Measuring Consumer Confidence via Google Trends

Holly Hannah Smith

Queen Mary University

Georg Von Graevenitz

Queen Mary University, CCP and CREATe

WORD COUNT: 9606

ABSTRACT

The Psychological Consumption theory informs us that consumers are expected to spend or save in relation to their confidence surrounding their expectations of their future personal finances and environment (Katona, 1975). As consumers become more sophisticated and risk adverse measuring consumer confidence accurately is becoming more difficult. We show how Google Trends data can help explain consumption in England, Scotland, Wales and Northern Ireland. Our results show that different expenditure items hold significance in explaining consumers' increases or decreases in expenditure. Importantly, the results show that Google search data is able to explain an additional 61.0% of average total expenditure in Northern Ireland, an extra 48.0% in Wales, 47.9% extra in Scotland and 34.2% more in England. This is greater than the efforts of consumer confidence indexes when considering the whole of the UK. Therefore, Google Trends data holds potential for local authorities and economists to capture a greater amount of consumers' expenditure to denote their consumer confidence, and towards gaining a greater insight into when and where regional shifts in the UK's consumption and GDP occur.

Table of Contents

1.0. Introduction and Background

Consumer confidence is an economic signal that denotes the level of assurance consumers have surrounding the economy (Van-Raaij and Gianotten, 1990; Dominitz and Manski, 2003 and Bram and Ludvigson, 1998). Several different consumer confidence indexes exist as independent surveys that collect information surrounding economic conditions, financial expectations, employment and business prospects (Garner, 1991 and Ludvigson, 2004). Countries have their own confidence surveys undertaken by specialist research, financial services and/or private companies (European Commission, 2014).

The Psychological Consumption theory informs us that consumers are expected to spend or save in relation to their confidence surrounding their expectations of their future personal finances and environment (investment prospects and credit market conditions) (Katona, 1975; Van Raaij and Gianotten, 1990 and Ven, 2011). However, the assumption that if consumers have greater availability of funds they will automatically spend more does not hold (Cadman, 2016). A leading reason for this is due to consumers being ever more sophisticated and risk adverse after the financial housing crisis in 2008 (Fornell et al, 2008). This is supported by a senior Economist at Oxford whom stated; ‘it is one thing giving the consumer that extra money, and it is another getting them to spend it’ (Cadman, 2016; 1). This presents a problem as critics recognise that consumers have different time preferences (the act of spending now, or saving in order to consume more in the future) which are continuously changing with their consumer confidence (Ven, 2011). Therefore, it is fundamental that we are able to measure consumer confidence to firstly capture changes in consumers' time preferences, and secondly to understand the catalyst factors that contribute to their decision to make alterations to spending (Acemoglu and Scott, 1994).

This is an important current economic issue that faces businesses, local authorities and governments as consumer confidence dictates consumers' expenditure and investment by companies (Van-Raaij and Gianotten, 1990; Ludvigson, 2004 and Della-Penna and Huang, 2009). The macroeconomic component consumption (C) accounts for 60% of the UK's economic welfare measured by its Gross Domestic Product (GDP). This is amongst other contributing components such as investment, government spending and net exports (Y=C+I+G+NX) (Boundless, 2016; Cadman, 2016 and Goossen, 2007). Therefore, changes in consumer spending is a vital aspect of macroeconomics (Kupelian and Hawksworth, 2013). For this reason, consumer confidence is closely monitored by government officials and economists as reduced or abrupt changes in spending has the ability to considerably stump economic growth, and decrease the value of all goods and services produced in the UK (European Commission, 2014 and Fornell et al, 2008). As a result, this could lead to negative outcomes for communities and local businesses surrounding their opportunities for development (Fornell et al, 2008).

The is because for businesses it matters how consumers distribute their money amongst different spending categories (Kupelian and Hawksworth, 2013). Possible categories being; food and drink, clothing and footwear, housing, communication, recreational and restaurants and hotels (ONS, 2012). Consumers allocation of funds to each category is likely to change as their attitudes towards spending and saving alter. For example, they are likely to spend less on clothing and footwear, recreational and restaurants and hotels in times of economic recessions rather than booms (Ven, 2011 and Kupelian and Hawksworth, 2013). This was the case for the period between 2007 and 2009 which marks the UK's financial crisis and was a time where inflation exceeded wage growth (Kupelian and Hawksworth, 2013). A recent ONS report demonstrates that UK households have not fully recovered from the financial crisis as family spending in 2015 remained below that of 2007 (ONS, 2015).

In light of this issue, exploring new methods that capture more accurate measures of consumption to denote consumer confidence is vital for policy makers, to enable them to be responsive in times of economic downfall towards diminishing the negative externalities of considerably reduced consumption (Goossen, 2007).

1.1. Added Value of Google Trends Data to Consumer Confidence Indexes

Economists and authors such as Van-Raaij and Gianotten (1990) have praised the predicting power of consumer confidence surveys in times of sharp fluctuations in economic expansion and contraction (Bram and Ludvigson, 1998 and Fornell et al, 2008). Their research has demonstrated this by showing that consumer confidence can inform us about consumers' actual and forecasted consumption levels (Bram and Ludvigson, 1998), discretionary purchases (Van-Raaij and Gianotten, 1990) and the amount of consumer debt (Fornell et al, 2008) in the UK both before and after the financial crisis. Despite this, some have shown that consumer confidence measures could predominantly be a mere summary mechanism of raw economic components in explaining consumption, and therefore its added value is poor (Garner, 1991 and Berry and Davey, 2004). This is because consumer confidence is correlated with GDP, earnings, house prices, exchange rates and the FTSE All-share (Berry and Davey, 2004).

In addition, traditional confidence surveys have been criticised for failing to recognise or only partially capturing consumers' attitudes that influence their confidence (Ludvigson, 2004). This is due to the generalisation of the questions leaving their interpretation in the hands of the participant (Ludvigson, 2004; Cadman, 2016 and Dominitz and Manski, 2003). Following this, past studies have gauged into the differences in questions asked and demonstrate that the conclusions drawn from the data is heavily dependent on the survey in question (Bram and Ludvigson, 1998). For example, both the Michigan survey and Conference Board survey in the US are widely relied upon to denote consumer confidence. However, they both give conflicting results (Bram and Ludvigson, 1998). This is because, the Michigan survey asks whether it is a 'good time to purchase major household items' with the possible responses; 'good time to buy', 'uncertain depends' and 'bad time to buy' (Michigan survey as cited in Bram and Ludvigson, 1998). In comparison, the Conference Board survey asks how consumers would rate 'present general business conditions' with the answers 'good', 'normal' and 'bad' in their area (Conference Board survey as cited in Bram and Ludvigson, 1998).

Furthermore, each of these questions surrounding present conditions of the economy account for 20% of each overall index, and due to the differences the indexes peak at different times and by varying amounts (Bram and Ludvigson, 1998). In addition, the answers available for consumers to select from are criticised for being too general and vague (Dominitz and Manski, 2003). This is further evidence to highlight the inconsistences in trying to accurately explain the picture of consumer confidence at any one time (Dominitz and Manski, 2003). We will overcome this current issue by capturing consumers' Google search data on keywords from Google Trends. This data is collected from consumers themselves to eliminate the possibility of survey questions introducing bias (McLaren and Shanbhogue, 2011).

Previous research has restricted its insights of consumer confidence to the country level and whilst we argue that this is too general, we must also recognise the difficulty in obtaining regional data regarding consumer confidence (Ven, 2011; Acemoglu and Scott, 1994; Van-Raaij and Gianotten, 1990 and Fornell et al, 2008). However, using regional data will help to make more realistic and meaningful suggestions (Dominitz and Manski, 2003). This is because localised events such as business closures and relocation of jobs have varying and contrasting impacts on different regions of the UK, therefore looking at the UK as a whole is less valuable in directing government policies to stimulate the economy (Kirkup, 2008; Wray, 2003; Wearden et al, 2009; LSE, 2011; Dominitz and Manski 2003 and Martin, 2012). Dominitz and Manski (2003) amongst few others have recognised that looking at different sub groups of a population would add value to the research field, and be essential in helping local authorities become flexible to each of their regions' needs by providing a greater understanding of what and how these events effect their consumer confidence levels. Della-Penna and Huang (2009) looked at 49 US states, Canadian proveniences and advanced economies and found that Google searches help predict consumers' willingness to spend better than standard confidence indexes by an extra 9%. Despite this, their paper does not discuss differences in US states or Canadian proveniences (Della-Penna and Huang, 2009). We will contribute to the undeveloped regional literature surrounding consumer confidence by exploring consumption in the regions of the UK; England, Scotland, Wales and Northern Ireland.

A further weakness of traditional consumer confidence indexes is the time delay between collecting consumers' information and publishing the results (Van-Raaij and Gianotten, 1990; Della-Penna and Huang, 2009; Choi and Varian, 2011 and McLaren and Shanbhogue, 2011). Consumer confidence is typically measured each month and published monthly, quarterly and/or annually depending on the source (Van-Raaij and Gianotten, 1990 and European Commission, 2014). This is a limitation of consumer confidence indexes as macroeconomic variables and consumers' attitudes are constantly changing with financial markets and everyday life events. Therefore, consumer confidence data can become quickly outdated (Ludvigson, 2004). We will eliminate this weakness by using Google Trends data which is provided on a weekly, monthly and annually basis, and therefore can now-cast (predict the present) because of its time advantage in providing data, unlike survey based indexes (Della-Penna and Huang, 2009; Chamberlin, 2010 and McLaren and Shanbhogue, 2011). For this reason, consumer confidence data informed by Google Trends may not only improve our ability to explain consumers' consumption better than traditional surveys, but it could also be a fundamental tool in observing early warning signs for changes in confidence of consumers via their consumption responses (Della-Penna -Penna and Huang, 2009 and McLaren and Shanbhogue, 2011).

