 = B_{m\bullet}^{-1}b^{*} > 0$, to maintain feasibility, (not violate $x_m^0 \geq 0$ we must have

$$x_m^0 = B_{m\bullet}^{-1}b^{*} - (B_{m\bullet}^{-1}A_{\bullet s}^{*})x_s \geq 0 \tag{203}$$

$$\frac{B_{m\bullet}^{-1}b^{*}}{B_{m\bullet}^{-1}A_{\bullet s}^{*}} \geq x_s \tag{204}$$

Thus equation (204) says that the maximum amount we can increase x_s and maintain feasibility is $\frac{B_{m\bullet}^{-1}b^{*}}{B_{m\bullet}^{-1}A_{\bullet s}^{*}}$. We then perform the same above calculation *for all indexes m* such that $B_{m\bullet}^{-1}A_{\bullet s}^{*} > 0$, and we take the amount we can increase x_s from 0 to be the minimum of all such ratios, which allows us to maintain feasibility, that is

$$x_s^{*} = \min_{m} \{\frac{B_{m\bullet}^{-1}b^{*}}{B_{m\bullet}^{-1}A_{\bullet s}^{*}} \ : B_{m\bullet}^{-1}A_{\bullet s}^{*} > 0\} \tag{205}$$

Lastly, we add column $A_{\bullet s}^{*}$ to the basis matrix and we drop a column r such $A_{\bullet r}^{*}$ achieves the minimum defined by equation (205). The above procedure defines the core of the simplex algorithm.

Note: When dropping a column $A_{\bullet r}^{*}$ from the basis matrix B, if there are multiple r that achieve the minimum defined by (205), choose the r with the *smallest index*. For example if $A_{\bullet 3}^{*}$ and $A_{\bullet 6}^{*}$ can both be dropped (they both achieve the minimum of (205)), we drop $A^*_{\bullet 3}$. The reason for doing so is otherwise the simplex algorithm may *cycle* (not terminate). This is referred to as *Bland's Rule*.

We now briefly show two example LPs, the first where the simplex algorithm gives us a finite optimal solution, and the second which has an unbouned solution.

Example 1: LP with a finite optimal solution.

$$\begin{aligned} &\max\ 5x_1 + 4x_2 + 3x_3 \ \ s.t. \\ &\begin{pmatrix} 2 & 3 & 1 \\ 4 & 1 & 2 \\ 3 & 4 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \leq \begin{pmatrix} 5 \\ 11 \\ 8 \end{pmatrix} \\ &x_1, x_2, x_3 \geq 0 \end{aligned}$$

We first add slack variables, x_4, x_5, x_6 so our constraint goes from $Ax \leq b$ to $Ax + I_3 x_s = b$ Re-writing our LP in the form of equation (199) we see:

$$\max\{x_0 : \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} x_0 + \begin{pmatrix} -5 \\ 2 \\ 4 \\ 3 \end{pmatrix} x_1 + \begin{pmatrix} -4 \\ 3 \\ 1 \\ 4 \end{pmatrix} x_2 + \begin{pmatrix} -3 \\ 1 \\ 2 \\ 2 \end{pmatrix} x_3 \quad + \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} x_4 + \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} x_5 + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} x_6$$

$$= \begin{pmatrix} 0 \\ 5 \\ 11 \\ 8 \end{pmatrix} \}$$

I will "guess" that our optimal basis matrix consists of e_0 and the first, third and fifth columns of the above, $B = (e_o, A^*_{\bullet 1}, A^*_{\bullet 3}, A^*_{\bullet 5})$, that is

$$B = \begin{pmatrix} 1 & -5 & -3 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 4 & 2 & 1 \\ 0 & 3 & 2 & 0 \end{pmatrix}$$

and a routine calculation shows that:

$$B^{-1} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 2 & 0 & -1 \\ 0 & -3 & 0 & 2 \\ 0 & -2 & 1 & 0 \end{pmatrix}$$

To see if this is indeed an optimal basis matrix, we must check that $B^{-1}_{0\bullet} A^*_{\bullet n} \geq 0$ for all columns n not in our optimal basis matrix ($n = 2, 4, 6$). Since $B^{-1}_{0\bullet} = (1, 1, 0, 1)$ so we have:

$$B^{-1}_{0\bullet} A^*_{\bullet 2} = 3$$

$$B_{0\bullet}^{-1}A_{\bullet 4}^{*} = 1$$
$$B_{0\bullet}^{-1}A_{\bullet 6}^{*} = 1$$

Thus our optimality conditions are satisfied. Our optimal solution is given by $x^* = B^{-1}b^*$ with all non-basic variables set to 0, so we calculate

$$x^* = B^{-1}b^* = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 2 & 0 & -1 \\ 0 & -3 & 0 & 2 \\ 0 & -2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 5 \\ 11 \\ 8 \end{pmatrix} = \begin{pmatrix} 13 \\ 2 \\ 1 \\ 1 \end{pmatrix}$$

so our optimal objective function value is $z = 13$ and our optimal decision variables are $x_1^* = 2$, $x_2^* = 0$ and $x_3^* = 1$

Example 2: LP with an unbounded solution.

$$\max \ -3x_1 + x_2 + 2x_3 \ \ s.t.$$
$$\begin{pmatrix} 1 & -1 & 0 \\ -2 & 0 & 1 \\ 0 & -2 & 1 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \le \begin{pmatrix} 0 \\ 1 \\ 2 \\ 6 \end{pmatrix}$$
$$x_1, x_2, x_3 \ge 0$$

We first add slack variables, x_4, x_5, x_6, x_7 so our constraint goes from $Ax \le b$ to $Ax + I_4 x_s = b$ Re-writing our LP in the form of equation (199) we see:

$$\max\{x_0 : \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} x_0 + \begin{pmatrix} 3 \\ 1 \\ -2 \\ 0 \\ 1 \end{pmatrix} x_1 + \begin{pmatrix} -1 \\ -1 \\ 0 \\ -2 \\ 1 \end{pmatrix} x_2 + \begin{pmatrix} -2 \\ 0 \\ 1 \\ 1 \\ -1 \end{pmatrix} x_3 \quad + \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} x_4 + \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} x_5 + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} x_6 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} x_7$$
$$= \begin{pmatrix} 0 \\ 0 \\ 1 \\ 2 \\ 6 \end{pmatrix} \}$$

I will "guess" that the optimal basis matrix consists of e_0 and the first, third, sixth and seventh columns of the above, that is $B = (e_o, A_{\bullet 1}^*, A_{\bullet 3}^*, A_{\bullet 6}^*, A_{\bullet 7}^*)$ Indeed we then have

$$B = \begin{pmatrix} 1 & 3 & -2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 & 1 \end{pmatrix}$$

and a routine calculation shows that

$$B^{-1} = \begin{pmatrix} 1 & 1 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & -2 & -1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$

We have $B^{-1}_{0\bullet} = (1,1,2,0,0)$ and we have optimality only if $B^{-1}_{0\bullet} A^*_{\bullet n} \geq 0$ for $n = 2,4,5$ (our non-basic variables). But

$$B^{-1}_{0\bullet} A^*_{\bullet 2} = (1,1,2,0,0) \begin{pmatrix} -1 \\ -1 \\ 0 \\ -2 \\ 1 \end{pmatrix} = -2 < 0$$

To see what our dropping variable will be we first calculate

$$\begin{aligned} B^{-1}_{1\bullet} A^*_{\bullet 2} &= -1 < 0 \\ B^{-1}_{2\bullet} A^*_{\bullet 2} &= -2 < 0 \\ B^{-1}_{3\bullet} A^*_{\bullet 2} &= 0 \leq 0 \\ B^{-1}_{5\bullet} A^*_{\bullet 2} &= 0 \leq 0 \end{aligned}$$

Thus we can increase x_2 without bound, and we have met our criterion for an unbounded solution described above. We calculate our optimal solution as

$$B^{-1} b^* = \begin{pmatrix} 1 & 1 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & -2 & -1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 2 \\ 6 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 1 \\ 1 \\ 7 \end{pmatrix}$$

Thus we parameterize and let $x_2 = t$. We then read off our optimal solution as $z^* = 2 + t$ for $t \in \mathbb{R}$, where t can be made as large as we please. With this parameterization from the above we see $x_1^* = t$, $x_2^* = t$, $x_3^* = 1 + 2t$ which fully characterizes the solution to Example 2.

In the above analysis the simplex algorithm was easy to initiate, but this may not always be the case. Indeed a linear program may actually be infeasible (meaning no basic feasible solution exists), or finding a basic feasible solution to initiate the algorithm may be complex (for example we often use the origin as our starting basic feasible solution, but in many problems this point may be infeasible). When this is the case, one must use the two-phase simplex method, which proceeds as follows. First introduce a vector of artificial variables s and use the simplex method on the problem

$$\min_{x,s} \; \{1_M^T s | \;\; Ax + I_M s = b; \;\; x \geq 0_N; \;\; s \geq 0_M\} \tag{206}$$

where here our initial basic feasible solution is letting $s = b \geq 0_M$ and $x = 0_N$. If the minimum for (206) is greater than 0 then the original LP is *infeasible*. If the minimum for (206) is 0 and no columns corresponding to the artificial variables are in the final basis matrix, then this x^0 is a basic feasible solution to the original LP and we can initiate the simplex method with it. If one or more columns corresponding to the artificial variables are in the final basis matrix, then we solve the LP

$$\max_{x_0, x \geq 0, s \geq 0} \{x_0 \mid x_0 + c^T x = 0;\ Ax + I_M s = b\ ;\ 1_M^T s = 0\} \tag{207}$$

which would be the same as the original LP, but the constraints have been re-written in terms of the artificial variables. Clearly $s \geq 0_M$ and $1_M^T s = 0$ imply all s_i are 0 for all iterations of the simplex algorithm.

A large part of the success of the simplex algorithm must be attributed to the fact that it is *efficient*. By this we mean it will run in polynomial time, where for a general LP we can expect to reach the optimal basis matrix in $\mathcal{O}(mn^2)$ iterations of the algorithm. To see what an improvement over naive methods this is, consider the problem of "guessing" the optimal basis matrix over and over. The number of ways we can choose the basis matrix is:

$$\begin{aligned}
\binom{n+m}{m} &= \frac{(n+m)!}{m!n!} && (208)\\
&= \frac{(n+m)(n+m-1)...(n+1)n!}{m!n!} && (209)\\
&= \frac{(n+m)(n+m-1)...(n+1)}{m!} && (210)\\
&\leq \frac{1}{\sqrt{2\pi m}}\left(\frac{e}{m}\right)^m (n+m)^m && (211)\\
&\leq \left(\frac{e}{m}\right)^m (n+m)^m && (212)\\
&= \left(\frac{e(n+m)}{m}\right)^m && (213)
\end{aligned}$$

where above the inequality in (211) follows from *Stirling's formula*, that is:

$$\begin{aligned}
\sqrt{2\pi m}\left(\frac{m}{e}\right)^m &\leq m! \\
\frac{1}{m!} &\leq \frac{1}{\sqrt{2\pi m}}\left(\frac{e}{m}\right)^m
\end{aligned}$$

and letting $m = n$ in (213) we see that

$$\binom{2n}{n} \leq e^n 2^n \leq 3^n 2^n$$

so this naive method can never be worse than $\mathcal{O}(6^n)$. It is clear that going from exponential time (combinatorially guessing) to polynomial time (simplex method) is a massive improvement. In a famous example, Klee and Minty (1972) showed that the simplex algorithm will sometimes run in exponential time, where they proposed the problem:

$$\begin{aligned}
\max \sum_{j=1}^{n} 10^{n-j} x_j \quad s.t. & \\
2\sum_{j=1}^{i-1} 10^{i-j} x_j + x_i \leq 100^{i-1} & \qquad\qquad i = 1, ..., n \\
x_j \geq 0 & \qquad\qquad j = 1, ..., n
\end{aligned}$$

on which the simplex will take $\mathcal{O}(2^n)$ iterations to terminate, that is the simplex algorithm will visit all 2^n vertices. Rare special cases aside, the simplex method will for all intents and purposes always run in polynomial time, making it practical for problems even where n and m are extremely large as they can be in operations research.

