University of Warwick
Department of Economics
EC331: Research in Applied Economics

Nowcasting private consumption in the US using insights from Google Trends

University ID: 1610375
Supervisor: Dr. Piotr Jelonek[1]
Word count: 4996

Abstract

This paper investigates the forecasting relationship between a Google Trends indicator and real private consumption expenditure in the US. The indicator is constructed by applying Kernel Principal Component Analysis to consumption-related Google Trends search categories. The predictive performance of the indicator is evaluated in relation to two conventional survey-based indicators: the Conference Board Consumer Confidence Index and the University of Michigan Consumer Sentiment Index. The findings suggest that in both in-sample and out-of-sample nowcasting estimations the Google indicator performs better than survey-based predictors. The results also demonstrate that the predictive performance of survey-augmented models is no different than the power of a baseline autoregressive model that includes macroeconomic variables as controls. The results demonstrate an enormous potential of Google Trends data as a tool of unmatched value to forecasters of private consumption.

[1] I am grateful to Dr. Piotr Jelonek for his insights and constructive criticism that allowed me to fully immerse into this research paper. His assistance alongside continuous feedback are much appreciated.

Table of Contents

1. Introduction

Since private consumption accounts for about 70% of the GDP in the US, making an accurate and fast assessment of short-term fluctuations in consumption expenditure is important when assessing overall economic performance (Vosen & Schmidt, 2011). For instance, if there is a sudden slowdown in economic activity, the Federal Reserve would prefer to diagnose it as soon as possible so that expansionary monetary policy is conducted in a timely manner in order to avoid further recessionary pressures. A major issue faced by policy-makers is a reporting lag of a month on data on private consumption. The existence of this lag creates a need for high frequency leading indicators that can provide insights into consumption dynamics well in advance of the official release of consumption figures. The practice of predicting the present and short-term future in the absence of timely economic data is referred to as economic nowcasting, it is of particular importance in times of macroeconomic uncertainty when past statistics on macroeconomic variables no longer resemble the present economic outlook (Vosen & Schmidt, 2011).

There is a variety of both conventional and unconventional leading indicators that are used to nowcast consumption. Survey-based sentiment indicators such as the University of Michigan Consumer Sentiment Index (MCSI) and the Conference Board's Consumer Confidence Index (CCI) ask consumers to evaluate both their own and overall economic conditions of the present and near future in an attempt to take into account both psychological and economic aspects of household behaviour[2]. According to the empirical literature, there is a significant correlation between the survey-based indicators and the private consumption figures in the US (Woo & Owen, 2019). However, most of the information within these indicators, is already captured by macroeconomic variables such as interest rates, wealth and income. In line with the recent technological advancements, there has been a rapid emergence of data sources on real-time economic dynamics provided by private companies that include Google, MasterCard, UPS and many others (Choi & Varian, 2012). This paper investigates Google Trends (GT), which is a collection of real-time indices of the aggregated search queries submitted by users into Google's search engine. Features such as a large sample size, observation of social interests in real time and the increasing popularity of the Internet as a tool for pre-consumption research for durable goods in the US, make it a solid candidate to improve upon the performance of survey-based nowcasting models (Carrière-Swallow & Labbé, 2013). Even in cases, where the purchase is not conducted online, consumers can obtain information about a product by utilising search engines, hence, their queries can potentially hold households' purchasing intentions (Combes & Bortoli, 2016).

This paper presents an indicator for private household consumption which is assembled using the search volume data provided by GT. Its main purpose is to assess the extent to which the GT data provides useful unmatched insights about the dynamics of consumption in the US. Before feeding a vast dataset of time series of GT search categories into the nowcasting equation, a dimensionality reduction procedure in the form of Kernel Principal Component

[2] For a discussion of economic psychology arguments on the relation between consumers' expectations and consumption decisions see Eppright *et al.* (1998)

Analysis (KPCA)[3] (Scholkopf *et. al.,* 1999) is conducted to make sure the model generalises well when presented with unseen data it was not trained on. The usefulness of the GT indicator to an economic forecaster is evaluated by measuring the extent to which the Google Components enhance a parsimonious AR(1) model compared to conventional survey-based sentiment indicators. In accordance with the empirical literature on leading consumption indicators, the GT Indicator[4] is assessed to test its capacity to boost the forecasting performance of the models that already incorporate other macroeconomic variables that act as a form of a robustness check. Hence, this research augments the above-mentioned AR(1) model with macroeconomic fundamentals relevant to consumer spending to capture the true predictive power of the Google Factors. This paper conducts both in-sample and out-of-sample nowcasting procedures using monthly year-on-year (yoy) growth rates of variables that range from February 2005 to November 2018. Both in-sample and out-of-sample nowcasting experiments reveal that the model augmented with GT data performs better than the survey-based indicators. They also suggest that both MCSI and CCI contain no valuable information for the nowcasts in presence of macroeconomic variables.

This research paper contributes to the existing literature in two crucial ways. To my knowledge, this study is the first to employ the method of KPCA to the current and one-month-lagged dataset of GT variables. Projections of non-linear GT series prove to be extremely helpful in boosting the nowcasting performance of the GT-based model to levels comparable with state-of-the-art GT specifications for US consumption nowcasting. A major deficiency of past

[3] Refer to Appendix 3
[4] Terms GT/Google indicator/Factors/Components/data/series are used interchangeably throughout the paper

literature (mostly published before 2011) on GT consumption nowcasting is the lack of observations since GT data is only available from 2004. This issue is of further importance since GT sample period happens to contain the turbulent times of the 2007-08 financial crisis. In this paper, I employ 165 monthly observations of the GT series which allows drawing more solid conclusions regarding the predictive performance of the GT indicator.

The remainder of the paper starts off by reviewing the most relevant literature. The third and fourth sections describe the dataset and the empirical approach to evaluate the usefulness of the Google Factors and survey-based indicators. The fifth section presents the discussion of the findings and is followed by a conclusion.

2. Literature Review

This work is related to two different branches of the literature. The first one is concerned with the use of survey-based consumer sentiment indicators in forecasting fluctuations in private consumption. The second one applies GT data to enhance the performance of a selection of short-term forecasting models.

Survey-based indicators have been extensively used in the consumption nowcasting models. However, a vast number of papers question the ability of the MCSI and CCI to provide information that is not already present in the macroeconomic variables. Fuhrer (1993) finds that around 70% of the variance in the MCSI can be explained by a mixture of macroeconomic fundamentals. This implies that a large portion of sentiment might be a reflection of the households' knowledge of the overall economic outlook. A potential deficiency of the survey-based indicators might be that they do not precisely capture the relationship between expectations and real spending (Vosen & Schmidt, 2011). With the help of in-sample regressions Carroll *et al*. (1994) and Ludvigson (2004) find that sentiment indicators nevertheless have predictive power for US consumption expenditure over and above to that captured by other macroeconomic variables. Opposingly, other research papers including Croushore (2005), who utilise real-time observations for out-of-sample forecasting trials, discover that CCI and MCSI are not helpful in forecasting consumer expenditure.

There is a considerable amount of literature that concerns itself with a variety of applications of GT data to fields not necessarily related to economics. GT series[5], in simple words, refer to the relative popularity of a given search keyword or category. So far, the first publication that claims that search query data is useful in forecasting economic statistics is Ettredge *et al.* (2005); it investigated the unemployment rate in the US. One of the most successful applications of GT data was in the field of epidemiology, Ginsberg *et al.* (2009) and Polgreen *et al.* (2008) demonstrated that search data could be helpful in predicting the occurrence of influenza-like diseases. Choi & Varian (2012) can be regarded as the first seminal paper in the field of GT nowcasting since they successfully employed GT variables to forecast short-term values of economic variables that include automotive sales, unemployment claims and the consumer confidence index. Askitas & Zimmermann (2009) discovered that a selection of search keywords linked with job search activity could form a useful index for forecasting the German unemployment rate.

Given these implications, this paper utilises GT as an instructive predictor for private consumption spending in the US. While macroeconomic variables signal consumers' ability to make purchases and survey-based indicators attempt to capture households' willingness to spend (Wilcox, 2007), the GT indicator's intention is to provide an index for consumers' preparatory steps to spend by incorporating the volume of consumption-related search keywords and categories into forecasting models (Vosen & Schmidt, 2011). Furthermore, in contrast to most conventional survey-based indicators, GT data is a by-product of normal

[5] See Appendix 2

activity and is continually collected on a wider range of topics rather than on a number of pre-compiled questions. As a result, search data is more suited for analysing unanticipated issues (McLaren & Shanbhogue, 2011).