Whilst the research into consumer confidence is wide, the research thus far exploring Google searches as an indicator of consumer consumption and confidence is small but growing. Not only have many studies already discovered that Google searches are statistically significant in predicting consumer spending (sometimes at the 1% level), but also that they outperform traditional consumer confidence surveys (Della-Penna and Huang, 2009; Kholodilin et al, 2009; Choi and Varian, 2011; McLaren and Shanbhogue, 2011 and Chamberlin, 2010). By including Google searches in regressions studies have shown the R-squared value to increase by as much as 20% depending on the independent variables (Della-Penna and Huang, 2009; Kholodilin et al, 2009 and Choi and Varian, 2011). For instance, Choi and Varian (2011) show that they are able to explain a further 9.3% of consumer confidence in Australia compared to standard surveys by adding Google Trends data. In addition, Kholodilin et al (2009) found that Google statistics improves the explanation of growth in private consumption in the United States of America between 2008 and 2009 by 20% more than consumer indexes. Here, the time period over which this relationship is observed is short. However, it shows despite the fact the financial crisis ‘played havoc with consumer spending’ (Fornell et al, 2008; 3) policy makers would have been able to explain a greater proportion of consumer spending more quickly had they used Google Trends.

Many are confident that this additional explanatory power that Google Trends has is because consumers may act differently to how they claim to feel (McLaren and Shanbhogue, 2011). Therefore, consumer confidence surveys are limited to only capturing the latter (McLaren and Shanbhogue, 2011). Whereas, Google searches are able to encompass contributions from impulse purchases, consumer satisfaction and consumers' spending habits (Fornell et al, 2008 and Fornell and Stephan, 2002), and therefore incorporate consumers' private information (Acemoglu and Scott, 1994).

This poses opportunities to build on past research and highlights a relatively new area of research in which we will contribute to. The purpose of this study is twofold, firstly to investigate whether Google Trends data can better explain consumer consumption compared to a traditional consumer confidence indexes in the UK. In addition, to see whether Google Search data can explain average household expenditure regionally in the UK between 2004 and 2014.

The structure of our paper is as follows: firstly, we will introduce the aggregate and regional data. Secondly, we present the methods we will use. For both datasets, this will be simple regression models, in the aggregate data we will replicate a similar model to that in the existing literature. Thirdly, we will present our results. This will show the value consumer confidence indexes offer in explaining consumer consumption. In addition, we show how Google Trends data can help explain consumption in England, Scotland, Wales and Northern Ireland. This analysis adds to the literature by demonstrating that consumer confidence is not fixed across different regions of the UK. Our results show that different expenditure items hold significance in explaining consumers' increases or decreases in expenditure. Importantly, the results demonstrate that Google Trends data is able to explain an additional 61% of average total expenditure in Northern Ireland, an extra 48% in Wales, 47.9% extra in Scotland and 34.2% more in England. This is greater than the efforts of consumer confidence indexes when considering the whole of the UK. Therefore, our results show that Google Trends data holds potential for local authorities and economists to capture a greater amount of consumers' expenditure to denote their consumer confidence. In turn, our results are promising towards gaining a greater insight into when and where shifts in the UK's consumption component of GDP occur.

2.0. Data

We construct two datasets: the first contains aggregate data on consumption, labour income and other macroeconomic aggregates; the second contains regional data from the Living Costs and Food Survey (LCFS)[1] and regional data from Google Trends.

2.1. Aggregate Data

We obtain quarterly data on aggregate domestic consumption at current prices from ONS, monthly data on the LIBOR rate, UK consumer price inflation (CPI), the Chicago Board of Exchange volatility index (VIX), and OECD's weekly earnings index covering all activities in the UK from the Federal Reserve Bank of St Louis (Fred, 2017 and ONS, 2015). All data we use is not seasonally adjusted. We supplement this with monthly data from GfK and MORI on consumer confidence (GfK, 2017 and Ipsos MORI, 2016). Due to the lack on monthly data on UK consumption we aggregate the data to quarterly frequency and differenced against the same quarter of the previous year to ensure the data is stationary. The aggregate dataset variables are outlined below.

Consumption

The UK's household final aggregate consumption expenditure in millions of pounds is extracted from the Office for National Statistics

[1] Before 2008 we use data from the Expenditure and Food Survey.

(ONS) for 2004 to 2014. The data includes the total consumption expenditure worldwide of UK households, this is arrived at by adding imports and removing exports from the UK’s domestic concept data (ONS, 2014). This data uses chain linking whereby changes in the structure of the economy and prices on an annual basis are accounted for from year to year (ONS, 2014).

Consumer Confidence Index: GfK on behalf of the European Commission

We will use two of the UK’s main consumer confidence indexes and that are commonly used in the literature (Berry and Davey, 2004). The first consumer confidence index used is sourced from Gfk on behalf of the European Commission, which is home to consumer confidence and economic sentiment information for 33 countries. Gfk conduct the research via ‘harmonised questionaries’ per the methodology and calendar dates set out by the European Commission (European Commission, 2014, 5). The survey asks 15 questions in total; 12 monthly and 3 on a quarterly basis (January, April, July and October) (European Commission, 2014). These questions are asked in the 1st and 3rd weeks of the month. We will use consumer confidence extracted from 2004 to 2014 monthly for the UK. The final confidence indicator is calculated as a simple average of the questions shown in table 1 (Berry and Davey, 2004).

Table 1: Questions asked in obtaining the GfK Consumer Confidence Index

1. How has the financial situation of your household changed over the last twelve months? (Possible answers: ‘no change’ and ‘a little’ or ‘a lot’ better or worse)

2. How do you expect the financial position of your household to change over the next twelve months? (Possible answers: ‘no change’ and ‘a little’ or ‘a lot’ better or worse)

3. How do you think the general economic situation in this country has changed over the last twelve months? (Possible answers: ‘no change’ and ‘a little’ or ‘a lot’ better or worse)

4. How do you expect the general economic situation in this country to develop over the next twelve months? (possible answers: ‘no change’ and ‘a little’ or ‘a lot’ better or worse)

5. In view of the general economic situation, do you think now is the right time for people to make major purchases such as furniture or electrical goods? (Possible answers: ‘yes’, ‘no’ or ‘evenly balanced’)

In the UK 2000 surveys are conducted randomly per month amongst people aged 16 and above[2]. However, the response rate of the surveys is 5.9%, and whilst the consumers that do not respond are not used again, the number of actual surveys obtained decreases to 118 (European Commission, 2014). One explanation for this may be the increased reluctance of consumers to disclose highly sensitive information about themselves. This further highlights the need for policy makers to find an alternative method because the low response rate indicates that consumer confidence surveys can only ever be used as an insight rather than to make generalisations to act upon (Dominitz and Manski, 2003). Although the low response rate indicates a likelihood of bias because the position and attitude of non-respondents may contrast with those people who participated, the European Commission body deems the results representative and reliable. This is because they conduct a stratified random sampling method, whereby they group the UK population into non-overlapping sub populations. After this, a stratification criteria involving gender, age, social class and region is used to ensure the results are representative (European Commission, 2014). This is achieved by weighting coefficients to reflect the relative significance of each sub population in the UK (European Commission, 2014).

[2] Regression is still independent because the survey samples different people each year.

Consumer Confidence Index: MORI

The Ipsos MORI monthly consumer confidence survey is conducted by interviewing a sample of 1,124 adults in the UK (Ipsos MORI, 2016). The responses are then weighted in order to represent the population (Ipsos MORI, 2016). Table 2 shows the questions asked in the MORI index (Ipsos MORI, 2016). Past literature has shown that the MORI index is the most useful indicator in the US when predicting growth in durable consumption (Easaw et al, 2005).

Table 2: Questions asked in obtaining the MORI Consumer Confidence Index

1. Looking ahead six months from now, do you expect your personal financial situation to be much stronger, somewhat stronger, about the same, somewhat weaker, or much weaker than it is now?

2. Still thinking about your personal financial situation over the next six months. How concerned, if at all, are you about each of the following over the next six months? Your job prospects? Your ability to pay the bills?

3. And over the next six months, do you expect to increase the amount of your income you put into savings, reduce the amount you put into savings, or will it not change?

2.1.1. Aggregate Descriptive Statistics

Table 3 provides the summary statistics for the UK aggregate data. All figures are reported on a quarterly basis. The Mori consumer confidence has a greater range than the GfK by 45, showing greater variation.

Table 3: Descriptive Statistics of Aggregate Data in UK (1997-2015)

VARIABLES	N	Mean	St. Dev	Min	Max
Aggregate Consumption / £ Millions	79.00	219,652.73	44,140.71	140,099.00	300,857.00
CPI	79.00	83.54	10.64	69.37	100.87
LIBOR	79.00	3.56	2.46	0.43	7.67
VIX	79.00	21.03	7.70	11.03	58.60
Earnings Index	73.00	0.87	0.15	0.61	1.11
GfK Consumer Confidence	80.00	-8.27	12.01	-35.67	8.50
Mori Consumer Confidence	76.00	-16.11	19.49	-60.33	29.67

2.2. Regional Data

We obtained data on regional expenditure, which is collected monthly and is added together and published annually (ONS, 2012). In addition, we use monthly search data from Google Trends (Google Trends, 2017). The regional dataset variables are outlined below.

