The COVID-19 shock and challenges for inflation modelling☆

Elena Bobeica *, Benny Hartwig

European Central Bank, Sonnemannstrasse 22, 60314 Frankfurt am Main, Germany

A B S T R A C T

We document the impact of COVID-19 on inflation modelling within a vector autoregression (VAR) model and provide guidance for forecasting euro area inflation during the pandemic. We show that estimated parameters are strongly affected, leading to different and sometimes implausible projections. As a solution, we propose to augment the VAR by allowing the residuals to have a fat-tailed distribution instead of a Gaussian one. This also outperforms with respect to unconditional forecasts. Yet, what brings sizeable forecast gains during the pandemic is adding meaningful off-model information, such as that entailed in the Survey of Professional Forecasters. The fat-tailed VAR loses part, but not all of its relative advantage compared to the Gaussian version when producing conditional inflation forecasts in a real-time setup. It is the joint fat-tailed errors and multi-equation modelling that manage to robustify models against extreme observations; in a single-equation model the same solution is less effective.

© 2022 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

'In reality, however, the distribution of shocks hitting the economy is more complex.... [The shocks] might exhibit excess kurtosis, commonly referred to as "tail risk" in which the probability of relatively large disturbances is higher than would be implied by a Gaussian distribution.'

[Mishkin, 2011]

As the pandemic unfolded worldwide, macroeconomists struggled to make sense of their models after adding the new COVID-19 observations. The variation exhibited by some macroeconomic series was so large (e.g. real activity

and labour market indicators) that this sufficed to distort the estimated coefficients since March 2020. This was shown for the U.S. (Carriero et al., 2021; Lenza & Primiceri, 2020; Schorfheide & Song, 2020) and we show that this is also true for the euro area.

In this paper we investigate the impact of the COVID-19 shock on one of the most popular time series models—Sims' vector autoregression (VAR) model—and discuss the results by focusing on euro area inflation. Our lessons are valid for a wide range of empirical analyses, but we choose to focus on inflation because this has been a particularly hard-to-grasp indicator since the Great Financial Crisis on both sides of the Atlantic. In the decade prior to the pandemic, inflation across advanced economies consistently surprised economists on the downside. The euro area in particular struggled with low inflation to the extent that inflation expectations became less well anchored to the ECB's inflation aim, adding to existing disinflationary pressures (Baumann et al., 2021). Inflation continued to surprise during the pandemic, starting in 2020. Initially, while energy inflation fell quickly, the response of euro area core inflation was modest relative to the decline in activity. From the middle of the year onwards, headline inflation fell further as core inflation increasingly reflected disinflationary tendencies,

and overall inflation surprised on the downside. Starting in 2021, the situation completely reversed and inflation once again took centre stage in the economic debate, as its surge took central banks and other professional forecasters by surprise. The normalisation of the economy, ensuing supply-side bottlenecks, and rising commodity prices put upward pressure on prices. In this context of highly atypical economic developments—marked by the economic lockdown and the subsequent re-opening—the question is which tools can be still employed to model inflation.

We show that using a standard Gaussian BVAR to model inflation distorts economic inference. The incoming data after March 2020 heavily impact the parameters of conventionally estimated models and this is something to be aware of when forecasting or conducting empirical analyses covering this period, irrespective of the scope. Here we focus on the analysis of inflation, but the critique is generally valid.

The contribution of our paper is twofold: (i) first, we propose solutions for the problem that the change in parameters poses for both in-sample and out-of-sample inference; (ii) second, we show what brings forecast gains when trying to project inflation during the pandemic. To the best of our knowledge, we are the first to discuss real-time inflation forecasting for the euro area for this particular abnormal period. We validate our proposed model by looking at a longer period of time (in a pseudo-real-time inflation forecasting exercise) and by providing evidence for a large set of advanced economies.

We tackle the problem of VAR parameters becoming unstable when adding the COVID-19 observations by proposing a tractable and elegant solution, namely to relax the assumption that the errors are normally distributed and assume instead that they follow a Student's t-distribution. Intuitively, unusually large shocks are more likely to occur under t-errors as Gaussian errors, since the former has fatter tails. This allows the residuals to soak up the abnormal variation, stabilising the parameter estimates.

Our paper is related to the strand of literature emerging after the Great Recession trying to accommodate tail events in macro-models. For DSGEs, Chib and Ramamurthy (2014), Cúrdia et al. (2014), and Ascari et al. (2015) argue that models with a t-distributed shock structure are strongly favoured by the data over standard Gaussian ones.

For VARs, Chiu et al. (2017) and Chan (2020) show that allowing for fat-tailed errors improves in-sample fit and forecasting properties.[1] We show that euro area macroeconomic data exhibit substantial tail risk even before the pandemic. Moreover, this extension suffices to deal with the impact of the pandemic (which is a more rare and extreme event than recessions).

We also contribute to the rising literature on how standard models can be adapted to withstand the impact of the COVID-19 observations. Lenza and Primiceri (2020)

and Carriero et al. (2021) propose downweighing the impact of abnormal observations through assumptions on the associated variance of the residuals. Our solution is similar in spirit. We believe it is still too soon to ascertain a break in the economic transmission mechanisms, and thus parameters should not display dramatic changes with respect to pre-COVID-19 times, especially if a normalisation in economic developments is to be expected. At the same time, we do not exclude the possibility that the COVID-19 observations might affect macroeconomic relationships in the future (as reflected by model parameters). From this point of view, our approach is more flexible than simply cutting the sample to the pre-COVID-19 period or including dummy variables for each quarter over the pandemic. The observations in 2020 will remain an issue for time series models going forward, so time series models have to be adapted for a longer period of time.

In a standard Gaussian BVAR, the response of inflation to a shock in real activity gradually weakens and appears to be stabilising at lower levels compared to the pre-pandemic period. In our fat-tailed BVAR there is still some weakening, but to a smaller extent. There are several reasons that call for caution when it comes to ascertaining changes in established economic relationships, such as the one between inflation and real activity, for instance, known as the Phillips curve. The pandemic is not the usual macroeconomic shock when it comes to its composition and its size. In terms of its composition, it is a multifaceted shock, actually a combination of demand and supply shocks acting at the same time. In terms of its size, it is an abnormally large shock which distorts any estimate not equipped to deal with tail events. On top, there have been unusual data distortions and unusually high uncertainty regarding slack.

We show that it is the joint multivariate and fat-tailed errors that manage to robustify models against extreme observations; in a single-equation model the same solution is less effective. We take a simple Phillips curve model and show that estimated parameters change notably when adding the COVID-19 observations to the sample, implying a drastic flattening in the euro area. A fat-tailed error distribution does not appear to be a panacea this time around. In the multiple-equation system of a VAR the fatness of tails pertaining to the error distribution (and hence the extent to which the errors can soak up abnormal observations in order not to distort coefficients) is informed by the simultaneous developments in all variables. In single-equation models it is the development of the dependent variable that matters more in informing the properties of the residuals; extreme developments occurred in the first part of the pandemic on the real activity front and less so on the nominal side of the economy. Thus, the residuals of a Phillips curve do not have such fat tails as the residuals of the inflation equation in a VAR model that also includes GDP, for instance. Hence, multi-equation models are more flexible and better equipped to deal with the impact of abnormal observations.

As an alternative to fat-tailed errors, we find that adding more variables to the Gaussian BVAR, together with a reasonably tighter prior to control for the larger

[1] Even before the Great Recession some economists warned that the unconditional distribution of macro-variables is not Gaussian (see Christiano (2007) or Fagiolo et al. (2008)).

520

dimension, yields more stable results than a smaller-scale BVAR. This suggests that in large-dimensional models the impact of variables with abnormal dynamics is mitigated; also, an appropriately informative prior helps in disciplining the results. Still, the model with *t*-distributed errors outperforms in delivering more stable parameters.

The pandemic-related change in coefficients also mechanically affects unconditional forecasts in a Gaussian BVAR. The fat-tailed and the large-scale BVAR alleviate this problem. Yet, we find that this problem of parameter and forecast instability is also mitigated in a standard Gaussian BVAR when relevant off-model information is included, as conditional forecasts appear to be well behaved.

Turning to the second contribution on how our proposed fat-tailed BVAR would have fared when forecasting euro area inflation in real time in this turbulent period, we find that a fat-tailed BVAR produces more accurate forecasts than a Gaussian BVAR. Our results also hold when performing a forecast evaluation over a longer period (but in a pseudo-real-time fashion due to data constraints) and are confirmed for a large set of advanced economies. Yet, in abnormal times unrecorded in the available estimation sample, it is hard to trust forecasts based solely on exploring historical regularities. We argue that these times are best suited to add information from outside the model to inform purely model-based results. Lenza and Primiceri (2020), Primiceri and Tambalotti (2020), also incorporate some kind of off-model information to produce more plausible forecasts during the pandemic.

We find that, indeed, including off-model information brings sizeable gains to forecasting inflation during the pandemic. We make use of the information included in the ECB Survey of Professional Forecasters (SPF), first by looking at 'hard conditional' forecasts where we impose the expected GDP path (together with a constant exchange rate and oil price futures), and second by looking at 'soft conditional' forecasts where we tilt the unconditional forecasts towards expectations of inflation and growth from the SPF. Tilting towards SPF expectations can be a valid way to provide the model with some degree of informed judgement, as professional forecasters do not solely rely on models (which might have been affected by COVID-19 observations), but use a substantial degree of judgement when forming their beliefs about the future. Also, there is evidence that such tilting improves inflation forecasting in certain challenging times, such as in the post-Great Recession period, as a way of indirectly accommodating structural changes (see Tallman and Zaman (2020) and Ganics and Odendahl (2021)). Both the considered hard and soft conditional inflation forecasts would have been superior to unconditional ones in a standard Gaussian BVAR. Compared to these, the BVAR with fat-tailed errors largely loses its comparative advantage, suggesting that, indeed, meaningful off-model information makes a difference in forecasting in a standard BVAR.