The idea of applying the vast dataset offered by GT in the context of nowcasting consumption expenditure is not new. Kholodilin *et al.* (2010) employ a dataset comprised of 220 consumption-related Google searches to conduct US private consumption nowcasting experiments. A major drawback of this paper is that it uses specific search terms instead of GT categories that aggregate keywords with similar meaning into indices. Specific keywords are likely to be more exposed to isolated shocks that are not relevant to consumption, which could introduce biases to the GT indicator (Combes & Bortoli, 2016). Not to mention, Ross (2013) claims that single keywords might be contaminated by a substantial amount of noise that can be caused by public campaigns, changes in trends or due to the emergence of searches entered by automated code scripts. When conducting out-of-sample forecasting experiments Kholodilin *et al.* (2010) estimate the model parameters using a 24-month rolling window estimation technique[6]. The paper justifies that procedure by arguing that models with limited memory can better follow a time series that contains structural changes not accounted for in the model (Giacomini & White, 2006). On the contrary, Vosen & Schmidt (2011) employ the expanding window technique in their estimations arguing that it leads to higher parameter stability and accuracy, and it makes more sense for forecasters to utilise all of the available data. When it comes to constructing the Google indicator, Kholodilin *et al.* (2010) applies the method of

[6] See Appendix 4

Principal Components (PC) to raw GT time series variables. However, unlike the Vosen & Schmidt (2011) and many other papers on GT-based consumption nowcasting[7], the PC extraction procedure is performed iteratively using the sample available at the time of the nowcast so that the indicator is in tone with the real-time vintages of private consumption expenditure (Kholodilin *et al.*, 2010).

Recent research at the intersection of consumption nowcasting and the utilisation of GT data in has revealed a number of remarkable results. For instance, the model augmented by the GT indicator in Vosen & Schmidt (2011) paper was the only one to accurately point out the turning point after the consumption plummeted to its long-time lowest value in December 2008. Furthermore, by constructing separate forecasting models for durable, non-durable and services consumption as dependent variables, Woo & Owen (2019) were able to achieve a further boost in the predictive performance, on average their nowcast errors were 7.14 % lower than the baseline model that only used sentiment indicators. Notably, forecasting experiments conducted by Woo & Owen (2019) provide evidence that durable goods models based on GT data perform better in one-month-ahead forecasting than nowcasting, whereas for non-durables the results are mixed, implying that in general pre-consumption research of durables happens much more in advance when compared to non-durables or services.

Overall, the majority of papers suggest that forecasting models should not be solely based on survey-based indicators. Moreover, they do not claim that GT based methods are good for

[7] Data transformation conducted ahead of estimations

forecasting, they suggest that GT is about diagnosing the present economic outlook. The literature lead to two hypotheses:

- GT-based models are capable of outperforming consumption nowcasting specifications based on sentiment indicators.
- Survey-based indicators do not entail predictive power that is not already captured by macroeconomic fundamentals.

3. Data[8]

Google Trends[9] is a database of recorded search queries entered into Google's search engine for a large number of keywords. The dataset consists of time series indices of the volume of Google search queries in a given geographic area and time period. Prior to public release, Google Trends transforms the raw data according to the following procedure[10]:

- The volume of a given search keyword is divided by total searches in a given geographic area and time interval it represents. This forms the query share of a given query index.
- The maximum query share within the given time period is then rescaled to an index that ranges from 0 to 100.

This procedure ensures that any trends that result from a change in the relative popularity of Google's search engine or the Internet overall are not present in the dataset.

GT supplies aggregated indices of search queries, which are funneled into a set of categories via a natural language classification engine. I have selected 48 consumption-related categories, these best match the product categories of Bureau of Economic Analysis (BEA) national income and product accounts (Vosen & Schmidt, 2011). Table 1 demonstrates the GT categories employed for three components of consumption.

[8] See Appendix 1 for variable descriptions, sources and summary statistics, Appendix 5 for stationarity tests
[9] See Appendix 2 and/or https://trends.google.com/trends/?geo=US
[10] Raw series are not disclosed by Google

The time series provided by Google do not undergo seasonal adjustment. Throughout the paper, year-on-year growth rates are used instead of seasonally adjusted series in levels or monthly growth rates. This decision is made because of the following reasons. It is unreasonable to calculate precise seasonal factors, as the observations are only available since 2004 and the range of the dataset includes the uncertain times of the real-estate crisis. Not to mention, most of the mainstream literature uses year-on-year growth rates, hence this specification makes comparisons to past literature easier. However, a drawback of this approach is that 12 months of data points are lost in subsequent analysis.

Table 1 - Consumption-related GT categories

Durable Consumption	Nondurable Consumption	Services Consumption
Automotive Industry	Food & Drink	Home Financing
Auto Insurance	Grocery & Food Retailers	Home Improvement
Vehicle Brands	Non-alcoholic Beverages	Home Insurance
Vehicle Shopping	Apparel	Homemaking & Interior Decor
Computer Electronics	Apparel Services	Drugs & Medications
Consumer Electronics	Footwear	Health Insurance
Home Appliances	Undergarments	Medical Facilities & Services
Home Financing	Energy Utilities	Retirement Pension
Home Gardening	Oil & Gas	Internet & Telecom
Home Improvement	Beauty & Fitness	Social Services
Homemaking and Interior Decor	Chemical Industry	Waste Management
Book Retailers	Drugs & Medications	
Entertainment Industry	Face & Body Care	
Movies	Hair Care	
Computer & Video Games	Health	
Mobile Wireless	Newspapers	
Internet & Telecom	Tobacco Products	

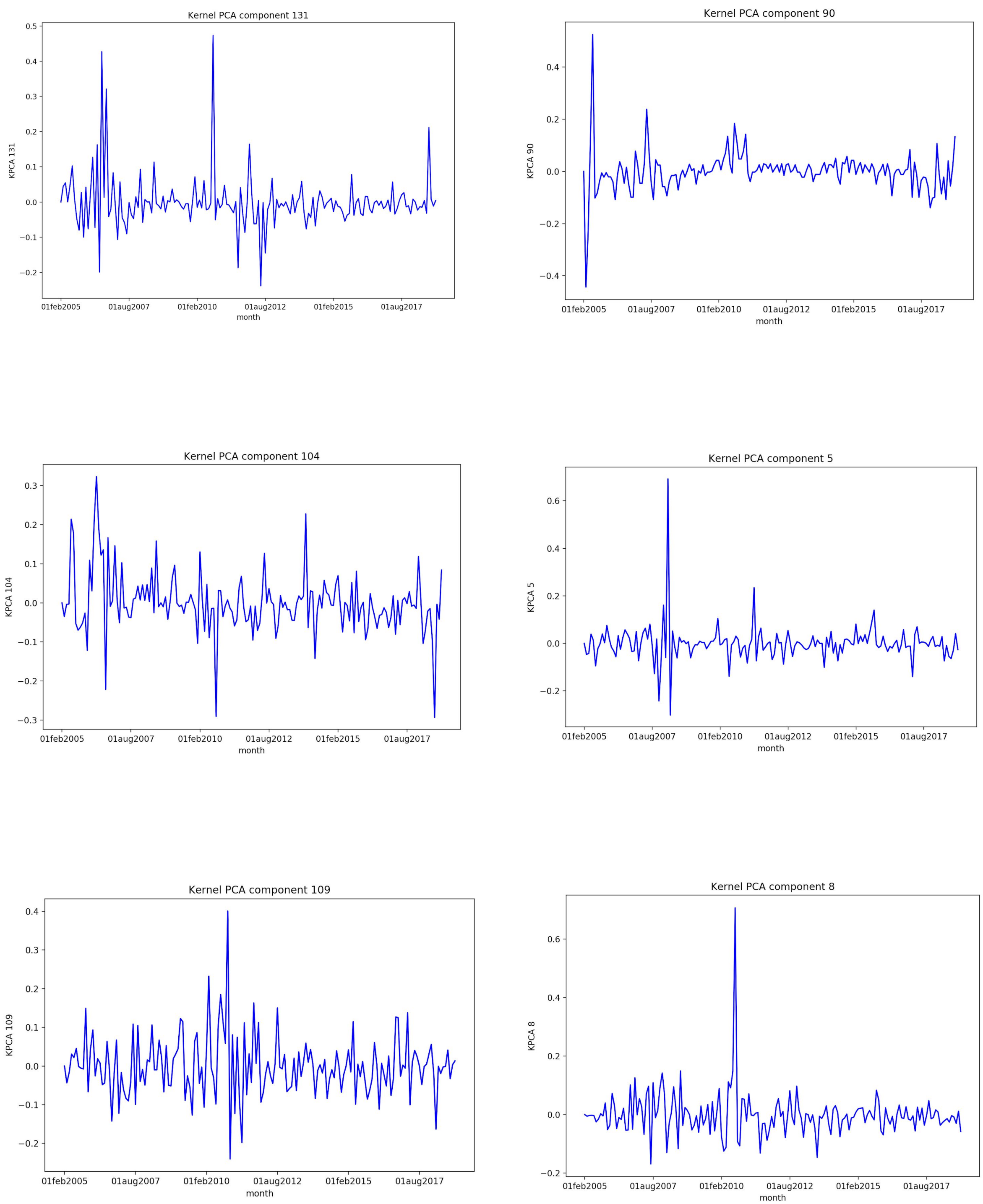
The GT tool possesses several drawbacks which diminish the confidence in the stability of the predictive performance of the GT indicator. The main weakness is the lack of transparency, the information on the processing and normalisation applied to GT keywords and categories is not disclosed by Google. Moreover, the composition of categories over time, especially when a new popular search term appears on a given day, is not reported (Combes & Bortoli, 2016).

