ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH

Office for National Statistics

Can sectoral output and trade data improve early estimates of consumption and investment? A preliminary investigation

Rory Macqueen, Ana Beatriz Galvão and Stephen Wright

ESCoE Discussion Paper 2023-11

June 2023

ISSN 2515-4664

DISCUSSION PAPER

Can sectoral out and trade data improve early estimates of consumption and investment? A preliminary investigation

Rory Macqueen, Ana Beatriz Galvão, and Stephen Wright

ESCoE Discussion Paper No. 2023-11
June 2023

Abstract

Input-output tables, currently published by the ONS with a lag of at least four years (and distinctly longer in earlier data), provide an accounting relation that decomposes final expenditure components into the contribution of sectoral outputs and imports. In this paper, we report on a preliminary investigation into whether past values of the associated weighting matrices can be used to nowcast real growth of Consumers' Expenditure and Gross Fixed Capital Formation (GFCF) from timely data on quarterly growth rates of sectoral GVA and aggregate imports. For 2010-2019, we find a stark, puzzling, contrast. The nowcasts for GFCF predict roughly as well as the ONS's first estimates of GFCF when considered as forecasts of the latest data, but the nowcasts have close to zero predictive value for consumption: the implied predictions are excessively volatile and almost uncorrelated with actual consumption growth on the latest data. We conduct a preliminary investigation of the source of these nowcast errors. Further restrictions on the nowcasting system and exploitation of other existing datasets to improve nowcast accuracy are proposed.

Keywords: consumption investment; nowcasting; input-output models

JEL classification: C53, C67

Rory Macqueen, Birkbeck, University of London
rorymacqueen81@gmail.com

Published by:
Economic Statistics Centre of Excellence
King's College London
Strand
London
WC2R 2LS
United Kingdom
www.escoe.ac.uk

Can sectoral output and trade data improve early estimates of consumption and investment? A preliminary investigation[1]

Rory Macqueen,[2 3] Ana Beatriz Galvão,[4] and Stephen Wright[3]

June, 2023

Abstract

Input-output tables, currently published by the ONS with a lag of at least four years (and distinctly longer in earlier data), provide an accounting relation that decomposes final expenditure components into the contribution of sectoral outputs and imports. In this paper, we report on a preliminary investigation into whether past values of the associated weighting matrices can be used to nowcast real growth of Consumers' Expenditure and Gross Fixed Capital Formation (GFCF) from timely data on quarterly growth rates of sectoral GVA and aggregate imports. For 2010-2019, we find a stark, puzzling, contrast. The nowcasts for GFCF predict roughly as well as the ONS's first estimates of GFCF when considered as forecasts of the latest data, but the nowcasts have close to zero predictive value for consumption: the implied predictions are excessively volatile and almost uncorrelated with actual consumption growth on the latest data. We conduct a preliminary investigation of the source of these nowcast errors. Further restrictions on the nowcasting system and exploitation of other existing datasets to improve nowcast accuracy are proposed.

Keywords: consumption investment; nowcasting; input-output models
JEL codes: C53, C67

This research has been funded by the Office for National Statistics as part of the research programme of the Economic Statistics Centre of Excellence (ESCoE).

[1] The authors would like to thank Eric Crane, Sanjiv Mahajan and Andrew Walton from the Office for National Statistics, and Nick Vaughan, formerly of the Office for National Statistics.

[2] ESCoE, corresponding author
[3] Birkbeck, University of London
[4] Warwick Business School

1. Introduction

The Office for National Statistics (ONS) estimates UK GDP measured in three ways: by sector (GDP output, or GDPO), by expenditure category (GDPE) and by income (GDPI). At the early stages of data collection, GDPO is the measure for which there is most comprehensive information available (McLaren, 2022) and, since 2018, estimates of growth rates for each GDPO sector have been published at monthly frequency.

GDPE is first published at quarterly frequency with a lag of around 40 days from quarter-end, coincident with initial GDPO estimates for the same quarter, and separated by a published 'statistical discrepancy'. Both measures are subject to repeated revisions over subsequent years: for a similar sample period, Galvão & Macqueen (2023, forthcoming) document the standard deviation of revisions to investment growth of 2.2% and consumption growth of 0.55%.

After GDPO and GDPE (and GDPI) are 'balanced' by reducing the statistical discrepancy to zero in the annual Blue Book, input-output tables are constructed, which consist of a series of matrices populated by cross-relationships between their constituent parts i.e. for each category of expenditure (final use), the ONS estimates in which industries of the British economy – or abroad – the value of the goods and services purchased was added. Around four years after the fact, therefore, at annual frequency, the value of goods and services in consumption[5] and Gross Fixed Capital Formation (GFCF) can be related precisely to the contribution of sectoral domestic value added and imports.

We use the matrices obtained from the relationships in earlier years' input-output tables to estimate latest quarter consumption and GFCF growth rates using the same quarter's sectoral output growth and import growth. This pseudo-out-of-sample exercise, which effectively assumes a Leontief economy in the short term, yields GFCF nowcasts that can compete with early ONS estimates in predicting the most recent vintage of ONS data for the same quarters; but the method essentially fails to predict consumption growth at all. One clear feature is that the forecasts for consumption are demonstrably inefficient, excessively volatile, and correlated both with the target variable and the nowcast errors on other variables, suggesting a number of areas for refinement.

A key question is whether the shifting sectoral weights in consumption and GFCF make the use of earlier years' input-output tables a major disadvantage to real-time nowcasting (matrix instability). We consider this and other potential sources of nowcast errors, including:

- The use of matrices constructed from nominal data to predict real variables;
- The use of matrices constructed from annual data to predict quarterly variables;
- Measurement error in early vintages of GDPO data;
- An implicit assumption that changes in taxes minus subsidies on expenditure components does not contribute to nowcasting the component (i.e. that GVA growth = GDP growth);
- GDPO data in Blue Books not matching GDPO data in the corresponding input-output table prior to 2017.

Sector nowcasts are not constrained to sum to any estimate of total GDP, given the bottom-up nature of this approach to nowcasting, as each variable's nowcast is a linear combination of the input variables. Though the results are not documented here, our approach also yields nowcasts for the remaining constituent parts of GDPE, including government and exports. One possible explanation for the failure to predict consumption is that we are not, at this stage, taking into

[5] Consumption by households and non-profit institutions serving households: henceforth just 'consumption'.

account the correlation structure in the prediction errors for different components of TFE. Given that timely data is available on government consumption and exports in particular, so that prediction errors for these can be observed at the time nowcasts are made, future research is likely to find that adding further constraints improves consumption and GFCF nowcasts further. We also consider other possible improvements to the preliminary methods described here in our Conclusion.

Section 2 introduces the methodology, the Office for National Statistics (ONS) input-output tables and the data used to form the nowcasts. Section 3 reports results, Section 4 concludes, and Section 5 contains suggestions for further research.

2. Methodology

The model

The national accounting identity between nominal GDPO and GDPE for year T can be expressed by the following:

$$\boldsymbol{y^e}_{\boldsymbol{T}} \equiv \boldsymbol{y^o}_{\boldsymbol{T}} \, \boldsymbol{R_T} \, \boldsymbol{C_T} + \boldsymbol{m_T} \, \boldsymbol{Q_T} + \boldsymbol{T_T} \qquad (1)$$

or, simplified due to limited data on $\boldsymbol{m_T}$ (discussed in the Conclusion):

$$\boldsymbol{y^e}_{\boldsymbol{T}} \equiv \boldsymbol{y^o}_{\boldsymbol{T}} \, \boldsymbol{R_T} \, \boldsymbol{C_T} + \boldsymbol{p_T} \, M_T + \boldsymbol{T_T} \qquad (2)$$

where

$\boldsymbol{y^e}_{\boldsymbol{T}}$= a 1x6 row vector of the expenditure aggregates comprising Total Final Expenditure (GDPE plus imports) in year T;
$\boldsymbol{y^o}_{\boldsymbol{T}}$ = a 1x20 vector of Gross Value Added in each domestic top-level SIC sector A-T in year T;[6]
$\boldsymbol{R_T}$ = a 20x20 diagonal matrix whose diagonal elements are the ratios between GVA by sector and GVA by product in year T;
$\boldsymbol{C_T}$ = a 20x6 matrix whose element c_{ij} corresponds to the domestically added value of products in sector (row) i which have their final use in expenditure aggregate (column) j for year T – for example: the expenditure in pounds sterling of consumption on agriculture products in 2018 – divided by the total GVA for products in sector (row) I;
$\boldsymbol{m_T}$ = a 1x20 vector of the value of imported products categorised by SIC industrial sector A-T in year T;
$\boldsymbol{Q_T}$ = a 20x6 matrix, the imports counterpart to $\boldsymbol{C_T}$, whose element q_{ij} corresponds to the imported value of products in sector (row) i which have their final use in expenditure aggregate (column) j for year T, divided by the total GVA for products in sector (row) i;
$\boldsymbol{p_T}$ = a 1x6 vector whose elements, which sum to 1, represent the share of total imports (whether consumed directly or indirectly) finding their final use in the corresponding final expenditure aggregate to $\boldsymbol{y^e}_{\boldsymbol{T}}$ in year T;
M_T = total imports in year T;
$\boldsymbol{T_T}$ = a 1x6 vector, each of whose elements is the total taxes minus subsidies on products in each expenditure category in year T.

Input-output tables, published now on an annual basis with a lag of four years, show how this identity holds for the vintage of $\boldsymbol{y^e}_{\boldsymbol{T}}$, $\boldsymbol{y^o}_{\boldsymbol{T}}$ and M_T data in the same Blue Book as used to estimate $\boldsymbol{R_T}$, $\boldsymbol{C_T}$and $\boldsymbol{p_T}$.

[6] SIC 2007 industrial classifications.

In this paper we are concerned with two of the elements of $\boldsymbol{y^e}$: consumption and GFCF. Their quarter-on-quarter (time t) growth rates, along with those of the other elements of $\boldsymbol{y^e}$, are estimated using the following equation:

$$\widehat{\widehat{\boldsymbol{y}}^{\boldsymbol{e}}}_{\boldsymbol{t}} = \widehat{\widehat{\boldsymbol{y}}^{\boldsymbol{o}}}_{\boldsymbol{t}}\, \boldsymbol{R_S}\, \boldsymbol{K_S} + \boldsymbol{\pi_S}\, \widehat{M}_t \qquad (3)$$

where

$\widehat{\widehat{\boldsymbol{y}}^{\boldsymbol{e}}}_{\boldsymbol{t}}$= a 1x6 vector of nowcasts of the quarterly growth rates of elements of $\boldsymbol{y^e}_t$;
$\widehat{\widehat{\boldsymbol{y}}^{\boldsymbol{o}}}_{\boldsymbol{t}}$ = a 1x6 vector of estimated GVA growth in each domestic industrial sector in the latest vintage of data available for quarter t;
$\boldsymbol{R_S}$ = a 20x20 diagonal matrix whose diagonal elements are the ratios between GVA by industrial sector and GVA by product in year S, which is always at least four years earlier than the year containing quarter t, owing to the time taken to produce the input-output tables;
$\boldsymbol{K_S}$ = a 20x6 matrix whose element c_{ij} corresponds to the share of aggregate (column) j expended on products in sector (row) i for year S – for example: the expenditure in pounds sterling of consumption on agriculture products in 2018 – divided by the total GVA for expenditure aggregate (column) j;
$\boldsymbol{\pi_S}$= a 1x6 vector, each of whose elements represents the share made up by imports in the corresponding expenditure aggregate in year S;
$\widehat{M}_t$ = estimated imports in the latest vintage of data available for quarter t.

