ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH Office for National Statistics

Tracking Weekly Activity using New Data Sources

Andrea De Polis, Ana Beatriz Galvão and Ivan Petrella

ESCoE Discussion Paper 2025-19

November 2025

ISSN 2515-4664

DISCUSSION PAPER

Tracking Weekly Activity using New Data Sources
Andrea De Polis, Ana Beatriz Galvão and Ivan Petrella
ESCoE Discussion Paper No. 2025-19
November 2025

Abstract

Policymakers increasingly rely on real-time measures of economic activity to inform decisions, yet official statistics are typically available only with substantial delay. This paper develops a methodology to extract information from high-frequency indicators in order to produce weekly estimates of official monthly statistics in real time. Building on real-time tracking and nowcasting models, our approach addresses two challenges inherent in alternative indicators: the absence of seasonal adjustment and the prevalence of outliers. We incorporate seasonal components directly into the model, allowing the seasonal structure of low-frequency data to inform high frequency proxies, and employ fat-tailed distributions to mitigate the influence of large, infrequent shocks. Applying our methodology to UK data, we track retail sales, monthly GDP, and vacancies using proxies such as debit-card spending (Revolut) and online job advertisements. Our results show that weekly estimates improve real-time prediction of official releases and highlight the usefulness of high-frequency alternative data.

Keywords: real-time tracker, mixed-frequency models, seasonality, state-space models, high-frequency data

JEL classification: C32, C53, E01, E37

Ana Beatriz Galvão, Bloomberg Economics, CEPR, Warwick and ESCoE
ana.b.galvao@proton.me

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

Tracking Weekly Activity using New Data Sources *

Andrea De Polis† Ana Beatriz Galvão‡ Ivan Petrella§

November 2025

Abstract

Policymakers increasingly rely on real-time measures of economic activity to inform decisions, yet official statistics are typically available only with substantial delay. This paper develops a methodology to extract information from high-frequency indicators in order to produce weekly estimates of official monthly statistics in real time. Building on real-time tracking and nowcasting models, our approach addresses two challenges inherent in alternative indicators: the absence of seasonal adjustment and the prevalence of outliers. We incorporate seasonal components directly into the model, allowing the seasonal structure of low-frequency data to inform high-frequency proxies, and employ fat-tailed distributions to mitigate the influence of large, infrequent shocks. Applying our methodology to UK data, we track retail sales, monthly GDP, and vacancies using proxies such as debit-card spending (Revolut) and online job advertisements. Our results show that weekly estimates improve real-time prediction of official releases and highlight the usefulness of high-frequency alternative data.

Keywords: real-time tracker, mixed-frequency models, seasonality, state-space models, high-frequency data.

*This research has been funded by the Office for National Statistics as part of the research programme of the Economic Statistics Centre of Excellence. Andrea de Polis thanks the Economics Departament of the University of Strathclyde for their hospitality and support during the development of this research. We thank Andrew Walton for his comments, suggestions and support on our earlier methods and results, and Rhiannon Yapp for her advice on the ONS real-time economic indicators.

†University of Strathclyde and ESCoE. andrea.de-polis@strath.ac.uk

‡Bloomberg Economics, CEPR, Warwick and ESCoE. ana.b.galvao@proton.me

§Collegio Carlo Alberto, University of Turin, CEPR and ESCoE. ivan.petrella@carloalberto.org

1 Introduction

Policymakers increasingly rely on real-time trackers of economic activity to fine-tune policy, as these provide timely evidence of both policy effects and the shocks they are designed to mitigate. The Aruoba-Diebold-Scott Business Conditions Index (Aruoba et al. (2009), published weekly by the Philadelphia Fed since 2008, is one of the earliest examples. More recently, the NY/Dallas Fed has introduced a Weekly Economic Index Lewis et al. (2021), which uses alternative data sources that become available more quickly than official releases from statistical agencies. Similarly, the Chicago Fed's Advanced Retail Trade Summary Brave et al. (2021) incorporates alternative data—particularly payments data—to provide weekly information on retail trade.

There are typically two sources of data used to track real-time economic activity. The first consists of official monthly statistics, such as measures of industrial production, retail sales, and unemployment. In the UK, these are published by the Office for National Statistics (ONS). While such data are informative about the state of the economy, they are released only once per month and with a delay of between 20 and 40 days after the reference period. A second source comprises higher-frequency economic indicators that fall outside the scope of statistical office surveys, as they are directly obtained from digital recording of economic-related activity by consumers and businesses. The ONS publishes a set of real-time indicators, including number of UK flights and debit card transactions, labelled as "official statistics in development."

In this paper, we develop a methodology to extract information from high-frequency real-time indicators in order to track official statistics at a weekly frequency. Our approach builds on the methodology of Aruoba et al. (2009) and Brave et al. (2021) for real-time economic trackers. While our model incorporates elements of nowcasting frameworks such as Bańbura et al. (2013) and Antolin-Diaz et al. (2024), our primary aim is to generate earlier estimates of official statistics using high-frequency proxies. For example, we use debit-card spending data to track nominal retail sales. A by-product of our framework is that it allows us to assess the usefulness of real-time indicators for producing faster estimates of official, lower-frequency series.

Our methodology addresses two key features of alternative real-time indicators. First, they are typically seasonal and generally not published in seasonally adjusted form. Whereas Brave et al. (2021) apply seasonal adjustment prior to model estimation, we incorporate seasonal components directly within the model. This is particularly valuable when only a limited time series is available, since identifying seasonal patterns—especially annual seasonality—can be difficult. We propose a method that models seasonality jointly in both low- and high-frequency data, allowing the seasonal structure of the former to inform that of the latter.

A second challenge with high-frequency indicators, such as payments platform data (as in Brave et al. (2021)), is the presence of outliers. Instead of discarding such observations, we follow Antolin-Diaz et al. (2024) and employ a fat-tailed distribution, which reduces the impact of infrequent but large shocks on model parameters.

We apply our methodology to track UK retail sales, monthly GDP, and vacancies. The real-time indicators, or proxies, we use include debit-card spending using Revolut cards and the number of online job adverts. Our approach delivers seasonally adjusted estimates of the three official monthly statistics at a weekly frequency. A real-time exercise shows that our weekly estimates outperform predictions based solely on past monthly data. We also provide strong evidence that debit-card spending data contain useful information for tracking both retail sales and monthly GDP.