A series of procedures are conducted on the raw GT variables to extract as much information from them as possible without running out of degrees of freedom in the forecasting specifications. KPCA[11] is applied to the dataset containing current and one-month lagged values of the GT series. The justification behind augmenting the initial dataset with time-lagged GT series is the hypothesis of users submitting their searches up to one month before purchasing a product (Combes & Bortoli, 2016). There are several reasons that make KPCA a better dimensionality reduction technique than a more traditional Principal Component Analysis (PCA), employed by most of the mainstream literature on GT nowcasting. In particular, KPCA can find principal components of non-linear transformations of the original feature space to represent the data in a lower dimension which proves to enhance the nowcasting performance. The KPCA suggests 165 components, however, with regards to a relatively limited sample period at hand, it is crucial to eliminate a large proportion of these factors to circumvent overfitting the predictive model. Nowcasts are initially performed with every component individually to select the factor with the highest predictive power and append it to the final set that resembles the GT indicator. This

[11] Refer to Appendix 3

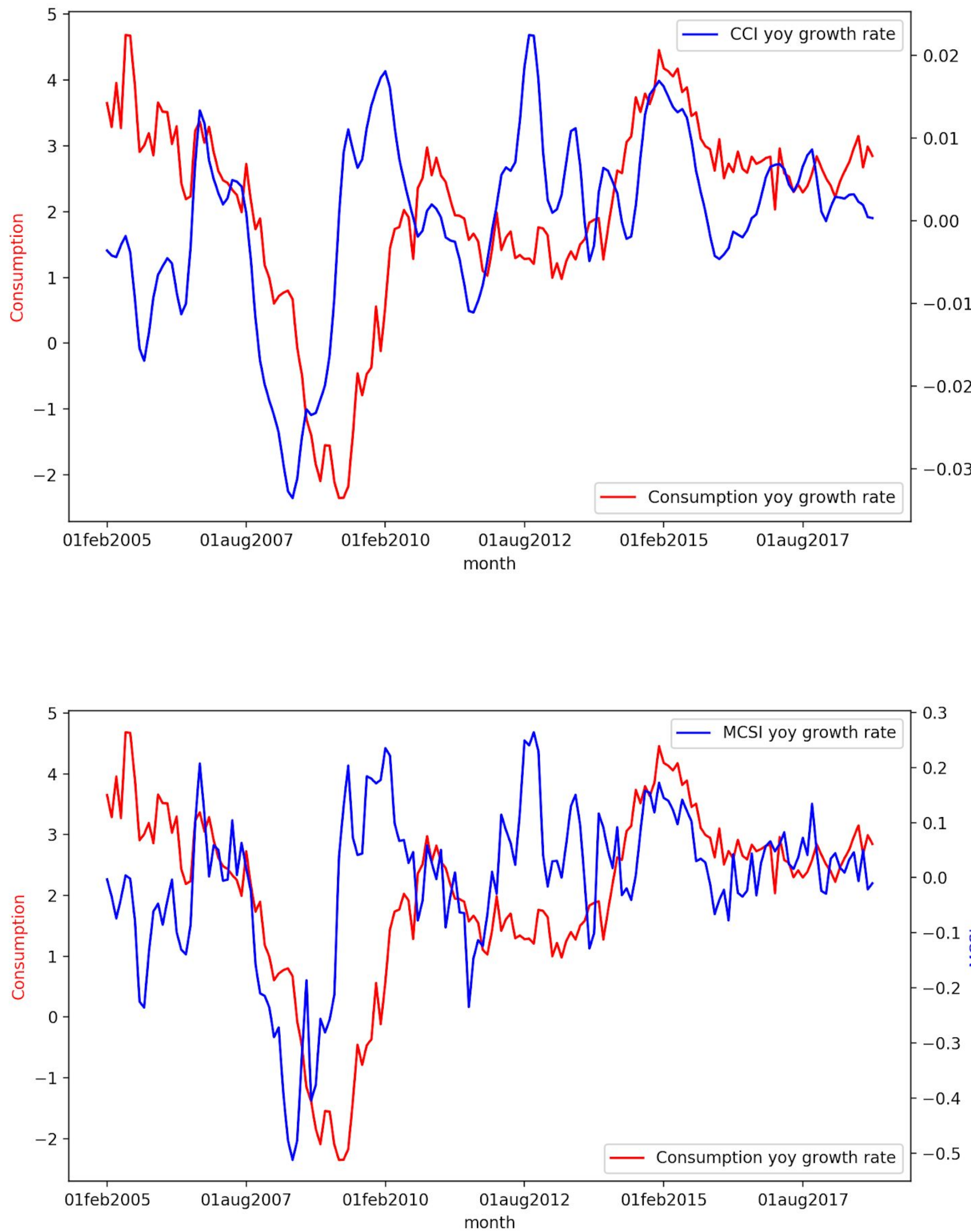
procedure is then continued in an iterative fashion with the remaining factors until an additional component has no enhancing impact on the model predictions. The best results were obtained when employing 25 out of 165 available factors, in what follows these 25 components will correspond to the GT indicator. Figure 1 displays six top performing KPCA components. A limitation of the above approach is that the transformation of the raw GT series into a compact set of predictive components is conducted in advance and is not applied at every iteration of the nowcaster in the further analysis. The issue of this approach is that one GT category might become a better predictor as time passes, and the model should ideally account for that.

Figure 1 - Top Performing KPCA Components



Moving on to more conventional predictors, MCSI and CCI are survey-based sentiment indicators that act as benchmark predictors in subsequent analysis. Both indices act as a proxy for consumer confidence which they attempt to extract by asking five questions that involve both current and expectation components. MCSI questions households on their financial position as well as their current view towards major purchases, whereas CCI puts a greater emphasis on the situation in the labour market. Therefore, CCI slightly lags behind MCSI since it is more linked to the unemployment rate, which has a tendency to lag the economic cycle. Moreover, CCI is more volatile than MCSI as it employs a different construction procedure. The result of these dissimilarities is occasional conflicting signals provided by the indicators, however, overall they are highly correlated (Woo & Owen 2019). Year-on-year growth rates of survey-based indicators are used instead of their levels to allow for better comparability to the Google indicator. The co-movement of CCI and MCSI and real private consumption looks quite incredible, although the time lead of the predictors appears to vary (Figure 2).

Figure 2 - Consumption and survey-based indicators (monthly year-on-year growth rates)



Throughout this paper's methodology, the dependent variable is the monthly year-on-year growth rate of real private consumption. In further analysis, forecasting models will be augmented by a selection of macroeconomic variables to assess whether the information provided by the predictors is beyond that already present in macroeconomic fundamentals. The choice of the variables takes inspiration from the models used by Croushore (2005); control macroeconomic series are the real personal income, interest rates on 3-month Treasury Bills and the S&P 500 index. The last two variables have a one-month publication lead over consumption and can, therefore, be used for nowcasting.

