

Big Data Team - Research projects

- Traffic flow as an early indicator for GDP growth (and other economic measures)

Edward Rowland

- Extracting data from online CVs

Hazel Martindale

Out[21]: [Click here to toggle on/off the raw code.](#)

Out[1]: [Click here to toggle on/off warnings.](#)

Talk aims

- The work has been largely driven internally
- Promote the work
- Give opportunity for others to collaborate and drive the aims

If you think the work is relevant to your area, please get in touch!

Traffic flow as an early indicator for GDP growth (and other economic & labour market measures)

Talk Structure

1. Background
2. Other work
3. Aims
4. Data
5. Overall Trends
6. Finding the lag
7. Correlations
8. Time series modelling
9. Summary
10. Future work - High Frequency traffic counts

1. Background

Traffic flow is thought to be elastic to a number of economic factors

- GDP: More stuff being made that needs to be moved about
- Salary: More money means people might travel more for work, business and leisure
- Employment: More people in work may mean more commuting
- Inflation: If the cost of living goes up, that might mean less traffic as people cut back on expenses

How is this useful?

Traffic could be used as a proxy for these figures. This could provide the following benefits..

1. More localised estimates
2. Potential early indicator
3. Identify impacts of specific events

Where could this be used? - Some examples

1. More localised estimates

Identify areas of low or negative growth that would be lost if looking at the national figures

2. Potential early indicator

A recession is defined as having two or more quarters of negative GDP growth. The UK won't officially know if it is in recession until around 6-7 months after it started.

3. Identify impacts of specific events

With high frequency data, it could identify the impacts of events (Weather, terrorist attacks, Brexit etc.)

This aim is to find relationships between traffic and economic indicators to see if this could work

2. Other work

Have others looked at this...?

...Yes!

Stats Netherlands

Research using data from an extensive road sensor network in The Netherlands shows correlations with traffic flow and a number of economic measures with a lag of 3 months

Measure	Total Traffic	Cat 1 (< 5.6m) Traffic	Cat 2 (5.6m =< >= 12.2m) Traffic	Cat 3 (> 12.2m) Traffic
Inflation	-0.42	-0.43	- 0.19	-0.43
Unemployment	-0.47	-0.41	-0.55	-0.22
Income	0.74	0.74	0.45	0.65
GDP	0.54	0.63	-0.01	0.70

- Income followed by GDP look to be the strongest correlates
- Larger vehicles (HGVs) correlate most strongly with GDP
- Perhaps some weaker correlations with inflation and Unemployment

We can look and see if this matches with the data used here later

Killan, Ros, "Road Traffic Correlations with Economic Variables: The Big Data Perspective.", 2017, <https://pdfs.semanticscholar.org/1f1d/b563d229bdd4fd8c90ad8dd6c5cd3487f76b.pdf>
(<https://pdfs.semanticscholar.org/1f1d/b563d229bdd4fd8c90ad8dd6c5cd3487f76b.pdf>)

Some issues

Variable	Normally Distributed?	Method
Daily traffic intensity	No	Spearman
Monthly traffic intensity	Yes	Pearson
Quarterly traffic intensity	Yes	Pearson
GDP	Yes	Pearson
Income	No	Spearman
Unemployment	Yes	Pearson
Inflation	No	Spearman

Not convinced on the use of Spearman's

- Shapiro–Wilk test for normality is more (too) sensitive for big data sets
- Income data is likely to be skewed, but hard to tell as histograms aren't given
- No indication they tried transforming the data
- Relationships look to be linear
- Spearman's tends to exaggerate coefficients and p-values compare to Pearson's where the latter is valid so can be used for p-hacking (I don't think this was deliberate here)

Would this change the conclusions? - Probably not hugely, so the work is still useful

Slovenia

- Used traffic flow to estimate GDP figure 45 days before its publication.
- PCA with linear regression to estimate GDP
- Found that Cargo vehicles were the best predictor of GDP, within 1% of the GDP figure

Črt Grahonja, 2018, Use of alternative data sources as flash estimates of economic indicators, European conference on quality in official statistics June 2018:

https://www.q2018.pl/papers-presentations/?drawer=Sessions*Session%2022

(https://www.q2018.pl/papers-presentations/?drawer=Sessions*Session%2022)

Finland

- Uses Company level data and traffic loops to produce two nowcasts and a 16-day backcast of GDP
- Reduce factors by two-step PCA and shrinkage step
- 16 day backcast was as accurate as the first estimate of GDP when comparing both to the revised figure
- Traffic flow and company data were similar when estimating GDP

Henri Luomaranta, 2018, Nowcasting Finnish Real Economic Activity: a Machine Learning Approach, European conference on Quality in official statistics June 2018:
https://www.q2018.pl/papers-presentations/?drawer=Sessions*Session%2022
(https://www.q2018.pl/papers-presentations/?drawer=Sessions*Session%2022)

3. Aims

- Examine the feasibility of applying this work to UK economic and labour market measures
- Looking at correlations between traffic flow and these; mostly replicates the Stats Netherlands approach
- Apply some simple time series models to see if traffic flow improves predictions

4. Data - Traffic flow

Department for Transport data

- Annual average daily flow (AADF) for major and minor roads is used as a measure of traffic flow from 2003 to 2015
- Split into different vehicle categories
- Daily flow is the number of vehicles passing a point on a road on a day. This is averaged across the year to produce the average daily flow
- This measure is based upon approximate 10,000 manual counts per year, between March and October on non-school and public holidays
- These counts are used to estimate AADF figures for major roads
- A representative sample of minor road sites are selected as observations points
- These figures are combined with the change on the previous year to estimate counts for all minor roads