Circumflexes on vectors signify that data is subject to revision and will not match the vintage used to construct $\boldsymbol{R_S}$, $\boldsymbol{K_S}$ and $\boldsymbol{\pi_S}$.

Note the important difference between $\boldsymbol{C_T}$ and $\boldsymbol{K_S}$, and between $\boldsymbol{p_T}$ and $\boldsymbol{\pi_S}$. The elements of the first matrix in each case represent the proportion of value added in a sector (or abroad) which is attributed to each expenditure category; each row sums to 1. The elements of the second in each case represent the proportion of value in each expenditure category – including taxes minus subsidies in the case of Identities 1 and 2 but not Equation 3 – which was added in each sector or abroad; each column of the matrix formed from vertically concatenating $\boldsymbol{K_S}$ and $\boldsymbol{\pi_S}$ sums to 1. Both are derived from the same original input-output table in each year (see Section 'Input-output tables' for more detail).

$$\begin{bmatrix} & \boldsymbol{G} & \boldsymbol{C} & \dots & \boldsymbol{X} \\ \boldsymbol{A} & G_{1\,1} & G_{1\,2} & \cdots & G_{1\,6} \\ \boldsymbol{B} & G_{2\,1} & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{T} & G_{20\,1} & \vdots & \vdots & G_{20\,6} \\ \boldsymbol{Imports} & G_{21\,1} & \cdots & \cdots & G_{21\,6} \end{bmatrix}$$

20x6 'GVA by final uses' matrix in GBP (current prices) vertically concatenated with Imports totals (1x6) from 'Primary input by final use' matrix in GBP (current prices)

$$\begin{bmatrix} \frac{G_{1\,1}}{\sum_{i=1}^{21} G_{i\,1}} & \frac{G_{1\,2}}{\sum_{i=1}^{21} G_{i\,2}} & \cdots & \frac{G_{1\,6}}{\sum_{i=1}^{21} G_{i\,6}} \\ \frac{G_{2\,1}}{\sum_{i=1}^{21} G_{i\,1}} & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{G_{20\,1}}{\sum_{i=1}^{21} G_{i\,1}} & \vdots & \vdots & \frac{G_{20\,6}}{\sum_{i=1}^{21} G_{i\,6}} \\ \frac{G_{21\,1}}{\sum_{i=1}^{21} G_{i\,1}} & \cdots & \cdots & \frac{G_{21\,6}}{\sum_{i=1}^{21} G_{i\,6}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{K_S} \\ \boldsymbol{\pi_S} \end{bmatrix}$$

N.B. Columns sum to 1.

$$\begin{bmatrix} \frac{G_{1\,1}}{\sum_{j=1}^{6} G_{1\,j}} & \frac{G_{1\,2}}{\sum_{j=1}^{6} G_{1\,j}} & \cdots & \frac{G_{1\,6}}{\sum_{j=1}^{6} G_{1\,j}} \\ \frac{G_{2\,1}}{\sum_{j=1}^{6} G_{2\,j}} & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{G_{20\,1}}{\sum_{j=1}^{6} G_{20\,j}} & \vdots & \vdots & \frac{G_{20\,6}}{\sum_{j=1}^{6} G_{20\,j}} \end{bmatrix} = [\boldsymbol{C_T}]$$

N.B. Rows sum to 1.

Taxes and subsidies

Consumption and GFCF measured at purchasers' prices include taxes minus subsidies on products, the so-called 'basic price adjustment': $\boldsymbol{T_T}$ in Identities 1 and 2. When estimating quarterly growth rates for consumption and GFCF, the contribution from taxes minus subsidies is assumed to be zero. This is the same assumption which is used when proxying GDP growth with GVA growth, as the ONS does in its monthly GDP releases. An official statistician may be privy to policy or HMRC information on taxes minus subsidies but, other than at times of extreme economic instability, such as the early days of the Covid-19 pandemic, taxes minus subsidies' contribution to consumption or GFCF growth are unlikely to be either significant or systematically forecastable by anyone else.

Input-output tables

Input-output tables are now published by the ONS at annual frequency with a significant lag from the end of the reference year: most recently, at the time of writing in early 2023, the input-output tables for 2018 were published on 1st April 2022 and used data from Blue Book 2021 (published October 2021). This release has only occurred annually since 2013, before which tables used in this paper were published in 1995, 2005 and 2010.

For our out-of-sample nowcasting exercises, matrices are taken strictly from the latest input-output table that would have been available at the time, which generally means a lag of four years (e.g. when 2022Q1 was first estimated in May 2022, the input-output table for 2018 had just been published). There is a long lag in the early years of the exercise, due to the less frequent publication of input-output tables at the time. Indeed, this may disadvantage our nowcasts for the early years of the nowcasting period, given the extra-long lag (see Appendix A) between year *S* and quarter *t*. Appendix A also lays out the correspondence between data series and vintages, making use of the assumption that both initial GDPE and GDPO estimates were published around 40 days after quarter-end throughout the test period, as is the current practice.

Figures 1 and 2 illustrate the industrial composition of consumption and GFCF expenditure in each year, corresponding to the respective column sum of the matrix formed by vertically concatenating $\boldsymbol{K_S}$ and $\boldsymbol{\pi_S}$. The greater the shifts between sectors here, the greater the challenges faced by the nowcaster who only has matrices from four years ago available to them.

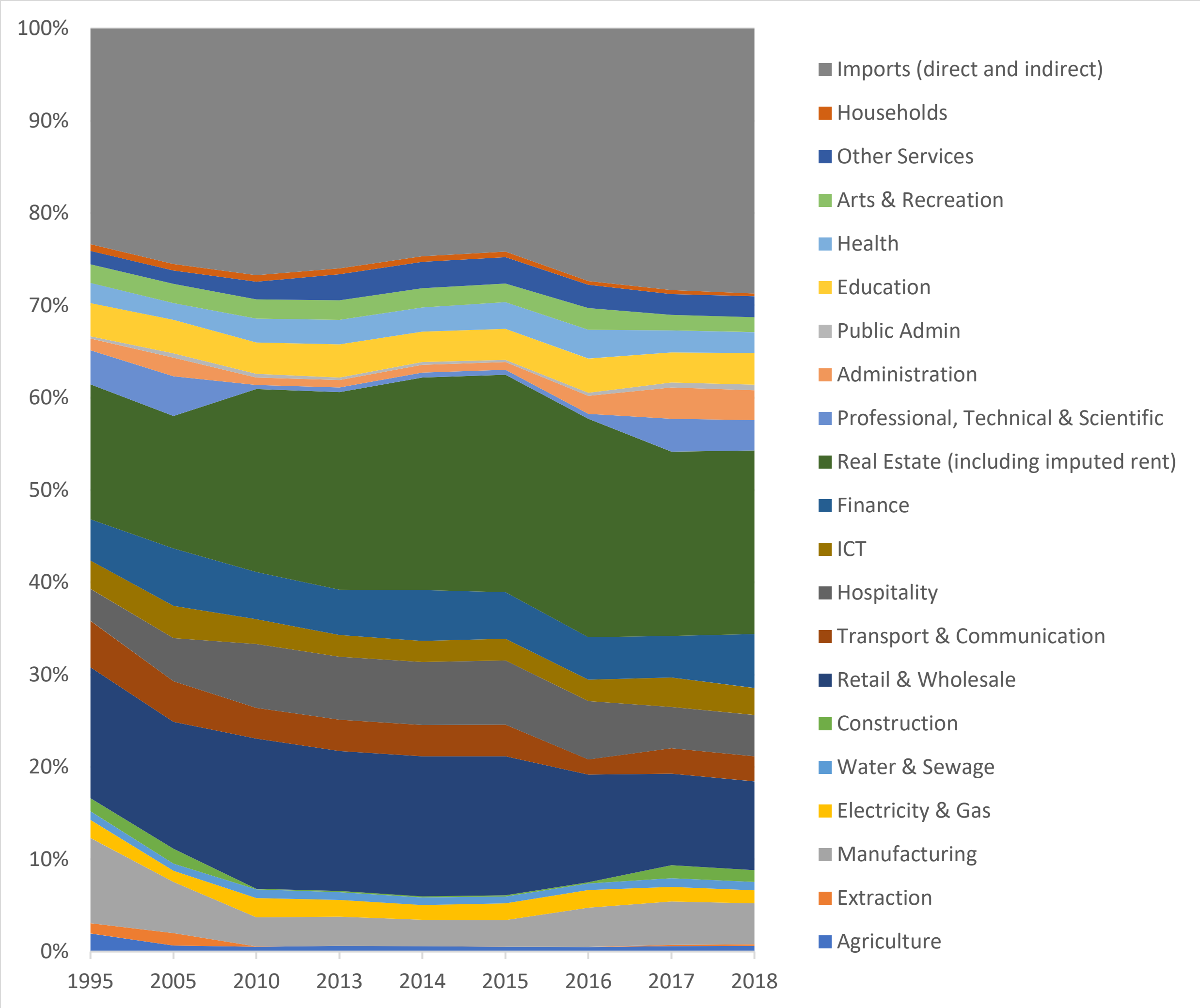

Figure 1: Source of GVA in UK consumption expenditure according to ONS input-output tables

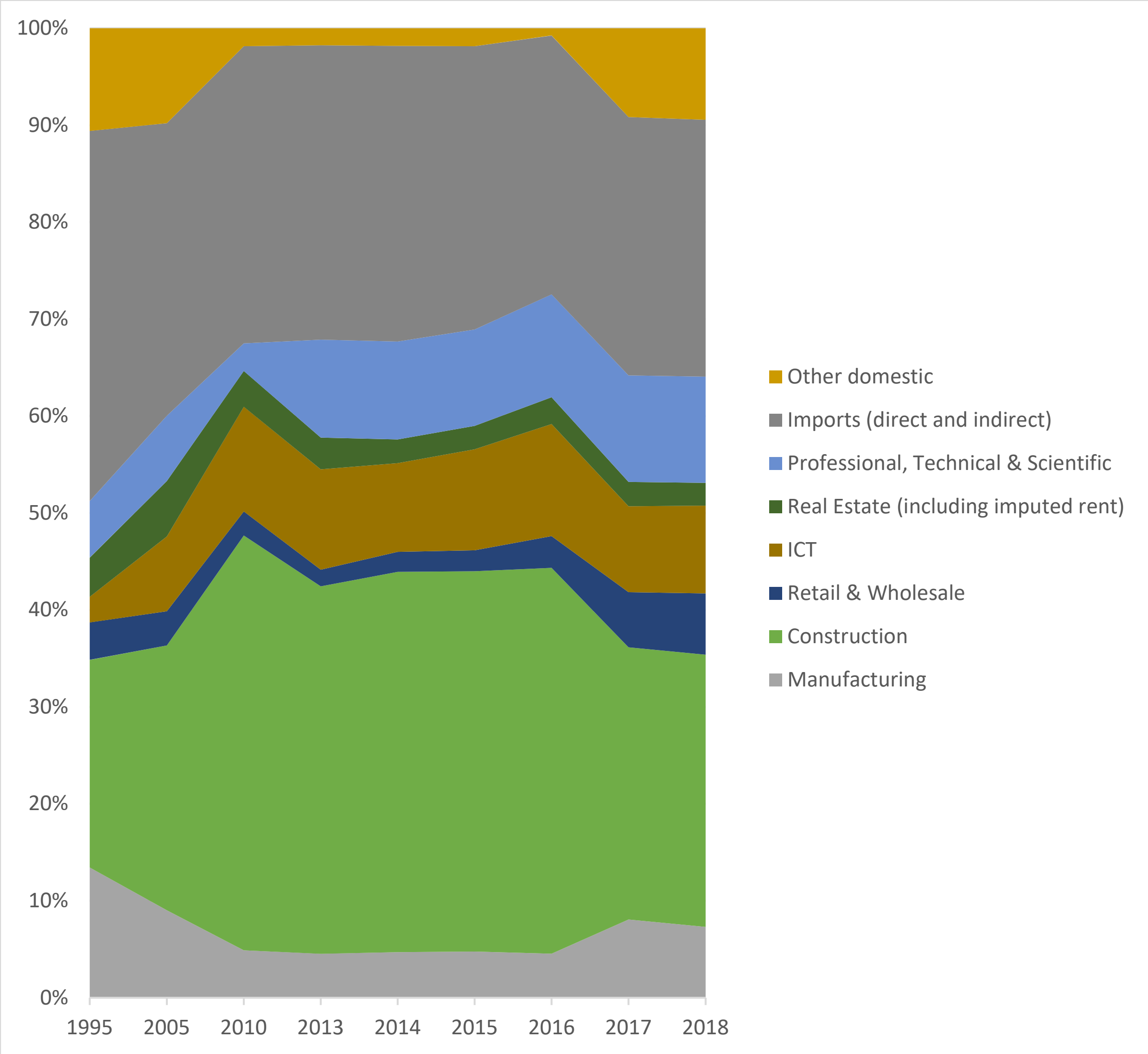

Figure 2: Source of GVA in UK GFCF expenditure according to ONS input-output tables

The derivation of these matrices from the input-output tables is described in Appendix B with an example in Appendix C.