Given an LP, there is an associated *dual linear program*. These problems are usually referred to as the "primal problem" and "dual problem" respectively. We will see that they are "connected" in such a way that any feasible solution of one gives a bound on the objective value of the other. The dual problem is of particular importance in economics, as dual variables are often interpreted via the complementary slackness condition (shadow prices).

Definition: If the primal problem is:

$$\begin{aligned}
\max_x \; & c^T x \quad s.t. \\
& Ax \leq b \\
& x \geq 0_N
\end{aligned}$$

then its associated dual is:

$$\begin{aligned}
\min_y \; & b^T y \quad s.t. \\
& A^T y \geq c \\
& y \geq 0_M
\end{aligned}$$

The reader may note that by definition the dual of the dual is the primal, a fact frequently used in proofs.

Theorem (Weak Duality): If x is a feasible solution of the primal LP defined above and y is a feasible solution of its corresponding dual, then

$$z = \sum_{j=1}^{N} c_j x_j \leq \sum_{i=1}^{M} b_i y_i = w$$

Proof: Let x be a feasible solution for the primal. Then

$$\begin{aligned} c^T x &\leq (A^T y)^T x \\ &= y^T A x \\ &\leq y^T b \end{aligned}$$

where the first inequality follows from the inequality constraints in the dual, the second equality follows from the definition of the transpose, and the last inequality follows from the constraints of the primal.

An immediate consequence of this theorem is then that *any* dual feasible solution provides an upper bound on the objective of the primal. To see this just note that the optimal x^* is feasible, so $z^* = c^T x^* \leq b^T y$ for all dual feasible y. By the same logic, *any* primal feasible solution provides a *lower bound* on the objective of the dual. Weak duality gives us $z^* \leq w^*$. The next question is then whether $z^* = w^*$, which we will answer in the affirmative with the strong duality theorem. In the proof of this theorem, we will apply the following lemma of Farkas. One can prove Farkas Lemma with the strong duality theorem very easily, but here we will just use Farkas Lemma to prove strong duality.

Lemma (Farkas): Let A be an M by N matrix with $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^M$ Then exactly one of the following "alternatives" has a solution:

1. $Ax \leq b$ and $x \geq 0_N$
2. $y^T A \geq 0_N$ and $y^T b \leq 0$

Theorem (Strong Duality): If the primal problem has a finite optimal solution x^*, then the dual problem has a finite optimal solution y^*, and the optimal objective values of the primal and dual coincide, that is:

$$z^* = c^T x^* = b^T y^* = w^*$$

Proof: Let the primal problem have a finite optimal solution x^* such that

$$c^T x^* = \delta$$

Let $\epsilon > 0$ be arbitrary. Define

$$A' = \begin{pmatrix} A \\ -c^T \end{pmatrix} \qquad\qquad b' = \begin{pmatrix} b \\ -\delta - \epsilon \end{pmatrix}$$

that is A' is now an $M+1$ by N matrix where we have placed $-c^T$ as the bottom row and b' is $M+1$ by 1 where we have placed $-\delta - \epsilon$ as the last element. Now consider

$$A'x \le b'$$

the last row of this multiplication gives

$$\begin{aligned} -c^T x &\le -\delta - \epsilon \\ c^T x &\ge \delta + \epsilon \end{aligned}$$

But we assumed the maximum of $c^T x$ subject to $Ax \le b$ was δ. Thus the system $A'x \le b'$ has no solution for $x \ge 0$! Thus we can apply Farkas Lemma, which says there exist $y' = (y, \alpha)$ such that $y'^T A \ge 0$ and $y'^T b \le 0$. The first expression gives us

$$(y, \alpha)^T \begin{pmatrix} A \\ -c^T \end{pmatrix} \ge 0 \tag{214}$$

$$y^T A \ge \alpha c \tag{215}$$

and the second expression gives us

$$y'^T b' < 0 \tag{216}$$

$$(y, \alpha)^T \begin{pmatrix} b \\ -\delta - \epsilon \end{pmatrix} < 0 \tag{217}$$

$$y^T b - \alpha\delta - \alpha\epsilon < 0 \tag{218}$$

$$y^T b < \alpha(\delta + \epsilon) \tag{219}$$

We note that if $\epsilon = 0$, then $A'x \le b'$ *would* have a solution and we would be in the other alternative of Farkas. But $\epsilon > 0$ so we are in the second alternative and this implies $\alpha > 0$. The fact that $\alpha > 0$ allows us to scale y' so that $\alpha = 1$. Setting $\alpha = 1$ in (215) we get

$$y^T A \ge c \tag{220}$$

Thus y is dual feasible! As stated above, every dual feasible solution gives an upper bound on the objective of the primal (Weak Duality):

$$\delta \le y^T b \tag{221}$$

But now examine (219) with $\alpha = 1$. This states:

$$y^T b < \delta + \epsilon \tag{222}$$

Combining (219) and (222) we see that

$$\delta \le y^T b < \delta + \epsilon \tag{223}$$

and taking $\lim \epsilon \to 0$ in (223) we

$$\delta \leq y^T b \leq \delta$$

So by the squeeze theorem $\delta = y^T b$ and thus our optimal objective value of the primal and dual coincide, so we have proven the strong duality theorem.

It is important to note that the strong duality theorem will only hold if a *finite* optimal solution exists to either the primal or dual.

Corollary 1: If the primal problem is unbounded, then the dual is infeasible.

This follows from the fact that every dual feasible solution provides an *upper bound* on the primal. But if there is no upper bound on the primal, there can be no dual feasible solutions.

Corollary 2: If the dual problem is unbounded, then the primal problem is infeasible.

This follows from the fact that every primal feasible solution provides a lower bound on the dual. But if the dual is unbounded from below, then there can be no primal feasible solution.

It also turns out that if we have solved the primal by the simplex method, then we have our optimal dual variables at our finger tips!

Fact: Let the primal LP have a finite optimal solution x^*, with associated optimal basis matrix B. Then the first row of B^{-1} is necessarily of the form:

$$B^{-1}_{0\bullet} = (1, y^{*T})$$

The dual variables as defined above are rich in economic significance. It can be seen that y^*_m represents the marginal increase in the primal objective function due to a small increase in the m-th component of b, b_m. Thus y^*_m represents the *shadow price* of the m-th resource constraint, or how much we would be willing to pay to relax it a little. One may note this is the exact same as the interpretation of the Lagrange multiplier in classical optimization theory, where λ represents the marginal increase in the objective when we marginally relax the constraint! To aid our understanding of dual variables as shadow prices, we will prove the following complementary slackness for linear programs result which is of great importance in economics. It says that if a optimal dual variable is 0, its corresponding resource constraint must not be binding, and if a resource constraint is binding then its associated dual variable must be non-zero.

Theorem (Complementary Slackness): If an optimal dual variable $y^*_m > 0$, then $\sum_{i=1}^N a_{mi} x^*_i = b_m$. If $y^*_m = 0$, then $\sum_{i=1}^N a_{mi} x^*_i < b_m$.

Proof: Let x^* be a finite optimal solution for the primal. Recall our proof of the weak-duality theorem where by substituting in constraints from the primal and dual we showed:

$$c^T x^* \leq (A^T y^*)^T x^* = y^{*T} A x^* \leq y^{*T} b \tag{224}$$

But x^* is a finite optimal solution so the strong duality theorem gives us $c^T x^* = y^{*T} b$, so we can replace all less than or equals to in (224) with just equalities, where the right hand side gives us

$$y^{*T} A x^* = y^{*T} b \tag{225}$$

$$y^{*T}(Ax^* - b) = 0 \tag{226}$$

which using scalar notation can be expressed as

$$\sum_{i=1}^{M} \left(y_i^* \left(\sum_{j=1}^{N} a_{ij} x_j^* - b_i \right) \right) = 0 \tag{227}$$

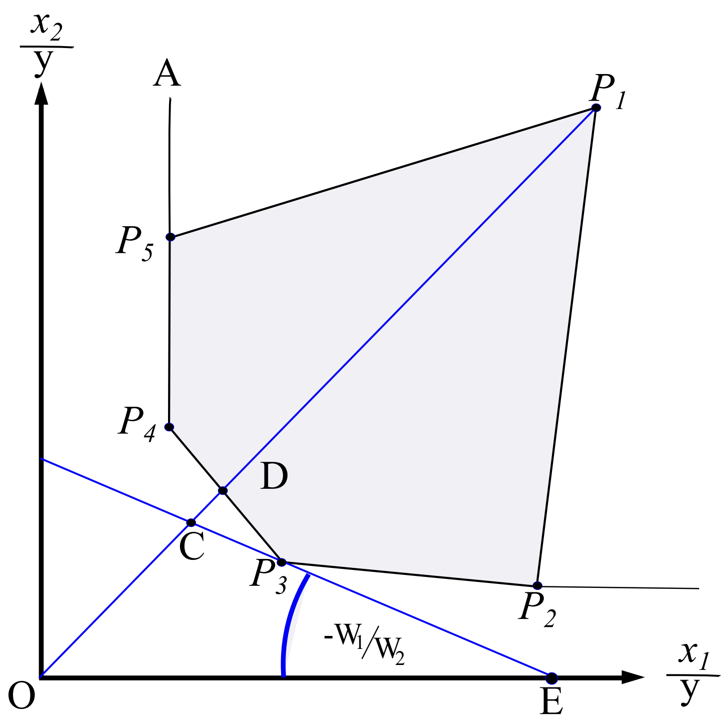
Now, every y_i^* is positive (because y^* is dual feasible), and every element of the sum of $\sum_{j=1}^{N}(a_{ij}x_j^* - b_i)$ is necessarily negative (since $Ax^* \leq b$), thus (227) represents a sum of all *non-positive* numbers that sum to 0. This can only be the case if every term in the sum (227) is therefore 0. Looking at the m-th term and noting that by the proceeding logic it must be equal to 0, we see:

$$y_m^* \left(\sum_{j=1}^{N} a_{mj} x_j^* - b_m \right) = 0 \tag{228}$$

Thus if $y_m^* > 0$ then we necessarily have $\left(\sum_{j=1}^{N} a_{mj} x_j^* - b_m \right) = 0$ and if $\left(\sum_{j=1}^{N} a_{mj} x_j^* - b_m \right) < 0$ then we necessarily have $y_m^* = 0$, which are precisely the two statements we set out to show, so we are done. This concludes our brief presentation of Linear Programming.

5.2 Farrell's methodology

We begin by briefly discussing the theory of Farrell (1957) and then its implementation in our empirical work. Consider the simplified case of K firms producing 1 output from 2 inputs. For firm K, let the output be y_1^k and let the inputs be x_1^k and x_2^k. For each firm we can calculate input-output co-efficients as x_1^k/y_1^k and x_2^k/y_1^k. Noting that these co-efficients can form tuples $(x_1^k/y_1^k, x_2^k/y_1^k)$ in the input output space, they can then be plotted as in the following figure, where we let $K = 5$ for concreteness. P_1 represents input-output co-efficients of firm 1, P_2 for firm 2, and so on. The interior of the set AP_4P_3B forms the *convex free disposal hull* for this observed production technology, and we take the boundary of AP_4P_3B itself to be the production possibilities frontier. We define the Farrell measure of technical efficiency of firm 1 as the ratio of the distances OD/OP_1, for this measures the fraction of inputs an efficient firm (on the PPP) would need to produce the same amount of output as firm 1. In this setup we note that firms 2,3,4,5 are all regarded as being efficient. This definition is clearly quite coarse as we have firm 4, P_4 is producing the same as P_5 while using strictly less of x_2, and firm 2 is producing the same as P_3 while using strictly less x_1. Ideally we would like firm 4 to be regarded as more technically efficient than firm 5, and firm 2 to be regarded as more technically efficient than firm 3.