4. Data - Economic and labour market measures

Annual figures are taken to match with traffic flow

- **GDP:** The measure used here is the National GDP growth figure as contained in the UK National Accounts Blue Book.
- **CPIH:** The annual UK Consumer Price index including owner occupied housing costs is used here. Note that this time series only dates to 2005, so no figure is available before this date
- **CPI:** The annual UK Consumer Price index is used here as CPIH is not available before 2005
- **Unemployment:** The seasonally adjusted UK unemployment rate for over 16s is used here
- **Earnings:** Average weekly earnings is the figure used that gives the money paid per week, per job before tax and other deductions to employees in the UK

4. Overall trends in the data (file:///home/eddr/Documents/Projects/report/figures/time_series_multi_metrics.html)

This shows the timeseries of the data to get an overview

Out[11]:

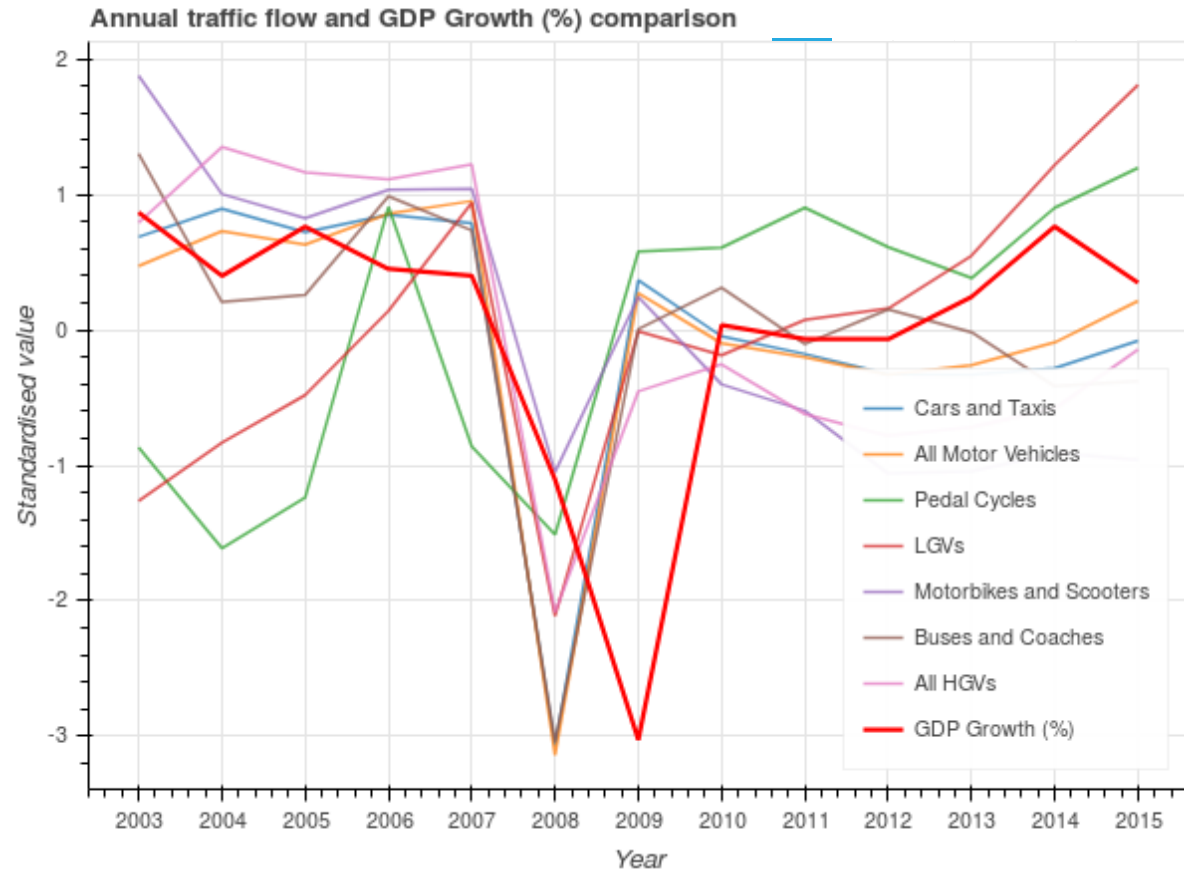
GDP Growth (%)

CPIH

CPI

Change in Unemployment (% pts)

Change in weekly earnings (£)



5. Finding the lag

Method

- Compute the cross-correlation function
- The peak gives the point of maximum similarity between traffic flow and the chosen measures
- This is used to find the time lag between the time series

Results

Variable	Peak/Lag with traffic
GDP	1 year
Income	1 year
Unemployment	5 years*
Inflation	0 years

* little overlap at 1 year indicates negative correlation

GDP Growth

Out[10]:

Buses and Coaches

Pedal Cycles

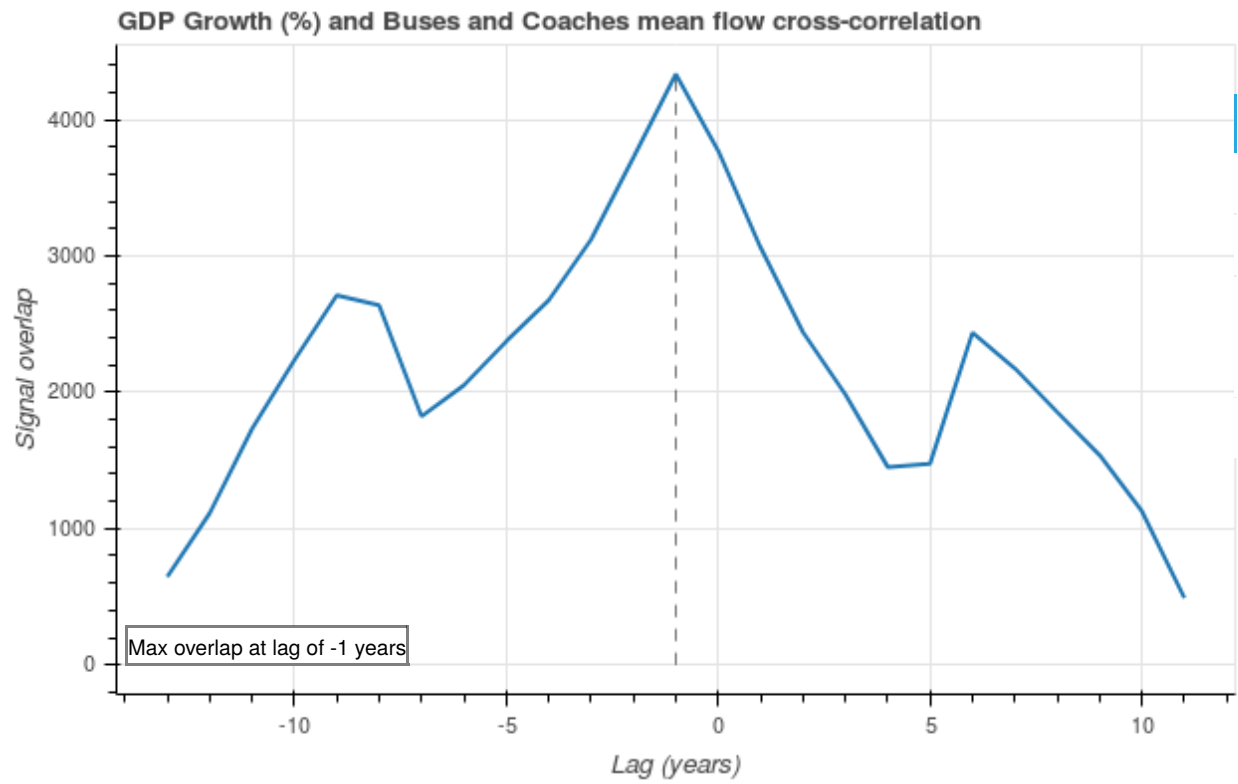
All Motor Vehicles

Motorbikes and Scooters

LGVs

All HGVs

Cars and Taxis



CPIH

Out[12]:

Buses and Coaches

Pedal Cycles

All Motor Vehicles

Motorbikes and Scooters

LGVs

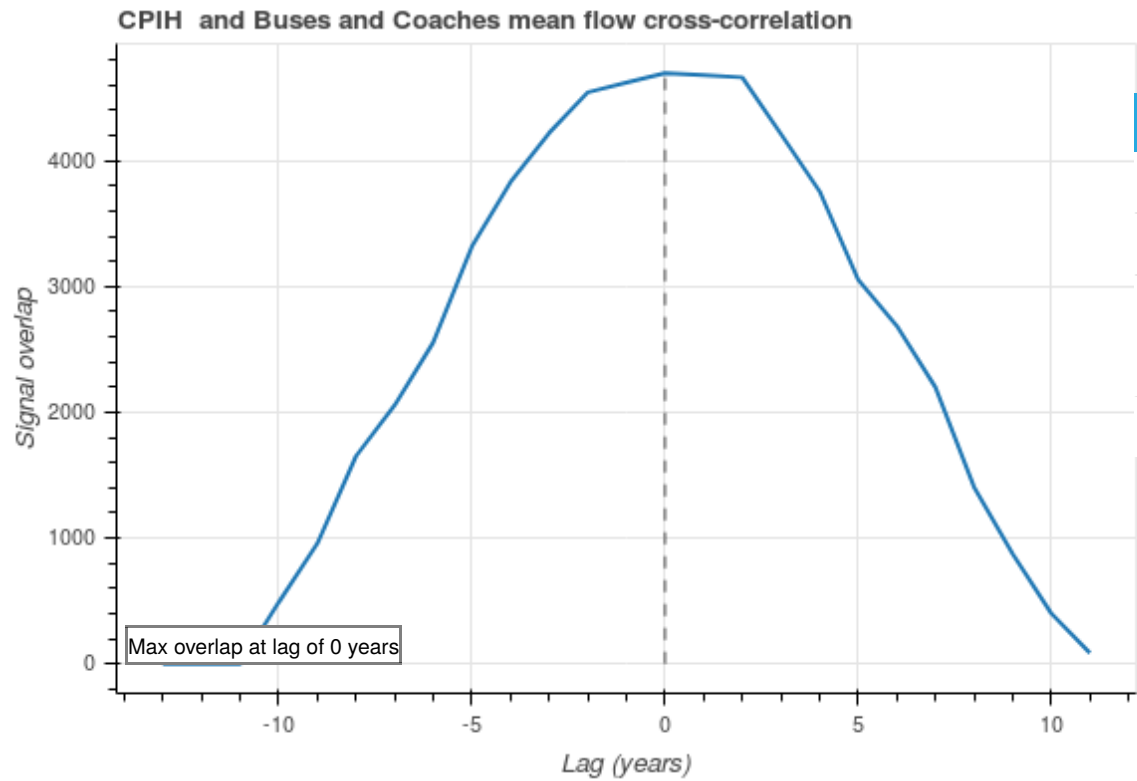
All HGVs

Cars and Taxis

GDP Growth (%)