The Living Costs and Food Survey

The ONS's LCFS collects information surrounding regular expenditure items such as mortgage payments and larger infrequent expenditures like vehicles, and can be obtained from the UK Data Service (ONS, 2012 and Ministry of Agriculture, Fisheries and Food, Office for National Statistics, 2004). This data is heavily relied upon by the Statistical Office of the European Communities (EUROSTAT), Her Majesty's Revenue and Customs (HMRC) and the Department for Transport (DfT) (ONS, 2012). This indicates its high validity and reliability.

The LCFS is conducted via interviews and a 2-week food diary each month of the year (ONS, 2012). An average of 11,484 households are selected at random using the post office's database of all delivery points (ONS, 2012). The sample for Great Britain is a multi-stage stratified random sample using clustering and Northern Ireland's database is organised by district council, the sample is systematic and stratified geographically (ONS, 2012). The average response rate of the LCFS is 52% from 2004 to 2014, although ONS recognise that the response rate is declining (ONS, 2014). This is because it is becomingly harder for interviewers to obtain an interview with each member of the household. To overcome this, ONS allow for a family member of the same household to provide information on behalf of another as a proxy, this enhances sample size and the precision of the results (ONS, 2014). Undoubtedly, there is a risk that the proxy may not provide exactly the same information, however ONS show how this enables them to capture higher earning households, and therefore fully represent the UK in the survey to improve the quality of the data (ONS, 2013).

Google Trends

Google was ranked the most popular search engine in the UK in October 2016 and firmly holds the largest market share of 86.94% amongst other search engines (Statista, 2017). Unlike its competitors, Google's 3.5 billion daily searches provides an enormous pool of data across an array of categories (Internet Live Stats, 2017). Google Trends generates data of keyword searches for a given time in the range of 0 to 100; this is calculated by dividing the total daily number of searches for the keyword by the entire number of searches for that day in the same location (Choi and Varian, 2011 and Google, 2004). Following this, Google categorises the search terms by connecting them to a topic and erases personal information (Google, 2004). End users of Google Trends are able to find the search activity of a keyword in a country or region over a chosen time period as early as 2004 (Google, 2004). They are then able to download this data for multiple keywords to enable comparisons to be made between terms. Due to this, studies using Google search data have complimented its timeliness, convenience and flexibility to meet users' needs

(Graevenitz et al, 2016; Goel et al, 2010 and McLaren and Shanbhogue, 2011). In addition, they have highlighted it's high persistence over time (Google, 2004; Graevenitz et al, 2016 and Goel et al, 2010).

However, we must also recognise the weaknesses of using Google Trends. Firstly, Google Trends (2004) removes duplicate searches made in a narrow time frame from the same person. Past literature discusses this as a disadvantage because it interferes with the end data that is used to predict consumer consumption (Askitas and Zimmermann, 2009). Despite this, it could be argued that removing duplicate searches can be seen as an advantage (Google, 2004). This is because Google filters out consumers repeated searches that often come with making purchase decisions or seeking after care services for a particular product (Choi and Varian, 2011). This diminishes but does not eliminate the problem of high search volume for a product when in fact unit sales may be lower (Graevenitz et al, 2016).

Furthermore, keywords that are searched by too few people generate an overall search volume of zero for that particular region (Google, 2004 and Choi and Varian, 2011). This is an issue we have had to work around and is a limitation for all papers in this field. For example, this is the case when searching for the descriptions of vegetables in Northern Ireland in table 4 as they are too specific. In addition, the search term 'soft drink' was not recognised, therefore to overcome this the trending keywords suggestions from Google Trends; ‘coca cola’ and ‘lemonade’ will be used. In addition, as consumers’ behaviour and trends change so may the keywords they type into Google to generate the same search results. Therefore, over time the group of keywords are not fixed and should be reconsidered (Graevenitz et al, 2016).

Despite these weaknesses, studies have shown that Google searches reflect demand, purchasing power and consumer confidence because consumers use Google to research into items they want to buy and that they care about (Della-Penna and Huang, 2009 and McLaren and Shanbhogue, 2011). We will capture these motivations by using keywords surrounding necessities and luxury items that act as proxies for consumer consumption. Some of the keywords represents basic non-durable items in the consumers’ consumption basket, and therefore they are purchased by most households and seen as a stable product (Van-Raaij and Gianotten, 1990). Others are widely regarded as luxury items that consumers would forgo if the economy was contracting because instead of spending they would be more inclined to save (Van-Raaij and Gianotten, 1990; Ven, 2011 and Ludvigson, 2004). We will use the keyword ‘board games’ which may seem outdated compared to ‘video games’ in this technology advanced age, however the ONS (2015) have reported that households still spend on ‘more traditional pastimes’ (Monaghan, 2017, 1). Keywords ‘restaurants’, ‘pub’ and ‘beer’ may be seen as luxury items that are important to capture as households in the UK are increasingly spending their money on dining experiences (Monaghan, 2017). Therefore, luxury goods are significant in explaining consumer confidence (Della-Penna and Huang, 2009).

Selecting Google Trends Keywords

Due to the ‘young but promising’ research surrounding Google Trends and consumer confidence there is no one preferred or popular procedure when it comes to selecting keywords, with many papers picking keywords based on their common sense and instinct (Kholodilin, 2009, 1; Askitas and Zimmermann, 2009 and McLaren and Shanbhogue, 2011). Ordinary Least Squares (OLS) is incompatible with our dataset because there are more variables than observations. Therefore, we use a relatively new variable selection technique called Double Lasso (least absolute shrinkage and selection operator). This method identifies a smaller subset of independent variables that most explain consumers’ expenditure from over 2000 independent expenditure variables in the LCFS dataset (Belloni et al, 2014; Urminsky et al, 2016 and Tibshirani, 1996). Lasso systematically adds these variables one by one into multiple regressions. This method drops variables if they are insignificant and adds new ones until a set of significant variables are established (Tibshirani, 1996). This will ensure that strong predictors of expenditure are included in the analysis to reduce errors and improve the statistical power (Urminsky et al, 2016).

The expenditure and regional variables identified by the lasso shooting were taken from the LCFS's household characteristics and expenditure files[3]. Data from these files are collected at the household level. The LCFS defines a household as comprising of 'one person living alone or a group of people (not necessarily related, living at the same address) who share cooking facilities and share a living room' (ONS 2014;13). Using household data may be regarded as a limitation to exploring consumer consumption. However, we will reasonably assume that the spending behaviour between family members living in the same household would be consistent as if one member is trying to save it is highly unlikely that another will be extravagantly spending.

The LCFS sorts data into 12 broad expenditure categories and each individual item has a unique coding. The codes generated from Lasso that most explain consumer expenditure covers 6 of these categories, they are; food and non-alcoholic beverages (C1), clothing and footwear (C3), transport costs (C7), recreational (C9), restaurants and hotels (CB) and miscellaneous goods and services (CC) (ONS, 2015). Therefore, our keywords encompass a reasonable range of items from necessities to luxuries. It is important to note that Google

[3] File name: dvhh_ukanon

differentiates between the keyword as a topic or a search term. We enter each of the keywords in table 4 as search terms to obtain the data in its rawest form rather than categorised by Google. Table 4 shows the codes generated from Lasso as the most explanatory of consumer consumption, and therefore are used.

Table 4: LCFS Categories and Google Search Keywords

Category	Google Search Keywords[4]
Food and non-alcoholic beverages, (C1)	
C11121t Bread/C11151t Other breads[5]	Bread, Toast
C11241t Poultry	Poultry, Chicken and Duck
C11251t Sausages	Pork and Sausages
C11711t Leaf and stem vegetables (fresh or chilled) /C11731t Vegetables grown for their fruit (fresh, chilled or frozen) /C11741t Root crops, non-starchy bulbs and mushrooms (fresh or frozen) /C11761t Other preserved or processed vegetables	Vegetable, Tomato and Potato
C11911t Sauces, condiments	Ketchup and Mayonnaise
C12221t Soft drinks	Coca Cola and Lemonade
C12231t Fruit juices	Orange Juice and Juice
Clothing and Footwear, (C3)	
C31221t Women’s outer garments	Women’s Coats and Women’s Jackets

[4] Each Google keyword is searched for separately and combine in Stata

[5] Where we had multiple expenditure codes for the same broad item, they were aggregated together in a Google index.

Transport costs, (C7)	
C72211t Petrol	Petrol
C72212t Diesel oil	Diesel
Recreational, (C9)	
C93111t Games, toys and hobbies	Games, Toys, Video Games and Board Games
Restaurants and hotels, (CB)	
CB1111t Catered food non-alcoholic drink eaten / drunk on premises	Restaurants
CB1115t Hot food eaten on premises	Pub
CB111It Beer and lager (away from home)	Beer
CB1127t Hot take away meal eaten at home	Takeaway, Chinese Takeaway and Pizza Delivery
Miscellaneous goods and services, (CC)	
CC1312t Toiletries (disposables - tampons, lip balm, toothpaste, etc.)	Toiletries and Soap
CC1316t Cosmetics and related accessories	Cosmetics and Lipstick
Regional Variables	
‘gorx’ (government office regions)	England, Scotland, Wales and Northern Ireland

The regional variables in the LCFS allow for the above expenditures to be broken down. The regional variables and expenditure codes are kept in the file along with the ‘case’, ‘year’ variables, and the ‘a055’ variable which codes for the sampling month[6]. The median and mean of each expenditure code were generated for each region and month. The datasets for all years were then appended and a real date variable created.

2.2.1. Regional Descriptive Statistics

Below table 5 shows the descriptive statistics for the regional data.