Data

The data used for nowcasting is quarterly growth rates for each of the 20 output indices of the UK economy. These relate to the 2007 Standard Industry Classification (SIC) sectors. For the initial exercise covering 2010Q1-2019Q4, input data are from the Monthly GDP time series published by the ONS in February 2023. For the later real-time nowcasts, a series of real-time GDPO growth estimates from the first release of each quarter (2018Q2-2022Q3) is constructed from revisions triangles provided by the ONS.

Total imports in the 2010-2018 estimation exercise using 'latest vintage' data are taken from First Quarterly Estimate data tables published by the ONS in February 2023. For the real-time exercise a series of real-time first estimates of total imports is constructed from the 'GDP expenditure components – real-time database' spreadsheet published by the ONS: specifically the 'Quarter 4 (Oct to Dec) 2022, first estimate edition of this dataset' published in February 2023. The industrial distribution of goods imports is not used but is discussed on page 17.

Consumption and GFCF nowcasts produced from the above GDPO data are assessed for accuracy against what is termed the 'final estimate' of GDPE components, which here is taken to be the First Quarterly Estimate data tables published by the ONS in February 2023.

To compare the accuracy of the nowcasts against the accuracy of the ONS first estimates, a database of first estimates of consumption and GFCF was constructed from the 'GDP expenditure components – real-time database' spreadsheet published by the ONS, using the same vintage as used for imports data (see above).

Pre-nowcasting checks

The "R^2=1 check" and the effects of diverging deflators

As a preliminary exercise, we take Blue Book data for $\boldsymbol{y^e}_{\boldsymbol{T}}$ and $\boldsymbol{y^o}_{\boldsymbol{T}}$ and used input-output tables to re-construct Identity 2. Taxes and subsidies are taken from the Blue Book. As this holds by definition, it constitutes a check of calculations, using data and matrices which would not be feasible in real time.

After confirming that the identity holds for nominal consumption and GFCF, we repeat the exercise with real terms data (not only are chained volume measures aggregates of greater general interest but GDPO indices are only available in real terms in real-time).[7] Figures 3 and 4 show for 2013 onwards[8] the in-sample i.e. non-feasible (in real time) estimates for real terms growth rates each year – derived from calculations in levels – for consumption and GFCF. Recalling that the errors in each case are zero for nominal data, errors here can be interpreted loosely as the degree to which differential price growth across the economy would hinder the use of input-output tables for nowcasting real aggregates (if not corrected for) even if we had perfect information about current and past data and in-year input-output tables. Because we do not have the input-output tables for *t-1* which are consistent with the same vintage of Blue Book data, growth rates are calculated against the contemporary vintage of data for GDPE components, rather than estimate GDPE components.

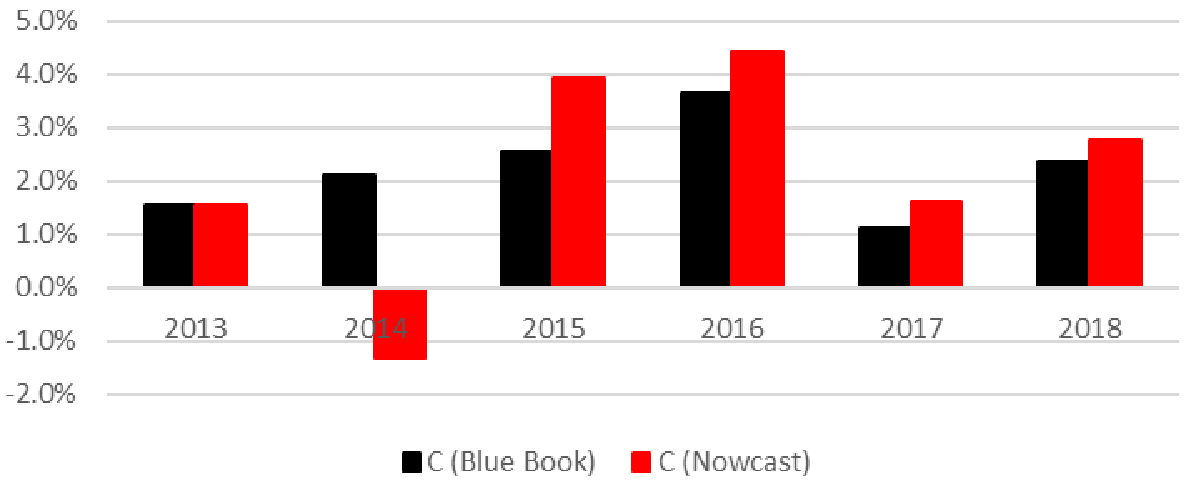

Figure 3: Infeasible estimates of year-on-year real consumption growth produced from the same year's input-output tables and Blue Book data

[7] Because the expenditure disaggregation of 'taxes minus subsidies' is taken from the input-output tables, and therefore only available in nominal terms, for the real terms in-sample exercise the corresponding nominal values were deflated by the GDP deflator.

[8] In-sample 'nowcasts' for 2010-2013 have not been produced as Blue Book tables appear not to be available online in the same format as Blue Book 2014 onwards.

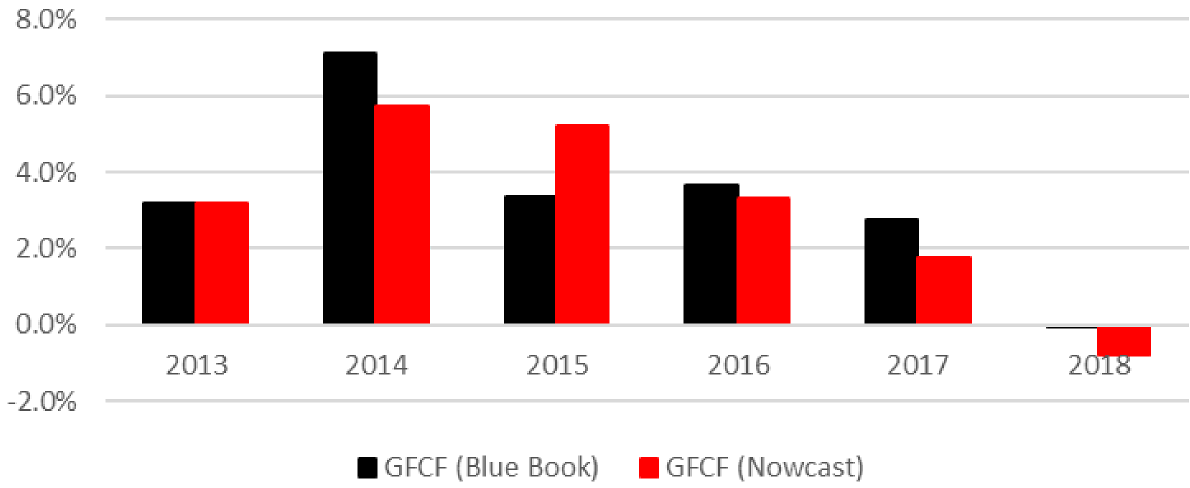

Figure 4: Infeasible estimates of year-on-year real GFCF growth produced from the same year's input-output tables and Blue Book data

In most cases, these errors seem small; the obvious exception is the estimate for consumption in 2014, which under-predicts by around 3 percentage points – a counterpart to large over-predictions for inventories and exports. Other than this, there appears to be a slight positive bias to the consumption estimates. Future research ought to include consideration of divergence between expenditure deflators: something we will return to later in the paper.

However, the divergence between nominal and real estimates increases with the distance between the reference year and the ONS base year for chained volume measures, which varies through this period between zero and two years, meaning that this is not a pure exercise in estimating the contribution of relative price shifts to forecasting errors.

Furthermore, it became clear in the process of checking the results for years in which nominal = real (in other words, the year in question was also the reference year for chained volume measures in the corresponding Blue Book) that the chained volume measure GDPO data in earlier Blue Books is inconsistent with GVA data in input-output tables constructed from the same Blue Book. This has been confirmed to us by the ONS as arising from the fact that only from Blue Book 2021 onwards has the supply-use framework fully integrated into the production of chained volume measure GDP – in other words, that, based on Blue Book data, the identity in Equation 1 does not hold (except by exception) in real terms, even though it holds in nominal terms and the choice of base year makes real equal to nominal values.

Clearly this introduces a further not insignificant source of potential forecasting error and one which is impossible to sense the magnitude of. However, while it disadvantages (perhaps severely) nowcasts in this paper for years up to and including 2016, the shift to integrating real and nominal GDP construction in 2021 means it can be disregarded as a problem for the future nowcaster.

Matrix instability and t+1 nowcasts

As an indicative investigation into the role which matrix instability plays in errors, we conducted a similar exercise on *t+1* annual nominal data. That is: using the second suffix T on $\boldsymbol{y^o}_{T,T}$ to denote that data is taken from the same Blue Book as $\boldsymbol{R_T}$ and $\boldsymbol{m_T}$ are constructed, we estimate

$$\boldsymbol{\hat{y}^e}_{T+1} = \boldsymbol{y^o}_{T+1,T}\, \boldsymbol{R_T}\, \boldsymbol{C_T} + \boldsymbol{p_T}\, M_{T+1,T} \qquad (4)$$

Note that net taxes on products are excluded, making this an exercise in recreating GDP at basic prices rather than market prices, because vector $\boldsymbol{T}_{\boldsymbol{T+1}}$ is not available at time *t* and, unlike ratio vectors $\boldsymbol{R}_{\boldsymbol{T}}$ $\boldsymbol{C}_{\boldsymbol{T}}$ and $\boldsymbol{p}_{\boldsymbol{T}}$, we know that $\boldsymbol{T}_{\boldsymbol{T}}$ would not constitute an unbiased estimator.

This (again, infeasible) approach – by concentrating on nominal growth rates and based on the same vintages of GDPO data – reduces errors to the sum of those caused by matrix instability and those attributable to the non-availability of taxes minus subsidies on products by expenditure components.

Perhaps surprisingly, if read in light of the results for quarterly growth in the following section, estimates for consumption (Figure 5) appear to be significantly more accurate than those from GFCF (Figure 6). This may be attributable to the smoother path of consumption in level terms but points towards the fact that matrix instability is unlikely to be the principal source of consumption nowcast errors.