To acknowledge problems of allocative (price) efficiency as well as technical efficiency, now suppose all firms face input price costs of w_1 and w_2. The cost of producing y is $x_1w_1 + x_2w_2$ and so our isocost line has slope $-w_1/w_2$. We can take the family of all such isocost lines, and find the smallest (lowest) one, that just intersects our PPP. From our diagram, we see that this occurs at P_3, which is the only firm that is truly cost minimizing. We can now define a measure of allocative efficiency for each firm. Suppose P_1 became *technically efficient* as previously described, so it is now at point D on the PPP. At this point, it is still not cost minimizing. Indeed, its relative distance to the isocost line is the ratio OC/OD, which we define as its *price efficiency.* Thus, for any firm, we can scale its output to make it technically efficient (put it on the PPP) and then calculate how far it still is away from the isocost line to get a measure of pure price inefficiency. We can then calculate a measure of overall efficiency as the distance from input-output coefficients to the corresponding point on the isocost line. For firm 1 this overall measure of efficiency is OC/P_1. One should note that once we begin to consider prices, we find that firm 1 is even more inefficient than when we considered just quantities. Indeed, technical efficiency will always be strictly greater than overall efficiency in this setup. Also note that now firm 3, P_3, is the only overall efficient producing firm ($e = 1$), even though other firms are on the PPP as well.

We will now generalize this theory in various ways. The above model had two inputs into production as our data does (capital and labour), but only a single output, and we would like to have multiple outputs. Indeed we will aggregate $C + I + G$ to *domestic output* using a chained Fisher Ideal index, which we will refer to as y_1, and we will also treat exports and imports as outputs y_2 and y_3 respectively. In all analysis that follows always assume that domestic output and exports (y_1, y_2) have the same sign, and that imports, capital and labour have the opposite sign of these (y_3, y_4, y_5). Next, instead of making observations across firms, we will make observations across *years*, so we go from $k = 1, ...K$ in the above to $t = 1, ...25$, thus carrying out our empirical work for years 1993-2017 of our sample period. We initially assume that in each year we have access to the same basic technology except for efficiency differences. An approximation to the basic technology is defined to be the convex free disposal hull of the observed quantity data

$$Conv\Bigg(\{(y_1^t, y_2^t, y_3^t, x_1^t, x_2^t)\ :\ t = 1, ..., 25\}\Bigg)$$

where we adopt the notation $Conv(S)$ is the convex hull of a set S. This technology assumption is consistent with decreasing returns to scale (and constant returns to scale) but it is not consistent with increasing returns to scale. We will measure the efficiency of year i by the smallest positive fraction δ_i^* of the year i input vector (x_1^i, x_2^i) such that $(y_1^i, y_2^i, y_3^i, \delta_i^* x_1^i, \delta_i^* x_2^i)$ is on the efficient frontier spanned by the convex free disposal hull of the 25 observations defined above. If the ith year is efficient relative to this frontier, then $\delta_i^* = 1$, the smaller δ_i^* is the lower the efficiency of this year. Thus we would like to solve the following linear program for years i=1,...25

$$\begin{aligned}\delta_i^* = \min_{\delta_i \geq 0, \lambda_1 \geq 0, \ldots, \lambda_{25} \geq 0} \Bigg\{\delta_i \mid \sum_{t=1}^{25} y_1^t \lambda_t \geq y_1^i\ ;\ \sum_{t=1}^{25} y_2^t \lambda_t \geq y_2^i\ ; \sum_{t=1}^{25} y_3^t \lambda_t \geq y_3^i \quad (229)\\ \sum_{t=1}^{25} x_1^t \lambda_t \leq \delta_i x_1^i\ \sum_{t=1}^{25} x_2^t \lambda_t \leq \delta_i x_2^i\ ;\ \sum_{t=1}^{25} \lambda_t = 1\Bigg\}\end{aligned}$$

To elucidate the structure of the problem we will rewrite it in matrix form which shows exactly how it was solved in our empirical work. Let $t \in \{1, ...25\}$.

Construct the A matrix as

$$A_t \equiv \begin{pmatrix} y_1^1 & y_1^2 & \dots & y_1^{25} & 0 \\ y_2^1 & y_2^2 & \dots & y_2^{25} & 0 \\ y_3^1 & y_3^2 & \dots & y_3^{25} & 0 \\ x_1^1 & x_1^2 & \dots & x_1^{25} & -x_1^t \\ x_2^1 & x_2^2 & \dots & x_2^{25} & -x_2^t \\ 1 & 1 & \dots & 1 & 0 \\ -1 & -1 & \dots & -1 & 0 \end{pmatrix} \tag{230}$$

The reader should note this is a 7x26 matrix, and we have one such for each year, giving us 25 such A matrices. Next define

$$b_t \equiv \begin{pmatrix} y_1^t \\ y_2^t \\ y_3^t \\ 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} \tag{231}$$

$$c^T = \begin{pmatrix} 0 & 0 & \dots & 0 & -1 \end{pmatrix} \tag{232}$$

$$\lambda_t = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{25} \\ \delta_t \end{pmatrix} \tag{233}$$

where we note that b_t is 7x1, c^T is 1x26 and λ_t is 26x1. Thus for each year t to get our measure of efficiency δ_t^* we solve the LP:

$$\max_{\lambda_t \geq 0_{26}} c^T \lambda_t \ \ s.t. \tag{234}$$

$$A_t \lambda_t \leq b_t$$

There are two final things to note here. The first is that the two rows of $1's$ and $-1's$ at the bottom of A_t may look strange. Indeed when one does the multiplication $A_t \lambda_t$ it is seen that these combined with our constraint yield

$$\sum_{t=1}^{25} \lambda_t \leq 1$$

and

$$\sum_{t=1}^{25} \lambda_t \geq 1$$

but together these imply

$$\sum_{t=1}^{25} \lambda_t = 1$$

which is precisely what we want. This is a common method of turning equality constraints into inequality constraints so linear programs can be written in "standard form". The other is that to bring our LP into standard form, we also rewrote our minimization problem as a maximization problem. This is dealt with by simply taking the negative of our optimized objective function after we have solved the problem.

We now continue to maintain that the underlying technology is convex, but also assume that it is subject to constant returns to scale. The approximation to the underlying technology set is the gree disposal hull of the convex cone spanned by our 25 data points. The efficiency of observation i is measured by the positive fraction δ_i^{**} of the year i input vector (x_1^i, x_2^i) such that $(y_1^i, y_2^i, y_3^i, \delta_i^* x_1^i, \delta_i^* x_2^i)$ is on the efficient frontier spanned by the convex free disposal hull of the 25 observations defined above. The efficiency of the ith year relative to this technology set can be calculated by the following Linear Program:

$$\delta_i^{**} = \min_{\delta_i \geq 0, \lambda_1 \geq 0, \ldots, \lambda_{25} \geq 0} \left\{ \delta_i \mid \sum_{t=1}^{25} y_1^t \lambda_t \geq y_1^i \; ; \; \sum_{t=1}^{25} y_2^t \lambda_t \geq y_2^i \; ; \; \sum_{t=1}^{25} y_3^t \lambda_t \geq y_3^i \right. \tag{235}$$

$$\left. \sum_{t=1}^{25} x_1^t \lambda_t \leq \delta_i x_1^i \sum_{t=1}^{25} x_2^t \lambda_t \leq \delta_i x_2^i \right\}$$

In matrix form we can define

$$A_t \equiv \begin{pmatrix} y_1^1 & y_1^2 & \ldots & y_1^{25} & 0 \\ y_2^1 & y_2^2 & \ldots & y_2^{25} & 0 \\ y_3^1 & y_3^2 & \ldots & y_3^{25} & 0 \\ x_1^1 & x_1^2 & \ldots & x_1^{25} & -x_1^t \\ x_2^1 & x_2^2 & \ldots & x_2^{25} & -x_2^t \end{pmatrix} \tag{236}$$

which we note is now a 5x26 matrix as we have eliminated the last two rows because we no longer have the equality constraint $\sum \lambda_t = 1$ entering our LP. We also define

$$b_t \equiv \begin{pmatrix} y_1^t \\ y_2^t \\ y_3^t \\ 0 \\ 0 \end{pmatrix} \tag{237}$$

$$c^T = \begin{pmatrix} 0 & 0 & \ldots & 0 & -1 \end{pmatrix} \tag{238}$$

$$\lambda_t = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{25} \\ \delta_t \end{pmatrix} \tag{239}$$

where we note that b_t is 5x1, c^T is 1x26 and λ_t is 26x1. Thus we can compactly write out the linear program defined by (235) as:

$$\max_{\lambda_t \geq 0_{26}} c^T \lambda_t \;\; s.t. \qquad (240)$$
$$A_t \lambda_t \leq b_t$$

It should be noted that the linear programs defined by (229) and (235) are the same except the latter has had the constraint $\sum \lambda_t$ removed. The implication of this is then that

$$\delta_i^* \geq \delta_i^{**} \qquad\qquad i = 1, ..., 25 \qquad (241)$$

which follows because *any* feasible solution for (229) (including our optimal solution), is also feasible for (235), so the set of solutions to (229) is a subset of the solutions to (235), and since these are minimization problems this implies the inequality in (241).

In the results section it will be seen that by only considering quantity data, we are regarding too many years as efficient. To deal with this, we now incorporate measures of *allocative efficiency* into our overall measure of efficiency. We first consider cost minimizing behaviour. We assume that in year t the input prices for labour and capital services were (w_1^t, w_2^t). To determine whether the economy was cost minimizing subject to our convex technology assumption in the i-th year, we solve the linear program

$$\min_{\lambda \geq 0_{25}} \Big\{ w_1^i (\sum_{t=1}^{25} x_1^t \lambda_t) + w_2^i (\sum_{t=1}^{25} x_1^t \lambda_t) \quad | \sum_{t=1}^{25} y_1^t \lambda_t \geq y_1^i \,;\, \sum_{t=1}^{25} y_2^t \lambda_t \geq y_2^i \qquad (242)$$
$$\sum_{t=1}^{25} y_3^t \lambda_t \geq y_3^i \,;\, \sum_{t=1}^{25} \lambda_t = 1 \Big\}$$
$$= \epsilon_i^* (w_1^i x_1^i + w_2^i x_2^i) \qquad (243)$$

Thus we define the overall efficiency measure ϵ_i^* for year i by equating (243) with the value of the optimized objective function above. One should note that setting $\lambda_i^* = 1$ and $\lambda_j^* = 0$ for all $j \neq i$ we obtain a feasible solution for the LP above. The value of the objective is then simply $w_1^i x_1^i + w_2 x_2^i$ so equating this with (243) see we must have $\epsilon_i^* \leq 1$, because we have found a feasible solution that yields $\epsilon_i^* = 1$, and as this a minimization problem there may be other feasible solutions that yield a lower value of the objective. Also by our non-negativity constraint we get that $0 < e_i^*$, combining these two inequalities we note

$$0 < e_i^* \leq 1$$

We can re-write the linear program defined by (242) in matrix form. We let

$$A = \begin{pmatrix} y_1^1 & y_1^2 & \dots & y_1^{25} \\ y_2^1 & y_2^2 & \dots & y_2^{25} \\ y_3^1 & y_3^2 & \dots & y_3^{25} \\ 1 & 1 & ... & 1 \\ -1 & -1 & ... & -1 \end{pmatrix} \tag{244}$$

where we note A is now a 5x25 matrix and will remain constant in all 25 linear programs that we end up running. Also define

$$b_t = \begin{pmatrix} y_1^t \\ y_2^t \\ y_3^t \\ 1 \\ -1 \end{pmatrix} \tag{245}$$

$$c^T = \begin{pmatrix} \frac{-(w_1^1 X1 + w_2^1 X2)}{w_1^T X1 + w_2^T X2} & \frac{-(w_1^2 X1 + w_2^2 X2)}{w_1^T X1 + w_2^X 2} & \dots & \frac{-(w_1^{25} X1 + w_2^{25} X2)}{w_1^T X1 + w_2^T X2} \end{pmatrix} \tag{246}$$

where above we let $X1^T = (x_1^1, .., x_1^{25})$ and $X2^T = (x_2^1, .., x_2^{25})$ Thus we can write our linear programs as (1 for each year):

$$\max_{\lambda_t \geq 0_{25}} c^T \lambda_t \;\; s.t. \tag{247}$$

$$A\lambda_t \leq b_t$$

and our efficiency measure in year t is the *negative* of the optimized objective in that year.

