Wealth Inequality and the Velocity of Money in the UK

Samuel Forbes

10.9.2018

University of Warwick

Supervisors: Dr Stefan Grosskinsky and Dr Alexander Karalis Isaac

Abstract

The velocity of money relates two fundamental macroeconomic quantities of gross domestic product (GDP) and money supply and is a measure of the frequency of transactions in the economy. Initially it was expected to be a constant value, however data showed that this is not the case. Partially due to this reason from the 1990s, velocity took a back seat in economic theory. Since the 2008-2009 financial crises there has been renewed interest in the role of money in the economy. We introduce in detail the concept of the velocity of money and find for UK data that its trend has a strong inverse relationship to rising wealth inequality from the 1980s until the financial crash of 2008-2009. We hypothesise why this may be the case and attempt to find a prefactor to the velocity equation to stabilise the velocity during this time period.

Contents

1 Introduction

Modern economics is often quoted to have its beginnings with the publication of the *The Wealth of Nations* in 1776 by Adam Smith [28]. From these foundations the unprecedented rise in industrial and technological advancement has led to an ever more complex economic society and many competing economic theories. Macroeconomics is concerned with factors at play in the economy as a whole. John Meynard Keynes is attributed to revolutionising macroeconomic theory during the great depression of the 1930s [13]. In particular he attacked claims by the quantity theorists who regarded the money supply in the economy as the main driving force of prices [25]. One of their claims that Keynes criticised was that the velocity of money, a measure of the frequency of transactions in the economy, is constant. As we shall see it has since been shown empirically that Keynes was correct in this regard. After Keynes' attack an adapted version of monetary theory was laid out by the Monetarists led by Milton Friedman [25]. Monetary theory again fell out of favour after the experience of stagflation in the 1980s and since then increasingly complex mathematical models based on rational expectations came to the fore [13]. However many of these models failed to predict the 2008-2009 financial crises and since then there has been a renewed focus on the role of money [12].

We start by introducing the velocity of money from a historical perspective within the quantity theory of money. Then we give an explanation of GDP and money supply followed by presenting UK data and calculating the velocity. Some basic time series analysis is performed suggesting that velocity follows random walk properties. We then discuss income/wealth inequality and how it has been rising since a low in the 1970s across most of the world. Two measures of income/wealth inequality, the top 1% income share and the Gini coefficient for the UK, are presented. The contribution of econophysics to the study of income/wealth inequality is discussed briefly.

We then present the main result of this project where we find a strong inverse relationship between income/wealth inequality and velocity in the UK from 1981-2008. Finally we attempt to find a prefactor based on this relationship that stabilises velocity. We discuss briefly some possible reasons why, although flattening the trend in velocity, the prefactor does not completely remove it.

2 Velocity of Money

2.1 The Velocity Equation and the Quantity Theory of Money

The velocity equation of money in it's modern form is stated as [25, 28]

$$\mathrm{V} = \frac{\mathrm{PY}}{\mathrm{M}} \tag{1}$$

where P is a measure of prices such as the consumer price index (CPI) or the gross domestic product (GDP) deflater, Y is real GDP and M is the money supply. Details of these macroeconomic terms will be explained in section 2.3. Velocity can be thought of as a measure of the frequency of monetary transactions in the economy [28], thus high velocity means there are more transactions taking place than low velocity and the money supply is being used more efficiently. If we are to think about velocity having units as in a physical system it would be $[\mathrm{P}] = 1$, $[\mathrm{Y}] = \frac{\text{money}}{\text{time}}$, $[\mathrm{M}] = \text{money}$, and hence $[\mathrm{V}] = \frac{1}{\text{time}}$. Hence we can see that velocity of money is a per time measurement like frequency in the normal physical sense.

A formulation of equation (1) was first postulated by Irving Fisher and called the *equation of exchange* [26]. In his formulation y was the actual physical number of transactions. As the number of transactions in the economy is an impossible value to measure precisely this was later replaced by gross national product (GNP) and now GDP.

The velocity of money was a mathematical realisation inspired by the quantity theory of money, one of the oldest surviving economic theories going back to at least the mid-16th century with the French social philosopher Jean Bodin [26]. He thought the inflation affecting the Western world at the time could be attributed to the abundance of monetary metals imported from the South American colonies by the Spanish. The quantity theory hypothesises that changes in the value of money is determined primarily by changes in the quantity in circulation. Five key ideas in the early quantity theory were:

1. the prices P will vary in exact proportion to the quantity of M, this was called the proportionality postulate, this postulate led to the belief that velocity is stable or constant,
2. the direction of causation goes from M to P: monetary changes precede and cause price changes, suggested first by David Hume,
3. the neutrality postulate which surmises that apart from transitional adjustment periods, monetary changes have no influence on real economic variables such as total output, employment and the product mix (the total range of products offered by a company),
4. the price level is affected predominantly by changes in the quantity of money, which leads to the proposition that inflation or deflation is caused primarily by the money supply rather than other factors,
5. the nominal money supply is independent of the demand for money, called the exogeneity of the nominal stock of money.

The quantity theory had an early influence on monetary policy in matters such as international finance and the role of the money supply in inflation.

Keynes attacked many views in the quantity theory such as assumptions on full employment and the constancy of velocity. There was a counter attack by the Moneterists led by Friedman where emphasis was put on the demand of money.

Since 1980 the quantity theory of money and velocity has fallen out of favour in part because velocity changed substantially from 1980 until the present day [28] and also due to the dominance of the theory of rational expectations [12].

As the 2008-2009 financial crises was largely not predicted due to insufficient focus on financial institutions, notably the lending of subprime mortgages [13, 28], it may be beneficial for a renewed focus on how the role of money and in particular velocity affects the economy [12].

2.2 Recessions

Recessions play a major part in macroeconomics as understanding them often drives economic theory. We present a brief summary of the main recessions in the UK since the great depression in Table 1 as we shall refer to them in our subsequent analysis of time series data and especially in velocity.

A simple definition of a recession is a period of negative GDP growth, usually two quarters or more. However how a recession is actually characterised is a lot more complicated [8].

Recession	Summary
The great depression: 1929-1933	Originated in the US which affected world trade. The outdated monetary system of the gold standard may have also had influence. UK was not hit as badly as other developed nations, particularly the US. Worst recession in history of Western world.
Stagflation: 1973-1976	Postwar boom came to an end. Unemployment and inflation rose simultaneously and cost of living hit a postwar peak.
Manufacturing meltdown: 1980-1981	Margaret Thatcher came to power and raised interest rates to try to tackle inflation, however inflation still rose by over 20%. Inflation, high borrowing costs, and cheap imports drove manufacturing down 20%.
Lawson's legacy: 1990-1992	Following the Lawson boom of the second half of the 1980s, interest rates, inflation and unemployment rose to close to 3 million.
Financial crises: 2008-2009	Collapse of the Lehman Brothers in the US and the run on Northern Rock in the UK due to the lending of subprime mortgages amongst other factors caused the largest downturn in modern history.

Table 1: Brief summaries of the main UK Recessions since the great depression [22]

2.3 Data

2.3.1 GDP

Gross domestic product or GDP is the total value of final goods and services produced in the economy during a given period [28]. It is usually measured quarterly or yearly and is a difficult quantity to estimate. GDP comes in two forms, real and nominal. Real GDP (Y) adjusts for inflation, nominal GDP does not and takes the value of that year's prices. Nominal GDP is equal to real GDP multiplied by the price (PY) so velocity (1) can also be expressed as

$$\mathrm{V} = \frac{\text{Nominal GDP}}{\mathrm{M}} \tag{2}$$

where M is the money supply as in (1). The price P can be taken as the GDP deflator or the consumer price index. These price measures have subtle differences between them [28]. Figure 1 shows quarterly values of nominal GDP from 1955-2018 and real GDP from 1975-2018 at 2010 prices on a log scale. As can be seen GDP has been increasing exponentially over this time period. Notice the dip around the vertical line passing through 2008 around the 2008-2009 financial crises.

A common measure is the yearly figure of public sector debt as a percentage of GDP. The government wants to keep this ratio low with an aim of an upper limit of 40% [31]. See Figure 2 for this percentage from 1948-2017 for the UK. The trend is U-shaped during this period with a dramatic decrease from 1948 (just after World War II) and then an increasing trend starting around the 2008-2009 financial crash indicated by the vertical line.