This paper is organized as follows. The next section describes our proposed methodology that describes an official monthly statistic into components as level, cycle, seasonal and irregular. Section 2 also provides details of our Gibbs sampling algorithm employed to estimate the model parameters and filter the unobserved components. Section 3 describes the data we employed in the empirical exercises in Section 4. Section 4 provides an analysis of the full-sample estimates of the unobserved states. An evaluation of the performance of our proposed modeling approach in real-time is discussed in Section 5.

2 Mapping Weekly Data to Lower Frequency Data: An Empirical Framework

Consider a simplified setting where an official statistic, y_t^O, is published at a lower frequency, while alternative indicators/proxies of the same variable are available at higher frequency, y_t^W. In this setting one can use the higher frequency indicator, y_t^W, to interpolate the official statistic, y_t^O, and to provide an early signal for this variable in real-time.

2.1 Model to Obtain the Seasonal Adjusted High-Frequency Signal from y_t^O

Let y_t^* denote the unobserved weekly estimate of the official statistic, which is linked to the observed official statistic as follows

$$y_t^O \approx \sum_{j=0}^{k-1} w_j y_{t-j}^* \text{ for } t = k, 2k, 3k, ..., Tk \tag{1}$$

where $\{w_j\}_{j=0}^{k-1}$ are a set of predetermined aggregation weights and T is the number of low frequency observations available at each k_{th} observations of the high frequency y_t^*. Examples of aggregations are provided in, e.g., Mariano and Murasawa (2003) and Bańbura et al. (2013). As we consider official statistics that are likely to have a trend, such as retail sales and monthly GDP, we set aggregation weights such as $w_0 = 1, w_j = 0$ for $\forall j > 0$ in line with aggregation based on end-of-period values.

The unobserved high frequency official statistic, y_t^* has two components, a level, L_t

and a seasonal, s_t^A, component:

$$y_t^* = L_t + s_t^A \text{ for } t = 1, 2, ..., Tk \tag{2}$$

The level component has a stochastic trend such that the cycle is modeled using its first difference as:

$$\Delta L_t = \phi_0 + \phi_1 \Delta L_{t-1} + \phi_2 \overline{\Delta L_{t-1}^4} + u_t. \tag{3}$$

In addition to the first-order autoregressive component, the cycle also depends on the lagged average over four periods, equivalent to the monthly average if the higher frequency is weekly and the lower is monthly. This is a parsimonious method to accommodate long memory first proposed by Corsi (2009).

The innovations follow a conditional t-student distribution with scale varying over time based on a random walk process:

$$u_t = \sigma_{u,t}\eta_t, \text{ where } \eta_t \sim t_\nu(0,1) \tag{4}$$

$$log\sigma_{u,t} = log\sigma_{u,t-1} + \omega_t, \text{ where } \omega_t \sim N(0, \sigma_\omega) \tag{5}$$

where ν is the number of degrees of freedom of the t-distribution. The random walk process for the volatility captures slow-moving changes in the variance of the shocks to the cycle component.

The seasonal component captures regular variation within a year as:

$$s_t^{(A)} = \sum_{j=1}^{\lfloor P_A/2 \rfloor} \theta_{j,t}^{(A)},$$

$$\begin{bmatrix} \theta_{j,t}^{(A)} \\ \theta_{j,t}^{(A)*} \end{bmatrix} = \begin{bmatrix} cos\lambda_j^{(A)} & sin\lambda_j^{(A)} \\ -sin\lambda_j^{(A)} & cos\lambda_j^{(A)} \end{bmatrix} \begin{bmatrix} \theta_{j,t-1}^{(A)} \\ \theta_{j,t-1}^{(A)*} \end{bmatrix} + \begin{bmatrix} \xi_{j,t}^{(A)} \\ \xi_{j,t}^{(A)*} \end{bmatrix},$$

$$\lambda_j^{(A)} = \frac{2\pi j}{P_A} \tag{6}$$

$$\xi_{j,t}^{(A)} \sim N(0, \sigma_{\xi_A}); \xi_{j,t}^{(A)*} \sim N(0, \sigma_{\xi_A^*}), \tag{7}$$

where we set $P_A = 6$.

The inclusion of seasonal components implies we do not need to apply a seasonal filter to the data before estimation, as in Brave et al. (2021). A by-product of our modelling approach is a seasonally adjusted series for the official statistic using the level L_t.

2.2 Adding a High Frequency Indicator

We aim to improve the estimate of y_t^* by including a set of n observed high frequency indicators y_t^W. The high frequency indicators are linked to the level and seasonality components of y_t^* as follows:

$$y_t^W = \gamma_0 + \gamma_f L_t + \gamma_s s_t^A + s_t^M + e_t \tag{8}$$

where s_t^M is the within-month seasonal component and e_t is an $n \times 1$ vector of irregular (purely transitory/noise) components. The intercept and coefficients on the level and annual seasonality are $n \times 1$ vectors. The component s_t^M picks up the seasonal variation within the month as:

$$s_t^{(M)} = \sum_{j=1}^{\lfloor P_M/2 \rfloor} \theta_{j,t}^{(M)},$$

$$\begin{bmatrix} \theta_{j,t}^{(M)} \\ \theta_{j,t}^{(M)*} \end{bmatrix} = \begin{bmatrix} cos\lambda_j^{(M)} & sin\lambda_j^{(M)} \\ -sin\lambda_j^{(M)} & cos\lambda_j^{(M)} \end{bmatrix} \begin{bmatrix} \theta_{j,t-1}^{(M)} \\ \theta_{j,t-1}^{(M)*} \end{bmatrix} + \begin{bmatrix} \xi_{j,t}^{(M)} \\ \xi_{j,t}^{(M)*} \end{bmatrix},$$

$$\lambda_j^{(M)} = \frac{2\pi j}{P_M} \tag{9}$$

$$\xi_{j,t}^{(M)} \sim N(0, \sigma_{\xi_M}); \xi_{j,t}^{(M)*} \sim N(0, \sigma_{\xi_M^*}), \tag{10}$$

where $P_M = 2$. The identification of $s_t^{(A)}$ and $s_t^{(M)}$ is achieved through an appropriate choice of harmonic functions.

