ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH Office for National Statistics

Nowcasting in the presence of large measurement errors and revisions

Paul Labonne and Martin Weale

ESCoE Discussion Paper 2022-05

March 2022

ISSN 2515-4664

DISCUSSION PAPER

Nowcasting in the presence of large measurement errors and revisions
Paul Labonne and Martin Weale
ESCoE Discussion Paper No. 2022-05
March 2022

Abstract

This paper extends the temporal disaggregation approach of Labonne and Weale (2020) to tackle another feature of the VAT data: the delay and highly noisy nature of the early figures. The main contribution of this paper lies in the presentation and illustration of a cleaning method which can deal with non-Gaussian features in the distribution of measurement errors such as asymmetry and extreme observations.

Keywords: Cleaning, nowcasting, measurement errors, score driven model, fat tails

JEL classification: C32, C53

Pau Labonne, Centre for Applied Macroeconomics and Commodity Prices, BI Norwegian Business School, paul.labonne@bi.no.

Published by:
Economic Statistics Centre of Excellence
National Institute of Economic and Social Research
2 Dean Trench St
London SW1P 3HE
United Kingdom
www.escoe.ac.uk

Nowcasting in the presence of large measurement errors and revisions

Paul Labonne*[1] and Martin Weale[2]

[1]Centre for Applied Macroeconomics and Commodity Prices, BI Norwegian Business School

[2]King's College London, Economic Statistics Centre of Excellence, Centre for Macroeconomics

March 2022

Abstract

This paper extends the temporal disaggregation approach of Labonne and Weale (2020) to tackle another feature of the VAT data : the delay and highly noisy nature of the early figures. The main contribution of this paper lies in the presentation and illustration of a cleaning method which can deal with non-Gaussian features in the distribution of measurement errors such as asymmetry and extreme observations.

JEL codes : C32, C53

Keywords : Cleaning, nowcasting, measurement errors, score driven model, fat tails

1 Introduction

Value-added tax (VAT) returns in the United Kingdom contain business turnover data similar to those collected in the Monthly Business Survey (MBS) run by ONS to measure GDP. However, while ONS publish monthly GDP figures about two

*Email: paul.labonne@bi.no. This work was carried out during my PhD at King's College London and has been funded by the ONS as part of the research programme of the Economic Statistics Centre of Excellence (ESCoE). We thank Ivan Petrella for his comments and suggestions on this project.

months after the end of the reference period, at this time only a relatively small proportion of the VAT data are available, and these take the form of rolling quarterly aggregates. VAT returns begin accruing shortly after the end of the reference period but take several months to be complete. Therefore, an important question is whether one can derive precise estimates of monthly output from the rolling quarterly data available two months after the end of the reference period.

By nature the VAT-based turnover data become more precise as more respondents fill their VAT returns, and early releases can be subject to biases. It should be possible to capture these revisions across releases as well as their differing levels of measurement errors by modelling them together. Labonne and Weale (2020) show how monthly figures can be derived from noisy rolling quarterly aggregates using an unobserved component model. Here this model is extended with successive releases, or snapshots, of the VAT data, in order to capture the revision pattern and thus forecast monthly output using early data.

But this nowcasting exercise is complicated by the extremely noisy nature of the early figures. Applying the Kalman filter and smoother to the original VAT releases, even when accounting for the heterogeneity across releases, produces very erratic results. It is necessary to clean the data from very large measurement errors before estimating the nowcasting model.

Two approaches are explored for cleaning the VAT data prior to nowcasting. The first approach, which is the common strategy when using state space models, consists of carrying t-tests on standardised disturbances to detect outlying observations. These observations are discarded and replaced with missing values and the model is re-estimated until no new outliers are detected. The Kalman filter - the recursive algorithm used for estimating state space models - handles missing values straightforwardly.

But estimation with original figures typically yields residuals which fail to be normally distributed. And the state space techniques presented in Labonne and Weale (2020) relies on the data being conditionally normally distributed. To model openly for non-Gaussian features in the data a second cleaning strategy relying on score driven methods introduced by Creal et al. (2013) and Harvey (2013) is exploited. This method deals with large measurement errors without systematically discarding them; instead it uses a complex weighting scheme for downweighting large prediction errors which is linked to their conditional distribution.

The uncertainty surrounding the nature of the noise affecting the VAT figures implies that combining both cleaning approaches could be beneficial. Therefore, both approaches are compared with a third strategy consisting of nowcasting separately with both cleaning approaches and averaging the resulting nowcasts.

The VAT series begin in March 2011 for the seventy-five industries for which they are available, but vintages are observable only from January 2012. The last month in which all releases are observable is September 2019; after that the data are gradually missing. The missing observations at the beginning and end of the sample are not problematic because the filtering techniques used for estimation below handle them straightforwardly by producing recursively optimal forecasts for missing values.

The next section investigates the extend of the revisions across VAT releases and show that they are biased significantly. Using a simple trimming rule it also illustrates the magnitude of the noise affecting the data and its asymmetric nature. Accordingly, section 3 shows how the bivariate model presented in Labonne and Weale (2020) can be extended with additional components for capturing these revisions, and section 4 sets out and illustrates the two approaches considered for cleaning the data of extreme measurement errors before nowcasting. Using a set Monte Carlo experiments it is shown that the score driven cleaning method outperforms the standard approach of discarding outliers through t-tests when measurement errors follow non-Gaussian processes.

The nowcasting exercise is carried out in pseudo real-time to mimic a publication schedule of a National Statistical Institute, and section 5 discusses the results. These show that the score driven cleaning approach yields smaller revisions to the monthly VAT-based output nowcasts than the t-test method, and that averaging both approaches does not help. Secondly, despite producing early estimates which tends to be revised downward, the VAT-based nowcasts can indicate timely an economic recession, as is illustrated with the Covid-19 pandemic period. Finally, although the VAT returns and MBS provide similar picture of monthly output changes overtime, there can be some persistent discrepancy in their levels, confirming the finding of Labonne and Weale (2020).

2 Descriptive analysis of the revisions and noise

Each month a new series is released which shows a VAT-based quarterly turnover observation for a new month as well as revised values for past months. The most recent observation of this vector is stored in a vector showing the first releases over time; the second observation is stored in a vector showing the second releases over time, and so on. These successive snapshots of a same variable are usually represented with a 'revision triangle' like Figure 1. The log quarterly turnover for a given industry in month t, and observable in release i, is defined as $y_{i,t}$. It is assumed that the eleventh release is the final release such that $i = 1, ..., 11$. Observations are still subject to some revisions after the eleventh release but only on a very small scale. The VAT figures are observable consistently starting from two months after the end of the reference period; hence the first release ($i = 1$) has a two-month lag.

$$\begin{bmatrix} y_{11,1} & y_{10,1} & y_{9,1} & \cdots & y_{2,1} & y_{1,1} \\ y_{11,2} & y_{10,2} & y_{9,2} & \cdots & y_{2,2} & y^{1}_{i,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{11,t} & y_{10,t} & y_{9,t} & \cdots & y_{2,t} & y_{1,t} \\ & y_{10,t+1} & y_{9,t+1} & \cdots & y_{2,t+1} & y_{1,t+1} \\ & & \ddots & \vdots & \vdots & \vdots \\ & & & y_{3,t+8} & y_{2,t+8} & y_{1,t+8} \\ & & & & y_{2,t+9} & y_{1,t+9} \\ & & & & & y_{1,t+10} \end{bmatrix}$$

Figure 1: Illustration of the revision triangle of the VAT data for a given industry. Columns and first subscripts indicate the releases while rows and second subscripts indicate the month the figures relate to. The maturity of the data increases from left to right and their timeliness increases from top to bottom. This matrix show the data observable at $t + 10$.

The VAT releases are weighted by ONS to account for the proportion of missing respondents, but they remain biased and differ in their level of measurement errors. To illustrate this fact a preliminary analysis is carried out on the revisions between the early releases and the last release, twelve months after the reference period at $i = 11$. These ten revisions are given by

$$rev^{j}_{i,t} = y^{j}_{i,t} - y^{j}_{11,t}, \quad i = 1, ..., 10, \tag{1}$$

where here $j = 1, ..., 75$ indicates the industry index.

The first two rows of table 1 show the aggregate mean revision and its standard

error for each early release. Revisions are computed at industry level using (1) and weighted using the industries' shares in total gross value added and shares of turnover covered by firms in size bands one to three. Hence the weighted mean is

$$\overline{rev}_i = \frac{\sum_{j=1}^{N}\sum_{t=1}^{T} w_{j,t} rev_{i,t}^{j}}{\sum_{i=1}^{N}\sum_{t=1}^{T} w_{j,t}}, \tag{2}$$

where

$w_{j,t}$ = Share of turnover covered by firms in size bands 1 to 3 in industry i in month $t\times$ Contribution of industry i in total gross value added in month t.

Gatz and Smith (1995) discuss several methods to compute the standard error of the mean when the data are weighted. Here the simplest approach is adopted and the standard error are computed as

$$se_i = \sqrt{\frac{1}{NT}\frac{\sum_{j}^{N}\sum_{t}^{T} w_{j,t}(rev_{i,t}^{j} - \overline{rev}_i^{j})^2}{\sum_{i}^{N}\sum_{t}^{T} w_{j,t}}}. \tag{3}$$

Table 1: Weighted mean and standard error of revisions at each maturity level with the implied z-scores.

Original data										
j	1	2	3	4	5	6	7	8	9	10
Mean	0.19	-1.27	-1.02	-0.85	-0.69	-0.58	-0.37	-0.17	-0.13	-0.08
S.E.	0.31	0.19	0.13	0.11	0.1	0.1	0.07	0.04	0.04	0.03
Z-score	0.59	-6.63	-7.78	-7.49	-6.97	-5.72	-5.27	-4.15	-2.84	-2.69
Trimmed data										
j	1	2	3	4	5	6	7	8	9	10
Mean	2.34	-0.5	-0.44	-0.39	-0.32	-0.22	-0.17	-0.11	-0.06	-0.05
S.E.	0.09	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.03	0.03
Z-score	25.11	-11.61	-11.05	-10.78	-10.26	-6.47	-5.36	-4.39	-2.11	-1.7

Note: Weighted mean of the revisions $rev_{i,t}^{j} = y_{i,t}^{j} - y_{11,t}^{j}, \quad i = 1, ..., 10$, using equation (2) with GVA weights. Standard error of the mean using equation (3)

To get an indication of the potential biases present in the early releases the third row of table 1 shows z-scores for weighted means. In aggregate, revisions are biased at all maturities and their standard errors typically decrease with maturity. As more data become available the releases become more precise and less volatile with

a decreasing bias.

The presence of these biases in the VAT data requires a flexible nowcasting approach capable of capturing large and potentially dynamic biases. The next section shows how the bivariate model presented in Labonne and Weale (2020) may be extended with an additional set of unobserved components aimed at capturing these specific features of the data.

3 Nowcasting approach

The early VAT data are incomplete because most firms submit their VAT returns with a delay. As more data accrue over time, the VAT figures are revised and their precision improves. To produce a timely monthly estimate of turnover from the VAT data it is necessary to use the earliest release, available with a two-month lag, and since the data are revised over time, the model should also take into account the accrual of new information and update past estimates. Both tasks can be achieved by modelling the eleven VAT releases together in a multivariate framework.

Building on the bivariate model presented in Labonne and Weale (2020), each VAT release is modelled as

$$\begin{aligned} y_{i,t} &= \tilde{y}_{1,t} + r_{i,t}, \\ &= \ln 3 + \frac{1}{3}x_{1,t} + \frac{1}{3}x_{1,t-1} + \frac{1}{3}x_{1,t-2} + \gamma_{1,t} + b_{1,t}^{(j)} + \beta_1 h_{1,t}^{a} + r_{i,t}, \end{aligned} \tag{4}$$

where $\tilde{y}_{1,t}$ is the VAT-based quarterly turnover *signal* representing the economically relevant information common to all release, $x_{1,t}$ the VAT-based monthly seasonally adjusted figure, $\gamma_{1,t}$ a seasonal effect, $b_{1,t}^{(j)}$ the bias of stagger j, β_1 an Easter effect and $r_{i,t}$ the observation error affecting release i. This measurement error is modelled as the sum of a bias specific to each early release $c_{i,t}$, $i = 1, ..., 10$ and white noise $\epsilon_{i,t}$:

$$r_{i,t} = c_{i,t} + \epsilon_{i,t}, \tag{5}$$

where the biases take the form of a random walk:

$$c_{i,t} = c_{i,t-1} + \tau_{i,t}, \quad \tau_{i,t} \sim N(0, \sigma_{\tau,i}), \tag{6}$$

and $\epsilon_{i,t}$ is white noise independent across releases:

$$\epsilon_{i,t} \sim N(0, \sigma_{\epsilon,i}). \tag{7}$$

By construction the last release in assumed to be unbiased. Importantly, the presence of a measurement error in the final release means that the 'true' quarterly turnover might never be observed; this is important because a substantial amount of noise remains in the final release.

For clarity the covariate is denoted as z_t and is modelled as

$$z_t = x_{2,t} + \gamma_{2,t} + \beta_2 h^a_{2,t}. \tag{8}$$

The covariate is as timely as the first release and is used to improve the measure of the seasonally adjusted VAT-based monthly turnover $x_{1,t}$. This is achieved by modelling seasonally adjusted figures, $x_{1,t}$ and $x_{2,t}$, together in a bivariate local linear trend model. Specifically the logs of the seasonally adjusted covariate and interpoland follow:

$$\begin{aligned} x_t &= \mu_t + e_t, & e_t &\sim \mathrm{N}(0, \Sigma_e), \\ \mu_{t+1} &= \mu_t + \nu_t + \xi_t, & \xi_t &\sim \mathrm{N}(0, \Sigma_\xi), \\ \nu_{t+1} &= \nu_t + \zeta_t, & \zeta_t &\sim \mathrm{N}(0, \Sigma_\zeta), \end{aligned} \tag{9}$$

where $x_t = (x_{1,t}, x_{2,t})'$, $\mu_t = (\mu_{1,t}, \mu_{2,t})'$ is the vector of dynamic trends, $\nu_t = (\nu_{1,t}, \nu_{2,t})'$ is the vector of dynamic slopes and $e_t = (e_{1,t}, e_{2,t})'$ is the vector of irregular components. Unlike the measurement error, the irregular component carries economic meaning, although both are modelled as white noise. The irregular variation in the covariate are used to help separating the irregular component in the interpoland from the measurement error. The vectors $\xi_t = (\xi_{1,t}, \xi_{2,t})'$ and $\zeta_t = (\zeta_{1,t}, \zeta_{2,t})'$ refer to the disturbances of the trend and slope components respectively.