GDP Growth (%)

GDP Growth (%)



CPI

Out[13]:

Buses and Coaches

Pedal Cycles

All Motor Vehicles

Motorbikes and Scooters

LGVs

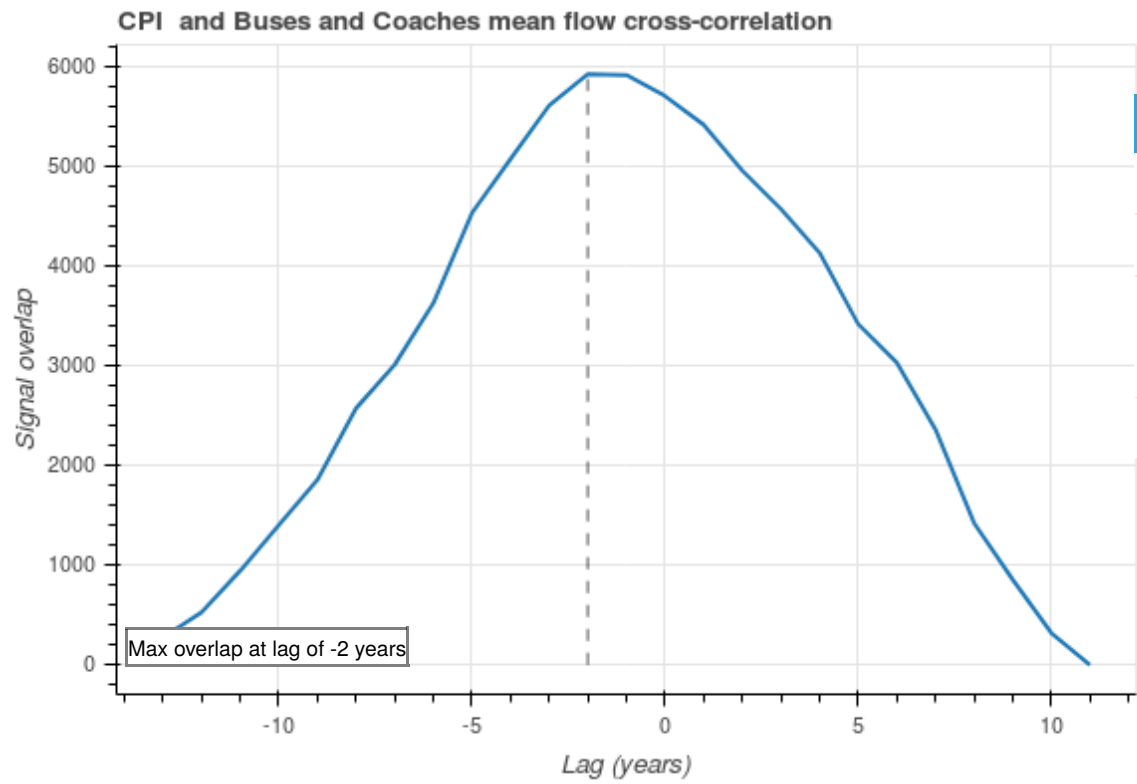
All HGVs

Cars and Taxis

GDP Growth (%)

GDP Growth (%)

GDP Growth (%)



Unemployment

Out[14]:

Buses and Coaches

Pedal Cycles

All Motor Vehicles

Motorbikes and Scooters

LGVs

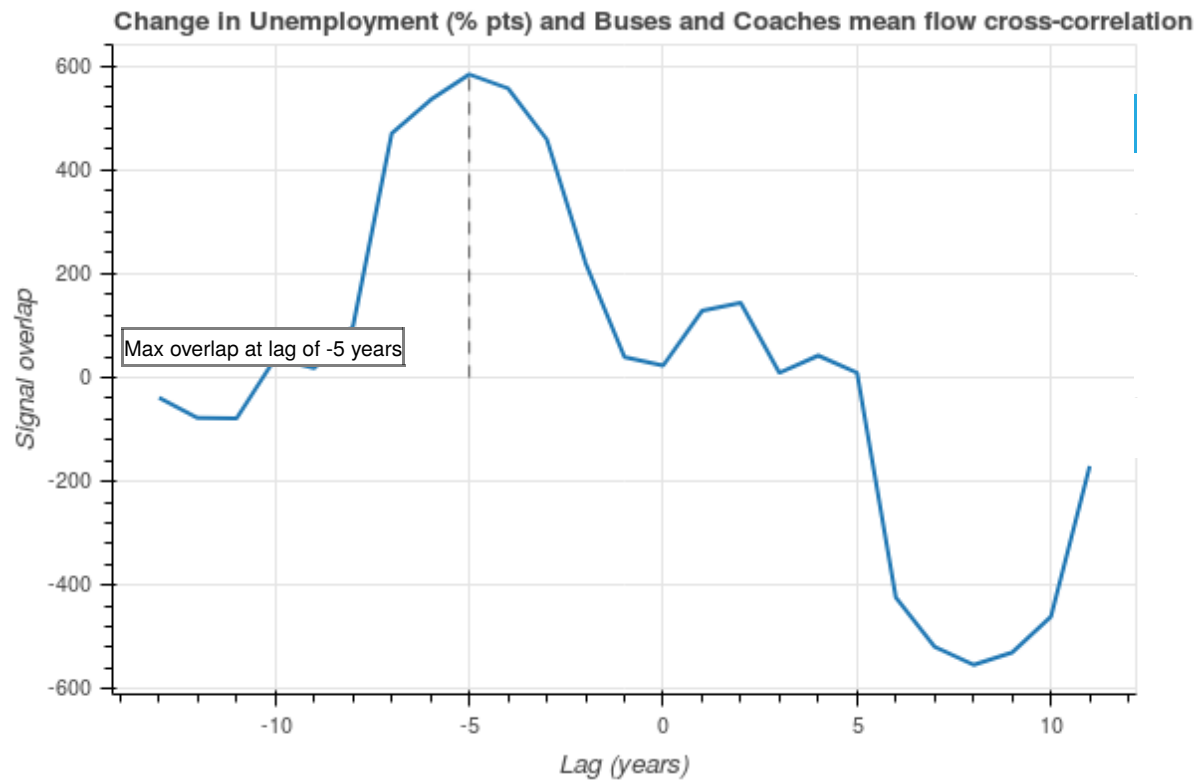
All HGVs

Cars and Taxis

GDP Growth (%)

GDP Growth (%)

GDP Growth (%)



Average Earnings

Out[15]:

Buses and Coaches

Pedal Cycles

All Motor Vehicles

Motorbikes and Scooters

LGVs

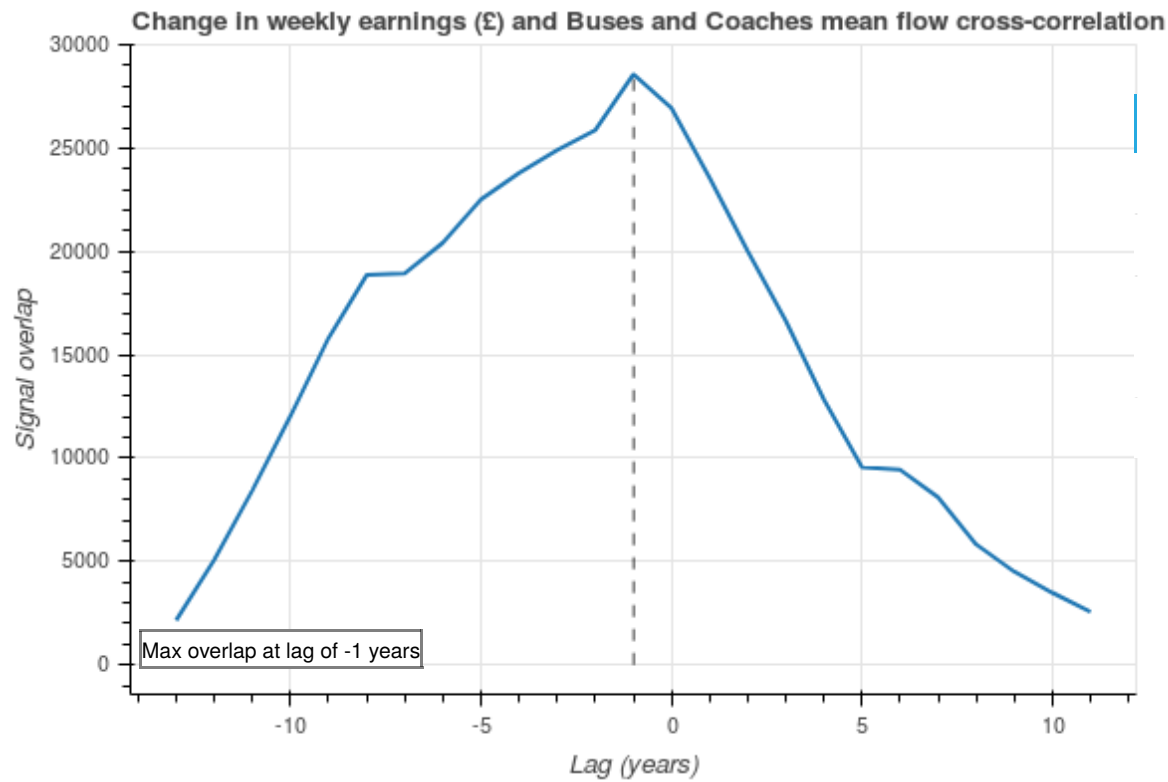
All HGVs

Cars and Taxis

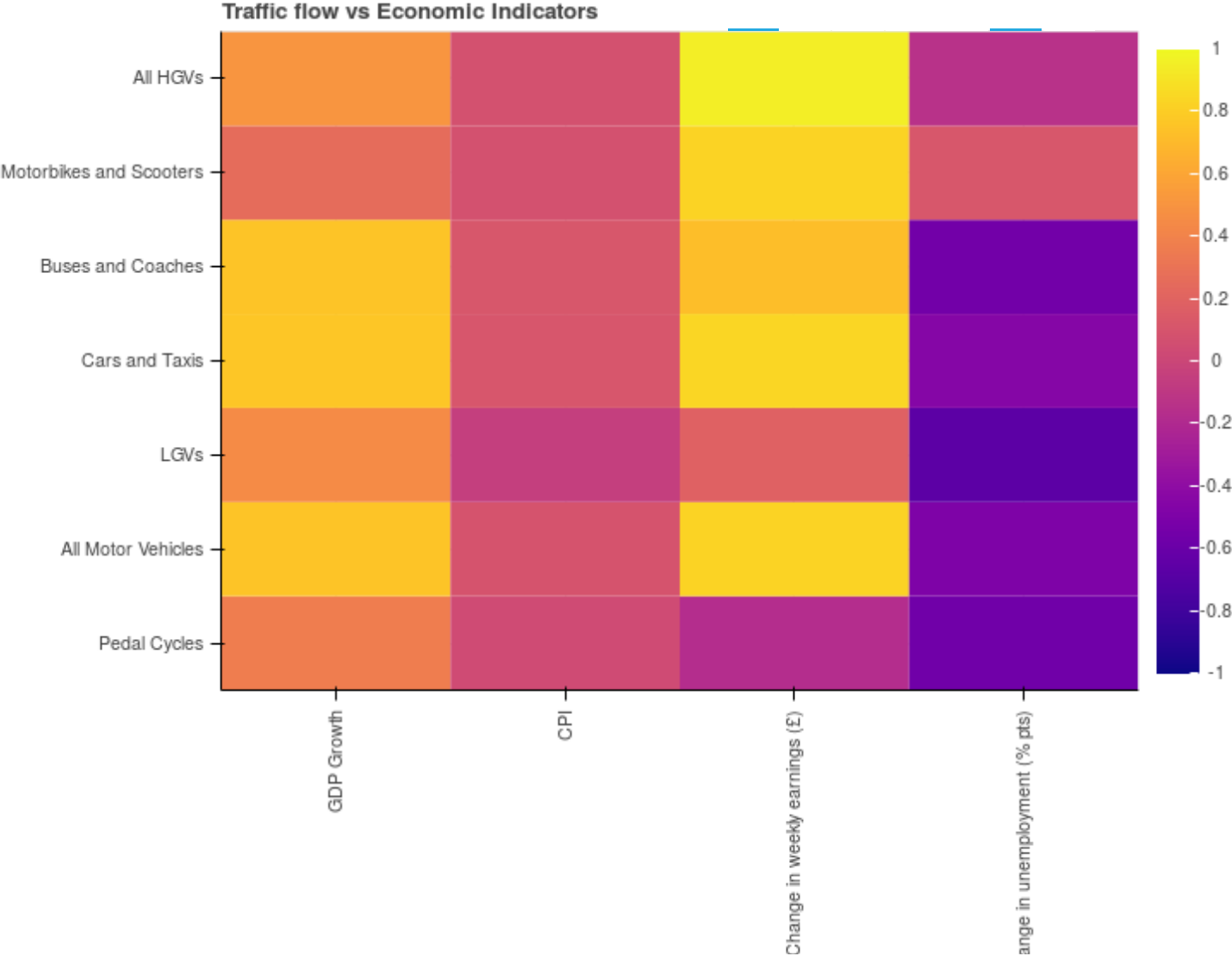
GDP Growth (%)

GDP Growth (%)

GDP Growth (%)



Out[21]:



Coefficients

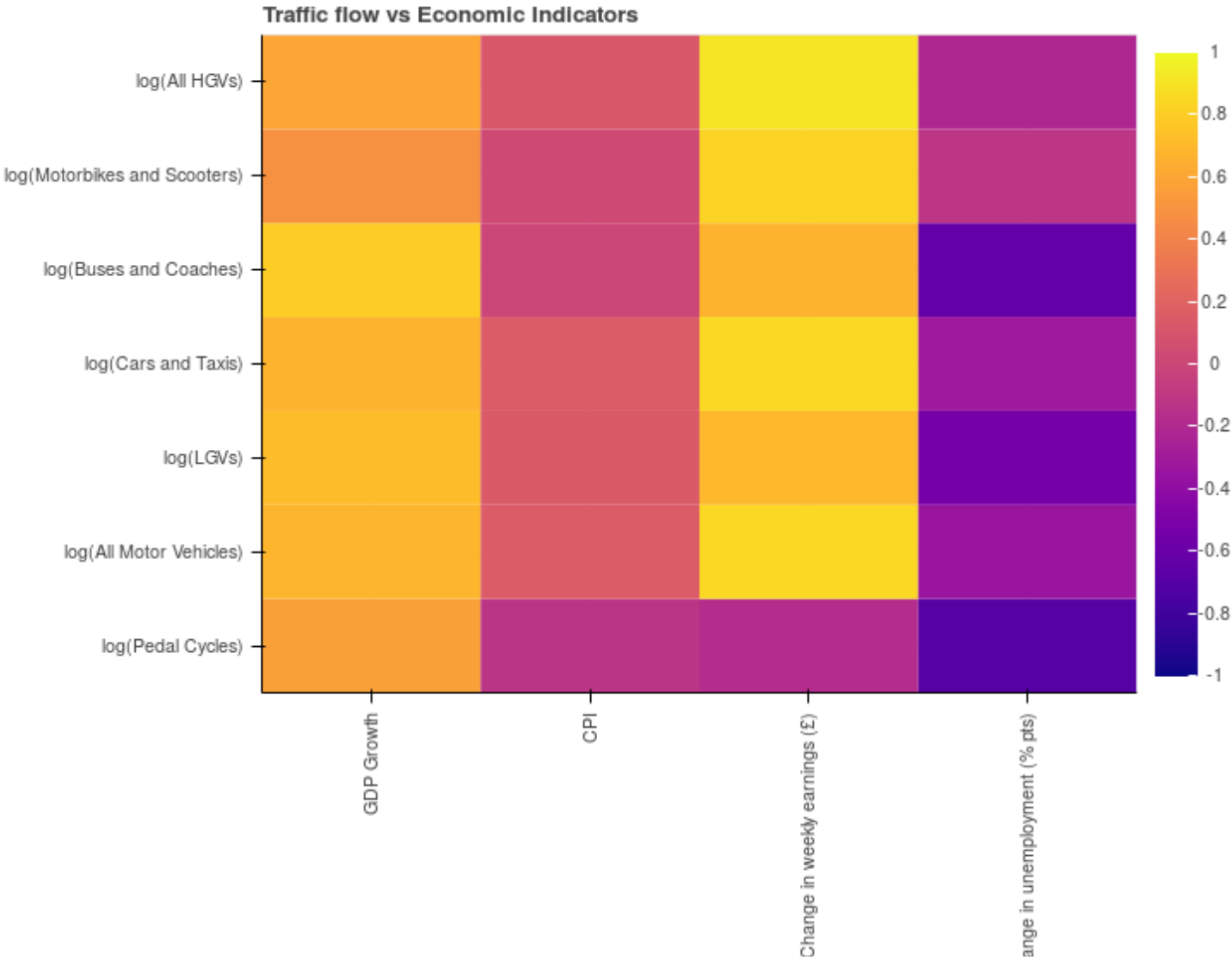
Economic indicator	CPI	Change in unemployment (% pts)	Change in weekly earnings (£)	GDP Growth
Vehicle type				
All HGVs	0.077045	-0.145477	0.946293	0.507384
All Motor Vehicles	0.088426	-0.494078	0.826505	0.757739
Buses and Coaches	0.111941	-0.559256	0.718989	0.756430
Cars and Taxis	0.104417	-0.455441	0.837060	0.765201
LGVs	-0.048951	-0.670309	0.183337	0.447924
Motorbikes and Scooters	0.073761	0.116624	0.825611	0.255856
Pedal Cycles	0.036957	-0.570229	-0.177683	0.366898