Table 5: Descriptive Statistics Regional Data UK (2004-2014)

VARIABLES	N	Mean	St. Dev	Min	Max
Average LCFS Expenditure: Bread	527.00	2.73	0.62	0.50	5.39
Average LCFS Expenditure: Other Breads	527.00	2.00	0.49	0.31	4.49
Average LCFS Expenditure: Poultry	527.00	2.02	0.79	0.00	6.62
Average LCFS Expenditure: Sausages	527.00	0.80	0.31	0.00	2.44
Average LCFS Expenditure: Leaf and Stem Vegetables	527.00	0.67	0.26	0.00	1.76
Average LCFS Expenditure: Vegetables grown for their fruit	527.00	1.14	0.37	0.00	3.01
Average LCFS Expenditure: Root crops	527.00	1.24	0.32	0.49	2.65
Average LCFS Expenditure: Other preserved or processed vegetables	527.00	1.17	0.32	0.00	3.41

[6] In the monthly variable additional monthly data with the name ‘reis’ were dropped as they were inconsistent amongst years and months.

Average LCFS Expenditure: Sauces and Condiments	527.00	1.13	0.32	0.18	2.57
Average LCFS Expenditure: Soft drinks	527.00	1.95	0.66	0.00	4.80
Average LCFS Expenditure: Fruit juices	527.00	1.06	0.33	0.00	2.82
Average LCFS Expenditure: Women’s outer garments	527.00	9.17	5.08	0.00	49.81
Average LCFS Expenditure: Petrol	527.00	15.04	4.89	0.00	55.00
Average LCFS Expenditure: Diesel oil	527.00	7.55	5.56	0.00	39.94
Average LCFS Expenditure: Games, toys and hobbies	527.00	1.95	2.26	0.00	34.53
Average LCFS Expenditure: Beer and lager away from home	527.00	3.49	1.73	0.00	11.01
Average LCFS Expenditure: Catered food and drink on the premises	527.00	13.57	4.48	0.43	39.08
Average LCFS Expenditure: Hot food eaten on premises	527.00	1.05	0.66	0.00	8.24
Average LCFS Expenditure: Cold food	527.00	0.94	0.49	0.00	5.72
Average LCFS Expenditure: Hot take away meal eaten at home	527.00	4.33	1.96	0.00	15.95
Average LCFS Expenditure: Toiletries	527.00	1.37	0.61	0.14	4.86
Average LCFS Expenditure: Cosmetics	527.00	2.65	1.78	0.00	16.83
Average LCFS Total Expenditure	527.00	383.14	70.77	106.65	903.68
Google Search: Bread and Toast	524.00	45.46	13.49	14.50	88.50
Google Search: Poultry, Chicken and Duck	524.00	15.90	13.88	-19.00	48.67
Google Search: Pork and Sausages	524.00	35.50	16.26	0.00	81.50
Google Search: Vegetable, Tomato and Potato	524.00	43.23	13.02	14.67	79.00
Google Search: Ketchup and Mayonnaise	524.00	29.70	19.45	-33.33	87.67
Google Search: Coca Cola and Lemonade	524.00	13.05	9.40	0.00	78.00
Google Search: Orange Juice and Juice	524.00	27.81	27.39	-45.50	87.00
Google Search: Women’s Coats and Women’s Jackets	524.00	14.45	12.46	0.00	63.00
Google Search: Petrol	524.00	48.97	15.91	10.00	100.00
Google Search: Diesel Oil	524.00	63.54	15.63	13.00	100.00
Google Search: Games, Toys, Video Games and Board Games	524.00	44.04	13.39	17.25	89.25
Google Search: Beer	524.00	61.16	12.51	21.00	100.00
Google Search: Restaurants	524.00	51.21	15.01	13.00	100.00
Google Search: Pub	524.00	62.25	13.82	13.00	100.00
Google Search: Takeaway, Chinese Takeaway and Pizza Delivery	524.00	34.98	15.97	4.33	89.00
Google Search: Toiletries and Soap	524.00	27.24	29.07	-46.50	91.50
Google Search: Cosmetics and Lipstick	524.00	32.04	7.94	10.00	71.00

The average expenditure is quoted on a weekly basis. Table 5 shows that the highest average weekly expenditure items in the food and non-alcoholic beverages category in the UK are bread and poultry. Whereas, leaf and stem vegetables are the lowest expenditure. Also, in the restaurant and hotel category households on average are more likely to spend on takeaway food at home rather than going out for beer or lager. Despite this, the second highest average expenditure other than petrol is catered food and drink on the premises. On average a household spends £15.91 a week on items in the food and non-alcoholic beverages category, this is only £2.34 more than their weekly average expenditure on catered food and drink on the premises of food establishments. The minimum and maximum values show the range of average expenditure our data captures. The minimum value of £0 shows on several items that some households choose not to purchase them items. Reasons for this could be consumers' dietary requirements, choosing to cycle or use public transport as oppose to car and the general composition of the households asked. For example, a household of all males may not spend their money on women's outer garments. Overall, the average total expenditure captures a reasonable range for a weekly basis with a difference of £797.03.

As discussed previously the nature of Google Trends data means that table 5 does not show absolute figures. Instead they are relative measures to enable comparison. Some of the minimum values in the Google search data generate negative figures, this is due to the transformation of the data when using the Cronbach's Alpha. This is a measure of internal consistency (Tavakol and Dennick, 2011). The expenditure items in the food and non-alcoholic beverages category in the UK with the highest search scores are; bread and toast, vegetable, tomato and potato and pork and sausages. This is interesting as the average LCFS expenditure shows that households spend the least on vegetables and sausages. In addition, on average games, toys, video games and board games are searched for more than takeaways. Women's jackets and women's coats along with Coca Cola and lemonade were searched for the least on average.

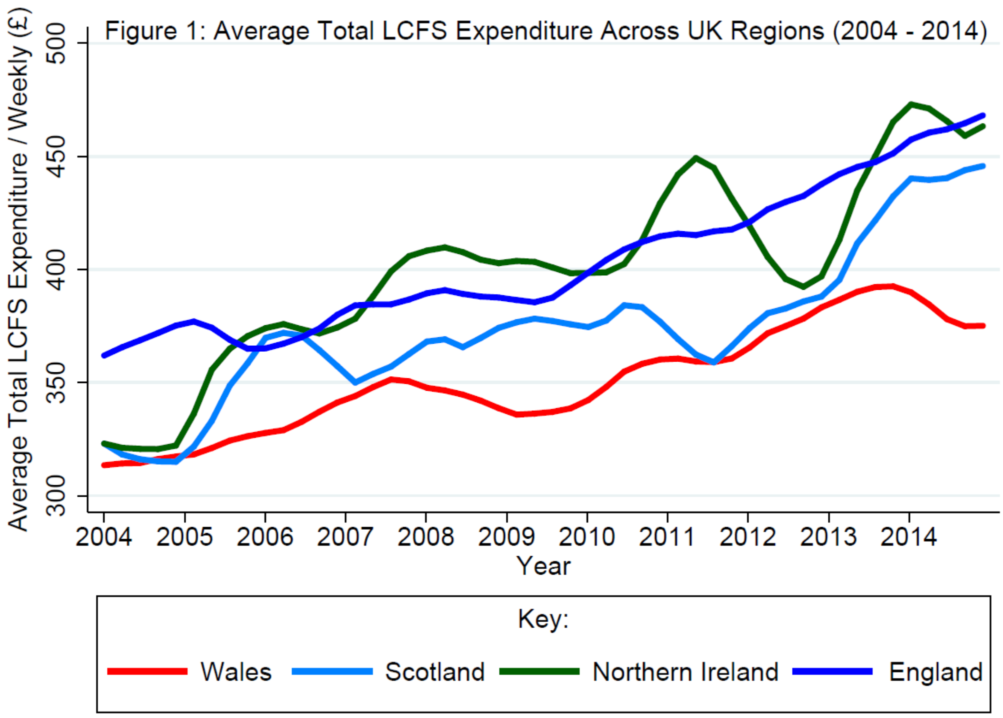


Figure 1: Average Total LCFS Expenditure Across UK Regions (2004 - 2014)

Figure 1 shows the average total LCFS expenditure for all four regions of the UK. Overall, across all regions there is a general upwards trend of weekly total expenditure between 2004 and 2014. In 2004 total expenditure in Wales, Scotland and Northern Ireland start at roughly the same point below England. However, Northern Ireland's average total expenditure shows a lot of variation and increases in a similar steepness with England. Therefore, by 2014 Northern Ireland has a similar weekly average total expenditure to England. Meanwhile, the gap of total expenditure between Wales and Scotland widens in 2014. Between 2007 and 2010 the weekly total expenditure for England levels off compared to other years, whereas, in Northern Ireland and Wales it decreases. However, in Wales over the same period the total expenditure dips in 2007 and then begins to increase. Interestingly, the total expenditure for both Northern Ireland and Scotland move together between 2004 and 2014, however this is with the exception of a two-year period between mid-2010 and mid-2012, whereby they move in opposite directions. Wales shows to have the lowest weekly average total expenditure amongst all the regions between the entire observed period.