The irregular components, $e_{t,i}$, account for the measurement error that the high frequency proxy i has for the economic concept measured by the official statistic. The irregular component has some limited persistent accounted by an AR(1) process:

$$e_{t,i} = \rho_{1,i} e_{t-1,i} + \zeta_{t,i};\ \zeta_{t,i} \sim N(0, \sigma_{\zeta_i});\ i = 1, ..., n \tag{11}$$

2.3 Dealing with Irregular Calendar: weekly to monthly

When incorporating weekly indicators into a mixed-frequency state space framework with monthly variables as the reference frequency, a central challenge arises from the misalignment between the weekly and monthly calendars. Since some calendar weeks span two months, a straightforward temporal aggregation of weekly observations into monthly totals is not well defined. For example, a weekly series recording payments may include days from both the end of the month m and the beginning of the month $m + 1$.

To address this issue, we construct a regularized calendar that enforces a one-to-one correspondence between weeks and months. Let $m = 1, 2, \ldots, M$ index months. Each month m is partitioned into exactly four weeks, indexed by $w = 1, \ldots, 4$, where each week consists of seven consecutive days. Denote by $\mathcal{D}_{m,w}$ the set of calendar days belonging to week w of month m, with

$$|\mathcal{D}_{m,w}| = 7 \quad \text{for } w = 1, \ldots, 4.$$

Since most calendar months have more than 28 days, residual days $\mathcal{R}_m$ remain after this partition. These trailing days are reassigned to the first week of the following month:

$$\mathcal{D}_{m+1,1} \leftarrow \mathcal{D}_{m+1,1} \cup \mathcal{R}_m.$$

For instance, if month m has 30 days, then days 29 and 30 are added to $\mathcal{D}_{m+1,1}$, while the remaining partition of month $m+1$ again proceeds with four seven-day weeks. This convention ensures that each week is uniquely associated with a single month, eliminating ambiguity in the construction of monthly aggregates. The seasonal component of the state space model is defined with respect to this regularized calendar, so that seasonal factors remain consistent with the artificial four-week month structure. In real time, when new weekly data are released on days d of month m with $d \notin \{7, 14, 21, 28\}$, the observation is assigned to the following week. This rule ensures that no information is discarded even if releases occur on irregular calendar dates.

2.4 Bayesian Estimation

The complete model, including high-frequency indicators, is set in state-space form with measurement equations (1) and (8) and state equations (3), (7), (10), and (11).

Parameters are estimated by Gibbs sampling. Table 1 describes the prior distributions. Φ collects the parameters of the cycle equation (3). Γ includes the parameters of the measurement equation for the high-frequency indicators, eq. (8). The parameters of the irregular component process are in ρ, and σ_ζ is the innovation standard deviation. The parameters to describe the t-distributed innovations to the cycle with stochastic volatility are the standard deviation of the innovation to the stochastic volatility, σ_ω, the number of degrees of freedom of the t-distribution, ν, and the mixture parameter λ_u.

The parameters estimated to describe the seasonal components are the standard deviation of the innovations σ_{ξ_A} and σ_{ξ_M}. As the seasonal components are stacked, we need to estimate just one innovation variance of each type of seasonality.

Large swings in the data, like those observed during COVID, might often be confounded with seasonal patterns. To avoid such behavior within our seasonal component, we temporarily reduce the variation of innovations to the seasonal component during 2020, such that the model is able to attribute the large variation in the data to the common factor, which features t-distributed errors with stochastic volatility. In practice, for the 12 months in 2020 we draw the seasonal variance from an Inverse-Wishhart distribution with a smaller scale compared to the one assumed for σ_{ζ^A}.

Table 1: Prior Distributions

Parameter	Distribution	p_1	p_2
Φ	$\mathcal{N}$	$\left[\frac{3}{52}\ \ 0\ \ 0\right]'$	$0.01 I_3$
Γ	$\mathcal{N}$	$\mathbf{0}_3$	$0.05 I_3$
$\boldsymbol{\rho}$	$\mathcal{N}$	0.95	0.001
σ_ζ	$\mathcal{IW}$	0	10
σ_ω	$\mathcal{IW}$	1	1
λ_u	$\mathcal{IG}$	$\frac{\nu_\omega}{2}$	$\frac{2}{\nu_\omega}$
ν	$\mathcal{U}$	3	40
σ_{ξ_A}	$\mathcal{IW}$	$T\sigma^0_{\xi_M}$	T
$\sigma_{\xi_A^{2020}}$	$\mathcal{IW}$	0	$12\times\{\#2020\}$
σ_{ξ_M}	$\mathcal{IW}$	0	1

Note: p_1 and p_2 represent the two parameters of the respective prior distribution. For the Normal distribution $\mathcal{N}$, they are the mean and variance; for the Inverse-Wishart $\mathcal{IW}$, they represent the scale and degrees of freedom; for the Inverse-Gamma $\mathcal{IG}$, they are the scale and shape; for the Uniform $\mathcal{U}$ they are the min and max. #2020 indicates the number of observations in 2020.

2.4.1 Gibbs Sampler Algorithm

The Gibbs sampler iteratively samples from the joint posterior of latent states, stochastic volatilities, and parameters. The latent states are $\{L_t, s_t, e_t\}$ and represented by f_t.

1. Set initial values for model parameters $\theta^{(0)}$ and volatility path for $\{\sigma^{(0)}_{u,t}\}_{t=1}^T$.

2. $p\left(f_t^{(j)}|y_t, \sigma_{u,t}^{(j-1)}, \theta^{(j-1)}\right)$. Draw latent states $\{L_t^{(j)}, s_t^{(j)}, e_t^{(j)}\}_{t=1}^T$ conditional on pa-

rameters $\theta^{(j-1)}$ via precision sampling.

3. $p\left(\Gamma^{(j)}, \rho^{(0)} | f_t^{(j)}, y_t\right)$. Draw the free elements of $\Gamma^{(j)}$ and $\rho_0^{(j)}$ by regressing $y_t^{SA,M,(j)} = y_t - s_t^{M,(j)} - e_t^{(j)}$ on $f_t^{(j)}$, $s^{A,(j)}$ and a constant.

4. $p\left(\rho_{1,...,p}^{(j)} | f_t^{(j)}\right)$. Draw the serial correlation, $\rho_{1,...,p}^{(j)}$ in the irregular by fitting an $AR(p)$ to $e_t^{(j)}$. We set $p = 1$.

5. $p\left(\Phi^{(j)} | f_t^{(j)}, \sigma_{u,t}^{(j)}\right)$. Draw $\Phi^{(j)}$ following Equation (3). We include a rejection step in case $|I - \Phi^{(j)} z| = 0$ or $|z| > 1$.