Stats Netherlands

Measure	Total Traffic	Cat 1 (< 5.6m) Traffic	Cat 2 (5.6m =< >= 12.2m) Traffic	Cat 3 (> 12.2m) Traffic
Inflation	-0.42	-0.43	- 0.19	-0.43
Unemployment	-0.47	-0.41	-0.55	-0.22
Income	0.74	0.74	0.45	0.65
GDP	0.54	0.63	-0.01	0.70

Log transformed traffic flow

Out[20]:



6. Correlations - summary

- Weekly earnings has the strongest correlation, similar to The Netherlands
- GDP shows some strong correlations, though not with HGVs like in The Netherlands
- Unemployment shows weaker correlations, similar to The Netherlands
- No correlation with inflation

6. Correlations - caveats

1. Small dataset
2. Not clear if variables are normally distributed
3. Correlations for GDP are dependent on outliers (i.e may not exist)
4. Time series only contains one recession
5. Not clear how recession events change distribution of variables

7. Time series models

Approach

1. Try some basic Auto-regressive (AR) models
 - These contain one variable, where you are trying to predict future values from past (lagged) values
 - These shouldn't work well, otherwise it would be easy to predict GDP etc!
2. Add in All Vehicles variable in a Vector AR (VAR) model to predict the economic variable
 - If traffic flow is a good predictor, then this should give a better estimate of GDP

Auto and partial correlations

Out[22]:

GDP Growth

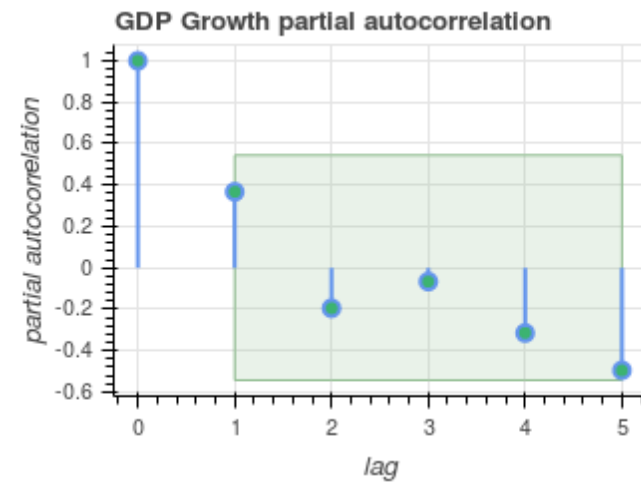
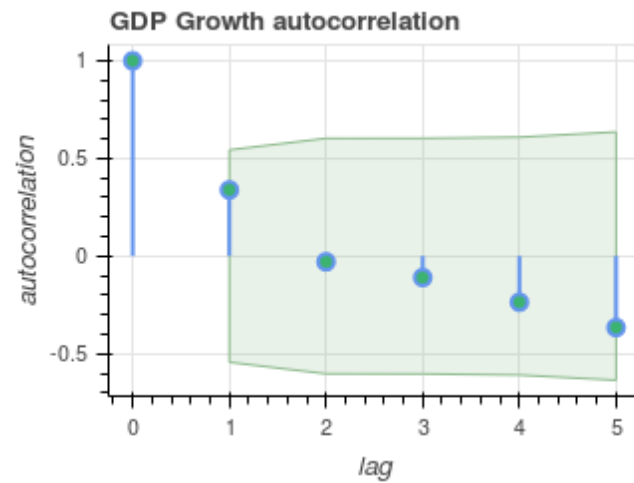
Change in weekly earnings (£)

Change in unemployment (% pts)