4. Methodology

To contrast the predictive performance of the Google indicator to that of the survey-based sentiment variables, I first construct a parsimonious autoregressive model of consumption growth. Graphical inspection of the autocorrelation and partial autocorrelation functions alongside formal testing confirmed that the autoregression is of the first order. The inspection of the model residuals does not reveal statistically significant autocorrelations, which further signals the correctness of the specification[12]. The simple AR(1) is then extended by the monthly year-on-year growth rates of real personal income y, interest rates on 3-month Treasury Bills i and stock prices[13] s to ensure the true predictive power of the indicators is extracted. Below, model (1) is the baseline; it is then augmented with CCI (2), $MCSI$ (3) and the $Google$ indicator (4). In equation (4) $Google$ refers to a matrix of 25 KPCA factors, γ refers to the matrix of the respective 25 coefficients. The operation between these two matrices is a dot product[14].

$$C_{\mathrm{t}} = \alpha + \beta_1 C_{\mathrm{t\text{-}1}} + \beta_2 y_{\mathrm{t\text{-}1}} + \beta_3 i_{\mathrm{t}} + \beta_4 s_{\mathrm{t}} + \epsilon_{\mathrm{t}} \tag{1}$$

$$C_{\mathrm{t}} = \alpha + \beta_1 C_{\mathrm{t\text{-}1}} + \beta_2 y_{\mathrm{t\text{-}1}} + \beta_3 i_{\mathrm{t}} + \beta_4 s_{\mathrm{t}} + \gamma CCI_{\mathrm{t}} + \epsilon_{\mathrm{t}} \tag{2}$$

$$C_{\mathrm{t}} = \alpha + \beta_1 C_{\mathrm{t\text{-}1}} + \beta_2 y_{\mathrm{t\text{-}1}} + \beta_3 i_{\mathrm{t}} + \beta_4 s_{\mathrm{t}} + \gamma MCSI_{\mathrm{t}} + \epsilon_{\mathrm{t}} \tag{3}$$

[12] See Appendix 6
[13] Measured by S&P 500 index
[14] Models from (1) to (4) do not include further lags of variables as they are not statistically significant

$$C_{\mathrm{t}} = \alpha + \beta_1 C_{\mathrm{t\text{-}1}} + \beta_2 y_{\mathrm{t\text{-}1}} + \beta_3 i_{\mathrm{t}} + \beta_4 s_{\mathrm{t}} + \gamma Google_{\mathrm{t}} + \epsilon_{\mathrm{t}} \tag{4}$$

The main objective of this paper is to test if Google Factors hold information beyond that which is carried by lagged values of the dependent variable C and other explanatory variables. This conceptual framework for determining the predictive link between two series was suggested by Granger (1969). Ashley *et al.* (1980) claim that the procedure of out-of-sample prediction is more aligned with the concept of Granger's theoretical framework since it only employs data available at the time the forecast is produced. Chen (2005) finds that tests performed in out-of-sample setting have a higher power in the presence of structural breaks, however, in their absence, in-sample tests offer a higher power. I test the hypothesis of Granger causality using forecasting results of in-sample estimation and out-of-sample estimation schemes to ensure robustness of the findings.

This research conducts in-sample and out-of-sample experiments to determine the extent to which the predictors are helpful in nowcasting dynamics of consumption expenditure. In-sample nowcasts estimate model parameters over the complete sample period that ranges from February 2005[15] to November 2018. To determine which indicator enhances the baseline specification the most, the adjusted R^2 of the indicator-augmented model is compared to that of the baseline model (1). I also calculate the F-statistics to test the joint significance of the respective predictors. The statistical significance of the relative increase in adjusted R^2 can thus be confirmed or disproved by the F-test.

[15] January 2005 is lost since one-month-lagged GT raw data was used to construct KPCA Google Factors

Out-of-sample experiments are conducted via expanding window recursive procedures[16]. Although rolling window models with limited memory are better at tracking time series in presence of structural shifts, the expanding window procedure results in higher parameter stability and accuracy, and it is more sensible for forecasters to employ all data at hand (Vosen & Schmidt, 2011). Initially, I fit forecast specifications to data from February 2005 to July 2010. Next, I perform out-of-sample predictions from August 2010 to November 2018, appending one month at each iteration, re-fitting the model and computing a sequence of forecasts for the current period. The predictions produced by indicator-based specifications are evaluated by their corresponding root mean squared forecast errors (RMSFE) and compared to that of the other models. The statistical significance of the results is tested using the Harvey *et. al.* (1997) alteration of the Diebold & Mariano (1995) statistic.

The robustness of the GT-augmented specification can potentially be improved by utilising an ensemble of methodological techniques. Blending the nowcasts produced by different models better operates all of the available information and also manages to be more robust when surprised by a shock produced by an isolated category. This phenomenon is referred to as the "forecast combination puzzle" by Watson & Stock (2004).

[16] See Appendix 4

5. Results

Table 2 demonstrates the results of the in-sample estimation for the models augmented by the indicators relative to the baseline specification for February 2005 to November 2018 interval. It states the increment to the adjusted R^2, the result of enhancing the baseline model with the corresponding predictor, and the F-statistic for a test of the significance of the indicator. All of the improvements to the baseline model proposed by the indicators are statistically significant. However, increases in the adjusted R^2 are all of minor size since the first lag of the dependent variable already accounts for a large portion of the variation. The model based on the GT indicator reaches the highest incremental adjusted R^2 of 1.3 percentage points. The information content of CCI and MCSI is substantially inferior, they both yield a 0.3 percentage point increase which is rather negligible.

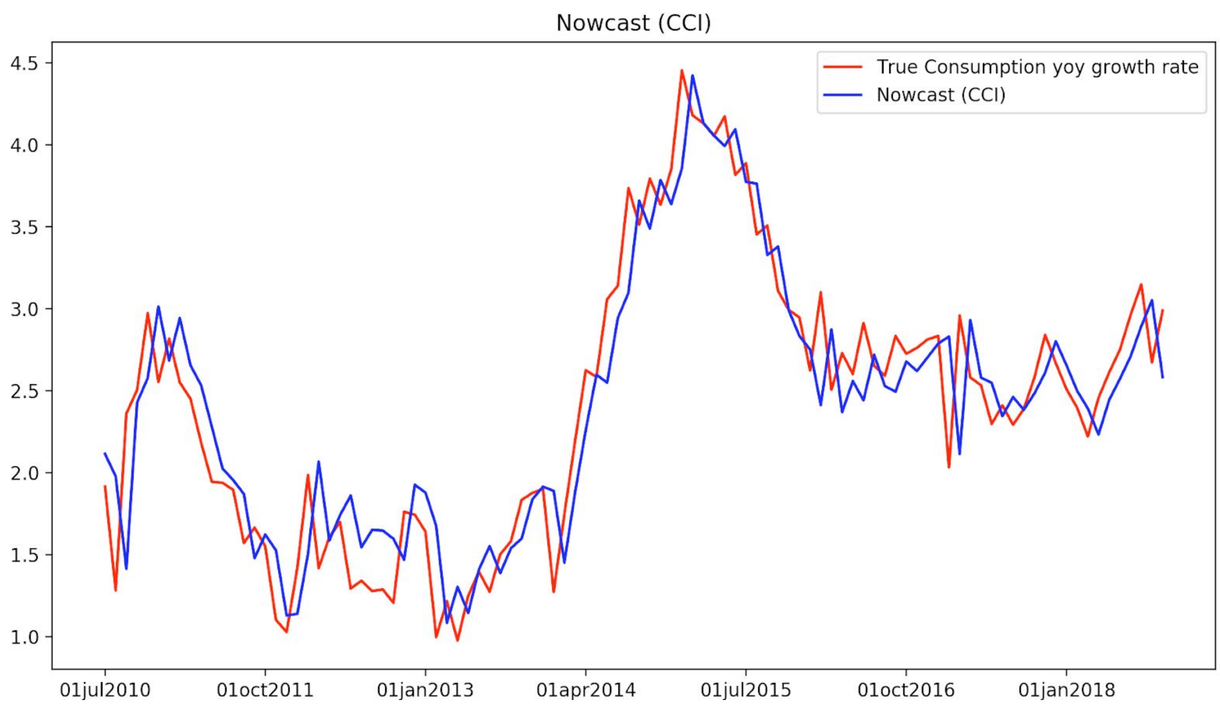
Table 2 - Information content of the indicators

