Nowcasting Economic Activity with Fat Tails and Outliers *

Seokki Simon Hong †

First version: 14 September 2023
This version: 26 October 2023

Abstract

The COVID-19 presented macroeconomic models with unique challenges, marked by extreme outliers in economic data. This paper extends dynamic factor models by explicitly incorporating outliers, moving beyond conventional data screening practices. The methodological contribution includes introducing fat tails and outliers multiplicatively into innovation volatility, and two distinct approaches for modeling outliers are presented to address large jumps. Empirical findings demonstrate that outlier-augmented models consistently outperform benchmark models in point and density forecasting, with the most significant improvements observed in nowcasting horizons. Incorporating outliers becomes particularly crucial during major crises, enhancing forecasting accuracy by 44% compared to the benchmark. The uniform-mixture approach is found to be more robust than the student-t models, as it targets extreme variations without disrupting the smoothness of the stochastic volatility process. Overall, this paper enhances macroeconomic modeling by explicitly addressing outliers, improving forecasting accuracy, and providing insights into economic dynamics during and after major crises like the COVID-19 pandemic.

Keywords: Now-casting, Dynamic factor models, Bayesian Methods.
JEL Classification: C11, C32, C38, C53, E37.

*I am grateful to Catherine Doz, Riccardo Degasperi, Francesco del Prato, Thomas Drechsel, Laurent Ferrara, Livia Paranhos, Ivan Petrella, Pierre-Alain Pionnier, Giovanni Ricco, and seminar participants at the PSE Workshop on Nowcasting for insightful comments. I acknowledge support from the Measurement in Economics, Nowcasting-Beyond GDP chair at Paris School of Economics - École d'Économie de Paris.

†Paris School of Economics, 48 Boulevard Jourdan, Paris, France. Email: simon.hong@psemail.eu

1 Introduction

The COVID-19 pandemic, with its widespread lockdowns and social distancing measures, presented macroeconomic models with an array of unique challenges. Economic indicators reached unprecedented levels during this period; for instance, in the week ending March 21, 2020, a staggering 3.28 million Americans filed for unemployment claims, surpassing the previous record set during the 1982 recession. The U.S. real GDP plummeted by a staggering 31.4% during the second quarter of 2020, marking an all-time low since the Bureau of Economic Analysis (BEA) began tracking these figures in 1947.[1] Such extreme observations left the New York Federal Reserve's benchmark nowcasting model in disarray, as this constant-parameter model resulted in implausible forecast paths.

Although the economic turbulence witnessed during the COVID-19 pandemic was unparalleled, the presence of outliers within macroeconomic data is a phenomenon with historical antecedents. It is, in fact, a recurring feature within the macroeconomic data landscape, as a manifestation of exceptional events such as labor strikes and natural disasters. Illustratively, Figure 1 portrays selected macroeconomic indicators from the French context spanning the period of 1980 to 2019. A cursory examination of this figure underscores the omnipresence of conspicuous, one-off outliers across four series. It is worth noting that the early contributions in the field of macroeconomic forecasting, including Stock and Watson (2002), recognized this phenomenon and proposed the screening of outliers via replacement with missing values.[2]

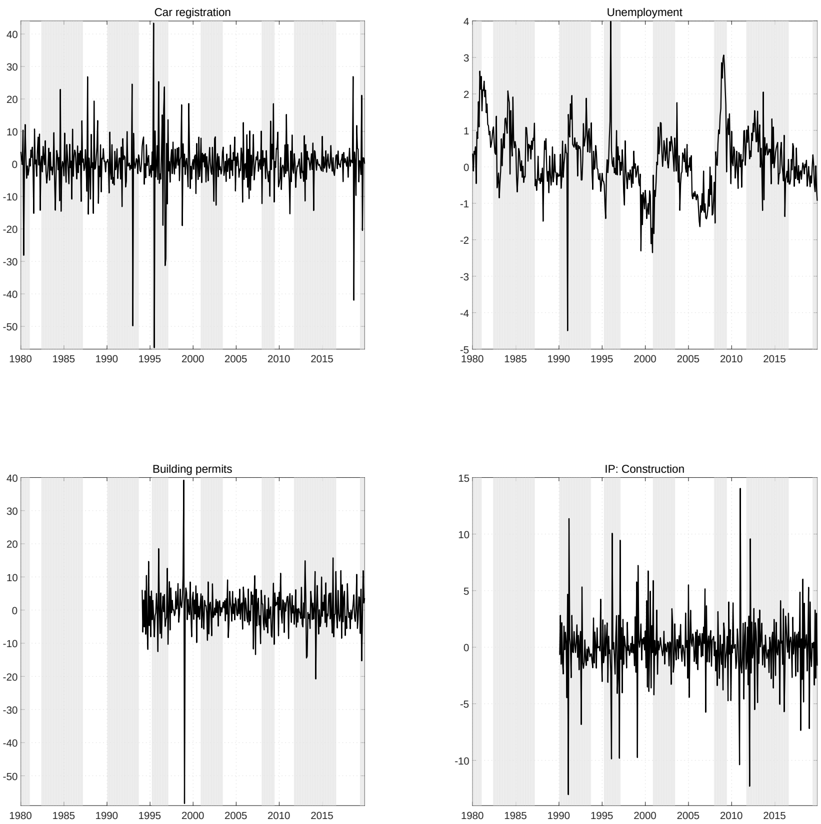
This paper extends the dynamic factor models (DFMs) paradigm by introducing explicit modeling of outliers, moving beyond data screening practices. Within the DFM framework, I take the view that outliers primarily represent transient spikes in volatility, not permanent disturbances. The model specification comfortably accommodates conventional models while providing a more intuitive way of dealing with outliers: I introduce and compare two distinct modeling strategies to address large jumps. Importantly, the estimation process is entirely data-driven, allowing the model to estimate outliers based solely on the data without prior constraints.

Subsequently, I take on the role of policymakers and market participants to conduct

[1] The figures are based on preliminary estimates: the FRED database now shows 2.91 million and 28% for initial claims and real GDP growth, respectively, after data revisions.

[2] Stock and Watson (2002) replaced observations exceeding 10 times the interquartile range from the median by missing values.

Figure 1: Existence of one-off outliers in French data

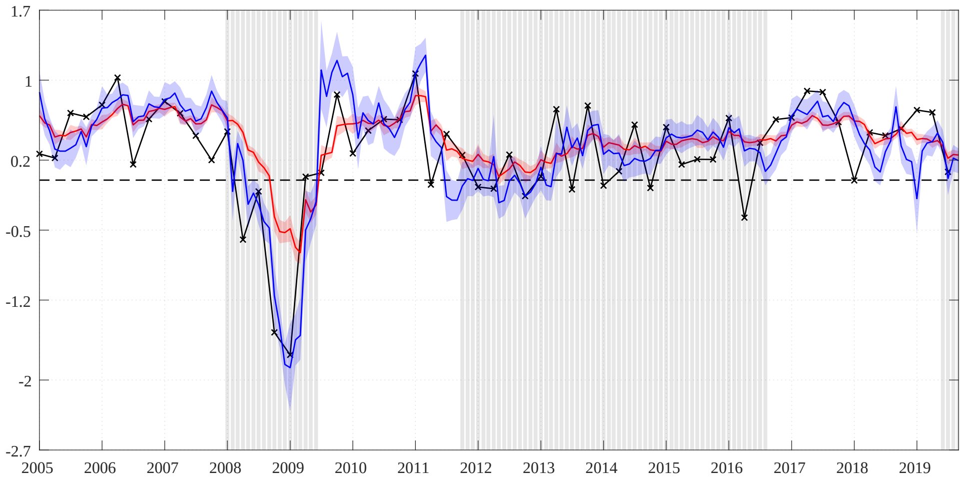


Note: This figure plots four monthly indicators for the French economy: passenger car registrations, unemployment, building permits, and industrial production: construction index. Sample: 1980.1 – 2019.12. The data are transformed in terms of month-on-month % changes. OECD recessions in gray. Source: DBnomics.

pseudo real-time out-of-sample forecasting analyses, comparing the model's performance to the benchmark model for both pre- and post-COVID periods.[3] In the context of nowcasting French GDP, I show that the explicit modeling of outliers is particularly beneficial during crises, as shown in Figure 2: while the benchmark model tends to underestimate the severity of recession and overestimate economic dynamics during normal times, the model with outliers adeptly captures economic activity in a more timely manner. As a result, the forecasting accuracy improves by 44% compared to the benchmark even when post-2020 samples are included. Following the recent literature dealing with econometric challenges after COVID-19 (e.g. Lenza and Primiceri, 2020; Schorfheide and Song, 2021), I also examine the model's ability to capture the shape of recovery right after the pandemic, revisiting the debate on the "U v. V-shaped" recovery.

[3]Throughout the paper, the benchmark model refers to constant-parameter DFM models of Bańbura and Modugno (2014) and Bańbura et al. (2010), among others.

Figure 2: Nowcasting the French GDP, 2005 – 2019



Note: This plot shows the realizations (in black) and pseudo real-time nowcasts of French GDP from 2005 to 2019 constructed from the model with outliers(in blue) and the benchmark DFM model (in red). Shaded areas denote the 68% bands. OECD recessions in gray.

The methodological contribution of this paper consists of two key components. Firstly, I introduce fat tails and outliers into the DFM framework by incorporating them multiplicatively into innovation volatility. The standard stochastic volatility model represents innovations with two separate error processes: conditional volatility following a log-normal auto-regressive model and a standard normal error. Inspired by the finance literature, particularly Jacquier et al. (2004), I further decompose the latter into an outlier state and noise, enhancing the model's resistance to outliers. This approach introduces pronounced jumps in the data rather than escalating stochastic volatility itself, resulting in a more stable volatility trajectory. While it is possible to model outliers additively, I demonstrate that the multiplicative approach achieves the same goal from a measurement-error perspective and remains independent of the specification of other components in the model. Consequently, outliers genuinely appear as one-off anomalies within this framework.

Secondly, I explicitly model outliers using two distinct approaches, discussing their different treatments of large jumps and interpretations of priors. Outliers are introduced into the model in two ways. One approach assumes that innovations exhibit fat tails and follow the Student-t distribution, a method more commonly employed in the literature. However, this is not the only way to model outliers. In the spirit of Stock and Watson (2016), the other approach assumes a uniform mixture distribution that transitions between regular observations and outlier states. I show that this modeling strategy is better suited for addressing large jumps akin to the COVID-19 shock, assigning more probability mass to events causing volatility to increase by more than twofold compared to the Student-t approach, which concentrates most probability mass on values close to one. Furthermore, this approach is more intuitive in terms of implementing and interpreting priors, as one can easily determine the prior probability of landing in outlier states by counting the occurrences of outliers in the pre-sample, even variable by variable. As we explore richer economic dynamics, we face the typical trade-off: estimating more parameters increases estimation uncertainty. Therefore, I adopt a Bayesian approach, which shrinks uninformative parameters toward a more parsimonious specification to address the curse of dimensionality.

Upon applying these models to forecast the French economy for both pre- and post-COVID periods, I arrive at three key findings. First, outlier-augmented models, irrespective of the distributional assumptions, consistently outperform the benchmark model in both point and density forecasting. In the pre-pandemic sample, the full model incorporating uniform outliers demonstrates $5-18\%$ improvements compared to the benchmark model, as indicated by reduced Root Mean Squared Error (RMSE) and Continuous Ranked Probability Score (CRPS) values across all forecasting horizons. The student-t approach displays a similar trend, except for the longest forecasting horizon. However, the 'screening' method, which replaces pre-defined outliers with missing values, induces a negligible effect on accuracy. Notably, the most significant enhancements are observed in the nowcasting horizons, where models with outliers outperform both the benchmark and other stochastic volatility specifications in point and density nowcasts. This performance extends even to the full sample up to 2022, and the findings remain robust when alternative metrics for assessing prediction accuracy are employed.

Second, the significance of incorporating outliers becomes even more pronounced during major crises such as the Great Recession and the COVID-19 pandemic. For instance, when considering post-2020 data, the model with uniform outliers enhances the Root Mean Squared Error (RMSE) of the benchmark specification by an impressive 44%. This represents a substantial increase compared to the 13% improvement observed in the pre-COVID scenario. In times of extreme events, outlier-augmented models distinctly differentiate between short-term and long-lasting spikes in uncertainty. Consequently, they adequately capture the peaks and troughs characterizing the Great Recession and the

COVID-19 crisis with greater timeliness and accuracy, while the benchmark model tends to underestimate the severity of the economic downturn and the subsequent rebound. The explicit modeling of outliers enhances the forecasting of economic developments, particularly the 'V-shaped' recovery in the initial phase of the pandemic, by effectively identifying early signs of recovery from data collected in April 2020.

Finally, when comparing different methods to incorporate outliers, the uniform mixture approach emerges as a more advantageous choice compared to the commonly employed student-t models. The uniform approach specifically targets extreme variations without unduly disrupting the smoothness of the stochastic volatility (SV) process when it is already behaving as expected. In contrast, student-t-distributed outliers tend to excessively suppress the SV process, unconditionally decreasing the level of volatilities for all indicators in the dataset. It further flattens already smooth time series, such as the SV estimates of GDP, resulting in nearly constant values over 40 years. These characteristics are less desirable in light of the model assumptions and the established narrative in the existing literature. Despite these drawbacks, the student-t method offers some benefits in terms of forecasting, particularly during times of crises and for density forecasting. Ultimately, the choice between these approaches should be made based on the specific research objectives and the particular context of the analysis.

The structure of the paper is as follows: Section 2 describes the econometric framework including the treatment of outliers. Section 3 introduces the data and provides results from the in-sample analyses using the pre-COVID sample, with a focus on comparing two different approaches to incorporate the outliers. Section 4 conducts pseudo real-time forecasting exercises with pre and post-COVID samples. Section 5 studies the ability of the model in terms of capturing the shape of recovery during the early phase of COVID-19. Section 6 concludes.

Related Literature. This paper aims to achieve methodological advances in macroeconomic forecasting with DFMs. After seminal papers such as Stock and Watson (2002), Evans (2005), and Giannone et al. (2008) formalized the application of DFMs to nowcast the economy, there have been many efforts to incorporate structural changes within the DFM framework.[4] Del Negro and Otrok (2008) allow for time-varying factor

[4]See Bańbura et al. (2010) for the survey on nowcasting with standard DFMs.

loadings and volatilities, and Marcellino et al. (2016) introduce the stochastic volatility in innovations to factors and idiosyncratic components. Antolin-Diaz et al. (2017) and Doz et al. (2020) feature shifts in the trend GDP growth and volatility over time. Antolin-Diaz et al. (2023) is the closest to this paper, where they introduce the outliers as an additive component that is assumed to follow the student-t distribution. While building upon their foundation, this paper also examines the limitations of their methodology: their additive specification inadvertently introduces a dependency of outliers on other components within the model. Moreover, the use of the student-t distribution for modeling outliers tends to dampen the stochastic volatility process in inherently smooth series. In response to these challenges, I propose alternative approaches to mitigate these issues by (i) incorporating outliers in a multiplicative fashion within the innovation volatility component and (ii) exploring two different strategies for modeling outliers, student-t and uniform distributions.

More broadly, it also speaks to an empirical literature that explores the role of time-variation, non-linearities, and non-gaussian shocks in macroeconomic models. This includes, but not limited to, Cogley and Sargent (2005), Primiceri (2005), Fernández-Villaverde et al. (2015), Carriero et al. (2022), and Chan (2023) in Bayesian Vector Autoregression (BVAR) models. This paper also takes a Bayesian perspective that enables the inclusion of more parameters and provides probabilistic predictions, particularly to take account of time-varying components and fat-tailed outliers.

The focus on the role of fat tails and outliers also relates to the studies addressing new challenges to empirical models after 2020, such as Lenza and Primiceri (2020), Marcellino et al. (2021), Schorfheide and Song (2021), and Cascaldi-Garcia (2022). Diebold (2020) studies the real-time performance of the business conditions index by Aruoba et al. (2009) during the early periods of the pandemic, and Lewis et al. (2022) provide a weekly economic index that tracks the rapid economic developments during the first 10 months of the pandemic. Ng (2021) interprets the variations around the outbreak as outliers and attempts to clean the data by using COVID indicators as controls. This paper also takes a similar view and explores the 2020 episode as a useful case study.

2 Econometric Framework

In this section, I introduce the dynamic factor model (DFM) which incorporates fat tails and outliers. Building on the models of Antolin-Diaz et al. (2017) and Marcellino et al. (2016), that introduce the shifts in long-run growth and stochastic volatility (SV) in innovations, I describe how to model outliers from two distinct approaches: the student-t and uniform mixture distribution. While the outliers are augmented by the SV of innovations in a multiplicative way in both cases, I discuss how they differ in dealing with large jumps and the interpretation of priors. In order to let the data speak, I impose the standard Minnesota-type of priors based on the stylized facts of the data. Finally, I briefly comment on the problem of mixed frequency and provide a sketch of the estimation algorithm.

2.1 The dynamic factor model

Let y_t denote n $\times$ 1 vector of macroeconomic indicators. The main idea of DFM is that a small number of latent common factors, f_t capture the majority of the joint dynamics across variables. Hence, it is assumed that the dimension of f_t is k $\times$ 1 and k $<<$ n. The factors are loaded via coefficients Λ(L), which represent the response of each indicator to a common shock. u_t is an idiosyncratic component that captures variable-specific movements or measurement errors. Eq (1) is what we usually call the "measurement equation" in the standard DFM literature.

$$\Delta y_t = c_t + \Lambda(L) f_t + u_t \tag{1}$$

In addition to the standard formulation, we allow the long-run growth rate of real GDP, and possibly other series, to drift gradually over time: the time-varying intercept c_t catches shifts in the long-run mean of Δy_t.[5] This specification is based on accumulating evidence that the long-run growth rate of GDP in advanced economies is lower than it has been over the past decades (Fernald, 2014; Gordon, 2014, for example). Antolin-Diaz et al. (2017) observe that standard DFM forecasts quickly revert to the unconditional mean of GDP, so taking account of the variation in long-run GDP growth substantially

[5] Macroeconomic indicators are denoted as Δy_t, as I apply appropriate transformation according to Table 1. Note that some indicators enter the estimation in levels, particularly the surveys, even though I take the first differences in most of the series.

improves point and density GDP forecasts even at very short horizons. Specifically, the time-varying intercept c_t is specified in the following way:

$$c_t = \begin{bmatrix} B & 0 \\ 0 & c \end{bmatrix} \begin{bmatrix} a_t \\ 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

where c_t is composed of two components: the time-varying long-run growth rate a_t and a vector of constants c. Note that the selection vector B only loads onto the first two variables, GDP and consumption. It implies that a_t captures the time-variation in long-run real GDP growth, which is shared by real consumption growth. This specification is based on the permanent income hypothesis: consumers aim to smooth their consumption throughout the lifetime, so they react more to permanent changes in their income than transitory ones. Hence, rather than simply allowing a time-varying intercept only in GDP by setting B = 1, taking GDP and consumption together may help separate growth into long-run and cyclical fluctuations.[6]

Next, the latent factors and idiosyncratic components evolve with the following law of motion:

$$(1 - \phi(L))f_t = \sigma_{\epsilon_t}\epsilon_t \tag{2}$$

$$(1 - \rho_i(L))u_{it} = \sigma_{\eta_{it}} o_{it}\eta_{it}, \quad i = 1, ..., n \tag{3}$$

where $\phi(L)$ and $\rho_i(L)$ are lag polynomials of order p and q, respectively. Following the standard assumption from the literature, the residuals of the latent factor and idiosyncratic components are orthogonal and uncorrelated with each other, i.e. $\epsilon_t \sim iidN(0, 1)$ and $\eta_{it} \sim iidN(0, 1)$. Importantly, each indicator displays outliers o_{it}, which are incorporated in the stochastic volatility (SV) of innovations in a multiplicative way. A more detailed discussion on the specification of o_{it} follows in the next subsection.

Finally, following Primiceri (2005) and others, the time-varying parameters of the model follow driftless random walks:

$$a_t = a_{t-1} + v_{a,t}, \quad v_{a,t} \sim N(0, \omega_a^2) \tag{4}$$

$$log\sigma_{\epsilon_t} = log\sigma_{\epsilon_{t-1}} + v_{\epsilon,t}, \quad v_{\epsilon,t} \sim N(0, \omega_\epsilon^2) \tag{5}$$

[6]While the low-frequency component of growth could be shared by other series, one also faces the risk of misspecification. So I leave any slow-moving components in other indicators to be absorbed by the idiosyncratic components. More discussion on this can be found in Antolin-Diaz et al. (2017).

$$log\sigma_{\eta_{i,t}} = log\sigma_{\eta_{i,t-1}} + v_{\eta_i,t}, \quad v_{\eta_i,t} \sim N(0, \omega^2_{\eta,i}), \quad i = 1, ..., n \tag{6}$$

where σ_{ϵ_t} and $\sigma_{\eta_{i,t}}$ capture the SV of innovations to factors and idiosyncratic components. Modeling the trend as a random walk may sound unrealistic since it means that the movement of real GDP growth could be unbounded. However, the drift takes place only for a finite time, and the variance of this process is small, so they alleviate the problem.[7]

Note that this specification easily nests the DFMs previously proposed in the literature. By shutting down the outliers and loadings on lagged factors, i.e. $o_{it} = 1$ and $\Lambda(L) = \Lambda$, the model shrinks towards Antolin-Diaz et al. (2017). If we further limit the time-variation in the intercepts, $c_t = c$, we recover the model of Marcellino et al. (2016), which features the DFM with SV. Finally, this model coincides with a more commonly used DFM model of Bańbura and Modugno (2014) by imposing dogmatic priors $\omega^2_a = \omega^2_\epsilon = \omega^2_{\eta,i} = 0$, that turn off the SV features.