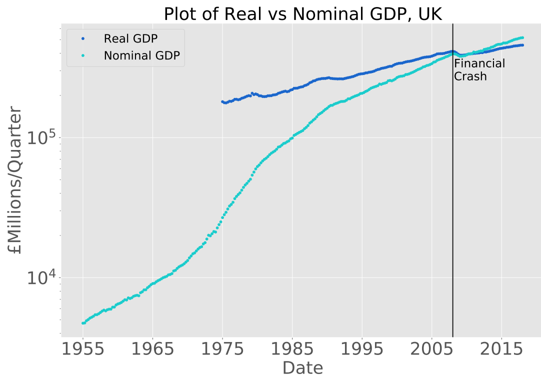


Figure 1: Quarterly real (1975-2018) and nominal GDP (1955-2018) for the UK. Vertical line indicates 2008-2009 financial crises.
Data from [3]

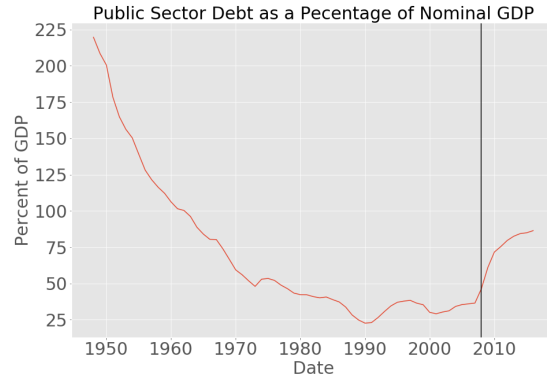


Figure 2: 1948-2017 yearly percentage of public sector debt to nominal GDP for the UK. Vertical line indicates 2008-2009 financial crises. Debt data from [3], yearly nominal GDP data from [5]

2.3.2 Money Supply

There are several aggregate money supply measures [29]. Over time new measures have arisen to account for the increasing ways money can exist. The most common measures in the UK over the last century have been notes and coin, M0, M1, M2, M3 and M4 which forms the containment:

$$\{\text{notes and coin}\} \subset \text{M0} \subset \text{M1} \subset \text{M2} \subset \text{M3} \subset \text{M4}.$$

The smallest measure, notes and coin, first calculated in 1870 are the notes and coin outside of the bank of England and can be used as an indicator of cash based transactions. M0, also established in 1870, is notes and coin and central bank reserves. M1, M2 and M3 add consecutively more monetary measures in the form of deposits. M4, first measured in 1963, is the broadest measure of the monetary supply and includes repositories and securities. Since 2007 there is a slightly modified version of M4 called M4^{ex} which excludes deposits of intermediate other financial corporations (IOFCs). See Figure 3 for the log scale plot of quarterly values of M0 from 1950-2017 and M4 from 1963-2018. Note the rapid rise of M0 which is indicative of the increase in bank reserves from the policy of quantitative easing (QE) in response to the 2008-2009 financial crises [10]. In this report we will focus on the broadest measure M4 as it encapsulates all money and wealth that contributes to GDP [16], which will become useful in section 4.

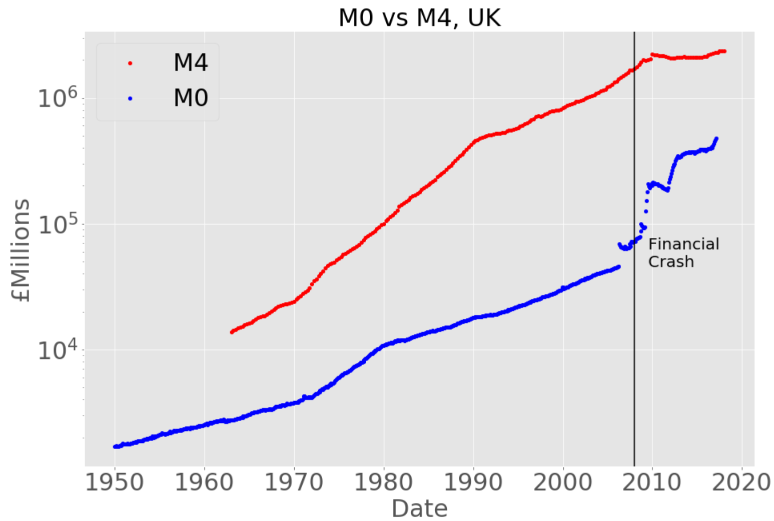


Figure 3: Log scale plot of quarterly values of the money supply M0 (1950-2017) and M4 (1963-2018). Data from [1]

2.4 Velocity

With the data for nominal GDP and the money supply we can now divide the two to find the velocity as in equation (2). The velocities with quarterly nominal GDP and quarterly M0 and M4 are shown in a dual scale plot in Figure 4. The vertical lines are the four recessions in the UK since the mid-1970s recession. As can be seen velocity for M0, let us call V0, does not always follow the same trend as velocity for M4, which we now call V4. As is evident the velocities are not constant which was supposed by early quantity theorists.

From this point forward we will concentrate on V4. The trend for V4 is oscillating from 1963-1980, at which point it follows a downwards trend till 2010, at which point it starts rising. Notice how V4 changes trend at or soon after each of the four recessions.

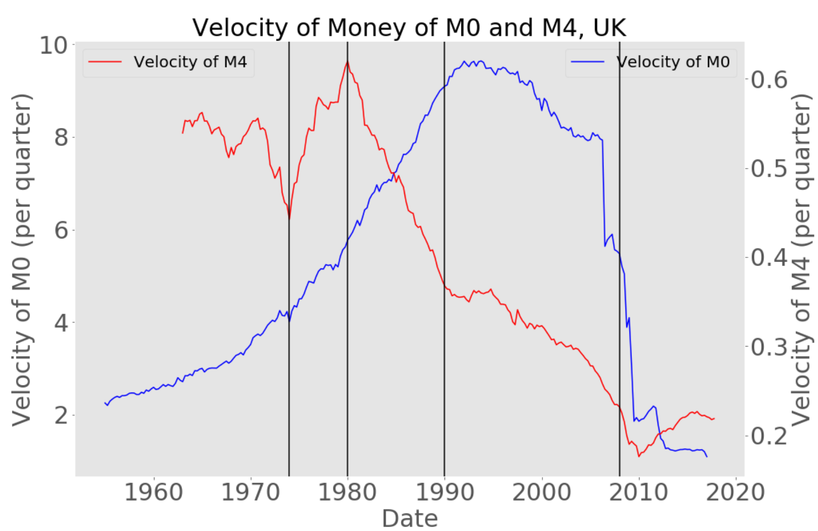


Figure 4: Velocity per quarter of M0 (1955-2017) and M4 (1963-2017) for the UK. Vertical lines indicate beginning of recessions, see Table 1

2.5 Analysis

Macroeconomic measures are time series data: each data point is indexed by time, usually monthly, quarterly or yearly. As the data points are taken at equally space intervals they are called **discrete**. In the case of V4 we have quarterly data. We proceed with some relevant definitions from time series analysis [19].

2.5.1 Definitions

Definition 1 (Stochastic process)**.** A stochastic process is a collection of random variables taking values in $\mathbb{Z}$ or $\mathbb{R}$, ordered in time and defined at a set of time points which may be continuous or discrete.

For short we shall call a stochastic process simply a process.

Definition 2 (Purely random process)**.** A discrete-time process is called a purely random process if it consists of a sequence of random variables $\{Z_t\}$ which are mutually independent and identically distributed.

It follows from the definition that a purely random process has constant mean and variance. One simple example of a model of a time series is the autoregressive process. There are several other models that we shall not go into here such as mixed ARMA models and integrated ARIMA models.

Definition 3 (Autoregressive process)**.** Suppose $\{Z_t\}$ is a purely random process with mean zero and variance σ^2. Then a process $\{X_t\}$ is said to be an autoregressive process of order p, abbreviated an AR(p) process, if

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t \tag{3}$$

Equation (3) is of the same form as a multiple regression model however the independent variables are instead past values of X_t hence the name autoregressive.

A particular example of an autoregressive process is the random walk. This is an AR(1) process with $\alpha_1 = 1$. With $\{Z_t\}$ and $\{X_t\}$ as in Definition 3 and $\mathbb{E}[X_t] = \mu$, the random walk is defined as

$$X_t = X_{t-1} + Z_t. \tag{4}$$

which can take values in $\mathbb{Z}$ or $\mathbb{R}$. A random walk starting at $X_0 = 0$ gives $X_t = \sum_{i=1}^{t} Z_i$ with mean $\mathbb{E}[X_t] = t\mu$ and variance $\mathrm{Var}(X_t) = \mu\sigma^2$.

A measure of correlation of a time series and a delayed version of itself of k time units is the autocorrelation.

Definition 4 (Autocorrelation of lag k)**.** The autocorrelation of a sample $\{x_t\}$ with mean $\bar{x}$ of a discrete time series $\{X_t\}$ is defined:

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2}. \tag{5}$$

Note that $r_0 = 1$.

Definition 5 (Correlogram)**.** A plot of r_k against the lag k is called a correlogram.

In general terms a time series is said to be stationary if there is no systematic change in mean (no trend) and no systematic change in variance, and if there is no periodicity. Formally:

Definition 6 (Stationary)**.** A time series $\{X_t\}$ is said to be (strictly) stationary if the joint distribution of $X_{t_1}, \ldots, X_{t_n}$ is the same as $X_{t_1+\tau}, \ldots, X_{t_n+\tau}$ for all $t_1, \ldots, t_n, \tau$ for all $n \in \mathbb{N}$.

A time series is called non-stationary if it is not stationary. Often non-stationary time series can be made stationary by differencing as is the case with a random walk.