6. $p\left(\sigma_{\zeta^A}^{(j)}, \sigma_{\zeta^{2020}}^{(j)}, \sigma_{\zeta^M}^{(j)} | f_t^{(j)}, \theta^{j-1}\right)$. Draw each seasonal variance $\sigma_{\zeta^A}^{(j)}$, $\sigma_{\zeta^{2020}}^{(j)}$ and $\sigma_{\zeta^M}^{(j)}$, from Inverse-Wishart distributions, by pooling the $P_{A,t}$, $P_{2020,t}$ and $P_{M,t}$ components.

7. $p\left(\sigma_{u,t}^{(j)} | f_t^{(j)}, \sigma_{u,t}^{(j-1)}, \theta^{(j-1)}\right)$. Draw the stochastic volatility path $\{\sigma_{u,t}^{(j)}\}_{t=1}^{T}$ using the Kim et al. (1998) mixture-of-Normals algorithm.

8. $p\left(\lambda_{u,t}^{(j)} | \nu_{\omega}^{(j)} f_t^{(j)} \sigma_{u,t}^{(j1)}, \theta^{(j-1)}\right) p(\nu)$. Following Jacquier et al. (2002), draw the t-distributed error $\eta_t^{(j)}$ as a scale-mixture $\eta_t^{(j)} = \sqrt{\lambda_{u,t}^{(j)}} \epsilon_t^{(j)}$, where $\epsilon_t \sim \mathcal{N}(0,1)$, $\lambda_{u,t} \sim \mathcal{IG}\left(\frac{\nu_w}{2}, \frac{2}{\nu_w}\right)$, and $\nu \sim \mathcal{U}[3, 40]$.

9. Increment j and repeat steps 2–8 until convergence of the chains.

2.5 Evaluation of Real-Time Performance

As the high-frequency indicators are available before the release of the official statistics, we use a real-time forecasting exercise to evaluate the accuracy of the model in anticipating the official statistics value. The out-of-sample exercise evaluates predictions for y_τ^O for $\tau = 1, ..., P$ in months. These are computed based on four different information sets: $\tau | w_1, ..., \tau | w_4$, where w_1 denotes the first week in the reference month. We also consider forecasts computed within month τ for the next quarter, $y_{\tau+1}$. This implies that we evaluate the real-time performance of the modeling for prediction horizons from one to eight weeks.

We assume a squared loss function to evaluate the model's performance. The relative model performance is computed relative to a model that does not include the high-frequency indicator and applies filtering based on a structural state-space model with level and seasonal components. This helps us evaluate whether the high-frequency indicators are informative for tracking monthly official statistics.

3 ONS Official Statistics and Real-Time Indicators

We use a set of monthly official statistics and real-time indicators from the UK Office for National Statistics (ONS). The ONS classifies the real-time indicators as 'official statistics in development'.

3.1 Official Statistics

We consider three monthly statistics in our empirical exercises. The first one is monthly retail sales, measured in nominal terms and seasonally unadjusted, covering the period from January 1996 to May 2025. We also use the monthly real Gross Domestic Product (GDP), which is seasonally adjusted, available from January 1997 to May 2025. In addition, we examine data on job vacancies over the last three months, reported monthly in thousands by the ONS, spanning from June 2001 to May 2025.

The retail sales series and vacancies are typically published with a 20-day delay. Monthly GDP is published with a 40-day delay.

3.2 Real-Time Weekly Indicators

Alongside these official statistics, we incorporate high-frequency indicators to provide weekly, real-time estimates of the official statistics.

The first series we use is daily Revolut debit-card spending. The data is available as a seven-day moving average index across nine mutually exclusive categories, and is published by the ONS with a 4-day delay. For tracking retail sales, we consider the focus on the categories of groceries and shopping. Data is available from 2020 to week 33 of 2025.

Figure 1 shows the data on debit card spending on groceries and shopping against the monthly retail sales. The figure also includes overall spending. Even if debit card spending on the two categories highlighted is highly correlated with the overall spending, as shown in Figure 2, there is still some potentially useful additional information, as indicated by the number of observations far from the 45-degree line. Both the official statistics and retail sales, and the real-time indicators exhibit seasonal behavior.

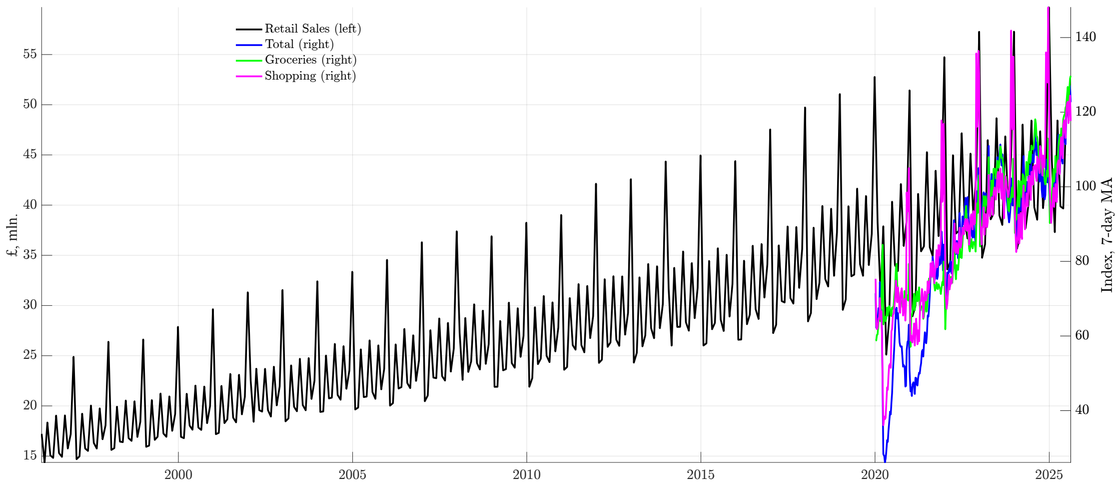


Figure 1: Retail sales and debit card spending

We also consider the use of the debit card spending data to estimate a high-frequency tracker for UK monthly GDP. Figure 3 shows how debit card spending behaves compared with the monthly GDP series. In comparison with tracking retail sales, two adjustments need to be made. The monthly GDP series is seasonally adjusted and in volume terms, so the annual seasonal factor is likely to capture the possibility of residual seasonality over the period before 2020, when spending data is not available. We divide the weekly spending by the monthly CPI to obtain a measure of consumer spending in real terms.