Again examining (243) we see that e_i^* is the fraction of (x_1^i, x_2^i) such that $\epsilon_i^*(x_1^i, x_2^i)$ is on the minimum isocost line for observation i. We also note that any solution which is feasible for (229), our first measure of efficiency using just convex technology, is also feasible for (242) (convex technology with cost minimization). Thus we get the bound

$$0 < e_i^* \leq \delta_i^* \tag{248}$$

This is a very nice inequality as it formalizes some of the theory discussed earlier. δ_i^* is a measure of just technical efficiency, whereas e_i^* is a measure of overall efficiency which incorporates technical efficiency *as well* as allocative efficiency, so we see that when we begin to incorporate allocative efficiency, we get more observations overall being inefficient, hence the inequality in (248)

The next assumption we make is cost minimization with constant returns to scale imposed on our convex technology set. To do so, all we must do is drop the constraint $\sum \lambda_t = 1$ from our previous LP defined by (242):

$$\min_{\lambda \geq 0_{25}} \left\{ w_1^i(\sum_{t=1}^{25} x_1^t \lambda_t) + w_2^i(\sum_{t=1}^{25} x_1^t \lambda_t) \mid \sum_{t=1}^{25} y_1^t \lambda_t \geq y_1^i \; ; \; \sum_{t=1}^{25} y_2^t \lambda_t \geq y_2^i \right. \tag{249}$$

$$\left. \sum_{t=1}^{25} y_3^t \lambda_t \geq y_3^i \right\}$$

$$= \epsilon_i^{**}(w_1^i x_1^i + w_2^i x_2^i) \tag{250}$$

Where to solve for e_i^{**} we simply equate the value of the optimized objective in (249) with (250). Again, one should note that for the i-th year, $\lambda_i^* = 1$ and $\lambda_j^* = 0$ for $i \neq j$ is feasible, and solving our objective with this in it gives $e_i^{**} = 1$ The non-negativity constraint implies $e_i^{**} \geq 0$ so combining these two inequalities we must have

$$0 < \epsilon_i^{**} \leq 1$$

We can express (249) in matrix form very easily to see how the problem was solved in our empirical work. Indeed Let

$$A = \begin{pmatrix} y_1^1 & y_1^2 & \dots & y_1^{25} \\ y_2^1 & y_2^2 & \dots & y_2^{25} \\ y_3^1 & y_3^2 & \dots & y_3^{25} \end{pmatrix} \tag{251}$$

Where note A is now a 3x25 matrix and will be the same for all 25 years. Also define

$$b_t = \begin{pmatrix} y_1^t \\ y_2^t \\ y_3^t \end{pmatrix} \tag{252}$$

$$c^T = \begin{pmatrix} \frac{-(w_1^1 X1 + w_2^1 X2)}{w_1^T X1 + w_2^T X2} & \frac{-(w_1^2 X1 + w_2^2 X2)}{w_1^T X1 + w_2^X 2} & \dots & \frac{-(w_1^{25} X1 + w_2^{25} X2)}{w_1^T X1 + w_2^T X2} \end{pmatrix} \tag{253}$$

where above we let $X1^T = (x_1^1, .., x_1^{25})$ and $X2^T = (x_2^1, .., x_2^{25})$ Thus we can write our linear programs as (1 for each year):

$$\max_{\lambda_t \geq 0_{25}} c^T \lambda_t \;\; s.t. \tag{254}$$

$$A\lambda_t \leq b_t$$

and our efficiency measure in year t is the *negative* of the optimized objective in that year.

We also note that the constraint set of (249) is that same as that of (242) but we have dropped the $\sum \lambda_t = 1$ equality constraint. Therefore any solution which is feasible for (242) is also feasible for (249). Noting that these are minimization problems, this gives us the inequality

$$e_i^{**} \leq \epsilon_i^* \tag{255}$$

By the same logic, in comparing (249) to (235) we get that

$$\epsilon_i^{**} \leq \delta_i^{**} \tag{256}$$

The inequality in (255) shows that making stronger assumptions on the underlying technology will *decrease* the efficiency measure, that is the constant returns to scale measure of efficiency for year i, ϵ_i^{**}, will be equal or less than that of the convex technology measure of efficiency of year i, ϵ_i^*. Inequality (256) is the analog of (248). It shows that once we begin to incorporate allocative efficiency, our overalll level of efficiency will be less than that of if we were just considering technical efficiency (only quantities), with the only difference from (248) being that now we are assuming constant returns to scale in both case.

Finally we consider our most stringent assumption, *profit maximization*. If there are positive output prices in year i, (p_1^i, p_2^i, p_3^i), to determine whether in year i the economy was maximizing profits subject to our convex technology assumptions, we solve the following linear program:

$$\max_{\lambda_t \geq 0_{25}} \left\{ \sum_{m=1}^{3} p_m^i (\sum_{t=1}^{25} y_m^t \lambda_t) - \sum_{n=1}^{2} w_n^i (\sum_{t=1}^{25} x_n^t \lambda_t) \mid \sum_{t=1}^{25} \lambda_t = 1 \right\} \tag{257}$$

$$= p_1^i y_1^i + p_2^i y_2^i - \alpha_i^* (w_1^i x_1^i + w_2^i x_2^i) \tag{258}$$

Equating (257) and (258) defines our efficiency measure α_i^* for year i. As before, if we set $\lambda_i = 1$ and all other $\lambda_j = 0$ for that *year*, we obtain a feasible solution for the objective function equal to $p_1^i y_1^i + p_2^i y_2^i - (w_1^i x_1^i + w_2^i x_2^i)$. Equating this with (258) gives us $\alpha_i^* = 1$. Thus as this is a minimization problem we see that $\alpha_i^* \leq 1$. It is also clear that any feasible solution for our cost minimzation LP defined by (242) is also feasible for for our profit maximization problem defined by (257), so we get the bound

$$\alpha_i^* \leq \epsilon_i^* \tag{259}$$

which states that profit maximization on a convex technology set is a more stringent test of efficiency than cost minimization on the same convex technology set.

Lastly we would like to impose constant returns to scale on the problem defined by (257). The problem with this is that when we drop the constraint $\sum \lambda_t = 1$, the problem became unbounded. To make the problem tractable, we now consider *conditional profit maximization* instead of full profit maximization. To this end, we will assume the level of our second input to production (capital) is fixed in the short run, which is an easily justified assumption. Thus, to determine whether in year i we are maximizing variable profits subject to our convex, conical technology assumption, we solve the LP:

$$\max_{\lambda_t \geq 0_{25}} \left\{ \sum_{m=1}^{3} p_m^i (\sum_{t=1}^{25} y_m^t \lambda_t) - \sum_{n=1}^{2} w_n^i (\sum_{t=1}^{25} x_n^t \lambda_t) \mid \sum_{t=1}^{25} x_2^t \lambda_t \leq x_2^i \right\} \tag{260}$$

$$= \max_t \left\{ \left(\sum_{m=1}^{3} p_m^i y_m^t - \sum_{n=1}^{2} w_n^i x_n^t \right) (\frac{x_2^i}{x_2^t}) \mid t = 1, ..., 25 \right\} \tag{261}$$

$$= p_1^i y_1^i + p_2^i y_2^i - \alpha_i^{**}(w_1^i x_1^i + w_2^i x_2^i) \tag{262}$$

where our efficiency observation α_i^{**} can be gotten from equating (261) with (262). Noting $\lambda_i = 1$ and all other $\lambda_j = 0$ is feasible for the problem (260), this yields $\alpha_i^{**} \leq 1$. In comparing (235) with (260) (recall the former is our convex technology constant returns to scale measure of *technical efficiency*), we see that any feasible solution for (235) is also feasible for (260), and as these are minimization problems this gives us the bound:

$$\delta_i^{**} \geq \alpha_i^{**} \tag{263}$$

which again states that once we begin to introduce allocative efficiency (profit maximization) we get a lower overall measure of efficiency than when we considered just technical efficiency (with constant returns to scale in both cases). Unfortunately, since (260) does not entirely drop the constraint $\sum \lambda_t = 1$, we cannot get an immediate inequality between (260) and (257), but it will be seen in our empirical work that ϵ_i^* and ϵ_i^{**} are "close" and move strongly together.

To summarize our work and remind the reader exactly what has been done so far we present the following table and tabulate our previously derived inequalities.

	Nonpar1	Nonpar2	Nonpar3	Nonpar4	Nonpar5	Nonpar6
Technology	Convex	Convex CRS	Convex	Convex CRS	Convex	Convex CRS
Optimizing			Cost Min.	Cost Min.	Profit Max	V. Profit Max
Efficiency	δ_i^*	δ_i^{**}	ϵ_i^*	ϵ_i^{**}	α_i^*	α_i^{**}

$$\delta_i^* \geq \epsilon_i^* \geq \alpha_i^* \tag{264}$$

$$\delta_i^{**} \geq \epsilon_i^{**} \tag{265}$$

$$\delta_i^* \geq \delta_i^{**} \tag{266}$$

$$\epsilon_i^* \geq \epsilon_i^{**} \tag{267}$$

$$0 < \delta_i^*, \delta_i^{**}, \epsilon_i^*, \epsilon_i^{**}, \alpha_i^*, \alpha_i^{**} \leq 1 \tag{268}$$

We are now in a position to present our findings in the following table. Indeed, the inequalities from equations (264)-(267) are all easily verified by noting that entries of the fifth column are less than the third column are less than the first column. Entries in the fourth column are less than corresponding entries in the second column. Entries in the second column are less than corresponding entries in the first column. And finally entries in the fourth column are less than corresponding entries in the third column.

It is seen from Nonpar1 and Nonpar2 that when we only consider technical efficiency, we classify far too many years as efficient. Indeed Nonpar1 only picks up 6 observations as inefficient and nonpar2 classifies 9 observations as inefficient. Indeed, the method does show promise though as it immediately classifies the two major recessions of this sample period, 2001-02 and 2007-08 as the USA not being on the production possibilities frontier. Interestingly both Nonpar1 and Nonpar2 classify 2016 as inefficient, which is seen in GDP statistics where real GDP growth fell from 2.88% in 2015 to 1.56% in 2016. Lastly Nonpar2 classifies the stagnant recovery of 2010-2012 as inefficient, which is good as unemployment was very high during this period and the productive capacity of the economy was clearly underutilized.

When we begin to consider allocative efficiency as well, such as in Nonpar3 and Nonpar4 where we assume cost minimizing behaviour, we begin to pick up many more observations as inefficient. By the time we reach Nonpar5 and Nonpar6 all observations are regarded as inefficient except 2017. From Nonpar5 and Nonpar6 it can be seen that there is a trend upwards in efficiency over the sample period.