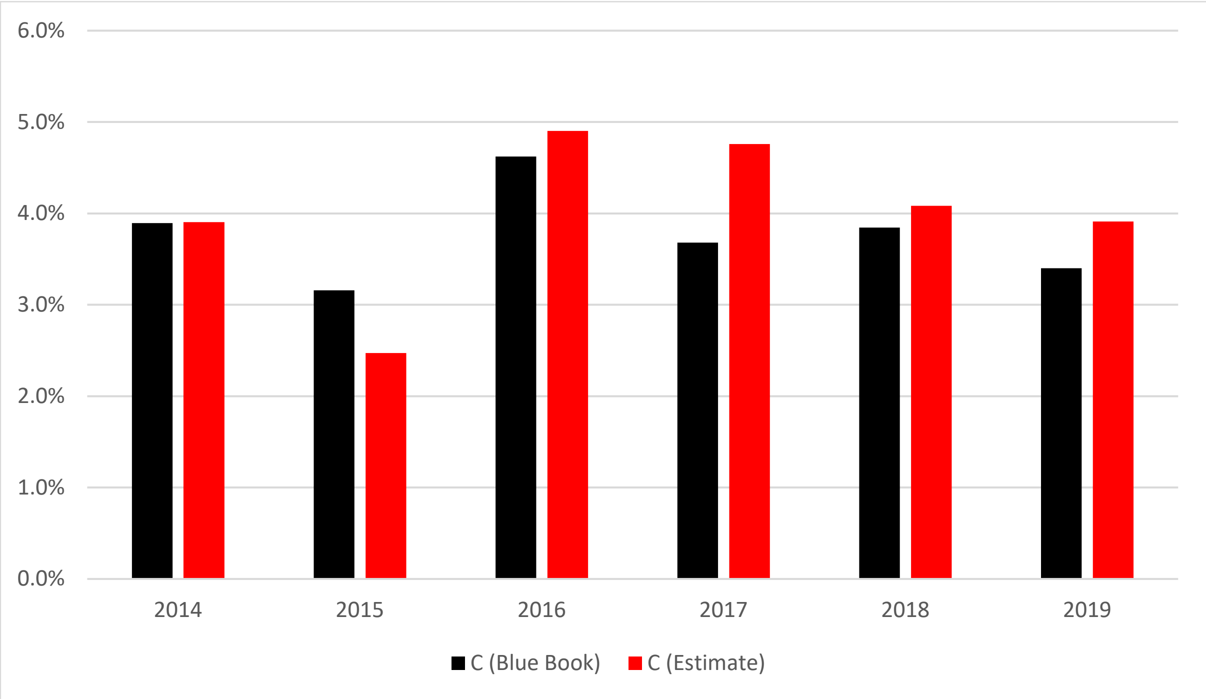

Figure 5: Infeasible estimates of year-on-year nominal consumption growth produced from the previous year's input-output tables and Blue Book data

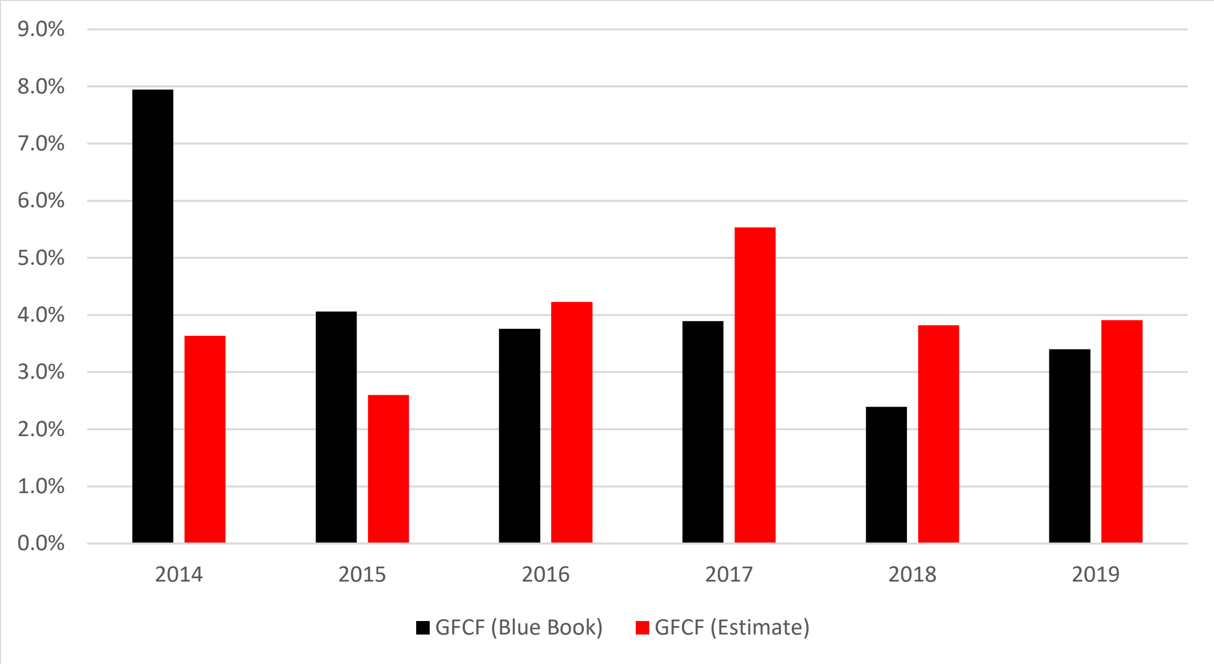

Figure 6: Infeasible estimates of year-on-year nominal GFCF growth produced from the previous year's input-output tables and Blue Book data

3. Results

Nowcasting consumption and GFCF 2010-2019

Table 1 shows the bias and root mean squared error of the consumption and GFCF nowcasts, using the latest available vintage of GDPO data, when considered as forecasts of the latest consumption and GFCF estimates for the same quarter, as well as comparisons with the ONS first estimates treated as forecasts in the same way. Figures 7 and 8 graph the series.

	Consumption		GFCF	
	Nowcast	First estimate	Nowcast	First estimate
RMSFE	1.14%	0.61%	2.09%	2.08%
Correlation coefficient	0.03	0.54	0.41	0.39
Mean error (bias)	0.13%	-0.14%	-0.09%	-0.56%

Table 1: 2010-2019 root mean squared errors and mean errors for consumption and GFCF. Nowcasts based on latest vintage GDPO data and ONS first estimates.

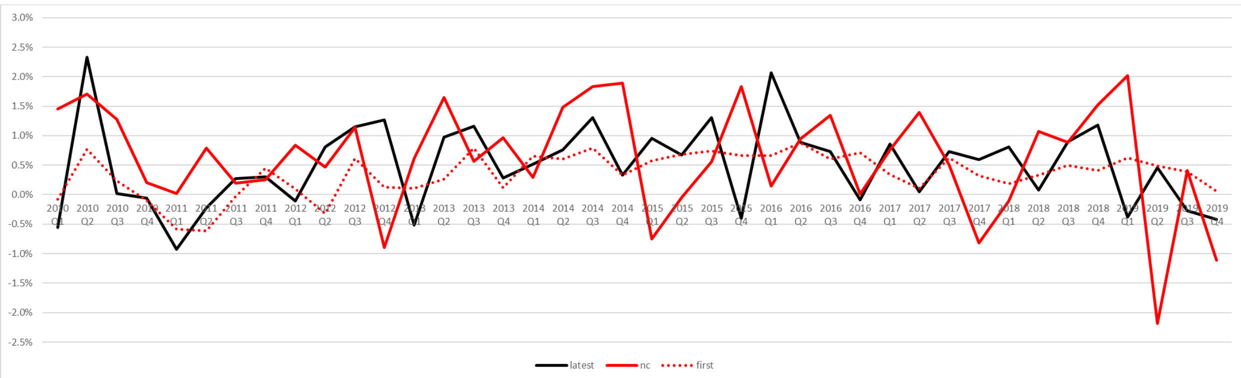

Figure 7: 2010-2019 latest estimates of consumption growth, nowcasts and ONS first estimates

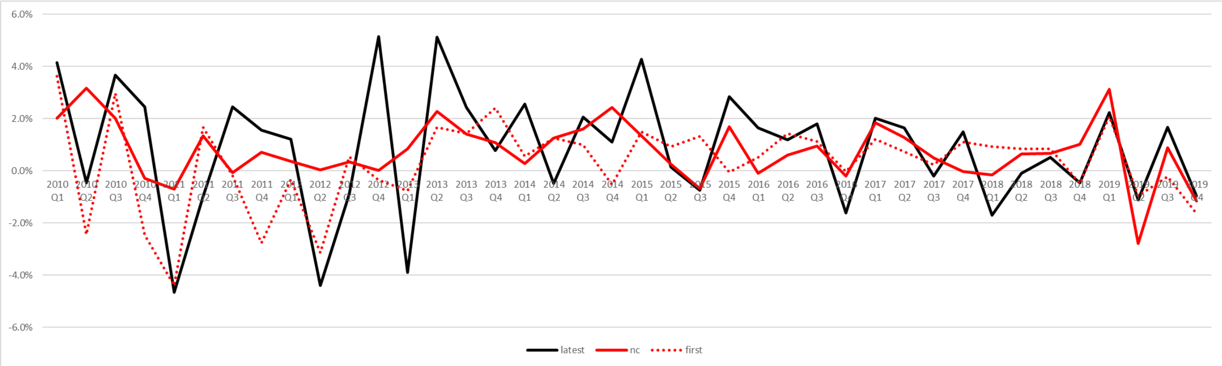

Figure 8: 2010-2019 latest estimates of GFCF growth, nowcasts and ONS first estimates

For GFCF, the GDPO-based nowcasts are less biased and have an almost identical RMSFE to the ONS first estimate, when both are considered as forecasts of the latest vintage data, and a higher correlation coefficient. Figure 8 suggests that the fit of the nowcasts improves during the period under consideration; this could be related to a change in the relationships or could also be attributed to the fact that early nowcasts were constructed with very outdated input-output tables. In quarters when GFCF rose or fell dramatically, the forecast errors of the nowcasts and of the first estimates appear correlated, suggesting that some important data may not have been available at the time of the first release.

For consumption, the nowcasts are significantly less accurate; indeed, the correlation between them and the final estimates is close to zero for the 2010-19 period. While the first estimate appears biased towards the unconditional mean, the nowcast is significantly more volatile than the latest vintage data, and is clearly inefficient (see page 17). Given the use of contemporaneous data, this poses something of a puzzle, and the Conclusion considers a number of areas in which it could be explored further.

Nowcasting with real-time data from 2018

Section 2 notes that the use of matrices constructed from nominal data to nowcast from and to real variables is likely to introduce an element of error, compared with the initial perfect fit. A further source of error when forecasting consumption and GFCF may be measurement error in early vintage GDPO data.

A real-time database of input data, i.e. quarterly growth rates for GDPO constituent sectors, was made available by the ONS. While this, unfortunately, covers only a short period, including major Covid-19 pandemic lockdowns, it may give us some indicative information on how much additional error the use of early-vintage GDPO data hinders the accuracy of the nowcasts.

Table 2 repeats the exercise of Table 1 but now uses real-time GDPO data to construct the nowcasts, instead of the latest vintage. Data is limited by availability to the period since Q2 2018.

	Consumption		GFCF	
	Nowcast	First estimate	Nowcast	First estimate
RMSFE	10.27%	1.00%	4.37%	2.27%
Correlation coefficient	0.88	0.99	0.98	0.95
Mean error (bias)	1.86%	-0.13%	-0.39%	-0.43%

Table 2: 2018Q2-2022Q4 root mean squared errors and mean errors for consumption and GFCF. Nowcasts based on real-time GDPO data.