It is assumed that the disturbances are uncorrelated across time and across unobserved components, but there can be a contemporaneous correlation within each unobserved component. It is through these contemporaneous correlations that the covariate can be useful in estimating the interpoland. Specifically, the covariance matrix Σ_h, $h = \xi, \zeta, e$, is defined as

$$\Sigma_h = \begin{pmatrix} \sigma^2_{1,h} & \rho_h \sigma_{1,h} \sigma_{2,h} \\ \rho_h \sigma_{1,h} \sigma_{2,h} & \sigma^2_{2,h} \end{pmatrix},$$

with $\sigma^2_{1,h}$ the variance of the interpoland's h component and $\sigma^2_{2,h}$ the variance of the covariate's h component.

A deterministic rolling quarterly seasonal modelis used for the VAT data :

$$\gamma_{1,t} = -\gamma_{1,t-2} - \gamma_{1,t-5} - \gamma_{1,t-8}, \tag{10}$$

where $\gamma_{1,t}$ is a three-month seasonal effect.

The seasonality in the covariate, which unlike in the VAT data, is observed monthly and is not subject to measurement errors, can be captured with a stochastic trigonometric model:

$$\gamma_{2,t} = \sum_{j=1}^{6} \gamma_{2,j,t}, \qquad \begin{bmatrix} \gamma_{2,j,t+1} \\ \gamma^*_{2,j,t+1} \end{bmatrix} = \begin{bmatrix} \cos\lambda_j & \sin\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{2,j,t} \\ \gamma^*_{2,j,t} \end{bmatrix} + \begin{bmatrix} \omega_{2,j,t} \\ \omega^*_{2,j,t} \end{bmatrix}, \tag{11}$$

where $\lambda_j = 2\pi j/12$, for $j = 1, ..., 5$, and $\gamma_{2,6,t+1} = -\gamma_{2,6,t} + \omega_{2,6,t}$. The disturbances $\omega_{2,j,t}$ and $\omega^*_{2,j,t}$ are independently and normally distributed with zero means and variance $\sigma^2_{2,\omega}$ for $j = 1, ..., 5$, and $\sigma^2_{2,\omega}/2$ for $j = 6$; for a detailed exposition of the trigonometric model and a comparison with other seasonal models see Proietti (2000).

Separately, the stagger biases are modelled explicitly using a dynamic specification taking the form of random walk:

$$b^{(j)}_{1,t+1} = b^{(j)}_{1,t} + \kappa^{(j)}_{1,t}, \qquad \kappa^{(j)}_{1,t} \sim N(0, \sigma^2_{1,\kappa j}), \quad j = 2, 3. \tag{12}$$

It is not possible to identify a bias in all of the three staggers, so the biases in the second and third staggers are defined with respect to the first stagger, whose bias is fixed to zero. The covariate is not subject to this bias.

Calendar effects can be classified in two broad categories: moving festivals and trading days effects. The Easter period is the only significant moving festival in the UK. Easter can fall either in March or in April and can overlap both months. The common approach to estimate the Easter effect is due to Bell and Hillmer (1983) and consists of setting $E_t = \beta h_t$, with E_t the Easter effect at period t, and h_t the proportion of the total number of days in the Easter period (H_t) that falls in month t. For the Easter effects to add up to zero over a year, a representation similar to Harvey (2006) is adopted:

$$E_{i,t} = \beta_i(h_{i,t} - \sum_{t=1}^{s} h_{i,t}/s) = \beta_i h^a_{i,t}, \qquad i = 1, 2, \tag{13}$$

where s is the frequency, which is twelve here, and $\sum_{t=1}^{s} h_{2,t} = 1$ while $\sum_{t=1}^{s} h_{1,t} = 3$ (the three-month aggregates ending in April and May include both March and April; hence two months for which h_t is equal to one).

In the US, businesses can see their turnover fluctuate considerably round the seven days preceding Easter and it is usual to set $H_t = 7$. However, Russ and Tariq (2017) suggest using a specific Easter period for the UK. Their results show that it is better to account for the period Good Friday to Easter Monday, the entire bank holiday period, or from Monday before Easter Sunday to Friday following it. Here the former is chosen and H_t is set to four.

State space representation and estimation

The observation and state equations can be cast together in state space form as

$$\begin{aligned} y_t &= Z_t \alpha_t, \\ \alpha_{t+1} &= T\alpha_t + R\eta_t, \qquad \eta_t \sim \mathrm{N}(0, Q), \\ \alpha_1 &\sim \mathrm{N}(a_1, P_1), \end{aligned} \tag{14}$$

where $y_t = (y_{1,t} - \log 3, ..., y_{11,t} - \log 3, z_t)'$ and $\alpha_t = (\alpha'_{1,t}, \alpha'_{2,t})'$;
$\alpha_{1,t} = (\mu_{1,t}, \mu_{1,t-1}, \mu_{1,t-2}, \nu_{1,t}, \gamma_{1,t}, ..., \gamma_{1,t-8}, e_{1,t}, e_{1,t-1}, e_{1,t-2}, \beta_1, b^1_{1,t}, b^2_{1,t}, b^3_{1,t},$
$c_{2,t}, ..., c_{10,t}, \epsilon_{1,t}, ..., \epsilon_{11,t})'$, $\alpha_{2,t} = (\mu_{2,t}, \nu_{2,t}, \gamma_{2,t}, ..., \gamma_{2,t-10}, e_{2,t}, \beta_2)'$. Appendix A shows the full matrix representation of the model.

The unknown variance parameters are estimated via maximum likelihood, where the log likelihood function is evaluated with the prediction error decomposition using the Kalman filter's output. The Kalman filter is a recursion initialised with the mean vector a_1 and covariance matrix P_1 of the initial state vector which are unknown. The trends, slopes, seasonal effects, Easter effects, stagger biases and release biases are initialised with an exact diffuse initialisation, that is to say with arbitrary means and infinite variances, while the stationary states (the irregular components and noise) are initialised with zero means and their unconditional variances.

While the Kalman filter yields optimal predictions of the state vector, which is suited to estimation and forecasting, the estimate of the state vector and its error variance given the entire sample is given by the Kalman smoother. The Kalman smoother is a backward recursion which makes use of the Kalman filter output when the parameters are set to their maximum likelihood estimates. The interpolands in

levels are constructed from the smoothed estimates of the states as

$$\exp(\hat{x}_{1,t}) = \exp(\hat{\mu}_{1,t} + \hat{e}_{1,t}), \qquad t = 1, ..., T. \tag{15}$$

The Kalman filter and smoother algorithms with exact diffuse initialisation of Durbin and Koopman (2012) and the associated log likelihood function is used.

When revisions are gradually missing towards the end of the series, as is illustrated with the revision triangle in figure 1, the dimension of the observation vector y_t is adapted accordingly. Hence, the information from early revisions can be used although not all revisions are available. The Kalman filter produces optimal forecasts for these missing values.

4 Cleaning strategies

Estimating the nowcasting model presented in section (3) using the original VAT data produces poor results. This is because the VAT releases are subject to extremely large measurement errors; these are illustrated for the aggregated data in the first panel of figure 8. Some of these errors are likely to come from firms reporting turnover figures with decimal errors. Indeed, since only small and medium size businesses are analysed, a single business cannot have an important impact on the industry aggregate unless it is a reporting error. This also means that outlying observations are most often positive because abnormally low figures do not affect the aggregate very much.

To improve the VAT-based monthly output nowcasts it is necessary to clean the data from these very large measurement errors beforehand. For this two separate approaches are explored because there is a limited understanding regarding the behaviour of the measurement affecting the VAT data.

4.1 Method 1: Sequentially discarding outlying observations

The first cleaning method consists of carrying out t-tests sequentially on the standardised observation errors. These are given by

$$\hat{\epsilon}^s_{i,t} = \hat{\epsilon}_{i,t}/\sqrt{H_{i,t}}, \quad i = 1, ..., 11, \tag{16}$$

where $H_{i,t}$ is the estimated variance. To derive the standardised observation errors the bivariate model of Labonne and Weale (2020) is estimated at industry level for each release, where the VAT data is modelled together with the covariate. After estimation the Kalman smoother is used to retrieve the smoothed observation errors and their estimated variance.

Each standardised observation error can be tested for significance using a two-tailed t-test with the null hypothesis that the corresponding observation is not an outlier. Outlying observations thus detected are discarded and replaced with missing values. Next the bivariate model is estimated again and a new set of t-tests are carried out on the standardised disturbances. This process is iterated until no standardised disturbances reach the threshold indicating outliers.

It is common to choose a confidence interval of 95% which is associated with critical values of ± 1.96. However, this is a multiple testing problem and too many outliers would be detected by chance. Because of the large number of industries studied, analysing each outlying observation individually to decide whether or not it should be excluded is not feasible. Therefore the confidence interval is set to a very conservative range of 99.9% which is associated with critical values of ± 3.3.

Outliers in these errors indicate observations which cannot be explained by the model. These can be due to extremely large measurement errors or misspecification in the model. Setting a high critical value increases the confidence that the outlying observation detected is a measurement error and not a structural break or one-off economically relevant event. In addition, since the data are rolling quarterly aggregates, an outlier carrying economic meaning should appear in three consecutive observation errors.

An important drawback of the method relying on t-tests is that it requires choosing critical values indicating outlying observations, and this inevitably generates a trade-off between discarding too many observations and taking the risk of not detecting large measurement errors. It also assumes that measurement errors are normally distributed. However, as discussed above, measurement errors in the VAT data are most often positive; this could be one source of non-normality. The second cleaning method proposed below deals with outlying observations without discarding them completely and can handle non-Gaussian features.

4.2 Method 2: Downweighting observations with a score driven model

Creal et al. (2013) and Harvey (2013) derive predictive filters based on the score of the conditional (or predictive) likelihood, which can arise from a wide range of distributions. These score driven models provide an alternative approach for cleaning where outlying observations are downweighted instead of being discarded completely. The downweighting scheme depends on the magnitude of the prediction errors and their estimated distribution. Using this approach it is possible to model openly the non-Gaussian features in the distribution of the measurement errors. In a cleaning context this method is used to derive pseudo observations where measurement errors are downweighted depending on their magnitude.

A general location model

The model is based on the general location Dynamic Conditional Score (DCS) model of Harvey (2013) and Harvey and Luati (2014). The model is univariate; each release is modelled separately. The general location model is

$$\begin{aligned} y_{i,t} &= Z_t a_t + v_t, \\ a_{t+1} &= T a_t + \kappa u_t, \end{aligned} \tag{17}$$

for releases $i = 1, ..., 11$, where κ is a vector of unknown parameters and u_t is the scaled score vector defined as

$$u_t = \frac{\partial \ell_t}{\partial a_t}.s_t^{-1}, \tag{18}$$

where $\ell_t = \ln p(y_{i,t}|Z_t a_t; \Theta)$ is the predictive log likelihood with Θ a vector of fixed parameters. The scaling factor s_t is typically related to the information matrix. Conditional on past observations $y_{i,t}$ follows a generalised asymmetric student-t (AST) distribution of Zhu and Galbraith (2010). Hence

$$\begin{aligned} \ell_t = &- \ln\sigma - \frac{\nu_1+1}{2}\ln\Big[1 + \frac{1}{\nu_1}\Big(\frac{y_{i,t} - Z a_t}{2\alpha\sigma K(\nu_1)}\Big)^2\Big]\ 1(y_{i,t} \le Z_t a_t) \\ &- \frac{\nu_2+1}{2}\ln\Big[1 + \frac{1}{\nu_2}\Big(\frac{y_{i,t} - Z_t a_t}{2(1-\alpha)\sigma K(\nu_2)}\Big)^2\Big]\ 1(y_{i,t} > Z_t a_t) \end{aligned} \tag{19}$$

where σ is the scale parameter, α is the skewness parameter which can take values in $[0, 1]$, ν_1 and ν_2 are respectively the left and right tail parameters, $K(\nu) = \Gamma((\nu+1)/2)/(\sqrt{\nu\pi}\Gamma(\nu/2))$ ($\Gamma(.)$ being the Gamma function) and $1(x)$ is an indicator

variable equal to one if statement x is true and zero otherwise. The distribution is skewed to the right if $\alpha < 1/2$ and to the left if $\alpha > 1/2$.

It is possible to recover well-known Student's t-distributions by restricting and redefining the parameters of the AST. With $\nu_1 = \nu_2$ (symmetric tails), $\alpha = 0.5$ (no skewness) and redefining the scale as $\sigma = \sigma K(\nu)$, the AST reduces to the Student's t-distribution. With $\nu_1 = \nu_2$ and redefining the shape and scale parameters as $\alpha = 1/(1+\gamma^2)$ and $\sigma = (\gamma + 1/\gamma)\sigma K(\nu)/2$ the AST is equivalent to the skewed t-distribution of Fernández and Steel (1996). The skewed Student's t-distributions of Azzalini and Capitanio (2003) and Gomez et al. (2007) cannot be recovered by simple reparametrisation.

An attractive feature of (19) is that each tail can have distinct rates of decay captured by the tail parameters. This is particularly useful for the VAT data because extreme measurement errors are most often positive. Hence the right side of the distribution is likely to exhibit a fat tail but not necessarily the left side.

The score vector is

$$\begin{aligned}\frac{\partial \ell_t}{\partial a_t} = &\frac{\nu_1 + 1}{1 + \frac{1}{\nu_1}\left(\frac{v_t}{2\alpha\sigma K(\nu_1)}\right)^2} \cdot \frac{v_t}{\nu_1 (2\alpha\sigma K(\nu_1))^2}\ 1(y_{i,t} \leq Z_t a_t) \\ &+ \frac{\nu_2 + 1}{1 + \frac{1}{\nu_2}\left(\frac{v_t}{2(1-\alpha)\sigma K(\nu_2)}\right)^2} \cdot \frac{v_t}{\nu_2 (2(1-\alpha)\sigma K(\nu_2))^2}\ 1(y_{i,t} > Z_t a_t).\end{aligned} \tag{20}$$

and the scaling factor is

$$\begin{aligned}s_t = &\frac{\nu_1 + 1}{\nu_1 (2\alpha\sigma K(\nu_1))^2}\ 1(y_{i,t} \leq Z_t a_t) \\ &+ \frac{\nu_2 + 1}{\nu_2 (2(1-\alpha)\sigma K(\nu_2))^2}\ 1(y_{i,t} > Z_t a_t).\end{aligned} \tag{21}$$

The scaled score (18) thus takes the form

$$\begin{aligned}u_t = &\frac{v_t}{1 + \frac{v_t^2}{\nu_1 (2\alpha\sigma K(\nu_1))^2}}\ 1(y_{i,t} \leq Z_t a_t) \\ &+ \frac{v_t}{1 + \frac{v_t^2}{\nu_2 (2(1-\alpha)\sigma K(\nu_2))^2}}\ 1(y_{i,t} \leq Z_t a_t).\end{aligned} \tag{22}$$

The form of (22) is close to the scaled score of Harvey (2013) (page 96), where the scaling factor diverges slightly from the information quantity. Notably $u_t = v_t$ in the Gaussian case when $\nu_1, \nu_2 \to \infty$ and $\alpha = 1/2$. The distance between the prediction error and the scaled score thus indicates the degree of nonlinear weighting

and divergence from the Gaussian model.