2.2 Treatment of Outliers

In this paper, I construct an outlier-adjusted stochastic volatility (SV) model within the DFM framework. While a typical SV-DFM model assumes changes in volatility to be highly persistent, as in Eq (6), many macroeconomic time series display large, one-off outliers due to policy changes, strikes, or natural disasters, even in the pre-COVID period.[8] By definition, these extreme observations are more reflective of transitory, rather than permanent, spikes in volatility.

Hence, following Jacquier et al. (2004), I model the outliers in a multiplicative fashion as in the Eq (3). If we set $o_{it} = 1$, it corresponds to the standard SV model where the innovations consist of two components with separate error processes: the conditional volatility, $\sigma_{\eta_{i,t}}$ follows a log-normal auto-regressive model, and the standard normal error. To allow for fat tails or outliers, I further separate the latter into o_{it} and the standard noise $v_{\eta_i,t}$. In the basic SV model, a spike in innovation means the volatility is high, but introducing fat-tails provides an additional source of flexibility: it takes a spike in data by introducing a large o_{it} before increasing $\sigma_{\eta_{i,t}}$. Hence, the intuition is that the basic model results in a more variable sequence of the SV, while this model is able to resist

[7] An alternative strategy would be the model with discrete breaks. However, this is less likely to be robust to misspecification than the reverse case, if the true data generating process is a random walk.

[8] Stock and Watson (2002) already noted the existence of such extreme observations.

outliers and results in more steady dynamics of $\sigma_{\eta_{i,t}}$.

I introduce outliers to the model in two ways. One way is to assume that innovations display fat tails and follow the Student-t distribution, i.e. $\nu_i/o_{i,t} \sim \chi^2_{\nu_i}$, as in Jacquier et al. (2004). The degree of freedom ν_i is jointly estimated with other parameters of the model, and the latent variable o_{it} can be obtained by a scale mixture. While this specification has been adopted more extensively in the literature, Student-t is not the only way to model outliers. The other way is to assume the following uniform mixture distribution for outliers, in the spirit of Stock and Watson (2016):

$$o_{it} = \begin{cases} 1 & \text{with probability} \quad 1 - p_i \\ U(2, 10) & \text{with probability} \quad p_i \end{cases} \tag{7}$$

where the probability of ending up in outlier states, p_i, follows the beta distribution with parameters α, representing the number of occurrences of outliers, and β, that of the normal times. This specification enables a highly persistent volatility state to infrequently and temporarily jump to the outlier states above 1.[9]

This modeling strategy is more advantageous than the student-t approach in at least two aspects. First, it is more geared towards modeling large jumps, like the COVID. Since the t-distribution assigns most probability to values close to one, most of the outliers are of moderate size. It also assigns some mass to values below one; this is not the perfect tool to deal with the COVID period, where the drop in real activity was at least 5 times deeper than any other recession since 1960. In the alternative approach, on the other hand, the outlier states cannot take values below one. There is equal probability for outlier states between 2 and 10, putting relatively more mass on the events that increase volatility by more than twofold: so it is more adequate for modeling large jumps, like COVID.

Moreover, this approach is more intuitive in terms of implementing and integrating priors. In the case of the t-distribution, it is not straightforward to justify the variable-specific prior on the degree of freedom ν_i.[10] However, since the parameter p_i follows the beta distribution in the latter approach, it is easier to implement and justify the priors:

[9] I discretize the distribution of o_{it} using a grid, [1:1:10].

[10] For instance, Antolin-Diaz et al. (2023) impose $\nu_i = 1$ for monthly and 30 for quarterly variables, just to reflect the fact that outliers are observed less in case of the lower frequency. Such priors are not based on exact derivation, so it is hard to extend them to take account of variable-specific rates.

for instance, one can form the prior by counting the number of occurrence of outliers in the pre-sample, variable by variable. I put a prior on p_i based on the stylized fact observed by Stock and Watson (2016) that the outlier occurs once every four years in 10 years of the pre-sample.

Note that despite such differences, both approaches share the same latent state representation: residuals are written as the product of noise and outlier state. Antolin-Diaz et al. (2023) propose an alternative way, where the outliers are present in the additive, rather than multiplicative fashion:

$$\Delta y_t - o_t = c_t + \Lambda(L) f_t + u_t, \quad o_{it} \sim t_{vi}(0, \sigma_{oi}^2) \tag{8}$$

and construct the outlier-adjusted data via the Kalman filter. I show that the multiplicative approach achieves the same goal, while being free from the unwanted dependency of outliers on the AR process of the idiosyncratic components that is present in the additive specification. First, by rearranging, we can rewrite the above equation as:

$$\Delta y_t^* = c_t + \Lambda(L) f_t + u_t, \quad u_t \sim N(0, 1)$$

$$\Delta y_t = \Delta y_t^* + o_t, \quad o_t \sim t_v(0, \sigma_o^2)$$

and this corresponds to the standard regression with measurement errors: since the residuals are the sum of idiosyncratic components u_t and outlier states o_t, one can interpret this as if the true data were sometimes observed with large errors. Then, identifying the measurement error and subtracting from the data is not the only way to remove it. Instead, we can formulate it in another way. The residual variance is sometimes larger due to the outliers, and the aim is to downweight them when estimating parameters: this corresponds to the multiplicative approach.

Meanwhile, the additive specification creates an unwanted dependency issue. After quasi-differencing the Eq (8), we can express the variance of residuals as $\mathrm{V}[(1 - \rho_1 L - \rho_2 L^2)(u_{it} + o_{it})]$[11]. In order to simplify the problem, I add two additional assumptions:

Assumption 1. *The idiosyncratic components u_{it} and outlier states o_{it} are cross-sectionally orthogonal and uncorrelated to each other.*

[11] In this case, I model the idiosyncratic components as the AR(2) process.

Assumption 2. *The outlier states o_{it} are serially uncorrelated.*

which are mostly innocuous by the definition of "one-off" outliers. Then, the variance of residuals corresponds to:

$$V[(1-\rho_1 L-\rho_2 L^2)(u_{it}+o_{it})] = \sigma^2_{\eta_{it}} + (1+\rho_1^2+\rho_2^2)V(o_{it}), \quad V(o_{it}) = \sigma^2_{oi}(\frac{\nu_i}{\nu_i-2})$$

I provide a more detailed derivation in Appendix A.3. Here the role of outlier clearly depends on ρ, the autoregressive coefficients of the idiosyncratic component u_{it}, not the outlier states o_{it}. This is not an intended feature of the model, hence the 'unwanted' dependency. In the likely case of $\rho > 0$, it amplifies the variance of the outlier component, resulting in an underestimation of the SV components in innovations. This is a by-product of the quasi-differencing that only applies to the additive modeling of outliers: the multiplicative approach of this paper is free from such dependency issues.

2.3 Priors and model settings

Based on previous empirical findings, the order of the lag polynomial $\Lambda(L)$ is set to 1, i.e. it loads onto the contemporaneous and the first lag of the factor. Camacho and Perez-Quiros (2010) reported that the survey data were better aligned with a distributed lag of GDP; Antolin-Diaz et al. (2023) found that adding heterogeneous dynamics 'rebalance' the panel by allowing more weight to the "hard" variables, which are otherwise under-weighted compared to the "soft" indicators. The number of lags in the autoregressive coefficients of factors and idiosyncratic components, $\phi(L)$ and $\rho_i(L)$, is set to 2, to allow for medium-term frequency fluctuations as in Stock and Watson (1989).

I adopt a Bayesian approach with informative "Minnesota" style priors (Litterman, 1986) for the coefficients in $\Lambda(L)$, $\phi(L)$ and $\rho_i(L)$.[12] I set the prior mean of $\phi(L)$ to 0.9 for the first lag and zero in other lags, to incorporate a belief that the common, contemporaneous latent factor captures a highly persistent but stationary process. The prior mean of $\Lambda(L)$ is set to the estimate of the standard deviation of each variable for the first lag and zero otherwise, to reflect the belief that the factor corresponds to the cross-sectional average of standardized variables. The prior for $\rho_i(L)$ centers on zero for

[12] These are the most commonly adopted macroeconomic priors for VARs and formalize the view that an independent random-walk model for each variable in the system is a reasonable ground for beliefs about their time series behavior (Sims and Zha, 1998).

all lags, the parsimonious model with no serial correlation. This prior aims to shrink the idiosyncratic components towards an iid measurement error. I impose the variance on the priors to be $\frac{\gamma}{l^2}$ for all of the coefficients: the hyperparameter that governs tightness of the prior, γ is set to 0.2, the standard choice from the VAR literature, and l is the lag order. The idea is to shrink distant lags more strongly than the contemporaneous ones.

To express a preference for the more parsimonious model, I impose priors that shrink variances of the time-varying parameters, $\omega_a^2, \omega_\epsilon^2, \omega_{\eta,i}^2$, close to zero. Specifically, I set an inverse gamma (IG) prior with one degree of freedom for these parameters – just the minimum to obtain proper prior distributions. Regarding the scale parameter of the IG distribution, I set $\omega_a^2, \omega_\epsilon^2$ to 0.001 and $\omega_{\eta,i}^2$ to 0.0001. Following the approach of Primiceri (2005), I incorporate a prior about the time-variation: for example, the value of 0.001 corresponds to the conservative belief that the posterior mean of the long-run growth rate fluctuates with a standard deviation of around 0.4 percentage points in annualized terms over a period of ten years. In the end, the prior specification is conservative enough to shrink the model towards a commonly used DFM model of Bańbura and Modugno (2014) without the time-varying means and the SV, while still being loose enough to let the data speak.

2.4 Mixed frequency and missing data

Macroeconomic indicators are measured at different frequencies. For instance, the real GDP and national accounts variables are usually observed every quarter, while survey data and monthly indicators of real activity are released every month. To efficiently integrate information from the data observed at different frequencies, the standard way is to follow Mariano and Murasawa (2003) which links the *observed* growth rates of quarterly variables y_t^Q to *unobserved* monthly growth y_t^M via the first-order Taylor expansion:

$$y_t^Q = \frac{1}{3}y_t^M + \frac{2}{3}y_{t-1}^M + y_{t-2}^M + \frac{2}{3}y_{t-3}^M + \frac{1}{3}y_{t-4}^M \tag{9}$$

The mixed frequency issue collapses into the problem of missing data, where the model is specified at monthly frequency and we only observe every third observation of y_t^Q. The standard approach is to estimate the latent factors, parameters, and missing data points jointly using the state space representation of the DFM via the Kalman filter. Note that substituting the Eq (1) into (9) yields:

$$y_t^Q = \frac{1}{3}\lambda_y' f_t + \frac{2}{3}\lambda_y' f_{t-1} + \lambda_y' f_{t-2} + \frac{2}{3}\lambda_y' f_{t-3} + \frac{1}{3}\lambda_y' f_{t-4} \tag{10}$$

$$+ \frac{1}{3}u_{yt} + \frac{2}{3}u_{y,t-1} + u_{y,t-2} + \frac{2}{3}u_{y,t-3} + \frac{1}{3}u_{y,t-4} \tag{11}$$

This drastically increases the size of the state vector, which includes 4 lags of f_t and u_t. While the dimension of the state vector is only k $\times$ (p+1) if all indicators are at the monthly frequency (so no approximation is needed), the use of this approximation requires the dimension to be max(p,5)$\times n_Q \times$ k. For instance, in the case of p = 2, k = 1, and n_Q = 4, the size of the state space rises from 3 to 20, more than a six-fold increase. Such an expansion leads to unnecessarily complicated state-space representation and significant rise in computation costs.

To remedy this issue, I specify the model entirely at the monthly frequency by using the interpolated monthly values for quarterly indicators. Specifically, instead of applying the Mariano and Murasawa (2003) approximation to the quarterly variables y_t^Q, we can also implement this for the unobserved quarterly idiosyncratic components, u_t^Q. Then Eq (1) and (3) yield the following state-space representation:

$$u_{i,t}^Q = Hx_{i,t} + \xi_{i,t}, \quad \xi_{i,t} \sim N(0, R) \tag{12}$$

$$x_{i,t} = Ax_{i,t-1} + e_{i,t}, \quad e_{i,t} \sim N(0, Q) \tag{13}$$

with $x_{i,t} = [u_{i,t}^M \; u_{i,t-1}^M \; u_{i,t-2}^M \; u_{i,t-3}^M \; u_{i,t-4}^M]'$, H = $[\frac{1}{3} \; \frac{2}{3} \; 1 \; \frac{2}{3} \; \frac{1}{3}]$, A = $[\rho_1 \; \rho_2 \; 0_{1\times 3}; \; I_4 \; 0_{4\times 1}]$, Q = $[\sigma_{\eta_{it}} o_{it}; \; 0_4]$, and R is a small number. We estimate this system using the Kalman filter at the beginning of the algorithm to obtain the latent variable $\hat{u}_{i,t}^M$. Then, together with $\hat{\lambda}_y$ and $\hat{f}_t$ obtained from the previous iteration, we obtain an *unobserved* monthly-interpolated quarterly value $y_t^M = \hat{\lambda}_y \hat{f}_t + \hat{u}_{i,t}^M$. A more detailed explanation follows in Appendix A.2.

2.5 Estimation algorithm

I estimate the eq (1) – (7) with the Bayesian approach using a hierarchical Gibbs sampler (Moench et al., 2013), to obtain the posterior distribution of parameters and factors.

So whenever possible, I parallelized the steps to reduce the computational costs. Since the idiosyncratic components in the model feature autocorrelation, the state space is rewritten in terms of quasi-differences. The sampling method for the SVs follows Kim et al. (1998), which provide us with faster speed than the exact Metropolis-Hastings algorithm of Jacquier et al. (2004). For the identification of factors, I adopt the steps proposed by Bai and Wang (2015). The full details of the algorithm can be found in Appendix A.2, and below I provide a sketch.

Algorithm. Let $\theta = \{\lambda, \Phi, \rho, \omega_a, \omega_\epsilon, \omega_\eta, \nu\}$ be the underlying parameters of the model, and Φ, ρ represent the autoregressive coefficients for the factor and idiosyncratic components. The latent states to be estimated are $\{a_t, f_t, \sigma_{\epsilon,t}, \sigma_{\eta i,t}, o_{it}\}_{t=1}^T$. The superscript j denotes a current draw. The algorithm consists of the following steps:

1. *Construct monthly-interpolated values for quarterly variables.* For the quarterly variables i = 1...n_Q, compute $\Delta y_{it}^Q - c_{it} - \lambda_i(L)f_t$, given $a_t^{j-1}, f_t^{j-1}, \lambda^{j-1}$ from the previous iteration. Then, using the state space in Eq (12) and (13) and conditional on $\rho^{j-1}, \{\sigma_{\eta i,t}^{j-1}\}_{t=1}^T$, draw $\hat{u}_{i,t}^M$ by employing the Kalman filter and smoother. Then we obtain the monthly-interpolated values, $\Delta y_{it}^{M,Q} = c_{it} + \lambda_i(L)f_t + \hat{u}_{i,t}^M$.

2. *Draw the latent factors and autoregressive coefficients.* Conditional on Δy_{it}^M and the parameters θ^{j-1},

   (a) Draw the factor and the trend, $p(\{a_t^j, f_t^j\}_{t=1}^T | \theta^{j-1}, \{\sigma_{\epsilon,t}^{j-1}, \sigma_{\eta i,t}^{j-1}\}_{t=1}^T, y)$ using the Kalman filter and smoother.The state space is rewritten in terms of quasi-differences at this step, conditional on ρ^{j-1} and $\{\sigma_{\epsilon,t}^{j-1}, \sigma_{\eta i,t}^{j-1}\}_{t=1}^T$.

   (b) Draw the variance of the time-varying GDP growth component from the Inverse-Gamma (IG) posterior, $p(\omega_a^{2,j} | \{a_t^j\}_{t=1}^T)$.

   (c) Given $\{f_t^j\}_{t=1}^T$ from the previous step, draw the autoregressive parameters of the factor VAR, Φ^j from the Normal posterior $p(\Phi^j | \{f_t^j, \{\sigma_{\epsilon,t}^{j-1}\}_{t=1}^T)$. Then draw the SV component of innovation to the factors, $p(\{\sigma_{\epsilon,t}^j\}_{t=1}^T | \Phi, \{f_t\}_{t=1}^T)$, using a mixture of normals following Kim et al. (1998). Draw $\omega_\epsilon^{2,j}$ conditional on $\{\sigma_{\epsilon,t}^j\}_{t=1}^T$ and the IG prior.

3. *Draw the factor loadings and serial correlation coefficients of idiosyncratic components.* For each variable i = 1...n and conditional on $\{f_t^j\}_{t=1}^T$,

(a) Draw the loadings λ^j from $p(\lambda^j|\rho^{j-1}, \{f_t^j, \sigma_{\eta i,t}^{j-1}\}_{t=1}^T, y)$. They can be estimated via GLS, conditional on ρ_i^{j-1} and $\sigma_{\eta i,t}^{j-1}$. Following Bai and Wang (2015), I restrict the loading of GDP on f_t to be unity for the identification.

(b) Draw serial correlation coefficients ρ^j from $p(\rho|\lambda^j, \{f_t^j, \sigma_{\eta i,t}^{j-1}\}_{t=1}^T, y)$ from the Normal posterior, based on the $\{u_{i,t}^j\}_{t=1}^T = y_t - \lambda^j f_t^j$ and the autoregression with heteroskedasticity.

4. *Draw the outlier-adjusted SV of innovations.* For each variable i = 1...n, I obtain the residuals $u_{i,t}^* = (1 - \rho^j(L))u_{i,t}^j$. Then,

(a) (in the case of t-distribution) draw p($\nu^j|u_{i,t}^*, \sigma_{\eta i,t}^{j-1}$) following Jacquier et al. (2004), and also draw p($o_{i,t}^j|\nu^j, u_{i,t}^*, \sigma_{\eta i,t}^{j-1}$), using $\nu^j/o_{i,t}^j \sim \chi_{v+1}^2$.

(b) Obtain mixture states conditional on $(u_{i,t}^*, o_{i,t}^j, \sigma_{\eta i,t}^{j-1})$ in logs, following Kim et al. (1998). Construct the state space with $log\sigma_{\eta_{it}}$ as the latent variable, and apply the Kalman filter and smoother to draw new $\{\sigma_{\eta i,t}^j\}_{t=1}^T$.

(c) (in the case of uniform mixture distribution) conditional on the $\{\sigma_{\eta i,t}^{j-1}, o_{i,t}^{j-1}\}_{t=1}^T$, obtain the mixture states and draw p($o_{i,t}^j|u_{i,t}^*, \sigma_{\eta i,t}^j$). Update p_i^j after calculating the number of outliers based on the cdf of the mixture normals.

(d) Based on the new estimates of the SV, draw p($\omega_{\eta i}^{2,j}|\{\sigma_{\eta i,t}^j\}_{t=1}^T$) from the IG posterior distribution.

5. Repeat the steps until convergence has been reached.

After estimating 7000 draws from the above Gibbs-sampling algorithm, I discard the first 2000 as burn-in draws. The rest 5000 draws of the model parameters and latent variables are used for inference.

3 Implications of outliers: Application to France

In this section, I explore the implications arising from explicit modeling of outliers and conduct a comparative analysis of two distinct approaches by applying these models to the French economy. I begin by describing the datasets closely monitored by market participants. Next, I demonstrate how outlier-augmented models improve the in-sample fit in comparison to the incumbent Dynamic Factor Models (DFMs). Finally, I present the in-sample estimates of various model features, including idiosyncratic volatilities and long-term growth, to facilitate a comparison between the student-t and uniform approaches for incorporating outliers. This paves the way for a detailed examination of the role of outliers in out-of-sample forecasting in the following section.

3.1 Data

Table 1 describes the dataset which consists of 27 variables in total: 11 soft and 16 hard indicators. Among the latter, three are measured at a quarterly frequency, and the rest are monthly indicators. The choice of variables is based on the evidence from the literature rather than a data-driven approach. The three quarterly indicators – output, investment, and total hours worked – are known to be strongly procyclical and key indicators of the business cycle. Investment and hours could be informative for capturing the cyclical movements, as they are sensitive to business cycle fluctuations (Stock and Watson, 1999). As the indicator of consumption, I use the monthly household consumption of manufactured goods, which exclude the durable goods spendings. While it lacks the services consumption compared to the quarterly measure in the national accounts, the monthly data have shorter publication lags that may provide a more timely assessment of not only the economic activity but also the trend, as the permanent income hypothesis suggests that taking the GDP and consumption together may help separate the long-run and component from the growth.

Inspired by Cascaldi-Garcia et al. (2021), I select monthly indicators based on the Bloomberg relevance index, among the large number of candidate series available. This index indicates the percentage of Bloomberg users who set an automatic alert for the release of specific data. Specifically, I take the variables whose relevance index is above 50% and the ones classified as the "Market Moving Indicators". The second column in Table 1 reports this index for each indicator in the dataset.[13] It is well known that market participants closely monitor macroeconomic data releases to extract signals on the current state of the economy, since it heavily affects the performance of their asset portfolios.[14] Hence, I believe this is a reasonable starting point to select the monthly

[13] Since the index does not vary dramatically over the sample, I take the relevance index reported in December 2019, which is the last observation before the COVID period.