Notwithstanding our documented forecast gains over a standard Gaussian BVAR, also in our best models, in the first phase of the pandemic, inflation tended to surprise on the downside, and subsequently it turned out to be notably higher than foreseen.[2] The puzzling behaviour of inflation during the pandemic is also reflected in the marked revisions in short-term inflation expectations of professional forecasters. Throughout 2020, analysts lowered their inflation expectations, particularly for the short term. Part of this revision was driven by oil price developments. In addition, in the perceptions of SPF panellists, demand factors were largely considered dominant relative to possible supply-side and scarcity effects. In the second year of the pandemic, professional forecasters, who revised their near-term outlooks quite markedly, were caught off guard by the extent of the inflationary pressures that global supply-side bottlenecks and increases in input costs created. This pattern in the revision of expectations is also valid for other advanced economies, such as the U.S. (see the discussion in Meyer et al. (2021)).

The rest of the paper is structured as follows. Section 2 discusses parameter instability in VAR models when the abnormal COVID-19 observations are added to the sample. Section 3 assesses the impact of this parameter instability on unconditional and conditional inflation forecasts, and Section 4 conducts a formal forecast evaluation for euro area inflation during the pandemic. Section 5 explains why multi-equation models (such as VARs) are better equipped to deal with the impact of abnormal observations than single-equation models (such as Phillips curves), and Section 6 concludes.

2. The impact of the COVID-19 shock on estimates within VAR models

In this section we consider one of the most popular models employed in macroeconomics, namely Sims' vector autoregression (VAR) model. For our analysis, we use the small-scale VAR that was employed to understand the drivers of euro area inflation by Bobeica and Jarociński (2019). Fig. 1 shows the data over the sample 1980:Q1–2021:Q2. There is some variation during the COVID-19 crisis in the considered variables, but nothing unprecedented, with the exception of the real GDP and the euro area share in world GDP. The existing literature dealing with the impact of the pandemic on time series modes has focused on the U.S., where some economic variables, in particular the unemployment rate, have recorded formidable spikes. Yet, we show that the variation recorded by the euro area variables, even if less pronounced than in the U.S. case, is large enough to distort the parameter estimates in a standard BVAR.
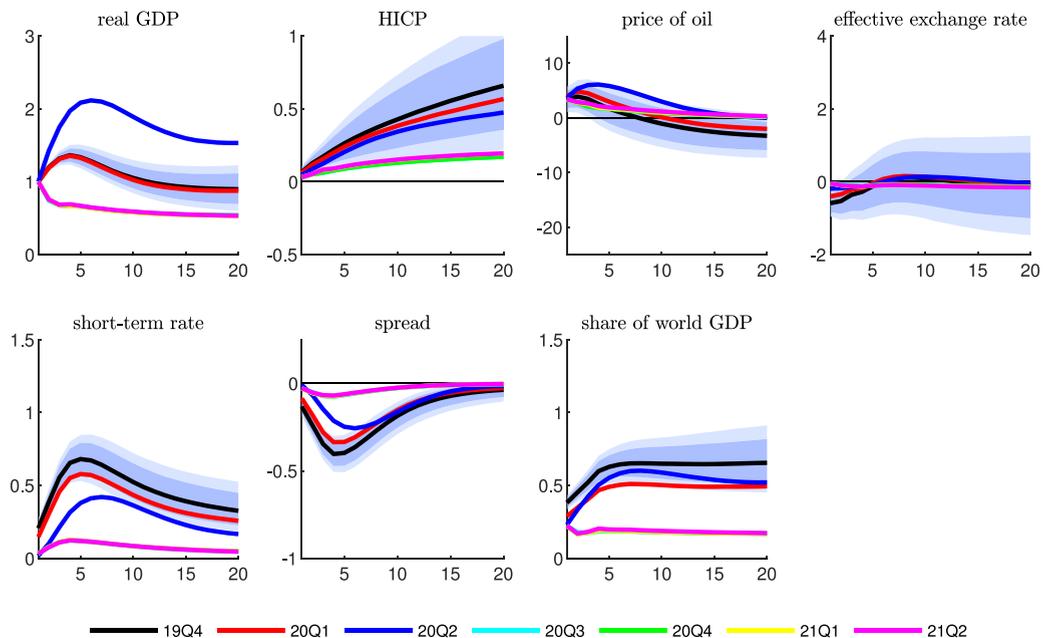
Fig. 2 shows impulse response functions (IRFs) to a one-standard-deviation shock in real GDP estimated over an expanding window until 2019:Q4, 2020:Q1, …, 2021:Q2, starting in 1980:Q1.[3] Estimates are obtained using a Sims

---

[2] In 2020, the sharp fall in oil prices was a result of global lockdowns and not captured by oil price futures. In 2021, apart from the economic drivers, a set of statistical effects (e.g. changes in the weights of the consumption basket and the reversal of a value-added tax decrease in Germany) not captured by our model significantly increased inflation.

[3] We do not attach any structural interpretation to this shock, but merely use it to study the dynamic properties.
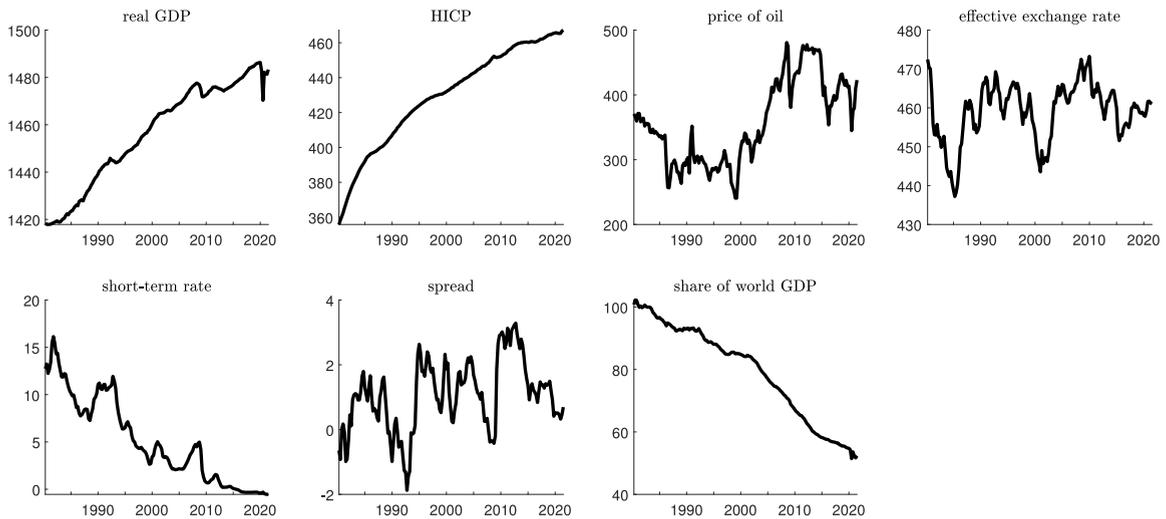
Fig. 1. Data set for the baseline small-scale euro area BVAR. Note: Real GDP, HICP, price of oil, and nominal effective exchange rate are in log-levels times 100; short-term rate, 10-year government bond spread, and share in world GDP are in percentage points.

Fig. 2. Impulse response functions in a small BVAR with Gaussian errors. Note: Impulse response functions to a one-standard-deviation shock in real GDP from a BVAR with Gaussian errors using a Sims and Zha prior. Coloured thick lines are median estimates, and the dark (light) blue area is the 68% (90%) credible interval for the estimation window until 2019:Q4.

and Zha (1998) prior; which we label as 'Sims and Zha prior'.[4] The inclusion of the COVID-19 observation substantially affects the IRF estimates as compared to the pre-COVID-19 times. Specifically, with the inclusion of the 2020:Q2 observation, the IRF of real GDP changes substantially, first it becomes more persistent, afterwards

it subsides. The transmission of the shock to inflation gradually weakens and appears to be stabilising at lower levels; the reaction of inflation to the generic GDP shock is flattened. The impact on financial variables is also dampened. This illustrates that parameter estimates are sensitive to unusually large innovations in a standard Gaussian BVAR, which is a major shortcoming for analysis in times of COVID-19, even after the large swings in real activity have passed.

The prior choice makes a difference, with more informative priors having an upper hand in delivering more stable coefficients once the COVID-19 observations are

---

[4] We construct dummy observations along the lines of Bańbura et al. (2010) and Doan et al. (1983), which reflect Minnesota-type shrinkage on VAR coefficients coupled with two sum-of-coefficients restrictions. For estimation, we use dummy observations to construct the moments of the implied normal-inverse-Wishart prior.

added. Lenza and Primiceri (2020), Schorfheide and Song (2020), and Carriero et al. (2021) document using monthly U.S. data that estimated VAR models are explosive when the recent COVID-19 observations are included. While the estimated IRFs in Fig. 2 are sensitive to the inclusion of the COVID-19 observations, they are not explosive for our chosen prior specification. In Online Appendix B, Panel (a) in Fig. B.1 shows that this explosive behaviour is present in the euro area case as well if we adopt a weakly informative prior for the VAR coefficients.[5] This suggests that the design of the prior, the chosen amount of shrinkage, as well as the data set at hand dictate whether a Bayesian VAR model becomes explosive once COVID-19 observations are included in the sample. Moreover, we also explore whether a higher degree of shrinkage applied to our baseline Sims and Zha prior dampens the impact of COVID-19 on the change in the VAR coefficient estimates; we label this the 'strong Sims and Zha prior'.[6] Panel (b) in Fig. B.1 shows that the higher degree of shrinkage mechanically disciplines the coefficients in the VAR to some extent, but there is still notable instability in the IRFs once the COVID-19 period is included in the sample, which stems from a change in residual correlations.[7] Yet, imposing a very dogmatic prior comes with the drawback of silencing the information coming from the data.

2.1. Solutions for estimation within a VAR

2.1.1. Accommodating COVID-19 observations via alternative error structures

Several proposals on how to tackle the problem of changing parameters due to the COVID-19 observations have been put forward in the literature.[8] The approach taken by Lenza and Primiceri (2020) and Carriero et al. (2021) have in common the belief that macroeconomic relationships which held in the past are assumed to be still informative. That is, they propose downweighing the abnormal observations by allowing the variance of the residuals to adjust quickly to the COVID-19 observations.