Given a time series of length N $\{x_1, \ldots, x_N\}$ from a process $\{X_t\}$, first differencing results in a new time series $\{y_1, \ldots, y_{N-1}\}$ where

$$y_t = x_{t+1} - x_t = \nabla x_{t+1}$$

and ∇ is the difference operator.

First differencing can be continued to form a new time series again of one less in length. For instance second order differencing of $\{x_1, \ldots, x_N\}$ results in the time series $\{w_1, \ldots, w_{N-2}\}$ where

$$w_t = y_{t+1} - y_t = (x_{t+2} - x_{t+1}) - (x_{t+1} - x_t) = x_{t+2} - 2x_{t+1} + x_t = \nabla^2 x_{t+2}.$$

Definition 7 (Integrated)**.** A non-stationary time series that becomes stationary after differencing d times is called integrated of order d and denoted I(d).

The AR(p) process can be rewritten $(1 - \alpha_1 L - \cdots - \alpha_p L^p)X_t = Z_t$ where L is the lag operator giving $LX_t = X_{t-1}$.

Definition 8 (Characteristic equation)**.** The characteristic of the AR(p) process is defined as the polynomial

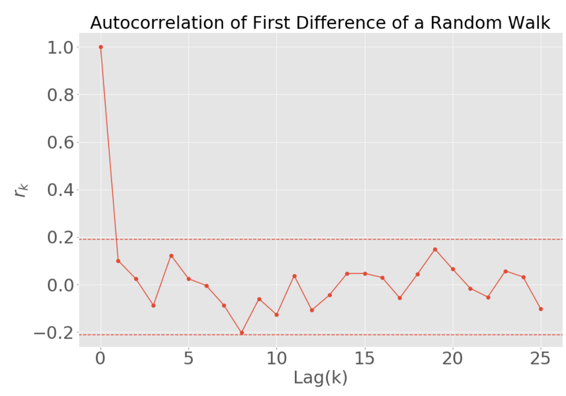
$$1 - \alpha_1 m - \cdots - \alpha_p m^p = 0.$$

Definition 9 (Unit root)**.** An AR(p) process is said to have a unit root if it's characteristic equation has a root m_r of ± 1.

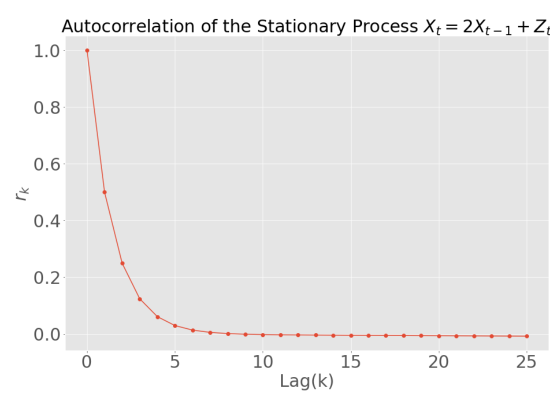
It can be shown that an AR(p) process is stationary if all the roots m_r of its characteristic equation lie outside the unit circle i.e. $|m_r| > 1$ and non-stationary if the absolute value of all the roots are less than one i.e. $|m_r| < 1$ or it has a unit root.

A random walk is I(1) and has a unit root as its characteristic equation is $m - 1 = 0$.

If a time series $\{x_1, \ldots, x_N\}$ is a sample from a purely random process then $r_k \approx 0$ for all non-zero values of k. For large N it can be proved that if $\{x_1, \ldots, x_N\}$ are independent and identically distributed then $\mathbb{E}[r_k] \approx -1/N$, $\mathrm{Var}(r_k) \approx 1/N$ and r_k is normally distributed [19]. This means a 95% confidence interval for r_k is approximately $-1/N \pm 2/\sqrt{N}$. Stationary time series often give short term correlation characterised by a fairly large value of r_1 followed by a few more consecutive coefficients greater than but tending to 0. The correlation coefficient r_k becomes approximately 0 for longer lags k. See Figure 5 for correlograms of the first difference of a random walk purely random process and of the stationary AR(1) process with $\alpha_1 = 2$.



Correlogram of the first 25 lags of a purely random process of the first difference of a random walk (4) of length 100. Dotted lines is the 95% confidence interval for r_k



Correlogram of the first 25 lags for the stationary process $X_t = 2X_{t-1} + Z_t$ of length 100

Figure 5: Correlogram Examples

2.5.2 Random Walk Properties of V4

If the first differences of V4 is a purely random process then there is evidence that V4 has similar properties to a random walk such as being a memoryless quantity. The plot of the first differences of V4 is shown in Figure 6 and the correlogram for these first differences are shown in Figure 7. As mentioned in the last section for a purely random process $r_k \sim \mathcal{N}(-1/N, 1/N)$ for large N. We have $N = 219$ first difference V4 values. For the first 20 lags, as shown in Figure 7, we would then expect one value of r_k to be out with the 95% confidence interval $(-1/N - 2/\sqrt{N}, 1/N + 2/\sqrt{N}) \approx (-0.14, 0.13)$ however 5 of the first 6 values for $k > 0$ are out with the interval. However for all lags up to 218 we do have 95% of r_k are within the interval for $k > 0$. The use of a correlogram is only a rough visual test for a purely random process.

We also know that a random walk has a unit root. A test for a unit root is the augmented Dickey-Fuller test with null hypotheses that a unit root is present and alternative hypothesis that the series is stationary [24], see Appendix A. Using this test on V4 gives a p-value of 0.837 meaning we accept the null hypothesis and so there is evidence that V4 has a unit root as is the case with a random walk.

By implementing the augmented Dickey Fuller test on the first differences of V4 we find an almost 0 p-value meaning there is evidence the first differences are stationary and thus is I(1) also in line with a random walk.

As was stated in section 2.4, V4 exhibits a change in trend after each of the recessions (1973, 1980, 1990 and soon after 2008, see Table 1) indicating that V4 may not be a pure random walk but may have a drift term a_t that changes. Formally $\{\mathrm{V4}_t\}$ may be of the form

$$\mathrm{V4}_t = a_t + \mathrm{V4}_{t-1} + Z_t$$

where a_t is a constant that is different before and after the change in trends and Z_t is from a random process.

More sophisticated testing for the velocity of M2 was done for 5 countries (UK, US, Canada, Sweden and Norway) from the late 19th century to the mid-1970s in [15] also giving evidence velocity has properties of a random walk.

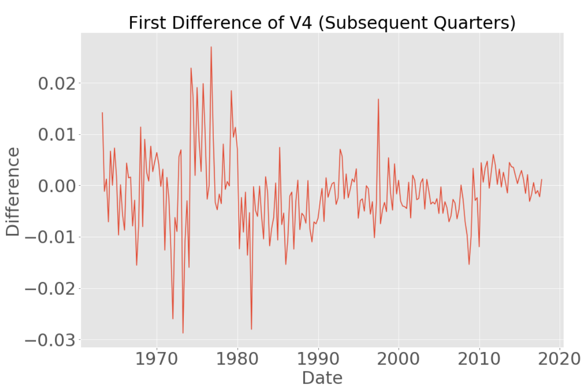


Figure 6: First difference of quarterly V4 from 1963-2017, UK

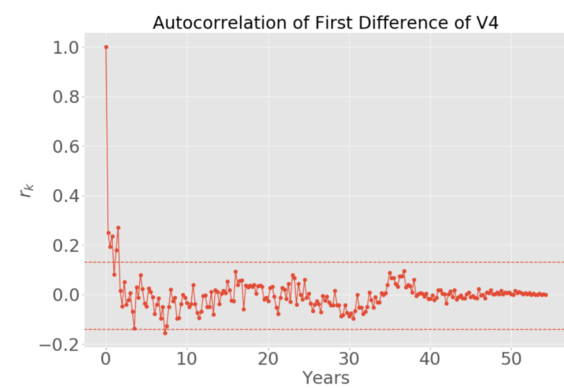


Figure 7: Correlogram of all lags for first differences of quarterly V4 (4 lags represent a year). Dotted lines is the 95% confidence interval for r_k

3 Wealth Inequality

3.1 Introduction

Inequality has a long history [14] and in recent times according to the World Inequality Report (WIR) [11], income inequality has increased in nearly all parts of the world from 1980-2016, see Figure 8.

Calculating overall inequality is difficult due to the lack of publicly available data and the complexity of combining data to get estimated income or wealth measures. Also some measures of inequality rely on household surveys which are known to underestimate the income and wealth inequality of those at the top. The data presented in the WIR was contributed to by over 100 researchers in all the major continents.

Europe as a whole has faired better than other regions. For Western Europe in particular we can see that income share of the bottom 50% decreased by 2% while the top 1% income share increased by roughly 2% from 1980-2016, see Figure 9. Western Europe has faired substantially better than the United States where the top 1% income share has roughly doubled and the bottom 50% income share has roughly halved. Exceptions to this pattern are in the Middle East, sub-Saharan Africa and Brazil where inequality has remained fairly stable but extremely high [11].