[14] In fact, Altavilla et al. (2017) reports that asset prices strongly react to data releases when the outcome differs from the expectations of market participants.

TABLE 1: Indicators Used, France

Type	Variable	Relevance	Start	Transformation
Quarterly	Real GDP	89	Jan-1980	% QoQ
	Real Investment (Gross Fixed Capital Formulation)		Jan-1980	% QoQ
	Total volume of hours worked (employees)		Jan-1980	% QoQ
Survey (11)	BdF survey: change in output, manufacturing industry		Jan-1980	Level
	BdF survey: expected production		Jan-1980	Level
	BdF survey: new orders		May-1981	Level
	BdF survey: sentiment indicator for manufacturing	51	May-1981	Level
	Composite business climate indicator	11	Jan-1980	Level
	Manufacturing: order books and demand		Jan-1980	Level
	Manufacturing: general outlook	(97)	Jan-1980	Level
	Manufacturing: probable trend		Jan-1980	Level
	Services: expected activity		Jan-1991	Level
	Services: expected demand	(77)	Jan-1991	Level
	Consumer confidence	80	Jan-1980	Diff
Consumption	Household consumption: manufactured goods	17	Jan-1980	% MoM
Output (7)	Industrial production	60	Jan-1980	% MoM
	Capacity utilization		Jan-1981	Diff
	Retail sales	55	Jan-1980	% MoM
	Car registration	90	Jan-1980	% MoM
	Building permits		Jan-1994	% MoM
	Industrial production: construction	60	Jan-1990	% MoM
	Turnover Index: manufacturing	11	Jan-1999	% MoM
Labor (3)	Registered unemployment level for France	37	Jan-1980	% MoM
	Active job seekers (A,B,C)		Jan-1996	% MoM
	New vacancies		Jan-1989	% MoM
Trade (2)	Exports: value goods for France	51	Jan-1980	% MoM
	Imports: value goods for France	54	Jan-1980	% MoM

Note: This table provides the list of variables in our dataset. It also provides the following additional details: the Bloomberg relevance index, the beginning of each series, and the transformation applied to the data. All series are publicly available from the FRED and DBnomics website. % QoQ: quarter-on-quarter changes, % MoM: month-on-month changes, Diff: first differences.

indicators.

While in principle I could construct a larger dataset by including broader categories of data and more disaggregated series, Boivin and Ng (2006) found that more data are not always better for factor analysis: a strong correlation in the idiosyncratic components across the disaggregated series within the same category may worsen both the in-sample fit and out-of-sample forecasting performance. Alvarez et al. (2012) reaches a similar conclusion by finding high persistence in either the common factor or the idiosyncratic errors for the large-scale DFMs, concluding that a small-scale dynamic factor model that uses one representative indicator of each category yields better forecasting results.

The choice of monthly indicators in Table 1 reflects the above findings from the literature. The market participants closely monitor variables across many different sectors, such as the labor market (e.g., the unemployment rate), the industrial sector (e.g., the industrial turnover), the construction sector (e.g., the index of production in construction), private consumption (e.g., retail sales and car registrations), and the external sector (e.g., exports and imports of goods). At the same time, most of the hard indicators are the headline indicators for each category: so the economists focus less on the disaggregated data to track the state of the economy. In the end, we end up with a medium-sized panel with representative indicators for each sector.[15] Importantly, I include monthly surveys that represent the perceptions of economic agents about current and future economic prospects from the two major sources: Banque de France and the National Institute of Statistics (INSEE). These "soft" data have a high signal-to-noise ratio and provide timely information to track the economy.

The dataset spans the period from January 1980 to December 2022, including the recent COVID-19 episode. Since I focus on the ability of this model in a traditional now-casting setting, I first utilize only the pre-COVID sample that ends in December 2019 and then employ the full series up to 2022, in the next section. Moreover, I exclude prices and monetary and financial indicators that are often available daily or weekly. Bańbura et al. (2013) have shown that these types of data do not improve the performance of a now-casting model due to their noisy nature. I also abstain from "alternative" data, such as web searches and text-based measures, at this point and leave these avenues for future research.[16]

3.2 In-sample fit

I start by presenting the in-sample fit across models, with the sample ending in 2019. Here I compare six models: (1) the model with t-distributed outliers, (2) the model with outliers following uniform mixture distribution, (3) the model without outliers but the time-varying long-run growth and stochastic volatilities (SV) are still present (Antolin-Diaz et al., 2017), (4) the benchmark DFM model of (Bańbura and Modugno, 2014), and (5–6) the latter two models with the outliers replaced by missing values ('SV miss' and 'Basic miss').[17] Since they correspond to a simple approach to treat outliers, we employ simpler models by moving right across the columns, until reaching the benchmark.

Here I evaluate the models in two aspects: point and density forecasts. I use two

[15]Beyond this criterion, I include two extra indicators, i.e. the capacity utilization and building permits, which may lead the existing headline indicators in the industrial and construction sectors.

[16]Evidences on the ability of alternative data to improve nowcasting the economy is still mixed: for instance, Aaronson et al. (2022) claim that Google Trends leads to superior real-time forecasts of initial UI claims compared to other models. However, Larson and Sinclair (2022) find the opposite result.

[17]Here I define the outliers as the observations five-interquartile range away from the median.

Table 2: In-sample fit, pre-COVID

Measure/model	Full (t-dist)	Full (uniform)	SV miss	SV	Basic miss	Basic DFM
(a) point forecasting						
RMSE	0.254***	0.277***	0.277***	0.277***	0.362	0.362
MAE	0.211***	0.228***	0.228***	0.229***	0.281	0.282
(b) density forecasting						
Log score	−3.688***	−4.716***	−4.984***	−5.029***	-28.020	-28.103
CRPS	0.172***	0.191***	0.192***	0.192***	0.257*	0.258

Note: This table provides the point-forecasting performance of different models: the DFM model with outliers following the student-t and the piecewise-uniform distribution, respectively; the model with stochastic trend and volatility (Antolin-Diaz et al., 2017); and the basic DFM model (Bańbura and Modugno, 2014). 'SV miss' and 'Basic miss' are the latter two models with the outliers, which are defined as the observations five-interquartile range away from the median, replaced by missing values. Sample: 1980.1 – 2019.9. The asterisks are related to the p-value of the null hypothesis that the basic model performs as well as others against the alternative that the other model performs better, based on the test statistic of Diebold and Mariano (1995). *** is significant at the 1% level, ** at the 5% level, and * at the 10% level.

commonly used criteria, the root mean squared error (RMSE) and mean absolute error (MAE) for the former. A lower RMSE or MAE implies that the model generates more accurate point forecasts. Since I adopt the Bayesian approach, density forecasts are also easily obtainable from the DFM models. While there are several measures for density forecast evaluation, I take the one of the most popular metric, the average log score: it assigns a higher value for the model that provides the highest probability to the realizations. I also use another metric, the continuous rank probability score (CRPS), which provides more robustness to outliers and the values that are close but not equal to the realization. In the end, a total of four metrics are used, as shown in the rows of Table 2.

Results from Table 2 suggest that the model with outliers provide better in-sample fit than the basic DFM. All of them reject the Diebold and Mariano (1995) test of the null hypothesis that the performance of the basic and alternative model is comparable, at the 1% significance level. The student-t and uniform approaches result in roughly 30% and 23% reduction in the RMSE, compared to the benchmark model. Moreover, a simple approach to replace outliers by missing values turns out to be not so effective, except some improvement in terms of density forecasting: performance of the 'basic miss' model is not better than the benchmark except for the CRPS, and when the SV process is explicitly modeled, the model with and without replacement display identical performance for all metrics except the log score. Finally, the student-t approach provides the best in-sample fit, with the uniform approach being the second best. While the latter is just as good as the model with outliers but the SV present in terms of point forecasting, it still provides a better performance in density forecasting.

3.3 Implications of modeling choices: in-sample estimates

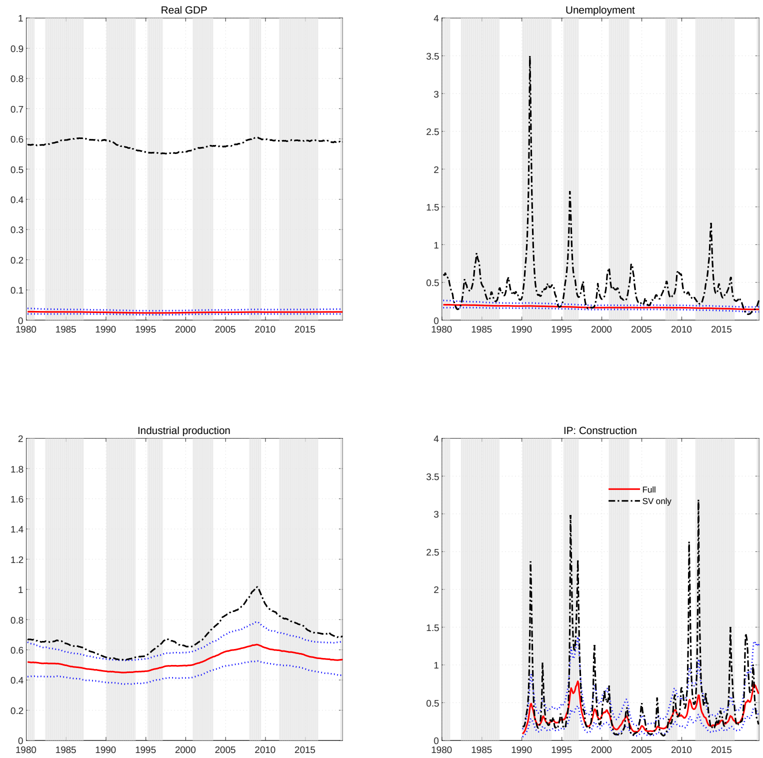
3.3.1 Stochastic Volatility

I begin by displaying the estimates of the SV of innovations to idiosyncratic components, $\sigma_{\eta_{i,t}}$, in Figure 3. Given that the two outlier modeling approaches, and even the standard DFM-SV model without outliers, diverge primarily in this aspect, it serves as an appropriate initial point to gain further insights into the role of outliers and the consequences of modeling choices. Here I select real GDP growth and three monthly indicators that are closely related to the business cycle: industrial production, construction, and unemployment (levels) for France. The full results from all indicators, including quarterly and soft indicators, can be found in the Appendix B.2.

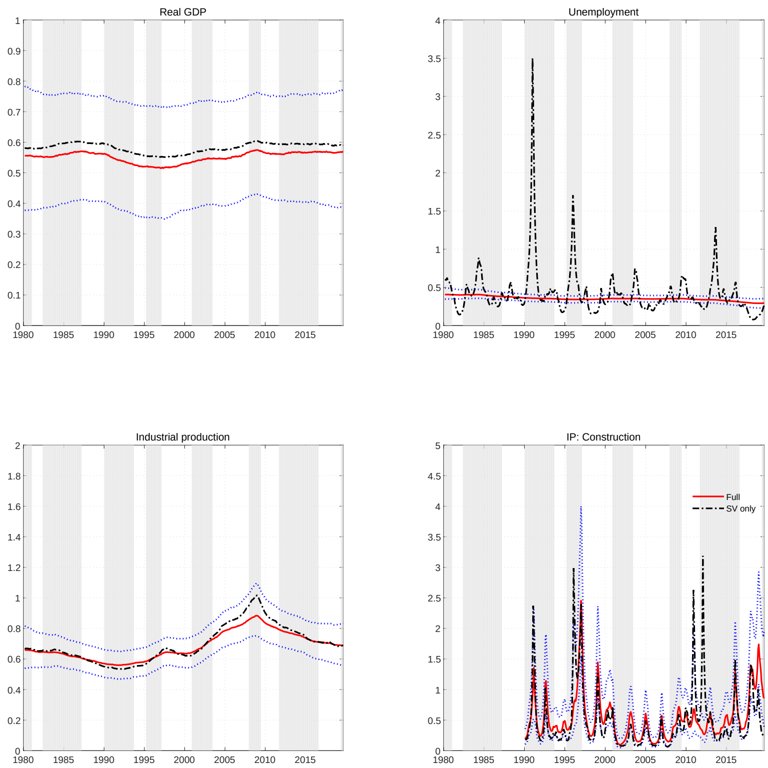
I comment on two observations. First, the t-distributed outliers (top panel) tend to excessively depress the SV process. The differences between the two approaches are clearly visible in Figure 3: the t-distributed outliers lower the level of idiosyncratic volatilities for the all of four indicators, which is not an intended feature of the model. As shown in the Appendix B.2, this phenomenon is not specific to selected variables but all indicators in the dataset, including the quarterly and monthly soft variables. While this approach results in a more persistent SV process, which is consistent with the driftless random walk assumption in Eq (6), the t-distributed outliers flatten out even the smooth series: for instance, the SV process of the industrial production (in black) is already free from strong cyclical variations, but it is unnecessarily suppressed even further. The SV of GDP is almost constant over 40 years, which is still consistent with the assumption in the model but not in line with the established narrative.

On the other hand, those from the uniform mixture distribution (bottom panel) effectively targets only extreme variations. This approach leaves the SV process untouched which it already lacks cyclical variations, resulting in similar dynamics to the baseline model. When the process exhibits strong cyclical variations, however, extreme spikes are removed, as shown in the case of the IP-construction and unemployment. Considering these aspects, the uniform approach seems more advantageous than the widely used

Figure 3: Stochastic volatility: with v. without outliers



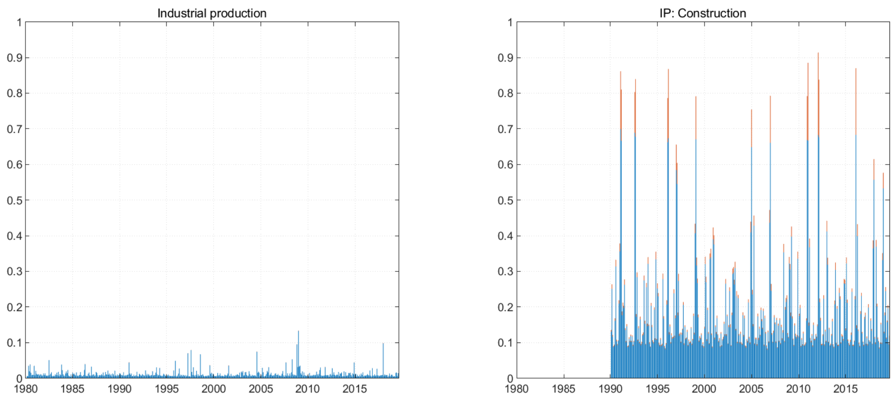
(a) student-t outliers v. no outlier



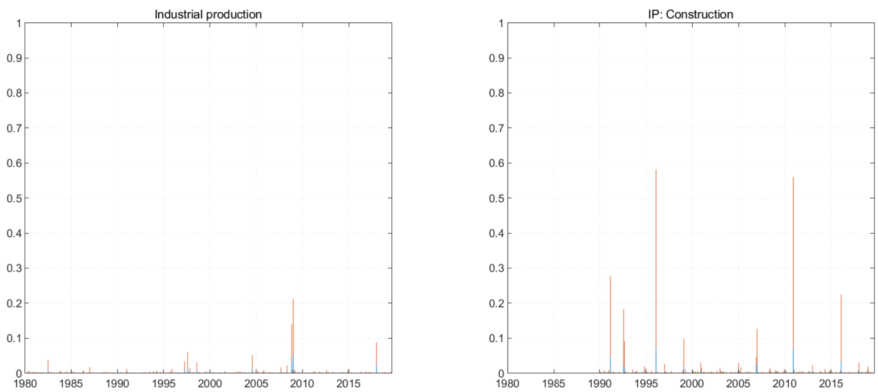
(b) uniform outliers v. no outlier

Note: This plot shows the posterior estimate of stochastic volatility in innovations to the idiosyncratic components. The black line represents the posterior median from the model without outliers. Red and blue lines are the posterior median and 68% bands from the model with outliers following the t-distribution and uniform mixture distribution, respectively. Sample: 1980.1 – 2019.9. OECD recessions in gray.

Figure 4: Posterior probabilities of outlier states



(a) student-t outliers



(b) uniform outliers

Note: The stacked bars represent posterior probabilities for realizations of outlier states that are larger than two. The blue bars correspond to the probability that the outliers are within the range between two and five, and the orange bars denote the probability of outlier states that are larger than five to take place. Sample: 1980.1 – 2019.9.

student-t models in delivering more model-consistent estimates. The Appendix B.2 shows that this also applies to other monthly and quarterly indicators.

To shed more light on the source of such differences, I present the posterior probabilities of outlier states estimated by both the student-t and uniform approaches in Figure 4. These outlier states, drawn from posterior estimates each period, are categorized into three regions: below 2, between 2 and 5, and above 5. These regions correspond to cases of no (or small) outliers, moderate outliers, and large outliers, respectively.[18] The figure directly illustrates the probability of outliers falling within the latter two regions, with the complement representing values below 2 which represent scenarios with no or

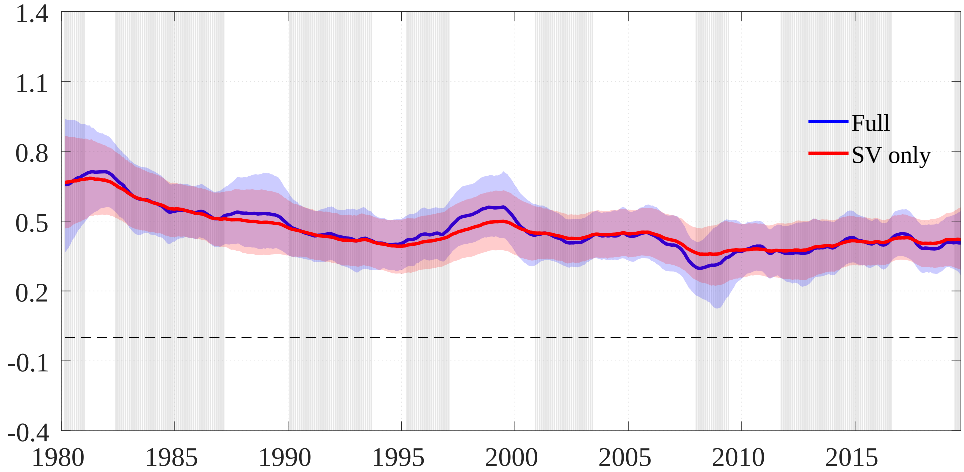
[18] I have chosen a cutoff value of two to facilitate the comparison between the two modeling choices. Note that while the support for outliers is a positive real line in the student-t case, the uniform approach discretizes it between 1 and 20.

minor outliers. For this analysis, I have selected two monthly indicators, namely industrial production and construction, which display relatively smooth and volatile processes, respectively. Additional results for other monthly and quarterly indicators, including posterior estimates of outlier states, can be found in Appendix B.3.

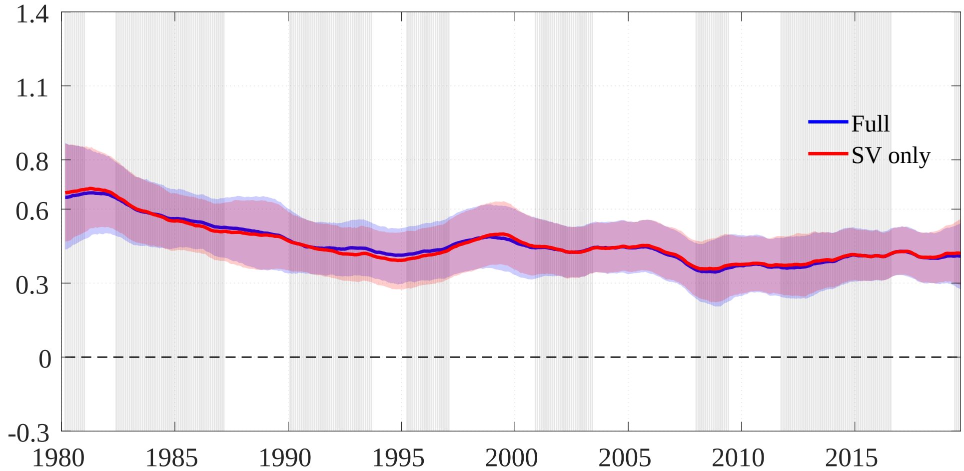
Figure 4 clearly illustrates a difference between the two approaches. First, the student-t approach tends to produce more moderately sized outliers, and they occur with greater regularity. The model consistently generates moderate-sized outliers in every period, even for the industrial production index, which already lacks high-frequency variations in its stochastic volatility process. In the case of the IP-construction index, the model frequently identifies moderate-to-high outliers across time, often ranging from 40 to 90 out of 100 draws. However, when outliers do occur, they are mostly of moderate size. On the contrary, the uniform approach identifies outlier states less frequently, but when they do occur, they are predominantly large in size. It is noticeable that outliers are drawn more frequently for the IP-construction index, which displays frequent spikes and strong cyclical variations. Industrial production, with its slower-moving SV process, features infrequent outliers over the sample, occurring less than 20 times over the entire period of 477 months. Consequently, while the the student-t assumption lowers the level of idiosyncratic volatilities across all indicators, the uniform approach leaves the SV process mostly untouched, targeting only extreme spikes.