Lenza and Primiceri (2020) propose an explicit volatility model for the residuals corresponding to the COVID-19 observations. Their solution is tractable and easy to understand, but requires specifying exactly when the abnormal observations start and estimating the decay of volatility based on few observations (especially when working with quarterly data). Carriero et al. (2021) propose a more flexible solution that allows for variable-specific outliers in the volatility process, leaving fewer choices to the modeller. Here volatility is a function of its

past values, but it is also allowed to jump with abnormal observations with a certain probability. One possible limitation of this model class is related to the sensitivity to the ordering of variables (see Primiceri (2005) and Cogley and Sargent (2005)). Specifically, when volatility changes idiosyncratically, estimates of the time-varying covariance matrix may be extremely sensitive to the ordering of variables, due to the structure imposed by the triangular factorisation of the time-varying covariance matrix in their model, see Hartwig (2020).

Our proposal does not involve making ad hoc choices and is thus more in line with the approach of Carriero et al. (2021). We propose to abandon the Gaussian error structure and assume a multivariate t-distribution for the errors instead; see Ni and Sun (2005) and Chan (2020). Intuitively, extremely large shocks are more likely to occur under the Student's t-distribution, as it has fatter tails than the Gaussian one. Consequently, not all observations are equally informative for the parameter estimates, and the impact of rare observations such as those related to the COVID-19 pandemic are downweighed, which stabilises the relationship between variables. The idea that the data might favour models with errors that haver fatter tails is inspired by the literature emerging after the Great Recession (see Chan (2020), Chiu et al. (2017), Cúrdia et al. (2014)), when Gaussian models failed to properly account for the strong variation in some variables; we show that t-distributed errors work also during the pandemic. This solution shares the spirit of the aforementioned ones in the sense that the COVID-19 observations are downweighed by some form of conditional variance within the estimation framework. Similar to Lenza and Primiceri (2020), we treat the change in shock volatility as being common across all variables instead of considering variable-specific volatility as in Carriero et al. (2021). This is a simple and tractable assumption but neglects potential change of the residual correlations during the pandemic. As opposed to the latter paper, in our VAR, residuals have constant second-order moments (see Appendix A for more technical details on the VAR with t-errors).[9] Compared to standard stochastic volatility models, we argue that Student's t-distributed errors—or other fat-tailed error distributions—are better suited during the pandemic, as the COVID-19 observations may inflate the volatility process for too long in widely employed stochastic volatility models (without outlier correction or t-distributed errors). This may lead to rather imprecise density forecasts; see Carriero et al. (2021) and Hartwig (2021).[10]

Fig. 3 shows the counterpart to Fig. 2 under the assumption that the errors in the Bayesian VAR are multivariate t-distributed. In general, the IRFs are much more

---

[5] The weakly informative prior is calibrated by centring the VAR coefficients around zero and setting the variance of all coefficients to one, except for the constant, which we set to 10. The scale of the inverse Wishart distribution is set to the identity matrix.

[6] Specifically, we divide the prior variances for the coefficients by a factor of four. A higher level of shrinkage is typically recommended for very large VARs with over 100 variables; see Bańbura et al. (2010).

[7] For brevity, we do not report changes in the residual correlation structure for our considered VAR models.

[8] See Carriero et al. (2021), Huber et al. (2020), Lenza and Primiceri (2020), and Ng (2021) in the context of VARs, and Antolín-Díaz et al. (2021) for dynamic factor models.

---

[9] Antolín-Díaz et al. (2021) suggest that the typical comovement between different variables changed with the pandemic, and within a dynamic factor model, the authors distinguish between persistent increases in aggregate volatility and transitory ones specific to individual variables. We do not account for such change explicitly; indirectly, while conducting conditional forecasting, we also inform the comovement between certain variables.

[10] Inference about macroeconomic volatility also changes, as the COVID-19 observations inflate the estimated error volatility of the stochastic volatility process due to the ex post smoothing of the spike during the pandemic.
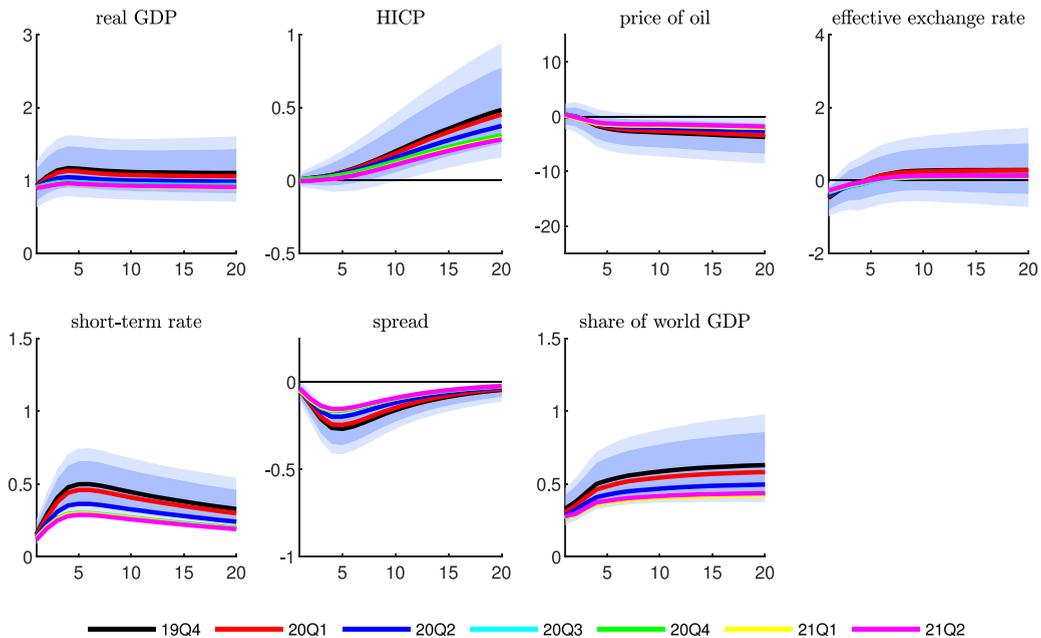
Fig. 3. Impulse response functions in a small BVAR with fat-tailed errors. Note: Impulse response functions to a one-standard-deviation shock in real GDP from a BVAR with fat-tailed errors using a Sims and Zha prior. Coloured thick lines are median estimates, and the dark (light) blue area is the 68% (90%) credible interval for the estimation window until 2019:Q4.

stable across different estimation samples compared to their Gaussian version.[11] Thus, this solution allows for the impact of COVID-19 observations to be soaked up by the heavy tails of the VAR residuals, leaving parameter estimates largely unaffected.[12] Zooming in on inflation, for the pre-COVID-19 sample, the reaction of inflation to real GDP is somewhat weaker in the VAR with fat-tailed errors compared to the one with Gaussian residuals. When adding the COVID-19 observations, the response of inflation to a shock in real GDP also weakens, but to a lesser extent than in the Gaussian VAR case.

We estimate the Bayesian VAR with t-errors by making use of an equivalent scale mixture of normal representations of the VAR model (see Appendix A). This estimation technique yields, as a by-product, an interesting latent variable, which we label *lambda*.[13] This variable reflects the common volatility of the VAR residuals and is independently inverse-gamma distributed. As such, *lambda* is a noisy variable that can be interpreted as the high-frequency common volatility of the data (Chiu et al.,

2017) or used as an outlier detection tool (Jacquier et al., 2004).[14,15]

Panel (a) of Fig. 4 depicts the latent high-frequency volatility in our data set alongside a measure of macroeconomic uncertainty.[16] *Lambda* spiked during the Great Recession and sky-rocketed during the COVID-19 crisis, capturing the fact that there was abnormal variation or statistical outliers in the data. The timing of these spikes coincides with the peaks in measured macroeconomic

---

[11] The IRFs of the fat-tailed VAR are stable under the weakly informative prior and under the strong Sims and Zha prior; see Panels (a) and (b) in Fig. B.2 in Online Appendix B.

[12] The dynamic propagation of the shock is somewhat dampened due to the revised prior distribution—which implies a tighter prior for some variables—in the expanding sample estimation; see also Hartwig (2021). Fig. B.3 shows that under a pre-COVID-19 Sims and Zha prior, the IRFs of the Gaussian model become even more unstable (Panel (a)), while those of the fat-tailed model are virtually unchanged (Panel (b)).

[13] In Appendix A, λ_t is defined in the variance scale. In the text, we use $lambda = \sqrt{\lambda_t}$. Technically, this variable mixes multivariate normal distributions such that it mimics the multivariate t-distribution with a particular degree of freedom; see Geweke (1993) and Ni and Sun (2005).

[14] Importantly, *lambda* is used as an estimation device to simulate the fat tails of the t-distribution. It is not to be confused with a traditional model for time-varying volatility such as stochastic volatility or GARCH.

[15] Since $\lambda_1, \ldots, \lambda_T$ are conditionally independent given the model parameters and the data, we sample each of them independently, as in Chan (2020). The information sources used to estimate *lambda* at each quarter are the sum of squared normalised VAR residuals for that quarter (the residuals are conditional on the VAR parameters utilising the entire sample), the number of variables included in the VAR, the degrees of freedom of the t-distribution, and the assumed prior inverse-gamma distribution; see the conditional posterior distribution of *lambda* in Eq. (9) in Appendix A. The influence of this prior specification on the width of the estimated uncertainty band for *lambda* depends on the degrees-of-freedom parameter (see Eq. (5) in Appendix A), which ensures that the errors are t-distributed (see Geweke (1993)). The lower the degrees-of-freedom parameter (ν), the more widespread the prior inverse-gamma distribution of *lambda*. Our chosen prior distribution for the degrees-of-freedom parameter is uninformative about the posterior shape of the t-distribution, as we assume a uniform prior distribution (see Appendix A). Therefore, we are agnostic about the heavy-tailedness of the t-distribution.

[16] We use the U.S. macroeconomic uncertainty index of Jurado et al. (2015) as an imperfect proxy for the euro area because the available Economic Policy Uncertainty index for the euro area derived by Baker et al. (2016) focuses too much on policy and does not reflect the macroeconomic uncertainty embedded in our variables.