Unsurprisingly, during the initial stages of the pandemic, when whole sectors of the economy were virtually shut down, a statistical model based on fixed shares of consumption and GFCF has significantly larger RMSFEs than in earlier periods. An implicit assumption in using fixed weights is that the preceding quarter saw consumption and GFCF distributed as described in the input-output table. At a period when – to take one example – hospitality's share of consumption shrank dramatically, this assumption will plainly be problematic for statistical nowcasting. Not only are the nowcasts likely to perform worse due to the pandemic but errors for the first estimates are likely to be far smaller than in the longer estimation period, because they have undergone relatively few rounds of revisions so far.

Despite the large errors during 2020, the nowcasts have a smaller RMSFE compared to the latest vintage than the ONS first estimate in 7 out of 19 periods for which we have data, in the case of GFCF, and 9 periods for consumption. Both sets of nowcasts have significantly improved correlations with the data, but improvements over the 2010-2018 period ought to be treated with caution: it may reflect the greater volatility of the period, or the fact that much of the data has been through few rounds of revisions; indeed some periods have not yet been 'balanced'. Future work should consider the time profile of errors between the first and final estimates to see if they increase and when.

With the important caveat that the latest vintage data has been subject to very few revisions, Appendix D gives nowcast errors for 2018-2022 when using latest GDP data and using real-time GDPO data. The real-time data nowcasts have an RMSFE around 0.6pp larger for consumption and around 0.9pp for GFCF. A good deal of the RMSFE difference is accounted for by the volatile Covid-19 period: limiting our focus to 2018 and 2019, the years which have seen the most revisions between the RTD and the latest vintage, the RMSFFEs for nowcasts with real-time data are only greater by 0.25pp (consumption) and 0.39pp (GFCF).

Matrix instability and quarterly nowcasts

Considering the additional source of error arising from the use of lagged matrices, a final set of nowcasts were produced for the period 2013-2018 (i.e. when tables began to be published annually until the latest available) using data from the initial exercise but $\boldsymbol{R_T}$, $\boldsymbol{C_T}$ and $\boldsymbol{p_T}$ matrices i.e. from the same year as the data. Significantly smaller errors here might suggest instability in the matrix parameters, limiting the potential use of input-output tables out of sample, and may argue for a shorter lag in producing tables, if this is feasible.

Table 3 compares errors for the period 2010-2018 from this (unfeasible) approach with those from the pseudo out-of-sample nowcasts in Figures 5 and 6. The RMSFE is actually slightly larger for GFCF nowcasts made with in-year matrices than when using feasible lagged matrices, and that for consumption is only slightly smaller. This suggests that the changing industrial composition of expenditure may not be a major source of nowcast errors, and the delay in publishing input-output tables may not be a significant hindrance to their use in nowcasting – a surprising but useful result if confirmed by further investigation and more data.

	Consumption		GFCF	
	Lagged matrices	In-year matrices	Lagged matrices	In-year matrices
RMSFE	1.01%	0.98%	1.66%	1.67%
Mean error (bias)	0.10%	0.05%	-0.07%	-0.05%

Table 3: 2013-2018 root mean squared forecast errors and mean errors for consumption and GFCF. Nowcasts created using lagged matrices and in-year matrices.

4. Conclusions

We started from the knowledge that it was possible to recreate nominal GDPE aggregates perfectly from nominal GDPO data and input-output tables, provided that a strong, indeed unfeasible, set of conditions are fulfilled.

Figures 3 and 4 suggest that, at annual frequency, the use of real rather than nominal data and the pre-2017 incompatibility of real GDPO data between Blue Book and input-output table construction introduce errors which may – particularly in the case of consumption – be biased and therefore partially predictable, but which otherwise do not appear large enough to hinder the nowcasts unduly.

Table 2 and Figure 9 demonstrate that the use of real-time data, with its known measurement errors, is likely to increase the magnitude of errors, though initial results tentatively suggest that this measurement error may not be too significant.

Figures 5 and 6 suggest that matrix instability is unlikely to be the principal source of nowcasting error for consumption, though it appears larger for GFCF. It may be instructive to note for future reference that instability may be genuine – a result of the world being closer to Cobb-Douglas than to Leontief – but it may also be exaggerated by the aggregation of lower-level SIC sectors' GVA to the 20 sections. Table 3 again illustrates the degree to which matrix instability disadvantages the nowcaster in real-time: outside volatile periods such as the Covid-19 lockdowns, the deterioration in nowcast accuracy down to the use of lagged matrices appears small. Using lagged matrices dating back to 1995 for the early years of the nowcast period, we are subjecting our approach to a larger disadvantage than may be strictly necessary, given that input-output matrices are now produced with a much shorter lag; however, repeating the early nowcasts using the 2005 tables from the start of the nowcast period instead does not materially alter the results.

Despite these known and partially quantifiable sources of error, we find that for GFCF, the nowcasts produced by a feasible nowcasting technique based on input-output matrices can – even when using real-time data – compete with existing ONS first estimates in predicting later estimates.

However, our contribution in this paper is not to claim an optimal approach to nowcasting with input-output tables but rather to demonstrate the potential for further academic or statistical office research. For example, if diverging expenditure deflators are known to be causing a systematic bias in nowcasting real variables, a bias correction can be easily applied. A real-time forecaster would also not be 'fooled' in the way that a fixed weights model is at times of major divergence in sectoral growth rates, such as in the early quarters of the Covid-19 pandemic.

More generally, the nowcasts produced are plainly far from efficient, as evidenced by the fact that – bearing in mind that efficient forecasts are orthogonal to their errors – nowcast errors for 2010-2019 have correlation coefficients of 0.8 (consumption) and 0.1 (GFCF) with their respective variables. There is also a degree of correlation between nowcast errors for different variables, including those which are not the subject of this paper (see Table 4).

	Government consumption	Consumption (HH+NPISH)	GFCF	Inventories	Valuables	Exports
Government consumption	1.00	0.45	0.37	0.06	0.10	0.23
Consumption (HH+NPISH)		1.00	0.28	0.03	0.09	0.03
GFCF			1.00	-0.23	0.38	-0.09
Inventories				1.00	-0.01	-0.24
Valuables					1.00	-0.07
Exports						1.00

Table 4: Correlation coefficients between nowcast prediction errors 2010-2019

In the case of consumption, in particular, the use of the fixed matrix appears to impose a variance for nowcasts which is clearly at odds with the variance of the data. Further research in decomposing the sources of the errors is a priority for future stages of research.

5. Future research

As noted above, clearly the nowcasts described above could be improved upon in a number of ways within the input-output framework. This section elaborates a number of directions in which future research could progress.

Levels or growth rates

When it comes to the pseudo-out-of-sample exercise in quarterly growth rates, we do not know the cross-industry distribution of consumption or GFCF at the start of the reference quarter and how far it has moved since the input-output table was constructed. As a result, the implicit assumption is that the cross-industry variation of each expenditure component matched that of the input-output table in the previous quarter: an assumption which will become more problematic the further it is from being true, as noted above about the start of the recent pandemic.

Instead of working purely in quarterly growth rates, one alternative would be to re-specify equations in terms of levels. More concretely: cumulative growth between the fourth quarter of year *S* and *t* in each element of $\boldsymbol{y^o}_{\boldsymbol{t}}$ can be allocated with $\boldsymbol{R_S}$ and $\boldsymbol{C_S}$ to nowcast $\widehat{\boldsymbol{y}}^{\boldsymbol{e}}{}_{\boldsymbol{t}}$, and the quarterly growth rates calculated by reference to the corresponding element of $\widehat{\boldsymbol{y}}^{\boldsymbol{e}}{}_{\boldsymbol{t-1}}$. Concretely, and considering for now only domestic value added:

$$\widehat{\widehat{\boldsymbol{y}}}^{\boldsymbol{e}}{}_{\boldsymbol{t}} = (\boldsymbol{y^o}_{\boldsymbol{S,4}}\, \boldsymbol{R_S}\, \boldsymbol{C_S} + \left(\boldsymbol{y^o}_{\boldsymbol{t}} - \boldsymbol{y^o}_{\boldsymbol{S,4}}\right) \boldsymbol{R_S}\, \boldsymbol{C_S}) - \widehat{\boldsymbol{y}}^{\boldsymbol{e}}{}_{\boldsymbol{t-1}} \qquad (5)$$

where the first term and last term are, to a degree, known. Especially in the presence of significant economic volatility, this approach is likely to be advantageous in removing the implicit assumption that value is distributed according to $\boldsymbol{C_S}$ in period *t-1*.

Price divergence

Note that in the above approach, there will remain potential divergences between real and nominal variables, which itself will need to be examined. A further alternative that could be considered would require at least some provisional estimates of sectoral price changes. If these were available, an alternative nowcasting methodology would not rely on the Leontief assumption, but would instead exploit the equivalence, shown in Burress (1993), between real changes in a Leontief economy and nominal changes in a Cobb-Douglas economy. In this alternative procedure, shares in nominal, rather than real output would be held constant (i.e. the impact of price and volume changes would be assumed to offset).

However, it is unclear whether relevant price data would be available in a sufficiently timely fashion to make this viable as GDPO is measured in real terms, though price data for goods and services is clearly available with a short lag for inflation indices, as well as implicit deflators for expenditure aggregates which could be combined.

Exports and government spending

In this paper, all aggregates of GDPE are nowcast in an unconstrained system of linear equations. This produces, inter alia, nowcasts for exports and government consumption. However, in real-time, at the point when the GDPO and imports data are known, estimates are available for exports (published at the same time as imports and GDPO around 40 days after quarter-end) and government consumption (published with around a 20-day lag from quarter-end). Information is therefore available about the extent to which these two variables are greater or less than the values produced by the nowcasting system, and an efficient nowcast would take this into consideration. Figure 9 highlights the divergence between the nowcast values for exports and actual values, known at the time of the nowcast but not presently used.

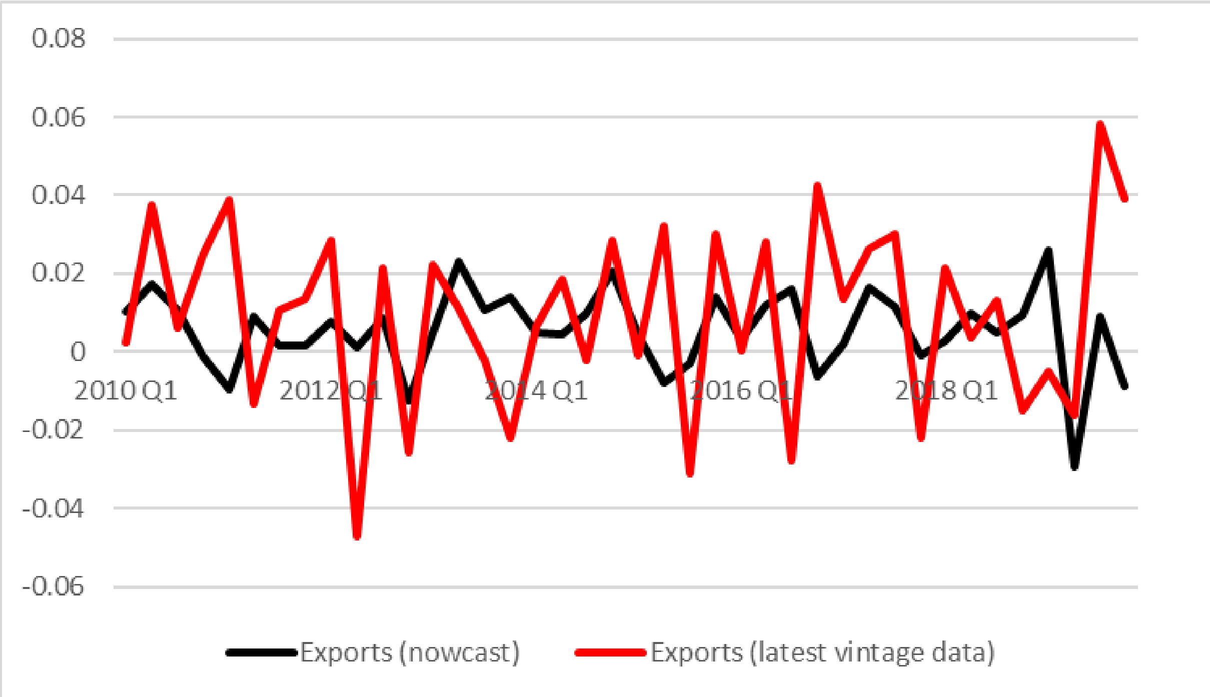

Figure 9: Nowcast (using latest vintage data) and actual exports 2010-2019

A further discussion of future avenues exploiting contemporary knowledge of exports follows on page 17.