Finally, we use online jobs advert data from TextKernel, which is collected through

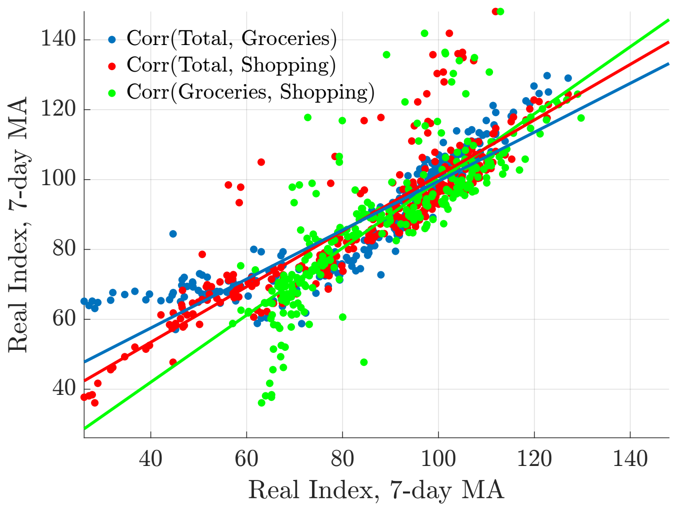


Figure 2: Correlation across Revolut Debit Card Spending Categories

Figure 3: Monthly GDP and Revolut Debit Card spending

web scraping and processed using a methodology designed to remove duplicates. For tracking vacancies, we employ the overall weekly number of adverts, covering the period from 2018 to week 20 of 2025.

Figure 4 illustrates the online job advertisement data in comparison to the 3-month moving sum of vacancies. The real-time indicator is clearly more volatile than the official statistics, with some apparent jumps. The irregular component is designed to capture this type of behavior of the high-frequency indicator.

4 Estimates of Unobserved Components

The key state variable to track official statistics in real time is L_t, as it excludes the seasonality of monthly and weekly data, as well as the excessive noise of the weekly data. As we estimate the model using Gibbs sampling, we can fully characterize the uncertainty of these estimates via a density distribution at each t. The variance of the density is allowed to change over time as we allow for a slowly moving variance change in the cycle process in Eq. (3).

The model is estimated using y_t^O in log-levels, so our tracker estimates are in logs.

4.1 Tracking Retail Sales and Monthly GDP with Debit Card Spending Data

Figure 5 shows our tracker of retail sales using the overall debit-card transactions data. The tracker captures the swings during the Pandemic, with the excessive variation of the weekly transaction data accommodated by the irregular component.

The within-year seasonality displayed in Figure 6 is typically regular. Still, one can see how the model captures changes in the seasonal pattern during 2020-2021, caused by the impact of the COVID-19 restrictions. The model also captures within-month seasonality related mainly to the weekly card transaction data.

As a preliminary check on how our weekly tracker behaves compared to retail sales,

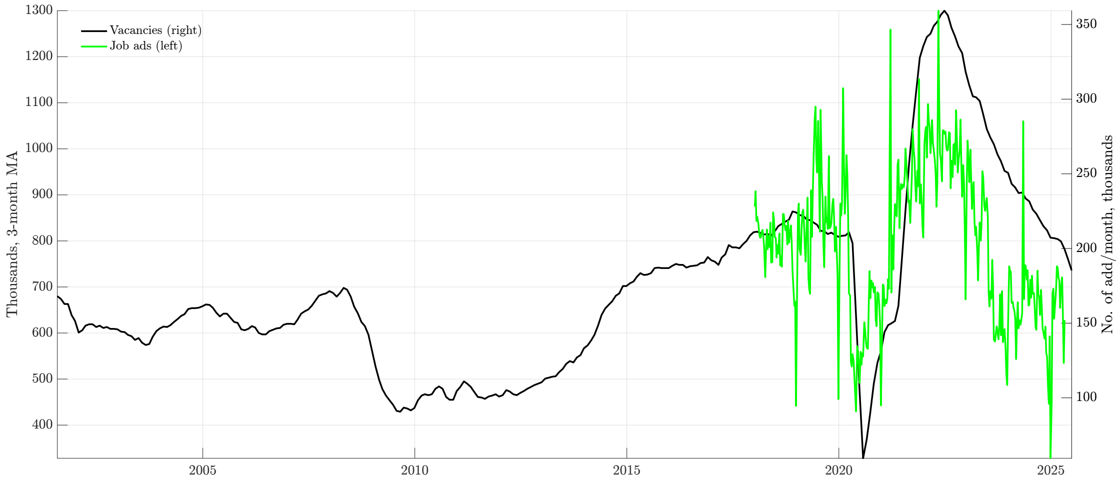
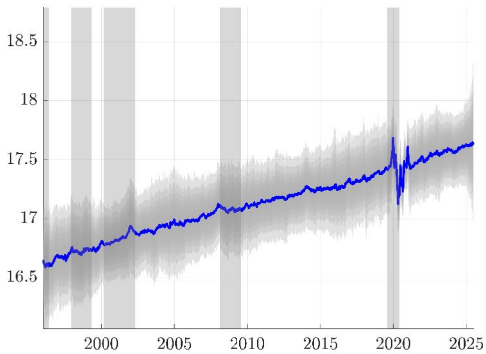
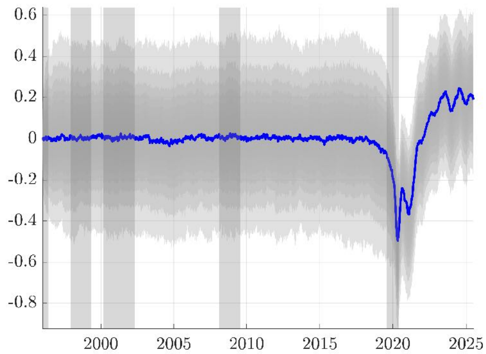


Figure 4: Retail sales and debit card spending



(a) Level (L_t)



(b) Irregular (e_t)

Figure 5: Retail Sales: Common Level and Irregular Components

we show in Figure 7 the year-on-year changes in the level estimate vs the actual YoY changes in retail sales. The red circles represent the observed monthly data, and the blue line displays the model estimates. One can see that for the period that we do not have weekly data, the model misses some large swings in the data. This changes from 2020 as then we find the model able to capture most of the data variations with the aid of the weekly data.