3.0. Methods

3.1. Aggregate Methods

We will work from the following baseline regression equation (1) at the aggregate level to demonstrate that by adding consumer confidence indexes (Gfk and MORI) into a model with aggregate consumption, the adjusted R-squared increases. This will replicate the existing literature. As previously, Bram and Ludvigson (1998) used a very similar model to show that adding the Conference Board's confidence index to the baseline equation predicts a further 9% of the next periods consumption growth. In addition, Acemoglu and Scott (1994) demonstrates that by lagging confidence by one or two periods, consumption growth can be predicted.

$$\text{Aggregate Consumption} = \text{CPI} + \text{LIBOR} + \text{Exchange Volatility Index (VIX)} + \text{Earnings Index} + \text{Gfk} + \text{MORI} + Y^{*7} + Q^{*} + \mu \quad (1)$$

Year dummies were included to take out the averages for the year, and quarterly dummies were used to take out seasonality. A simple testing down method whereby statistical power is enhanced by excluding insignificant lags and variables will be used for all models in this study (Mantel, 1970). The testing down method has been challenged in the literature as it can cause problems later on if more data is added

[7] The * shows that all possible yearly, quarterly and/or monthly dummies are used.

(Sribney, 1996). This is because as variables and lags are dropped the coefficients of all the other variables change. Therefore, Elastic Net a variable selection method could be used next time which treats all variables as whole groups and addresses the weakness of testing down (Zou and Hastie, 2005 and Caner and Zhang, 2014). However, due to time constraints and the purpose of our study, a testing down method will be used throughout.

For all variables in this study we take the difference against the year before and then take the difference of that against the previous period as shown in equation (2) (DS4). By transforming our data in this way, we eliminate seasonality effects, which reduces the need for quarters in the regressions as controls. Had we ignored this issue our R-squared values would have been inflated due to overstated t-tests.

$$(C_t - C_{t-4}) - (C_{t-1} - C_{t-5}) \tag{2}$$

By doing this, we also address our problem of our data being non-stationary. This is because our data is subject to sudden shifts in the mean and variances of the data, unlike stationary data (Hendry and Pretis, 2016). These shifts may be caused by evolution, policy changes and more specifically to our data, the financial crisis in 2008 (Hendry and Pretis, 2016). As a result, findings may show ‘nonsense correlations’ (Hendry and Pretis, 2016; 11). We check for stationarity in our data using Stata via the following unit root tests; Phillips Perron (pperron), Dickey-Fuller (dfgls) and Phillips, Schmidt and Shin test (kpss) (appendix A). The first two tests indicate we can reject the null hypothesis that the data is non-stationary for all variables (Kwiatkowski et al, 1992). The third test supports this as we cannot reject the null hypothesis that the data is stationary (Schwert, 1989). The above tests all allow for trends in the data. The fact that authors in the previous literature ignore these unit root tests leads us to consider their R-squared figures critically for heightened results.

Next, in equation (3) we run a regression to show that aggregate consumption is correlated with the average total expenditure from the LCFS which we will go on to use to generate regional results.

$$\text{Aggregate Consumption} = \text{LCFS Expenditure} \qquad (3)$$

3.2. Regional Methods

The second purpose and the main contribution of our paper is to show that Google search data can help to predict regional average expenditure levels from the LCFS. To do this, we use a fixed effects baseline model as shown in equation (4), which will allow us to show how much our Google index can add to explaining consumption for each region of the UK in equation (5).

$$\text{Total average LCFS Expenditure} = M^* + Y^* + R^* \mu \qquad (4)$$

$$\text{Total average LCFS Expenditure} = \text{Google Search Index} + M^{*} + Y^{*} + \mu \qquad (5)$$

A weakness of our paper is that there is not as many consumers searching on Google in Wales and Northern Ireland compared to England and Scotland, which creates more noise in our dataset. However, this is a weakness that every paper using Google Trends will have. As a result, we will add dummies to control for this.

Furthermore, it is important to recognise that the limitations of Google Trends as discussed previously will be built into this study. Firstly, we do not know the exact methods Google Trends use to transform their raw data into the final values that are extracted. When they alter these methods and how much they effect the final data is also unknown. For example, it is important to consider what no other paper using Google Trends have; that Google Trends informs users of an improvement in its location data collection post 2011 (Google, 2004 and Martin, 2011). In order to control for this a year dummy variable will be used. Other weaknesses of our study are that the results are not generalisable outside of the scope of the UK data. However, it can be replicated for the 30 other countries featured on Google Trends with their own country's consumer confidence index. In addition, the limited degrees of freedom have restricted our regression models, therefore a dataset with larger observations would address this. Finally, the nature of our data is non-stationary and the weaknesses of this as discussed previously can only be transformed.

4.0. Empirical Results

4.1. Aggregate Results

Table 6 shows the aggregate results from equation (1). Including consumer confidence indexes in regression (2) generates an adjusted R-squared of 47.8%. This shows that with our inputs around 48.0% of consumption in the UK can be explained. Although not all variables and lags are significant, lag 2 of the MORI consumer confidence and GfK lag 1 are significant at the 1% level, and GfK lag 2 is significant at the 5% level. Most importantly, the R-squared decreases to 29.2% when running the same regression (1) without both consumer confidence indexes. This demonstrates that consumer confidence is important in explaining around 18.0% of consumption. However, even when consumer confidence indexes are included into the model there is still a remaining 52.2% of total consumption left unexplained. Therefore, there is room for improvement.

Comparing regression (1) without consumer confidence to regression (2) with consumer confidence, the results show that when consumer confidence indexes are included in the model the lags of the earnings index are no longer significant at the 1% and 10% level. In addition, lag 2 of CPI becomes less significant. This demonstrates the combined effect whereby the earnings index is better explained by the other variables in the model when combine in a regression, and therefore is no longer significant in explaining the dependent variable, total aggregate consumption in the UK.

Table 6: Aggregate Results / £ (1999-2015)

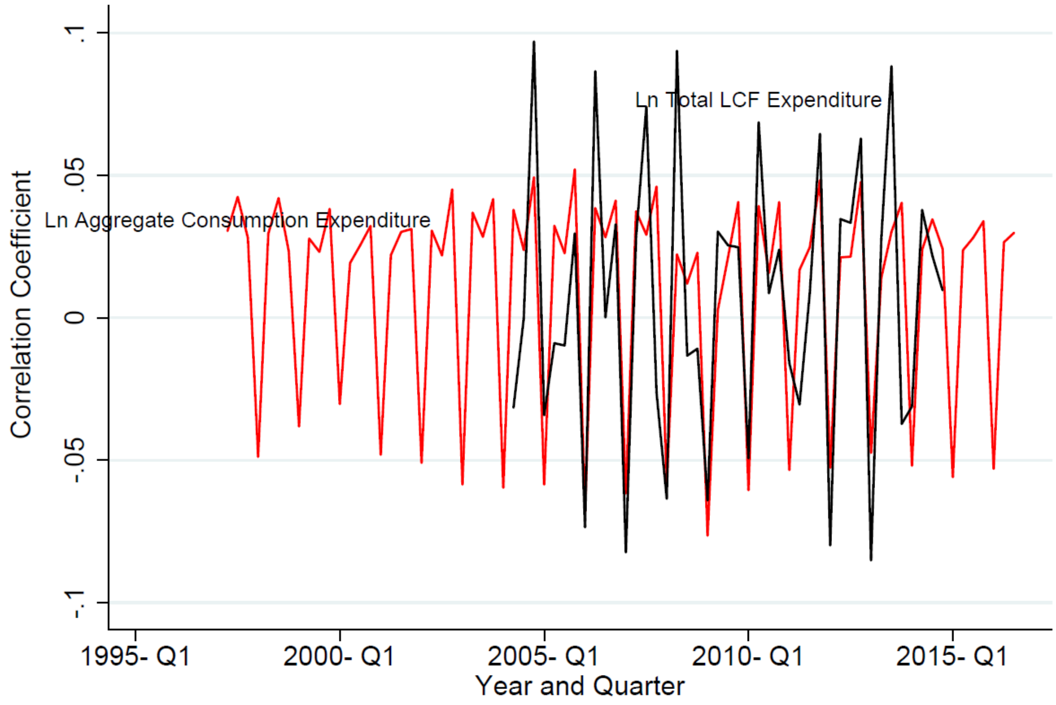
VARIABLES	(1) Total Aggregate Consumption UK	(2) Total Aggregate Consumption UK
Lag 1 Earnings Index	58,019***	
	(21,617)	
Lag 2 Earnings Index	39,743*	
	(21,816)	
Lag 1 LIBOR	1,110***	1,540***
	(402.7)	(385.4)
Lag 2 CPI	-1,916***	-1,623**
	(626.3)	(649.0)
Lag 3 CPI		1,377*
		(699.8)
Lag 2 VIX		-71.17**
		(30.93)
Lag 2 MORI CCI		-77.06***
		(22.92)
Lag 1 GfK CCI		164.6***
		(41.45)
Lag 2 GfK CCI		168.5**
		(65.02)
Constant	41.96	-129.8
	(247.4)	(221.5)
Observations	66	62
R-squared	0.335	0.538
Adjusted R-squared	**0.292**	**0.478**

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Figure 2 shows the results from equation (3). Correlation of the log of both variables is 0.93, and this decreases to 0.34 when the DS4 of the variables is used. This decrease is to be expected as the seasonality has been taken out. The logs of the variables are presented in figure 2 as it demonstrates more clearly than the DS4 variables that the troughs and peaks of both variables occur at the same times, and as discussed above shows strong yearly seasonality. The log of total LCFS expenditure is overstated in its peaks and troughs compared to that of the log of aggregate consumption expenditure. This correlation allows us to look at the regional data using the LCFS expenditure.