Inventories

As per Table 4, nowcasts for inventories are produced as a by-product of this exercise, which is unlikely to be accurate and may contribute to nowcast errors elsewhere. Specifically, we might expect a negative AR1 term for inventories: the quarters before and after the UK's initially planned, though missed, departure date (29 March 2019) from the European Union saw large over- and under-predictions respectively in both consumption and GFCF.

Industrial distribution of imports

Further to the above, there is industrial information available for goods imports, which – with some assumptions placed on services imports – would enable a more detailed nowcast of imported consumption and GFCF than is used in this paper. Given the very general allocation implied by $\boldsymbol{p_T}$, this is likely to yield further nowcasting improvements, even if the services element of the import vector has to be itself nowcast.

Secondary production

Finally, the authors believe that it ought to be possible to produce a more informative mapping from industries to products, currently performed by diagonal matrix $\boldsymbol{R_S}$, based on the relationships between 'Use bp PxI' and 'IOT' matrices. Unfortunately the ONS does not publish the matrices used to allocate secondary production but something close to them ought to be possible to recover. Note this is complicated though not made impossible by the methodological discontinuity discussed on page 17.

Exploring nowcastability further using Supply-Use Tables

The methodology above uses data on GVA and imports to attempt to nowcast expenditure components. As noted in the paper, a downside of this approach is that it does not impose consistency of the nowcasts with GDP forecasts. Appendix E is an exploration of an alternative approach which imposes consistency, based on the identity (setting the statistical discrepancy to zero):

$$Domestic\ Demand\ \equiv Total\ Consumption + Investment \equiv GVA + M + X \qquad (6)$$

which must hold, not just at the aggregate level, but also at the product level.

References

Burress, D. (1994) ‘Homeomorphism between Leontief and Cobb-Douglas input-output models’, Economic Letters 44, pp.49-53

Galvão, A. & Macqueen (R.) (2023) ‘Nowcasting UK consumption and investment with monthly output components and real-time indicators’, ESCoE Discussion Paper, forthcoming

McLaren, C. (2022) ‘What is GDP and how do we measure it?’, National Statistical blog, December 12 https://blog.ons.gov.uk/2022/12/12/what-is-gdp-and-how-do-we-measure-it/

Appendix A

Year (t)	Quarter (t)	Nowcast in	Latest IO tables (S)	Based on Blue Book
2010	1	May-10	1995	2001
	2	Aug-10	1995	2001
	3	Nov-10	1995	2001
	4	Feb-11	1995	2001
2011	1	May-11	1995	2001
	2	Aug-11	2005	2009
	3	Nov-11	2005	2009
	4	Feb-12	2005	2009
2012	1	May-12	2005	2009
	2	Aug-12	2005	2009
	3	Nov-12	2005	2009
	4	Feb-13	2005	2009
2013	1	May-13	2005	2009
	2	Aug-13	2005	2009
	3	Nov-13	2005	2009
	4	Feb-14	2005	2009
2014	1	May-14	2010	2013
	2	Aug-14	2010	2013
	3	Nov-14	2010	2013
	4	Feb-15	2010	2013
2015	1	May-15	2010	2013
	2	Aug-15	2010	2013
	3	Nov-15	2010	2013
	4	Feb-16	2010	2013
2016	1	May-16	2010	2013
	2	Aug-16	2010	2013
	3	Nov-16	2010	2013
	4	Feb-17	2010	2013
2017	1	May-17	2013	2016
	2	Aug-17	2013	2016
	3	Nov-17	2013	2016
	4	Feb-18	2013	2016
2018	1	May-18	2014	2017
	2	Aug-18	2014	2017
	3	Nov-18	2014	2017
	4	Feb-19	2014	2017
2019	1	May-19	2015	2018
	2	Aug-19	2015	2018
	3	Nov-19	2015	2018
	4	Feb-20	2015	2018
2020	1	May-20	2016	2019
	2	Aug-20	2016	2019
	3	Nov-20	2016	2019
	4	Feb-21	2016	2019
2021	1	May-21	2017	2020
	2	Aug-21	2017	2020
	3	Nov-21	2017	2020
	4	Feb-22	2017	2020
2022	1	May-22	2018	2021
	2	Aug-22	2018	2021
	3	Nov-22	2018	2021

Alignment of GDP release dates with latest input-output table available at the time

Appendix B

Each input-output data release from the ONS contains the following tabs, amongst others, and with slightly different naming conventions in some years.

1. A 'GVA by final uses' matrix which describes, for each category of products within each of 60-100 lower-level SIC classifications, the form of their final use by expenditure category. Each row corresponds to the GVA in a group of products and each column to the amount of this GVA which finds its final use in each expenditure aggregate. (Note that, as described below, the treatment of indirect GVA differs between post-2017 and pre-2017 tables.) Each row corresponds to the total GVA in that sector of products and each column corresponds to the expenditure on each category of final use.

   From this, SIC-classified products are grouped into the top-level 20 SIC sectors to get matrices $\boldsymbol{C_T}$ and $\boldsymbol{K_S}$. An example, that for 2018, is in Appendix C. Additionally, the row totals are grouped and divided by their sum to form a 20x1 matrix $\boldsymbol{c_T}$', each element of which is the share of total GVA (direct and indirect) in each of the groups of products corresponding to the 20 top-level SIC sectors.

2. A 'Primary Input from final use' matrix which, in each column, adds the GVA which went into each form of final use (consumption, GFCF, etc) to its associated taxes and imported content, giving as column totals the total expenditure on each final use: GVA + taxes + imports. From the distribution of import values (direct plus indirect) vectors $\boldsymbol{p_T}$ and $\boldsymbol{\pi_S}$ are calculated as described above.

3. A 'Use bp PxI' matrix which relates supplied products (by lower-level SIC classification) to the industries (by lower-level SIC classification) in which they are used and the value added to them in each industry. These are grouped to top-level SIC sectors and the 1x20 vector of GVA for each industry is divided by total GVA and designated $\boldsymbol{a_T}$.

4. The 'IOGs' matrix used to group lower-level SIC codes into 20 top-level SIC sectors A-T for products and industries. Prior to 2007, different SIC groupings were in use, but the fact that 2-digit industrial classifications are used meant that for 1995 and 2005 tables these could be grouped into post-2007 SIC sections.

The vector of ratios between the corresponding elements of matrices $\boldsymbol{c_T}$ and $\boldsymbol{a_T}$, $\boldsymbol{r_T}$, gives the relationships between industry GVA and product GVA. Both 1x20 vectors sum across elements to total GVA but it is distributed differently between their elements. Diagonal matrix $\boldsymbol{R_T}$ = diag($\boldsymbol{r_T}$), so that, suppressing time subscripts, each element $R_{ii} = b_i/a_i$. By allocating this 'secondary production'[9] we move from estimating how much activity has taken place in each industry, $\widehat{\boldsymbol{y}}^{\boldsymbol{o}}{}_{\boldsymbol{t}}$, to an assumption about how much value of each corresponding product grouping (the rows of the 'GVA by final use' tab) has been added, $\widehat{\boldsymbol{y}}^{\boldsymbol{o}}{}_{\boldsymbol{t}}\, \boldsymbol{R_S}$.

[9] For example, if an oil company additionally sells management consultancy services in addition to its core business of extraction, the inputs supplied and GVA associated with the consultancy would be allocated in the 'IOT' matrix to sector M (which includes management consultancy) and in the 'Use bp PxI' matrix to sector B (which includes oil extraction).

Discontinuities in the tables arise in 2010 and 2017 related to intermediate production.[10] In the 'GVA by final uses' matrix between 2010 and 2017 inclusive, GVA in products used up in intermediate production was allocated to the SIC code of the final product containing them. Before and after this, intermediate GVA has been allocated instead to the SIC code of the intermediate product. The reallocation of secondary production in $\boldsymbol{R_S}$ therefore takes place using a different approach before and after this period. This does not affect the validity of our approach, because whichever approach to allocating indirect product GVA is taken, it is used to link industry GVA to expenditure categories via the 'GVA by final uses' matrix. However, the discontinuities are observable in Figures 1 and 2.

[10] The authors are grateful to Eric Crane of the ONS for explanation.

Appendix C

Sector	G	C(HH+NPISH)	GFCF	Inventories	Valuables	Exports	TFE
A	0	5,609	1,072	76	0	2,109	8,867
B	0	168	587	-464	0	12,263	12,554
C	2,134	31,804	14,048	-1,157	395	113,000	160,225
D	0	19,858	0	-1	0	122	19,979
E	3,887	8,447	77	9	0	3,015	15,435
F	0	1,246	115,460	4,568	0	1,668	122,942
G	1,796	166,111	6,392	0	-73	39,376	213,602
H	2,156	37,859	505	0	0	15,193	55,714
I	0	76,353	0	0	0	10,662	87,015
J	2,390	26,094	30,705	289	0	19,773	79,250
K	0	55,458	29	0	0	54,891	110,377
L	0	259,758	7,098	0	0	1,599	268,455
M	0	6,063	29,224	266	0	40,671	76,224
N	0	9,053	719	1,162	0	23,616	34,550
O	98,095	2,724	2,072	0	0	912	103,802
P	52,783	37,149	0	-82	0	6,398	96,248
Q	121,489	31,717	0	0	0	365	153,571
R	2,876	22,320	367	1,997	-99	6,333	33,794
S	0	31,554	0	0	0	1,203	32,757
T	0	6,658	0	0	0	18	6,676
Total GVA	287,607	836,001	208,356	6,664	224	353,188	1,692,039

'GVA by final use' matrix grouped vertically into 20 top-level SIC sectors and horizontally with consumer consumption and total exports grouped

Imports	49,422	266,281	86,125	-175	-423	145,364	546,594
Total GVA including imports	337,028	1,102,282	294,481	6,489	-200	498,552	2,238,633

From 'Primary Input by final use' matrix

0.0%	0.5%	0.4%	1.2%	0.0%	0.4%
0.0%	0.0%	0.2%	-7.1%	0.0%	2.5%
0.6%	2.9%	4.8%	-17.8%	-198.2%	22.7%
0.0%	1.8%	0.0%	0.0%	0.0%	0.0%
1.2%	0.8%	0.0%	0.1%	0.0%	0.6%
0.0%	0.1%	39.2%	70.4%	0.0%	0.3%
0.5%	15.1%	2.2%	0.0%	36.4%	7.9%
0.6%	3.4%	0.2%	0.0%	0.0%	3.0%
0.0%	6.9%	0.0%	0.0%	0.0%	2.1%
0.7%	2.4%	10.4%	4.4%	0.0%	4.0%
0.0%	5.0%	0.0%	0.0%	0.0%	11.0%
0.0%	23.6%	2.4%	0.0%	0.0%	0.3%
0.0%	0.6%	9.9%	4.1%	0.0%	8.2%
0.0%	0.8%	0.2%	17.9%	0.0%	4.7%
29.1%	0.2%	0.7%	0.0%	0.0%	0.2%
15.7%	3.4%	0.0%	-1.3%	0.0%	1.3%
36.0%	2.9%	0.0%	0.0%	0.0%	0.1%
0.9%	2.0%	0.1%	30.8%	49.5%	1.3%
0.0%	2.9%	0.0%	0.0%	0.0%	0.2%
0.0%	0.6%	0.0%	0.0%	0.0%	0.0%