These differences between the two approaches clearly stem from the model specifications detailed in Section 2.2. The key disparity lies in the assumed densities for outlier states. While the density of outlier states peaks around the value of 1 for both methods, the student-t concentrates most of the mass on values close to 1. In the case of the uniform approach, there is an equal probability of outlier states between 2 and 10. As a result, the former allocates relatively more mass to values above 1 in total, although these values are centered around 1. The latter, on the other hand, places relatively more mass on the far-right tail. From an empirical standpoint, it translates to the student-t approach characterizing outliers as more moderately sized but occurring more frequently, whereas the uniform approach tends to identify large outlier states. In the context of macroeconomic data, outliers are typically perceived as infrequent occurrences with a substantial impact when they do occur. Therefore, the latter approach aligns more closely with our notion.

Figure 5: Posterior estimate of trend GDP



(a) student-t outliers v. no outlier



(b) uniform outliers v. no outlier

Note: This plot shows the posterior estimate of the long-run growth from the two models, the one with (in blue) and the one without (in red) outliers. The solid lines and shaded areas denote the posterior median and 90% bands, respectively. Sample: 1980.1 – 2019.9. OECD recessions in gray.

3.3.2 Long-run growth

In the previous subsection, we noted that the student-t approach tends to generate moderate-size outliers too frequently due to its concentration of mass around values close to 1. This behavior may lead to unwanted characteristics in volatility estimation. However, is this the only side effect? To address this question, we now explore additional model estimates.

Figure 5 plots posterior estimates of the time-varying component in long-run real GDP growth, a_t, from the model. Here I show the estimates from the two different approaches to model outliers (the student-t and uniform mixture) and compare each of them to the model without outliers. Our prior is that they are likely to be identical: the trend is modeled as driftless random walks as in Eq(4), and the idiosyncratic components lack stochastic volatility by the definition of long-run shift. Since no additional components related to the trend were imposed in the model with outliers, we anticipate that the estimates are unlikely to vary. However, it is interesting to note that the way outliers are modeled results in slight differences in the trend estimates. In the case of the student-t outliers (top panel), trend estimates are less smooth, displaying more high-frequency variations. There is also a spurious IT boom in the late 1990s peaking around 2000, which is true for the U.S.economy but not France. These are less desirable features given the model assumptions and the established narrative in the literature.

On the other hand, when outliers are drawn from the uniform distribution (the bottom panel), the posterior estimates of the long-run growth from the two models, the one with (in blue) and the one without (in red) outliers, are almost identical. The estimated trend captures low-frequency, slow-moving components of growth: it shows the gradual slowdown until 1995, slight rise from 1995 to 2000, and another prolonged decline until the Great Recession. I also plot the posterior estimates of the trend and real GDP growth together in the Appendix B.1. This result is consistent with the narrative of Fernald (2014) that the bulk of the US economy slowdown took place between the turn of the century and the Great recession, as the IT boom faded. Note, however, the size of the IT boom seems much more modest in the case of France, in line with observations from Gordon (2004) that the level of productivity in Europe has been falling behind the US since 1995, due to regulatory barriers in ICT-using industries like wholesale and retail trade and in securities trading.

In Appendix B.1, I also provide posterior estimates of the volatility of the common factor, denoted as f_t, from both models. Despite our assumption that the volatilities of innovations in the factor and idiosyncratic components are independent, the estimates from both models exhibit similar dynamics but show slight differences in levels. These differences, while not statistically significant due to increased estimation uncertainty, are more pronounced in the student-t model compared to the uniform approach, similar to what we observed with long-term growth. In summary, the uniform approach appears to be a more robust choice in terms of estimating parameters that may offer structural insights.

4 Out-of-sample forecasting exercises

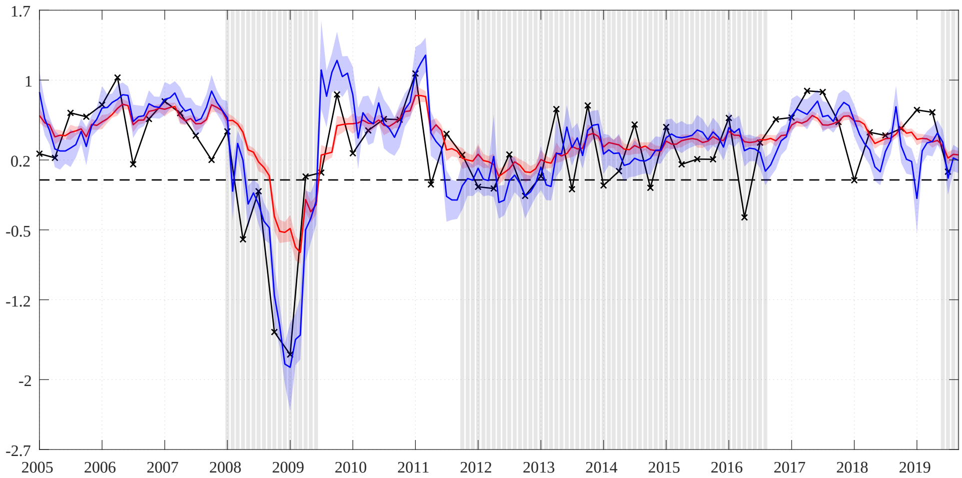
To replicate the situation of policymakers and market participants, I conduct out-of-sample forecasting analyses with the data available at each point in time. Specifically, using data going back to 1980, I produce nowcasts of French GDP growth from 2005 and evaluate the point forecasting performance of the models relative to the benchmark. Since the historical data vintages are not readily available for France, I run the exercise in pseudo-real time, i.e. based on the recent vintages but taking account of the publication delays. I adopt an expanding out-of-sample window: after January 2005, the model is re-estimated when the new data are released. Since repeating the estimation for hundreds of periods takes huge computational costs, I rely on modern cloud computing facilities for this exercise.[19] Then, I run similar exercises with an extended sample up to 2022, which includes the COVID-19 episode.

4.1 Out-of-sample forecasting: pre-COVID

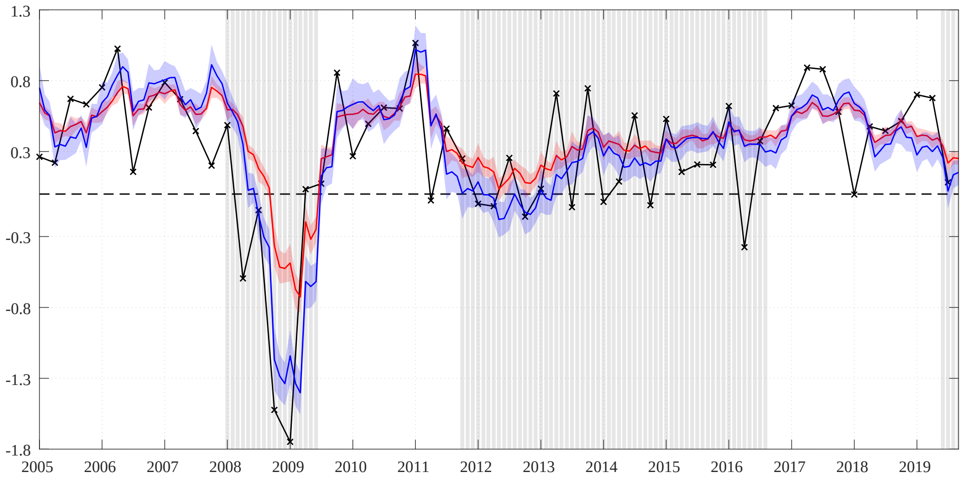
Figure 6 shows the pseudo real-time nowcasts of French GDP from 2005 to 2019, the pre-COVID period. On average, the nowcasts from the model with outliers (in blue) track the economic development over the last 15 years, including the Great Recession and the Euro Area debt crisis. The model with outliers perform better than the benchmark (in red), the basic DFM model, in two aspects. First, they capture the peaks and troughs during the Great recession more timely and accurately, while the benchmark model underestimates the severity of the recession and the rebound afterwards. Second, while the estimates from the benchmark model are overly persistent after 2012 with an upward bias, those from the model with outliers display more cyclical variations. Between two approaches to model the outliers, the student-t outliers are slightly better in capturing the peaks and troughs during the Great recession, while the uniform approach relatively underestimates the fall in 2009Q1 and the subsequent recoveries. The former, however, also has a drawback: it creates spurious cycles at the end of the sample. Overall, explicitly modeling outliers provide a more accurate picture of the current state of the economy, particularly during the extreme events.

[19] It takes around 40 minutes to complete the estimation for one period on a computer with an Intel i5 processor and 8GB of RAM. I thank Stephane Brice and the IT team at PSE for excellent technical support. One can also utilize a more popular alternative channel, e.g. the Amazon Elastic Compute Cloud (EC2) or Microsoft Azure.

Figure 6: Pseudo real-time nowcasts of French GDP, 2005 – 2019



(a) student-t outliers v. benchmark DFM



(b) uniform outliers v. benchmark DFM

Note: This plot shows the realizations and pseudo real-time nowcasts of French GDP from 2005 to 2019 constructed from the two models, the one with outliers(in blue) and the benchmark DFM model (in red). The black line represents the actual realizations of GDP growth, and the blue and red solid lines and shaded areas denote the posterior median and 68% bands, respectively. OECD recessions in gray.

To assess the forecasting ability of models with outliers, I conduct a more formal forecast evaluation in Table 3. As in the Section 3.2, here I compare six models: (1) the model with t-distributed outliers, (2) the model with outliers following uniform mixture distribution, (3) the model without outliers but the time-varying long-run growth and SV are still present (Antolin-Diaz et al., 2017), (4) the benchmark DFM model of (Bańbura and Modugno, 2014), and (5–6) the latter two models with the outliers replaced by missing values ('SV miss' and 'Basic miss'). Models become simpler as we move right across the columns, reaching closer to the benchmark model.

TABLE 3: Out-of-sample performance, pre-COVID

Horizon/model	Full (t-dist)	Full (uniform)	SV miss	SV	Basic miss	Basic DFM
(a) point forecasting: relative RMSE						RMSE
-5 month	1.004	0.948	0.952	0.951	0.996	0.485
-4 month	0.945	0.947	0.944	0.945	0.999	0.461
-3 month	0.954	0.929	0.924	0.925	1.00	0.425
-2 month	0.974	0.874	0.869	0.871	0.997	0.377
-1 month	0.880	0.874	0.871	0.872	1.01	0.348
0 month (end of reference Q)	0.901	0.870	0.873	0.874	1.007	0.341
1 month	0.906	0.868	0.866	0.868	0.994	0.338
(b) density forecasting: relative CRPS						CRPS
-5 month	1.000	0.950	0.954	0.952	0.996	0.268
-4 month	0.941	0.948	0.940	0.943	1.003	0.260
-3 month	0.949	0.924	0.920	0.924	0.997	0.250
-2 month	0.905	0.849	0.845	0.846	1.000	0.236
-1 month	0.802	0.829	0.826	0.827	1.013	0.234
0 month (end of reference Q)	0.808	0.823	0.828	0.832	1.014	0.234
1 month	0.825	0.821	0.820	0.822	0.989	0.233

Note: This table provides the forecasting performance of different models: the DFM model with outliers following the student-t and the piecewise-uniform distribution, respectively; the model with stochastic trend and volatility (Antolin-Diaz et al., 2017), compared to the basic DFM model (Bańbura and Modugno, 2014). 'SV miss' and 'Basic miss' are the latter two models with the outliers, which are defined as the observations five-interquartile range away from the median, replaced by missing values. Using data going back to 1980, I estimate the model at each point in time for 2005.1 – 2019.9 in an expanding window.

I evaluate the models in terms of point and density forecasts, and among many measures available, I use two commonly used metrics for each: the root mean squared error (RMSE) and the continuous ranked probability score (CRPS).[20] Here I present the values in relative terms compared to the benchmark basic DFM model, except the benchmark itself in the last column. The rows in Table 3 correspond to forecasting horizons: I set the last month of the reference quarter as month 0 and start producing forecasts from the first month of the last quarter. In other words, the first three rows are the one-quarter ahead predictions, the next three rows are "nowcasts" of the current quarter, and the last one is the forecasts on GDP of the last quarter, or "backcasts". Since the first release of GDP takes place 40 days after the end of the reference period in France, the backcast horizon covers only one month.

I present three takeaways from the results in Table 3. First, the outlier-augmented models, regardless of the distributional assumptions, outperform the benchmark model in terms of point and density forecasting. The full model with uniform outliers display lower

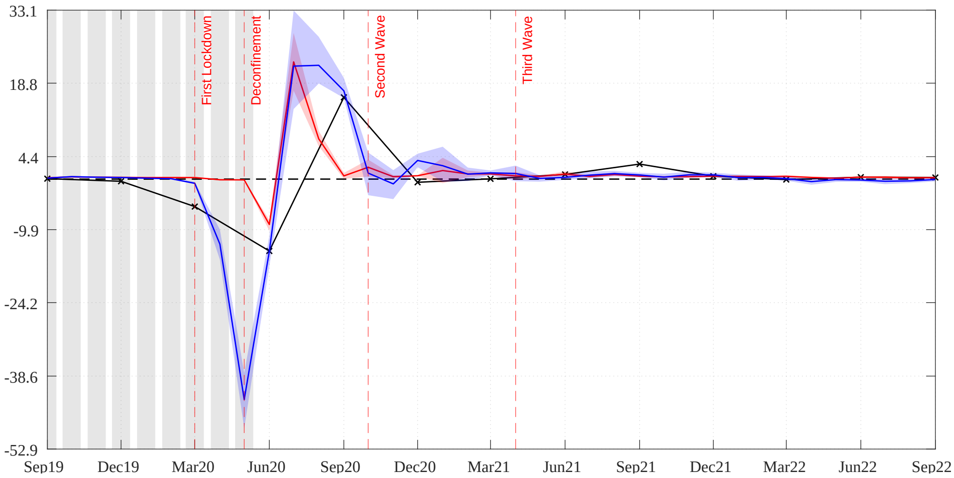
[20] Appendix C.1 reports the results based on all four measures, including the mean absolute forecast error (MAFE) and log score, in absolute terms.

RMSE and CRPS than the benchmark model in all horizons, and the same applies to the student-t approach except the longest horizon. While the 'screening' method, which replaces the pre-defined outliers by missing values, yields negligible impact, the SV model without outliers also dominate the basic DFM. Such results highlight the importance of modeling stochastic volatility in idiosyncratic components: as the sample includes stable expansions (Great Moderation) and sudden depressions (Great Recession), the constant volatility assumption is not enough to track the state of the economy.

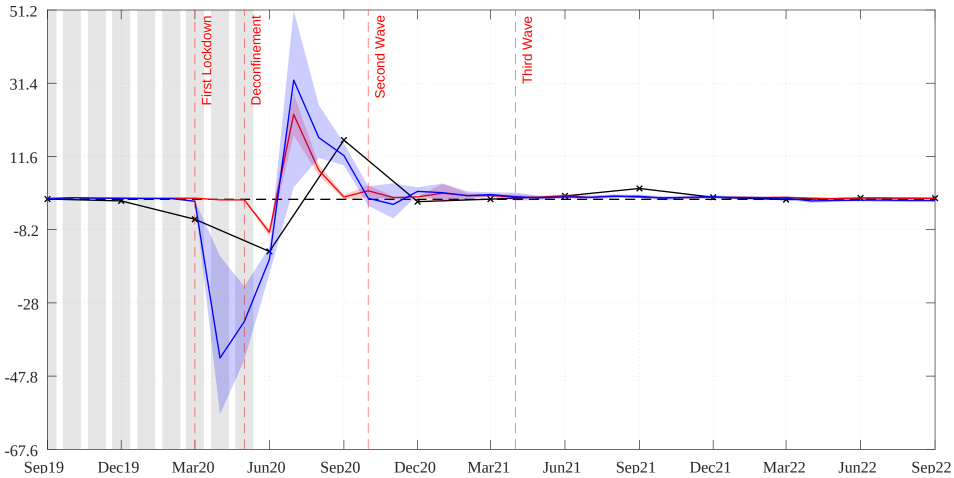
Second, improvements are particularly remarkable at the nowcasting horizons. For instance, at the end of the reference quarter, the SV model with uniform outliers improves on the RMSE of the benchmark specification by 13%, and this is the best model in terms of point forecasts. The student-t approach yields a 20% reduction in CRPS relative to the basic DFM, also outperforming all other modeling choices: the model with outliers outperform the benchmark and other SV specifications in point and density nowcasts. The last column and Appendix C.1 show that all models produce more accurate point forecasts as more information become available: the RMSE steadily decreases as the forecasting horizon approaches the end of the quarter, across all columns.

Finally, the explicit modeling of outliers appears to offer marginally more benefit in terms of density than point forecasts. The relative reduction in CRPS becomes noticeably larger than that of RMSE after the two month ahead horizon. When the accuracy of density forecasts are measured by the log score, as shown in the Appendix C.1, improvements upon the benchmark are also statistically significant at 5–10% levels in forecasting horizons, and 1% level in nowcasting and backcasting horizons.[21] The outlier-augmented models also yield consistently better density forecasts than other SV specifications in all horizons. Interestingly, the uniform approach outperforms the student-t method for point forecasting, but the situation reverses when it comes to density forecasting. Overall, the results highlight a strong forecasting performance of Bayesian DFM models with outliers compared to the benchmark in terms of point and density forecasts for the French GDP.

Figure 7: Pseudo real-time nowcasts of French GDP, 2020 – 22



(a) student-t outliers v. benchmark DFM



(b) uniform outliers v. benchmark DFM

Note: This plot shows the realizations and pseudo real-time nowcasts of French GDP for 2019.9 – 2022.9 constructed from the two models, the one with outliers(in blue) and the benchmark DFM model (in red). The black line represents the actual realizations of GDP growth, and the blue and red solid lines and shaded areas denote the posterior median and 68% bands, respectively. OECD recessions in gray.

4.2 Out-of-sample forecasting: post-COVID

I move on to conduct similar pseudo real-time exercises with an extended sample up to September 2022, which includes the COVID episode. Figure 7 shows the pseudo-real time nowcasts of French GDP from the third quarter of 2019 to the same quarter of 2022 with the ex-post realizations (in black) and the major events, e.g. the first nationwide

[21] The statistical significance is based on the Diebold and Mariano (1995) test. However, Diebold (2015) emphasizes that this test is for comparing forecasts, not models, and the optimal model comparison is based on full-sample residuals, not out-of-sample forecast errors. For this reason, I present the test results only in the Appendix, just for comparative predictive performance.

Table 4: Out-of-sample performance, full sample

Horizon/model	Full (t-dist)	Full (uniform)	SV miss	SV	Basic miss	Basic DFM
(a) point forecasting: relative RMSE						RMSE
-5 month	1.187	1.822	1.221	1.225	1.081	2.744
-4 month	2.502	1.609	2.171	2.372	1.020	2.685
-3 month	0.739	0.776	0.964	0.959	0.965	2.789
-2 month	0.658	1.786	1.126	0.762	1.201	2.053
-1 month	1.769	1.177	1.461	1.279	1.328	2.078
0 month (end of reference Q)	0.422	0.458	0.914	0.454	1.101	2.096
1 month	0.389	0.763	1.136	0.616	1.213	2.112
(b) density forecasting: relative CRPS						CRPS
-5 month	1.069	1.193	1.136	1.133	1.008	0.848
-4 month	1.625	1.206	1.495	1.524	0.974	0.802
-3 month	0.884	0.842	0.990	0.964	1.040	0.747
-2 month	0.675	1.120	1.062	0.813	1.141	0.631
-1 month	1.147	0.842	1.183	0.916	1.199	0.644
0 month (end of reference Q)	0.566	0.605	0.795	0.622	1.132	0.632
1 month	0.500	0.637	0.919	0.694	1.130	0.683

Note: This table provides the forecasting performance of different models: the DFM model with outliers following the student-t and the piecewise-uniform distribution, respectively; the model with stochastic trend and volatility (Antolin-Diaz et al., 2017), compare to the basic DFM model (Bańbura and Modugno, 2014). 'SV miss' and 'Basic miss' are the latter two models with the outliers, which are defined as the observations five-interquartile range away from the median, replaced by missing values. Using data going back to 1980, I estimate the model at each point in time for 2005.1 – 2022.9 in an expanding window.

lockdown and subsequent waves. It is clearly visible that the model with outliers (in blue) better tracks the actual dynamics during the year 2020. Nowcasts from the benchmark model (in red) decline only in June, and the degree of the drop and recovery is modest in both cases. The model with student-t or uniform outliers expects a large fall since April 2020, which is a sensible timing given the publication delay. They assign troughs in April and May 2020, respectively, and correctly capture the realization at the second quarter of 2020. The timing of rebound is also consistent with the actual monthly data during this period. Moreover, while the benchmark model reverts to the trend too early, the outlier-augmented models adjust the timing of slowdown to have the actual growth at the third quarter in the ballpark.

In conjunction with the preceding section, I assess the precision of point and density forecasts across various models in comparison to the benchmark in Table 4. Just as in the pre-COVID case, improvements are particularly noticeable in the nowcasting horizon for both point and density forecasting. At the end of the reference quarter, it is evident that the SV model with t-distributed outliers outperforms the benchmark and other SV specifications in point and density nowcasts. In the domain of density forecasting, all models incorporating outlier adjustments demonstrate superior forecasting accuracy compared to other models.