(a) Macro (implied) volatility

(b) Degrees of freedom
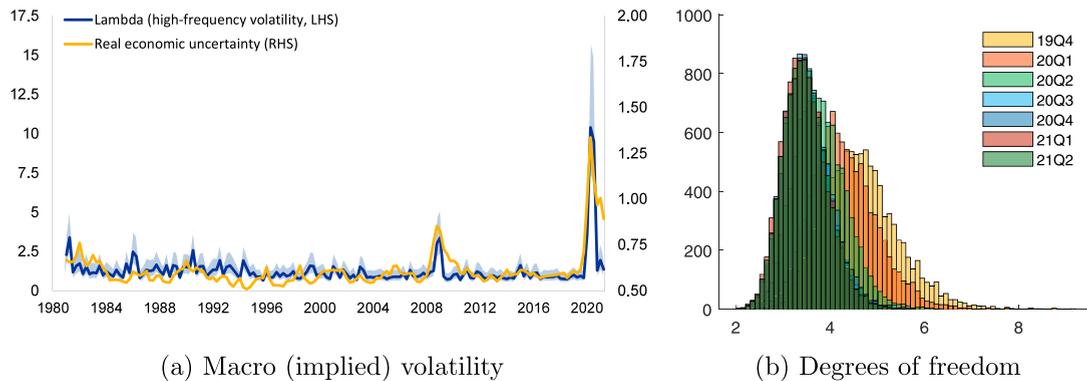
Fig. 4. Macroeconomic uncertainty and tail risk. Note: (a) Posterior median and 90% credible interval of the high-frequency volatility *lambda* in the BVAR with t-distributed errors and Sims and Zha prior. (b) Posterior distribution of the degrees-of-freedom parameter estimated on an expanding sample.

uncertainty.[17] *Lambda* capturing the spikes in macroeconomic uncertainty is another argument in favour of allowing for heavy tails: the model with t-distributed errors identifies these periods to be less informative for the parameter estimates compared to the Gaussian model, which treats all observations equally. Therefore, standard Gaussian models may yield distorted estimates of macroeconomic relationships in turbulent times.

Did the euro area economy exhibit heavy-tailed errors prior to the pandemic? Panel (b) of Fig. 4 shows that the posterior distribution of the degrees-of-freedom parameter (ν) becomes sharper and shifts to the left once the COVID-19 observations are included, indicating a higher tail risk. However, the decline of the posterior median for ν is moderate, from 4.6 in the pre-COVID-19 period to 3.5 in 2021:Q2. Therefore, the tail risk in the euro area economy increases somewhat with the COVID-19 observations.[18] We interpret these findings as evidence that the data not only exhibit fat tails due to the extreme COVID-19 observations but also exhibited substantial tail risk even before, which speaks in favour of the t-distribution assumption as a more appropriate treatment of the data in general.

An alternative way to check whether the data ask for fat-tailed errors is to compare the log marginal likelihood of the Bayesian VARs with Gaussian and t-distributed errors. Table 1 shows these estimates for various prior specifications and on an expanding estimation window.[19]

By comparing the log marginal likelihood figures in Panels (a) and (b) we find decisive evidence in favour of fat tails both before and during the COVID-19 pandemic, as well as across all different prior specifications.

The optimal degree of parameter shrinkage is not affected by the COVID-19 observations. Specifically, both models favour our standard calibration for the Sims and Zha prior over the weakly informative or the strong Sims and Zha prior with a higher degree of parameter shrinkage.[20] This is in contrast to the result reported in Lenza and Primiceri (2020), who document that the Gaussian VAR model likes less shrinkage after COVID-19.

2.1.2. Expanding the data set by adding more variables

As an alternative to fat-tailed errors in the VAR, we investigate whether employing a larger model ensures more stable parameters during the pandemic. Intuitively, in large-dimensional models, the impact of variables with abnormal dynamics might be mitigated. A larger set of data might offer a hedge when it comes to estimation and forecasting in the presence of instabilities; see Rossi (2020). To explore this possibility, we estimate our BVARs with 23 variables used in Bobeica and Jarociński (2019) to understand inflation dynamics in the euro area.[21]

Fig. 5 shows the impulse response functions for real GDP and HICP inflation to a one-standard-deviation shock in real GDP for both the Gaussian and fat-tailed BVAR with more variables, using the two variants of the Sims and Zha prior.[22] The IRFs of the large Gaussian BVAR with a

---

[17] By construction, our measure of high-frequency volatility (*lambda*) is of a short-lived nature, as under t-errors, large shocks are modelled as independent events and are not allowed to be serially correlated. This can explain why during the Great Recession, the proxy for macroeconomic uncertainty was elevated for longer than what *lambda* would imply. The same applies for the case of the COVID-19 shock.

[18] Nevertheless, the log probability of large tail events is rather similar for a degrees-of-freedom parameter between 3 and 5, while for larger values of ν, the tails become substantially thinner. Compared to the Gaussian case, a ten-standard-deviation shock receives a log probability of -50, i.e. 7.69×10^{-21} percentage points, which practically means that such an event will never materialise. Under t-errors with a degrees-of-freedom parameter between 3 and 5, the probability of such an event is roughly comparable to a four-standard-deviation shock under normal errors, which is about 0.0134 percentage points.

[19] For the VAR with t-errors, we use the algorithm in Chan (2020).

[20] We also conducted a more granular grid search on the tightness of the dummy observation prior and found that the t-error model generally favours a lower level of tightness as compared to the Gaussian model.

[21] The VAR includes HICP, real GDP, real private consumption, real private investment, total employment, unemployment rate, capacity utilisation, consumer confidence, PMI, the rest-of-the-world real GDP, nominal effective exchange rate, EUR/USD exchange rate, the oil price in dollars, non-energy commodity prices, U.S. real GDP, U.S. CPI, U.S. short-term interest rate, EONIA, two-year government bond spread, 10-year government bond spread, mortgage bank lending spread, VSTOXX index of financial volatility, and economic policy uncertainty. See Bobeica and Jarociński (2019) for details regarding the data set.

[22] In Online Appendix B, Figs. B.4, B.5, and B.6 show the IRFs for all variables and for all prior specifications.
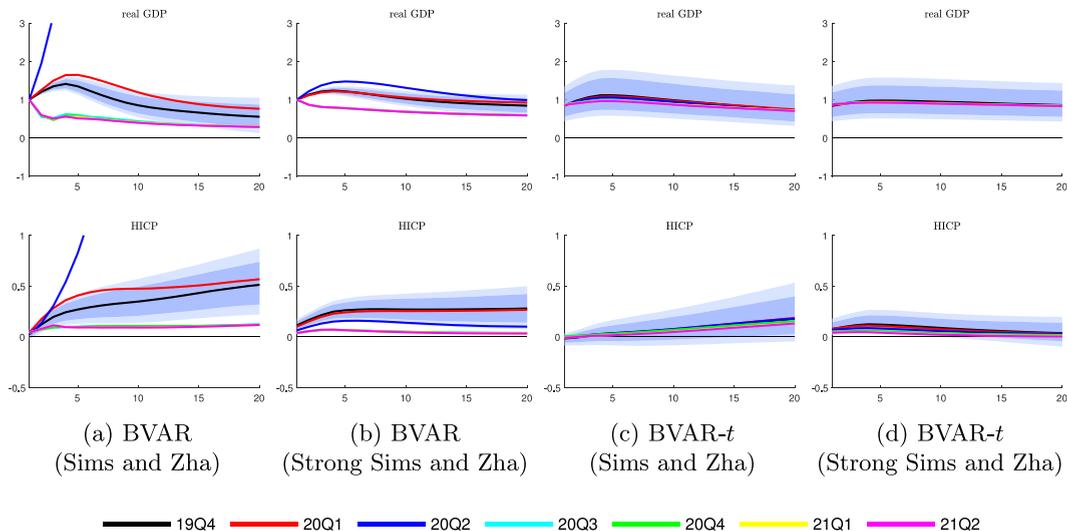
Table 1
Log marginal likelihood for the small BVAR model.

	Weak	SZ	Strong SZ		Weak	SZ	Strong SZ
19Q4	−1748.65	**−1336.97**	−1440.18	19Q4	−1686.01	**−1261.99**	−1381.34
20Q1	−1791.18	**−1376.40**	−1475.32	20Q1	−1707.68	**−1283.12**	−1402.46
20Q2	−1886.03	**−1484.85**	−1585.18	20Q2	−1740.65	**−1315.44**	−1434.05
20Q3	−1966.81	**−1539.32**	−1631.13	20Q3	−1772.91	**−1348.37**	−1463.95
20Q4	−1986.24	**−1550.60**	−1642.30	20Q4	−1786.26	**−1359.18**	−1475.18
21Q1	−2003.93	**−1568.70**	−1658.81	21Q1	−1802.27	**−1374.57**	−1490.36
21Q2	−2016.85	**−1577.25**	−1667.85	21Q2	−1813.50	**−1384.72**	−1501.50
	(a) BVAR with Gaussian errors				(b) BVAR with fat-tailed errors		

Note: The bold figure indicates the maximum log marginal likelihood for each model in a selected estimation window. 'Weak' stands for a weakly informative prior, and 'SZ' stands for Sims and Zha.

(a) BVAR   (b) BVAR   (c) BVAR-*t*   (d) BVAR-*t*
(Sims and Zha) (Strong Sims and Zha) (Sims and Zha) (Strong Sims and Zha)

— 19Q4 — 20Q1 — 20Q2 — 20Q3 — 20Q4 — 21Q1 — 21Q2

Fig. 5. Selected impulse response functions in a large BVAR. Note: Impulse response functions to a one-standard-deviation shock in real GDP from the large BVAR with Gaussian (BVAR) and fat-tailed errors (BVAR-*t*) using various priors. Coloured thick lines are median estimates, and the dark (light) blue area is the 68% (90%) credible interval for the estimation window until 2019:Q4.

Sims and Zha prior exhibit substantial instability as soon as the COVID-19 observations are included and become explosive in 2020:Q2; see Panel (a). The change in the IRFs persists after the initial COVID-19 shock. When imposing more shrinkage, as recommended for larger models, the BVAR coefficients become better behaved; see Panel (b). Compared to Fig. 2, the IRFs of this large BVAR are somewhat more stable when adding the COVID-19 observations. Nevertheless, the covariance matrix is still substantially affected by the extreme observation, which leads to a different propagation of the shock. In contrast, estimates from a large BVAR with heavy-tailed errors are well behaved across different prior specifications; see Panels (c) and (d). Therefore, even in large VARs, fat-tailed errors are a sufficient extension to make the model robust against the extreme variations triggered by the pandemic.