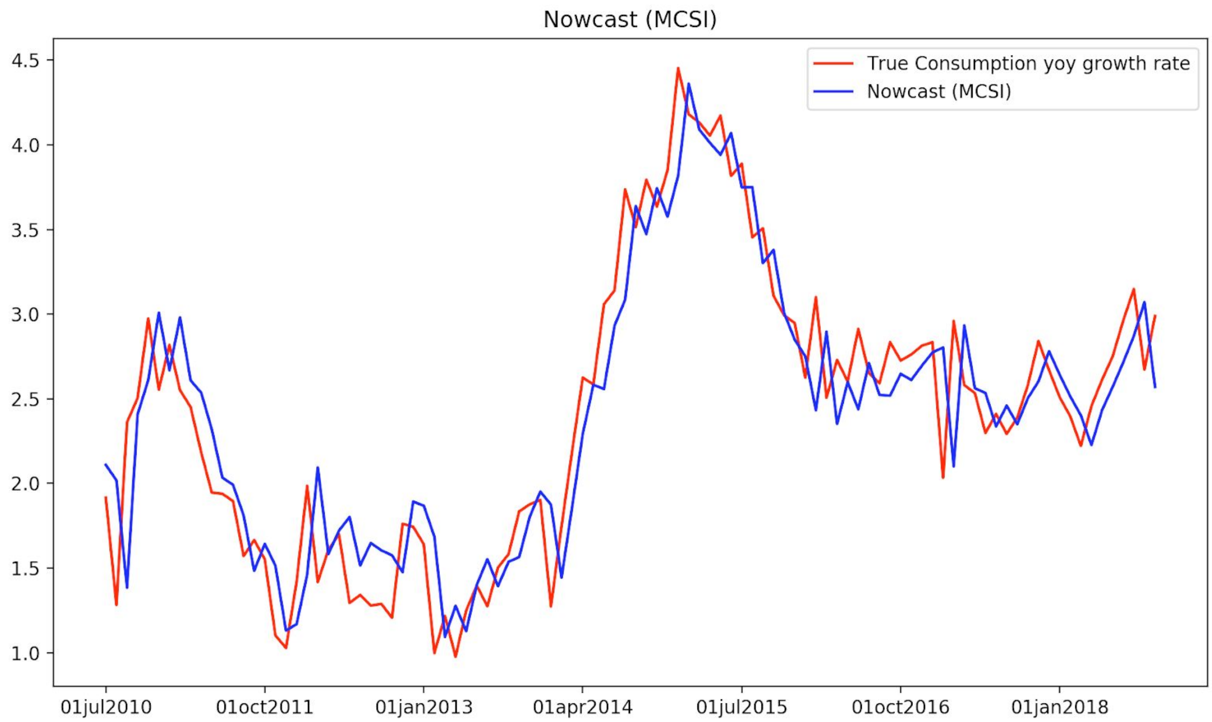
	Increm. adjusted R^2	F-statistic
CCI (2) vs. Baseline (1)	0.003	7.75***
MCSI (3) vs. Baseline (1)	0.003	7.10***
Google (4) vs. Baseline (1)	0.013	2.28***

Note: Asterisks document significance at the *10%, **5% and ***1% significance level. The sample ranges from February 2005 to November 2018.

Analysis performed thus far suggests that the introduction of the GT indicator outperforms univariate models based on CCI and MCSI. Below I investigate if the Google Factors are capable of increasing the nowcast precision in out-of-sample experiments. The out-of-sample nowcasting performance of models (2) to (4) is illustrated in Figure 3. The nowcast values are displayed in comparison to the actual yoy growth rates of US consumption expenditure.

Figure 3 - Out-of-sample nowcasts and true year-on-year growth rates of US consumption




Nowcast (MCSI)
True Consumption yoy growth rate
Nowcast (MCSI)
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
01jul2010
01oct2011
01jan2013
01apr2014
01jul2015
01oct2016
01jan2018

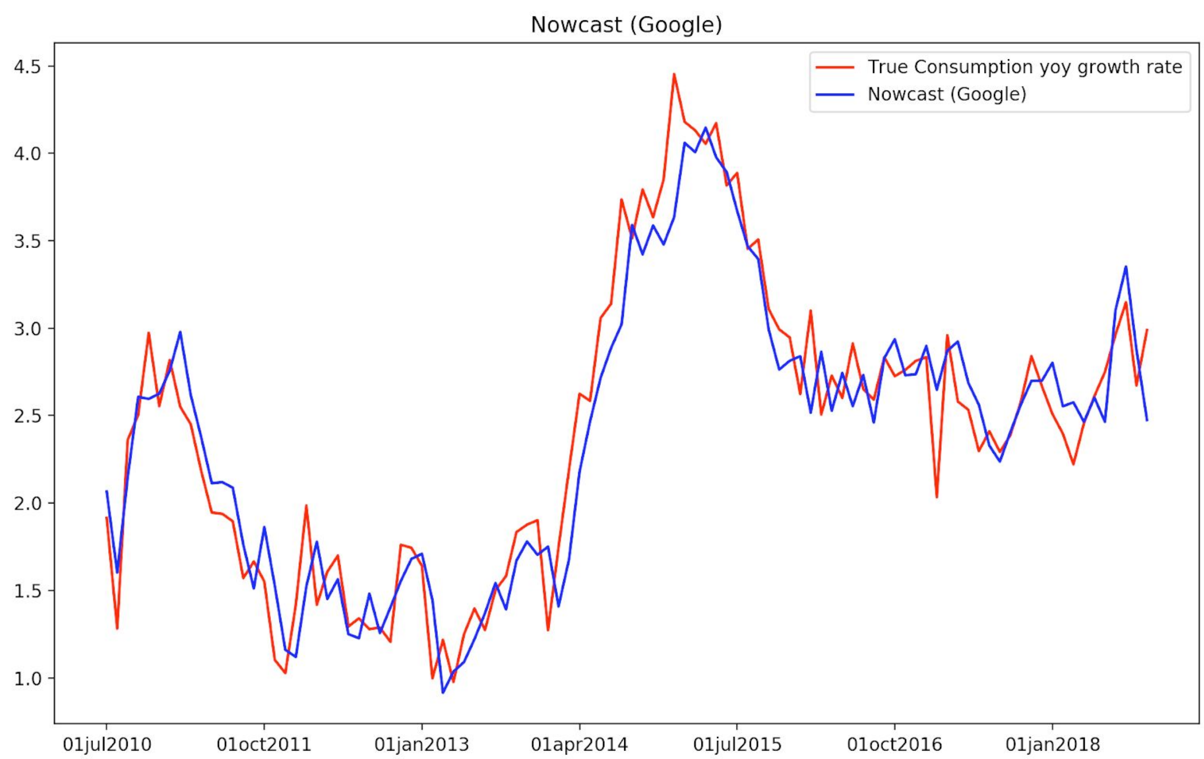

Nowcast (Google)
True Consumption yoy growth rate
Nowcast (Google)
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
01jul2010
01oct2011
01jan2013
01apr2014
01jul2015
01oct2016
01jan2018

Figure 3 shows that the GT-based model follows true consumption more closely when compared to MCSI and CCI augmented specifications. Table 3 supports the visual impressions and reports RMSFE for all indicator-based specifications. In presence of the macroeconomic variables both MCSI and CCI approximately produce the same RMSFE as the baseline model. GT indicator, on the other hand, manages to reduce the nowcast error by 20.2% when contrasted with both the baseline and survey-based specifications.

Table 4 reports relative RMSFE amongst the nowcasting specifications. Values lower than unity suggest that the first model is more accurate than the second one. In spirit of Granger Causality, the table also documents Harvey *et al*. (1997) modification of the Diebold & Mariano (1995) test statistic for equal forecast accuracy to test whether the forecast error improvements are due to statistical chance. The modified Diebold-Mariano (MDM) values are only applicable to non-nested models, therefore they are only reported for the comparisons of the indicator-based models. The MDM statistic follows a Student's t-distribution and determines whether differences in RMSFEs are statistically significant. If the first model has a lower prediction error than the second one, the MDM statistic has a negative sign. Table 4 demonstrates that the Google-augmented model significantly[17] outperforms all other models. The null hypothesis of the MDM test is that there is equal accuracy between two forecasting models, however, I am able to reject it at a 1% significance level when GT predictions are compared to both the CCI and MCSI nowcasts. Particularly, high negative MDM test statistics of -2.69 and -2.71 confirm that the GT-based indicator yields a lower nowcasting error.