Interestingly, the results indicate that the advantages of outlier-augmented models in nowcasting become even more pronounced when the post-COVID sample is included. For instance, the SV model with uniform outliers improves the RMSE of the benchmark specification by 44%, which is a more than threefold increase compared to the 13% improvement observed in the pre-COVID case. The difference in CRPS between the student-t approach and the basic DFM amounts to 0.43, whereas it was only 0.2 in the pre-COVID samples. The improvements over other SV specifications have also expanded. During the pandemic, the outlier-augmented models clearly distinguish between short-term and long-lasting increases in uncertainty, an advantage over other models. In contrast to the pre-COVID case, the student-t method outperforms the uniform approach for both point and density forecasting. However, the uniform approach still demonstrates comparable capability, especially in terms of density forecasting. The ability of outlier-augmented models to capture extreme events appears particularly advantageous during uncertain times.

In contrast to the pre-COVID case presented in Table 3, models with outliers no longer possess superiority over the benchmark in longer forecasting horizons. This outcome is not surprising, given that the emergence of COVID-19 represented a wholly unexpected event from an economic perspective. Still, when considering the log score as a measure of density forecasting accuracy, as illustrated in Appendix C.1, models incorporating outliers, regardless of their distributional assumptions, consistently generate more precise density forecasts compared to the standard model across all forecasting horizons. In summary, the findings suggest that explicitly modeling outliers significantly enhances forecasting performance, particularly in nowcasting horizons and in terms of density forecasting for both pre and post-COVID datasets. Consequently, it appears that outlier-augmented models are better suited to predict economic developments during the early phase of the COVID crisis. I will further examine this result in the following section.

5 A case study: U or V shaped recovery?

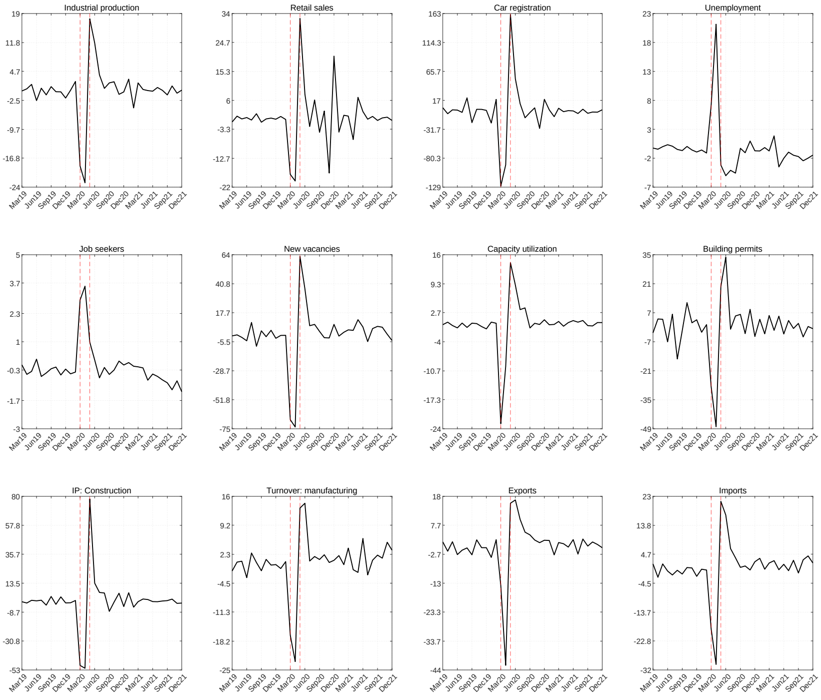
The economic turbulence after the pandemic posed unprecedented challenges to the estimation of macroeconometric models. In this section, I assess the role of models featuring outliers in stochastic volatilites, in terms of forecasting during this exceptional period: specifically, I explore the situation in May 2020 as a case study to test the ability of the model to capture the shape of recovery after COVID early on, revisiting the debate on "U v. V-shaped" recovery.

Going back in time to around the breakout of COVID, one of the most important economic debates was: what will the recovery look like? Even though there are more pessimistic views such as the L-shaped recovery, the two most popular scenarios in Wall Street were the U or V-shaped recoveries. The former view argues that even if we assume that COVID-19 has reached its peak, the easing of lockdown will happen only gradually, and people will keep social distancing for some time even after the lockdown ends. Hence, the recovery will take place slowly over time, i.e. in a 'U-shape', similar to the oil shock episodes in the US economy in the 1970s. The latter takes the view that while the COVID led to an unprecedented drop in economic activity, the size of fiscal and monetary stimulus was also the largest in history, so the policy easing could aid a swift 'V-shape' rebound, like the US recession in the early 1950s.

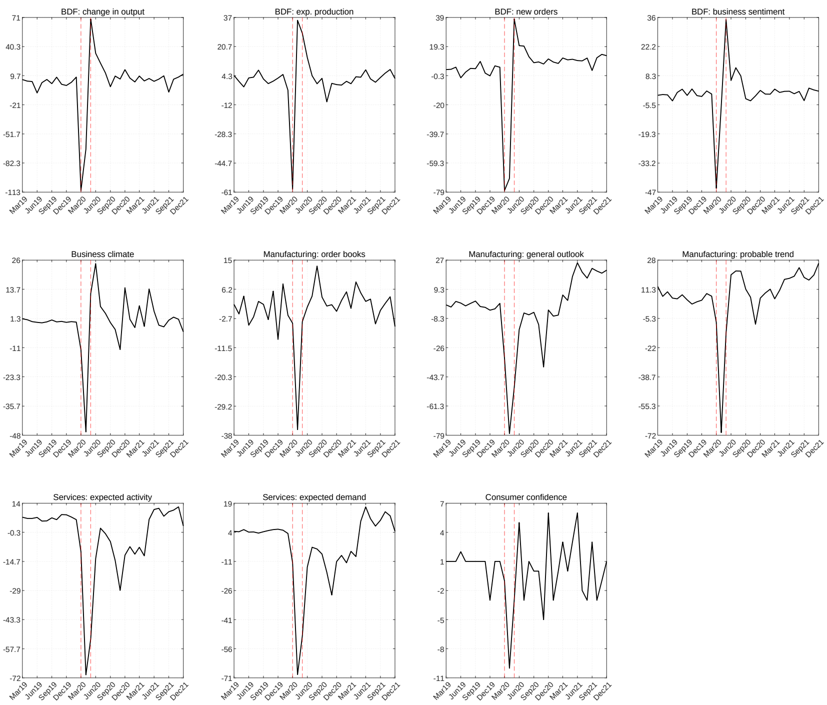
While we now know that the recovery was closed to the V-shape, the major opinion was actually the opposite in May 2020. For instance, here are some headlines from the Wall Street Journal: "Why the Economic Recovery Will be More of a 'Swoosh' Than V-Shaped" (May 11, 2020); "Corporate America Isn't Betting On a V-Shaped Recovery" (May 13, 2020). To understand a sudden shift in experts' opinions, I plot ex-post realizations of macroeconomic indicators in 2019 – 2021 in Figure 8. Looking at the top panel of the figure, which displays 12 monthly 'hard' indicators from the dataset, we can find a pattern consistent across variables: all of them experience an unprecedented drop in March 2020 and then an even larger rebound in May, as shown by the red dashed lines. Recalling that the hard indicators are usually published with one to two months of lag, at the time of May 2020, the market experts observed either the historical drop in March or an even further fall in April 2020. So it is not a surprise that the majority retained the more pessimistic view, the V-shaped recovery.

At the same time, however, more timely data are also available to experts: the surveys,

Figure 8: Ex-post realizations of macroeconomic indicators, 2019 – 2021



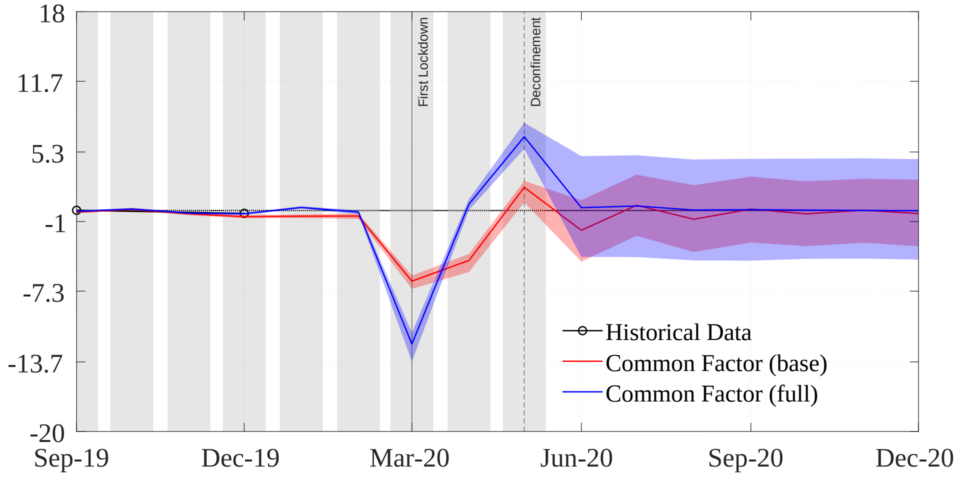
(a) Hard indicators



(b) Soft indicators

Note: This plot shows ex-post realizations of macroeconomic indicators in the dataset for 2019 – 2021. The top and bottom panels plot hard and soft indicators, respectively. The red dotted lines represent March and May 2020 for all variables.

Figure 9: Estimates of the economic activity, May 1 2020



Note: This plot shows pseudo real-time estimates of current economic activity for 2019 – 2020, based on the vintage available on May 1, 2020, constructed from the two models, the one with (in blue) and without (in red) outliers. The black line represents the actual realizations of GDP growth until 2019Q4, the last observation available at that time. The blue and red solid lines and shaded areas denote the posterior median and 68% bands, respectively. The vertical bars represent the first lockdown and deconfinement in France. OECD recessions in gray.

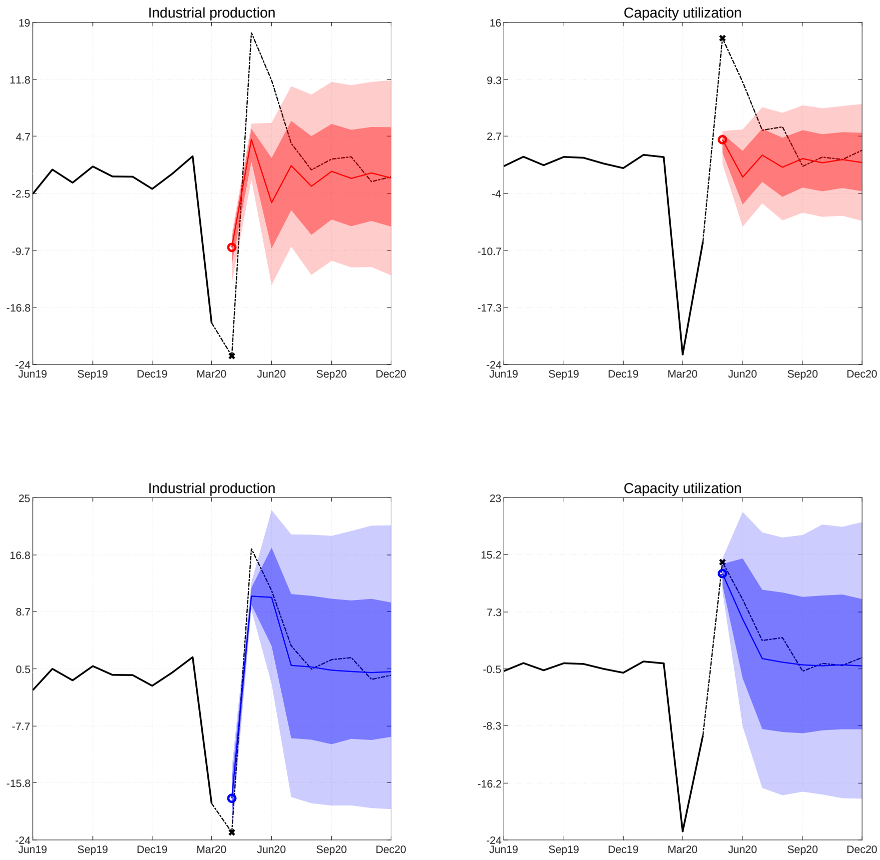
or 'soft' data, are published roughly at the end of the reference month or even earlier. The bottom panel of 8 shows 11 soft indicators from the dataset. Interestingly, we observe the same pattern as the hard indicators: a large decrease in March and a corresponding rebound in May for the majority of them, in particular the surveys from the Banque de France.[22] Then, while forecasts based on only qualitative data tend to be less reliable, it is proven that the DFMs have the ability to summarize common information across data with different publication lags and frequencies: so given some signs of recovery from the April survey data, markets could have learned that the V-shaped recovery is more likely in May. However, in reality, the experts only began to expect this in June.[23] Can this model beat the experts and expect the right shape of recovery on time? This makes nowcasting the economic activity in May 2020 an interesting case study.

Figure 9 shows the estimates of current economic activity, f_t, until the end of 2020 from the out-of-sample analysis based on the vintage available on May 1, 2020. Here I compare the forecasts from two models: the full model with t-outliers (in blue) and

[22] Actually, while the surveys from the Banque de France share almost the same pattern as the hard indicators, the INSEE surveys seem to be lagging by one month, i.e. drop in April and rebound in June. This is due to the different publication schedules between two institutions within a month: while BdF releases its surveys at the end of the month, INSEE publishes earlier, usually during the first week. I appreciate INSEE for pointing it out after the discussion on this issue.

[23] The headline of the Wall Street Journal changes to: "Signs of a V-Shaped Early-Stage Economic Recovery Emerge (June 13, 2022)".

Figure 10: forecasts with and without fat tails (May 1, 2020)



Note: This figure compares predictions on French industrial production and capacity utilization, generated by models with (in blue) and without (in red) the idiosyncratic Student-t component, based on the vintage available on May 1, 2020. Solid black lines represent observations available up to date, and the stars and dotted black lines are ex-post next period and future realizations, respectively. Shades denote the 68% and 90% posterior credible intervals.

the SV model without outliers (in red). The plot shows that our model better captures the stylized features across the data: an unprecedented drop in March but a rebound in May 2020, hence the 'V-shaped' recovery. The model without outliers instead expects a shallower drop and rebound with fluctuations afterwards, more in line with the 'U-shaped' story. Note that at that time, due to publication lags, GDP data were only available up to 2019Q4. It implies that modeling outliers helps produce better forecasts during the early phase of the COVID by more effectively extracting signs of early recovery from the April survey data than the benchmark model. Our model delivers the right shape of recovery at least one month ahead of the market experts.

To further examine the role of modeling outliers, I present pseudo real-time estimates of monthly hard indicators for 2019 – 2020 in Figure 10. I select two measures representing real economic activities, the industrial production index (IPI) and capacity utilization.

While they are closely related, they differ in terms of the publication delay: the former is released two months later, but the latter is only delayed by one month. So on May 1st, 2020, the last observation (the black solid line) ends in March and April, respectively: it implies that while the IPI only informs the model about the historical drop, the capacity utilization also signals some signs of recovery. Figure 10 illustrates the importance of modeling outliers: it interprets a large decrease in May as the outlier and, by taking account of its transitory nature, expects a strong rebound in May. Then the forecasts (in blue) quickly revert to the trend, with no significant swing afterwards. However, when we turn off the outliers, the model misses extreme dynamics and suggests a more volatile path around the trend until the end of the year. In sum, the model with outliers improves forecasting the early economic developments after the breakout of COVID, in particular the 'V-shaped' recovery early on.

6 Conclusion

This paper examines the modeling of dynamic factor models in the presence of one-off outliers in the data. While it is of particular interest during unprecedented times like the COVID-19 pandemic, the occurrence of outliers is a common feature in macroeconomic data. Moving beyond conventional data screening practices, this study introduces an explicit modeling of outliers within dynamic factor models, which helps mitigate unintended features present in comparable approaches. The methodological contributions of this paper are twofold: firstly, it introduced the incorporation of fat tails and outliers multiplicatively into innovation volatility. Secondly, it explicitly modeled outliers using two distinct approaches including the student-t and uniform mixture distributed outliers.

The results of applying these models to forecast the French economy before and after the pandemic have yielded several key findings. First, the outlier-augmented models consistently outperformed the benchmark model, regardless of distributional assumptions, in point and density forecasting. These enhancements were most pronounced in nowcasting horizons, where models with outliers demonstrated superior performance, capturing economic activity more accurately and in a more timely manner. Secondly, the significance of incorporating outliers became more apparent during major crises, such as the Great Recession and the COVID-19. Outlier-augmented models distinctly differentiated between short-term and long-lasting spikes in uncertainty, effectively capturing economic peaks and troughs with greater accuracy and timeliness. Finally, when comparing two distinct methods to incorporate outliers, the uniform approach is possibly a more advantageous choice over the commonly employed student-t models. The uniform approach effectively targeted extreme variations without disrupting the smoothness of the stochastic volatility process, while the t-distributed outliers tend to overly suppress even the already smooth process.

I also demonstrate the value of incorporating outliers in forecasting the early stages of economic recovery following the COVID-19 outbreak. By effectively discerning early signs of recovery, the outlier-augmented model offers a more accurate depiction of economic developments, particularly in distinguishing between 'V-shaped' and 'U-shaped' recovery narratives. This ability to anticipate recovery dynamics one month ahead of market experts has important implications for policymakers and market participants.

In light of these findings, this research enriches the field of macroeconomic modeling by addressing the issue of outliers in economic data, enhancing the accuracy of forecasting the economic activity, in particular during times of significant crisis. It provides a strong analytical foundation for the treatment of outliers, opening new possibilities for improved economic modeling and forecasting. While the choice between modeling approaches should depend on specific research objectives and the context of the analysis, this work contributes to the ongoing dialogue on enhancing the robustness and adaptability of macroeconomic models in the face of exceptional economic events.

References

Aaronson, Daniel, Scott A. Brave, R. Andrew Butters, Michael Fogarty, Daniel W. Sacks, and Boyoung Seo, "Forecasting unemployment insurance claims in realtime with Google Trends," *International Journal of Forecasting*, 2022, *38* (2), 567–581.

Altavilla, Carlo, Domenico Giannone, and Michele Modugno, "Low frequency effects of macroeconomic news on government bond yields," *Journal of Monetary Economics*, 2017, *92* (C), 31–46.

Alvarez, Rocio, Maximo Camacho, and Gabriel Perez-Quiros, "Finite sample performance of small versus large scale dynamic factor models," Working Papers 1204, Banco de España February 2012.

Antolin-Diaz, Juan, Thomas Drechsel, and Ivan Petrella, "Tracking the Slowdown in Long-Run GDP Growth," *The Review of Economics and Statistics*, May 2017, *99* (2), 343–356.

_ , _ , and _ , "Advances in Nowcasting Economic Activity: The Role of Heterogeneous Dynamics and Fat Tails," CEPR Discussion Papers 17800, C.E.P.R. Discussion Papers January 2023.

Aruoba, S. Boragan, Francis X. Diebold, and Chiara Scotti, "Real-Time Measurement of Business Conditions," *Journal of Business & Economic Statistics*, 2009, *27* (4), 417–427.

Bai, Jushan and Peng Wang, "Identification and Bayesian Estimation of Dynamic Factor Models," *Journal of Business & Economic Statistics*, April 2015, *33* (2), 221–240.

Bańbura, Marta and Michele Modugno, "Maximum Likelihood Estimation Of Factor Models On Datasets With Arbitrary Pattern Of Missing Data," *Journal of Applied Econometrics*, January 2014, *29* (1), 133–160.

_ , Domenico Giannone, and Lucrezia Reichlin, "Nowcasting," Working Paper Series 1275, European Central Bank December 2010.

_ , _ , Michele Modugno, and Lucrezia Reichlin, "Now-casting and the real-time data flow," Working Paper Series 1564, European Central Bank July 2013.

Boivin, Jean and Serena Ng, "Are more data always better for factor analysis?," *Journal of Econometrics*, May 2006, *132* (1), 169–194.

Camacho, Maximo and Gabriel Perez-Quiros, "Introducing the euro-sting: Short-term indicator of euro area growth," *Journal of Applied Econometrics*, 2010, *25* (4), 663–694.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino, "Nowcasting tail risk to economic activity at a weekly frequency," *Journal of Applied Econometrics*, August 2022, *37* (5), 843–866.

Cascaldi-Garcia, Danilo, "Pandemic Priors," International Finance Discussion Papers 1352, Board of Governors of the Federal Reserve System (U.S.) August 2022.

_ , Thiago Revil T. Ferreira, Domenico Giannone, and Michele Modugno, "Back to the Present: Learning about the Euro Area through a Now-casting Model," International Finance Discussion Papers 1313, Board of Governors of the Federal Reserve System (U.S.) March 2021.

Chan, Joshua C.C., "Comparing stochastic volatility specifications for large Bayesian VARs," *Journal of Econometrics*, 2023, *235* (2), 1419–1446.

Cogley, Timothy and Thomas J. Sargent, "Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S," *Review of Economic Dynamics*, April 2005, *8* (2), 262–302.

Diebold, Francis X., "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests," *Journal of Business & Economic Statistics*, January 2015, *33* (1), 1–1.

— , “Real-Time Real Economic Activity: Exiting the Great Recession and Entering the Pandemic Recession,” NBER Working Papers 27482, National Bureau of Economic Research, Inc July 2020.

Diebold, Francis X and Roberto S Mariano, “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, July 1995, *13* (3), 253–263.