Since the large Gaussian BVAR produces more stable dynamics only with the strong Sims and Zha prior, a natural question is whether the marginal likelihood would also favour this model. Table 2 shows the log marginal likelihood for these VARs. Again, the models with fat-tailed errors yield higher marginal likelihoods, suggesting that they are preferred by the data. At the same time, the standard Sims and Zha prior, which was used for the

small-scale BVAR, is preferred over its strong variant for all estimation samples. This criterion would thus guide us to choose an explosive over a stable model, which is a clear caveat when selecting an appropriate model for inference. This suggests that model diagnostics cannot be blindly trusted in the case when the error distribution is heavily mis-specified—as it is in the case with the abnormal observations during the pandemic (see also Section 4.2). Therefore, we argue that in times of COVID-19, one has to consider other aspects as well when selecting the most appropriate model.

3. The impact of the COVID-19 shock on forecasts within VAR models

It is difficult to ascertain what works in forecasting during the pandemic, as there are only a few observations available as yet. To provide some guidance for inflation forecasters, we first investigate how unconditional and conditional forecasts are affected by the change in parameters previously documented. Then we study how our proposed models fare in a real-time forecasting exercise during the pandemic. We also check whether our extension to the standard VAR model of allowing for fat-tailed

Table 2
Log marginal likelihood for the large BVAR.

	Weak	SZ	Strong SZ		Weak	SZ	Strong SZ
19Q4	−6480.35	**−3756.80**	−3918.06	19Q4	−5915.25	**−3485.19**	−3695.15
20Q1	−6615.97	**−3864.18**	−4021.19	20Q1	−5982.10	**−3539.20**	−3751.03
20Q2	−6756.28	**−3978.87**	−4151.69	20Q2	−6062.13	**−3608.26**	−3810.56
20Q3	−6889.06	**−4067.78**	−4233.83	20Q3	−6152.90	**−3717.54**	−3905.92
20Q4	−6964.35	**−4114.55**	−4298.21	20Q4	−6190.53	**−3766.41**	−3957.56
21Q1	−7049.20	**−4151.13**	−4334.66	21Q1	−6299.66	**−3807.15**	−4000.44
21Q2	−7134.59	**−4199.77**	−4388.15	21Q2	−6323.87	**−3852.67**	−4045.27
	(a) BVAR with Gaussian errors				(b) BVAR with fat-tailed errors		

Note: The bold figure indicates the maximum log marginal likelihood for each model in a selected estimation window. 'Weak' stands for weakly informative, and 'SZ' stands for Sims and Zha.

errors affects unconditional forecasts before and after the pandemic. Finally, we show in a real-time setup that bringing off-model information to our VAR model would have improved the inflation forecast during the pandemic.

3.1. Impact of COVID-19 observations on unconditional forecasts

The change in parameters generated by the COVID-19 observations in traditional Gaussian VAR models has an impact on the forecast paths as well. Focusing again on inflation, Fig. 6 shows the unconditional forecasts starting 2021:Q3 for the BVAR models previously described. In producing these forecasts, the data and the starting date of the forecast are the same. The only thing that differs is the sample on which the coefficients have been estimated. In other words, the change in the forecast path solely reflects revisions in VAR parameter estimates. Panels (a) and (b) show that in the case of the small-scale VAR, unconditional forecasts for both a weakly informative prior and the Sims and Zha prior exhibit visible variation once the COVID-19 observations are added to the estimation sample. Specifically, the short-run and long-run forecasts for inflation bounce down when the abnormal 2020:Q2 observation is included in the sample (blue line) and up again when including more quarters over the pandemic period. The effect of changing parameters on the forecast path is more pronounced for the weakly informative prior. In contrast, unconditional forecasts are only mildly affected in the BVAR with fat-tailed errors (Panel (c)) or in the large-scale Gaussian BVAR with a strong Sims and Zha prior (Panel (d)).

Interestingly, the projected paths of inflation are quite similar for both the small-scale fat-tailed BVAR and the large-scale Gaussian BVAR. With actual data up to 2021:Q2, both models predict that inflation would eventually normalise after the spike in 2021. The forecast bands are somewhat wider for the fat-tailed BVAR than for the large-scale Gaussian model. Also, the bands under the fat-tailed BVAR become wider (hardly visible in the figures) as soon as the COVID-19 observations are included in the estimation sample. This is due to the fact that the estimated degrees-of-freedom parameter declines, which increases the unconditional variance of the forecast errors.

3.2. Impact of COVID-19 observations on conditional forecasts

Most policy institutions rely on some type of conditional forecast in order to form an opinion on the likely future inflation path. The previous section showed that the unconditional forecast paths are sensitive to the change in parameters brought about by the inclusion of the pandemic observations. A natural question arises whether the inclusion of off-model information while producing conditional forecasts mitigates this forecast instability problem.

'Hard conditioning' using external projections

We illustrate how the change in coefficients brought about by the COVID-19 shock affects conditional forecasts for euro area inflation when taking into account external projections. Specifically, we condition on a set of off-model projected paths for the following variables: real GDP growth based on the Survey of Professional Forecasters,[23] the price of oil based on futures contracts,[24] and the nominal effective exchange rate, which is kept constant over the projection horizon, mimicking a common practice in policy circles. The path of the variables is strictly imposed using the algorithm of Waggoner and Zha (1999) without the updating step. Again, as in the exercise showing unconditional forecasts, the aim is to study the change in the forecast path coming solely from the revisions in VAR parameters; the starting date of the forecast, the actual data, and the conditioning data are the same across all the coloured lines in the figure.

Fig. 7 shows that conditional forecasts using the small-scale Gaussian BVAR are now less sensitive to the change in parameter estimates over the expanding estimation window for all prior specifications, even for the weakly informative one; see Panels (a) and (b). Including the 2020:Q2 information does not result in an explosive or a dramatically changed forecast anymore. Thus, the problem of unstable forecasts can be mitigated by conditioning on a sufficient set of variables that exhibit well-behaved future paths and are strongly interlinked with the target variable. Similarly, the conditional forecasts of the

---

[23] We use the one-, two-, and five-year-ahead real GDP growth expectations of the ECB SPF 2021:Q3 round (i.e. the latest available when the ECB projected inflation as of 2021:Q3) and we interpolate linearly for missing quarters.

[24] As of 16 August 2021, corresponding to the cut-off of the ECB inflation projections starting in 2021:Q3.
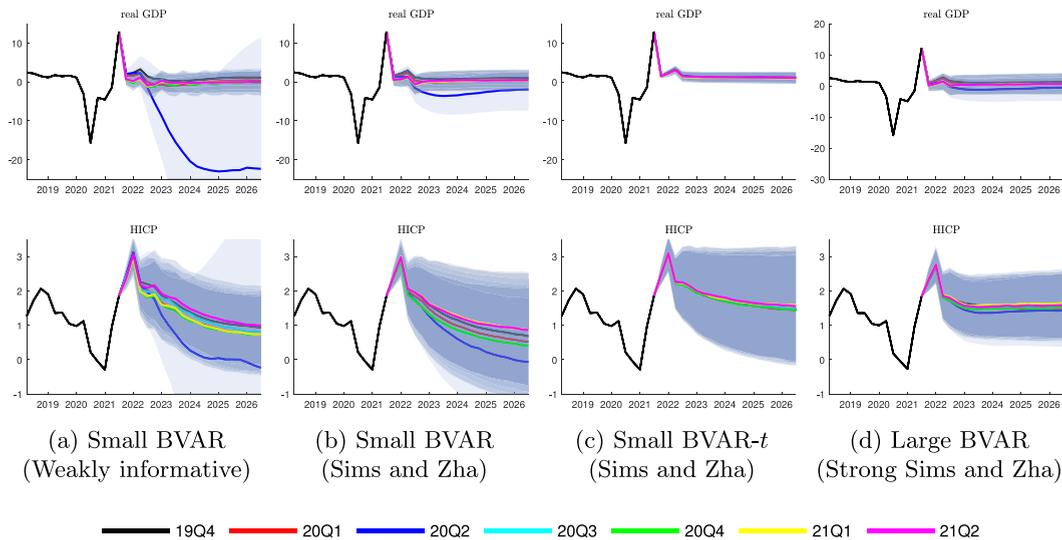
Fig. 6. Unconditional forecasts starting in 2021:Q3. Note: The forecast starts in 2021:Q3; parameters are estimated on an expanding estimation window. Coloured thick lines are median estimates, and the (overlapping) light blue areas are the 68% credible interval for each respective estimation window.
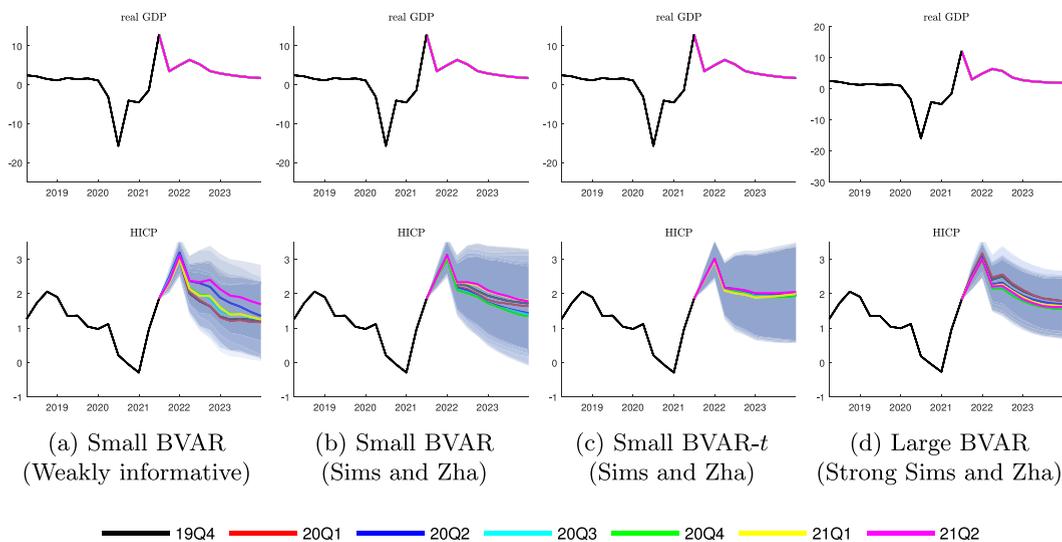
Fig. 7. 'Hard conditioning' forecasts starting in 2021:Q3. Note: Forecasts conditional on real GDP growth (as in the 2021:Q3 ECB Survey of Professional Forecasters), oil price futures (as of 16 August 2021), and a constant nominal effective exchange rate; parameters are estimated on an expanding window. Coloured thick lines are median estimates, and the (overlapping) light blue areas are the 68% credible interval for each respective estimation window.