The report suggests that the main driver of this inequality trend is the transfer of public to private wealth. Some potential solutions to this inequality phenomenon include more effective progressive taxation measures, a global financial register of financial assets to reduce tax evasion and money laundering and more equal education opportunities and well paying jobs [11]. It remains an open question how feasible it would be to put these potential solutions into practice.

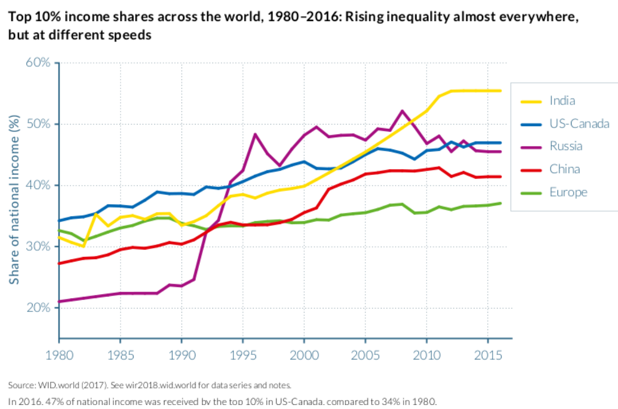


Figure 8: Top 10% income share across the world [11]

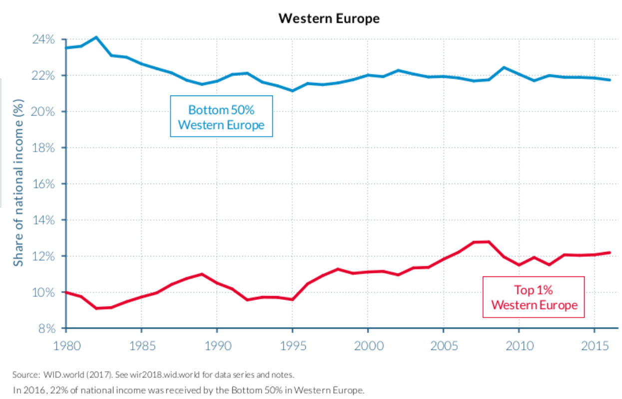


Figure 9: Top 1% and bottom 50% income share in Western Europe [11]

Another measure of wealth or income inequality is the Gini coefficient also called the Gini index named after the Italian statistician and sociologist Corrado Gini [23]. In it's simplest form the Gini coefficient is calculated as the ratio of the area between the line of perfect equality and the Lorenz curve, divided by the area under the line of perfect equality [18]. The Lorenz curve, call $L(x)$, shows the proportion of overall income or wealth share as a function of the proportion x of people.

Mathematically the Gini coefficient, call G, can be defined

$$G := 2\int_0^1 (x - L(x))dx = 1 - 2\int_0^1 L(x)dx. \tag{6}$$

Figure 10 shows the line of perfect equality and a Lorenz curve. G is then the ratio $\frac{A}{A+B} = 2A = 1 - 2B$. Figure 11 shows a fitted Lorenz function to data from 1996 [21].

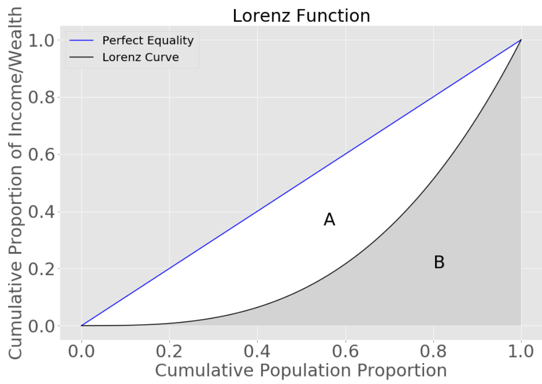


Figure 10: Line of perfect equality, blue ($y = x$) and example of a Lorenz function, black. The areas A and B form the Gini coefficient $G = \frac{A}{A+B}$

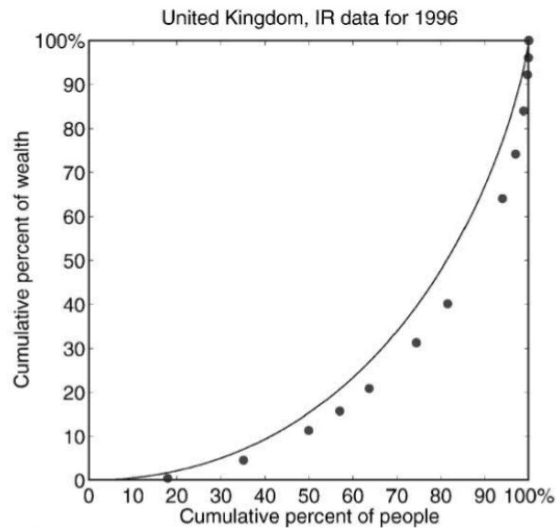


Figure 11: Lorenz function from 1996 fitted to Inland Revenue (IR) UK data with a quoted Gini coefficient of 0.68 [21]

G lies in the interval $[0, 1]$ with perfect equality at the lower bound $G = 0$ where everyone has the same income/wealth share to complete inequality of the upper bound $G = 1$ where one person has all the income/wealth. Increasing G means increasing inequality and the Lorenz curve is 'further away' from the perfect equality line.

3.2 UK Data

We now present wealth inequality data for the UK. Figure 12 shows the top 1% fiscal income share for the years 1915-2014 with some missing data in the early years and 1980. Figure 13 gives the Gini coefficient from 1976-2009.

As can be seen the top 1% fiscal income share was at its highest in 1915 at near 20% to a low of less than 6% in 1977. Since 1980 till around the 2008-2009 financial crises the top 1% fiscal income share increased to more than 15%. Since the crash it has dropped slightly but there is not enough time to see the long term trend.

The Gini coefficient also increases from 1980 till 2008 however there is more of a levelling off since 1990 according to the available data.

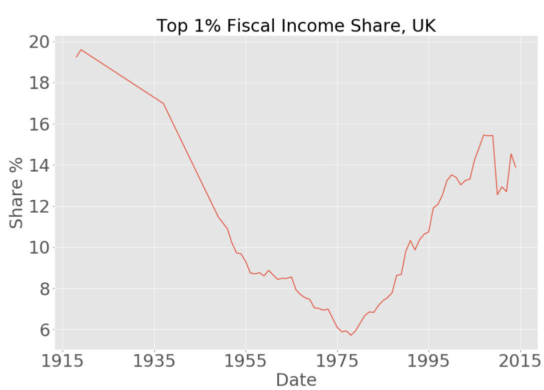


Figure 12: Top 1% fiscal income share. Data from [32]

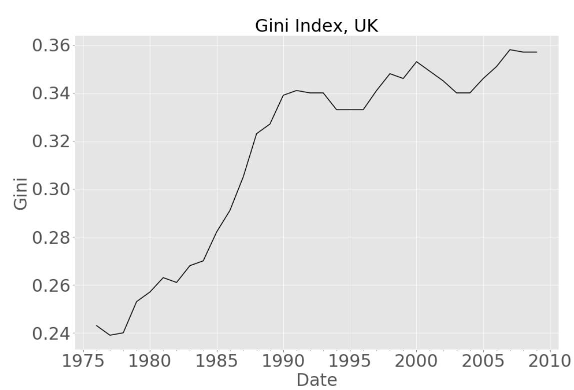


Figure 13: Gini coefficient from 1976-2009. Data from [4]

3.3 Analysis

We now test the UK inequality data for integration which will be useful for section 4.2. Applying the augmented Dickey Fuller test to the first, second and third differences of the top 1% UK fiscal income share indicates that the first and second differences have a unit root whereas the third difference is stationary. This implies the data is I(3). Using just the increasing trend from 1981-2006 gives the same result.

For the Gini coefficient the augmented Dickey Fuller test gives the first difference having a unit root and the second difference as stationary implying this time series is I(2). Thus statistical tests can indicate that two time series are of a different form even though they measure a similar quantity.

3.4 Econophysics

Econophysics is a relatively new field applying mathematical methods from statistical mechanics to economics and finance [35]. The term 'econophysics' was first introduced by the theoretical physicist Eugene Stanley in 1995. As in statistical physics econophysics is concerned with the analysis of large systems which makes it particularly relevant to the study of macroeconomics.

Statistical mechanics originated with the three physicists James Clerk Maxwell, Ludwig Boltzmann, and Josiah Willard Gibbs in the second half of the 18th century and was used to study the statistical properties of atoms. In fact Boltzmann alluded to the fact that on a large scale the behaviour of atoms or molecules in gases could have relations to social phenomena. Econophysics treats individuals in the economy, often called economic agents, as particles in a physical system.

A large subset of econophysics is concerned with the study of income/wealth distributions and thus wealth inequality. A law from statistical mechanics states that the probability of finding a physical system in a state with energy E is

$$\mathbb{P}(E) = \frac{1}{Z} \exp\left(-\frac{k_B E}{T}\right)$$

where k_B is the Boltzmann constant, T is the temperature and Z is a normalising constant. In Physics this exponential distribution is also called the Boltzmann-Gibbs distribution. From a static maximum entropy principle $T \sim \langle E \rangle$ is the average energy per particle. This distribution was derived under the assumption that energy is a conserved quantity; the total energy of a fixed system is constant over time.