Doz, Catherine, Laurent Ferrara, and Pierre-Alain Pionnier, “Business cycle dynamics after the Great Recession: An Extended Markov-Switching Dynamic Factor Model,” PSE Working Papers halshs-02443364, HAL January 2020.

Evans, Martin D. D., “Where Are We Now? Real-Time Estimates of the Macroeconomy,” *International Journal of Central Banking*, September 2005, *1* (2).

Fernald, John G., “Productivity and Potential Output before, during, and after the Great Recession,” in “NBER Macroeconomics Annual 2014, Volume 29” NBER Chapters, National Bureau of Economic Research, Inc, January 2014, pp. 1–51.

Fernández-Villaverde, Jesús, Pablo Guerrón-Quintana, Keith Kuester, and Juan Rubio-Ramírez, “Fiscal Volatility Shocks and Economic Activity,” *American Economic Review*, November 2015, *105* (11), 3352–3384.

Giannone, Domenico, Lucrezia Reichlin, and David Small, “Nowcasting: The real-time informational content of macroeconomic data,” *Journal of Monetary Economics*, May 2008, *55* (4), 665–676.

Gordon, Robert J., “Why was Europe Left at the Station When America’s Productivity Locomotive Departed?,” NBER Working Papers 10661, National Bureau of Economic Research, Inc August 2004.

— , “The Demise of U.S. Economic Growth: Restatement, Rebuttal, and Reflections,” NBER Working Papers 19895, National Bureau of Economic Research, Inc February 2014.

Jacquier, Eric, Nicholas G. Polson, and P.E.Peter E. Rossi, “Bayesian analysis of stochastic volatility models with fat-tails and correlated errors,” *Journal of Econometrics*, September 2004, *122* (1), 185–212.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib, “Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models,” *Review of Economic Studies*, 1998, *65* (3), 361–393.

Larson, William D. and Tara M. Sinclair, “Nowcasting unemployment insurance claims in the time of COVID-19,” *International Journal of Forecasting*, 2022, *38* (2), 635–647.

Lenza, Michele and Giorgio E. Primiceri, “How to estimate a VAR after March 2020,” Working Paper Series 2461, European Central Bank August 2020.

Lewis, Daniel J., Karel Mertens, James H. Stock, and Mihir Trivedi, “Measuring real activity using a weekly economic index,” *Journal of Applied Econometrics*, June 2022, *37* (4), 667–687.

Litterman, Robert B, “Forecasting with Bayesian Vector Autoregressions-Five Years of Experience,” *Journal of Business & Economic Statistics*, January 1986, *4* (1), 25–38.

Marcellino, Massimiliano, Mario Porqueddu, and Fabrizio Venditti, “Short-Term GDP Forecasting With a Mixed-Frequency Dynamic Factor Model With Stochastic Volatility,” *Journal of Business & Economic Statistics*, January 2016, *34* (1), 118–127.

— , Todd Clark, Andrea Carriero, and Elmar Mertens, “Addressing COVID-19 Outliers in BVARs with Stochastic Volatility,” CEPR Discussion Papers 15964, C.E.P.R. Discussion Papers March 2021.

Mariano, Roberto S. and Yasutomo Murasawa, “A new coincident index of business cycles based on monthly and quarterly series,” *Journal of Applied Econometrics*, 2003, *18* (4), 427–443.

Moench, Emanuel, Serena Ng, and Simon Potter, "Dynamic Hierarchical Factor Model," *The Review of Economics and Statistics*, December 2013, *95* (5), 1811–1817.

Negro, Marco Del and Christopher Otrok, "Dynamic factor models with time-varying parameters: measuring changes in international business cycles," Technical Report 2008.

Ng, Serena, "Modeling Macroeconomic Variations after Covid-19," NBER Working Papers 29060, National Bureau of Economic Research, Inc July 2021.

Primiceri, Giorgio E., "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 2005, *72* (3), 821–852.

Schorfheide, Frank and Dongho Song, "Real-Time Forecasting with a (Standard) Mixed-Frequency VAR During a Pandemic," NBER Working Papers 29535, National Bureau of Economic Research, Inc December 2021.

Sims, Christopher A and Tao Zha, "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, November 1998, *39* (4), 949–968.

Stock, James H. and Mark W. Watson, "New Indexes of Coincident and Leading Economic Indicators," in "NBER Macroeconomics Annual 1989, Volume 4" NBER Chapters, National Bureau of Economic Research, Inc, January 1989, pp. 351–409.

_ and _ , "Business cycle fluctuations in us macroeconomic time series," in J. B. Taylor and M. Woodford, eds., *Handbook of Macroeconomics*, Vol. 1 of *Handbook of Macroeconomics*, Elsevier, 1999, chapter 1, pp. 3–64.

Stock, James H and Mark W Watson, "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics*, April 2002, *20* (2), 147–162.

Stock, James H. and Mark W. Watson, "Core Inflation and Trend Inflation," *The Review of Economics and Statistics*, October 2016, *98* (4), 770–784.

A Details on the estimation of the model

A.1 The state space system

Let y_t denote n $\times$ 1 vector of macroeconomic indicators. While it includes n_Q quarterly and n_M monthly variables, I employ the interpolated monthly values $y_t^{Q,M}$ for the quarterly indicators as described in Section 2.4. So I specify in the model that the time index t is always monthly for all variables. The number of lags in the autoregressive coefficients of factors and idiosyncratic components is set to 2, i.e. p = q = 2. Since the idiosyncratic components in the model feature autocorrelation, the state space is rewritten in terms of quasi-differences:

$$\tilde{y}_t = \tilde{\Lambda} F_t + \tilde{\eta}_t, \quad \tilde{\eta}_t \sim N(0, R_t)$$
$$F_t = A F_{t-1} + e_t, \quad e_t \sim N(0, Q_t)$$

where the observables are defined as:

$$\tilde{y}_t = \begin{bmatrix} y_{1,t}^{Q,M} - \rho_{1,1} y_{1,t-1}^{Q,M} - \rho_{1,2} y_{1,t-2}^{Q,M} \\ \vdots \\ y_{n_Q,t}^{Q,M} - \rho_{n_Q,1} y_{n_Q,t-1}^{Q,M} - \rho_{n_Q,2} y_{n_Q,t-2}^{Q,M} \\ y_{1,t}^{M} - \rho_{n_Q+1,1} y_{1,t-1}^{M} - \rho_{n_Q+1,2} y_{1,t-2}^{M} \\ \vdots \\ y_{n_M,t}^{M} - \rho_{n,1} y_{n_M,t-1}^{M} - \rho_{n,2} y_{n_M,t-2}^{M} \end{bmatrix}$$

and the state vector $F_t = [a_t \ a_{t-1} \ a_{t-2} \ f_t \ f_{t-1} \ f_{t-2}]'$. As we extract the trend and factor, the factor loadings consist of two blocks, $\tilde{\Lambda} = [\Lambda_a \ \Lambda_f]$, which are respectively defined as the following:

$$\Lambda_a = \begin{bmatrix} 1 & -\rho_{1,1} & -\rho_{1,2} \\ 1 & -\rho_{2,1} & -\rho_{2,2} \\ b_c & -b_c\rho_{3,1} & -b_c\rho_{3,2} \\ & 0_{(n-3)\times 3} & \end{bmatrix}, \quad \Lambda_f = \begin{bmatrix} 1 & -\rho_{1,1} & -\rho_{1,2} \\ \Lambda_2 & -\Lambda_2\rho_{2,1} & -\Lambda_2\rho_{2,2} \\ \vdots & \vdots & \vdots \\ \Lambda_n & -\Lambda_n\rho_{2,1} & -\Lambda_n\rho_{n,2} \end{bmatrix}$$

where b_c loads on the consumption. The matrix of autocorrelation coefficients A also consists of two blocks, $[F_1; \ F_2]$, which are defined as:

$$F_1 = \begin{bmatrix} 1 & 0_{1\times 2} \\ I_2 & 0_{2\times 1} \end{bmatrix}, \quad F_2 = \begin{bmatrix} \phi_1 & \phi_2 & 0 \\ I_2 & 0_{2\times 1} & \end{bmatrix}$$

Finally, the innovations to the law of motion can be described as $e_t = [v_{a,t} \ 0_{2\times1} \ \epsilon_t \ 0_{2\times1}]$', which follows the normal distribution with mean zero and the covariance matrix Q_t = diag(ω_a^2, $0_{1\times2}$, $\sigma_{\epsilon,t}^2$,$0_{1\times2}$). The covariance matrix of the measurement equation R_t, is the diagonal matrix with $\omega_{\eta_i,t}^2$, for i = 1...n.

A.2 The estimation algorithm

Let $\theta = \{\lambda, \Phi, \rho, \omega_a, \omega_\epsilon, \omega_\eta, p, \alpha, \beta\}$ be the underlying parameters of the model, and Φ, ρ represent the autoregressive coefficients for the factor and idiosyncratic components.[24] The latent states to be estimated are $\{a_t, f_t, \sigma_{\epsilon,t}, \sigma_{\eta i,t}, o_{it}\}_{t=1}^T$. All variables are standardized, and the superscript j denotes a current draw. The algorithm consists of the following steps.

Initialization

The model parameters are initialized at the arbitrary starting values for the parameters and the stochastic volatilities, θ^0 and $\{\sigma^0_{\epsilon,t}, \sigma^0_{\eta i,t}, o^0_{it}\}_{t=1}^T$. Specifically, I set $\lambda = 1$, $\Phi = \rho = 0$, $\omega_a = \omega_\epsilon = \omega_\eta = 10^{-5}$, p = 0.1. I set α and β to reflect stylized facts observed by Stock and Watson (2016) that the outlier occurs once every four years in the 10 years of the pre-sample. I also select $\sigma_{\epsilon,t} = 0.1$, $\sigma_{\eta i,t} = 1$ and $o_{it} = 1$ for the latent states. Set j = 1.

1. Construct monthly-interpolated values for quarterly variables

For the quarterly variables i = 1...n_Q, compute $\Delta y^Q_{it} - c_{it} - \lambda_i(L) f_t$, given $a_t^{j-1}, f_t^{j-1}, \lambda^{j-1}$ from the previous iteration. Then, using the state space in Eq (12) and (13) and conditional on $\rho^{j-1}, \{\sigma^{j-1}_{\eta i,t}\}_{t=1}^T$, draw $\hat{u}^M_{i,t}$ by employing the Kalman filter and smoother. Following Bai and Wang (2015), I initialize the filter from a normal distribution, i.e. $x_0 \sim N(0, 10^4 I_5)$, and the covariance matrix of the measurement equation R_t is set to 10^{-4}. Then we obtain the monthly-interpolated values, $\Delta y^{M,Q}_{it} = c_{it} + \lambda_i(L) f_t + \hat{u}^M_{i,t}$. When j = 1, I interpolate quarterly variables based on the cubic splines using the MATLAB command *interp1*. In the case of the out-of-sample analysis, the unknown values are set to NaN.

2. Draw the latent factors and autoregressive coefficients

Based on the previous draws of $\phi^{j-1}, \Lambda^{j-1}, \rho^{j-1}, \omega_a^{j-1}, \{\sigma^{j-1}_{\epsilon,t}, \sigma^{j-1}_{\eta i,t}\}_{t=1}^T$, construct the state space in the quasi-differences as shown in Appendix A.1. Also apply the quasi-differencing to the monthly and monthly-interpolated quarterly indicators obtained from the previous step. Draw the factor and the trend, $p(\{a_t^j, f_t^j\}_{t=1}^T | \theta^{j-1}, \{\sigma^{j-1}_{\epsilon,t}, \sigma^{j-1}_{\eta i,t}\}_{t=1}^T, y)$ using the Kalman filter and smoother, which are initialized from a normal distribution, i.e. $x_0 \sim N(0, 10^4 I)$. Then, conditional on the new draws of $\{a_t^j\}_{t=1}^T$, obtain $v^j_{a,t}$ using Eq (4). Draw the variance of the time-varying GDP growth component, $p(\omega_a^{2,j} | \{a_t^j\}_{t=1}^T)$, from the Inverse-Gamma (IG) posterior with the prior of one degree of freedom and the scale of 0.001.

Given $\{f_t^j\}_{t=1}^T$ from the previous step, draw the autoregressive parameters of the factor

[24] These are the parameters of the model where outliers follow the piecewise-uniform distribution. In the case of the student-t distributed outliers, we estimate parameters ν instead of $\{p, \alpha, \beta\}$.

VAR, Φ^j from the Normal-Inverse Wishart posterior $p(\Phi^j|\{f_t^j, \{\sigma_{\epsilon,t}^{j-1}\}_{t=1}^T)$. Specifically, I run the standard Bayesian VAR routine with the Minnesota-style priors described in Section 2.3 and p = 2. Then, using the residuals from the BVAR, run the Kalman filter and smoother and use the mixture of normals following Kim et al. (1998) to draw the SV component of innovation to the factors, $p(\{\sigma_{\epsilon,t}^j\}_{t=1}^T|\Phi, \{f_t\}_{t=1}^T)$. Finally, draw $\omega_\epsilon^{2,j}$ conditional on $\{\sigma_{\epsilon,t}^j\}_{t=1}^T$ and from IG posterior with the prior of the one degree of freedom and the scale of 10^{-4}.

3. Draw the factor loadings and serial correlation coefficients of idiosyncratic components

Draw the loadings λ^j from $p(\lambda^j|\rho^{j-1}, \{f_t^j, \sigma_{\eta i,t}^{j-1}\}_{t=1}^T, y)$. They can be estimated via GLS, conditional on $\rho_i j - 1$ and $\sigma_{\eta i,t}^{j-1}$. To be more specific, for each variable i = 1...n, I first divide the indicator (monthly or monthly-interpolated quarterly) and the factor $\{f_t^j\}_{t=1}^T$ by the volatility of the idiosyncratic component from the previous draw, $\{\sigma_{\eta i,t}^{j-1}\}_{t=1}^T$. Then I also apply quasi-differencing to the resulting indicator and factor. After these steps, I stack all the indicators and factors and run the standard OLS: for the prior on λ^j, I set the mean and variance as the matrix of ones and 0.2, respectively. Finally, I correct the estimated loadings λ^j to reflect the restriction that the loading of GDP on f_t to be unity, following Bai and Wang (2015).

Having obtained λ^j and $\{a_t^j, f_t^j\}_{t=1}^T$, calculate the idiosyncratic components $u_{it}^j = \Delta y_{it} - c_{it} - \lambda_i(L) f_t$. Draw serial correlation coefficients ρ^j from $p(\rho|\lambda^j, \{f_t^j, \sigma_{\eta i,t}^{j-1}\}_{t=1}^T, y)$ in a similar way to the above, i.e. for each variable i = 1...n, first divide u_{it}^j by the volatility of the idiosyncratic component from the previous draw, $\{\sigma_{\eta i,t}^{j-1}\}_{t=1}^T$. Then estimate Eq (3) via OLS. I set the prior mean and the variance of ρ_i^j as the vector of zeros and 0.2 with the lag decay of 2. In the case of explosive roots, I discard the draw and repeat this step.

4. Draw the SV of innovations and the outlier states.

For each variable i = 1...n, I obtain the residuals $u_{i,t}^{*,j} = (1 - \rho^j(L)) u_{i,t}^j$. Then, (*in the case of the piecewise-uniform distribution*) conditional on $\{\sigma_{\eta i,t}^{j-1}, o_{i,t}^{j-1}, u_{i,t}^{*,j}\}_{t=1}^T$ in log terms, obtain the mixture states. Then obtain the new draws of $\{\sigma_{\eta i,t}^j\}_{t=1}^T$ using the Kalman filter and smoother, where we construct the observations by subtracting the mean of the mixture states and $\log(o_{i,t}^{j-1})$ from the $\log(u_{i,t}^{*,j})$. I initialize the filter from $x_0 \sim N(0, 10I)$, with the transition matrix and the loadings set to the Identity. The covariance matrices for the state and measurement equations consist of $\omega_{\eta i}^{2,j-1}$ and the volatility of the mixture states, respectively. Then draw p($o_{i,t}^j|u_{i,t}^*, \sigma_{\eta i,t}^j$) based on the cdf of standardized draws for each component of mixture normals and calculate the number of outliers to update α_i, β_i. Finally, draw p_i^j from the beta (α_i, β_i) posterior. The entire step can be parallelized and run in a univariate state-space system.

(*in the case of the student-t distribution*) conditional on $\{\sigma^{j-1}_{\eta i,t}, o^{j-1}_{i,t}, u^{*,j}_{i,t}\}_{t=1}^{T}$ in log terms, update the posterior distribution of the ν_i, which is proportional to the product of t-distribution ordinates, i.e. $p(\nu^j_i|u^*_{i,t}, \sigma^{j-1}_{\eta i,t}) = p(\nu)\Pi_{t=1}^{T}\frac{\nu^{-1/2}\Gamma(\nu+1/2)}{\Gamma(1/2)\Gamma(\nu/2)}(\nu + (u^{*,j}_{i,t}/\sigma^{j-1}_{\eta i,t})^2)^{-(\nu+1)/2}$, following Jacquier et al. (2004). I impose a weakly informative prior, i.e. $p(\nu) \sim \Gamma(2, 10)$ which is discretized on the support [3:40], to ensure the existence of a conditional variance at the lower bound. Then I draw $p(o^j_{i,t}|\nu^j, u^*_{i,t}, \sigma^{j-1}_{\eta i,t})$, using $\nu^j/o^j_{i,t} \sim \chi^2_{v+1}$. After this step, I follow the same process as in the previous case of the piecewise-uniform distribution, to update $\{\sigma^j_{\eta i,t}\}_{t=1}^{T}$ using the Kalman filter and smoother, where we construct the observations by subtracting the mean of the mixture states and $\log(o^{j-1}_{i,t})$ from the $\log(u^{*,j}_{i,t})$.

Then, in both cases, draw $p(\omega^{2,j}_{\eta i}|\{\sigma^j_{\eta i,t}\}_{t=1}^{T})$ from the IG posterior with the prior of the one degree of freedom and the scale of 10^{-4}.

Increase j by 1 and repeat the steps above until j = 7000. After discarding the first 2000 as burn-in draws, use the rest 5000 draws of the model parameters and latent variables for the inference.

A.3 The unwanted dependency

Antolin-Diaz et al. (2023) model the outliers in an additive rather than multiplicative way, with the following measurement equation:

$$\Delta y_t - o_t = c_t + \Lambda(L) f_t + u_t, \quad o_{it} \sim t_{vi}(0, \sigma_{oi}^2)$$

where Δy_t is the transformed data, which can be in levels or differences, and so is o_t. By re-arranging and quasi-differencing $(1 - \rho_i(L))u_{it} = \sigma_{\eta_{it}} o_{it} \eta_{it}$ with the lag order of 2, we obtain:

$$\tilde{y}_t = \tilde{\Lambda} F_t + (1 - \rho_1 L - \rho_2 L^2)(u_t + o_t)$$

where $\tilde{y}_t$ and $\tilde{\Lambda} F_t$ are in terms of the quasi-differences, e.g. $\tilde{y}_t = (1 - \rho_1 L - \rho_2 L^2)\Delta y_t$, as shown in the state-space representation in Appendix A.1. Then, the variance of residuals, the last term on the right-hand side, can be expressed as:

$$\begin{aligned} V[(1 - \rho_1 L - \rho_2 L^2)(u_{it} + o_{it})] &= V(\sigma_{\eta_{it}} \eta_{it} + (1 - \rho_1 L - \rho_2 L^2) o_{it}) \\ &= \sigma_{\eta_{it}}^2 + (1 + \rho_1^2 + \rho_2^2) V(o_{it}) \end{aligned}$$

where the first equality reflects $(1 - \rho_i(L))u_{it} = \sigma_{\eta_{it}} o_{it} \eta_{it}$, and the second line is based on the two extra assumptions (1) – (2): u_{it} and o_{it} are uncorrelated and o_{it} are serially uncorrelated. These are mostly innocuous by the definition of "one-off" outliers. Then, the role of outlier depends on ρ, the autoregressive coefficients of the idiosyncratic component u_{it}, not the outlier states o_{it}. This is not an intended feature of the model, hence the 'unwanted' dependency. In the likely case of $\rho > 0$, it amplifies the variance of the outlier component, resulting in an underestimation of the SV components in innovations. We can calculate further to obtain the variance of outlier components:

$$\begin{aligned} V(o_{it}) &= V(\sqrt{\psi_{it}} z_{it}) \\ &= V[E(\sqrt{\psi_{it}} z_{it} | \psi_{it})] + E[V(\sqrt{\psi_{it}} z_{it} | \psi_{it})] \\ &= V[\sqrt{\psi_{it}} \underbrace{E(z_{it} | \psi_{it})}_{=0}] + E[\psi_{it} \underbrace{V(z_{it} | \psi_{it})}_{=\sigma_{oi}^2}] \\ &= \sigma_{oi}^2 E(\psi_{it}) \end{aligned}$$

where we apply the specification of outliers from Antolin-Diaz et al. (2023), i.e. the latent scale mixture variable ψ_{it} follows $\nu_i / \psi_{it} \sim \chi_{\nu_i}^2$, and the noise $z_{it} \sim N(0, \sigma_{oi}^2)$, so that $o_{it} = \sqrt{\psi_{it}} z_{it} \sim t_{\nu_i}(0, \sigma_{oi}^2)$. Since $\nu_i / \psi_{it} \sim \chi_{\nu_i}^2$ implies that $\psi_{it} / \nu_i \sim IG(\frac{\nu_i}{2}, \frac{1}{2})$, we can complete the above expression by:

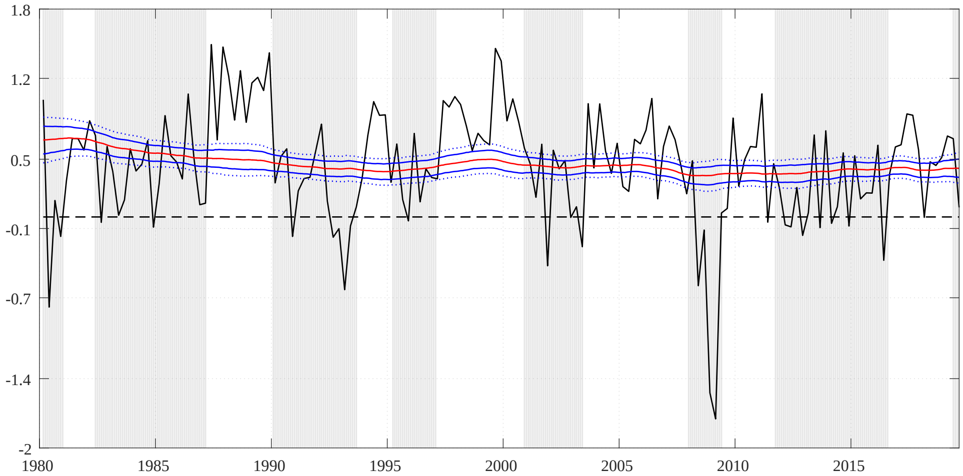
$$E(\frac{\psi_{it}}{\nu_i}) = \frac{1}{\nu_i - 2}, \quad E(\psi_{it}) \qquad\qquad = \frac{\nu_i}{\nu_i - 2}$$

We can reach the same expression for $V(o_{it})$ by both additive and multiplicative modeling of outliers. However, when the quasi-differencing is applied, the dependeny issue only arises in the former case.