BVAR with fat-tailed errors are not sensitive to the inclusion of the pandemic observations in the estimation sample (Panel (c)). This model yields a slightly higher medium-term inflation outlook than the Gaussian one, and compared to its unconditional version, inflation is now expected to stabilise at somewhat higher levels. Conditional forecasts are also stable for the large BVAR (Panel (d)). Overall, this suggests that conditioning on a relevant information set alleviates the problem of in-sample parameter instability.

'Soft conditioning' using expectations of the Survey of Professional Forecasters

Using the previously described methodology of producing conditional forecasts, the imposed future paths of the conditioning variables always hold exactly. However, in particularly uncertain times, one may want to consider off-model information with some degree of uncertainty around it. This is known as 'soft conditioning'; see the relative entropy method proposed by Robertson et al. (2005). The main idea of this approach is to derive a new predictive distribution 'tilted' to the mean or median of the distribution of the off-model information (and possibly

also to some quantile values) that is as close as possible to the initial unconditional forecast distribution (which might not be the case in the approach by Waggoner and Zha (1999)).

In principle, one can apply this procedure to produce forecasts for any variable included in the VAR by 'tilting' its unconditional forecast to some off-model information regarding its future path. Here we illustrate this approach by considering expectations of the Survey of Professional Forecasters (SPF) for the euro area.[25] This alternative approach alleviates the issue of having to come up with a set of strong assumptions, as in Primiceri and Tambalotti (2020), interpolation issues in case forecasts are available only for some quarters, and the problem of imposing the future path of many variables included in the model. It can also constitute a valid crosscheck for hard conditional forecasts.

We produce soft conditional forecasts by considering the median SPF inflation expectations for one, two, and five years ahead, but also long-term real GDP growth expectations; see Fig. 8. Panel (a) shows that tilting the unconditional inflation forecasts to median SPF values for inflation and GDP yields inflation forecasts which are robust to the parameter change induced by the COVID-19 observations.[26] Nevertheless, these conditions are not sufficient to ensure overall stability in the Gaussian VAR with a weakly informative prior. Furthermore, Panel (b) shows forecasts when the unconditional distribution is tilted to five-year-ahead inflation only. Thus, whereas long-run expectations are crucial to ensure stability, short-run expectations are important to ensure an informed profile of the future inflation path.

4. Forecasting inflation during the pandemic

The previous section documented the sensitivity of unconditional and conditional forecasts with respect to changing VAR parameters during the pandemic. In this section, we study how our proposed models would have fared when forecasting inflation in real time. We also investigate the comparative forecasting advantage of the fat-tailed BVAR model by looking at a longer period of time and at a larger set of countries in a pseudo-out-of-sample forecast evaluation.

4.1. Real-time forecast evaluation

Would a forecaster have been better equipped by using our fat-tailed BVAR compared to using a standard Gaussian model when trying to forecast inflation during the pandemic? We investigate this question by looking at our benchmark small-scale BVAR.[27]

Table 3 shows the root mean squared forecast error (RMSFE) for inflation over the period from 2020:Q1–2021:Q2. Given the very short period, the results can only provide some hints towards promising avenues to explore when forecasting inflation during the pandemic. The first column shows the figures corresponding to unconditional forecasts within a standard Gaussian BVAR. Compared to this model, it would have paid off to employ a model with fat-tailed errors, as doing so would have produced a smaller RMSFE over the pandemic period (second column). As an additional benchmark to the Gaussian BVAR, we employ the same VAR model augmented with dummy variables for each quarter during the pandemic, which we label *dummy BVAR*.[28] This can appear to be a simple fix to the problem of changing parameters when including the COVID-19 observations in the sample, as it basically fixes the parameter estimates to the pre-pandemic period (as done in Schorfheide and Song (2020)).[29] This benchmark has the advantage that the dummy variables are able to absorb the variation induced by the COVID-19 observations, leaving the remaining parameters of the BVAR largely unchanged. Nevertheless, it has the disadvantage that, unlike the BVAR with fat-tailed errors, it requires several ad hoc choices from the modeller, such as when pandemic observations start and when the pandemic will be over (which would be an indication to stop including dummy variables). Also, it is not a priori clear how an informative prior should be formulated for the dummies such that economic relations might be affected by the COVID-19 shock during the pandemic. Table 3 shows that unconditional forecasts based on a dummy BVAR would have outperformed the ones based on an unadjusted Gaussian BVAR, but the BVAR with fat-tailed errors beats this benchmark—especially at longer horizons.

Including off-model information brings sizeable gains to forecasting inflation during the pandemic. Unconditional forecasts based on models that only look at past historical patterns of the data are hard to trust during the pandemic, as COVID-19 is an unprecedented macroeconomic shock. It has complex effects on the demand- and supply-side of the economy (Guerrieri et al., 2020)), highly heterogeneous impacts across sectors and agents, and unknown behavioural consequences. For a stylised representation of the channels through which the COVID-19 shock affects inflation; see Fig. B.7 in Online Appendix B. As such, these times are ideally suited to adding information from outside the model to inform purely model-based forecasts, as also argued by Primiceri and Tambalotti (2020). Both the considered hard and soft conditional inflation forecasts would have been superior to unconditional ones in a Gaussian BVAR (columns 4 and 7, compared to column 1). Compared to these, the BVAR

(a) Small BVAR (Weakly informative)

(b) Small BVAR (Sims and Zha)

(c) Small BVAR-*t* (Weakly informative)

(d) Small BVAR-*t* (Sims and Zha)

19Q4   20Q1   20Q2   20Q3   20Q4   21Q1   21Q2
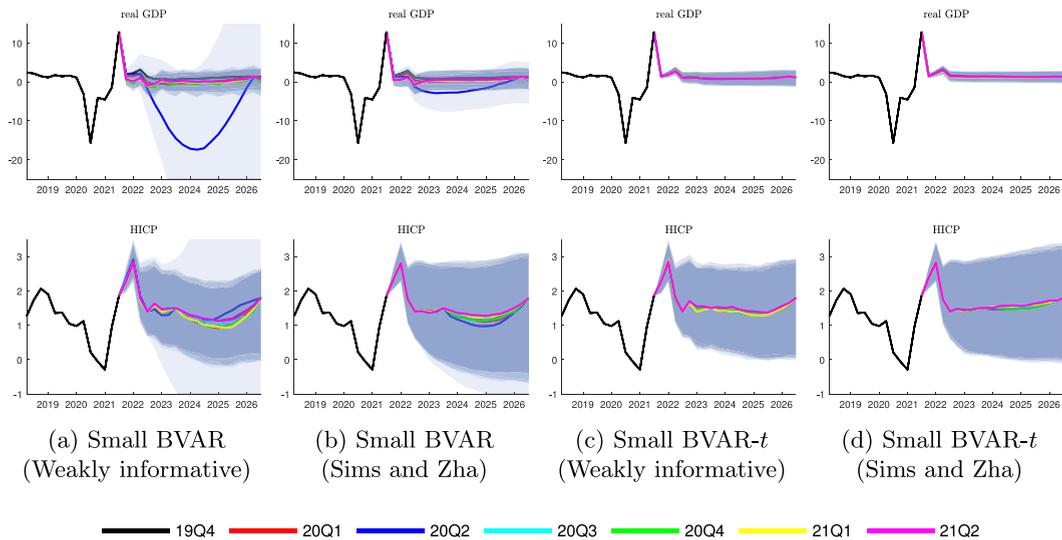
(I) Tilting to SPF expectations for inflation (short- and long-run) and GDP growth (long-run)

(e) Small BVAR (Weakly informative)

(f) Small BVAR (Sims and Zha)

(g) Small BVAR-*t* (Weakly informative)

(h) Small BVAR-*t* (Sims and Zha)

19Q4   20Q1   20Q2   20Q3   20Q4   21Q1   21Q2
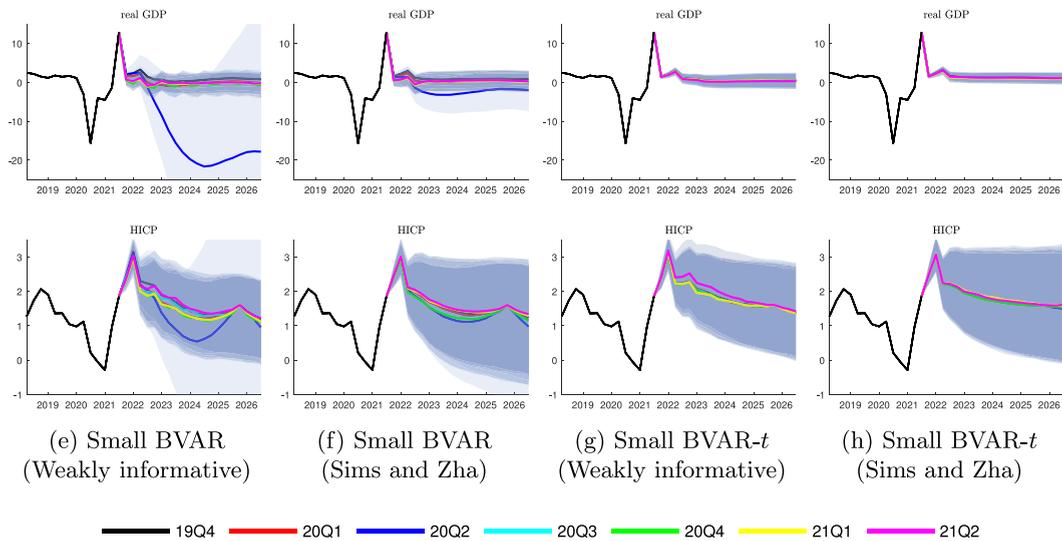
(II) Tilting to long-run SPF inflation expectations

Fig. 8. Conditional forecasts based on tilting to SPF expectations. Note: Forecasts tilted to SPF expectations for inflation and real GDP growth starting in 2021:Q3 and using parameters from an expanding window estimation. Coloured thick lines are median estimates, and the (overlapping) light blue areas are the 68% credible interval for each respective estimation window.

Table 3
Real-time forecast evaluation for euro area inflation: RMSFE.