[17] Significantly refers to both statistical and absolute significance

Table 3 - Out-of-sample nowcasting performance measured by RMSFE (percentage points)

	RMSFE
Baseline (1)	32.96
CCI (2)	32.78
MCSI (3)	32.95
Google (4)	**26.3**

Table 4 - Relative out-of-sample predictive power of nowcasting models

	Relative RMSFE	MDM statistic
CCI (2) / Baseline(1)	0.99	
MCSI (3) / Baseline(1)	1.00	
Google (4) / Baseline(1)	0.80	
Google (4) / CCI(2)	0.80	-2.69***
Google (4) / MCSI(3)	0.80	-2.71***

Note: Asterisks document significance at the *10%, **5% and ***1% significance level.

For MCSI and CCI in-sample and out-of-sample results are in line with the claim of Crushore (2005) that survey-based indicators do not enhance forecasting performance once other macroeconomic variables are controlled for in the model. This implies that a considerable part of sentiment could be a reflection of a consumer's comprehension of the general macroeconomic stance. On the other hand, following nowcasting experiments the GT-augmented model comes out as a leading specification. The nature of GT's contribution may be linked to GT's ability to signal pre-consumption research conducted by households in advance of purchasing durable goods (Woo & Owen, 2019).

6. Conclusion

This research paper reveals the enormous potential of Google Trends as a leading indicator for US consumption expenditure. Google Trends detects changes in contemporaneous consumption by providing valuable information about the pre-consumption research activity of US households conducted either on the month or one-month in-advance of purchasing a product. Both in-sample and out-of-sample nowcasting experiments proved that the predictive power of the GT indicator is much higher than that of survey-based sentiment indicators.

In particular, by utilising the method of Kernel Principal Component Analysis alongside an iterative procedure of factor selection this paper was able to extract valuable information from GT variables that resulted in a 20.2% reduction in the out-of-sample nowcasting RMSFE relative to both the baseline and survey-based specifications. This advancement of the model performance is proved to be statistically significant given the modified Diebold-Mariano statistic of equal forecasting accuracy. The introduction of the technique of KPCA coupled with a large sample size of GT observations relative to past literature resulted in a GT-augmented model more accurate (relative to conventional sentiment indicators) than most state-of-the-art specifications proposed by existing papers on GT-based US consumption nowcasting. Negligible improvement in the adjusted R^2 in in-sample experiments and RMSFE of out-of-sample nowcasts of CCI and MCSI augmented models relative to the baseline specification is demonstrated. This result further reinforces the idea that conventional sentiment indicators only reflect information already captured by macroeconomic variables that are accounted for in the baseline model.

This paper relies on Vosen & Schmidt (2011) strategy of GT category selection that compiles a dataset of GT series that best matches the product categories of the BEA's. The predictive performance of the GT indicator can potentially be enhanced by utilising an alternative procedure of category selection. Eventually, working with seasonally adjusted GT variables might be more suitable than working with year-on-year growth rates. The main weakness of transforming GT data into yoy monthly rates is their introduction of overlapping growth rates that can potentially introduce an MA(1) error into the estimation (Vosen & Schmidt, 2011). Nevertheless, accurate seasonal adjustment can only be performed when a vast number of time series observations are available at hand, this is not yet the case with GT data that is still considered to be at its infancy.

Since 2008 Google Trends also provides News Search data whose sample searches are limited to those entered to "news.google.com" or the Google News search function. This dataset provides information on the relative popularity of news articles that include specific keywords. Google News data corresponding to such terms as "recession" and "layoff" can potentially boost the accuracy of the current GT model. The incorporation of this dataset would, however, result in further shrinkage of the sample size. Further analysis can also potentially focus on testing the performance of GT data in times of economic turbulence.

The implications of this paper are significant. The confirmed hypothesis of Granger Causality between GT and consumption suggests that sudden changes in consumption figures correspond to changes in consumption-related GT series. Since consumption resembles the largest portion of the US GDP, GT should be a real-time diagnostic tool for the entire economy.

Economists and policy-makers can employ daily and weekly Google Trends series to detect high-frequency changes in consumption. With further research, Google Trends datasets can become a major indicator in the toolbox of policy-makers that will allow them to be more reactive to sudden economic events. By responding to real shocks to the economy in a timely manner, the policy-makers can better mitigate the long-term consequences of unanticipated dynamics in the economy.

7. References

Ashley, R., Granger, C.W.J. & Schmalensee, R., 1980. Advertising and Aggregate Consumption: An Analysis of Causality. *Econometrica*, 48(5), p.1149.

Askitas, N. & Zimmermann, K.F., 2009. Google econometrics and unemployment forecasting. *Applied Economics Quarterly*, 55, pp.107–120.

Scholkopf, B., Smola, A. & Muller, K.-R., 1999. Kernel principal component analysis. *Advances in kernel methods*, MIT Press, Cambridge, MA, USA, pp.327-352.

Carrière-Swallow, Y. & Labbé, F., 2013. Nowcasting with Google Trends in an Emerging Market. *Journal of Forecasting*, 32(4), pp.289–298.

Carroll, C.D., Fuhrer, J.C. & Wilcox, D.W., 1994. Does consumer sentiment forecast household spending? If so, why? *American Economic Review*, 84(5), pp.1397–1408.

Chen, S.-S., 2005. A note on in-sample and out-of-sample tests for Granger causality. *Journal of Forecasting*, 24(6), pp.453–464.

Choi, H. & Varian, H., 2012. Predicting the Present with Google Trends. *Economic Record*, 88, pp.2–9.

Combes, S. & Bortoli, C., 2016. Nowcasting with Google Trends, the more is not always the better. *Proceedings of the 1st International Conference on Advanced Research Methods and Analytics*.

Croushore, D., 2005. Do consumer-confidence indexes help forecast consumer spending in real time? *North American Journal of Economics and Finance*, 16, pp.435–450.

Diebold, F.X. & Mariano, R.S., 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, pp.253–263.

Eppright, D.R., Arguea, N.M. & Huth, W.L., 1998. Aggregate consumer expectation indexes as indicators of future consumer expenditures. *Journal of Economic Psychology*, 19(2), pp.215–235.

Ettredge, M., Gerdes, J. & Karuga, G., 2005. Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, 48(11), pp.87–92.

Fuhrer, J. C., 1993. What role does consumer sentiment play in the U.S. macroeconomy. *New England Economic Review. January,* pp.32–44

Giacomini, R. & White, H., 2006. Tests of Conditional Predictive Ability. *Econometrica*, 74(6), pp.1545–1578.

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S. & Brilliant, L., 2009. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), pp.1012–1014.

Granger, C.W.J., 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods". *Econometrica*, 37 (3), pp.424–438.

Harvey, D., Leyborne, S. & Newbold, P., 1997. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13, pp.281–291.

Kholodilin, K.A., Podstawski, M. & Siliverstovs, B., 2010. Do Google Searches Help in Nowcasting Private Consumption? A Real-Time Evidence for the US. *SSRN Electronic Journal*.

Ludvigson, S.C., 2004. Consumer confidence and consumer spending. *Journal of Economic Perspectives*, 18, pp.29–50.

McLaren, N. & Shanbhogue, R., 2011. Using Internet Search Data as Economic Indicators. *SSRN Electronic Journal*.

Polgreen, P.M., Chen, Y., Pennock, D.M. & Nelson, F.D., 2008. Using internet searches for influenza surveillance. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 47(11), pp.1443–1448.

Ross, A., 2013. Nowcasting with Google Trends : a keyword selection method. *Fraser of Allander Economic Commentary,* 37 (2). pp.54-64.

Vosen, S. & Schmidt, T., 2011. Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of Forecasting*, 30(6), pp.565–578.

Watson, M.W. & Stock, J. H., 2004. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6), pp.405–430.

Wilcox, J.A., 2007. Forecasting Components of Consumption with Components of Consumer Sentiment. *Business Economics*, 42(4), pp.22–32.

Woo, J. & Owen, A.L., 2019. Forecasting private consumption with Google Trends data. *Journal of Forecasting*, 38(2), pp.81–91.