$\boldsymbol{K_{2015}}$

14.7%	24.2%	29.2%	-2.7%	212.2%	29.2%

$\boldsymbol{\pi}_{\mathbf{2015}}$

0.0%	63.3%	12.1%	0.9%	0.0%	23.8%
0.0%	1.3%	4.7%	-3.7%	0.0%	97.7%
1.3%	19.8%	8.8%	-0.7%	0.2%	70.5%
0.0%	99.4%	0.0%	0.0%	0.0%	0.6%
25.2%	54.7%	0.5%	0.1%	0.0%	19.5%
0.0%	1.0%	93.9%	3.7%	0.0%	1.4%
0.8%	77.8%	3.0%	0.0%	0.0%	18.4%
3.9%	68.0%	0.9%	0.0%	0.0%	27.3%
0.0%	87.7%	0.0%	0.0%	0.0%	12.3%
3.0%	32.9%	38.7%	0.4%	0.0%	24.9%
0.0%	50.2%	0.0%	0.0%	0.0%	49.7%
0.0%	96.8%	2.6%	0.0%	0.0%	0.6%
0.0%	8.0%	38.3%	0.3%	0.0%	53.4%
0.0%	26.2%	2.1%	3.4%	0.0%	68.4%
94.5%	2.6%	2.0%	0.0%	0.0%	0.9%
54.8%	38.6%	0.0%	-0.1%	0.0%	6.6%
79.1%	20.7%	0.0%	0.0%	0.0%	0.2%
8.5%	66.0%	1.1%	5.9%	-0.3%	18.7%
0.0%	96.3%	0.0%	0.0%	0.0%	3.7%
0.0%	99.7%	0.0%	0.0%	0.0%	0.3%

$\boldsymbol{C}_{\mathbf{2015}}$

Appendix D

	Consumption		GFCF	
	Latest vintage data	Real-time data	Latest vintage data	Real-time data
2018 Q2	0.99%	0.10%	0.73%	0.36%
2018 Q3	-0.01%	-0.54%	0.15%	0.69%
2018 Q4	0.34%	-0.67%	1.47%	0.78%
2019 Q1	2.40%	2.48%	0.88%	0.50%
2019 Q2	-2.63%	-3.61%	-1.67%	-3.20%
2019 Q3	0.68%	0.62%	-0.79%	-1.03%
2019 Q4	-0.70%	0.17%	-0.21%	0.89%
2020 Q1	-0.96%	-0.66%	1.40%	0.15%
2020 Q2	-3.43%	-4.82%	-7.62%	-15.11%
2020 Q3	41.50%	44.04%	11.21%	10.32%
2020 Q4	2.71%	0.43%	3.75%	1.71%
2021 Q1	-3.11%	-3.72%	-0.35%	-0.96%
2021 Q2	-0.94%	-0.56%	-0.07%	0.28%
2021 Q3	0.33%	-0.16%	0.31%	-0.48%
2021 Q4	1.60%	-0.28%	2.43%	0.94%
2022 Q1	1.58%	2.23%	-3.09%	-2.50%
2022 Q2	-0.30%	-0.82%	1.22%	0.79%
2022 Q3	-0.64%	-0.77%	-1.13%	-1.19%
2022 Q4	0.29%	0.29%	-1.02%	-1.03%
RMSE	9.66%	10.27%	3.45%	4.37%
RMSE (2018-2019 only)	1.45%	1.70%	1.00%	1.39%

Nowcast errors for 2018Q2-2022Q3 for latest vintage and real-time data

Appendix E

While complete, and fully disaggregated data for the right-hand side of the identity in Equation 6 on page 17 are not available on a timely basis, data on an exhaustive set of GVA components, plus exports and imports of goods, are available on approximately the same timeline, so that in principle a reasonable estimate of total domestic demand should be available on a timely basis, at a reasonably disaggregated level.

What follows conditions on Blue Book 2022 data for the right-hand side of the identity, at the level of disaggregation in the Supply-Use tables. This would clearly not be a viable real-time exercise, but it is intended to provide a benchmark test of whether a simple linear extrapolation can at least yield reasonable forecasts under these idealised conditions. The provisional answer is that, at least on data for 2019 and 2020, both household consumption and GFCF appear to have considerable forecastability at the product level, based on an extremely simple prediction methodology.

The methodology is applied at each of the 105 product levels, and can be summarised as follows:

$$\sum_i D_{i,t}^j \equiv GVA_t^j + M_t^j - X_t^j \qquad (7)$$

where

$i = HH, NPISH, CG(Central\ Govt), LG(Local\ Govt), GFCF, II(Inventories), V(\ Valuables)$

$j = 1 \dots 105;$

$$\widehat{D}_{i,t+1}^j = s_{i,t}^j \times \sum_i D_{i,t+1}^j \qquad (8)$$

where

$i = HH, NPISH, CG, LG, GFCF, II, V$

$j = 1 \dots 105,$

$\widehat{D}_{i,t+1}^j = 0; i = Inventories, Valuables$

and

$$s_{i,t}^j = \frac{D_{i,t}^j}{\sum_{i \neq II,V} D_{i,t}^j}$$

For each product, each demand component (excluding inventory investment and valuables) is simply projected as a fixed share of total domestic demand in *t+1* (which is assumed observable from GVA and trade data) using shares from year *t.* Given large movements in inventories and valuables in some sectors, base year shares are calculated excluding inventory investment and valuables, and for consistency these are simply projected at zero.

The projection is carried out in nominal terms, which, as argued above, may be more robust than in real terms, since it allows in principle for offsetting price and volume effects. In practice, however, at a sufficiently disaggregated product level, as in this exercise, these effects are likely mostly to affect product totals, rather than individual demand component shares. But such effects could be more important with less detailed product disaggregation.

The results suggest strongly that, at least at this level of disaggregation, even this very crude projection methodology yields quite good predictions, even in the highly volatile conditions of 2020.

Table 5 summarises results for the aggregate components of domestic demand. One clear result is that there is (as indeed there must be, given the identity) a clear element of offsetting errors; thus, total consumption is forecast distinctly better than its component parts. The errors are not, however, perfectly offsetting, given that inventory investment and valuables are simply being forecast at zero.

	Households	Non-profit institutions serving households	Central government	Local government	Total	Gross fixed capital formation
	2019					
Actual	2.63%	5.84%	7.93%	4.49%	3.65%	4.46%
Predicted	3.39%	4.86%	5.80%	6.15%	3.94%	3.76%
	2020					
Actual	-12.48%	-7.58%	14.29%	5.79%	-6.84%	-9.29%
Predicted	-11.17%	-2.91%	9.70%	0.80%	-6.86%	-9.63%

Table 5: Annual growth in nominal components of domestic demand (2019 and 2020) when projected as a fixed share of total domestic demand for each low-level SIC product

Figures 10 to 14 illustrate for household consumption and GFCF at the product level, with all predictions expressed as contributions to growth of the total, bringing out two key features: first, the importance of a relatively few product categories which dominate the predictions; and, second, that a number of these products are dominated by a single expenditure component (the most notable example being construction, which dominates movements in GFCF in both years illustrated).

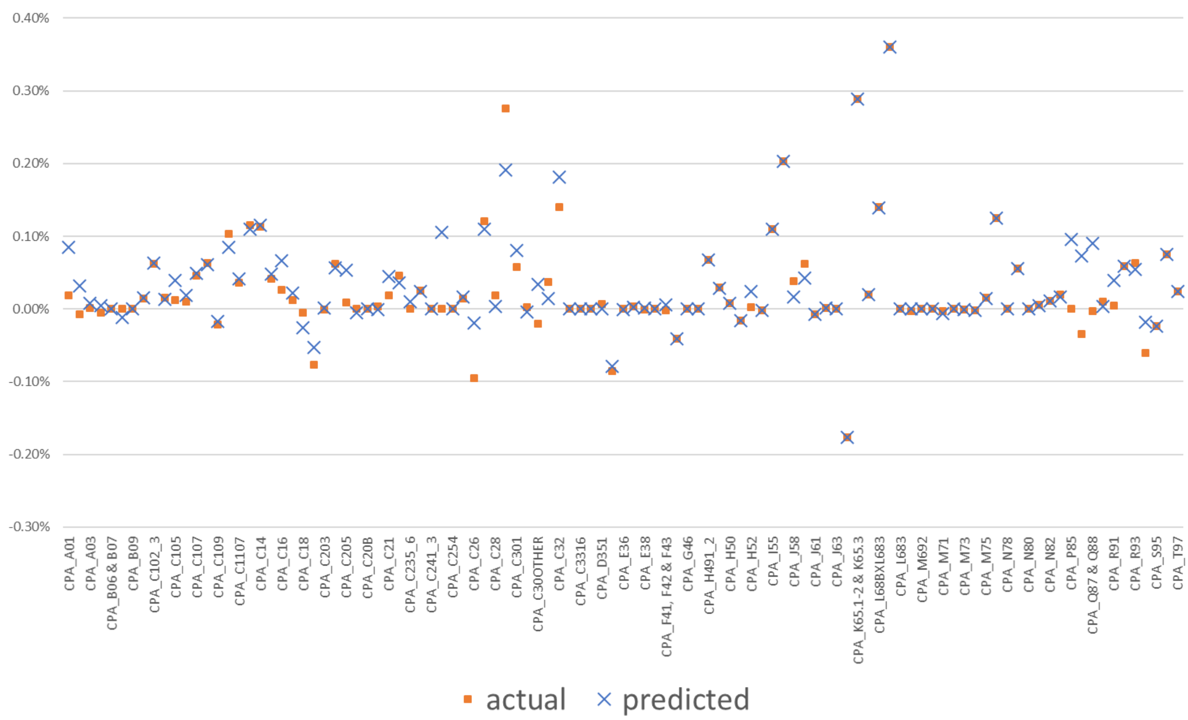

Figure 10: Actual and predicted contributions to growth in household consumption in 2019 by low-level SIC product