Figure 2 shows the scaled score as a function of the prediction error. Low tail parameters downweight the effect of large prediction errors, while the scaled score reduces to the prediction error when tail parameters tend to infinity, which is the behaviour intended with this specific choice of scaling factor. Note that low tail parameters do not induce an overweighting of small prediction errors, as is typically the case when downweighting the score with its information quantity (see notably Delle-Monache and Petrella (2017)). The absence of overweighting is consistent with a cleaning application where observation errors are downweighted whereas common observations remains unaffected.

To compare the score driven with the typical approach of discarding outliers, 2 also plots the response function driving the latent states when the distribution of the prediction error is constrained to be Gaussian (in this case, score driven models for location parameters can be equivalent to the Kalman filter at steady state, see Buccheri et al. (2021)) and outliers are discarded with t-tests. This response function is an approximation of the response function implied by Gaussian state space models; it is only an approximation because observation errors in state space models are derived using the Kalman smoother which, unlike the Kalman filter, makes use of the entire sample. While the use of t-tests generates a discontinuity in the response function, the scaled score downweights prediction errors gradually depending on their magnitude.

The benefit of the asymmetric Student's t-distribution is illustrated with a low tail parameters for negative prediction errors (dotted blue line). This set of parameters yield a larger downweighting of positive errors but a linear response for negative prediction errors. This behaviour of the scaled score with asymmetric tails is especially fitted for the VAT data because observation errors tend to be positive only. Finally, although low tail parameters never yield a response function larger than the prediction error, this is not a limitation when the model is estimated; indeed low tail parameters typically yield a smaller scale parameter which in turn gives a greater scaled score (see equation (22)).

The recursion is initialised with the initial state vector a_1 which is typically unknown. While with the standard Kalman filter one can resort to a diffuse initialisation, this is not possible with score driven models. Instead the initial state vector is estimated via maximum likelihood along the other unknown parameters. Specific-

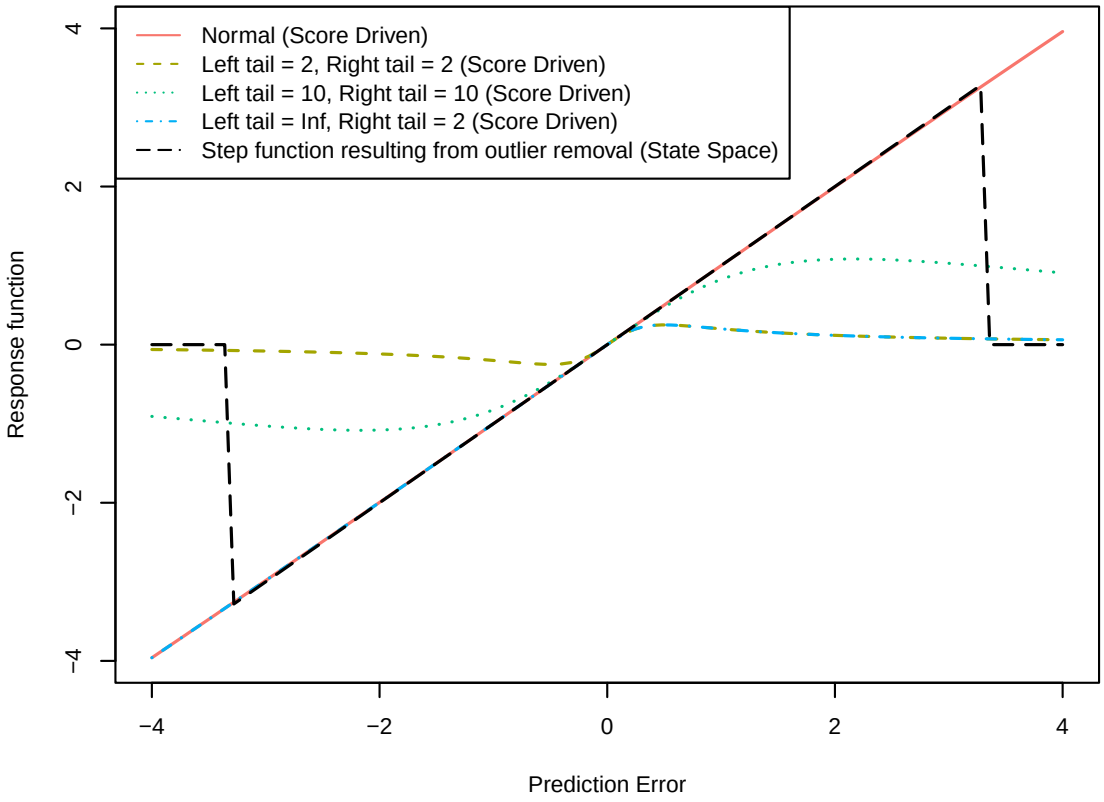

Figure 2: Illustration of the response function driving the unobserved components in score driven and state space models. In a score driven model the response function is the scaled score u_t; In a state space model discarding outliers yields a linear response function for prediction errors not detected as outliers and zero otherwise (critical value of +-3.3). The scale parameter is set to one ($\sigma = 1$).

ally, the vector of unknown fixed parameter $\Theta = (\nu_1, \nu_2, \alpha, \sigma, \kappa, a_1')'$ is estimated by numerical maximisation of the log likelihood as

$$\hat{\Theta} = \arg\max_{\Theta} \sum_{t=3}^{N} \ell_t.$$

To improve estimation an analogous Gaussian unobserved components model is estimated beforehand and the resulting initial smoothed state vector is used as starting values for a_1.

Smoothing

It is not possible to generate directly smoothed unobserved components with score driven techniques, but pseudo observations where non-Gaussian features have been alleviated can be retrieved. This is what Caivano et al. (2016) propose through an iterative algorithm which makes use of a Gaussian unobserved components model. This method consists of the following iterative scheme

1: Estimate the score driven model (17) with the original data $y_{i,t}$;

2: Generate pseudo observations as $\tilde{y}_{i,t} = Z_t a_t + u_t$;

3: Estimate the unobserved components model with the pseudo observations $\tilde{y}_{i,t}$ and retrieve the smoothed estimates $\hat{\alpha}_t$;

4: Set $v_t = y_{i,t} - Z_t \hat{\alpha}_t$ in (20) which yields new values of u_t through (22);

5: Generate pseudo observations as $\tilde{y}_{i,t} = Z_t \hat{\alpha}_t + u_t$;

6: Iterate steps 3 to 5 until convergence.

If the estimated distribution is Gaussian $u_t = v_t$, which implies that the pseudo observations constructed from the score driven recursion are equal to the original observations. In this case the smoothing algorithm is reduced to the Kalman smoother. Once the algorithm has converged the pseudo observations $\tilde{y}_{i,t}$ are used for estimating model (14).

The pseudo observations thus derived are not free of measurement errors, but the remaining noise should be normally distributed and can therefore be captured effectively with standard state space techniques.

4.3 Monte Carlo experiments

This section investigates the stability and efficiency of the score driven method presented above when applied to a relatively short sample (approximately 100 observations) while comparing different level of restrictions on the model generating the prediction errors. The most efficient model specification is then used in another experiment designed to analyse the effectiveness of the score driven cleaning approach compared to the more common strategy of discarding outliers through t-tests.

On the stability and efficiency of the score driven method

For studying the stability of the score driven estimation, synthetic data are generated using model (17) with different specifications for the distribution of the prediction error v_t. Specifically, in each specification the vector of prediction error v_t is generated using an asymmetric student-t distribution with different levels of restriction on its parameters. This is useful to understand if a greater flexibility in the distribution of the prediction error is worth the increase in the number of parameters to estimate, which can be difficult with a relatively short sample. The efficiency of the score driven estimation is investigated using the root mean square error between the estimated pseudo observation $Z_t\hat{a}_t + \hat{u}_t$ and its true value (stored when generating the data). The results are calculated from 100 simulations for each model specification.

The different model restrictions tested in the Monte Carlo experiment are shown in table 2. The most flexible model can be skewed and have different tail parameters. Restricting the skewness to zero but allowing for distinct tail parameters yields to a Student-t model with distinct tail parameters. Conversely, restricting the tail parameters to be identical but allowing the distribution to be skewed yields a Skewed student-t model. Restricting the tail parameters to be identical and constraining the skewness to zero yields the Student-t model, while constraining the tail parameters to very large or infinite values but letting the skewness parameter vary yields a Skew Normal distribution. Finally, restricting the tail parameters to very large or infinite values and fixing the skewness to zero gives the Normal distribution.

Tables 3 to 11 show the results. Overall having distinct tail parameters is important; however, constraining the skewness to zero yields to better results. Next a new set of simulations examines the performance of the score driven approach with this asymmetric distribution compared to the classic approach of discarding outliers through t-tests.

Table 2: Model specifications for the prediction error in the Monte Carlo experiments. All specifications are derived from the Asymmetric Student-t distribution with different degrees of restriction on the parameters.

Model	Location	Left tail	Right tail	Scale	Shape
FT Skew Asy.	$\mu > 0$	$\nu_1 > 0$	$\nu_1 > 0$	$\sigma > 0$	$1 > \alpha > 0$
FT Asy.	$\mu > 0$	$\nu_1 > 0$	$\nu_1 > 0$	$\sigma > 0$	$\alpha = 0.5$
FT Skew	$\mu > 0$	$\nu > 0$		$\sigma > 0$	$1 > \alpha > 0$
FT	$\mu > 0$	$\nu > 0$		$\sigma > 0$	$\alpha = 0.5$
Skew	$\mu > 0$	$\nu = \infty$		$\sigma > 0$	$1 > \alpha > 0$
Gaussian	$\mu > 0$	$\nu = \infty$		$\sigma > 0$	$\alpha = 0.5$

Table 3: Simulation results with true parameters set to $\boldsymbol{\alpha = 0.5}$ $\boldsymbol{\nu_2 = 1}$; $\sigma = 5$ $\nu_1 = 250$ $k_1 = 0.5$

Model	AIC	ν_1	ν_2	k_1	RMSE	σ	α
FT Skew Asy.	-375.87	157.96	1.12	1.13	0.76	4.14	0.42
FT Asy.	-373.91	146.72	0.93	0.98	0.66	4.2	0.5
FT Skew	-357.25	1.28	1.28	2.2	1.26	3.47	0.44
FT	-356.7	1.1	1.1	2.06	1.19	3.51	0.5
Skew	-191.2	250	250	0.12	4.36	171.37	0.03
Gaussian	-116.94	250	250	0.13	4.42	185.91	0.5

Table 4: Simulation results with true parameters set to $\boldsymbol{\alpha = 0.4}$ $\boldsymbol{\nu_2 = 1}$; $\sigma = 5$ $\nu_1 = 250$ $k_1 = 0.5$

Model	AIC	ν_1	ν_2	k_1	RMSE	σ	α
FT Skew Asy.	-361.77	167.55	0.96	1.21	0.69	3.71	0.35
FT Asy.	-356.27	174.73	0.76	0.77	0.51	3.94	0.5
FT Skew	-344.93	0.83	0.83	2.2	1.05	2.82	0.48
FT	-346.64	0.81	0.81	2.27	1.06	2.73	0.5
Skew	-147.93	250	250	0.1	4.35	31.58	0.02
Gaussian	-62.89	250	250	0.1	4.47	53.65	0.5

Table 5: Simulation results with true parameters set to $\boldsymbol{\alpha = 0.3}$ $\boldsymbol{\nu_2 = 1}$; $\sigma = 5$ $\nu_1 = 250$ $k_1 = 0.5$

Model	AIC	ν_1	ν_2	k_1	RMSE	σ	α
FT Skew Asy.	-342.97	230.21	2.25	0.9	0.75	4.33	0.17
FT Asy.	-327.04	225.32	1.43	0.47	0.6	4.33	0.5
FT Skew	-306.46	0.99	0.99	2.17	1.26	3.71	0.43
FT	-305.26	0.86	0.86	2.07	1.27	3.66	0.5
Skew	-39.34	250	250	0.1	10.77	483.02	0.03
Gaussian	39.79	250	250	0.1	10.77	506.4	0.5

Table 6: Simulation results with true parameters set to $\boldsymbol{\alpha = 0.5 \quad \nu_2 = 3}$; $\sigma = 5 \quad \nu_1 = 250 \quad k_1 = 0.5$

Model	AIC	ν_1	ν_2	k_1	RMSE	σ	α
FT Skew Asy.	-442.72	153.5	38.48	0.8	0.45	4.01	0.42
FT Asy.	-439.23	150.49	23.02	0.72	0.38	4.18	0.5
FT Skew	-443.4	49.55	49.55	1.06	0.59	3.85	0.37
FT	-439.37	41.72	41.72	1.04	0.6	3.92	0.5
Skew	-437.99	250	250	0.29	0.52	5.2	0.22
Gaussian	-422.38	250	250	0.3	0.53	5.74	0.5

Table 7: Simulation results with true parameters set to $\boldsymbol{\alpha = 0.4 \quad \nu_2 = 3}$; $\sigma = 5 \quad \nu_1 = 250 \quad k_1 = 0.5$

Model	AIC	ν_1	ν_2	k_1	RMSE	σ	α
FT Skew Asy.	-438.26	206.28	14.12	0.81	0.47	4.02	0.28
FT Asy.	-429.95	199.63	2.38	0.67	0.35	4.15	0.5
FT Skew	-437.39	22.88	22.88	1.26	0.65	3.63	0.26
FT	-426.02	6.01	6.01	1.11	0.63	3.87	0.5
Skew	-430.74	250	250	0.25	0.72	5.41	0.08
Gaussian	-401.75	250	250	0.26	0.73	6.39	0.5