The economy can be thought of as a large statistical system. If we assume that the amount of money is a conserved quantity econophysics hypothesises that we could treat money as energy and thus arrive at a Boltzmann-Gibbs distribution for money.

To test this theory a proxy, very basic economy was simulated based on the exchange of a fixed sum between two randomly selected agents repeated many times. When a fixed maximum debt is in place the simulations do indeed give a limiting Boltzmann-Gibbs type distribution. If a maximum debt is not in place then debt and income increase indefinitely.

There are more sophisticated models of the economy such as modelling firms instead of individual agents which again gives exponential type Boltzmann-Gibbs curves.

From data it has been shown that there is evidence that around the first 95% of people follow a Boltzmann-Gibbs distribution and around the last 5% follow a power law called the Pareto distribution, see Figure 14 for 1996 data [21].

More discussion of these wealth distributions fitted to real world data will occur in section 4.3.

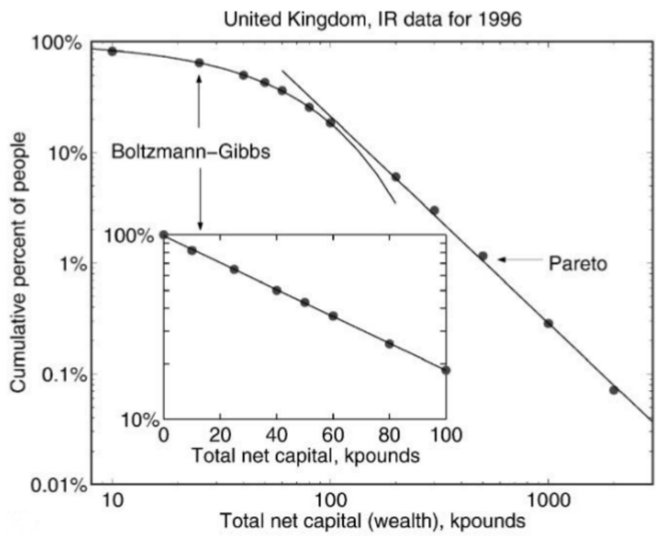


Figure 14: Log-log plot of wealth distribution of Boltzmann-Gibbs and Pareto fit to Inland Revenue (IR) data for 1996, inset log-linear plot of Boltzmann-Gibbs for first 90% of population. Taken from [21]

4 Wealth Inequality and Velocity

4.1 Introduction

This section contains our main results relating wealth inequality and velocity. First we study the relation on a statistical basis and then theoretically.

From observing V4 in Figure 4, and the top 1% fiscal income share in Figure 12, we can see that in the UK between the years of 1980 and before the 2007-2008 financial crises that as inequality rises velocity goes down.

A hypothesis of why both events are inversely related is that people with more income/wealth contribute less to transactions in the economy by holding on to a greater proportion of their income/wealth. Thus if the share of top 1% of income becomes greater, indicating higher inequality, then more wealth is being saved rather than used for transactions contributing to GDP and so the velocity goes down.

From a search of the economic literature there is mention of a possible relationship of wealth and velocity in [15, 16] and a research paper of the possible relation in the US from 1959-1981 [34].

4.2 Analysis

Figure 15 shows both a non-linear and Gaussian process regression [33], see Appendix B, of the top 1% fiscal income share against V4 for the years 1981-2009 which are indicated by the red values. The fits are both of a negative sigmoid shape. The blue values in the upper left are values for the years 2010-2014 and the blue values in the lower right are the years 1963-1979, the year 1980 is missing in the fiscal income data. These blue dots are not involved in the regression but still agree with the general trend of the higher the inequality the lower the velocity.

The R-squared values for the Gaussian process fit and the non-linear fit are both high at roughly 0.98 corresponding with the visually good fits of both curves. Applying the Durbin-Watson test for autocorrelation in the residuals [24], see Appendix C, gives a Durbin-Watson statistic of roughly 1.15 for the non-linear regression which is inconclusive and 1.03 for the Gaussion process regression which is just under the lower critical value for 29 data points and one explanatory variable at the 1% significance level [2] meaning there is evidence in this case that the residuals are autocorrelated. Thus the Durbin Watson test does not support a goodness of fit as in theory residuals should be random and have no autocorrelation [24]. This may suggest that there are other factors at play in influencing velocity which shall be discussed at the end of section 4.3.

It is to be noted that it is often the case that two time series that are highly correlated may be spurious correlations. One way to reduce the likelihood of a spurious relation is to check if the series are cointegrated [19, 20], see Appendix D. However in our case as V4 is I(1) and the top 1% fiscal income share is I(3) we cannot do this so we have to treat the high correlation with a grain of caution.

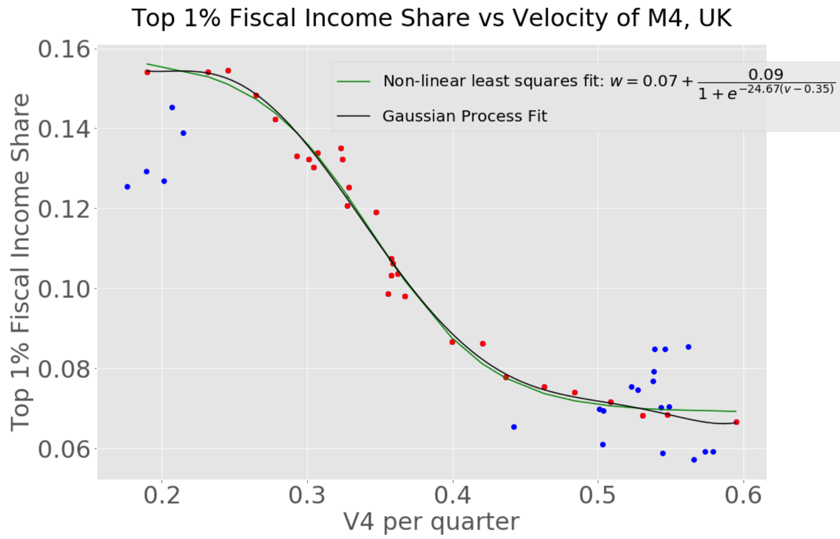


Figure 15: Non-linear least squares and Gaussian process regression of top 1% fiscal income share against quarterly V4 on the red values which are the years 1981-2009. Blue values in upper left years 2010-2014 and blue values in lower right 1963-1979 do not contribute to the regression. Missing the year 1980

4.3 Prefactor to Stabilise Velocity

In this section we attempt to explain the hypothesis of fewer transactions with rising wealth inequality by a theoretical time dependant prefactor to the velocity equation. We want a prefactor γ_t such that when $V4_t$ is divided by it the trend is taken away to become near stable or constant, thus removing the possible effect wealth inequality has on the trend of V4.

We define the prefactor γ_t as:

$$\gamma_t := \frac{1}{\mathbb{E}[M_t]} \int_0^\infty x m_t(x) g_t(x) dx. \tag{7}$$

M_t is a random variable of how much wealth an individual has at time t with corresponding wealth distribution $m_t(x)$ (the probability of having £x amount of wealth) and $g_t(x)$ is the fraction of wealth that is spent, similar the marginal propensity to consume (MCP) [17]. γ_t is the size biased expectation of g_t and can be thought of as the expected fraction of total pounds spent at time t.

We define the random variable M_t with probability density function $m_t(x) := f_{M_t}(x)$ as the mixture of the independent random variables X_t and Y_t which follow Boltzmann-Gibbs and Pareto distributions respectively:

$$M_t = \mu X_t + (1 - \mu) Y_t$$

where $\mu \in [0, 1]$, the probability density function of Boltzmann-Gibbs is defined

$$f_{X_t}(x) = \frac{1}{\beta_t} \exp\left(-\frac{x}{\beta_t}\right) \quad \text{for } x > 0$$

and the probability density function for Pareto is defined

$$f_{Y_t}(x) = \begin{cases} \dfrac{\alpha_t k^{\alpha_t}}{x^{\alpha_t+1}} & \text{for } x \geq k_t, \\ 0 & \text{for } x < k_t. \end{cases}$$

We choose this mixture distribution as there is evidence that the majority of the population's wealth follows a Boltzmann-Gibbs distribution, whilst the top percentage (≈ 5 %) follows a Pareto distribution [21].