CPIH

CPI

All Motor Vehicles



The models

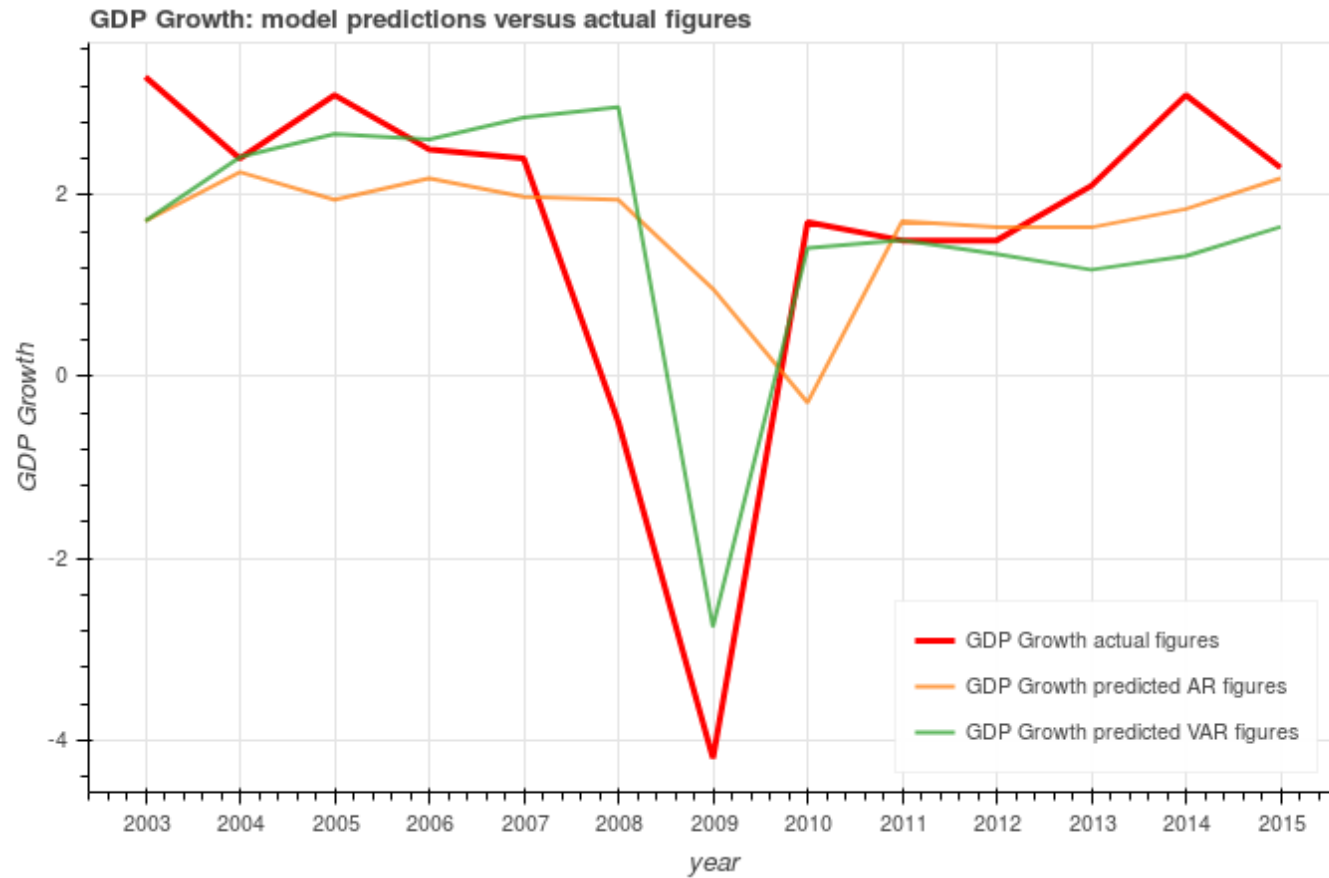
Out[24]:

GDP Growth

Change in weekly earnings (£)

Change in unemployment (% pts)

CPI



7. Time series models - caveats

1. Small timeseries
2. In-sample predictions (predicting the data it has been fitted with)
3. Time series' are not stationary - due to the recession
4. Does give some indication that this could work (with better data and methods)

8. Overall Summary

- Evidence that traffic flow can be used as an early indicator for economic measures
- Need more data
- Better data will allow for better methods
- Still, more data will not solve all our problems
- We need to test these methods' robustness even with new data

Speaking of better data...

9. Future work - High frequency traffic flow data

Highways England provides an open dataset containing traffic flow counts at 15 minute intervals for Motorways and Major A-Roads across England.

- The Data Engineering team in DaaS have scraped this and ingested into DAP-E
 - Over 2 million separate files
 - Over 200Gb of data
- Really exciting high frequency dataset - ideally suited for this work
 - Replicate these correlations with daily data
 - Look at the approaches used in Slovenia and Finland
 - Develop some ideas of our own

Thank you for listening, Questions?

- Contact: edward.rowland@ons.gov.uk
- Location: Newport 1.156

Analysis

A bit rough plus some visualisations may not work as I update/break things periodically - If you want to look at something that is broken, please email me

- GitHub: [https://github.com/ONSBigData/traffic as early indicator](https://github.com/ONSBigData/traffic%20as%20early%20indicator)
([https://github.com/ONSBigData/traffic as early indicator](https://github.com/ONSBigData/traffic%20as%20early%20indicator))
- MyBinder (run code without python/jupyter: [https://mybinder.org/v2/gh](https://mybinder.org/v2/gh/ONSBigData/traffic%20as%20early%20indicator/master)
[/ONSBigData/traffic as early indicator/master](https://mybinder.org/v2/gh/ONSBigData/traffic%20as%20early%20indicator/master) ([https://mybinder.org/v2/gh](https://mybinder.org/v2/gh/ONSBigData/traffic%20as%20early%20indicator/master)
[/ONSBigData/traffic as early indicator/master](https://mybinder.org/v2/gh/ONSBigData/traffic%20as%20early%20indicator/master))