B Additional figures

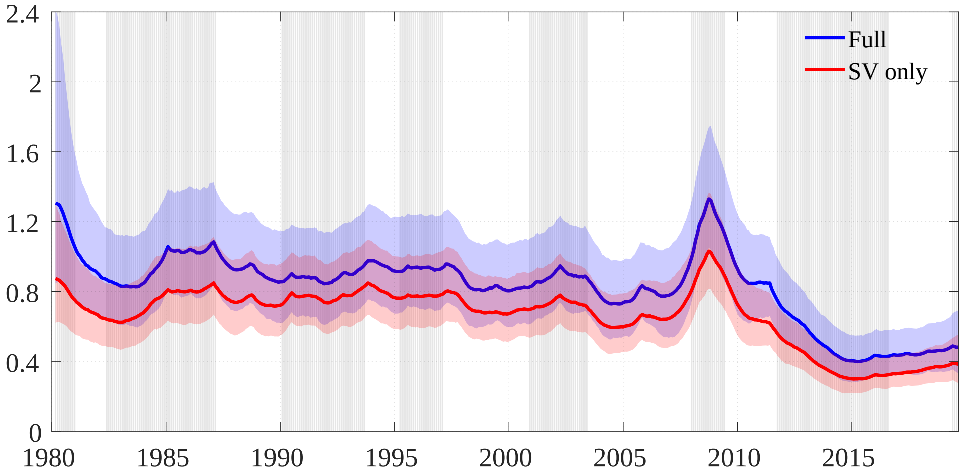
B.1 Posterior estimate of trend and factor volatility

FIGURE B.1: POSTERIOR ESTIMATE OF TREND V. REAL GDP GROWTH

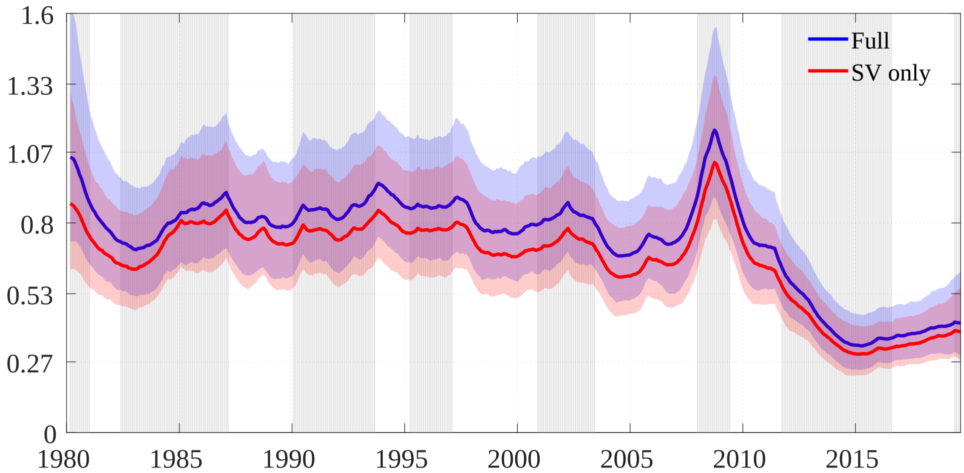


Note: This plot shows the posterior estimate of the long-run growth from the model with outliers following the uniform mixture distribution (in red and blue) and the real GDP growth of France in the black solid line. The solid red and blue lines denote the posterior median and 68% bands, respectively, while the dotted blue line represents 90% bands. Sample: 1980.1 – 2019.9. OECD recessions in gray.

Figure B.2: Volatility of the common activity factor



(a) student-t outliers v. no outlier

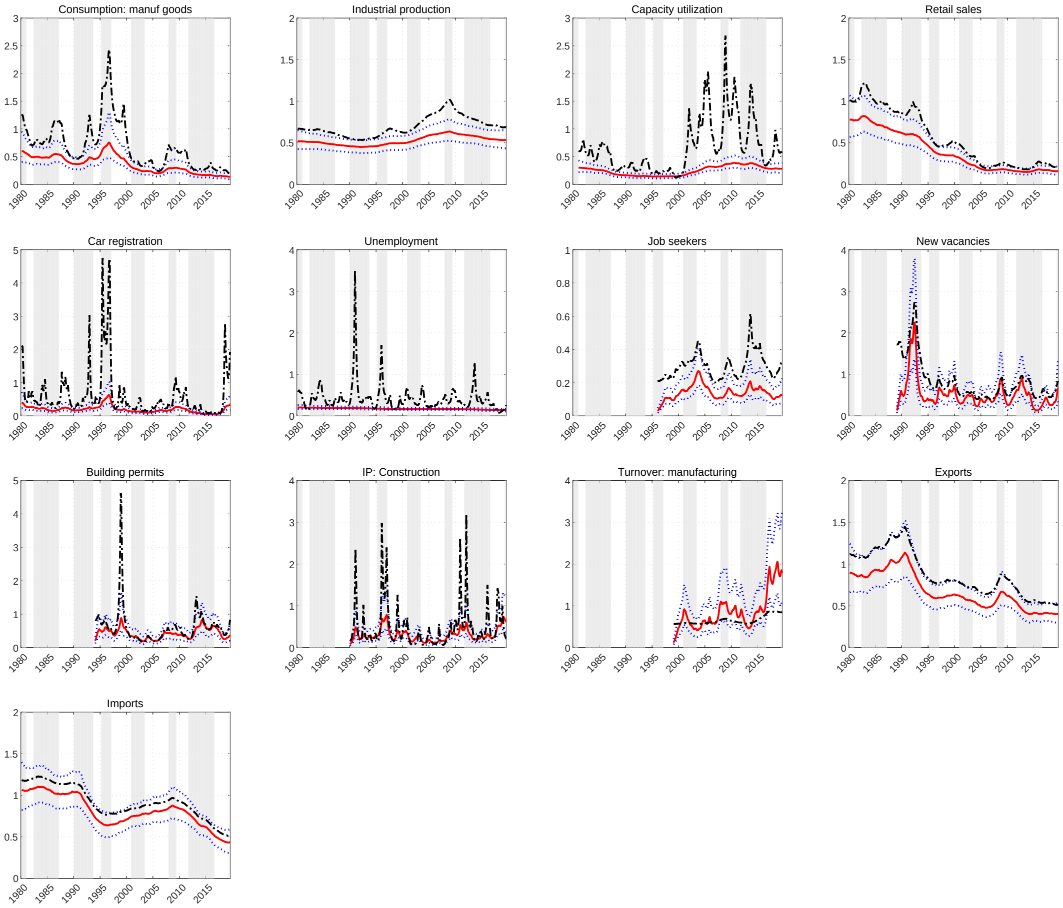


(b) uniform outliers v. no outlier

Note: This plot shows the posterior estimate of volatility of the common activity factor from the two models, the one with (in blue) and the one without (in red) outliers. The solid lines and shaded areas denote the posterior median and 90% bands, respectively. Sample: 1980.1 – 2019.9. OECD recessions in gray.

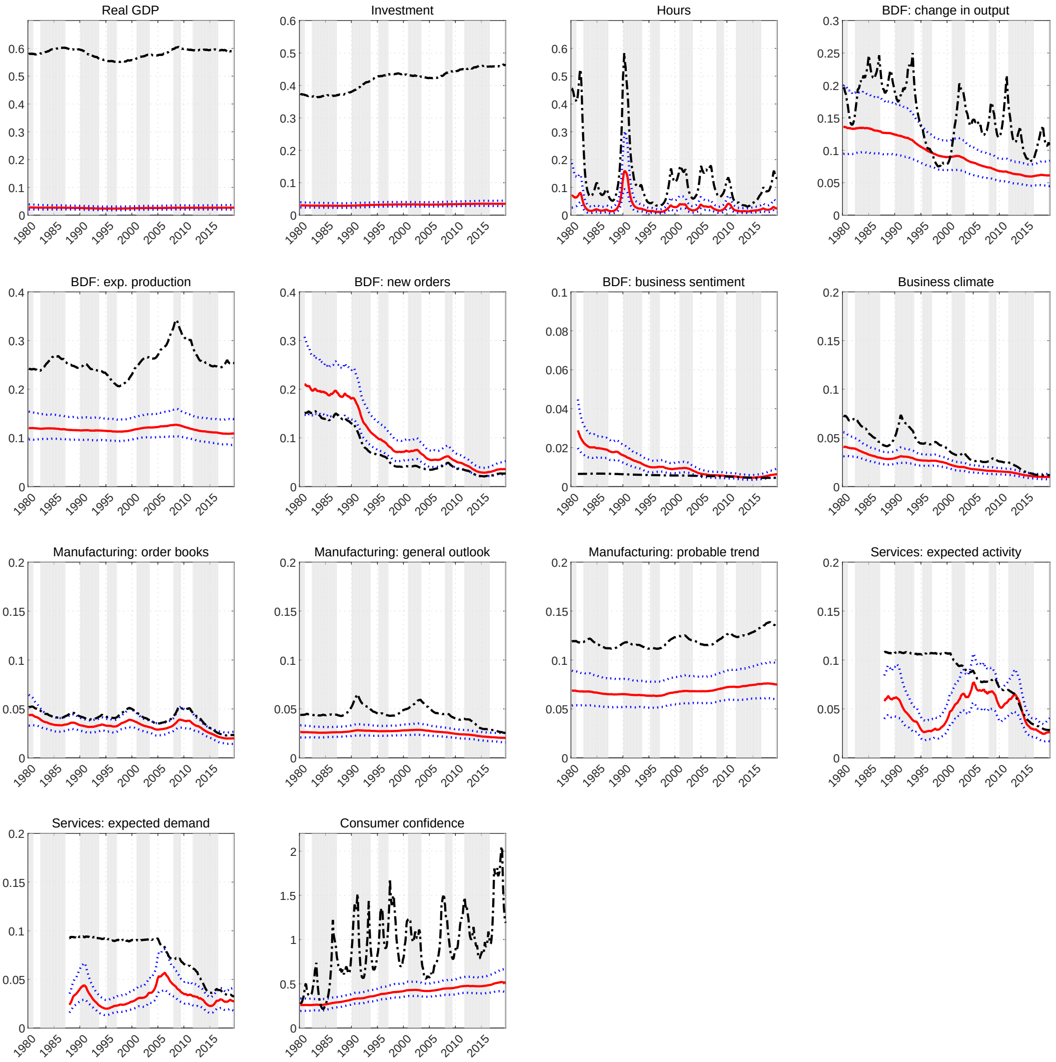
B.2 Stochastic volatility: with v. without outliers

Figure B.3: SV estimates: student-t v. without outliers, monthly hard



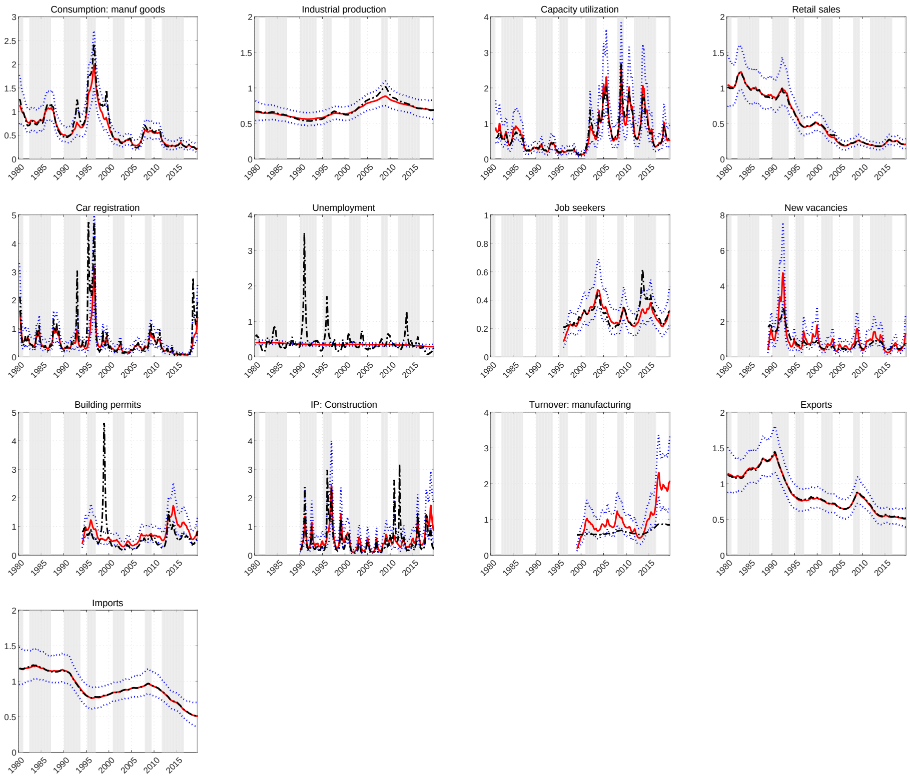
Note: This plot shows the posterior estimate of stochastic volatility in innovations to the idiosyncratic components of monthly hard indicators. The black line represents the posterior median from the model without outliers. Red and blue lines are the posterior median and 68% bands from the model with outliers following the t-distribution, respectively. Sample: 1980.1 – 2019.9. OECD recessions in gray.

Figure B.4: SV estimates: student-t v. without, quarterly and soft



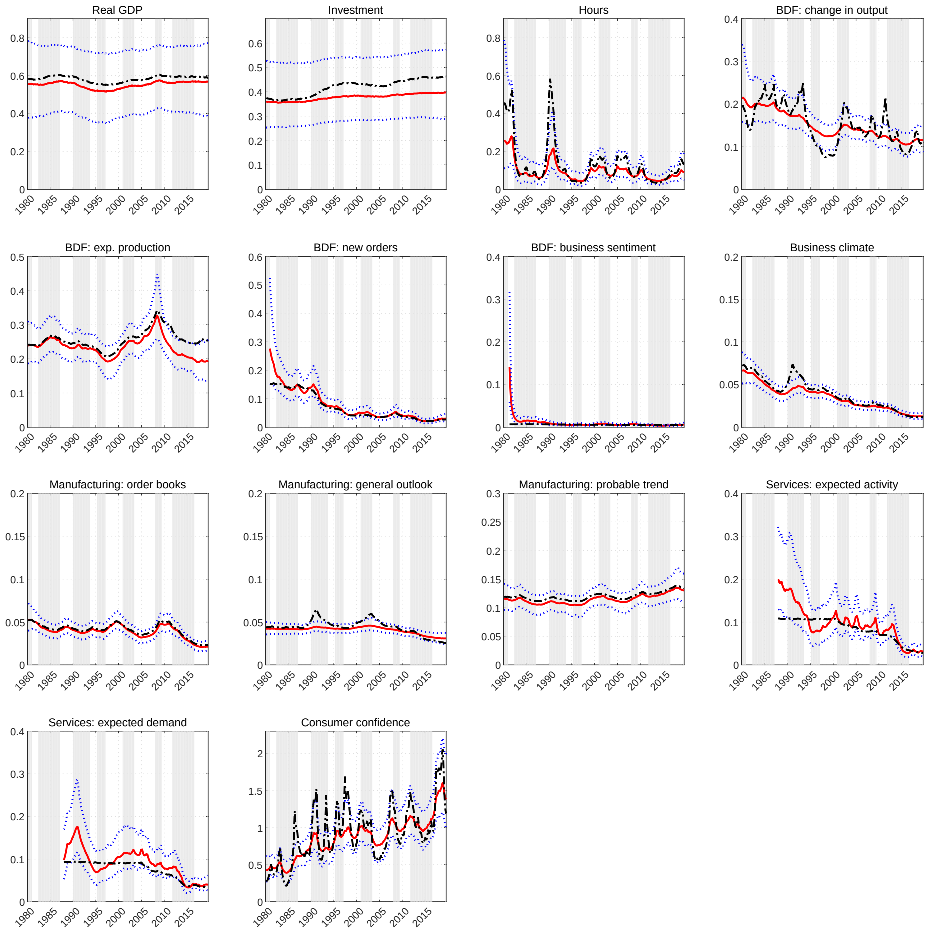
Note: This plot shows the posterior estimate of stochastic volatility in innovations to the idiosyncratic components of quarterly and monthly soft indicators. The black line represents the posterior median from the model without outliers. Red and blue lines are the posterior median and 68% bands from the model with outliers following the t-distribution, respectively. Sample: 1980.1 – 2019.9. OECD recessions in gray.

Figure B.5: SV estimates: uniform v. without, monthly hard



Note: This plot shows the posterior estimate of stochastic volatility in innovations to the idiosyncratic components of monthly hard indicators. The black line represents the posterior median from the model without outliers. Red and blue lines are the posterior median and 68% bands from the model with outliers following the uniform mixture distribution, respectively. Sample: 1980.1 – 2019.9. OECD recessions in gray.

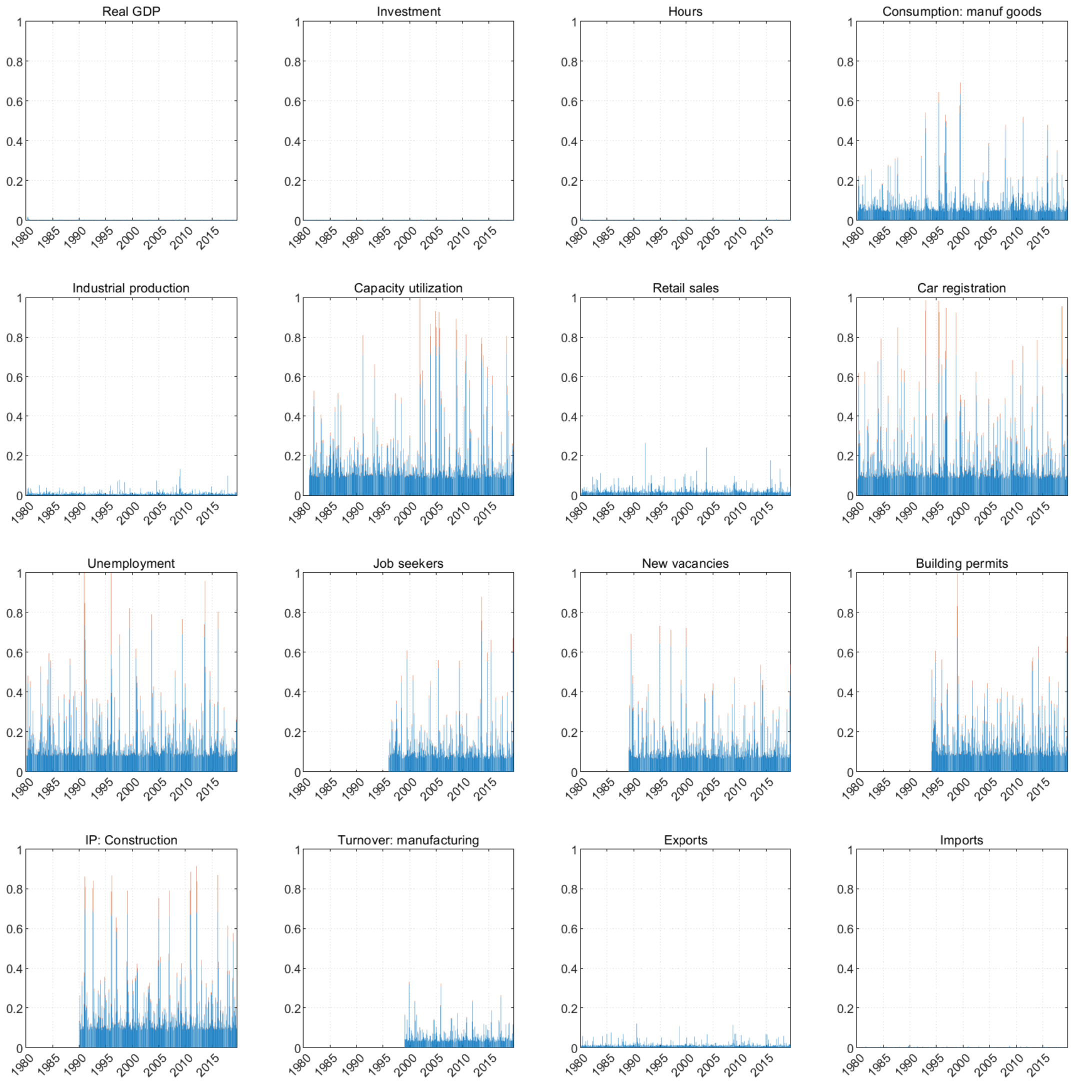
Figure B.6: SV estimates: uniform v. without, quarterly and soft



Note: This plot shows the posterior estimate of stochastic volatility in innovations to the idiosyncratic components of quarterly and monthly soft indicators. The black line represents the posterior median from the model without outliers. Red and blue lines are the posterior median and 68% bands from the model with outliers following the uniform mixture distribution, respectively. Sample: 1980.1 – 2019.9. OECD recessions in gray.