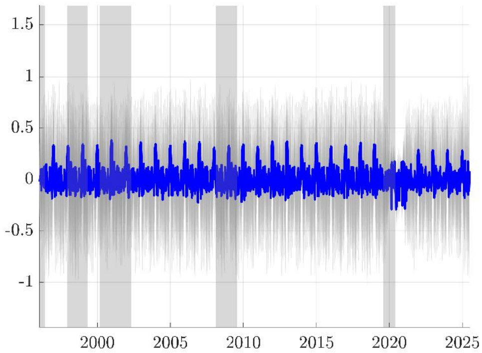
As our modeling approach accommodates a set of high-frequency indicators, we consider separately two categories of debit card spending that are likely to be more related to retail sales. These are Revolut debit card spending in groceries and shopping. Figure 8 shows the tracker using these two categories. The level estimates are not far from the ones in 5, but it is clear that the uncertainty around estimates is reduced after 2021 by the inclusion of the two high-frequency indicators. The irregular component shows evidence of residual seasonality as we restrict seasonal components to only capture the common seasonality among official and high-frequency indicators.

Figure 9 suggests that the improvement in the precision of the retail sales tracker is likely to be driven by improved estimates of the seasonal components. By using two categories instead of the aggregate index, both the within-year and within-month seasonality estimates improved.

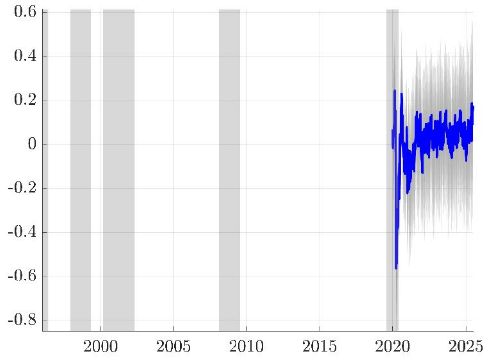
Figure 10 shows that the two-categories modeling performance in tracking retail sales is not far from the model with just the aggregate index in Figure 10.

4.2 Tracking Monthly GDP with Debit Card Spending Data

The ONS estimate of monthly GDP is based on value added by industry. Among these, the service industries are typically the main drivers of UK growth. Consequently, we also consider debit card spending as a high-frequency indicator to provide a weekly tracker of monthly GDP. The full-sample estimates for this tracker are presented in Figure 11. As in the case of retail sales, the irregular components capture variations in spending



(a) Annual seasonal (s_t^A)



(b) Monthly seasonal (s_t^M)

Figure 6: Retail Sales: Seasonal Components

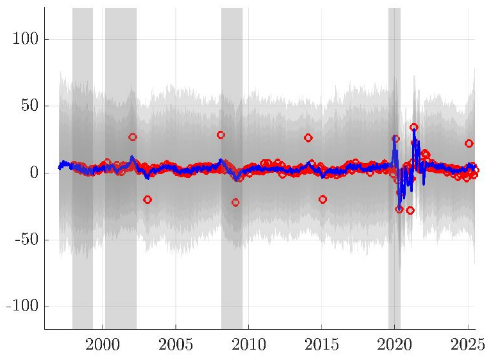
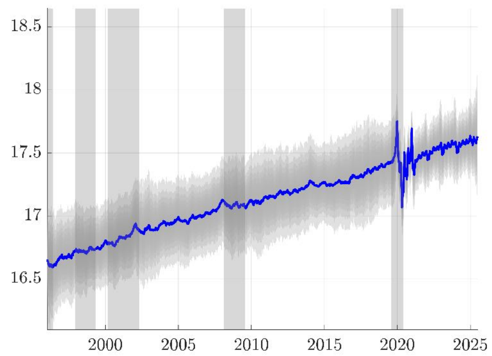
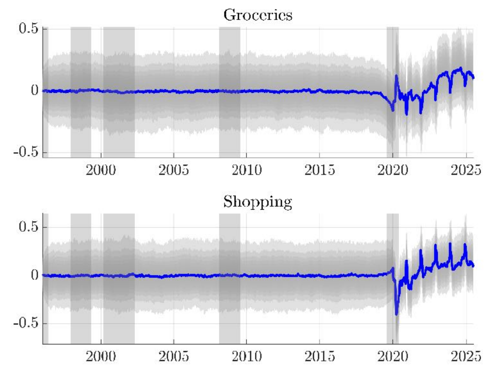


Figure 7: YoY Level Estimate (blue line) vs. YoY Retail Sales (red circles)

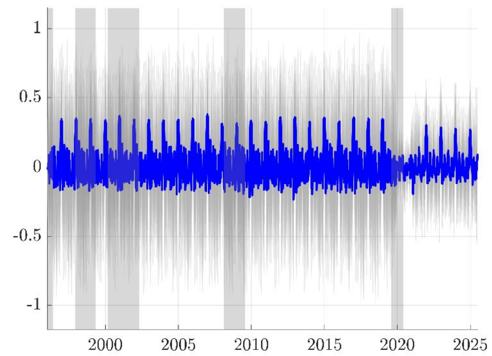


(a) Level (L_t)

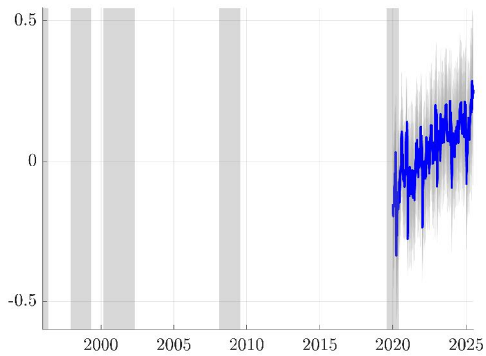


(b) Irregular (e_t)

Figure 8: Retail Sales with Two Spending Categories: Common Level and Irregular Components



(a) Annual seasonal (s_t^A)



(b) Monthly seasonal (s_t^M)

Figure 9: Retail Sales with Two Spending Categories: Seasonal Components

that do not contribute to tracking monthly GDP, indicating that spending accelerated in 2022 more rapidly than the GDP recovery actually observed.

Monthly GDP is seasonally adjusted, and the estimates up to 2020 in Figure 12 show no evidence of residual seasonality. From 2020 onwards, some annual seasonality is detected as the within-month seasonality, given the information added by the weekly spending data.