The tail function (often called survival function) of a random variable X with density $f(x)$ is defined

$$\overline{F}(x) := P(X > x) = 1 - F(x).$$

We note that

$$\frac{\mathrm{d}}{\mathrm{d}x}\overline{F}(x) = -f(x) \tag{8}$$

and

$$\overline{F}(x) = \int_x^\infty f(y)\mathrm{d}y.$$

We find the tail function for the Boltzmann-Gibbs distribution as

$$\overline{F}_{X_t}(x) = \exp\left(-\frac{x}{\beta_t}\right)$$

and the tail function for the Pareto distribution as

$$\overline{F}_{Y_t}(x) = \begin{cases} \left(\dfrac{k_t}{x}\right)^{\alpha_t} & \text{for } x \geq k_t, \\ 1 & \text{for } x < k_t. \end{cases}$$

The tail function of the mixture distribution of wealth is therefore

$$\overline{F}_{M_t}(x) = \begin{cases} \mu \exp\left(-\frac{x}{\beta_t}\right) + (1-\mu)\left(\dfrac{k_t}{x}\right)^{\alpha_t} & \text{for } x \geq k_t, \\ \mu \exp\left(-\frac{x}{\beta_t}\right) + (1-\mu) & \text{for } x < k_t. \end{cases} \tag{9}$$

Let the lower bound of wealth for the top 1% of wealthiest individuals be x_{*t} at time t. Then assuming $x_{*t} > k_t$

$$\begin{aligned} 0.01 = \overline{F}_{M_t}(x_{*t}) &= \int_{x_{*t}}^\infty m_t(x)\mathrm{d}x \\ &= \mu\overline{F}_{X_t}(x_{*t}) + (1-\mu)\overline{F}_{Y_t}(x_{*t}) \\ &= \mu \exp\left(-\frac{x_{*t}}{\beta_t}\right) + (1-\mu)\left(\frac{k_t}{x_{*t}}\right)^{\alpha_t}. \end{aligned} \tag{10}$$

The top 1% wealth share is the expected amount of wealth in the top 1% divided by the expected amount of overall wealth:

$$\text{share}_t = \frac{\int_{x_{*t}}^\infty x m_t(x)\mathrm{d}x}{\int_0^\infty x m_t(x)\mathrm{d}x}.$$

We have

$$
\begin{aligned}
\int_0^\infty x m_t(x)\mathrm{d}x &= \mathbb{E}[M_t] \\
&= \mu\mathbb{E}[X_t] + (1-\mu)\mathbb{E}[Y_t] \\
&= \mu\beta_t + (1-\mu)\frac{\alpha_t k_t}{\alpha_t - 1}.
\end{aligned}
$$

and by integration by parts and using (8)

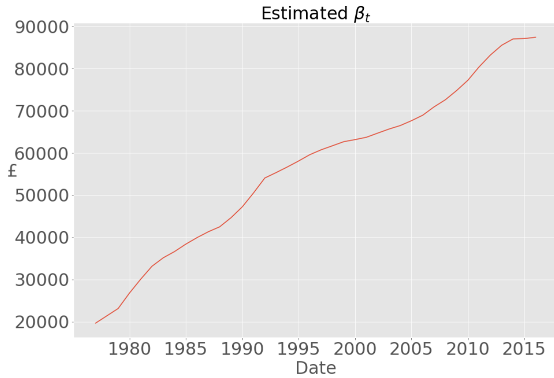
$$
\begin{aligned}
\int_{x_{*t}}^\infty x m_t(x)\mathrm{d}x &= \left[-x\overline{F}_{M_t}(x)\right]_{x_{*t}}^\infty + \int_{x_{*t}}^\infty \overline{F}_{M_t}(x)\mathrm{d}x \\
&= \left[-x\left(\mu\exp\left(-\frac{x}{\beta_t}\right) + (1-\mu)\left(\frac{k_t}{x}\right)^{\alpha_t}\right) - \mu\beta_t\exp\left(-\frac{x}{\beta_t}\right) - \frac{(1-\mu)k_t^{\alpha_t}}{(\alpha_t-1)x^{\alpha_t-1}}\right]_{x_{*t}}^\infty \\
&= x_{*t}\left(\mu\exp(-\frac{x_{*t}}{\beta_t}) + (1-\mu)\left(\frac{k_t}{x_{*t}}\right)^{\alpha_t}\right) + \mu\beta_t\ \exp(-\frac{x_{*t}}{\beta_t}) + (1-\mu)\frac{k_t^{\alpha_t}}{(\alpha_t-1){x_{*t}}^{\alpha_t-1}} \\
&= \mu(x_{*t}+\beta_t)\exp\left(-\frac{x_{*t}}{\beta_t}\right) + (1-\mu)\left(x_{*t} + \frac{x_{*t}}{\alpha_t-1}\right)\left(\frac{k_t}{x_{*t}}\right)^{\alpha_t}.
\end{aligned}
$$

Therefore

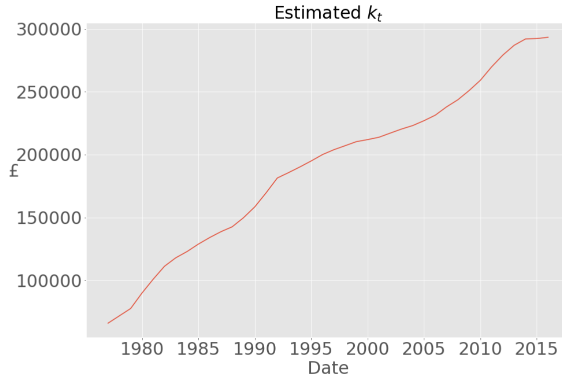
$$
\text{share}_t = \frac{\mu(x_{*t}+\beta_t)\exp\left(-\frac{x_{*t}}{\beta_t}\right) + (1-\mu)\left(x_{*t}+\frac{x_{*t}}{\alpha_t-1}\right)\left(\frac{k_t}{x_{*t}}\right)^{\alpha_t}}{\mu\beta_t + (1-\mu)\frac{\alpha_t k_t}{\alpha_t-1}}. \tag{11}
$$

Choosing the parameters for the year 1996 as $\mu = 0.95$, $k_t =$£200000, $\beta_t =$ £59600 and $\alpha_t = 1.9$ corresponds roughly with [21] (although not exactly: £200000 is roughly the top 8% lower bound in the mixture distribution whereas in [21] it is top 5%, also the mixture distribution bottom 10% upper bound is £6629 which is slightly lower than bottom 10% income from ONS data [5]). With these parameter values, numerically solving for the lower bound of the top 1% in equation (10) gives $x_{*t} =$£474856 (which does roughly correspond with [21]). Plugging all these parameters into (11) gives for 1996, $\text{share}_t = 0.127$. There is data from the World Inequality Database [32] of fiscal income share of the top 1% from 1977-2014 with a missing data point in 1980, see Figure 12. The fiscal income share for 1996 is 0.119, roughly 0.007 lower from the calculated wealth share which makes sense as wealth share is greater than income share [6]. To get an estimate of the 1% wealth share, 0.007 is added to each 1% fiscal income data share point, see Figure 19.

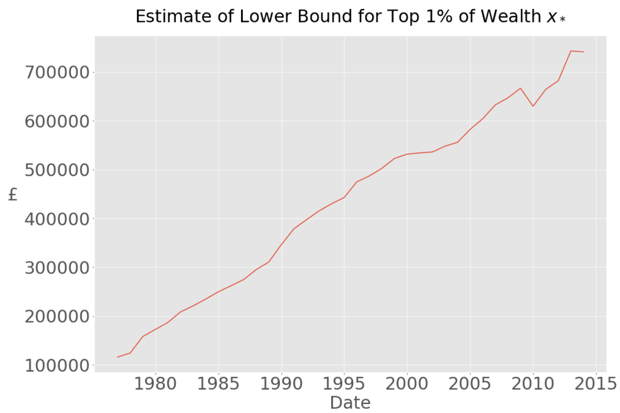
β_t is inferred using the consumer price index (CPI) to get £59600 in prices for other years. k_t is then estimated from this by keeping $\dfrac{k_t}{\beta_t} \approx 3.36$ the same in all years under the assumption that both k_t and β_t evolve with inflation. We fix $\mu = 0.95$ for all the years and set share_t to the estimated top 1% wealth share. It is noted that μ may also vary in time however we make the simplification that it is fixed. We then use these parameters and a non-linear solver to find x_{*t} and α_t for each year from (10) and (11) set to the estimated top 1% wealth share. See Figure 16 for these estimates. We see that α_t decreases as wealth inequality increases giving a larger contribution of the Pareto in $\overline{F}_{M_t}(x)$ (9) for the wealthy individuals with $x > k_t$.



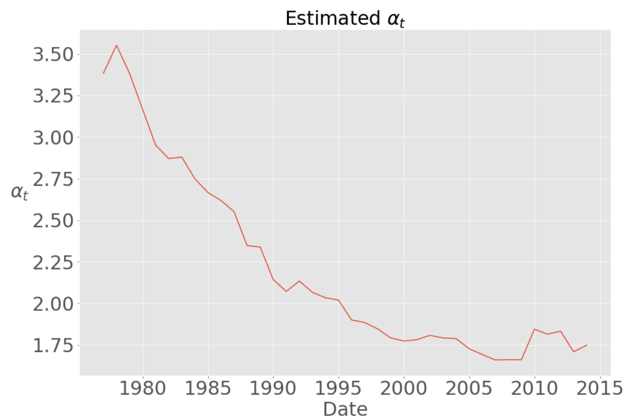
Estimated β_t



Estimated k_t



Estimated Top 1% Lower Bound x_{*t}



Estimated α_t

Figure 16: Parameter value estimates. β_t and k_t from 1977-2015, α_t and x_{*t} from 1977-2014 (excluding 1980)

From these parameters we can plot the tail function $\overline{F}_{M_t}(x)$ (9) of the wealth mixture distribution M_t at different years (the probability of having greater than £x of wealth), see Figure 17.