Table 8: Simulation results with true parameters set to $\boldsymbol{\alpha = 0.3 \quad \nu_2 = 3}$; $\sigma = 5 \quad \nu_1 = 250 \quad k_1 = 0.5$

Model	AIC	ν_1	ν_2	k_1	RMSE	σ	α
FT Skew Asy.	-436.09	219.83	8.96	0.82	0.53	3.95	0.15
FT Asy.	-419.19	215.88	1.63	0.64	0.39	4.04	0.5
FT Skew	-427.77	12.6	12.6	1.53	0.71	3.38	0.28
FT	-413.76	9.56	9.56	1.36	0.75	3.6	0.5
Skew	-414.27	250	250	0.19	1.02	5.88	0.03
Gaussian	-370.98	250	250	0.2	1.04	7.54	0.5

Table 9: Simulation results with true parameters set to $\boldsymbol{\alpha = 0.5 \quad \nu_2 = 250}$; $\sigma = 5 \quad \nu_1 = 250 \quad k_1 = 0.5$

Model	AIC	ν_1	ν_2	k_1	RMSE	σ	α
FT Skew Asy.	-466.59	147.93	162.4	0.77	0.23	4.08	0.47
FT Asy.	-462.55	127.95	134.07	0.75	0.24	4.25	0.5
FT Skew	-467.09	173.79	173.79	0.74	0.2	4.2	0.47
FT	-463.8	143.6	143.6	0.73	0.2	4.31	0.5
Skew	-467.49	250	250	0.53	0.07	4.45	0.48
Gaussian	-464.1	250	250	0.52	0.06	4.58	0.5

Table 10: Simulation results with true parameters set to $\boldsymbol{\alpha = 0.4}$ $\boldsymbol{\nu_2 = 250}$; $\sigma = 5$ $\nu_1 = 250$ $k_1 = 0.5$

Model	AIC	ν_1	ν_2	k_1	RMSE	σ	α
FT Skew Asy.	-471.89	193.63	180.92	0.66	0.18	4.07	0.24
FT Asy.	-459.99	178.46	116.8	0.67	0.21	4.35	0.5
FT Skew	-472.04	189.51	189.51	0.69	0.17	4.12	0.27
FT	-461.24	145.06	145.06	0.65	0.18	4.41	0.5
Skew	-474.47	250	250	0.52	0.06	4.3	0.23
Gaussian	-462.05	250	250	0.5	0.06	4.63	0.5

Table 11: Simulation results with true parameters set to $\boldsymbol{\alpha = 0.3}$ $\boldsymbol{\nu_2 = 250}$; $\sigma = 5$ $\nu_1 = 250$ $k_1 = 0.5$

Model	AIC	ν_1	ν_2	k_1	RMSE	σ	α
FT Skew Asy.	-469.5	208	162.17	0.69	0.23	4.07	0.17
FT Asy.	-454.32	204.48	95.99	0.69	0.26	4.41	0.5
FT Skew	-470.87	182.87	182.87	0.73	0.22	4.11	0.17
FT	-454.88	133.44	133.44	0.71	0.22	4.48	0.5
Skew	-473.94	250	250	0.5	0.06	4.31	0.13
Gaussian	-455.11	250	250	0.49	0.06	4.8	0.5

The score driven approach yields a clear benefit when one tail exhibits a very low rate of decay

In this section a score driven model with an AST distribution featuring distinct tail parameters is applied to clean the VAT data using the smoothing algorithm discussed above. This nonlinear approach is compared to the more common strategy of discarding outliers through t-tests after estimating a Gaussian state space model. For this experiment, the synthetic data are generated using the Kalman filter where the parameters of the model are chosen to be close to the parameters estimated with the VAT data. In a second step, noise following an asymmetric Student's t-distribution is added. Different specification for the asymmetric Student's t-distribution are compared.

Table 12 shows the results. The score driven approach performs significantly better when the tail parameters are low, while its relative advantage decreases when the tail parameters get close to 3. Experimentation with the VAT data shows that its tail parameters are close to one for many industries, suggesting that the score driven approach should be better strategy here.

Table 12: Simulations investigating the efficiency of the cleaning method. Root mean square log approximation error (×100). The log approximation error is given by the difference between the log observation y_t minus the log smoothed observation $\hat{y}_t$. With the score driven approach the pseudo observation is generated with the iterative score driven smoothing scheme presented in the previous section; when observations are discarded with t-tests the pseudo observation is the signal in that period given by the Kalman smoother. Results from 100 simulations. Bold figures indicate the lowest value.

ν_2	α	Score driven approach	Discarding outlying observations
1	0.3	**4.48**	5.38
1	0.4	**4.49**	5.31
1	0.5	**6.38**	6.51
3	0.3	2.35	**2.34**
3	0.4	**2.07**	2.07
3	0.5	**1.9**	1.91
250	0.3	**2.1**	2.1
250	0.4	**1.87**	1.88
250	0.5	**1.82**	1.83

4.4 Averaging both cleaning approaches

Although measurement errors decrease as more firms submit their VAT returns, the final estimate remains very noisy. Consequently, there is no benchmark series which can be used to compare the cleaned series with, complicating the evaluation of the cleaning methods. This is in contrast to the Monte Carlo experiments of the previous section where the data generating process is know and hence the evaluation straightforward. These experiments are useful for deciding which degree of flexibility in the distribution of the measurement error is reasonable, and how the approach compares to the standard strategy of using t-tests under a specific environment. But they should be used with care when drawing conclusions on the modelling approach most adapted to the VAT data because the nature of the measurement errors affecting them remains mostly unknown. In this context, where there is an important uncertainty regarding the appropriate modelling framework, a popular solution is to average estimates from a set of models instead of choosing a particular model assumed to be optimal.

The simplest approach to model averaging consists of using equal weights for all models, but weighting models using Bayesian Model Averaging (BMA) techniques typically works better, especially in an uncertain and changing environment. BMA

techniques rely on models' likelihood functions, that is the likelihood of observing the data given a particular model (which includes regressors, relationship between variables and prior distributions). As such models are weighted depending on the likelihood that the observed data arise from their processes; hence, importantly, the data used for model evaluation must be common to all models.

The model selection problem faced in this paper diverges from those typically encountered in the economic and forecasting literature because the uncertainty arises from the cleaning method, and the very nature of cleaning consists of altering the data. While BMA is used to overcome the uncertainty arising from model specification when there is no or very little uncertainty regarding the data, the uncertainty here arises from the data used for estimation and forecasting. It is difficult to know which dataset - the data when outliers are excluded using t-tests or the data arising from the score driven cleaning method - is closer to reality. In this context BMA is not applicable.

Given the high uncertainty surrounding the data and the inability to compare the different cleaning strategy to a common benchmark, since the true data are never observed, using equal weights in the averaging scheme is the only reasonable option.

4.5 Illustration of the cleaning strategies using one industry as a case study

This section illustrates both cleaning strategies using one industry as a case study. This industry is one of the largest in terms of gross value added for small and medium size firms.

T-test approach

The t-test approach is illustrated first. In this strategy outlying observations are discarded sequentially using t-tests. Outlying observations are detected by testing the standardised observation errors given by (16) for significance with a critical value of +- 3.3. The standardised observation errors resulting from the first round of estimation using the original VAT data are shown in the first panel of figure 3. Each of those points may be interpreted as two-tailed t-tests and the horizontal lines indicate the threshold of +/- 3.3. Three observations are outside this threshold and therefore considered as outliers.

A complimentary standard robustness check with Gaussian state space models consists of analysing the one-step ahead forecast errors given by

$$v_t^s = v_t/\sqrt{F_t}, \tag{23}$$

with F_t the variance of the prediction error.

The Kalman filter relies on normally distributed one-step ahead forecast errors, an assumption which can be verified by testing the standardised prediction errors given by (23) for normality using the skewness and kurtosis statistics. Bowman and Shenton (1975) show that if the normality assumption is respected these statistics should asymptotically be distributed as

$$\mathrm{S} \sim N(0, 6/n), \quad \mathrm{K} \sim N(0, 24/n),$$

where n is the sample size.

The second panel of figure 3 shows the standardised prediction error resulting from the first round of estimation. Panel three shows the histogram and kernel density estimate of these errors, while panel four shows their associated Q-Q plot. From these figures it is clear that the standardised prediction errors are not normally distributed. This is confirmed with skewness and kurtosis statistics of 3.33 and 15.56, which yield z-scores of 11.93 and 22.49, both clearly rejecting the null hypothesis of normally distributed errors.

In the second round of estimation the observations previously detected as outliers in the first round are discarded and replaced with missing values. The results are presented in figure 4. There is one outliers in the standardised observation errors, and the standardised prediction errors fail again to be normally distributed with skewness and kurtosis statistics of 5.2 and 14.

Discarding the outlier found in the second round of estimation and re-estimating the model yields the results presented in figure 5. No outliers are detected, and with skewness and kurtosis statistics of -0.06 and 2.62 (giving z-scores of -0.19 and -0.67) the null hypothesis that the standardised prediction error is normally distributed is not rejected any more. Since no new outliers are detected, the data after the second round of estimation can be used for estimating the nowcasting model.

Score driven approach

Score driven models provide an alternative approach for cleaning where outlying observations are downweighted instead of being discarded completely.

The estimated degrees of freedom with the score driven method are 250 for the left tail and 0.62 for the right tail. The degrees of freedom for the right tail is very low, but this is necessary to capture the extremely large outlying observations observable in the first plot. The left tail, on the other hand, is Gaussian, hence negative prediction errors are never downweighted and the pseudo observations are equal or below the original figures, but never above.

Figure 6 illustrates the downweighting behaviour of the score driven cleaning approach. The red line shows the downweighting applied to the data, which are shown using the vertical bars. Large outliers are downweighted whereas relatively small errors are not affected. Importantly negative prediction errors are not subject to any downweighting due to the very large tail parameter on the left side of the distribution.

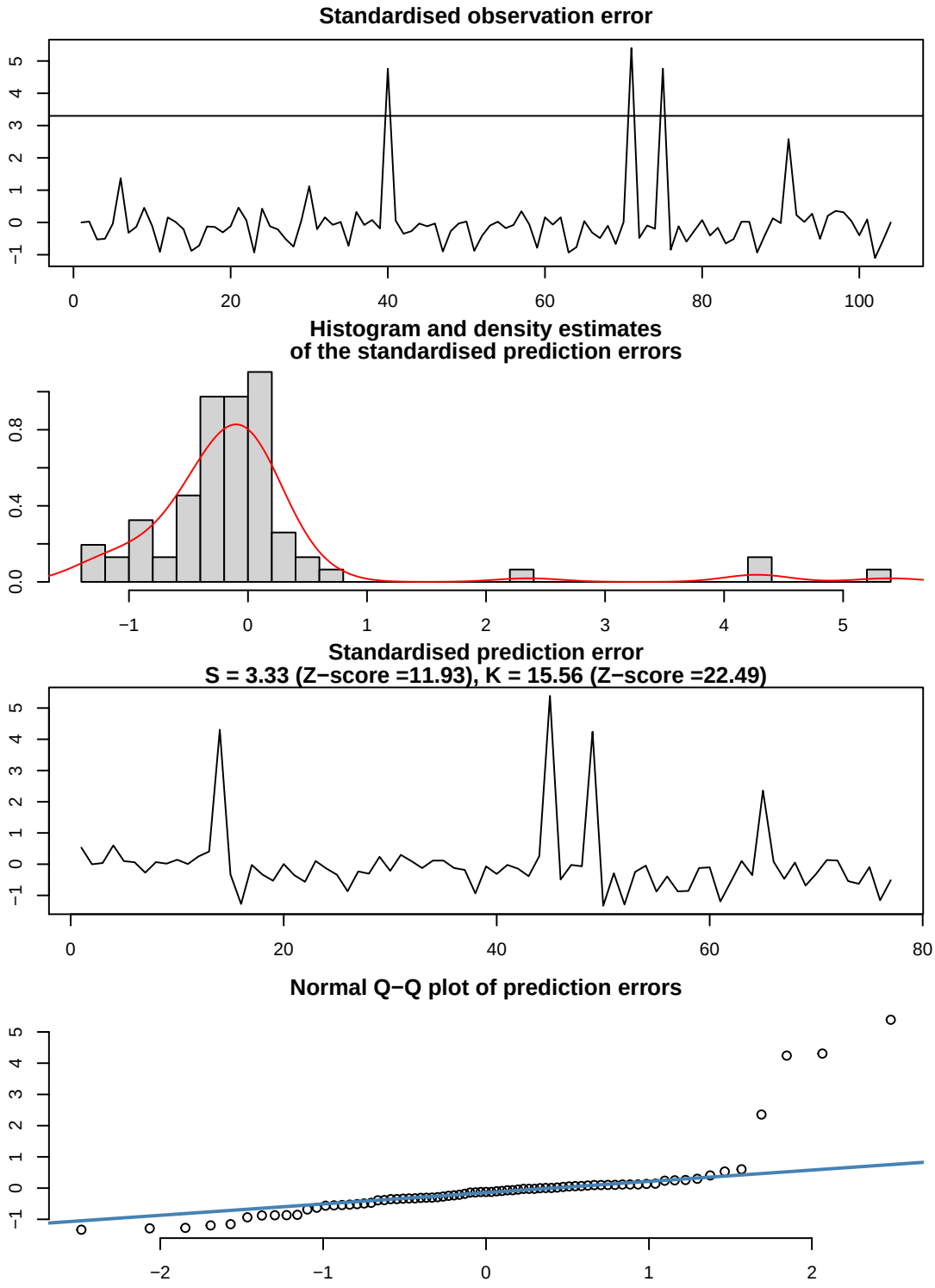

Figure 3: Results from the bivariate model estimated with VAT and MBS data for the transport industry. The VAT data are taken from the first release. The density of the standardised prediction errors is derived using a kernel density estimation with a Gaussian kernel (density() function in R). The first 25 standardised prediction errors are excluded because of the diffuse initialisation. Only outputs from the model with which it is not possible to re-construct the original data are shown to respect the confidentiality of the VAT respondents.