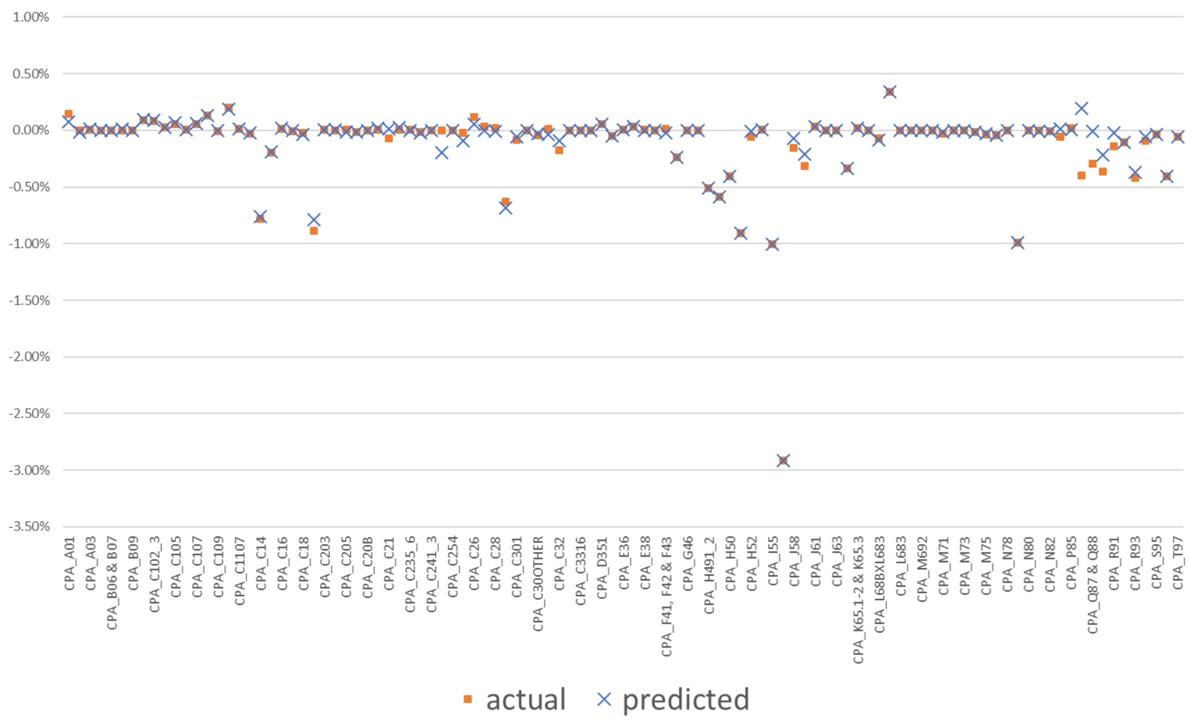

Figure 11: Actual and predicted contributions to growth in household consumption in 2020 by low-level SIC product

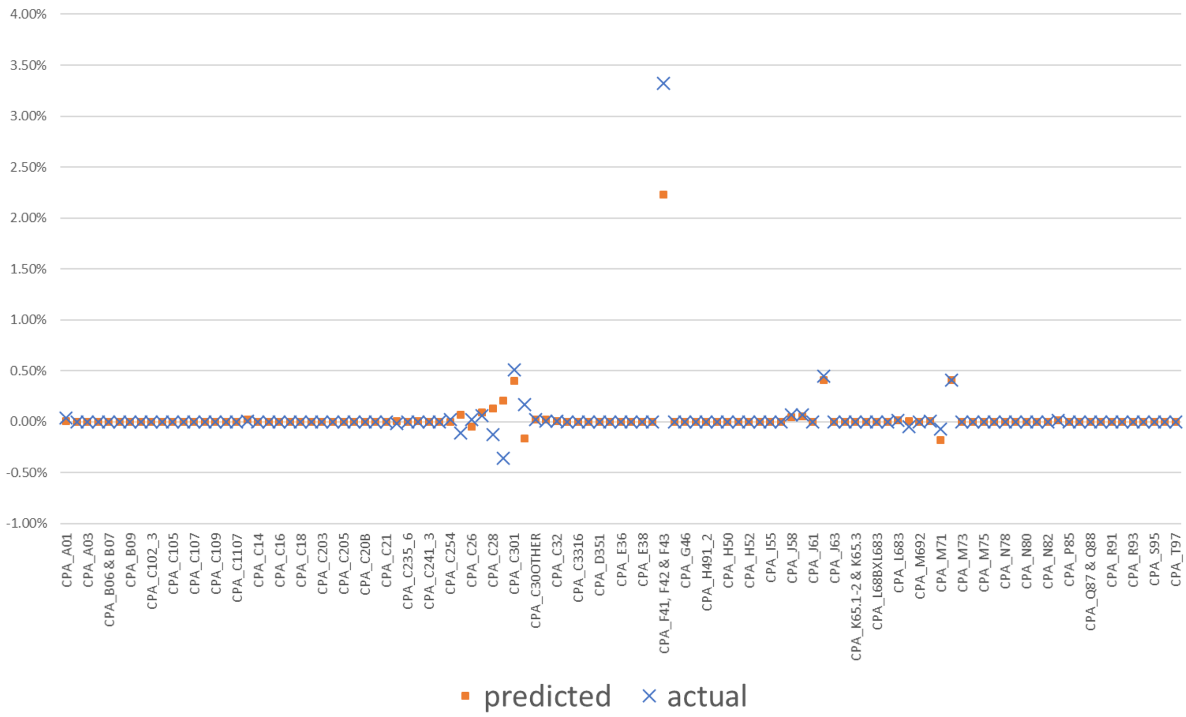

Figure 12: Actual and predicted contributions to growth in GFCF in 2019 by low-level SIC product

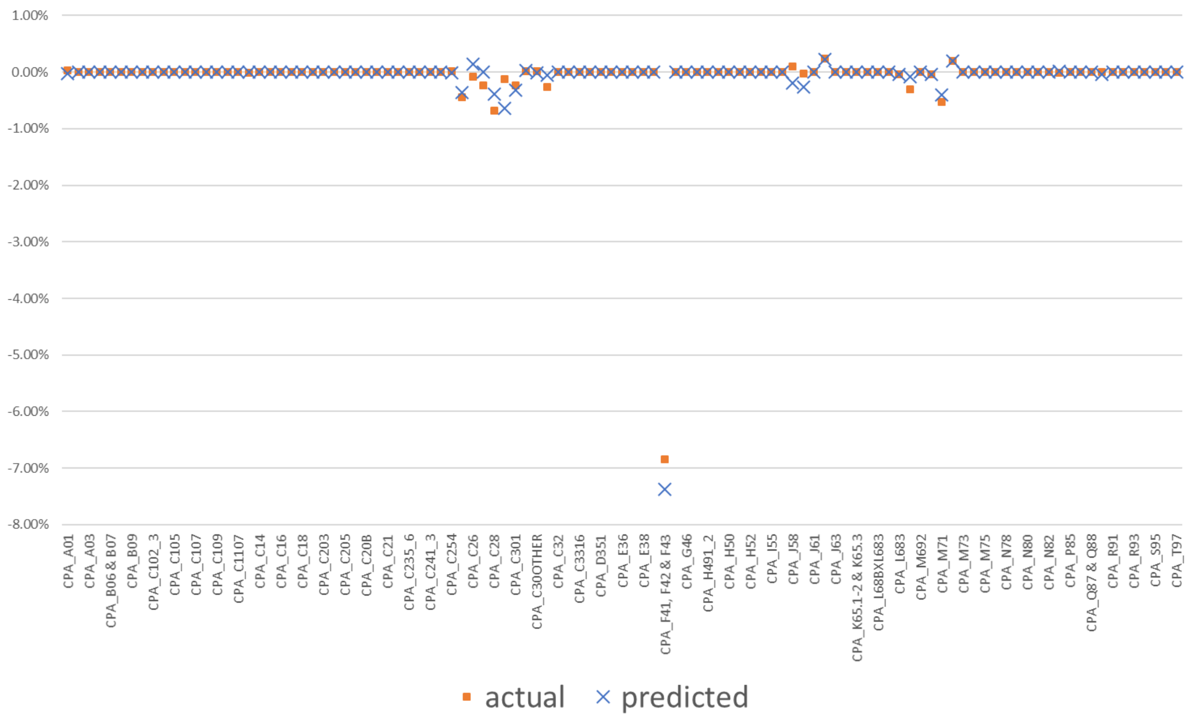

Figure 13: Actual and predicted contributions to growth in GFCF in 2020 by low-level SIC product

Figure x5

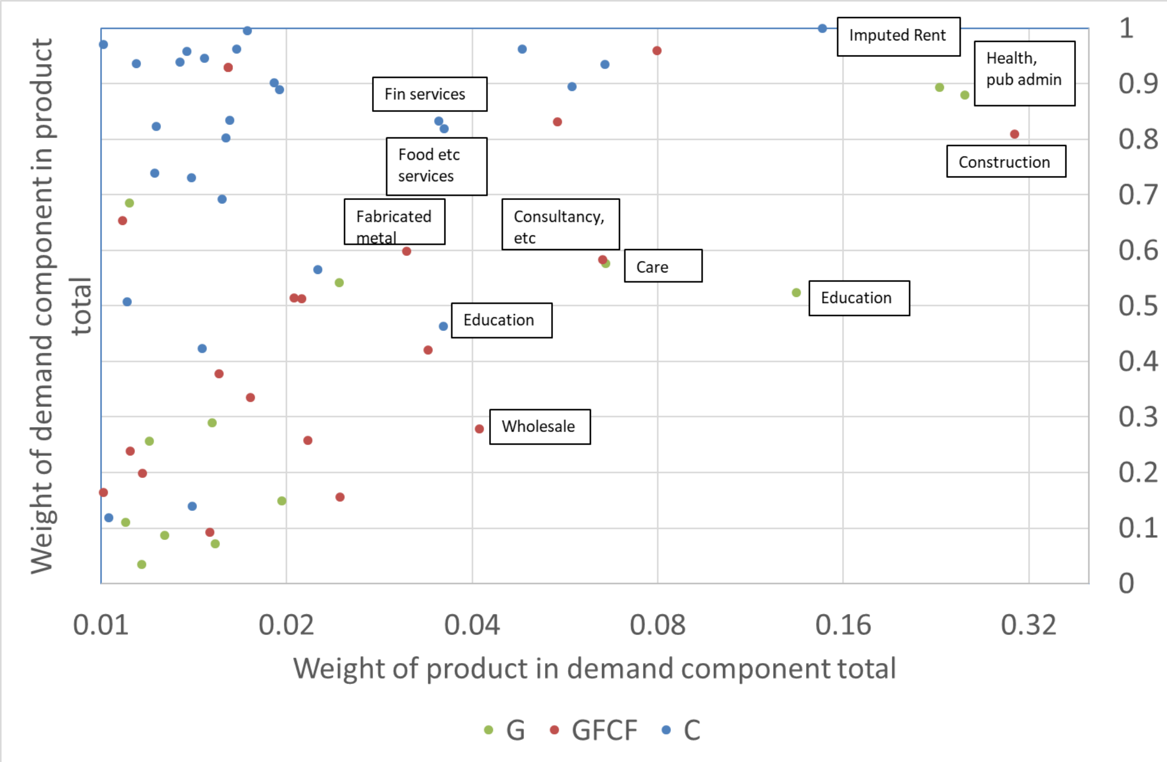

Figure 14: Share of low-level SIC product in demand component (government consumption, GFCF and household consumption) and vice versa

Figure 14 illustrates further, by comparing shares of a given product in any individual demand component (on the x axis) with the share of the same demand component in the product total. The chart brings out the feature that a few sectors make up a high share of government consumption and GFCF in particular; and these are also dominated by them, so such points appear in the top right of the chart. In such cases, growth of total domestic demand for this product implies a very high probability that government consumption and GFCF are growing. The most striking example is the high weight of construction in GFCF, which Figures 13 and 14 show result in a dominant role in predicting changes in GFCF.

In contrast consumers' expenditure has a larger number of sectors with lower weights (hence Figures 11 and 12); however, the chart illustrates that demand for many of these products is in turn dominated by consumption, so that movements in these sectors collectively gives a good handle on consumption growth.

Potential problem sectors show up as points roughly in the middle of Figure 14: these sectors are quantitatively important to, but not dominated by any particular demand, component. However, it is striking that in a number of these categories, most notably education, care and health, the breakdown between government spending and consumers' expenditure is presumably known in a fairly timely fashion, so that in practice it may prove relatively easy to derive a more accurate split between these demand components than can be achieved simply using the categories as here.

While this exercise has, due to time limitations, only been carried out for two consecutive years, the results do appear to reveal one key characteristic: at a sufficiently detailed level of disaggregation, and holding other things equal, extrapolating growth rates based on the assumption of constant shares in nominal terms does *not* appear to induce significant sources of error. This suggests that we shall need to look elsewhere for the source of the errors described above.