Figure 2: Changes in the Logarithm of Aggregate Consumption Expenditure and Total LCFS Expenditure (1997 - 2015)



4.2. Regional Results

Table 7 shows the fixed effects results for equation (4). Here, the testing down method only keeps specific months amongst and UK and its' regions. Year and region dummies are dropped by this method for being insignificant. Regression (1) for the UK generates a R-squared of 0.05%, which is extremely low. This figure increases to 5.4% as we consider Scotland in regression (3), and the highest R-squared of 29% for England in regression (2). No months were significant for Wales in regression (4) and Northern Ireland in regression (5), therefore both these R-squared values are zero. The table shows that the significant months are in the first half of the year.

Table 7: Results for Fixed Effects

VARIABLES	(1) Average Total LCFS Expenditure UK	(2) Average Total LCFS Expenditure England	(3) Average Total LCFS Expenditure Scotland	(4) Average Total LCFS Expenditure Wales	(5) Average Total LCFS Expenditure Northern Ireland
Month: January	-36.97*	-51.02***			
	(19.03)	(12.01)			
Month: March			-94.01**		
			(41.67)		
Month: April		-28.92**			
		(12.01)			
Month: May		61.94***			
		(12.01)			
Month: June			-88.84**		
			(39.90)		
Constant	3.823	1.351	14.71	3.323	0.862
	(5.337)	(3.670)	(12.23)	(11.57)	(12.30)
Observations	496	127	127	123	119
R-squared	0.008	0.311	0.069	0.000	0.000
Adjusted R-squared	**0.005**	**0.294**	**0.054**	**0.000**	**0.000**

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Firstly, we will consider the results for the entire UK. The R-squared results in regression (1) from table 7 will be our base standard to enable us to see how much our Google Trends data adds to explaining average total LCFS expenditure. (equation 5). The results from this are presented in table 8 in regression (1). Comparing the R-squared results; we show that by adding our Google Trends data an extra 11% of average total LCFS expenditure is explained. However, it is important to note that the testing down method in regression (1) eliminates the year and regional dummies. Therefore, we run another regression (2) whereby we force Stata to include all of these dummies in the model. This generates an R-squared of 9.1%, and therefore Google Trends data explains an extra 8.6% of average total LCFS expenditure. This decrease in the R-squared value in table 8 is expected as the testing down method excluded the dummies in order to improve the R-squared figure.

Table 8: Results for the UK from Adding Google Trends Data / £ (2004 – 2014)

VARIABLES	(1) Average Total LCFS Expenditure UK	(2) Average Total LCFS Expenditure UK
Lag 3: Pork and Sausages	-1.207***	-1.218***
	(0.444)	(0.451)
Lag 1: Vegetable, Tomato and Potato	-0.929*	-0.948*
	(0.555)	(0.563)
Lag 3: Orange Juice and Juice	-0.854**	-0.866**
	(0.422)	(0.430)
Lag 2: Petrol	0.551*	0.534*
	(0.299)	(0.304)
Lag 3: Petrol	-1.185***	-1.213***
	(0.306)	(0.315)
Lag 3: Games, Toys, Video Games and Board Games	-1.167**	-1.172**
	(0.540)	(0.548)
Lag 2: Restaurants	-0.964**	-0.985**
	(0.398)	(0.407)
Lag 2: Pub	-1.192***	-1.230***
	(0.414)	(0.423)
Lag 3: Pub	1.297***	1.277***
	(0.416)	(0.423)
Lag 1: Beer	-0.813*	-0.834*
	(0.431)	(0.441)
Lag 2: Cosmetics and Lipstick	1.833***	1.826***
	(0.646)	(0.655)
Month: January	-100.9***	-102.1***
	(30.84)	(31.30)
Month: February	11.31	10.97
	(26.59)	(26.95)
Month: March	-13.21	-13.45
	(26.75)	(27.12)
Month: April	-95.95**	-96.34**
	(40.94)	(41.55)
Month: May	-1.481	-1.088
	(31.31)	(31.78)
Month: June	-66.83*	-66.91*
	(34.40)	(34.96)
Month: July	-30.79	-30.85
	(33.75)	(34.27)

Month: September	-9.149	-8.966
	(28.33)	(28.83)
Month: October	-59.49**	-59.68*
	(30.14)	(30.76)
Month: November	-62.61**	-62.80**
	(29.68)	(30.21)
Month: December	-19.25	-19.35
	(28.87)	(29.44)
Year: 2005		2.200
		(36.53)
Year: 2006		-0.924
		(33.28)
Year: 2007		11.38
		(33.52)
Year: 2008		-1.324
		(33.30)
Year: 2009		14.14
		(33.41)
Year: 2010		-2.237
		(33.33)
Year: 2011		5.590
		(33.45)
Year: 2012		12.73
		(33.45)
Year: 2013		6.352
		(33.40)
Year: 2014		1.801
		(33.35)
Region: England		1.612
		(14.31)
Region: Northern Ireland		2.398
		(14.57)
Region: Scotland		1.079
		(14.31)
Constant	36.78	30.73
	(23.55)	(39.48)
Observations	476	476
R-squared	0.156	0.159
Adjusted R-squared	**0.115**	**0.091**

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In table 8 most of the variable coefficients are negative, this suggests that the more consumers search for the keyword, the less their household spends. For example, items in the food and non-alcoholic beverages category; pork and sausages, vegetable, tomato and potato and orange juice and juice all show negative signs. This could be because consumers search these items in an attempt to source less costly substitutes, and therefore they are associated with less expenditure. This rationale may also explain the negative coefficient for the search term petrol in lag 3 as consumers try to find the cheapest price. On the other hand, lag 3 of the keyword pub and lag 2 of cosmetics and lipstick both have positive coefficients and are significant at the 1% level. This may be because consumers consider these expenditure items as more of a luxury.

Table 9 shows the regional results of equation (5), which can be compared for each region to the R-squared results in the fixed effects model in table 7. When considering the Google Trends keywords the results show that the adjusted R-squared increases from 29.4% to 63.6% in England. Therefore, by adding the Google keywords an extra 34.2% of the LCFS total average expenditure is explained. In addition, when considering Scotland, the adjusted R-squared increases from 5.4% to 53.3%. This increase of 47.9% demonstrates that the Google Trends data does an even better job in Scotland at explaining more of the variance in average total LCFS expenditure than in England. Furthermore, the R-squared for Wales increases from 0% to

48.0% and from 0% to 61% in Northern Ireland. The results show that although the R-squared for England is the highest, adding Google Trends data brings the greatest improvement in statistical power in Northern Ireland.

Table 9: Regional Results (2004 – 2014)

VARIABLES	(1) Average Total LCFS Expenditure England	(2) Average Total LCFS Expenditure Scotland	(3) Average Total LCFS Expenditure Wales	(4) Average Total LCFS Expenditure Northern Ireland
Lag 1: Bread and Toast	-1.650***			
	(0.522)			
Lag 3: Bread and Toast			4.858***	
			(0.849)	
Lag 1: Poultry, Chicken and Duck	1.630**	-2.619**		
	(0.637)	(1.152)		
Lag 1: Pork and Sausages	2.063**			-1.190**
	(0.791)			(0.554)
Lag 3: Pork and Sausages			-2.653***	1.707***
			(0.963)	(0.594)
Lag 1: Vegetable, Tomato and Potato	-1.542*		-3.490***	-3.955***
	(0.883)		(1.167)	(0.692)
Lag 1: Ketchup and Mayonnaise			2.620**	3.029**
			(1.283)	(1.165)

Lag 2: Ketchup and Mayonnaise	-1.974***	2.874***		
	(0.386)	(0.877)		
Lag 3: Ketchup and Mayonnaise		2.740***		
		(0.841)		
Lag 1: Coca Cola and Lemonade			3.190***	-5.355***
			(0.967)	(0.789)
Lag 2: Coca Cola and Lemonade	-2.565**			
	(1.166)			
Lag 3: Coca Cola and Lemonade		6.856***	2.706***	-2.960***
		(1.804)	(0.988)	(0.913)
Lag 1: Orange Juice and Juice			-3.694***	
			(0.752)	
Lag 2: Orange Juice and Juice	0.807*	1.564**		-2.413**
	(0.465)	(0.619)		(1.035)
Lag 3: Orange Juice and Juice				
Lag 1: Women’s Coats and Women’s Jackets				1.849***
				(0.605)
Lag 2: Women’s Coats and Women’s Jackets			1.548**	-1.065*
			(0.663)	(0.616)
Lag 1: Petrol			3.406***	

			(0.721)	
Lag 2: Petrol	-0.511**			
	(0.211)			
Lag 3: Petrol				-1.713***
				(0.424)
Lag 1: Diesel				2.151***
				(0.685)
Lag 2: Diesel		-4.797***	-1.356**	1.160*
		(0.979)	(0.636)	(0.643)
Lag 3: Diesel	0.723*	-2.788***		
	(0.394)	(0.994)		
Lag 2: Games, Toys, Video Games and Board Games	1.663***			
	(0.516)			
Lag 3: Games, Toys, Video Games and Board Games			-4.764***	
			(1.105)	
Lag 1: Restaurants			2.264***	
			(0.665)	
Lag 3: Restaurants			1.151*	
			(0.594)	
Lag 1: Pub			2.569***	1.303**
			(0.780)	(0.588)

Lag 2: Pub		-2.741***		-1.901***
		(0.808)		(0.525)
Lag 3: Pub	-2.711***			
	(0.691)			
Lag 1: Beer	1.276**			
	(0.505)			
Lag 2: Beer	-1.047*	8.452***		-1.271*
	(0.576)	(1.342)		(0.752)
Lag 3: Beer	2.922***	3.708***	-3.954***	
	(0.671)	(1.128)	(0.991)	
Lag 1: Takeaway, Chinese Takeaway and Pizza Delivery			-2.317***	-1.485**
			(0.698)	(0.731)
Lag 2: Takeaway, Chinese Takeaway and Pizza Delivery				-2.117**
				(0.847)
Lag 3: Takeaway, Chinese Takeaway and Pizza Delivery				2.457***
				(0.907)
Lag 1: Toiletries and Soap	-0.967***		-1.816**	
	(0.340)		(0.796)	
Lag 2: Toiletries and Soap		2.652***		-3.354***
		(0.754)		(0.739)
Lag 3: Toiletries and Soap				2.805***