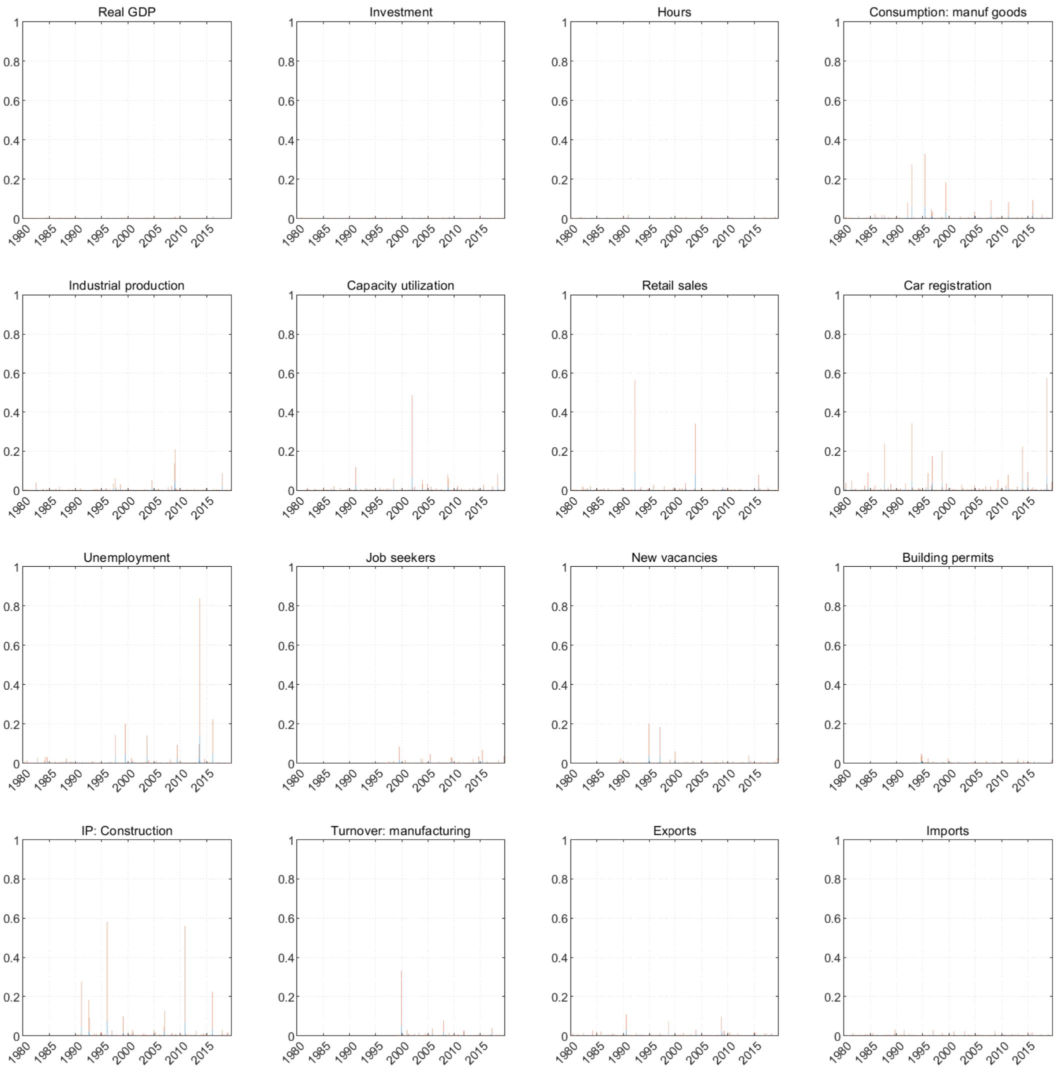
B.3 Posterior estimates of outlier states

Figure B.7: Posterior probabilities of outlier states: student-t model



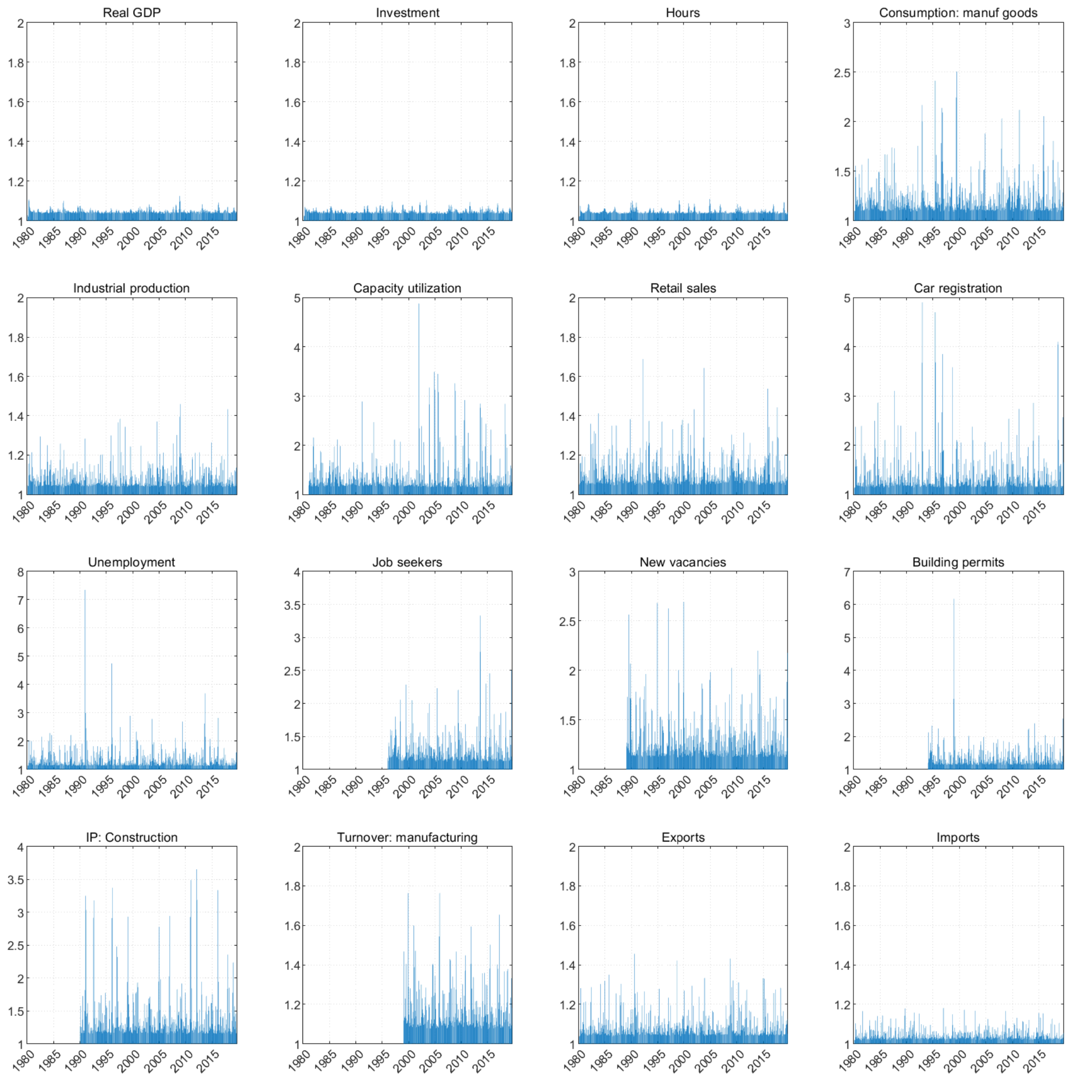
Note: The stacked bars represent posterior probabilities for realizations of outlier states that are larger than two. The blue bars correspond to the probability that the outliers are within the range between two and five, and the orange bars denote the probability of outlier states that are larger than five to take place. Monthly and quarterly hard indicators in the dataset are included. Sample: 1980.1 – 2019.9.

Figure B.8: Posterior probabilities of outlier states: uniform model



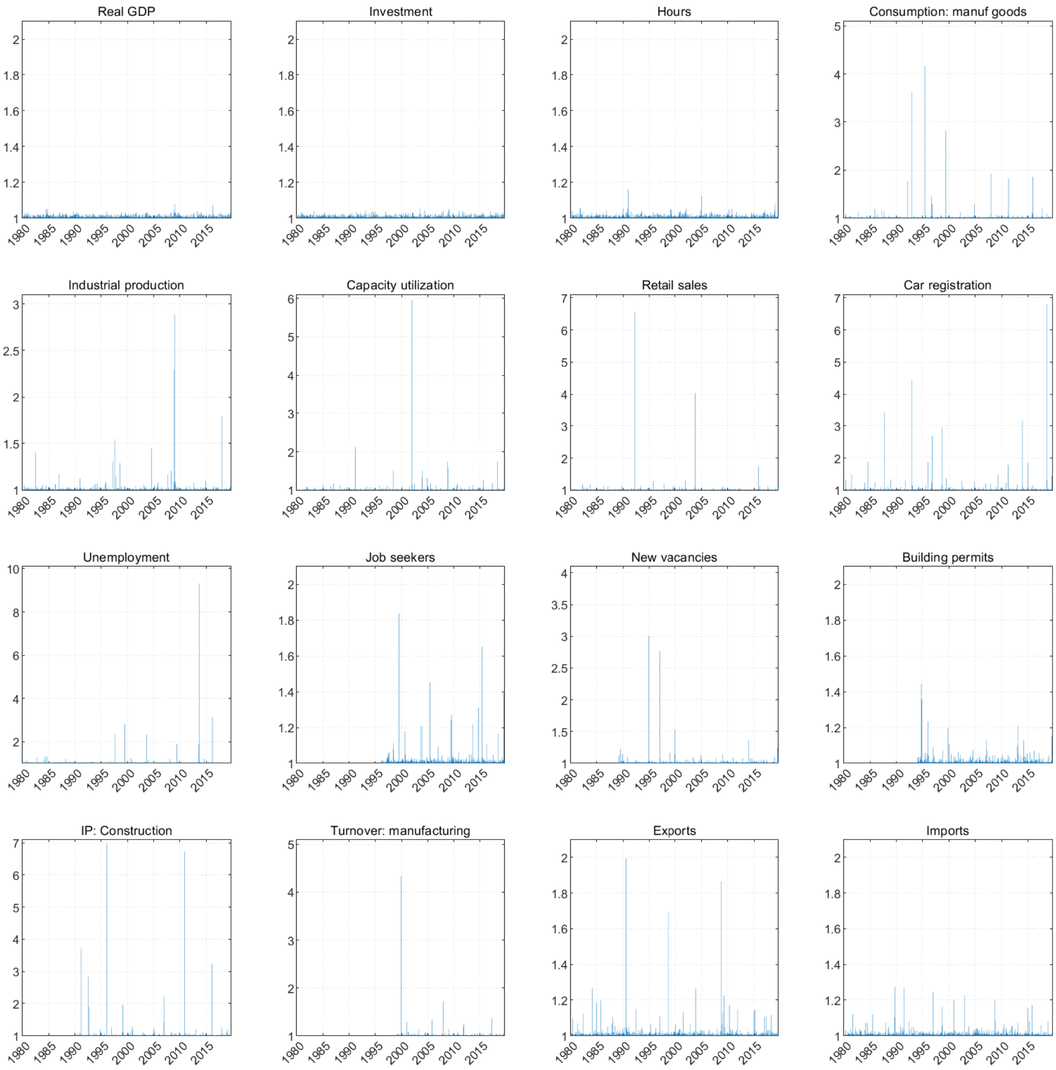
Note: The stacked bars represent posterior probabilities for realizations of outlier states that are larger than two. The blue bars correspond to the probability that the outliers are within the range between two and five, and the orange bars denote the probability of outlier states that are larger than five to take place. Monthly and quarterly hard indicators in the dataset are included. Sample: 1980.1 – 2019.9.

FIGURE B.9: POSTERIOR ESTIMATES OF OUTLIER STATES: STUDENT-T MODEL



Note: This plot shows the posterior mean estimates of outlier states for monthly and quarterly hard indicators in the dataset. Sample: 1980.1 – 2019.9.

Figure B.10: Posterior estimates of outlier states: uniform model



Note: This plot shows the posterior mean estimates of outlier states for monthly and quarterly hard indicators in the dataset. Sample: 1980.1 – 2019.9.

C Additional tables

C.1 Out-of-sample point forecasting evaluation

Table C.1: Out-of-sample point forecasting evaluation, pre-COVID

Horizon/model	Full (t-dist)	Full (uniform)	SV miss	SV	Basic miss	Basic DFM
(a) point forecasting: RMSE						
-5 month	0.487	0.460	0.462	0.461	0.483	0.485
-4 month	0.436	0.437	0.436	0.436	0.461	0.461
-3 month	0.406	0.395	0.393	0.393	0.425	0.425
-2 month	0.367	0.329	0.327	0.328	0.376	0.377
-1 month	0.306	0.304	0.303	0.304	0.352	0.348
0 month (end of reference Q)	0.307	0.297	0.298	0.298	0.343	0.341
1 month	0.307	0.294	0.293	0.294	0.336	0.338
(b) point forecasting: MAFE						
-5 month	0.366	0.345	0.347	0.346	0.339	0.340
-4 month	0.341	0.328	0.325	0.326	0.329	0.329
-3 month	0.326	0.300	0.297	0.299	0.305	0.305
-2 month	0.289	0.255	0.253	0.253	0.277	0.276
-1 month	0.250	0.243	0.241	0.242	0.266	0.264
0 month (end of reference Q)	0.252	0.237	0.239	0.240	0.263	0.260
1 month	0.249	0.235	0.235	0.235	0.255	0.258

Note: This table provides the forecasting performance of different models: the DFM model with outliers following the student-t and the piecewise-uniform distribution, respectively; the model with stochastic trend and volatility (Antolin-Diaz et al., 2017); and the basic DFM model (Bańbura and Modugno, 2014). 'SV miss' and 'Basic miss' are the latter two models with the outliers, which are defined as the observations five-interquartile range away from the median, replaced by missing values. Using data going back to 1980, I estimate the model at each point in time for 2005.1 – 2019.9 in an expanding window.

Table C.2: Out-of-sample density forecasting evaluation, pre-COVID

Horizon/model	Full (t-dist)	Full (uniform)	SV miss	SV	Basic miss	Basic DFM
(a) density forecasting: log score						
-5 month	-0.859^{*}	-0.869^{*}	-0.901^{*}	-0.884^{*}	-2.094^{*}	-2.203
-4 month	-0.674^{*}	-0.969^{*}	-0.971^{*}	-0.999^{*}	-2.697	-2.803
-3 month	-0.804^{**}	-1.161^{**}	-1.184^{**}	-1.234^{**}	-3.577	-3.960
-2 month	-0.757^{***}	-1.660^{***}	-1.703^{***}	-1.663^{***}	-7.541	-8.281
-1 month	-0.958^{***}	-2.486^{***}	-2.474^{***}	-2.491^{***}	-16.314	-15.538
0 month (end of reference Q)	-1.213^{***}	-2.971^{***}	-3.033^{***}	-3.009^{***}	-22.931	-20.279
1 month	-1.568^{***}	-2.960^{***}	-3.080^{***}	-3.098^{***}	-21.274	-21.784
(b) density forecasting: CRPS						
-5 month	0.267	0.254	0.255	0.255	0.266	0.268
-4 month	0.245	0.246	0.245	0.245	0.261	0.260
-3 month	0.237	0.231	0.230	0.231	0.249	0.250
-2 month	0.214	0.201	0.200	0.200	0.236	0.236
-1 month	0.187	0.194^{*}	0.193^{*}	0.193^{*}	0.236	0.234
0 month (end of reference Q)	0.189	0.193^{*}	0.194^{*}	0.195^{*}	0.238	0.234
1 month	0.192	0.191^{**}	0.191^{**}	0.192^{**}	0.231	0.233

Note: This table provides the forecasting performance of different models: the DFM model with outliers following the student-t and the piecewise-uniform distribution, respectively; the model with stochastic trend and volatility (Antolin-Diaz et al., 2017); and the basic DFM model (Bańbura and Modugno, 2014). 'SV miss' and 'Basic miss' are the latter two models with the outliers, which are defined as the observations five-interquartile range away from the median, replaced by missing values. Using data going back to 1980, I estimate the model at each point in time for 2005.1 – 2019.9 in an expanding window. The asterisks are related to the p-value of the null hypothesis that the basic DFM model performs as well as others, against the alternative that the other model performs better, based on the test statistic of Diebold and Mariano (1995). *** is significant at the 1% level, ** at the 5% level, and * at the 10% level.

Table C.3: Out-of-sample point forecasting evaluation, full sample

Horizon/model	Full (t-dist)	Full (uniform)	SV miss	SV	Basic miss	Basic DFM
(a) point forecasting: RMSE						
-5 month	3.258	5.002	3.352	3.361	2.966	2.744
-4 month	6.718	4.319	5.823	6.370	2.739	2.685
-3 month	2.060	2.166	2.690	2.674	2.692	2.789
-2 month	1.352	3.668	2.312	1.565	2.466	2.053
-1 month	3.674	2.445	3.035	2.659	2.759	2.078
0 month (end of reference Q)	0.885	0.961	1.917	0.952	2.309	2.096
1 month	0.824	1.555	2.400	1.302	2.562	2.112
(b) point forecasting: MAFE						
-5 month	1.066	1.277	1.1147	1.090	0.927	0.923
-4 month	1.607	1.231	1.412	1.423	0.840	0.877
-3 month	0.829	0.807	0.919	0.868	0.856	0.832
-2 month	0.562	1.041	0.776	0.591	0.780	0.731
-1 month	0.874	0.654	0.885	0.693	0.813	0.699
0 month (end of reference Q)	0.437	0.480	0.579	0.484	0.754	0.673
1 month	0.452	0.618	0.701	0.544	0.806	0.740

Note: This table provides the forecasting performance of different models: the DFM model with outliers following the student-t and the piecewise-uniform distribution, respectively; the model with stochastic trend and volatility (Antolin-Diaz et al., 2017); and the basic DFM model (Bańbura and Modugno, 2014). 'SV miss' and 'Basic miss' are the latter two models with the outliers, which are defined as the observations five-interquartile range away from the median, replaced by missing values. Using data going back to 1980, I estimate the model at each point in time for 2005.1 – 2022.9 in an expanding window.

Table C.4: Out-of-sample density forecasting evaluation, full sample

Horizon/model	Full (t-dist)	Full (uniform)	SV miss	SV	Basic miss	Basic DFM
(a) density forecasting: log score						
-5 month	-21.500	-29.690	-30.070	-30.696	-68.461	-163.700
-4 month	-25.484	-40.091	-39.032	-38.766	-87.874	-428.82
-3 month	-5.498	-8.174	-8.762	-9.049	-282.777	-344.83
-2 month	-8.332	-16.951	-18.884	-19.398	-86.944	-193.510
-1 month	-12.095	-26.112	-26.291	-24.729	-151.154	-557.547
0 month (end of reference Q)	-4.573	-8.401	-10.158	-10.906	-359.418	-339.861
1 month	-2.126	-3.459	-18.682	-14.156	-168.765	-425.400
(b) density forecasting: CRPS						
-5 month	0.906	1.011	0.963	0.961	0.855	0.848
-4 month	1.303	0.967	1.198	1.222	0.781	0.802
-3 month	0.660	0.629	0.740	0.720	0.777	0.747
-2 month	0.426	0.707	0.670	0.514	0.720	0.631
-1 month	0.739	0.542	0.762	0.590	0.773	0.644
0 month (end of reference Q)	0.358	0.382	0.503	0.393	0.716	0.632
1 month	0.341	0.435	0.628	0.474	0.772	0.683

Note: This table provides the forecasting performance of different models: the DFM model with outliers following the student-t and the piecewise-uniform distribution, respectively; the model with stochastic trend and volatility (Antolin-Diaz et al., 2017); and the basic DFM model (Bańbura and Modugno, 2014). 'SV miss' and 'Basic miss' are the latter two models with the outliers, which are defined as the observations five-interquartile range away from the median, replaced by missing values. Using data going back to 1980, I estimate the model at each point in time for 2005.1 – 2022.9 in an expanding window.