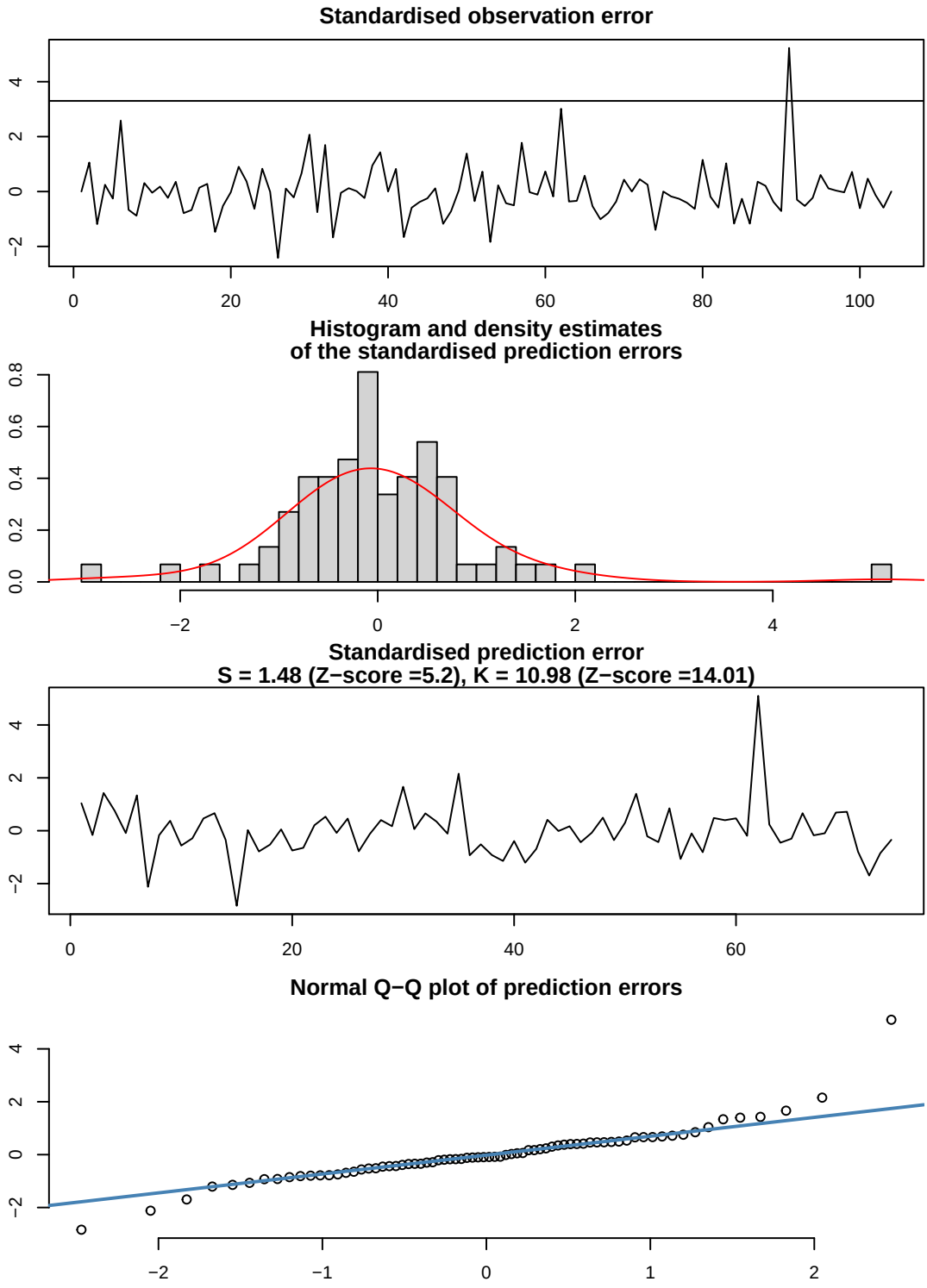

Figure 4: Results from the bivariate model estimated with VAT and MBS data for the transport industry. The VAT data are taken from the first release. **VAT figures at time points 40, 73 and 76 are discarded**. The density of the standardised prediction errors is derived with a kernel density estimation with a Gaussian kernel (density() function in R). Only outputs from the model with which it is not possible to re-construct the original data are shown to respect the confidentiality of the VAT respondents.

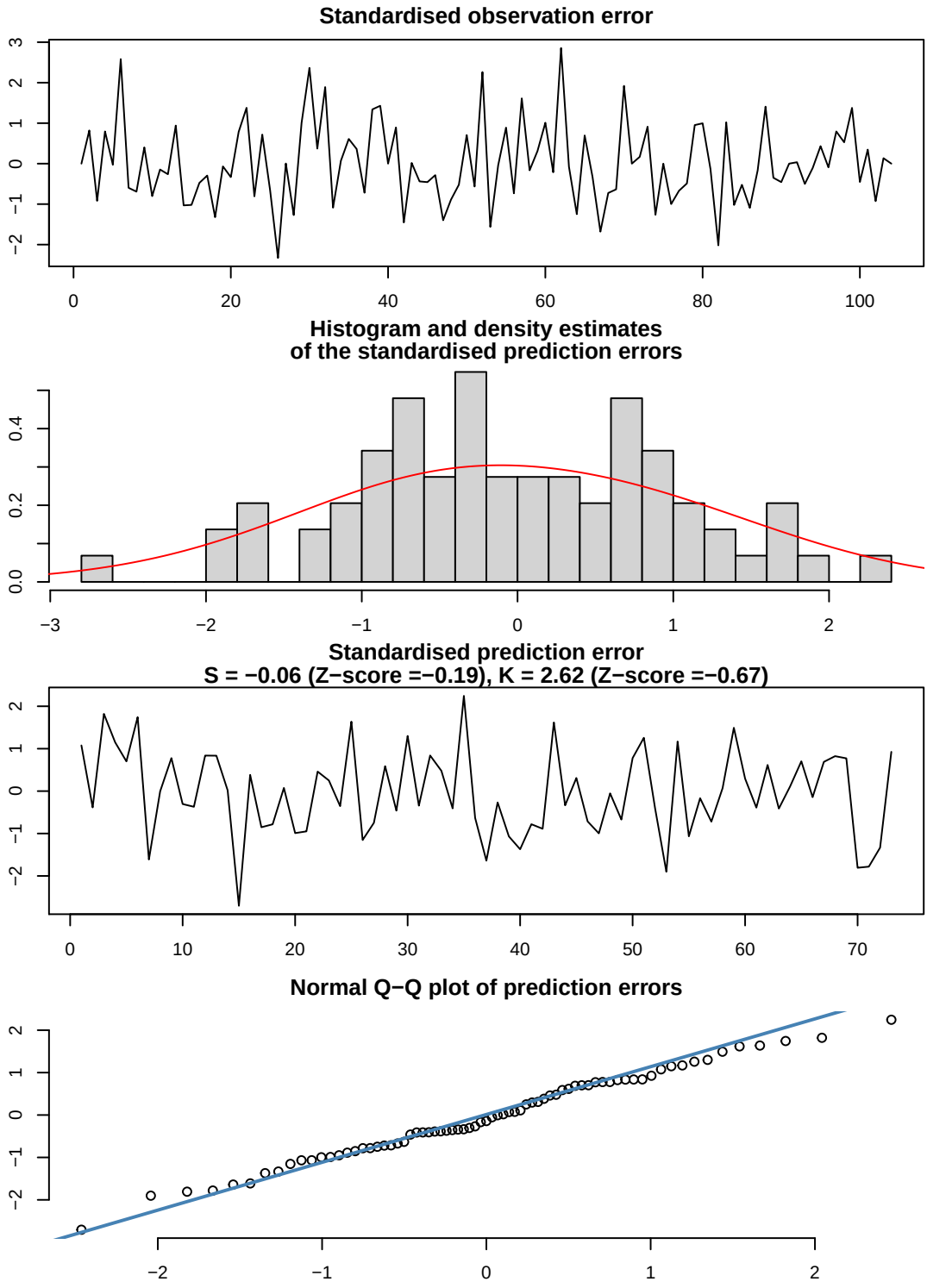

Figure 5: Results from the bivariate model estimated with VAT and MBS data for the transport industry. The VAT data are taken from the first release. **VAT figures at time points 40, 73, 76 and 93 are discarded**. The density of the standardised prediction errors is derived with a kernel density estimation with a Gaussian kernel (density() function in R). Only outputs from the model with which it is not possible to re-construct the original data are shown to respect the confidentiality of the VAT respondents.

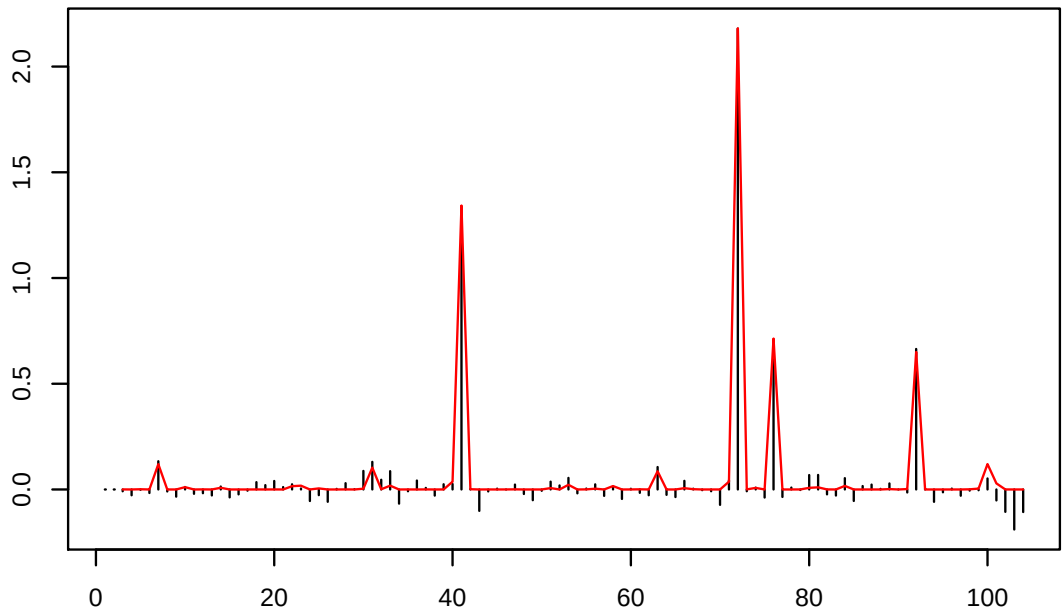

Figure 6: Illustration of the score driven cleaning method using VAT data from the first release (industry 493T495). The vertical bars show the prediction errors in log units derived from the score driven model with an asymmetric student distribution. The red line shows the difference in logs between the original VAT data and the pseudo "clean" observations derived using the score driven smoothing algorithm. Only outputs from the model with which it is not possible to re-construct the original data are shown in order to respect the confidentiality of the VAT respondents.

5 Pseudo real-time exercise

The VAT data are regularly revised over time; i.e. the first estimate of the VAT observation for March 2020 using the first release of the quarterly March VAT observation (released in May) differs from the estimate when estimation is carried in July when the second and third release for May have been published, and quarterly estimates for April and May have become available. To investigate how the model would have performed in real time in May 2020 when forecasting monthly output for March 2020 it is important to use only the data that were available at the time. Proceeding this way it is then possible to analyse the revisions to the nowcasts as more data become available.

To analyse the performance of the model over time, estimation is carried out over a period of 40 months using 40 vintages of the data. Figure 7 illustrates two successive vintages. This is a *pseudo* real-time exercise because all VAT releases are cleaned together using the last vintages. Cleaning the VAT observations at each step of the rolling estimation in addition to estimating the nowcasting model would slow done the analysis excessively.

$$\begin{bmatrix} y_{11,1} & y_{10,1} & y_{9,1} & \cdots & y_{2,1} & y_{1,1} \\ y_{11,2} & y_{10,2} & y_{9,2} & \cdots & y_{2,2} & y_{1,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{11,t} & y_{10,t} & y_{9,t} & \cdots & y_{2,t} & y_{1,t} \\ y_{11,t+1} & y_{10,t+1} & y_{9,t+1} & \cdots & y_{2,t+1} & y_{1,t+1} \\ & y_{10,t+2} & y_{9,t+2} & \cdots & y_{2,t+2} & y_{1,t+2} \\ & & \ddots & \vdots & \vdots & \vdots \\ & & & y_{3,t+10} & y_{2,t+10} & y_{1,t+10} \\ & & & & y_{2,t+11} & y_{1,t+11} \\ & & & & & y_{1,t+12} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_t \\ z_{t+1} \\ z_{t+2} \\ \vdots \\ z_{t+10} \\ z_{t+11} \\ z_{t+12} \end{bmatrix}$$

Figure 7: Illustration of the successive vintages data used for estimation in a pseudo real-time setting. $y_{i,t}$ represents the quarterly VAT data from release i in month t. Columns and first subscripts indicate the releases while rows and second subscripts indicate the month the figures relate to. The maturity of the data increases from left to right and their timeliness increases from top to bottom.z_t represents the monthly covariate (the MBS for large businesses) in month t. In step t of the real-time pseudo exercise only black terms are available for estimation, while in step $t+1$ both black and red terms are used.

At each step of the rolling estimation the vector of seasonally adjusted monthly

output estimates (both smoothed and filtered) in logs $\hat{x}^T = (\hat{x}_1^T, \hat{x}_2^T, ..., \hat{x}_N^T)'$, where T indicates the last month all maturities are observable and $N = T + 10$, are stored. By taking T forward gradually it is possible to simulate the real-time accrual of the data.

The revisions to the seasonally adjusted figures of interest are:

$$\text{revision}_{i,t}^{j} = \hat{x}_t^{11} - \hat{x}_t^{j}, \quad j = 1, ..., 10, \tag{24}$$

where $\hat{x}_t$ is the log of monthly seasonally adjusted output derived from the Kalman filter or smoother.

Two rounds of rolling estimation carried out, one for each data set (i.e. the one obtained when excluding outliers and the one obtained with the score driven cleaning approach). The revisions implied by both methods are analysed and compared with the revisions resulting from averaging nowcasts.

Sources of revision

There are four channels through which the accrual of releases can yield to revisions. First, noise in the aggregate figures should decrease with maturity, i.e. $y_{4,t}$ should be a better estimate of the signal $\tilde{y}_{1,t}$ than $y_{1,t}$ is. More precise aggregate figures should yield better estimates of the underlying seasonally adjusted figures.

Secondly, as revised aggregate figures for month t become available, early figures for periods succeeding t also become available. Notably, when all releases for month t are available, estimates of aggregate figures for periods up to $t + 10$ are also available (as illustrated in figure 7). Assuming that the target is the seasonally adjusted monthly estimate for month t, the data succeeding t will affect the smoothed estimate of the target. This is because the smoothed estimate is the best estimate given past, current and future observations.