				(0.718)
Lag 1: Cosmetics and Lipstick	-4.226***			-3.924***
	(0.997)			(1.021)
Lag 2: Cosmetics and Lipstick			2.672**	
			(1.048)	
Month: January	-51.77***	-95.16**	-184.2***	-311.6***
	(18.21)	(45.14)	(59.69)	(50.02)
Month: February	-11.08	118.1**	-34.76	-79.96
	(17.21)	(45.41)	(56.93)	(51.44)
Month: March	103.3***	-182.7***	-107.5**	119.2**
	(22.67)	(45.95)	(53.01)	(45.83)
Month: April	-4.207	-102.5*	-351.1***	-150.4***
	(22.13)	(61.64)	(99.41)	(45.26)
Month: May	105.4***	96.79*	-169.4***	84.96*
	(17.96)	(49.52)	(61.21)	(45.54)
Month: June	72.86***	-210.7***	-364.2***	-70.69
	(21.13)	(51.98)	(70.55)	(46.04)
Month: July	-44.53**	-38.81	-195.4***	-96.89**
	(21.45)	(55.80)	(70.10)	(46.27)
Month: September	8.721	-223.9***	-103.9*	-12.13
	(18.77)	(52.96)	(59.20)	(43.87)

Month: October	-0.710	-145.8***	-69.53	-55.84
	(23.16)	(54.76)	(69.19)	(44.11)
Month: November	-0.887	25.35	-158.6**	-199.0***
	(17.71)	(45.80)	(67.74)	(44.76)
Month: December	67.72***	-136.2***	-99.97*	-47.11
	(16.06)	(45.47)	(59.56)	(49.31)
Constant	-20.21*	72.68**	157.1***	71.17**
	(11.98)	(35.83)	(50.77)	(31.81)
Observations	124	124	115	113
R-squared	0.716	0.632	0.614	0.728
Adjusted R-squared	**0.636**	**0.533**	**0.481**	**0.618**

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

In table 9 the Google search keywords vegetable, potato and tomato has a negative significant coefficient in England, Wales and Scotland at the 10% and 1% levels. This shows that an increase in search activity by consumers translates into less expenditure. Higher coefficients in Wales and Scotland means this negative effect is greater. On the other hand, searching for ketchup and mayonnaise in lag 1 has a positive effect on average total LCFS expenditure in Wales, which is significant at the 5% level. Furthermore, this positive effect is greater in Scotland in lag 2 as the coefficient is larger and is significant at the 1% level. This shows that the null hypothesis should be rejected with 99% confidence that there is a significant relationship between the Google Trends search and average total LCFS household expenditure. Searching for Coca Cola and Lemonade is significant for all regions of the UK at different lags between 1 and 3. However, the effects of searching these items are positive on average total LCFS expenditure in Scotland and Wales and negative in England and Northern Ireland. This indicates that consumers from England and Northern Ireland may search for cheaper alternatives, unlike consumers in Scotland and Wales.

In addition, lag 2 of petrol searches are significant in England at the 5% level and lag 3 of petrol is significant at the 1% level in Northern Ireland. Whereas, lag 2 of diesel is significant in

Scotland at the 1% level and in Wales at the 5% level. This shows that Google Trends searches of both petrol and diesel act as predictors of average total expenditure across all regions of the UK.

Searches for restaurants have a positive significant effect on average total LCFS expenditure in Wales in lags 1 and 3. In addition, lag 1 of both Google searches pub and beer have positive significant coefficients in England, Wales and Northern Ireland at the 5% and 1% level. This shows that households' expenditure of these luxury items heightens as their searches on Google increase. However, this does not hold for other expenditure items that are considered a luxury for example, takeaways as it has negative coefficients in Wales and Northern Ireland.

5.0. Conclusion

To conclude, we highlight via the Psychological Consumption theory that consumers are expected to spend or save in relation to their confidence surrounding their expectations of their future personal finances and environment (Katona, 1975; Van Raaij and Gianotten, 1990; Della-Penna and Huang, 2009 and Ven, 2011). However, as consumers become more sophisticated and risk adverse new and improved methods to capture consumers' consumption to denote their confidence are needed. We propose that using Google Trends data has the potential to explain consumer consumption in a more timelier manner, whilst also eliminating the weaknesses discussed of traditional consumer confidence indexes. For example, the lack of regional consumer confidence data limits the ability of economists, local authorities and businesses to have an in-depth insight into their local consumers to be able to make targeted changes to stimulate the economy.

Our aggregate results support the existing literature discussed previously. Consumer confidence is part of an economic system that adds to the explanation of consumption. Despite this, broad and differing survey questions have been challenged for generating results that are less representative of consumer's true confidence. Our study seeks to address this problem and builds on the small amount of existing literature exploring measuring consumer confidence via Google Trends. Our results show that a greater proportion of average total expenditure can be explained when considering Google Trends data in all regions of the UK, as oppose to the 47.8% of aggregate consumption that is explained when adding consumer confidence for the UK. When we consider that consumption accounts for around 60% of the UK's GDP, the implications of our data have the potential to change the future of measuring consumer confidence and the way local authorities approach their strategies. This is because Google Trends data has the ability to explain a greater proportion of consumption to denote consumer confidence in a more timelier manner than traditional consumer confidence indexes. Furthermore, Google Trends will enable businesses to relate to their consumers on a localised level through consideration of its expenditure categories in different regions of the UK. As changes in spending amongst different categories can depict changes in consumer confidence (Della-Penna and Huang, 2009).

In the future, our study should be built upon by exploring forecasting vector correction (VAR) models and vector error correction models (VECMs) with Lasso. This will enable us to explore the ability of Google Trends data to forecast consumption and consumer confidence, and therefore its accuracy as a new tool to predict consumption further in the future.

6.0. References

Acemoglu, D., and Scott, A., (1994), 'Consumer Confidence and Rational Expectations: Are Agents' Beliefs Consistent with the Theory?', *Economic Journal*, Vol. 104 (422), p. 1-19.

Askitas, N., and Zimmermann, F, K., (2009), 'Google Econometrics and Unemployment Forecasting', *Applied Economics Quarterly*, Vol. 55(2), p.107-120.

Bell, F, N, D., and Blanchflower, G, D., (2010), 'UK Unemployment in the great recession', *National Institute Economic Review* No. 214, p. 3-25.

Belloni, A., Chernozhukov, V., and Hansen, C., (2014), 'High Dimensional Methods and Inference On Structural and Treatment Effects', *Journal of Economic Perspectives*, Vol.28(2), p.29-50.

Berry, S., and Davey, M., (2004), 'How Should we Think About Consumer Confidence?', *Bank of England Quarterly Bulletin*. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=700066. [Accessed on 07/04/2017]

Boundless., (2016), 'GDP Equation in Depth (C+I+G+X)', *Boundless Economics Boundless.* Available at: https://www.boundless.com/economics/textbooks/boundless-economics-textbook/measuring-output-and-income-

19/measuring-output-using-gdp-92/gdp-equation-in-depth-c-i-g-x-349-12446/. [Accessed on 29/01/2017].

Bram, J., and Ludvigson, S., (1998), 'Does Consumer Confidence Forecast Household Expenditure? A Sentiment Index Horse Race', *FRBNY Economic Policy Review*, p.59-78.

Cadman, E., (2016), 'Consumer Spending Drives UK Growth', *Financial Times*. Available at: https://www.ft.com/content/e194e464-6b65-11e6-ae5b-a7cc5dd5a28c. [Accessed on: 29/01/2017].

Caner, M., and Zhang, H, H., (2014), 'Adaptive Elastic Net for Generalized Methods of Moments', *Journal of Business and Economic Statistics,* Vol. 32(1), p.30-47, DOI: 10.1080/07350015.2013.836104

Chamberlin, G., (2010), 'Googling the Present', *Economic &Labour Market Review,* ONS, p.59-94

Choi, H., and Varian, H., (2011), 'Predicting the Present with Google Trends', *Technical Report Google,* p. 1-12.

Cingano, F and Sichivardi, F., (2004), 'Identifying the Sources of Local Productivity Growth', *Journal of the European Economic Association,* Vol. 2(4), p.720-742.

Della-Penna, N., and Huang, H., (2009)., 'Constructing Consumer Sentiment Index for US Google Searches', *Department of Economics University of Alberta*, p.1-20.

Dominitz, J., and Manski, F, C., (2003), 'How should we measure Consumer Confidence: Evidence from the Michigan Survey of Consumers', p.1-27.

European Commission., (2014), 'The Joint Harmonised EU Programme of Business and Consumer Surveys: List of 'best practice' for the conduct of business and consumer surveys'. Available at: http://ec.europa.eu/economy_finance/db_indicators/surveys/method_guides/index_en.htm. [Accessed on 21/01/2017].

Evening Standard., (2012), 'Car giant Ford closes last UK plant, announcing Dagenham job loses', *Evening Standard*. Available at: http://www.standard.co.uk/business/business-news/car-giant-ford-closes-last-uk-plant-announcing-dagenham-job-loses-8226947.html. [Accessed on 18/10/2016].