	NONPAR1	NONPAR2	NONPAR3	NONPAR4	NONPAR5	NONPAR6	PROD Normalized
1993	1	1	1	1	0.716	0.848	.8203
1994	1	1	0.9979	0.9979	0.7485	0.8622	.8307
1995	1	1	1	1	0.7631	0.8665	.8358
1996	1	1	1	1	0.7945	0.881	.8511
1997	1	1	1	1	0.8209	0.8931	.8654
1998	1	0.9998	0.9994	0.9919	0.843	0.9029	.8826
1999	1	1	1	0.9964	0.8758	0.9203	.9052
2000	1	1	1	0.975	0.8903	0.9267	.9130
2001	0.9945	0.9942	0.9945	0.9901	0.888	0.9221	.9139
2002	0.9959	0.9939	0.9956	0.9924	0.9009	0.9289	.9219
2003	1	1	1	1	0.9207	0.9414	.9348
2004	1	1	1	0.9988	0.9445	0.9577	.9517
2005	1	1	1	1	0.9588	0.9675	.9612
2006	1	1	0.9959	0.9931	0.9582	0.9659	.9604
2007	0.9933	0.9931	0.9889	0.9811	0.9544	0.9615	.9577
2008	0.9816	0.9747	0.9801	0.9722	0.9431	0.9505	.9486
2009	1	1	1	1	0.9334	0.9408	.9435
2010	1	0.994	1	0.9938	0.9608	0.965	.9652
2011	1	0.9867	1	0.9863	0.967	0.9702	.9698
2012	0.9996	0.9928	0.9995	0.9923	0.9737	0.976	.9753
2013	1	1	1	1	0.9788	0.9804	.9797
2014	1	1	1	1	0.9863	0.9871	.9865
2015	1	1	1	1	0.996	0.9961	.9960
2016	0.9985	0.998	0.998	0.9976	0.992	0.9921	.9921
2017	1	1	1	1	1	1	1
Inefficient	6	9	9	14	24	24	

As Nonpar6 shows an upwards trend in efficiency over the sample period, it is then natural to ask how this observed increase in efficiency of the US economy compares to our productivity estimates of earlier. Indeed, using the same data from 1993-2017 we can aggregate $C+I+G$ to $Y1$ using a chained Fisher, and then we can get an output aggregate Y, of domestic output, exports and imports again using a chained Fisher. We also form an input aggregate X using a chained Fisher over capital and labour and then we take our conventional index number estimate of productivity to be simply Y/X. To make this productivity series *comparable* with our efficiency metric we then normalize it by dividing all observations through by our highest (2017) TFP observation. Thus we now have a productivity series over the sample period 1993-2017 and it is normalized so our observation for TFP in 2017 is 1. We can then run the simple linear regression

$$\epsilon_t^{**} = \beta(Prod)_t + z_t \qquad\qquad t = 1, ..., 25$$

and it is seen that $\beta = 1.0019$ with an $R^2 = 0.9935$

	N	Mean	St. Dev	Variance	Min	Max	Corr. Coeff
ϵ_i^{**}	25	0.94015	4.40E-02	1.94E-03	0.84803	1	0.99674
Prod	25	0.9304	5.48E-02	3.00E-03	0.82037	1	0.99674

Above is the descriptive statistics for ϵ_i^{**} and our normalized productivity series. We note that over the 25 year sample period the mean of these two series differs by only .01, and they have a correlation co-efficient of .99674! This is very re-assuring as these two somewhat disparate methods of measuring productivity or efficiency give us almost exactly the same answer on average. If we are approaching the problem from two entirely different avenues and arriving at the same answer it means we are doing something fundamentally correct from a scientific standpoint. This should also be seen as a vindication of Farrell's theory and that it is indeed a robust and scalable way of measuring efficiency or productivity, not just across firms but across years of an entire economy. Although it is more complex and has more built in assumptions, this non-parametric method also provides valuable macroeconomic insight. By successively building more and more stringent conditions on technology sets and optimizing behavior into the problem, we witnessed the importance of allocative efficiency. Indeed, if we are to take ϵ_i^{**} to be a true measure of productivity, it shows us that prices (moving towards the isoprofit line) is just as important as quantities (being on the PPP) for measuring efficiency. That is, prices truly matter.

5.3 Diewert and Fox Cost Constrained Value Added Function

We will now obtain a decomposition of value added growth into efficiency changes, changes in output prices, changes in primary inputs, changes to input prices, technical progress, and returns to scale, replicating the work of Diewert and Fox (2016). To this end we will make use of the cost constrained value added function. We will not rely on any convexity assumptions as we did in the previous section, and instead use the Free Disposal Hull (FDH) approach. As previously mentioned, in a recession we will not be on the production possibilities frontier, so it is important to have a growth accounting methodology that accounts for technical and allocative efficiency, which is what we do here. A key feature of this model is that efficiency will be *distinct* from technical progress, and we will not allow for technical regress during our recession years.

Let $y \in \mathbb{R}^M$ be a vector of M net outputs and let $x \in \mathbb{R}^N$ be a vector of strictly positive inputs, with corresponding positive prices p and w respectively. Denote the production possibilities set in year t as S^t. We then define the cost constrained value added function as

$$R^t(p, w, x) = \max_{y,z} \{p^T y \mid (y, z) \in S^t \,;\, w^T z \leq w^T x\} \tag{269}$$

If (y^t, x^t) is our observed production vector, then we say production is *value added efficient* if our observed value added, $p^{tT} y^t = R^T(p^t, w^t, x^t)$. It will often be the case that production will not be efficient, so we will have the inequality $p^{tT} y^t \leq R^T(p^t, w^t, x^t)$ which leads us to define our metric of value added or net revenue efficiency during the period as

$$e^t = \frac{p^{tT} y^t}{R^T(p^t, w^t, x^t)} \leq 1 \tag{270}$$

If $e^t = 1$ then we say production is allocatively efficient. We also note that (270) is just a net revenue counterpart to the theory of Farrell's cost based measure of overall efficiency developed in the previous section.

We now begin the work of deriving our decomposition for the cost constrained value added function defined by (269). To this end we will make use of index numbers to "isolate" the effects independent factors that enter R^t. We first define the change in value added efficiency ϵ^t as

$$\epsilon^t = \frac{e^t}{e^{t-1}} = \left(\frac{p^{tT} y^t}{R^t(p^t, w^t, x^t)}\right) \Big/ \left(\frac{p^{(t-1)T} y^{t-1}}{R^{t-1}(p^{t-1}, w^{t-1}, x^{t-1})}\right) \tag{271}$$

where if $\epsilon^t > 1$ value added efficiency has improved from period t to $t-1$ and fallen if $\epsilon_t < 1$. We next define our Paasche and Laspeyres net output price indexes as

$$\alpha_L^t = \frac{R^{t-1}(p^t, w^{t-1}, x^{t-1})}{R^{t-1}(p^{t-1}, w^{t-1}, x^{t-1})} \tag{272}$$

$$\alpha_P^t = \frac{R^t(p^t, w^t, x^t)}{R^t(p^{t-1}, w^t, x^t)} \tag{273}$$

where note in both measures above all factors are held fixed except output prices p^t. As (272) and (273) are both equally valid measures as output price indexes, we will then take a geometric average of them to form:

$$\alpha^t = (\alpha_L^t \alpha_P^t)^{1/2} \tag{274}$$

which we define as our overall net output price index. For our input quantity index we take

$$\beta_L^t = \frac{w^{(t-1)T} x^t}{w^{(t-1)T} x^{t-1}} \tag{275}$$

$$\beta_P^t = \frac{w^{tT} x^t}{w^{tT} x^{t-1}} \tag{276}$$

and then we take their geometric average to form the overall input quantity index

$$\beta^t = (\beta_L^t \beta_P^t)^{1/2} \tag{277}$$

We next define our input mix index in the cost constrained value added function as

$$\gamma_{LPP}^t = \frac{R^t(p^{t-1}, w^t, x^t)}{R^t(p^{t-1}, w^{t-1}, x^t)} \tag{278}$$

$$\gamma_{PLL}^t \frac{R^{t-1}(p^t, w^t, x^{t-1})}{R^{t-1}(p^t, w^{t-1}, x^{t-1})} \tag{279}$$

Note: The "Laspeyres" input mix index (278) uses the period t cost constrained value added function R^t in the above definition, and the "Paasche" input mix index (278) uses the period $t-1$ cost constrained value added function R^t in the above definition. The reason for this abnormality is it will allow us to get an *exact* growth decomposition shortly. To form our overall input mix index we then take a geometric average of the above Paasche and Laspeyres measures.

$$\gamma^t = (\gamma^t_{LPP}\gamma^t_{PLL})^{1/2} \tag{280}$$

We will take technical progress to be defined in terms of upward shifts in the production function. If there is technical progress from period $t-1$ to t, then we expect R^t to be greater than R^{t-1} ceteris paribus. Our Paasche and Laspeyres measures of technical progress for the cost constrained value added function are:

$$\tau^t_L = \frac{R^t(p^{t-1}, w^{t-1}, x^t)}{R^{t-1}(p^{t-1}, w^{t-1}, x^t)} \tag{281}$$

$$\tau^t_P = \frac{R^t(p^{t-1}, w^t, x^{t-1})}{R^{t-1}(p^t, w^t, x^{t-1})} \tag{282}$$

Note: Above input vector x^t enters the Laspeyres measure (281) and input vector x^{t-1} enters the Paasche measure (282), where again this is done to aid us in getting an exact decomposition. This change can be justified because we will shortly assume constant returns to scale, and when this is the case it turns out that τ^t_L is independent of x^t and τ^t_P is independent of x^{t-1}, so (281) and (282) are then to be considered "true" Laspeyres and Paasche indexes.

The last family of indexes we will define will be our measures of returns to scale. Returns to scale will be defined as output growth divided by input growth from period $t-1$ to t but technology is held constant when we compute our output growth measure. Our input growth measure is $w^T x^t / w^T x^{t-1}$ and our measure of output growth will be $R^s(p, w, x^t)/R^s(p, w, x^{t-1})$. We are thus in a position to define our Laspeyres and Paasche measures of returns to scale as

$$\delta^t_L = \left(\frac{R^{t-1}(p^{t-1}, w^{t-1}, x^t)}{R^{t-1}(p^{t-1}, w^{t-1}, x^{t-1})}\right) \Bigg/ \left(\frac{w^{(t-1)T}x^t}{w^{(t-1)T}x^{t-1}}\right) \tag{283}$$

$$\delta^t_P = \left(\frac{R^t(p^t, w^t, x^t)}{R^t(p^t, w^t, x^{t-1})}\right) \Bigg/ \left(\frac{w^{tT}x^t}{w^{tT}x^{t-1}}\right) \tag{284}$$

We can then take a symmetric (geometric) average of the above two measures of returns to scale to form an overall index of returns to scale as

$$\delta^t = (\delta^t_L \delta^t_P)^{1/2} \tag{285}$$

It can be seen that if the period $t-1$ production possibilities set S^{t-1} is a *cone* (recall a subset S of a vector space V is a cone if $x \in S$ and c is a positive scalar, then $cx \in S$) so that production is subject to constant returns to scale, then $\delta_L^t = 1$, and if S^t is a cone then $\delta_P^t = 1$ Indeed, in our empirical work we will assume constant returns to scale and thus have $\delta_t = 1$.