8. Appendices[18]

Appendix 1 - Description of non-GT variables and summary statistics[19]

Table 5- Description and summary statistics of non-GT variables

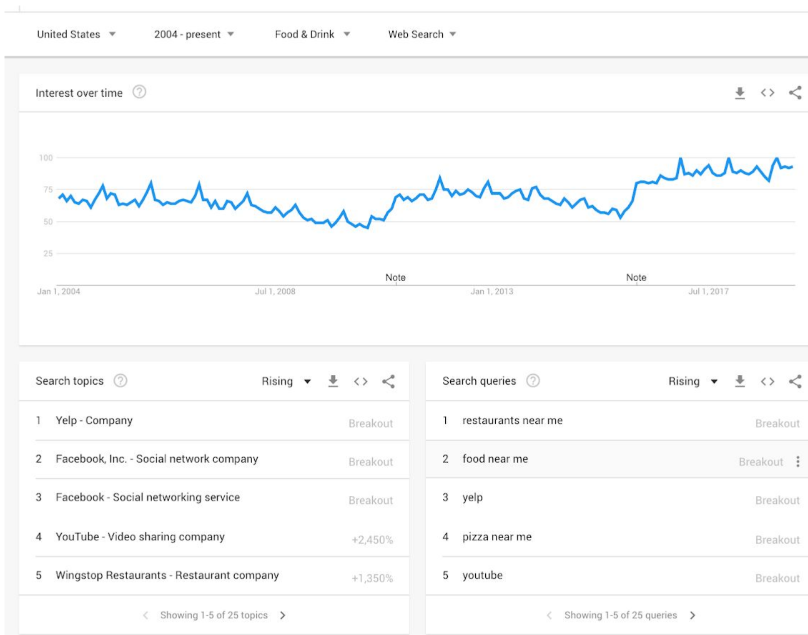
Variable (year-on-year monthly growth rate)	Description	T[20]	μ[21]	σ[22]
Consumption	Real Private Consumption Expenditure (Billions of Chained 2012 Dollars, seasonally adjusted)	166	2.03	1.47
Income	Real Personal Income (Billions of Chained 2012 Dollars, seasonally adjusted)	166	.021	0.021
Smp500	S&P 500 index (not seasonally adjusted)	166	.048	0.19
Tbill	Interest Rates on 3-month Treasury Bills (not seasonally adjusted)	166	-1.70	8.30
CCI[23]	Consumer Confidence Index - An indicator of future dynamics of households' savings and consumption. The survey is based on questions regarding their view about the overall economic stance, unemployment, and their expected financial position (not seasonally adjusted).	166	.0002	0.01
MCSI[24]	Michigan Consumer Sentiment Index - Attempts to find households' sentiment on three things: their own financial position, their sentiment on the short-term and long-term general economic situation. Calculated each month based on 500 telephone calls to a sample of US citizens (not seasonally adjusted).	166	-0.007	.14

[18] Most of the analysis was performed using Python, the relevant packages include Sklearn, Pandas and Numpy: https://scikit-learn.org/stable/, https://pandas.pydata.org/, https://www.numpy.org/
[19] Refer to Appendix 2 for a detailed description of GT data
[20] Number of observations: from January 2005 to November 2018
[21] Sample mean
[22] Sample standard deviation
[23] Source: https://data.oecd.org/leadind/consumer-confidence-index-cci.htm, not seasonally adjusted
[24] Source: http://www.sca.isr.umich.edu/tables.html: index of consumer sentiment table, not seasonally adjusted

- All of the macroeconomic series were extracted from FRED, Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/. For all macroeconomic variables, except for consumption, a transformation was applied to obtain year-on-year growth rates (consumption data was already retrieved in the form of yoy growth rates from FRED).

Appendix 2 - Google Trends[25]

Figure 4 - Google Trends interface



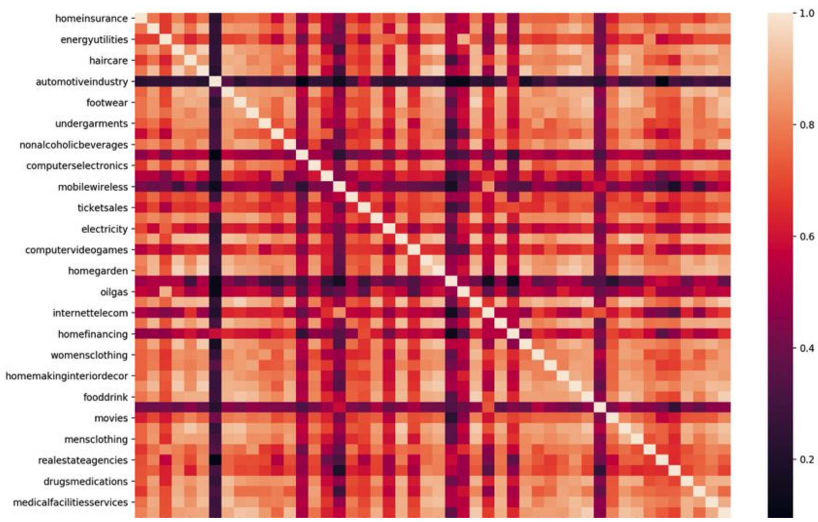
- Figure 4 is a screenshot from the Google Trends website. It illustrates 'interest over time' for the Food & Drink category of Google Searches.
- The 'interest over time' term represents search interest relative to the maximum point on the graph for the given geographic location and time period. For instance, a value of 100 signals peak popularity for a given category, 50 implies that a category is half as popular and so on.
    - Each observation is divided by the overall volume of searches of a given region and the time interval it resembles. The resulting dataset is then normalised to a range of 0 to

[25] Refer to https://trends.google.com/trends/?geo=US for more information

100. GT data is calculated using a sampling method, hence, the results can vary by a few percentage points from one day to another.

- Google Trends categories combine similar terms. They are classified via a natural language classification engine. The procedure of classification is probabilistic since every keyword such as [apple] could be assigned to different categories that can include Computers & Electronics, Food & Drink, and Entertainment.

Figure 5 - Correlation matrix of consumption-related GT categories



Appendix 3 - Principal Component Analysis and Kernel Principal Component Analysis

- Principal Component Analysis[26]
    - In cases where the dimensionality of data is high, the models might suffer from the so-called 'curse of dimensionality' which refers to the inability of a model to generalise well to unseen observations. This happens because when there is a high number of exogenous variables, the specification starts to fit the noise present in the training samples and fails to capture genuine relationships between variables.

[26] Refer to https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html for Sklearn implementation of PCA

- Hence, when confronted with high-dimensional feature spaces, it is reasonable to conduct a dimensionality reduction procedure, e.g. by utilising the method of Principal Component Analysis (PCA).
- PCA transforms the original features to a new set of variables called Principal Components, which are uncorrelated, and which are ordered so that the first few factors preserve most of the variance present in the original feature space.
- Conventional PCA works with zero-centered data, that is $\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i = \mathbf{0}$, where $\mathbf{x}_i$ is the vector of N multivariate samples. PCA diagonalises the covariance matrix, $C = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^\top$, in other words, it provides an eigenvalue decomposition of the covariance matrix of the features: $\lambda\mathbf{v} = C\mathbf{v}$, it can be rewritten as $\lambda\mathbf{x}_i^\top\mathbf{v} = \mathbf{x}_i^\top C\mathbf{v} \quad \forall i \in [1, N]$.

- Kernel Principal Component Analysis (KPCA)[27]
  - Extension of Principal Component Analysis that employs techniques of kernel methods.
  - PCA is a linear combination of the original dimensions, Kernel PCA is a nonlinear function of the original dimensions. When conducting KPCA we end up with a linear combination of non-linear transformations of the original feature space.
  - KPCA also operates on zero-centered data, for this reason before applying KPCA to GT data, Standard Scaling[28] was applied to the feature space in order to transform it to a mean zero and unit variance dataset.

[27] Refer to https://scikit-learn.org/stable/auto_examples/decomposition/plot_kernel_pca.html for Sklearn implementation of KPCA
[28] Refer to https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html for the Standard Scaler implemented in Sklearn

- In contrast to PCA, in KPCA, a non-trivial and arbitrary function Φ is 'selected' that is never computed explicitly. It allows the use of potentially very-high-dimensional Φ's since we never need to transform the original variables to that space. A 'kernel trick' is utilised to avoid tedious computation. In KPCA an N-by-N kernel is created: $K = k(\mathbf{x}, \mathbf{y}) = (\Phi(\mathbf{x}), \Phi(\mathbf{y})) = \Phi(\mathbf{x})^T \Phi(\mathbf{y})$ which resembles the inner product pace of the potentially high-dimensional transformed feature space. The dual form makes it possible for us to construct a version of PCA where we never explicitly solve for the eigenvalues and eigenvectors of the covariance matrix of the Φ based transformation of the original feature space.
- One drawback of KPCA is that when we operate with linear PCA, we can utilise the eigenvalues to sort eigenvectors according to the share of variation captured by them. It is useful for dimensionality reduction purposes. In KPCA, however, since eigenvalues and eigenvectors are not calculated explicitly, this is not possible. That was the main reason why in this paper the predictive KPCA components are selected on the basis of an iterative nowcasting algorithm that is outlined in the main text of the paper.
- When implementing KPCA on the raw GT dataset, a radial basis function (RBF) kernel[29] is applied to the original feature space. The main advantage of RBF over other kernels is its potential to implicitly map every point of the original dataset into an infinite dimensional space which proves to enhance the performance of the nowcasting algorithm the most (compared to other kernels). It provides the necessary flexibility to

[29] Refer to https://en.wikipedia.org/wiki/Radial_basis_function_kernel for further information on the RBF kernel

the transformed feature space to capture a lot of valuable information present in the GT dataset.