Fornell, C., and Stephan, J., (2002), 'Consumer spending growth predicted by buyer satisfaction', *The National Quality Research Center University of Michigan Business School*, p.1-19.

Fornell, C., T.Rust, R. and G.Dekimpe, M. (2008), *'The Effect of Customer Satisfaction on Consumer Spending Growth', Journal of Marketing Research, American Marketing Association*, Vol. XL VI, p.1-29.

Fred., (2017), 'CBOE Volatility Index', *Economic Research Federal Reserve Bank of St. Louis.* Available at:

https://fred.stlouisfed.org/series/VIXCLS/. [Accessed on 02/04/2017].

Garner, A, C., (1991), ‘Forecasting Consumer Spending: Should Economists Pay Attention to Consumer Confidence Surveys?’, *Economic Review,* p. 57-71.

GfK., (2017), ‘UK Consumer Confidence drops one point in February to -6’ GfK. Available at: http://www.gfk.com/en-gb/insights/press-release/uk-consumer-confidence-drops-one-point-in-february-to-6/. [Accessed on 27/03/2017].

Goel, S, Hofman, M, J., Lahaie, S., Pennock, M, D., and Duncan, J, W., (2010), ‘Predicting Consumer Behaviour with Web Search’, *PNAS,* Vol. 107 (41), p. 17486-17490.

Google., (2017), ‘How Trends Data is Adjusted’, *Google Trends.* Available at: https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052. [Accessed on 05/01/2017].

Goossen, Y., 2007), ‘Alternative Progress Indicators to Gross Domestic Product (GDP) as a Means Towards Sustainable Development’, *European Parliament Policy Department.*

Graevenitz, V, G., Helmers, C., Millot, V., and Turnbull, O., (2016), ‘Does online search predict sales? Evidence from Big Data for car markets in Germany and the UK’, *Centre for Competition Policy,* p. 1-32.

Hendry, F, D., and Pretis, F., (2016), 'All Change! The Implications of Non-Stationarity for Empirical Modelling, Forecasting and Policy' *Oxford Martin School Policy Paper Series,* p.1-28.

Ipsos MORI., (2016), 'Ipsos MORI Consumer Confidence Survey, *Ipsos MORI.* Available at: https://www.ipsos-mori.com/Assets/Docs/Polls/post-brexit-consumer-confidence-2016-topline.pdf. [Accessed on 02/04/2017]

Katona, G., (1975), '*Psychological Economics'*, Elsevier Scientific Publishing Company, p. 1-438.

Kholodilin, A, K., Podstawski, M., Siliverstovs, B., Burgi, S, R, c., (2009), 'Google searches as a means of improving the nowcasts of key macroeconomic variables', *DIW Berlin Discussion Paper* No. 946, p. 1-13.

Kirkup, J., (2008), 'Financial crisis: 50 small businesses closing each day', *The Telegraph*. Available at: http://www.telegraph.co.uk/finance/financialcrisis/3379487/Financial-crisis-50-small-businesses-closing-each-day.html.[Accessed on 18/10/2016]

Kupelian, B and Hawksworth, J., (2013), '4- Consumer Spending Trends to 2030', *UK Economic Outlook Price Waterhouse Coopers.* Available at:

http://pwc.blogs.com/files/consumer-spending-trends-to-2030.pdf. [Accessed on 03/03/2017]

Kwiatkowski, D., Phillips, P, C, B., Schmidt, P., and Shin, Y., (1992), 'Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure are we that Economic Time Series have a Unit Root?', *Journal of Econometrics*, Vol. 54, p. 159-178. Available at: http://www.portalaction.com.br/sites/default/files/SeriesTemporais/planilhas/kpps_test.pdf. [Accessed on 03/04/2017].

Lahirl, K., Monokroussos, G., and Zhao, Y., (2016), 'Forecasting Consumption: The Role of Consumer Confidence in Real Time with Many Predictors', *Journal of Applied Econometrics*, Vol.31, P.1254-1275.

Internet Live Stats., (2017), 'Google Search Statistics', *Live Stats*. Available at: http://www.internetlivestats.com/google-search-statistics/. [Accessed 11/01/2017].

LSE., (2011), 'London Briefing – The Economic Recession-Why London Escaped Lightly', *London School of Economics*.

Ludvigson, C, S., (2004), 'Consumer Confidence and Consumer Spending', *Journal of Economic Perspectives*, Vol.18 (2), P.29-50.

Mantel, N., (1970), 'Why Stepdown Procedures in Variable Selection', *Technometrics,* Vol.12 (3), p.621-625.

Martin, P., (2011), ‘Warning: Google Makes Insights for Search Data Useless’, *Epiphany.* Available at: https://www.epiphanysearch.co.uk/news/2011/warning-google-makes-insights-for-search-data-useless/. [Accessed on 05/01/2017].

Martin, R., (2012), ‘Regional economic resilience, hysteresis and recessionary shocks’, *Journal of Economic Geography*, p. 1-32.

McLaren, N., and Shanbhogue, R., (2011), ‘Using internet search data as economic indicators’, *Bank of England Quarterly Bulletin*.

Ministry of Agriculture, Fisheries and Food, Office for National Statistics. (2004). *National Food Survey, 2000*. [data collection]. *2nd Edition.* UK Data Service. SN: 4512, http://doi.org/10.5255/UKDA-SN-4512-1

Monaghan, A., (2017), ‘Britons Spend More on Food and Leisure, Less on Booze, Smoking and Drugs’, *The Guardian*. Available at: https://www.theguardian.com/business/2017/feb/16/britons-spending-more-on-food-and-leisure-than-booze-smoking-and-drugs. [Accessed on 16/02/2016]

ONS., (2012), ‘Office for National Statistics 1 Living Costs and Food Survey Technical Report for survey year: January – December 2012’, *ONS.* Available at:

https://www.ons.gov.uk/ons/about-ons/get-involved/taking-part-in-a-survey/information-for-households/a-to-z-of-household-and-individual-surveys/living-costs-and-food-survey/living-costs-and-food-survey-2012.pdf+&cd=1&hl=en&ct=clnk&gl=uk [Accessed on 29/01/2017].

ONS., (2013), 'Office for National Statistics 1 Living Costs and Food Survey Technical Report for survey year: January – December 2013', *ONS.* Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/methodologies/livingcostsandfoodsurvey. [Accessed on 29/01/2017].

ONS., (2014), ''Office for National Statistics 1 Living Costs and Food Survey Technical Report for survey year: January – December 2014', *ONS.* Available at: https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/methodologies/livingcostsandfoodsurvey/livingcostsfoodtechnicalreport20142tcm774263841.pdf. [Accessed on 30/01/2017].

ONS., (2015), 'Family Spending: 2015', *Office for National Statistics,* p.1-88.

Schwert, G, W., (1989), 'Tests for Unit Roots: A Monte Carlo Investigation', *Journal of Business and Economic Statistics,* Vol. 7, p. 147-160. Available at:

http://schwert.ssb.rochester.edu/jbes89.pdf. [Accessed on 03/04/2017]

Sribney, B., (1996), ‘Problems with Stepwise Regression’, *StataCorp.* Available at: http://www.stata.com/support/faqs/statistics/stepwise-regression-problems/. [Accessed on: 21/04/2017]

Statista., (2017), ‘Market share held by the leading search engines in the United Kingdom (UK) as of January 2017, *The Statistics Portal.* Available at: https://www.statista.com/statistics/280269/market-share-held-by-search-engines-in-the-united-kingdom/. [Accessed 11/01/2017].

Tavakol, M., and Dennick, R., (2011), ‘Making Sense of Cronbach Alpha’, *Int J Med Educ,* Vol. 2, p.53-55. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4205511/. [Accessed on 14/03/2017].

Tibshirani, R., (1996), ‘Regression Shrinkage and Selection via the Lasso’, *Journal of the Royal Statistical Society*. Series B (Methodological), Vol. 58(1), p. 267-288.

Urminsky, O., Hansen, C., and Chernozhukov, V., (2016), ‘Using Double Lasso Regression for Principled Variable Selection’, p.1-70.

Van-Raaij, F, W., and Gianotten, J, H., (1990), 'Consumer Confidence expenditure, saving and credit', *Journal of Economic Psychology*, Vol.11(2), p.269-290.

Ven, J., (2011), 'Expenditure and Disposable Income Trends of UK Households: Evidence from Micro-Data', National *Institute Economic Review,* No. 218, p.44-57.

Wearden, G., Wainwright, M., and Gibbs, G., (2009), 'More than 1,100 jobs go as Burberry and Vion close factories', *The Guardian.* Available at: https://www.theguardian.com/business/2009/jan/20/burberry-jobs-factory-closure. [Accessed on 18/10/2016].

Wray, R., (2003), '400 jobs to go as Gillette moves east', *The Guardian.* Available at: https://www.theguardian.com/business/2003/dec/12/manufacturing. [Accessed on 18/10/2016].

Zou, H., and Hastie, T., (2005), 'Regularisation and Variable Selection via the Elastic Net', *J.R. Statistic. Soc.* B, Vol. 62 part 2, p. 301- 320.

7.0. Appendices

Appendix A

Table 1: Unit Root Tests Aggregate Data UK

	(1) Consumption	(2) Earnings	(3) CPI	(4) LIBOR	(5) VIX	(6) GfK	(7) MORI
Observations	69	69	69	69	69	69	69
Perron	0.000	0.000	0.739	0.531	0.0458	0.872	0.173
Mod. Dickey-Fuller	-2.174	-0.679	-2.058	-2.470	-3.491	-1.570	1.510
KPSS	0.156	0.360	0.386	0.163	0.110	0.227	0.214