The year-on-year change in monthly GDP compared to our tracker equivalent is presented in Figure 13. The model finds no difficulties in tracking GDP growth.

4.3 Tracking Vacancies with Online Ads Data

The number of new vacancies is a key measure to track the stance of the UK labour market. A declining number of vacancies may signal a decrease in overall demand from firms for UK workers, leading to downward pressure on wage growth and possible consequences for aggregate prices and inflation. We use the number of online job adverts to anticipate changes in UK vacancies every month.

The tracker is presented in Figure 14. The irregular component shows that online job adverts recovered faster than the official statistics in 2021, and indicated a faster decline in vacancies in 2024.

Even if vacancies are published as a 3-month average, Figure 15 shows that the model still finds some annual seasonality in the data, with more information available after 2018 when the online jobs data is available.

Comparing the tracker for year-on-year changes in vacancies in Figure 16 reveals that online adds data helps identify an earlier decline in vacancy growth in 2022 and a faster increase in 2024.

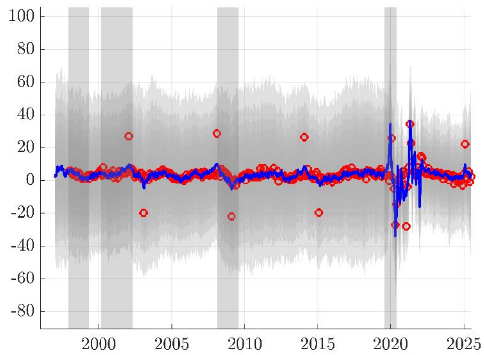
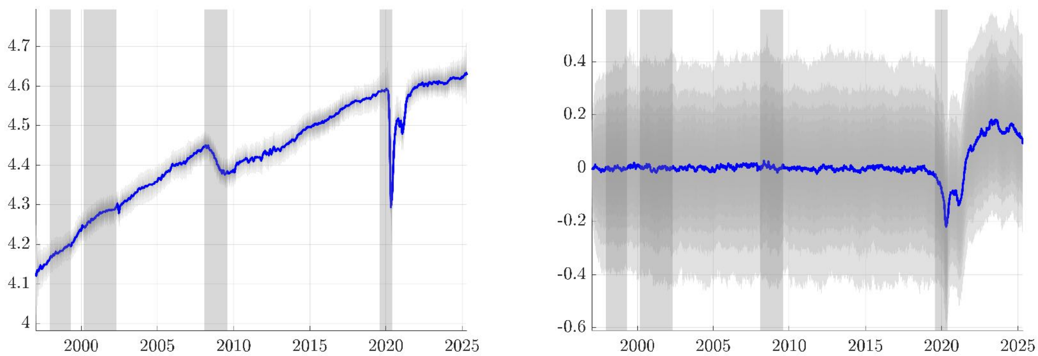


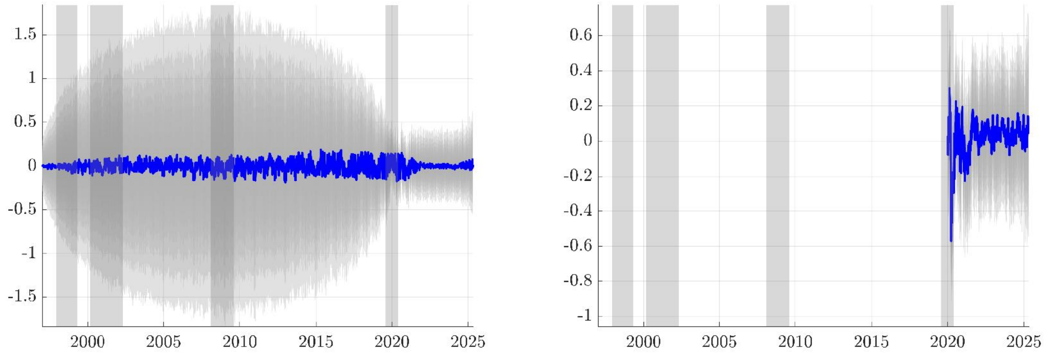
Figure 10: YoY Level Estimate (blue line) vs. YoY Retail Sales (red circles) from the two-categories specification



(a) Level (L_t)

(b) Irregular (e_t)

Figure 11: Monthly GDP: Common Level and Irregular Components



(a) Annual seasonal (s_t^A)

(b) Monthly seasonal (s_t^M)

Figure 12: Monthly GDP: Seasonal Components

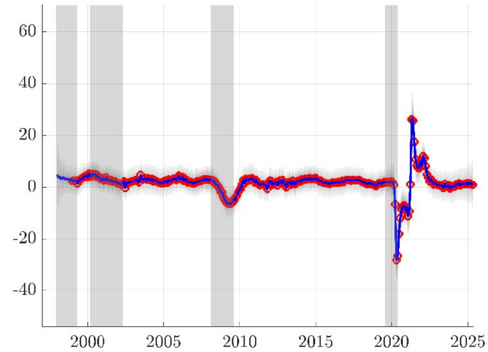
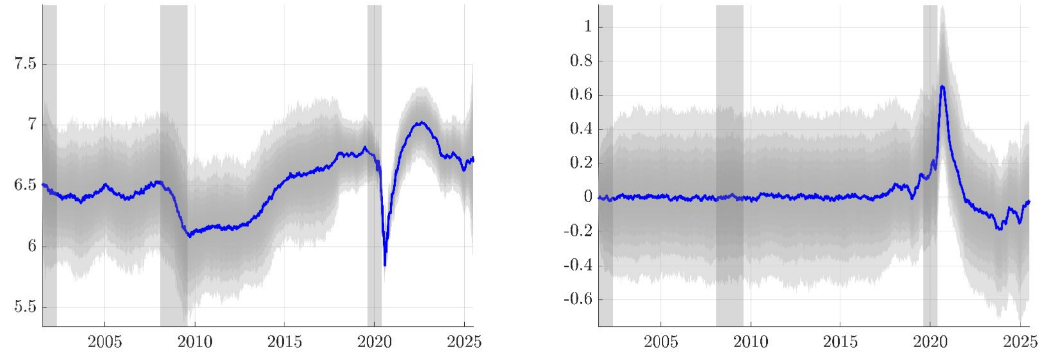