Indeed, using our previously defined indexes we now note that the observed value added ratio going from period $t-1$ to period t can be written in two ways:

$$\frac{p^{tT}y^t}{p^{(t-1)T}y^{t-1}} = \epsilon^t \alpha_P^t \beta_L^t \gamma_{LPP}^t \delta_L^t \tau_L^t \tag{286}$$

$$\frac{p^{tT}y^t}{p^{(t-1)T}y^{t-1}} = \epsilon^t \alpha_L^t \beta_P^t \gamma_{PLL}^t \delta_P^t \tau_P^t \tag{287}$$

where when the algebra of the above is written out in all its glory, one must again be careful to note that γ_{LPP} uses R^t and γ_{PLL} uses R^{t-1}, as well as that τ_L uses x^t and τ_P uses x^{t-1} Indeed we can multiply (286) and (287) together to get:

$$\left(\frac{p^{tT}y^t}{p^{(t-1)T}y^{t-1}}\right)^2 = (\epsilon^t)^2(\alpha_P^t\alpha_L^t)(\beta_P^t\beta_L^t)(\gamma_{LPP}^t\gamma_{LPP}^t)(\delta_P^t\delta_L^t)(\tau_P^t\tau_L^t) \tag{288}$$

and we can then take square roots to form our overall indexes as previously described:

$$\frac{p^{tT}y^t}{p^{(t-1)T}y^{t-1}} = \epsilon^t \alpha^t \beta^t \gamma^t \delta^t \tau^t \tag{289}$$

and assuming S^{t-1} and S^t are cones (constant returns to scale), we get that $\delta^t = 1$ so our value added growth decomposition we will work with going forward is

$$\frac{p^{tT}y^t}{p^{(t-1)T}y^{t-1}} = \epsilon^t \alpha^t \beta^t \gamma^t \tau^t \tag{290}$$

We are now in a position to discuss our metric of total factor productivity growth. We will define TFPG going from period $t-1$ to period t to be defined as an index of output growth divided by an index of input growth. Our output growth will be the value added ratio deflated by our value added price index α^t. Our index of input growth is just β^t. Thus we define our period t TFPG growth rate as

$$TFPG^t = \frac{\text{Output Growth}}{\text{Input Growth}} \tag{291}$$

$$= \left(\left[\frac{p^{tT}y^t}{p^{(t-1)T}y^{t-1}}\right]/\alpha^t\right)\Big/\beta^t \tag{292}$$

$$= \epsilon^t \gamma^t \tau^t \tag{293}$$

where the last equality follows by re-arranging equation (290). Thus we have decomposed total factor productivity growth from period $t-1$ to t as the product of value added efficiency change ϵ^t, an input mix index γ^t and technical progress τ^t

We will now perform a levels (Kohli) decomposition for the observed level of nominal value added in period t, $p^{tT}y^t$ relative to the observed level of nominal value added in period 1, $p^{1T}y^1$. This will be instructive when we present it in graphical form. Indeed, we proceed as in Section Three by first "telescoping" the value ratio as follows, and then applying equation (290):

$$\frac{p^{tT}y^t}{p^{1T}y^1} = \left(\frac{p^{tT}y^t}{p^{(t-1)T}y^{t-1}}\right)\left(\frac{p^{(t-1)T}y^{t-1}}{p^{(t-2)T}y^{t-2}}\right)\left(\frac{p^{(t-2)T}y^{t-2}}{p^{(t-3)T}y^{t-3}}\right)...\left(\frac{p^{2T}y^2}{p^{1T}y^1}\right) \quad (294)$$

$$= (\epsilon^t\alpha^t\beta^t\gamma^t\tau^t)(\epsilon^{t-1}\alpha^{t-1}\beta^{t-1}\gamma^{t-1}\tau^{t-1})\ ...\ (\epsilon^2\alpha^2\beta^2\gamma^2\tau^2) \quad (295)$$

$$= (\epsilon^t\epsilon^{t-1}...\epsilon^2)(\alpha^t\alpha^{t-1}...\alpha^2)(\beta^t\beta^{t-1}...\beta^2)(\gamma^t\gamma^{t-1}...\gamma^2)(\tau^t\tau^{t-1}...\tau^2) \quad (296)$$

Now we define and set our cumulated explanatory variables to be 1 at $t = 1$:

$$EC^1 \equiv 1 \;\; A^1 \equiv 1 \;\; B^1 \equiv 1 \;\; C^1 \equiv 1 \;\; T^1 \equiv 1$$

for $t = 2, ..., T$ we then define our cumulated variables recursively as:

$$\begin{aligned} EC^t &= \epsilon^t EC^{t-1} \\ A^t &= \alpha^t A^{t-1} \\ B^t &= \beta^t B^{t-1} \\ C^t &= \gamma^t C^{t-1} \\ T^t &= \tau^t T^{t-1} \end{aligned}$$

Recursively substituting into the above equations we get

$$\begin{aligned} EC^t &= \epsilon^t\epsilon^{t-1}...\epsilon^2 EC^1 \\ &= \epsilon^t\epsilon^{t-1}...\epsilon^2(1) \\ &= \epsilon^t\epsilon^{t-1}...\epsilon^2 \end{aligned}$$

and similarly

$$\begin{aligned} A^t &= \alpha^t\alpha^{t-1}...\alpha^2 \\ B^t &= \beta^t\beta^{t-1}...\beta^2 \\ C^t &= \gamma^t\gamma^{t-1}...\gamma^2 \\ T^t &= \tau^t\tau^{t-1}...\tau^2 \end{aligned}$$

We can then apply these formulas to (296) to get a *levels* decomposition for period t nominal value added relative to period 1, indeed:

$$\frac{p^{tT}y^t}{p^{1T}y^1} = A^tB^tC^tEC^tT^t \quad (297)$$

We can apply the same steps to get a levels decomposition for total factor productivity. Indeed, set the level of TFP at time $t = 1$ to be

$$TFP^1 = 1$$

and for time $t = 2, ..., T$ recursively define TFP^t as follows:

$$TFP^t = TFPG^t TFP^{t-1} \tag{298}$$

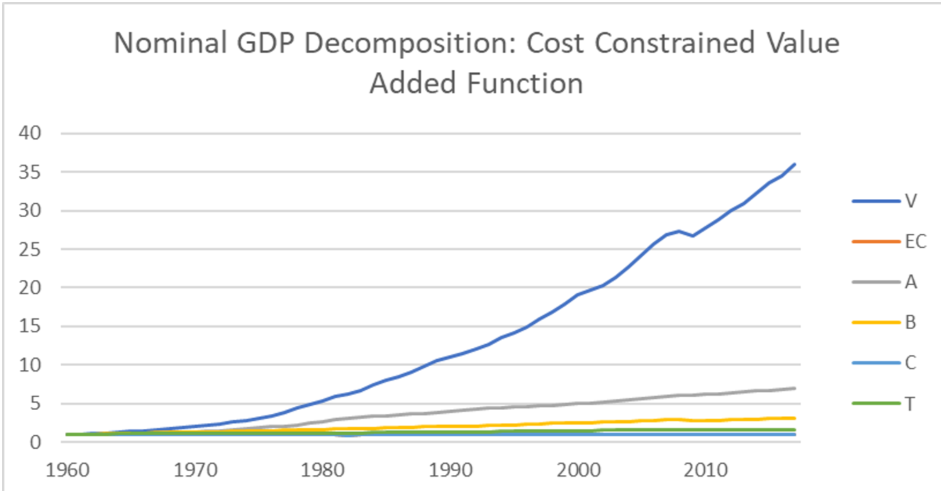
where $TFPG^t$ is defined by (293). We can recursively substitute into (298) to get

$$\begin{aligned} TFP^t &= TFPG^t TFPG^{t-1} ... TFPG^2 TFP^1 \\ &= TFPG^t TFPG^{t-1} ... TFPG^2 (1) \\ &= (\epsilon^t \gamma^t \tau^t)(\epsilon^{t-1} \gamma^{t-1} \tau^{t-1}) ... (\epsilon^2 \gamma^2 \tau^2) \\ &= (\epsilon^t \epsilon^{t-1} ... \epsilon^2)(\gamma^t \gamma^{t-1} ... \gamma^2)(\tau^t \tau^{t-1} ... \tau^2) \\ &= EC^t C^t T^t \end{aligned} \tag{299}$$

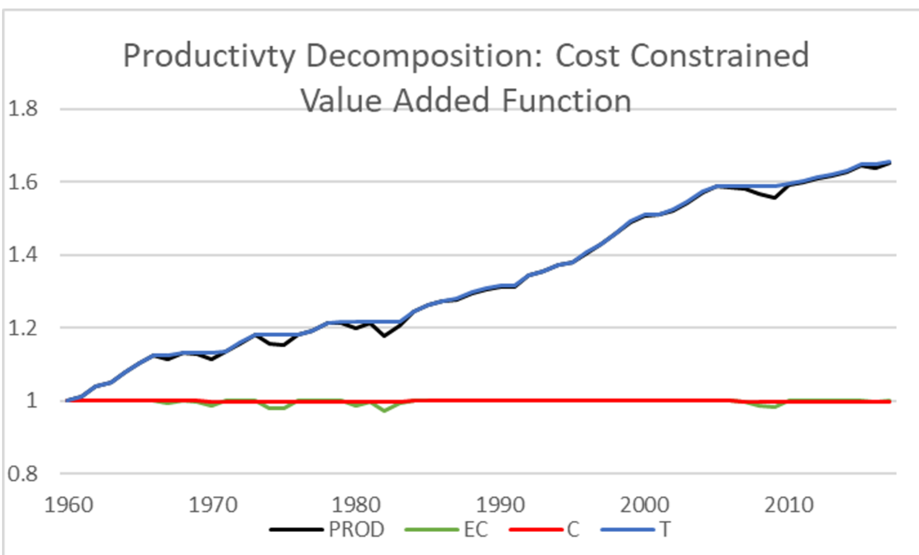
Equations (297) and (299) will form the basis of our empirical work which we now present. Indeed, consider the nominal GDP decomposition of the cost constrained value function defined by (297). The following table shows our cumulated explanatory facors as of 2017

	V	EC	A	B	C	T
2017	35.9382	1	6.92092	3.14685	0.99786	1.65364

It is seen that nominal GDP grew 36 fold over the sample period. Interesting recessions (our cumulated efficiency factor), did not contribute negatively to value added growth over the sample period by the time we reach 2017. If we had ended our sample period at 2009 we would have had $EC = 0.98194$. Nominal GDP growth is mostly explained by output price growth, $A = 6.92$, input quantity growth, $B = 3.14685$, and technical progress $T = 1.65364$. Finally, one should note that if we are to take the geometric mean of our cumulated level of technical progress, we get get .88% technical progress per year on average which is the exact same as our index number estimate from before. One should also note that the multiplicative decomposition $V = EC \cdot A \cdot B \cdot C \cdot T$ holds for the values in the table above, which is a check on our work.



We now examine our levels TFP decomposition defined by (299), $TFP^t = EC^tC^tT^t$. In this case what is more interesting than 2017 levels is the movements of the explanatory variables over the sample period. Indeed consider the following graph which shows the levels of our cumulated efficiency factor, mix input growth factor, and technical progress.



Here we see that there was a substantial decline in value added efficiency over the years 2006-09, as well as from 1974-76, 1980-83 and 1991. Thus our efficiency factor is picking up recession years as we would hope. The most important thing to note in the above graph is that during these recession years, total factor productivity will fall (substantially), but our measure of *technical progress* will not. That is, with this decomposition we have successfully decoupled technical progress from total factor productivity. This can be contrasted with the Diewert Morrison decomposition from before where we took TFP to be technical progress and this fell during recessions where ideally we would like our model to rule out technical regress as the Diewert and Fox decomposition does. Lastly, note this graph shows the US productivity slowdown since the great recesson, as well as the unsatisfactory productivity of the 70s. Otherwise US productivity growth has been robust over this period.