Another source of revision arising from using data succeeding the targeted period and affecting the smoothed estimate comes from the temporal aggregation constraints. All monthly figures are related through the temporal aggregation constraints because the observations are overlapping; modifying one quarterly figure affects the entire series of seasonally adjusted estimates. For these reasons large revisions to the smoothed series should not be taken to imply that the early releases are uninformative.

5.1 Aggregating then nowcasting

This section reports the results when cleaned VAT observations are aggregated across all seventy five industries before carrying out the real-time estimation exercise. This approach contrasts with nowcasting output for each industry and then aggregating these nowcasted figures.

The first plot of Figure 8 shows the original eleven releases of VAT observations taken from the most recent vintage. The figures represent all 75 industries and the data have been aggregated using gross value-added weights. Missing observations in the original series have been replaced with the average observation of the series for aggregation purpose. This plot illustrates well the extend and magnitude of the noise affecting the VAT data.

The second plot shows the same observations where extreme outliers have been discarded using t-tests, while the third plot shows the pseudo observations derived with the score driven method. From these two plots it is possible to get a much better picture of the VAT-based quarterly turnover.

The score driven cleaning approach produces smaller revisions and averaging does not help

Table 13 shows the mean absolute revision derived with (24) across each early release. These are revisions to the monthly seasonally adjusted output estimate in logs. The score driven method produces the lowest revisions in absolute values on average. Importantly the revisions' magnitude decreases monotonically as the maturity of the data increases, with a mean absolute revision of 1.3 percentage point for the first release, compared to 0.06 percentage point for the last.

Table 13: Mean absolute revision ($\times 100$) each cleaning approach and across releases. The revisions represent the differences in logs between the early estimates ($i = 1, ..., 11$) and last estimate ($i = 12$) of monthly seasonally adjusted output (equation (24)). Bold figures indicate the lowest mean absolute revision at each maturity.

Release	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$	$i=8$	$i=9$	$i=10$
Average	1.335	1.2	0.829	0.636	0.402	0.333	0.273	0.239	0.165	0.126
Score driven	**1.306**	**1.187**	**0.509**	**0.403**	**0.261**	**0.227**	**0.175**	**0.15**	**0.099**	**0.06**
T-test	1.692	1.532	1.315	1.006	0.626	0.508	0.42	0.37	0.278	0.213

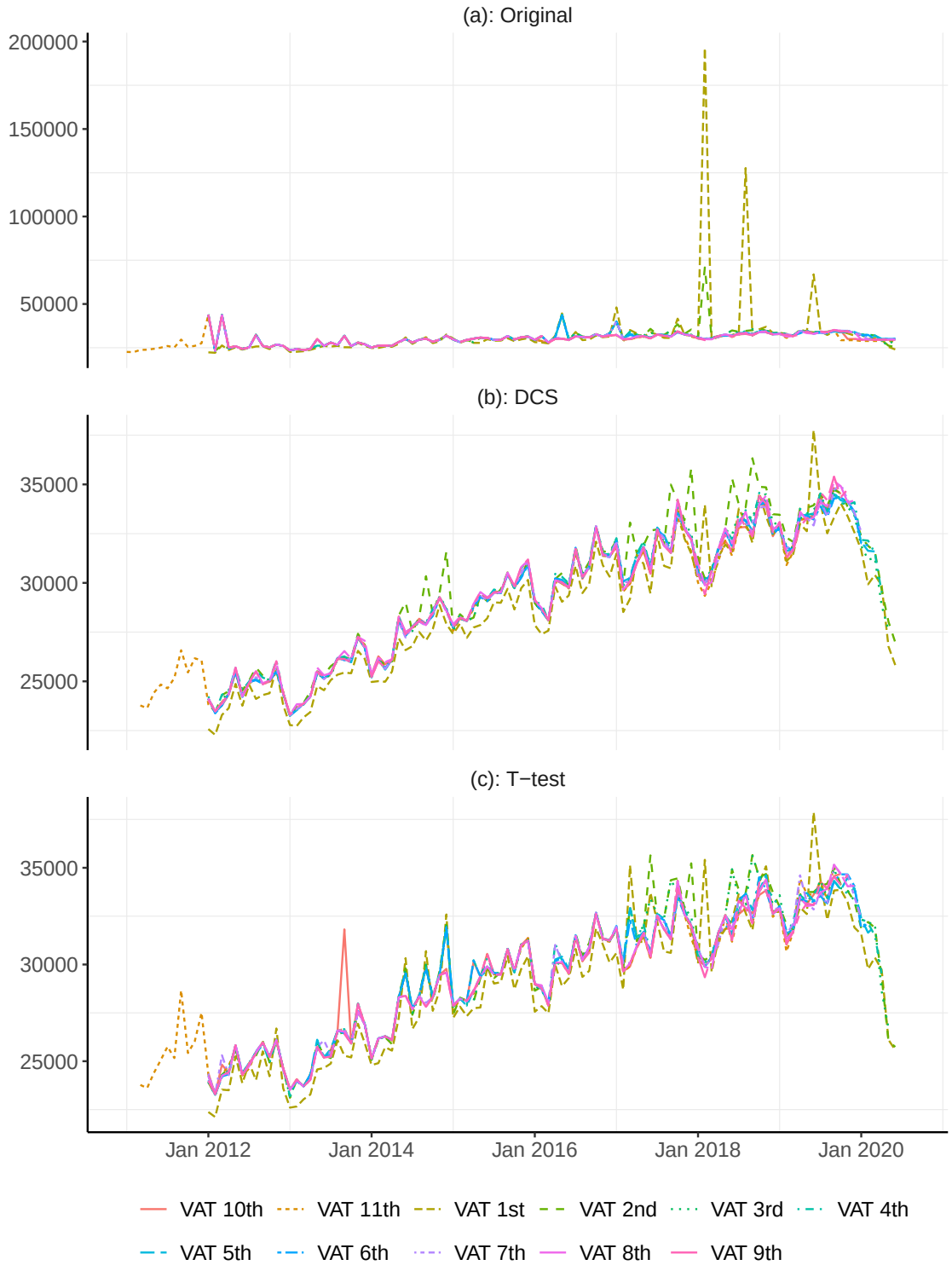

Figure 8: Aggregate non-seasonally adjusted VAT releases (representing approximately a quarter of GVA in the UK). Original VAT observations (most recent data vintage) compared with the observations obtained when using the t-test and score driven cleaning approaches.

Early VAT-based estimates tend to be revised towards

Table 14 shows the mean revision with its standard error and associated t-statistic across each early releases. While the absolute revision is useful to analyse the magnitude of the revisions, the mean revision and its standard error are used to investigate the biases in the early releases. Here again the score driven method performs better than the other two approaches as it tends to produce lower means and standard errors.

Table 14: Mean revision and standard error ($\times 100$) with implied t-statistic for each cleaning approach and across releases. The revisions represent the differences in logs between the early estimates ($i = 1, ..., 11$) and last estimate ($i = 12$) of monthly seasonally adjusted output (equation (24)). Bold figures indicate the lowest mean revision at each maturity.

Cleaning	Stat.	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	$j=8$	$j=9$	$j=10$
Average	Mean	-0.931	-0.906	-0.532	-0.336	-0.198	-0.155	-0.12	-0.115	-0.084	-0.067
Average	S.E.	0.274	0.214	0.16	0.12	0.067	0.059	0.053	0.046	0.036	0.032
Average	t-stat	-3.397	-4.227	-3.327	-2.795	-2.982	-2.616	-2.249	-2.522	-2.319	-2.082
Score driven	Mean	-0.977	**-0.79**	**-0.286**	**-0.175**	**-0.123**	**-0.075**	**-0.046**	**-0.033**	**-0.029**	**-0.021**
Score driven	S.E.	0.283	0.229	0.085	0.07	0.049	0.045	0.038	0.034	0.02	0.014
Score driven	t-stat	-3.454	-3.454	-3.359	-2.503	-2.484	-1.659	-1.215	-0.96	-1.42	-1.483
T-test	Mean	**-0.875**	-1.015	-0.773	-0.495	-0.275	-0.234	-0.193	-0.197	-0.138	-0.113
T-test	S.E.	0.336	0.294	0.272	0.202	0.113	0.101	0.09	0.081	0.071	0.063
T-test	t-stat	-2.609	-3.455	-2.846	-2.455	-2.443	-2.314	-2.137	-2.428	-1.949	-1.802

But unfortunately the first five releases exhibit a significant negative bias. This is not necessarily a bad feature because it can arise from relatively small standard errors, indicating a low volatility in the mean revisions which is desirable since these are small. However, this also implies that the monthly output estimates tend to be revised downwards as more data become available.

Figure 9 helps to shade light on the nature of this negative bias; it shows the first release of monthly output alongside the last release and the implied revision. While both estimates tend to be close, there are few discrepancies where the first release shows a large increase in output which does not materialise in the last release. This is probably due to the cleaning approach which fails to capture large measurement errors in these few months. The negative bias is likely to be driven by these few large discrepancies. These large negative revisions can also be seen in figure 10 which

shows output estimates from each releases across the entire real-time estimation period.

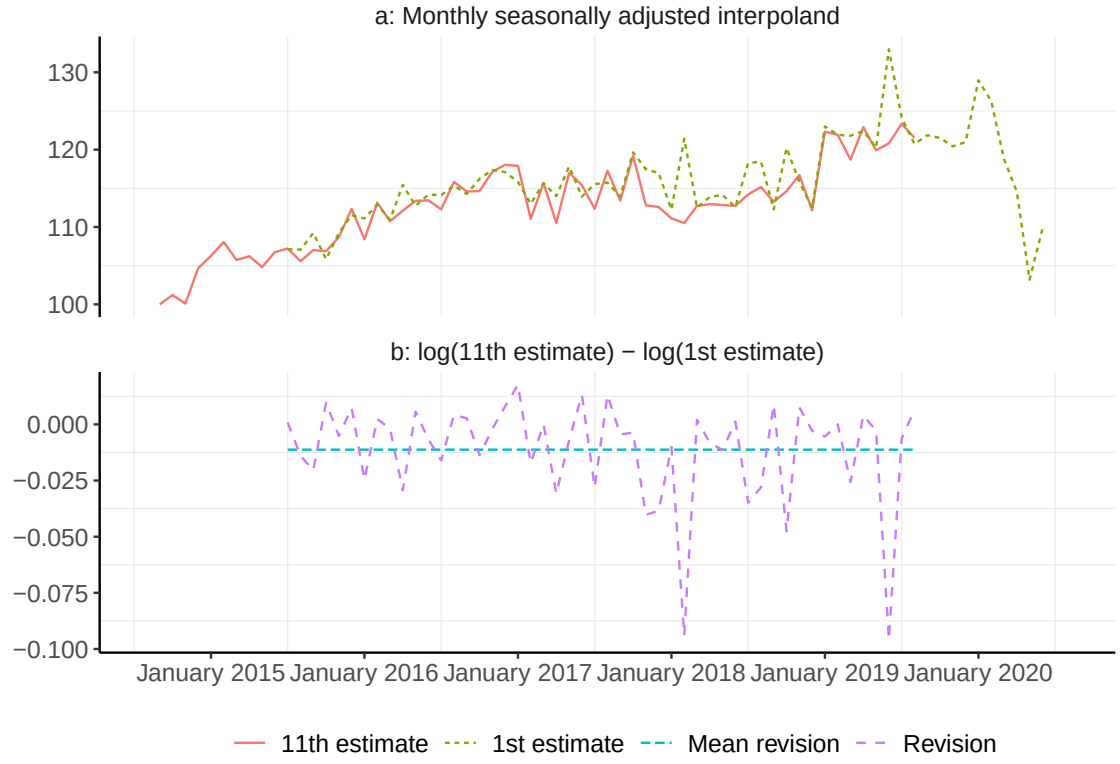

Figure 9: Seasonally adjusted figures representing small and medium size businesses in 75 industries (a quarter of GVA in the UK). First and last (eleventh) VAT estimates. Estimation using pseudo real-time data. The VAT-based figures are derived using the score driven method for cleaning. Index July 2014=100.

5.1.1 The VAT data can provide a timely indication of economic slumps

Although the VAT-based monthly output estimates tend to be revised towards, they can still provide a timely picture of economic activity. Figure 11 shows the first releases of the VAT-based monthly output estimates alongside the estimate arising from the MBS. The VAT-based estimate indicate a clear important reduction of economic activity starting from March 2020, albeit on of a smaller magnitude than the picture given by the MBS.

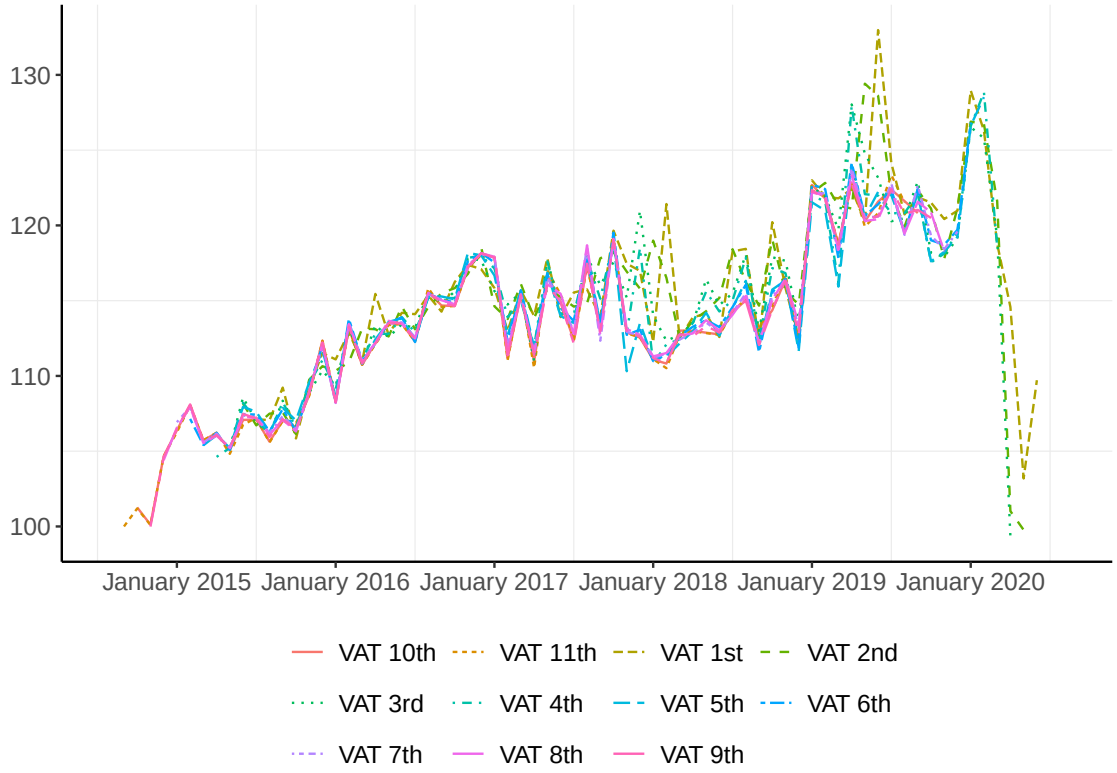

Figure 10: VAT-based nowcasted monthly output. Seasonally adjusted figures representing small and medium size businesses in 75 industries (a quarter of GVA in the UK). First VAT estimate with the next ten revisions. Estimation using pseudo real-time data. The VAT-based figures are derived using the score driven method for cleaning. Index July 2014=100.