	Unconditional			Hard conditioning			Soft conditioning		
	BVAR	BVAR *t*	BVAR *D*	BVAR	BVAR *t*	BVAR *D*	BVAR	BVAR *t*	BVAR *D*
$h = 1$	0.61	0.58	0.60	0.47	0.50	0.62	0.56	0.56	0.55
$h = 2$	1.07	0.98	1.02	0.84	0.91	1.05	0.88	0.91	0.88
$h = 3$	1.41	1.13	1.17	0.93	0.91	1.01	0.99	0.98	0.96
$h = 4$	1.66	1.19	1.25	1.01	0.96	0.94	1.01	1.01	1.00

Note: RMSFE during COVID-19 pandemic for HICP inflation. BVAR denotes a standard Gaussian BVAR, BVAR *t* is the fat-tailed BVAR, and BVAR *D* is the BVAR augmented with dummy variables each quarter over the COVID-19 period. The evaluation sample covers the period from 2020:Q1–2021:Q2.
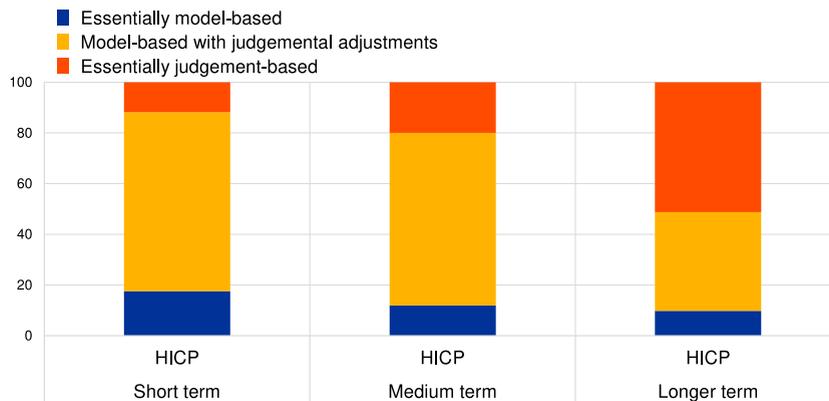
Fig. 9. Information used by professional forecasters of inflation at various forecast horizons. Note: Professional forecasters' response to the question: To what extent are your point forecasts model- or judgement-based? (percent of responses). *Source:* ECB (2019)

with fat-tailed errors largely loses its comparative advantage, suggesting that, indeed, meaningful off-model information makes a difference in forecasting in a standard VAR, as discussed in the previous section.

While the forecast gain from bringing in meaningful off-model information to the VAR is notable, our exercise suggests that it does not make a big difference whether one does this via hard conditioning on several variables (including real GDP, as in the SPF) or via soft conditioning on the inflation expectations of professional forecasters. It is clear in any case that the expectations of professional forecasters can discipline econometric models in turbulent times. This is also because professional forecasters do not solely rely on models, which have likely been adversely affected by COVID-19 observations, but use judgement when submitting their forecasts (see the discussion in de Vincent-Humphreys et al. (2019)). In a special ECB SPF questionnaire (see ECB (2019)), respondents indicated that a sizeable share of their forecast can be attributed to their judgement, especially when it comes to long-term forecasts; see Fig. 9.

Tilting model-based inflation forecasts to SPF expectations has been investigated recently, but for somewhat different reasons. Ganics and Odendahl (2021) focus on the euro area and show that survey forecasts can help mitigate the effects of structural breaks on the forecasting performance in a VAR. They also find that professional forecasters are better at forecasting than a standard Bayesian VAR model around the two recent euro area recessions, as well as the slow recovery thereafter. This is a particularly challenging period for forecasting euro area inflation as reduced-form time series models have a hard time providing accurate forecasts (as also discussed in Bańbura and Bobeica (2020)). Similarly, Tallman and Zaman (2020) show that such tilting produces superior inflation forecasts after the Great Recession in the U.S., acting as a way of indirectly accommodating structural changes (and mitigating misspecification issues in a VAR). Bańbura et al. (2021) document forecast gains for euro area inflation and GDP when an optimal forecast combination from many models is tilted to the more subjective SPF mean. It is worth noting that when tilting

to both first and second moments of the SPF, there is a general worsening of the forecasting performance, so the authors recommend tilting only to the mean.

Fig. 10 shows the real-time predictions of euro area inflation during the COVID-19 crisis using both unconditional and conditional forecasts. Panels (c) and (d) show that incorporating alternative information brings the projection for euro area inflation closer to its actual level, which is consistent with the RMSFE analysis. In addition to the information included in Table 3, it becomes obvious that in the first phase of the pandemic, inflation tended to surprise on the downside, and subsequently, it turned out to be notably higher than foreseen. The puzzling behaviour of inflation during the pandemic is also reflected in the marked revisions to the short-term inflation expectations of professional forecasters (see Fig. 11).

Throughout 2020, analysts lowered their inflation expectations, particularly for the short term. Part of this revision was driven by oil price developments. In addition, in the perceptions of SPF panellists (see ECB (2020)), demand factors were largely considered dominant relative to possible supply-side and scarcity effects. The combination of lower oil price assumptions and the drop in demand led SPF respondents to revise HICP inflation expectations down sharply for shorter horizons, and the downward adjustment also involved expected core inflation. The second year of the pandemic marked a reversal in inflation dynamics. Inflation in the euro area (and in other advanced economies) surged, on the back of the rebound in oil prices and rising cost pressures from supply-side bottlenecks which accompanied the reopening of the economies. Inflation exceeded the expectations of professional forecasters, who revised their near-term outlooks quite markedly (see Fig. 11). This pattern is valid also for other advanced economies, such as the U.S. (see the discussion in Meyer et al. (2021)). At the same time, despite the large swings in near-term inflation expectations, long-term expectations remained fairly stable. This is also reflected in Fig. 10, Panel (d), where soft conditioning disciplines inflation forecasts beyond the near term.
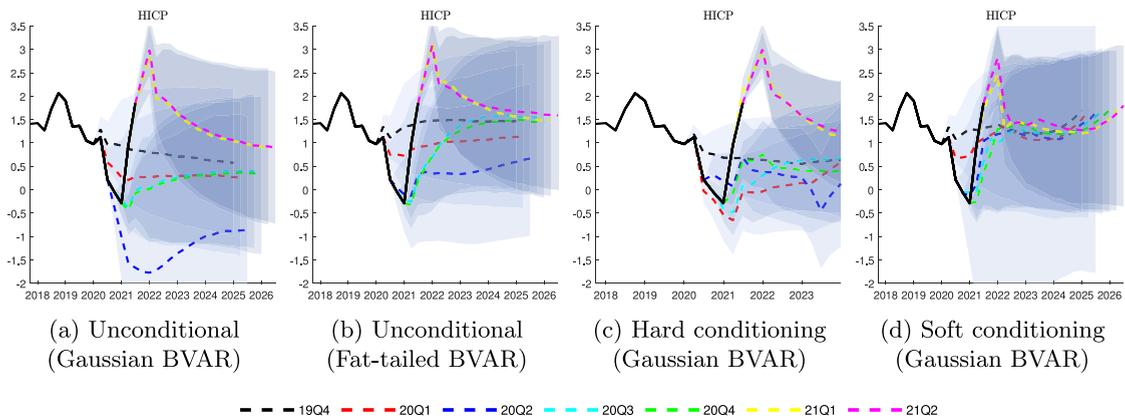
Fig. 10. Real-time inflation forecast. Note: Real-time forecast for inflation estimated over an expanding window. The thick black line is the actual inflation, the coloured dashed lines are median estimates, and the (overlapping) light blue areas are the 68% credible interval for each respective estimation window.
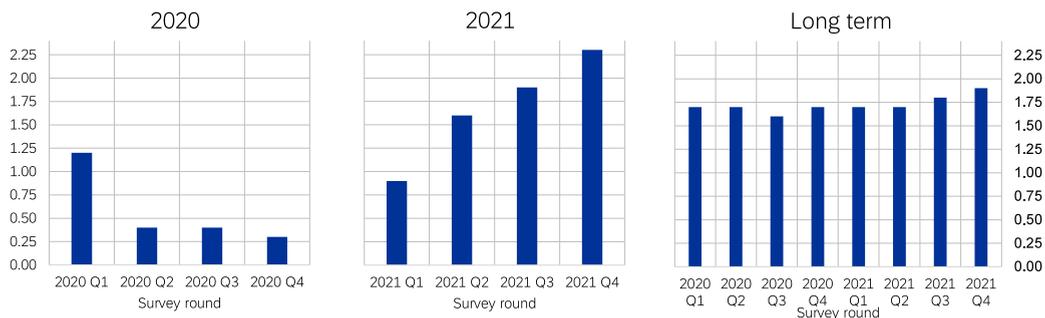
Fig. 11. Revisions in short- and long-term euro area inflation expectations for various forecast horizons. Note: ECB Survey of Professional Forecasters.

4.2. Pseudo-forecast evaluation and cross-country evidence

Our assessment over the pandemic period is that the solution that we propose for disciplining BVAR parameters in abnormal times is also superior to a standard Gaussian BVAR when it comes to unconditional forecasts. Caution is warranted, of course, given the very short available pandemic sample. In order to form a better view of whether this model is a valid solution in general, we conduct a forecast evaluation over a longer sample (starting in 1995), where we compare the fat-tailed BVAR with its Gaussian counterpart. For the euro area, such an exercise can be done only using a pseudo-real-time setup, due to the lack of real-time vintages going so far back into the past.

Specifically, we evaluate the accuracy of unconditional forecasts in the small BVAR with the Sims and Zha prior on the pre-pandemic sample from 1995:Q1 until 2019:Q4, and during the COVID-19 period from 2020:Q1 until 2021:Q2. To do so, we re-estimate each model over an expanding window using 1980:Q1 as a starting date, simulate a forecast, and evaluate the horizons from one quarter to four quarters ahead. The forecast performance is measured in terms of point and density predictions.[30]

For the point prediction, we compute the root mean squared forecast error (RMSFE) using the posterior mean of the predictive distribution. The performance of density forecasts is judged by two commonly used scoring rules: the average continuous ranked probability score (ACRPS), for which the lower the value the better; and the average log predictive likelihood (ALPL), for which the higher the value the better.[31] Even though various authors have argued in favour of either ACRPS or ALPL, Krüger et al. (2021) stress that this choice appears ultimately subjective. Particularly, we consider both scoring rules because they may yield different answers regarding the best performing model in the event of extreme observations; see Bjerregård et al. (2021). The CRPS measures the quadratic difference between the predicted and empirical cumulative distribution function. By construction, this metric rewards observations close to the median and is less sensitive to extreme observations. The log predictive likelihood, on the other hand, evaluates the likelihood of the observation based on its predictive density. As such, extreme observations are particularly punished if the tails of the predictive density are too thin.