6 Conclusion

Strong productivity growth has played a vital role in the large increase in per capita living standards in the USA in the postwar period. In this paper we have measured this productivity growth by way of Solow Residual, index numbers, econometrics and non-parametric techniques. It was seen that on average, all estimates ended up being very close to each other, in the range of .88-.93% per year. There was substantial variation about this mean, with all methods uncovering strong productivity growth until 1973, followed by a slowdown due to the oil shocks and high inflation of the 70s, and then the gestation period associated with newly discovered information technology being put into practice. This began to change in the mid 90s as productivity surged during the information technologies and communications revolution, which lasted until the mid 00s. More recently US productivity performance has been dismal. This has ignited a debate between researchers as to whether this slowdown is a temporary lull or a permanent structural shift. Among the most cynical is Bob Gordon, who has satirically said "When I go out I like to play a game: spot the robot" which epitomizes his view that current technological innovations are useless and that the USA has entered a permanent low growth period. Among the more optimistic is Erik Brynjolfsson who believes the history of the early 90s is being repeated and that the USA is on the precipice of another era of strong productivity growth, this time due to advances in artificial intelligence technologies.

We took the uniformity of our productivity estimates by different methods to be a triumph of productivity measurement as a science. Indeed, disparate methods of measurement all arriving at the same answer on average lends credence to the argument that we have defined and understood the problem in a way that is fundamentally correct. Although in the end "all roads lead to Rome", there are certainly reasons to apply or prefer one method to another. The Solow Residual was seen to be the easiest to derive and the fastest to apply to the data. Index Number methods made the least tenuous assumptions on underlying economic behaviour and were seen to provide our most "pure" form of measurement. Econometric methods smoothed out the index number estimates and estimated equations provided us with other valuable economic insights such as elasticities. Non-parametric methods nicely framed the problem in terms of inefficiencies and provided us with a decomposition that did not allow for technical regress. It is our opinion that index number methods provide the most rigorous and justifiable way of measuring TFP growth, though this debate is not central to our results.

Although our estimates of productivity growth were seen to be satisfactory, we have identified certain problems and areas for further research and inquiry. First, we believe the two good (labour-leisure) consumer model as presented in section 4.3.1 needs substantial amendments. A static model is not the correct framework to study labour-leisure choice. Indeed, this choice is driven largely by expectations of future incomes (as in a life-cycle earnings model) and any model which does not incorporate these expectations is not only insufficient but incorrect. We believe the income elasticities of demand for labour and leisure generated by this model should be thrown out and the model should be re-estimated in a dynamic setting. Next, as there is now a substantial emerging literature on "misallocation" problems driving productivity trends across countries and within countries, as seen in Hsieh and Klenow (2009) and Hsieh and Moretti (2018) we believe the methods described in this paper should be used to produce *regional* productivity estimates for the USA. Such estimates would provide key evidence in proving the hypothesis that the observed productivity slowdown may be driven by regional slowdowns acting in conjunction with a misallocation of resources across regions. Lastly, we believe the project of adding land and natural resources to our asset base should be undertaken. By not including these assets in our capital inputs, we calculated nominal and real rates of return which were "too high" and thus may have systematically biased our results.

Going forward, it is clear that no single factor is more important to the US economy than productivity growth. As population growth flatlines, the solvency of vital social programs such as Social Security and Medicare depend almost entirely on strong economic growth and thus productivity gains in order to pay for them. Wages are becoming increasingly stagnant, and productivity growth will ultimately be the only way to measurably raise them if current and future generations are to have any chance at obtaining a higher living standard than that of their parents. Furthermore rising wages from productivity gains would be the easiest way to hold enormous sociological problems such as inequality and identity politics at bay. This paper has made a modest attempt to measure US productivity growth, and going forward the task of both continually updating and improving upon these statistics and methods of measurement will be of critical importance.

References:

Abrams, Burton (2006), "How Richard Nixon Pressured Arthur Burns: Evidence from the Nixon Tapes" Journal of Economic Perspectives vol. 20, no. 4, 177-188

Brynjolfsson, Erik, Daniel Rock and Chad Syverson (2017), "Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics." NBER Working Paper No. 24001

Cowen, Tyler (2011) *The Great Stagnation: How America Ate All the Low-Hanging Fruit, Got Sick, and Will (Eventually) Feel Better* New York: Dutton

Diewert, W.E. (1971), "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function", Journal of Political Economy vol. 79, issue 3, 481-507

Diewert, W.E. (1976), "Exact and Superlative Index Numbers", Journal of Econometrics 4, 114-145.

Diewert, W.E. (1978), "Superlative Index Numbers and Consistency in Aggregation", Econometrica 46, 883-900.

Diewert, W.E. and C.J. Morrison (1986), "Adjusting Output and Productivity Indexes for Changes in the Terms of Trade", The Economic Journal 96, 659-679.

Diewert, W.E. and T.J. Wales (1988), "A Normalized Quadratic Semiflexible Functional Form", Journal of Econometrics 37, 327-42.

Diewert, W.E. (1992), "Exact and Superlative Welfare Change Indicators", Economic Inquiry 30, 565-582.

Diewert, W.E. (2009), ECON594: Applied Economics. Chapter 10: Linear Programming (Lecture Notes) Retrieved from `https://econ.sites.olt.ubc.ca/files/2016/05/pdf_course_erwin-diewert-ECON594Ch10.pdf`

Diewert, W.E. (2018), ECON580: Index Number Theory and Economic Measurement. Chapters 0-8 (Lecture Notes) Retrieved from `https://econ.sites.olt.ubc.ca/files/2015/02/pdf_course_erwin-diewert-ECON580Ch3.pdf`

Farrell, M.J. (1957), "The Measurement of Production Efficiency", Journal of the Royal Statistical Society, Series A, 120, 253-278.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. (2015) "The Next Generation of the Penn World Table." American Economic Review, 105 (10):

3150-82

Gordon, R.J. (2010), "Revisiting U. S. Productivity Growth over the Past Century with a View of the Future," NBER Working Papers 15834

Gordon, R.J. (2014). "The Demise of U.S. Economic Growth: Restatement, Rebuttal, and Reflections," NBER Working Papers 19895

Gordon, R.J. (2018). "Why Has Economic Growth Slowed When Innovation Appears to be Accelerating?," NBER Working Papers 24554

Griliches, Zvi (1988), "Productivity Puzzles and R&D: Another Nonexplanation." The Journal of Economic Perspectives 2 : 9-21.

Jorgenson, Dale, Frank Gallop, and Barbara Fraumeni. (1987) *Productivity and U.S. Economic Growth.* Cambridge: Harvard University Press

Jorgenson, Dale, Mun S Ho, and Kevin J Stiroh. (2008) "A Retrospective Look at the U.S. Productivity Growth Resurgence." Journal of Economic Perspectives 22, no.1 3-24

Jorgenson, Dale, Mun S. Ho, and Jon D. Samuels. (2015) "The Impact of Information Technology on Postwar U.S. Economic Growth." Telecommunications Policy

Hsieh, Chang-Tai, and Peter J Klenow. (2009). "Misallocation and Manufacturing TFP in China and India." The Quarterly Journal of Economics, 124(4): 1403–1448.

Hsieh, Chang-Tai, and Enrico Moretti. (2018)"Housing Constraints and Spatial Misallocation" NBER Working Paper No. 21154

Kaldor, Nicholas (1957). "A Model of Economic Growth". The Economic Journal. 67 (268): 591–624.

Karabarbounis, Loukas, and Brent Neiman. (2014) "The Global Decline of the Labor Share." The Quarterly Journal of Economics, 129(1): 61–103.

Krugman, Paul (1994) *The age of diminished expectations: U. S. Economic policy in the 1990s* Cambridge, MA: MIT Press

Konüs, A.A. (1939), "The Problem of the True Index of the Cost of Living", Econometrica 7, 10-29. [Originally published in 1924]

Nordhaus, William. (1982) "Economic Policy in the Face of Declining Productivity Growth." European Economic Review, pp.131-58.

Romer, Paul (1986), “Increasing Returns and Long-Run Growth” Journal of Political Economy”, 94(5), 1002-1037

Romer, Paul (1990). “Endogenous Technological Change. Journal of Political Economy”, 98(5), S71-S102.

Solow, Robert M. (1957). “Technical change and the aggregate production function”. Review of Economics and Statistics. The MIT Press. 39 (3): 312–20.

7 Appendix

7.1 Detailed Consumer Model Data Construction

Here we describe the construction of our data used in the detailed consumer model as presented in section 4.4. The source of this data was OECD.stat, Detailed National Accounts Table 5, (Dataset 5) Final Consumption Expenditures of Households. This dataset encompasses the years 1971-2016, and when compared with its older counterparts (revisions have been made) was seen to be almost identical so we chose it to be the sole source of data for this section; no linking was done between different sources or older revisions. Expenditure was first read in at both current prices (nominal) and constant prices (OECD base year) for twenty categories of consumer goods which were as follows:

VC1	P31CP010: Food and non-alcoholic beverages
VC2	P31CP021: Alcoholic beverages
VC3	P31CP022: Tobacco
VC4	P31CP030: Clothing and footwear
VC5	P31CP041: Actual rentals for housing
VC6	P31CP042: Imputed rentals for housing
VC7	P31CP043: Maintenance and repair of the dwelling
VC8	P31CP044: Water supply and miscellaneous housing services
VC9	P31CP045: Electricity, gas and other fuels
VC10	P31CP050: Furnishings, households equipment and routine maintenance
VC11	P31CP060: Health
VC12	P31CP071: Purchase of vehicles
VC13	P31CP072: Operation of personal transport equipment
VC14	P31CP073: Transport services
VC15	P31CP080: Communications
VC16	P31CP090: Recreation and culture
VC17	P31CP100: Education
VC18	P31CP110: Restaurants and hotels
VC19	P31CP120: Personal and Financial Services
VC20	P33: Final consumption expenditure of resident households abroad

Where the VC_i above are the nominal values, and the counterpart QC_i, $1 \leq i \leq 20$ were the corresponding real values for the exact same categories. These values were read in at millions of dollars. We then divided all series (real and nominal) to make our units billions. We next generated price series by deflating nominal values by real values, so that for each good i:

$$PC_i = \frac{VC_i}{QC_i} \qquad\qquad 1 \leq i \leq 20$$

The next task was to form a housing aggregate, that is to consolidate categories 5,6,7 and 8 above. We formed the 4x1 price and quantity vectors for each year t as

$$q^t = (QC_5^t, QC_6^t, QC_7^t, QC_8^t)$$
$$p^t = (PC_5^t, PC_6^t, PC_7^t, PC_8^t)$$

We chose a chained Fisher Ideal index to do the aggregation, that is for year t we let the *chain link* be:

$$P_{FCHL}(p^{t-1}, p^t, q^{t-1}, q^t) = \left(\frac{p^{tT} q^{t-1}}{p^{(t-1)T} q^{t-1}}\right)^{1/2} \left(\frac{p^{tT} q^t}{p^{(t-1)T} q^t}\right)^{1/2}$$

so the period t index then is

$$P_F(p^0, p^t, q^0, q^t) = P_{FCHL}(p^0, p^1, q^0, q^1) P_{FCHL}(p^1, p^2, q^1, q^2) ... P_{FCHL}(p^{t-1}, p^t, q^{t-1}, q^t)$$

which then along with the value ratio in any year defines the implicit quantity index. As a check on our work we present the 2016 values of this price index, as well as the Divisia (chained Törnqvist-Theil), Paasche and Laspeyres.

	Divisia	Fisher	Laspeyres	Paasche
Chained Index	6.417	6.417	6.424	6.410

Here we see our 2^{nd} order approximation property at work, the Törnqvist-Theil and Fisher are equal to 3 decimal places. We also note the spread between the Paasche and Laspeyres is negligible, so we were justified in chaining this smooth trending long run time series data. Lastly we see the Fisher, which we take to be our price index for this new series, falls perfectly between the Paasche and Laspeyres.