5.1.2 Overtime the VAT and MBS-based monthly output signals coincide but there remains some differences in their levels

Figure 12 shows the monthly output figures derived using the latest available vintage of the VAT data, that is in the last step of the real-time estimation exercise. Here both cleaning strategies are illustrated. These series are compared with the monthly output estimate derived with the MBS data. The first plot shows the levels while the second plot shows log differences.

Both VAT-based series show similar trends, but these diverge locally with the trend underlying the MBS figures. This is consistent with the result obtained in Labonne and Weale (2020). Separately, the score driven approach produces monthly changes closer to the MBS monthly changes, with a correlation coefficient of 0.87 when using the score driven approach compared to 0.76 when using the t-test cleaning. Hence, in addition to producing lower revisions, the score driven

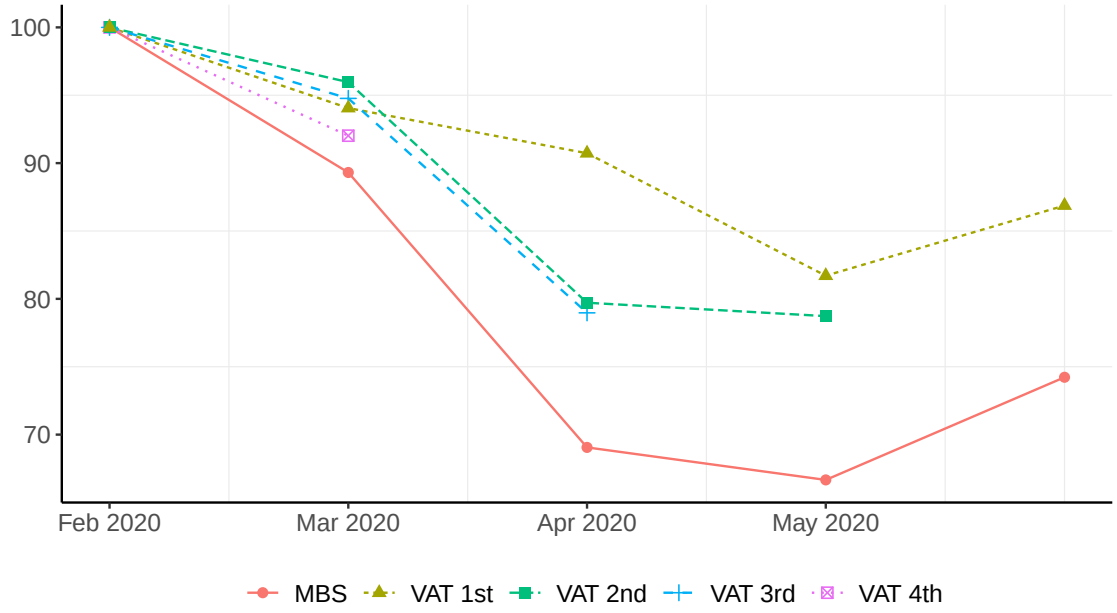

Figure 11: Real-time nowcasts of monthly output. Seasonally adjusted figures representing small and medium size businesses in 75 industries (a quarter of GVA in the UK). First VAT estimate and the four subsequent revisions. The VAT-based figures are derived using the score driven method for cleaning. Index February 2020=100.

approach yields a picture of economic activity closer to the MBS that the t-test cleaning method.

The covariate helps in estimating VAT-based monthly changes

Table 15 shows the estimation results when estimating the nowcasting model (14) with VAT observations cleaned using the score driven method, since the previous section shows that it is the cleaning method which yields the lowest revisions and therefore is preferable.

The first row shows variance and correlation estimates of the local linear trend model, which estimated smoothed components are plotted in figure 13. The VAT and MBS disturbances in the local linear trend model governing the monthly seasonal adjusted output estimates are strongly correlated. This highlights the importance of the MBS as covariate for improving VAT-based monthly output.

The bias in the first release changes rapidly

The second row of table 15 shows the variance estimates of the releases biases' disturbances, while the release biases are plotted in figure 14. These show that

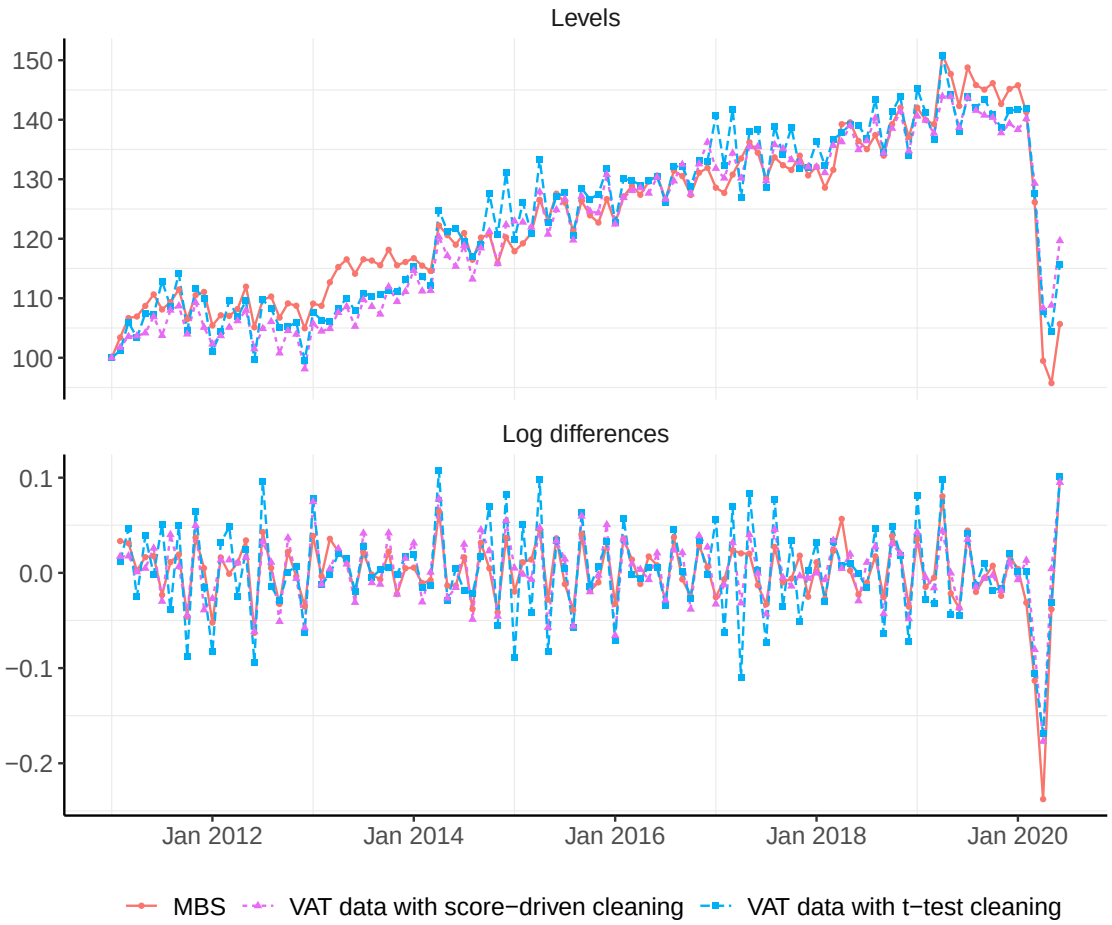

Figure 12: VAT-based and MBS-based nowcasted monthly output. Seasonally adjusted figures representing small and medium size businesses in 75 industries (a quarter of GVA in the UK). The VAT data are taken from the most recent vintage. Index January 2011=100. Corr log differences MBS and VAT when score driven cleaning = 0.87; Corr log differences MBS and VAT when t-test cleaning = 0.76.

the bias in the first release changes significantly over time, a phenomenon less pronounced in more mature releases.

The first two releases are significantly more noisy than later ones

The third row shows the estimated variance of the releases' observation errors. The first and second releases are much more noisy than later releases. Separately, estimated noise remains relatively strong in the last release, which shows that even mature figures are subject to non-negligible measurement errors.

Table 15: Estimation results from the nowcasting model using the latest vintage of VAT data cleaned with the score driven approach. The first line shows the estimated variance and correlation parameters of the bivariate local linear trend model at the core of the nowcasting model. The second line shows the estimated variance parameters of the disturbances in the release biases. The third line shows the estimated variance parameters of the release observation errors. Variance parameters are reported $\times 1e4$.

$\sigma_{\mu,1}$	$\sigma_{\nu,1}$	$\sigma_{e,1}$	$\sigma_{\mu,2}$	$\sigma_{\gamma,2}$	$\sigma_{\nu,2}$	$\sigma_{e,2}$	ρ_{ν}	ρ_{μ}	ρ_{e}	
0.620	0.090	7.990	3.970	0	0.110	5.220	1	1	0.930	
$\sigma_{c,1}$	$\sigma_{c,2}$	$\sigma_{c,3}$	$\sigma_{c,4}$	$\sigma_{c,5}$	$\sigma_{c,6}$	$\sigma_{c,7}$	$\sigma_{c,8}$	$\sigma_{c,9}$	$\sigma_{c,10}$	
0.130	0.030	0.010	0.010	0.010	0	0	0	0.010	0.060	
$\sigma_{\epsilon,1}$	$\sigma_{\epsilon,2}$	$\sigma_{\epsilon,3}$	$\sigma_{\epsilon,4}$	$\sigma_{\epsilon,5}$	$\sigma_{\epsilon,6}$	$\sigma_{\epsilon,7}$	$\sigma_{\epsilon,8}$	$\sigma_{\epsilon,9}$	$\sigma_{\epsilon,10}$	$\sigma_{\epsilon,11}$
5.620	5.860	0.170	0.270	0.040	0.080	0.210	0.370	0.380	0.090	0.210

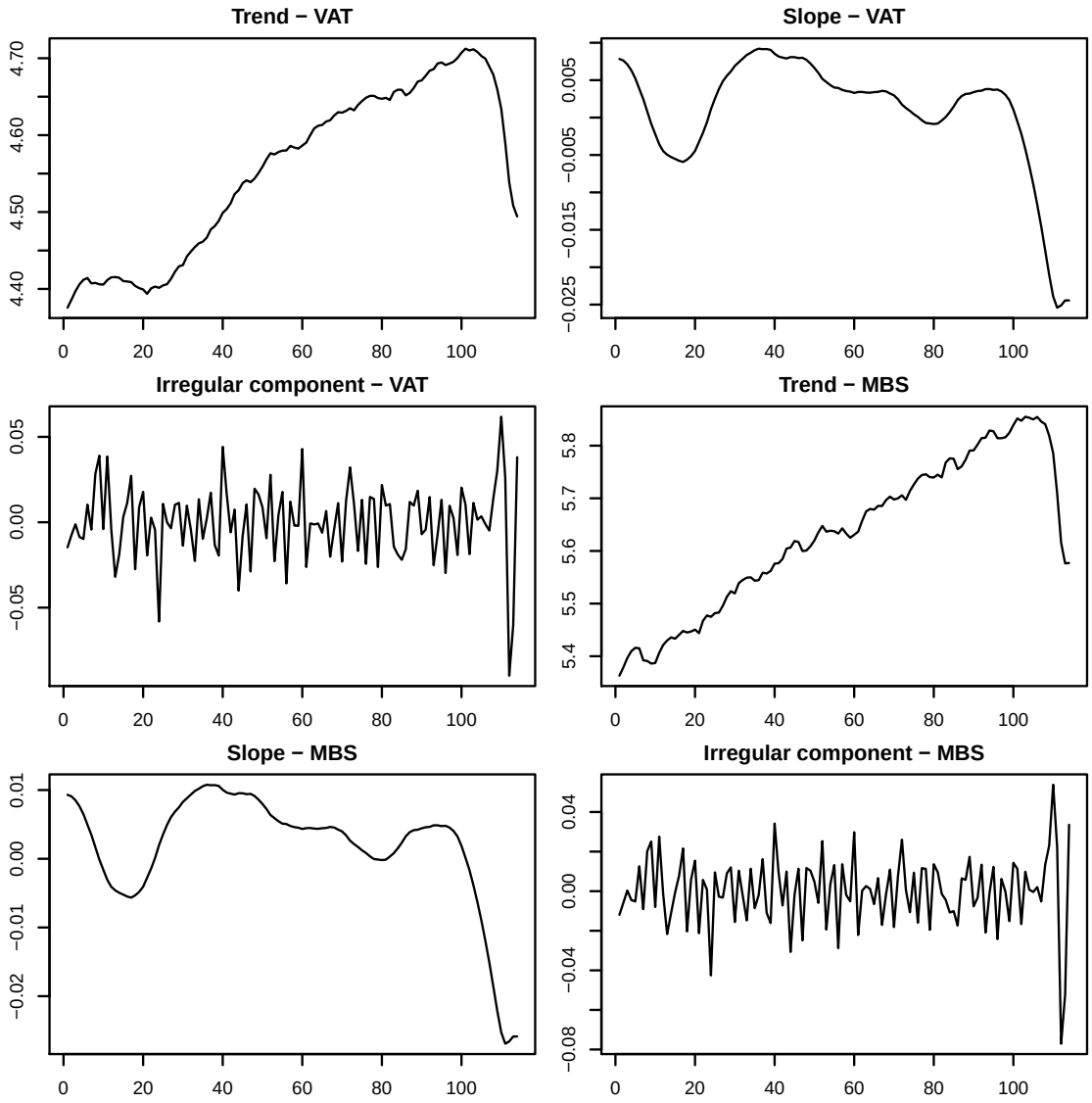

Figure 13: Smoothed estimates, in log units, of the bivariate local linear trend model at the core of the nowcasting model. Estimation with the most recent data vintages which have been cleaned using the score driven method and aggregated across all 75 industries using gross value-added weights. Monthly seasonally adjusted figures are derived by adding up the trend with the irregular component.