Appendix 4 - Rolling window vs. Expanding window Time Series Forecasting

- Rolling window forecasting
    - The procedure starts by selecting the initial set of training observations that a forecasting model would be estimated on and a window size which determines the 'memory' of the forecasting model. The concept can be explained by the means of an example.
    - A rolling window one-month-ahead forecasting procedure can be performed by initially training a specification on 12 months of observations, so from month 1 to month 12 including, and producing a forecast for the 13th observation. On the next iteration of the algorithm, the model would be estimated on the observations ranging from 2nd to 13th, and the prediction would be evaluated for the 14th month. This would imply that the window size is picked to be 12 observations, meaning that every subsequent forecast is evaluated based on a model trained on the 12 past observations.
    - The main weakness of this procedure is that it does not utilise all of the available data when performing a forecast, however, it appears to perform better in cases when the time series data has structural breaks and very old data points no longer resemble the current state of the variable under consideration
- Expanding window forecasting
    - Unlike rolling window forecasting models, an expanding window procedure utilises all of the available past data at the time of the forecast.
    - It proves to work better in cases when the series is relatively stable. It also leads to more parameter stability and accuracy.

Appendix 5 - Augmented Dickey-Fuller Tests for stationarity of non-GT variables[30]

Table 6 - ADF test results (non-GT variables)

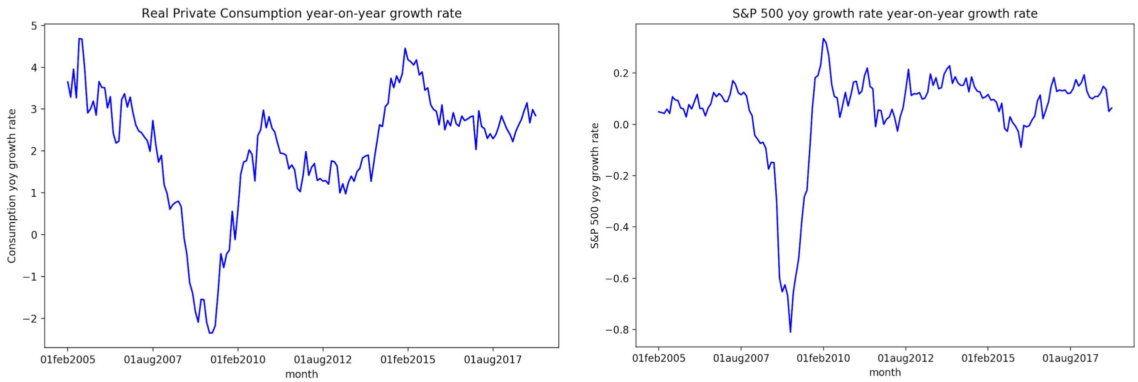
Variable (year-on-year monthly growth rate)	Augmented Dickey-Fuller test model	Test statistic	Critical Value (5%)	Stationary / Non-stationary	Order of integration
Real Private Consumption	Constant, no trend, 1 lag[31]	-1.641	-2.886	Non-stationary	I(1)
Real Personal Income	Constant, no trend, 1 lag	-2.967	-2.886	Stationary	I(0)
S&P 500 index	Constant, no trend, 1 lag	-2.477	-2.886	Non-stationary	I(1)
Interest Rates on 3-month Treasury bills	Constant, no trend	-8.435	-2.886	Stationary	I(0)
CCI	Constant, no trend, lags(3)	-2.984	-2.886	Stationary	I(0)
MCSI	Constant, no trend	-3.412	-2.886	Stationary	I(0)

- According to the ADF tests above, year-on-year growth rates of the S&P 500 index and Real Private Consumption appear to be non-stationary. This seems to go against mainstream macroeconomic literature. Below are the plots of the two variables under question to conduct a further visual investigation.

[30] Since the KPCA procedure is applied to Standard Scaled GT series, by definition the KPCA components that resemble the GT indicator are stationary since the Standard Scaled versions of the original variables have a zero mean and a unit variance

[31] Number of lags is determined by iterative elimination of statistically insignificant lags

Figure 6 - Year-on-year growth rates of Real Private Consumption and the S&P 500 index



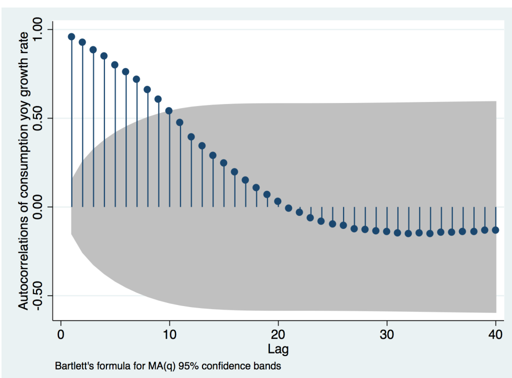
- Visual inspection of Figure 6 suggests that the year-on-year growth rate of Real Private Consumption and S&P 500 index are both stationary processes. The reason why ADF tests of these variables provide misleading results is the financial crisis of 2007-08, both variables plummet during that period. However, if we look at the periods before and after the crisis both indices fluctuate around a constant mean with a relatively constant variance. Not to mention, misleading results of the ADF tests can also be attributed to the lack of data which puts a higher weight on the crisis when determining the stationarity of the processes. Hence, it is reasonable to assume that the above-variables are stationary, which is also in line with mainstream literature on consumption nowcasting.
- ADF tests for the remaining variables suggest that they are stationary, hence, they can be employed in time series analysis since they cannot cause spurious results.

Appendix 6 - Real Private Consumption year-on-year growth rate - baseline specification

- Following the principle of Occam's razor which suggests a poor generalisation of complex models, a decision was made to build a univariate ARMA(p, q) model for the real private consumption.

- Partial Autocorrelation Function on Figure 7 reveals one significant lag term for the yoy growth rate of consumption and a near geometric decay of the autocorrelation function. These two graphs act as strong evidence of an AR(1) process.
- According to Table 7, AR(1) is also the best fit for the consumption yoy growth rate data since it has the lowest value of the Bayesian Information Criterion.

Figure 7 - Autocorrelation function (left-side) and partial autocorrelation function (right-side) of Real Private Consumption year-on-year growth rate



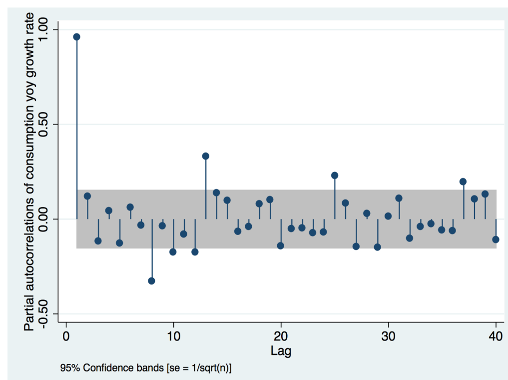


Table 7 - Univariate consumption yoy specifications - Bayesian Information Criterion (BIC)

Model	BIC
AR(1)	**187.1135**
AR(2)	189.8328
MA(1)	452.7677
MA(2)	346.2338
ARMA(1, 1)	190.3098
ARMA(1, 2)	193.3515
ARMA(2,1)	191.9348
ARMA(2,2)	196.3469

Table 8 - Breusch-Godfrey Lagrange Multiplier test for autocorrelation in residuals (null hypothesis of no serial correlation amongst AR (1) model residuals)

Lags(p)	Chi-squared statistic	p- value	Serial Correlation
1	2.298	0.1295	No
2	4.466	0.1072	No
3	4.693	0.1957	No
4	6.796	0.1470	No

- According to Table 8, AR(1) does not reveal serial correlations, meaning that test-statistics obtained from estimations based on the specification are reliable.