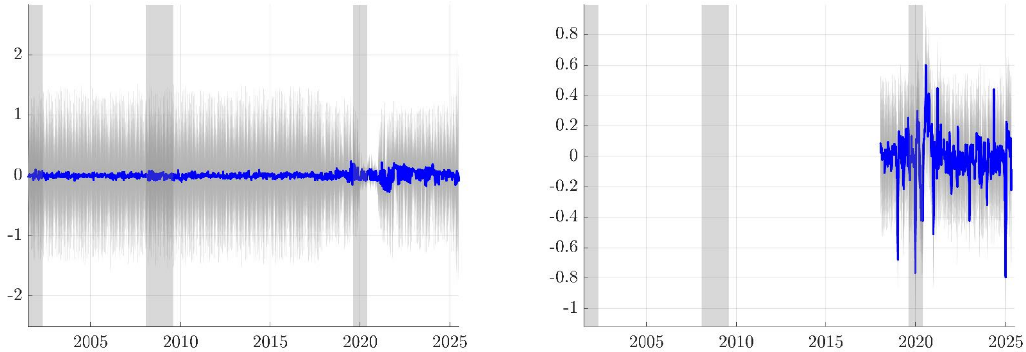
Figure 13: YoY Level Estimate vs. YoY Monthly GDP



(a) Level (L_t)

(b) Irregular (e_t)

Figure 14: Vacancies: Common Level and Irregular Components



(a) Annual seasonal (s_t^A)

(b) Monthly seasonal (s_t^M)

Figure 15: Vacancies: Seasonal Components

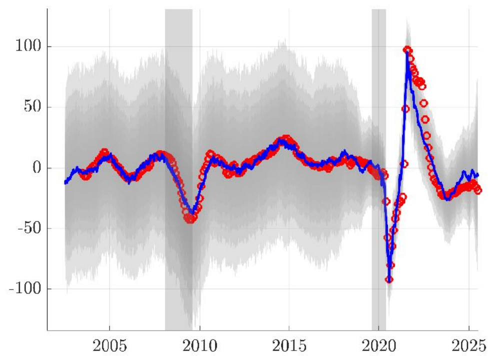


Figure 16: YoY Level Estimate vs. YoY Vacancies

5 Real-Time Forecasting Evaluation

The main question we address in this section is: Does the inclusion of high-frequency indicators improve the weekly tracking of the monthly official statistics? We answer this question by comparing the forecasting performance, measured by the mean squared error, of the model with a high-frequency indicator versus a model where the only observed variable is the official statistic.

As described earlier, at the current month τ, we compute predictions for y^O_τ. These are conditional on data available up to the first, second, third, and fourth weeks of the month. We use similar information sets to compute predictions for $y^O_{\tau+1}$. The sample employed in the evaluation is as such as $\tau = Jan2020, ..., May2025$.

Table 2 provides the mean squared error ratios between the model with a high-frequency indicator and the one with only monthly data. The gains are more substantial for using the total debit-card spending to track retail sales. Similar data reduces the error in predicting monthly GDP every week by half. The inclusion of two categories of spending does not improve real-time forecasting performance. The empirical results suggest that a more extended sample period is needed to handle both series effectively, as the performance remains relatively stable across weekly information sets.

For tracking vacancies, we find that the number of job adverts is helpful if we have observed at least two weeks of the high frequency indicator during the current quarter.

The findings from this real-time exercise demonstrate that high-frequency, real-time proxies for official statistics can effectively provide earlier estimates of official measures.

6 Conclusion

The unobserved components model developed in this paper enables us to extract information from seasonally unadjusted high-frequency data that is useful for producing accurate real-time estimates of low-frequency economic statistics. The proposed frame-

Table 2: Relative MSE Performance: Model with Weekly Indicator vs Monthly Model

Prediction for:	Current Month				Next Month			
Data up to:	Week 1	Week 2	Week 3	Week 4	Week 1	Week 2	Week 3	Week 4
HF Indicator:	Retail Sales							
Debit-Card Spending	0.393	0.413	0.427	0.437	0.400	0.417	0.420	0.417
Debit-Card Groceries & Shopping	0.489	0.581	0.611	0.521	0.464	0.586	0.600	0.496
	Monthly GDP							
Debit-Card Spending	0.509	0.499	0.540	0.500	0.551	0.562	0.583	0.555
	Vacancies							
Job Ads	1.085	0.972	0.778	0.827	1.147	0.987	0.718	0.797

Note: The results are mean squared error ratios. Values smaller than one indicate that the weekly frequency indicator adds value in real-time. The out-of-sample period is from January 2020 to May 2025. Week 1-4 indicate the time at which end-of-period forecasts are performed.

work accommodates several features of real-time alternative indicators: they are typically available over a shorter time span than official statistics, exhibit both intra-year and intra-month seasonality, and are noisier than the series they are intended to track.

Our empirical exercises highlight the value of debit-card spending data and online job advertisements for tracking monthly statistics such as retail sales, GDP, and vacancies. This research provides strong support for the usefulness of alternative data sources to provide earlier estimates of key UK monthly economic variables.

References

Antolin-Diaz, J., T. Drechsel, and I. Petrella (2024): "Advances in nowcasting economic activity: The role of heterogeneous dynamics and fat tails," Journal of Econometrics, 238, 105634.

Aruoba, S. B., F. X. Diebold, and C. Scotti (2009): "Real-time measurement of business conditions," Journal of Business & Economic Statistics, 27, 417–427.

Bańbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013): "Now-casting and the real-time data flow," in Handbook of economic forecasting, Elsevier, vol. 2, 195–237.

Brave, S. A., M. Fogarty, D. Aaronson, E. Karger, and S. D. Krane (2021): "Tracking US consumers in real time with a new weekly index of retail trade," .

Corsi, F. (2009): "A Simple Approximate Long-Memory Model of Realized Volatility," Journal of Financial Econometrics, 7, 174–196.

Jacquier, E., N. G. Polson, and P. E. Rossi (2002): "Bayesian analysis of stochastic volatility models," Journal of Business & Economic Statistics, 20, 69–87.

Kim, S., N. Shephard, and S. Chib (1998): "Stochastic volatility: likelihood inference and comparison with ARCH models," The review of economic studies, 65, 361–393.

Lewis, D. J., K. Mertens, J. H. Stock, and M. Trivedi (2021): "High-frequency data and a weekly economic index during the pandemic," in AEA Papers and Proceedings, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 111, 326–330.

Mariano, R. S. and Y. Murasawa (2003): "A new coincident index of business cycles based on monthly and quarterly series," Journal of applied Econometrics, 18, 427–443.