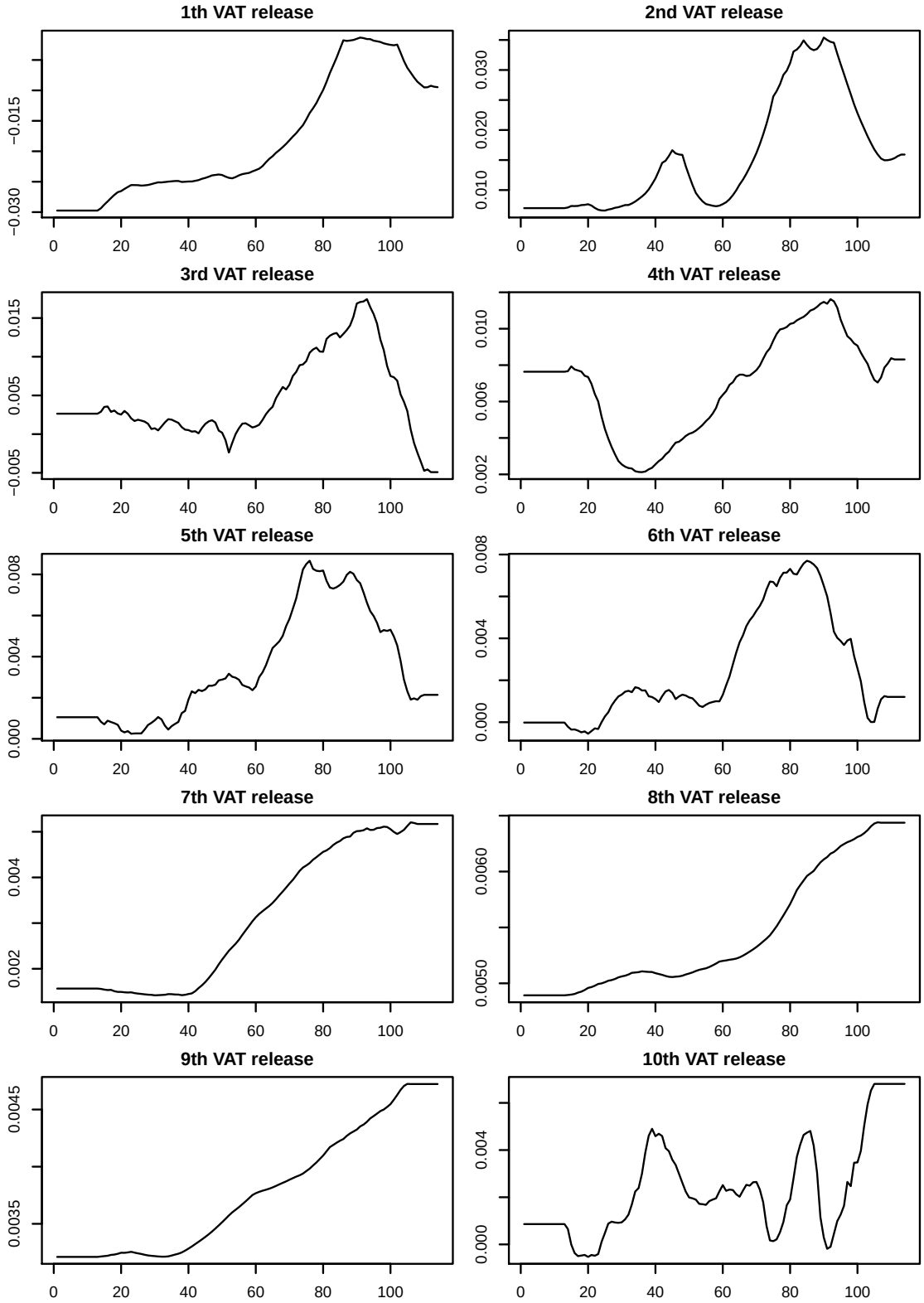

Figure 14: Bias of the first ten VAT releases with respect to the eleventh (and last) release. Log units. Estimation with the most recent data vintages which have been cleaned using the score driven method and aggregated across all 75 industries using gross value-added weights.

6 Conclusion

This paper has presented and illustrated a flexible nowcasting approach for data subject to large measurement errors and revisions. It has shown that, in the presence of extremely large and asymmetric measurement errors, using a score driven approach works better than discarding outlying observations. This finding is illustrated through the production of a monthly output series from noisy VAT-based quarter figures which can be used as a timely indicator of the economic recession resulting from the coronavirus pandemic.

References

Azzalini, A. and A. Capitanio (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(2), 367–389.

Bell, W. R. and S. C. Hillmer (1983). Modeling time series with calendar variation. *Journal of the American Statistical Association 78*(383), 526–534.

Bowman, K. O. and L. R. Shenton (1975). Omnibus test contours for departures from normality based on vb 1 and b 2. *Biometrika 62*(2), 243.

Buccheri, G., G. Bormetti, F. Corsi, and F. Lillo (2021). Filtering and smoothing with score-driven models.

Caivano, M., A. Harvey, and A. Luati (2016). Robust time series models with trend and seasonal components. *SERIEs 7*(1), 99–120.

Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*(5), 777–795.

Delle-Monache, D. and I. Petrella (2017). Adaptive models and heavy tails with an application to inflation forecasting. *International Journal of Forecasting 33*(2), 482–501.

Durbin, J. and S. J. Koopman (2012). *Time series analysis by state space methods*. Oxford University Press.

Fernández, C. and M. F. Steel (1996). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association 93*(441), 359–371.

Gatz, D. F. and L. Smith (1995). The standard error of a weighted mean concentration i. bootstrapping vs other methods. *Atmospheric Environment 29*(11), 1185–1193.

Gomez, H., F. Torres, and H. Bolfarine (2007). Large-sample inference for the epsilon-skew-t distribution. *Communications in Statistics - Theory and Methods 36*(1), 73–81.

Harvey, A. and A. Luati (2014). Filtering with heavy tails. *Journal of the American Statistical Association 109*(507), 1112–1122.

Harvey, A. C. (2006). Seasonality and unobserved components models: an overview. *Conference on seasonality, seasonal adjustment and their implications for short-term analysis and forecasting*.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series (Econometric Society Monographs)*. Cambridge University Press.

Labonne, P. and M. Weale (2020). Temporal disaggregation of overlapping noisy quarterly data: estimation of monthly output from uk value-added tax data. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 183*(3), 1211–1230.

Proietti, T. (2000). Comparing seasonal components for structural time series models. *International Journal of Forecasting 16*(2), 247–260.

Russ, B. and A. Tariq (2017). The impact of moving holidays on official statistics time series. *Proceedings of the 22nd GSS Methodology Symposium 2017*.

Zhu, D. and J. W. Galbraith (2010). A generalized asymmetric student- distribution with application to financial econometrics. *Journal of Econometrics 157*(2), 297–305.

Appendix A Multivariate model for nowcasting

The state space representation (**??**) is

$$\begin{aligned} y_t &= Z_t\alpha_t, \\ \alpha_{t+1} &= T\alpha_t + R\eta_t, \qquad \eta_t \sim \mathrm{N}(0, Q), \\ \alpha_1 &\sim \mathrm{N}(a_1, P_1). \end{aligned}$$

$0_{m\times n}$ as a matrix of zeros with m rows and n columns. The observation matrix is $Z_t = (Z_{0,t}, Z_n, Z_\chi, Z_c)$, where
$Z_{0,t} = (Z_{1,t}, Z_{1,t}, Z_{1,t}, Z_{1,t}, Z_{1,t}, Z_{1,t}, Z_{1,t}, Z_{1,t}, Z_{1,t}, Z_{1,t}, Z_{1,t}, Z_{2,t})'$;
$Z_{1,t} = (1/3, 1/3, 1/3, 0, 1, 0_{8\times1}, 0, 1/3, 1/3, 1/3, h^a_{1,t}, 1^1_{b,t}, 1^2_{b,t}, 1^3_{b,t}, 1^1_{\chi,t}, 1^2_{\chi,t}, 1^3_{\chi,t}, 0_{14\times1})$,
$Z_{2,t} = (0_{24\times1}, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, h^a_{2,t})$. The variables $1^j_{b,t}$, $j = 1, 2, 3$, are stagger dummy variables relating to the stagger biases, while $1^j_{\chi,t}$, $j = 1, 2, 3$, are stagger dummy variables relating to the persistent measurement errors linked to the eleventh release (which therefore affect all releases).

The state vector is $\alpha_t = (\alpha'_{1,t}, \alpha'_{2,t})'$, where
$\alpha_{1,t} = (\mu_{1,t}, \mu_{1,t-1}, \mu_{1,t-2}, \nu_{1,t}, \gamma_{1,t}, ..., \gamma_{1,t-8}, e_{1,t}, e_{1,t-1}, e_{1,t-2}, \beta_1, b^1_{1,t}, b^2_{1,t}, b^3_{1,t}, \chi^1_{1,t}, \chi^2_{1,t}, \chi^3_{1,t})'$,
$\alpha_{2,t} = (\mu_{2,t}, \nu_{2,t}, \gamma_{2,1,t}, \gamma^*_{2,1,t}, \gamma_{2,2,t}, \gamma^*_{2,2,t}, ..., , \gamma_{2,5,t}, \gamma^*_{2,5,t}, \gamma_{2,6,t}, e_{2,t}, \beta_2)'$.

$$Z_n = \begin{pmatrix} 1&1&1&1&1&1&1&1&1&1&1\\ 0&1&1&1&1&1&1&1&1&1&1\\ 0&0&1&1&1&1&1&1&1&1&1\\ 0&0&0&1&1&1&1&1&1&1&1\\ 0&0&0&0&1&1&1&1&1&1&1\\ 0&0&0&0&0&1&1&1&1&1&1\\ 0&0&0&0&0&0&1&1&1&1&1\\ 0&0&0&0&0&0&0&1&1&1&1\\ 0&0&0&0&0&0&0&0&1&1&1\\ 0&0&0&0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0 \end{pmatrix}; \; Z_\epsilon = \begin{pmatrix} 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0&0&0&0\\ 0&0&0&0&0&1&0&0&0&0&0&0\\ 0&0&0&0&0&0&1&0&0&0&0&0\\ 0&0&0&0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&0&0&1&0&0&0\\ 0&0&0&0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&0&0&0&0&0&0 \end{pmatrix}; \; Z_c = \begin{pmatrix} 1&0&0&0&0&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0&0&0&0\\ 0&0&0&0&0&1&0&0&0&0&0&0\\ 0&0&0&0&0&0&1&0&0&0&0&0\\ 0&0&0&0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&0&0&1&0&0&0\\ 0&0&0&0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&0&0&0&0&0 \end{pmatrix}.$$

The transition matrix is $T = \mathrm{diag}(T_{1,LLT}, T_{1,\gamma}, T_{1,e}, T_{1,\beta}, T_{1,b}, T_{1,\epsilon}, T_{2,LLT}, T_{2,\gamma}, T_{2,e}, T_{2\beta})$,

where $T_{1,LLT} = \begin{pmatrix} 1&0&0&1\\ 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0 \end{pmatrix}$; $T_{1,\gamma} = \begin{pmatrix} 0&0&-1&0&0&-1&0&0&-1\\ 1&0&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&1&0&0&0\\ 0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1&0 \end{pmatrix}$;

$T_{1,e} = \begin{pmatrix} 0&0&0\\ 1&0&0\\ 0&1&0 \end{pmatrix}$; $T_{1,\beta} = 1$, $T_{1,b} = I_2$; $T_{1,\epsilon} = 0_{3\times3}$; $T_{2,LLT} = \begin{pmatrix} 1&1\\ 0&1 \end{pmatrix}$; $T_{2,e} = 0$

$T_{2,\gamma} = \mathrm{diag}(T_{2,1,\gamma}, T_{2,2,\gamma}, T_{2,3,\gamma}, T_{2,4,\gamma}, T_{2,5,\gamma}, -1)$; $T_{2,j,\gamma} = \begin{pmatrix} \cos(2\pi j/12) & \sin(2\pi j/12)\\ -\sin(2\pi j/12) & \cos(2\pi j/12) \end{pmatrix}$;

$T_{2,\beta} = 1$. $R = \text{diag}(R_{1,\mu}, R_{1,\nu}, R_{1,e}, R_{1,b}, R_{1,\epsilon}, R_{2,\mu}, R_{2,\nu}, R_{2,\gamma}, R_{2,e})$, $T_\epsilon = 0_{10\times 10}$ $T_\chi = 0_{10\times 10}$, $T_c = 1_{10\times 10}$, where $R_{1,\mu} = R_{1,e} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}'$; $R_{1,\nu} = 1$; $R_{1,b} = I_2$;$R_{1,\epsilon} = I_3$ $R_{2,\mu} = R_{2,\nu} = R_{2,e} = 1$; $R_{2,\gamma} = I_{11}$.

Q is a 23×23 matrix with diagonal $(\sigma^2_{1,\xi}, \sigma^2_{1,\zeta}, \sigma^2_{1,e}, \sigma^2_{1,\kappa 2}, \sigma^2_{1,\kappa 3}, \sigma^2_{1,\epsilon 1}, \sigma^2_{1,\epsilon 2}, \sigma^2_{1,\epsilon 3}$, $\sigma^2_{2,\xi}, \sigma^2_{2,\zeta}, \sigma^2_{2,\omega}, ..., \sigma^2_{2,\omega}, \sigma^2_{2,\omega}/2, \sigma^2_{2,e})$, and $Q_{[1,5]} = \rho_\xi \sigma_{1,\xi}\sigma_{2,\xi}$, $Q_{[2,6]} = \rho_\zeta \sigma_{1,\zeta}\sigma_{2,\zeta}$, $Q_{[3,7]} = \rho_e \sigma_{1,e}\sigma_{2,e}$.

For estimation, Q is expressed using the Cholesky factorisation as $Q = LL'$ and minimise the negative log likelihood function with respect to the parameters in the lower triangular matrix L. Thus, the estimated variance-covariance matrix Q is positive semi-definite.