For comparison purposes we also present 2016 price indexes for our housing aggregate if we were not to have chained, that is

$$P_L(p^0, p^t, q^0, q^t) = \frac{p^{tT} q^0}{p^{0T} q^0} \qquad P_P(p^0, p^t, q^0, q^t) = \frac{p^{tT} q^t}{p^{0T} q^t}$$

$$P_F(p^0, p^t, q^0, q^t) = P_L^{1/2} P_P^{1/2}$$

where our empirical software Shazam automatically chains the Törnqvist-Theil so it will remain unchanged.

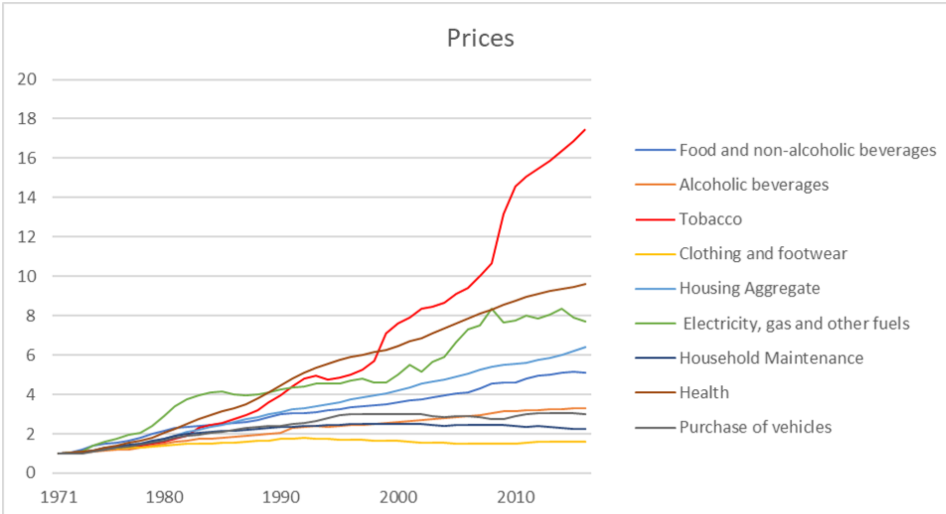
	Divisa	Fisher	Laspeyres	Paasche
Index	6.417	6.438	6.494	6.382

Now we have the inequality $P_L > P_P$ as we would expect without chaining and the spread between them is quite large. Although this data is at a lower level of aggregation, it is yearly and smooth trending so we see we were clearly justified in chaining (there is no bouncing or drifting when I look at the individual series).

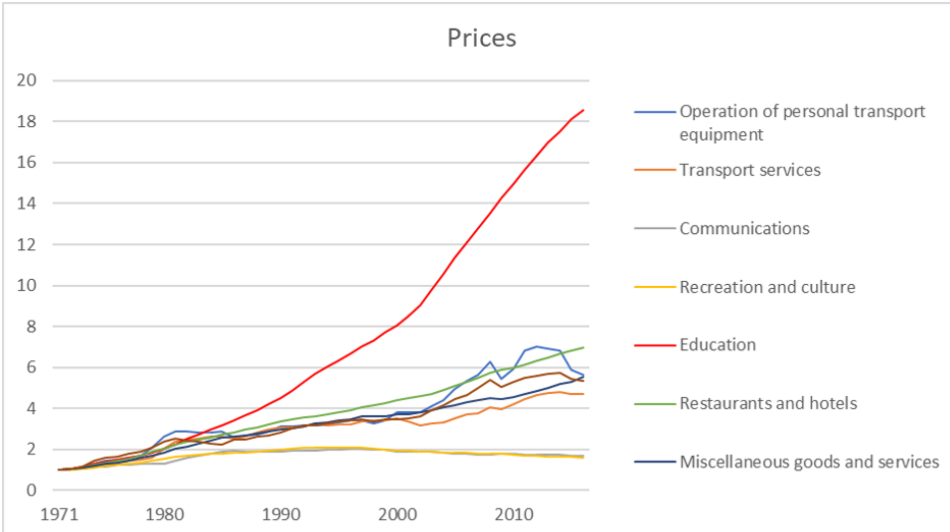
We let our chained Fisher Ideal price index be the price series for our housing aggregate and the quantity index by the quantity series for our housing aggregate. After re-labelling, we now have the *17 good model* where series are defined as follows, for which of course there are correpsonding prices.

Q1	Food and non-alcoholic beverages
Q2	Alcoholic beverages
Q3	Tobacco
Q4	Clothing and footwear
Q5	Housing Aggregate
Q6	Electricity, gas and other fuels
Q7	Furnishings, households equipment and routine maintenance
Q8	Health
Q9	Purchase of vehicles
Q10	Operation of personal transport equipment
Q11	Transport services
Q12	Communications
Q13	Recreation and culture
Q14	Education
Q15	Restaurants and hotels
Q16	Miscellaneous goods and services
Q17	Final consumption expenditure of resident households abroad

We finish up this section by providing plots for these series as they are of general interest and lend insight into the results of our normalized quadratic consumer preferences estimation.



We see that tobacco experienced enormous inflation over the sample period due to various sin taxes. Next is healthcare whose price increased almost *10 fold* in a span of less than 50 years. The USA has a notoriously inefficient healthcare system and also pays far more for pharmaceuticals than any country on the planet. We can see the oil shocks of the 70s and the surge in oil prices leading up to 2008 figure prominently into our fuels series, and it is also interesting to note that recently this series is in decline, most likely due to falling electricity and natural gas prices, as well as cheaper oil. Lastly we note that clothing and footwear experienced less inflation over the sample period than any other good, most likely due to the introduction of cheap Chinese and Vietnamese manufacturing.



Of all 17 consumer goods education experienced the most inflation over the sample period. This is most likely due to the enormous bureaucratic overheard and administrative costs at universities. Education becoming prohibitively expensive will certainly be a large problem for American society and will exacerbate inequality going forward. We lastly note that communications had almost no inflation over the sample period, due to massive technological improvements.

Good	2016 Price Level
Food and non-alcoholic beverages	5.09
Alcoholic beverages	3.31
Tobacco	17.45
Clothing and footwear	1.58
Housing Aggregate	6.42
Electricity, gas and other fuels	7.69
Household Furnishing and Maintenance	2.23
Health	9.63
Purchase of vehicles	3.01
Operation of personal transport equipment	5.64
Transport services	4.70
Communications	1.70
Recreation and culture	1.62
Education	18.57
Restaurants and hotels	6.98
Miscellaneous goods and services	5.53
Final consumption expenditure abroad	5.33

For a discussion of our quantity series the reader is referred to section 4.4.3 where we normalized these series and then discussed their geometric growth rates over the sample period

7.2 A Proof of Schlömilch's Inequality

Recall from Section 2.3 the definition of the weighted mean of order r as

$$M_r(x) = \Big(\sum_{n=1}^{N} \alpha_n x_n^r\Big)^{1/r}$$

where $\sum_{n=1}^{N} \alpha_n = 1$

Theorem (Schlömilch's Inequality): Let $x >> 0_N$ but $x \neq k1_N$ for any $k > 0$ and let $r < s$. Then,

$$M_r(x) < M_s(x)$$

That is, $M_r(x)$ is an i*ncreasing function of r*. In proving this result, we will use the following lemma.

Lemma: Let $\alpha, y \in \mathbb{R}^N$, with $\alpha, y >> 0_N$ such that $\sum_{i=1}^{N} \alpha_i = 1$. Then

$$\sum_{n=1}^{N} \alpha_n y_n \log(y_n) \geq (\alpha^T y) \log(\alpha^T y)$$

Proof: Consider the function $f(x) = x \log x$. Then $f'(x) = \log x + x(1/x) = \log x + 1$ and $f''(x) = (1/x) > 0$ for all x on its domain of deinition so $f(x)$ is a convex function. The left hand side of Lemma 1 can be re-written in terms of f, yielding

$$\sum_{n=1}^{N} \alpha_n y_n \log(y_n) = \sum_{n=1}^{N} \alpha_n f(y_n)$$

As our α_i sum to 1 and we have shown f to be convex, we can then apply Jensen's Inequality:

$$\begin{aligned}\sum_{n=1}^{N} \alpha_n f(y_n) &\geq f\big(\sum_{n=1}^{N} \alpha_n y_n\big)\\ &= f(\alpha^T y)\end{aligned}$$

which then re-substituting in our definition of f gives

$$\sum_{n=1}^{N} \alpha_n y_n \log(y_n) \geq (\alpha^T y) \log(\alpha^T y)$$

Proof of Schlömilch: As

$$\frac{\partial \log M_r(x)}{\partial r} = \frac{1}{M_r(x)} \frac{\partial M_r(x)}{\partial r}$$

and $M_r(x) > 0$ for $x >> 0_N$ as we have assumed, it will thus be sufficient to show $\frac{\partial \log M_r(x)}{\partial r} > 0$ for $r \neq 0$ and we will deal with the $r = 0$ case separately at the end. Indeed

$$
\begin{aligned}
\frac{\partial \log M_r(x)}{\partial r} &= \frac{-1}{r^2} \log\left(\sum_{n=1}^{N} \alpha_n x_n^r\right) + \frac{1}{r}\left(\frac{1}{\sum_{n=1}^{N} \alpha_n x_n^r}\right)\left(\sum_{n=1}^{N} \alpha_n x_n^r \log x_n\right) \\
&= \frac{-1}{r^2} \log\left(\sum_{n=1}^{N} \alpha_n x_n^r\right) + \frac{1}{r^2}\left(\frac{1}{\sum_{n=1}^{N} \alpha_n x_n^r}\right)\left(\sum_{n=1}^{N} \alpha_n x_n^r \log x_n^r\right) \\
&= \frac{-1}{r^2} \log\left(\sum_{n=1}^{N} \alpha_n y_n\right) + \frac{1}{r^2}\left(\frac{1}{\sum_{n=1}^{N} \alpha_n y_n}\right)\left(\sum_{n=1}^{N} \alpha_n y_n \log y_n\right) \\
&= \frac{-1}{r^2} \log\left(\alpha^T y\right) + \frac{1}{r^2}\left(\frac{1}{\alpha^T y}\right)\left(\sum_{n=1}^{N} \alpha_n y_n \log y_n\right) \\
&> \frac{-1}{r^2} \log\left(\alpha^T y\right) + \frac{1}{r^2}\left(\frac{1}{\alpha^T y}\right)\left(\alpha^T y \log \alpha^T y\right) \\
&= 0
\end{aligned}
$$

where in the third equality we let $y_n = x_n^r$, in the fourth equality we apply the definition of the inner product and the inequality follows by simply applying our lemma to the right most term in the fourth equality. We finish up by now letting $r = 0$. Indeed first consider $s > r = 0$:

$$
\begin{aligned}
(M_0(x))^s &= \left(\prod_{n=1}^{N} x_n^{\alpha_n}\right)^s \\
&= M_0(x_1^s, ..., x_N^s) \\
&< M_1(x_1^s, ..., x_N^s) \\
&= \sum_{n=1}^{N} \alpha_n x_n^s \\
&= M_s(x)^s
\end{aligned}
$$

where the inequality follows simply by AM-GM. Thus raising both sides to the positive power $1/s$ we get

$$M_0(x) < M_s(x)$$

Now consider $r < s = 0$

$$
\begin{aligned}
M_r(x) &= \frac{1}{M_{-r}(x_1^{-1}, ..., x_N^{-1})} \\
&< \frac{1}{M_0(x_1^{-1}, ..., x_N^{-1})} \\
&= M_0(x)
\end{aligned}
$$

where the inequality follows from the above $s > r = 0$ case since $-r > 0$ so $M_r(x) < M_0(x)$ and we are done.