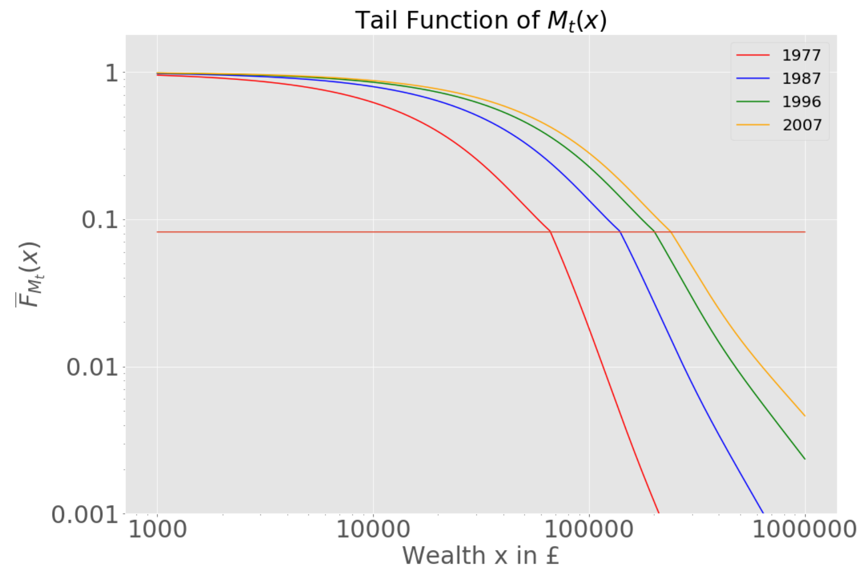


Figure 17: Tail Function $\overline{F}_{M_t}(x)$ of M_t on a log-log scale for the years 1977, 1987, 1996 and 2007. Horizontal line at ≈ 0.08 indicates when the Pareto function starts taking affect in the mixture distribution (for $\mu = 0.95$).

We also have all the necessary parameters to estimate γ_t (7) for 1977-2014 (missing 1980) with numerical integration. However the proportion of wealth spent function g_t still needs to be chosen.

We approximate a function $g_t(x)$ based on [17] as being 1 up to a lower bound B_t of the lower third of wealth and then decreasing non-linearly until it reaches 0.05 at say x_{m_t} at which point it stays for all increasing wealth. We estimate B_t as $\beta_t/3$. Three different functions $g_t(x)$ were tried. Each one is of the form

$$g_t(x) = \begin{cases} 1 & \text{for } x \leq B_t, \\ f(x) & \text{for } B_t < x \leq x_{m_t}, \\ 0.05 & \text{for } x > x_{m_t} \end{cases}$$

with each $f(x)$ in Table 2 corresponding to the colours of $g_t(x)$ in Figure 18 for the year 1996.

Colour	$f(x)$
Red	$\sqrt{\frac{B_t}{x}}$
Blue	$\frac{B_t}{x}$
Green	$\frac{B_t}{x} \log\left(1 + \frac{x}{B_t}\right)$

Table 2: Three choices of $f(x)$

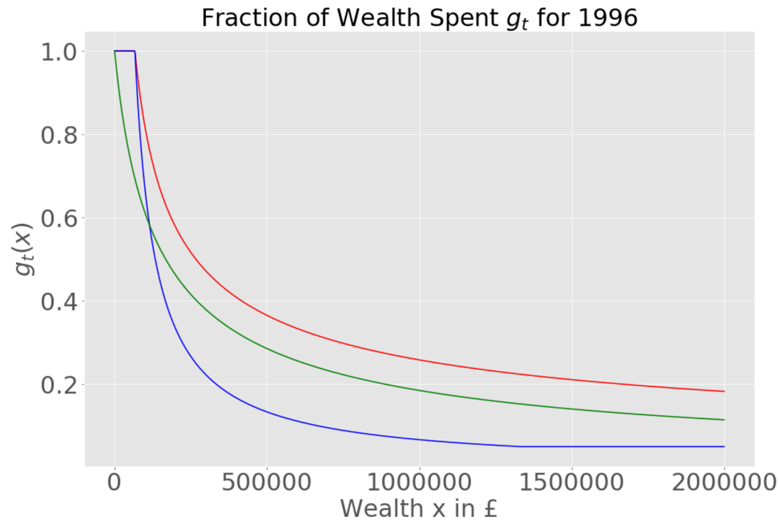


Figure 18: Three possible fraction of wealth spent functions g_t for t =1996

Using these functions we now calculate the prefactor γ_t for each year 1977-2014 (missing 1980). We find for each of the three g_t that γ_t has a very similar slightly decreasing trend until a flattening around 2010, see Figure 20. When $V4_t$ is divided by these three γ_t the downward trend is slightly flattened, with the red γ_t having the most flattening effect, see the dual plot Figure 21.

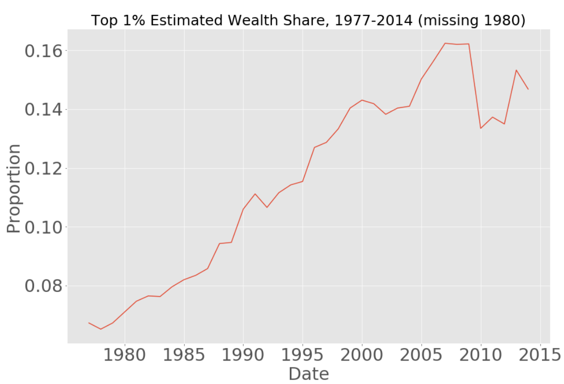


Figure 19: Estimated 1% wealth share

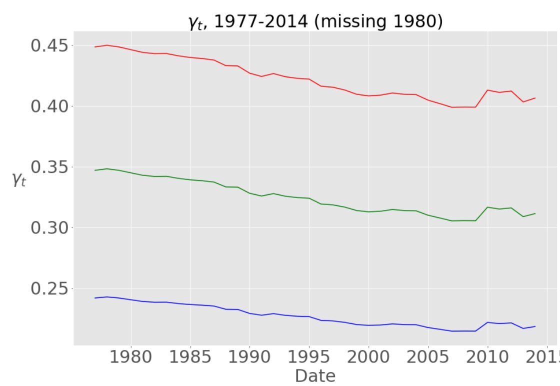


Figure 20: Estimated γ_t for the three proportion of wealth spent functions g_t. Colours correspond to Figure 18

From UK data, 1977-2014 (excluding 1980)

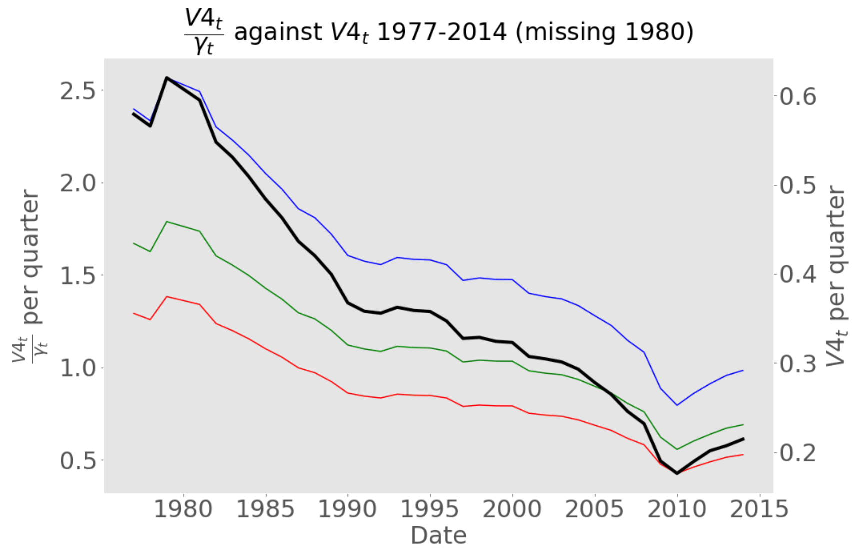


Figure 21: Dual plot of $V4_t$ (black) and $\dfrac{V4_t}{\gamma_t}$ for the three g_t proportion of wealth spent functions. Colours correspond to Figure 20

There are some possible reasons why the trend is not removed: the estimated parameters and functions in γ_t are not chosen accurately enough or there are more factors than wealth influencing velocity.

A factor that may affect velocity is that GDP may not take into account the true number of transactions. For example the growth of mortgages and the finance sector since the 1980s may not be fully captured in GDP [9] but increases the money supply M4 thereby decreasing V4 . Also how foreign trade is incorporated into GDP is complex [30] and as the net export to import (the current account) has decreased a great deal in the UK since the 1980s [7] this may also compromise the relation of GDP to money supply and hence velocity.

5 Conclusion

We have introduced the field of macroeconomics and focused on the velocity equation and income/wealth inequality. We have found a high inverse correlation between V4: the velocity of M4, and wealth inequality, hypothesising that this may be due to wealthier people contributing less to nominal GDP by saving a higher proportion of their wealth.

In the early days of the quantity theory of money it was thought that velocity was a constant value. As we saw this is not the case, however we attempted to correct for this by dividing by a time dependant prefactor based on the fraction of money in use contributing to GDP. This had a slight effect towards detrending the velocity but the overall trend remained. This may be because the parameters and functions were not estimated accurately enough or that there are more factors at play than wealth inequality influencing the trend of the velocity.

Future work could include finding or calculating more income/wealth data for different years other than 1996 for a better estimation of parameters in the prefactor γ_t. Factors other than wealth inequality that might influence velocity such as interest rates, the current account and measures of financial sophistication could be analysed using multivariate statistics. It would also be of interest to see if the inverse relationship between wealth inequality and velocity holds in other countries or areas of the world besides the UK such as the US and the EU.

Acknowledgements

I would like to thank both my supervisors for their helpful input and discussions. In particular Stefan's suggestion and support with the prefactor and Alex's assistance with various economic questions.

References

[1] Bank of England Statistics. `https://www.bankofengland.co.uk/statistics`.

[2] Durbin Watson significance tables. `https://www3.nd.edu/~wevans1/econ30331/Durbin_Watson_tables.pdf`.

[3] Federal Reserve Bank of St. Louis. `https://fred.stlouisfed.org/`.

[4] Gini Index Data from The Guardian. `https://www.theguardian.com/politics/datablog/2013/apr/08/britain-changed-margaret-thatcher-charts`.

[5] Office for National Statistics. `https://www.ons.gov.uk/`.

[6] The scale of economic inequality in the UK. `https://www.equalitytrust.org.uk/scale-economic-inequality-uk`.

[7] United Kingdom Current Account. `https://tradingeconomics.com/united-kingdom/current-account`.

[8] (2008). The definition of a recession. *Keybridge Reasearch*. `http://keybridgedc.com/wp-content/uploads/2013/11/Key-Insight-Definition-of-Recession-Dec-2008.pdf`.

[9] (2016). The trouble with GDP. *The Economist*. `https://www.economist.com/briefing/2016/04/30/the-trouble-with-gdp`.

[10] Allen, W. A. (2017). Quantitative easing and the independence of the Bank of England. *National Institute of Economic and Social Research*.

[11] Alvaredo, F., Chancel, L., Piketty, T., Saez, E., and Zucman, G. (2018). World inequality report.

[12] Benati, L., Lucas, R. E., Nicolini, J., and Weber, W. (2017). International evidence on long-run money demand. *Minneapolis Fed Research*.

[13] Blanchard, O. (2017). *Macroeconomics*. Pearson, 7th edition.

[14] Boix, C. (2010). Origins and persistence of economic inequality. *The Annual Review of Political Science*.

[15] Bordo, M. D. and Jonung, L. (1987). *The long-run behaviour of the velocity of circulation*. Cambridge University Press.

[16] Breedon, F. and Fisher, P. (1993). The determination of M0 and M4. *Bank of England*.

[17] Caroll, C., Slacalek, J., Tokuoka, K., and White, M. (2016). The distribution of wealth and the marginal propensity to consume.

[18] Catalano, M. T., Leise, T. L., and Pfaff, T. J. (2009). Measuring resource inequality: The gini coefficient.

[19] Chatfield, C. (1999). *The Analysis of Time Series*. Chapman & Hall, 5th edition.

[20] Dickey, D. A., Jansen, D. W., and Thornton, D. L. (1991). A primer on cointegration with an application to money and income. *Federal Reserve Bank of St. Louis*.

[21] Drăgulescu, A. and Yakovenko, V. M. (2001). Exponential and power-law probability distributions of wealth and income in the united kingdom and the united states. *Elsevier Science*.

[22] Elliott, L. (2012). British recessions: a short history. *The Guardian*.

[23] Giorgi, G. M. (2011). Corrado Gini: the man and the scientist. *International Journal of Statistics*.

[24] Greene, W. H. (2012). *Econometric Analysis.* Prentice Hall, 7th edition.

[25] Higgins, B. (1978). Velocity: Money's second dimension. *Economic Review.*

[26] Humphrey, T. M. (1974). The quantity theory of money: its historical evolution and role in policy debates. *Economic Review.*

[27] Karlsen, H. A., Myklebust, T., and Tjøstheim, D. (2007). Non-parametric integration in a non-linear type model. *The Annals of Statistics.*

[28] Krugman, P. and Wells, R. (2013). *Economics.* Worth Publishers, 3rd edition.

[29] McLeay, M., Radia, A., and Thomas, R. (2014). Money creation in the modern economy. *Bank of England Quarterly Bulletin.*

[30] McTeer, B. (2013). The role of foreign trade on GDP. *Forbes Magazine.*

[31] Miles, D. (2012). Government debt and unconventional monetary policy. *Bank of England Speech.*

[32] Piketty, T. World Inequality Database. `https://wid.world/country/united-kingdom/`.

[33] Rasmussen, C. E. (2006). *Gaussian Processes for Machine Learning.* The MIT Press.

[34] Santoni, G. J. (1987). Changes in wealth and the velocity of money. *Federal Reserve Bank of St. Louis.*

[35] Yakovenko, V. M. and J. Barkley Rosser, J. (2009). Colloquium: Statistical mechanics of money, wealth, and income. *Reviews of Modern Physics.*

Appendix

A Augmented Dickey Fuller Test

The augmented Dickey-Fuller test is a hypothesis test on whether a time series has a unit root (the null hypothesis) or is stationary (alternative hypothesis).

The test is carried out with the assumption that the series is of the form:

$$x_t = \mu + \beta t + +\gamma x_{t-1} + \gamma_1 \nabla x_{t-1} + \cdots + \gamma_p \nabla x_{t-p} + z_t.$$

The test statistic

$$\mathrm{DF}_\tau = \frac{\hat{\gamma} - 1}{\mathrm{SE}(\hat{\gamma})}$$

is tested against critical values from the Dickey-Fuller t-distribution for the null hypothesis $\mathrm{H}_0 : \gamma = 0$ against the alternative $\mathrm{H}_\mathrm{A} : \gamma < 0$.

B Gaussian Process

Definition. A Gaussian process is a collection of random variables any finite number of which have a joint Gaussian distribution.

Suppose we have a set $\mathcal{X}$ of inputs and $\mathcal{Y}$ outputs and a function between them $f : \mathcal{X} \to \mathcal{Y}$. A function evaluation $f(x_i) = y_i \in \mathcal{Y}$ can be thought of as a random variable. Let us define a vector of n function evaluations as $\mathbf{f} = (f(x_1), f(x_2), \ldots, f(x_n))^T$ and suppose

$$\mathbf{f} \sim \mathcal{N}(\boldsymbol{\mu}, K)$$

where $\boldsymbol{\mu} = (\mu(x_1), \mu(x_2), \ldots, \mu(x_n))^T$ and K is an n by n covariance matrix with entries $K_{ij} = k(x_i, x_j)$. The function $\mu : \mathcal{X} \to \mathbb{R}$ is the called the mean function and the function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called the covariance or kernel function. This function f (viewed as an infinite number of function evaluations) thus satisfies the above definition of a Gaussian process and we write

$$f \sim \mathcal{GP}(\mu, k).$$

Gaussian process regression is then Bayesian inference applied to the data with a Gaussian process prior. The mean function is usually chosen as $\mathbf{0}$, however the kernel function has to be chosen appropriately for a good fit. In our case a squared exponential kernel function was chosen with parameters that gave a good R^2 value without overfitting. The prediction line is the mean of the posterior Gaussian process based on the data.

C Durbin Watson Test

The Durbin Watson test is used to test the null hypothesis of no first order autocorrelation in the residuals e_t against the alternative of first order autocorrelation from regression of two time series of size N. The test statistic is

$$d = \frac{\sum_{t=2}^{N} (e_t - e_{t-1})^2}{\sum_{t=1}^{N} e_t^2}.$$

d is then compared to two critical values at significance value α: $d_L(\alpha, k, N)$ and $d_U(\alpha, k, N)$ where k is the number of explanatory variables in the regression. If $d < d_L$ there is evidence of first order autocorrelation, if $d > d_U$ there is evidence of no first order autocorrelation and if $d_L < d < d_U$ the test is inconclusive.

D Cointegration

Checking if two time series are cointegrated is often used in time series analysis to find if they have a long-run link of ‘trending together’ and so reducing the chances that any relation between them is spurious.

Definition ((Linear) Cointegration)**.** Two time series X_t and Y_t are said to be cointegrated if they are both I(d) and if a linear combination exists of order of integration less than d [19].

For the particular case where X_t and Y_t are I(1), a usual way to test for cointegration is to check if the residuals of a linear regression of the two time series is stationary. There is also the natural extension to non-linear cointegration of two time series of the same order where a non-linear combination of a lower order exists [27].