Table 4 reports the outcome of this forecast evaluation for real GDP growth and HICP inflation. Prior to the

---

[30] Appendix A provides some computational details for the evaluation metrics.

[31] For recent applications; see e.g., Chan (2020), Clark and Ravazzolo (2015), and Carriero et al. (2021).

Table 4
Pseudo-out-of-sample forecast evaluation.

(a) Real GDP growth

	RMSFE				ACRPS				ALPL			
	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
BVAR	0.52	0.95	1.37	1.75	0.26	0.48	0.71	0.93	−0.83	−1.38	−1.77	−2.07
BVAR t	0.51	0.94	1.36	1.74	0.25	0.46	0.69	0.90	−0.74	−1.24	−1.64	−1.92
					(I) Pre-COVID-19 period: 1995:Q1–2019:Q4							
BVAR	9.03	11.72	13.07	14.79	5.73	7.14	8.22	9.26	−34.09	−63.40	−55.06	−44.43
BVAR t	7.38	8.71	8.90	9.69	5.01	5.61	6.13	7.35	−6.67	−6.72	−7.43	−8.89
BVAR D	7.49	8.93	9.04	9.77	5.05	5.76	6.08	7.04	−61.45	−80.80	−69.11	−63.25
					(II) COVID-19 period: 2020:Q1–2021:Q2							

(b) HICP inflation

	RMSFE				ACRPS				ALPL			
	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
BVAR	0.30	0.54	0.75	0.99	0.16	0.28	0.39	0.52	−0.25	−0.82	−1.12	−1.39
BVAR t	0.30	0.53	0.72	0.96	0.16	0.28	0.38	0.50	−0.22	−0.72	−1.03	−1.33
					(I) Pre-COVID-19 period: 1995:Q1–2019:Q4							
BVAR	0.61	1.06	1.35	1.56	0.32	0.66	0.78	0.80	−1.09	−1.90	−1.68	−1.59
BVAR t	0.58	0.99	1.12	1.14	0.31	0.64	0.71	0.67	−0.86	−1.71	−1.67	−1.58
BVAR D	0.59	1.02	1.16	1.21	0.31	0.63	0.68	0.67	−1.05	−1.82	−1.55	−1.48
					(II) COVID-19 period: 2020:Q1–2021:Q2							

Note: Forecast evaluation metrics prior to and during the COVID-19 pandemic for (a) real GDP growth and (b) HICP inflation. BVAR denotes a standard Gaussian BVAR, BVAR t is the fat-tailed BVAR, and BVAR D is the BVAR augmented with dummy variables for each quarter in the COVID-19 period.

pandemic period, the BVAR with fat-tailed errors produces slightly more accurate point and density predictions for both real GDP growth and HICP inflation as compared to the standard Gaussian BVAR. During the pandemic, the fat-tailed model exhibits a clear upper hand over the standard Gaussian model. It produces substantially more accurate point and density predictions for real GDP growth and HICP inflation, especially at longer forecast horizons. Therefore, the unconditional forecast accuracy of a Gaussian BVAR model suffers particularly when extreme observations distort parameter estimates.

The dummy-augmented BVAR, on the other hand, is a competitive benchmark during the pandemic. Overall, the accuracy of point predictions is at a similar level to that of the fat-tailed BVAR, though the latter is somewhat more accurate at longer horizons. One driving force behind this relative improvement of the fat-tailed BVAR might be that it implies less persistent propagation of macroeconomic shocks as compared to the dummy-augmented Gaussian BVAR. In terms of density predictions, the evidence is more mixed. The ACRPS suggests that the quality of the density predictions is of similar accuracy for both real GDP and HICP inflation. However, the ALPL draws a different picture. Under this metric, the dummy-augmented BVAR produces the worst density forecast for real GDP, while the figures for HICP inflation are comparable. The reason behind this seemingly unintuitive result is that the predictive distribution of the dummy BVAR (excluding the dummies from the forecast exercise) is too far away from the actual realisation of real GDP. In other words, the tails of the predictive distribution are too thin to rationalise these abnormal observations. Updating the VAR parameters helps in this respect, but generates unreasonable

forecasts, as discussed above. Thus, the log predictive likelihood may discard a very competitive forecast model, as the predictive distribution is heavily mis-specified. Since the predictive likelihood and the marginal likelihood are closely connected (Geweke & Amisano, 2010, 2011), this also explains why the marginal likelihood may favour an overall explosive model—specifically, because it is unable to rationalise extreme observations with more stable parameters. Similar messages arise based on the large VAR. The forecast evaluation is presented in Table B.1 in Online Appendix B.

While the pandemic was a global shock, each country reacted differently when it came to containment measures. So one might ask whether the findings for the euro area are also valid for other countries. This is also a way to check the robustness of the finding that the BVAR with fat-tailed errors is a viable alternative to Gaussian BVARs when it comes to inflation forecasting before and after the pandemic. We found that for major euro area countries (France, Germany, Italy, and Spain), and for Canada, the United Kingdom, the U.S., and Japan, the model with fat-tailed errors is a valid alternative tool to a standard Gaussian BVAR when it comes to forecasting inflation. The differences in forecasting performance are marginal in the pre-COVID-19 period, but as soon as the extreme observations kick in, the model equipped with fat-tailed residuals outperforms; see Tables B.2, B.3, and B.4 in Online Appendix B.

5. Modelling the COVID-19 shock with univariate and multivariate t-distributed errors

A relevant question is whether the solution of relaxing the normality assumption of the residuals works only
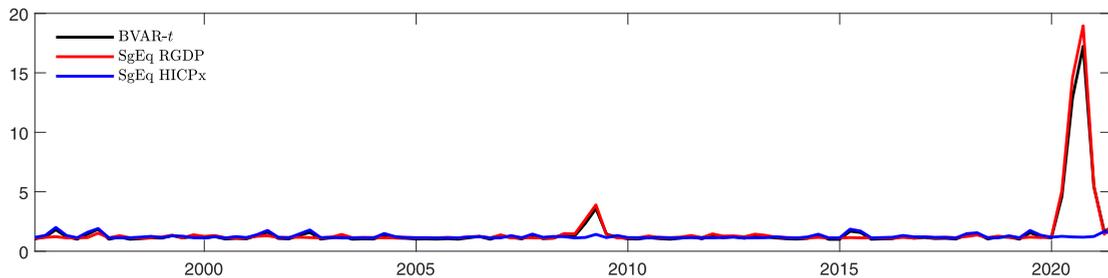
Fig. 12. Implied volatility under multivariate and univariate t errors. Note: Coloured thick lines are posterior median estimates of implied volatility in the following models with t-errors: (i) BVAR-t using one lag (BVAR-t) and the single-equation regression of the VAR model for (ii) real GDP growth (SgEq RGDP) and (iii) inflation (SgEq HICPx).

for multi-equation systems such as VARs, or whether it also works for single-equation models. We show that it actually is the joint multivariate and fat-tailed errors that manage to robustify models against extreme observations. We illustrate this by taking a simple Phillips curve model, a key macroeconomic relationship used to gauge past and future inflation dynamics. In central banking, it is one of the pillar frameworks in thinking about inflation, with reduced-form models often employed to grasp the link between inflation and real activity (see Yellen (2015) and Eser et al. (2020)).

In the single-equation framework, fatter tails are less effective at mitigating the impact of extreme observations, as they primarily capture abnormal variations associated with the dependent variable, but not so much those induced by an exogenous regressor (since its coefficients can always be shrunk to zero to mitigate the impact on the errors). Intuitively, in the multivariate setup, all variables are re-scaled by a common volatility factor, which is informed by the innovations in all endogenous variables, while in the single-equation setup, this volatility factor is informed only by the residuals of the dependent variable. For technical details; see Appendix A.

To illustrate this point, consider the following example with a simple Phillips curve model, using real GDP growth as a measure of slack, its own lag, and no other exogenous regressor (such as external price pressures or inflation expectations):

$$\pi_t = \alpha + \rho \pi_{t-1} + \beta y_{t-1} + \epsilon_t$$

where π_t is the seasonally adjusted quarterly growth rate of inflation excluding energy and food (HICPx), and y_t is the seasonally adjusted quarterly growth rate of the real gross domestic product (RGDP).

We estimate the coefficients using system and single-equation techniques. For the system estimation, we use: (i) a Bayesian VAR with Gaussian errors, (ii) a Bayesian VAR with multivariate t-distributed errors, and (iii) a counterfactual BVAR with exogenous time-varying variance Λ (as detailed below). For the single-equation estimation, we use a univariate version of the three models listed above. As a prior distribution, we assume a diffuse (non-informative) prior for all parameters. We obtain different estimates of λ from the models with t-errors: the BVAR-t (λ_{VAR}) and the single-equation regression of the VAR equation for the output gap (λ_{RGDP}) and HICP

inflation excluding energy and food (λ_{HICPx}), respectively. Subsequently, the posterior medians of these λ's are fed into the model with exogenous time-varying variance to study the impact of allowing for fat tails on the coefficient estimates.

Fig. 12 shows the implied volatility by these λ's. Both λ_{VAR} and λ_{RGDP} spike when the COVID-19 pandemic wreaks havoc on the euro area, whereas λ_{HICPx} remains relatively flat during this period. This illustrates that different information sets may crucially affect these λ's in the multi- and univariate setup with t-errors.

How do these different estimates of λ affect the inference about the coefficients in Eq. (5)? Table 5 shows the posterior mean of the coefficients using the estimation choices described above for (a) the pre-COVID-19 sample and (b) with the COVID-19 observations included in the estimation sample.

Focusing on the pre-COVID-19 sample (Panel (a)), the estimated slope of the Phillips curve β is the same for the Gaussian model estimated as a system or as a single equation (column 'Gaussian'). The slope coefficients of the fat-tailed model (column 'Fat-tailed') are also very similar. However, other coefficients such as the intercept and persistence differ slightly between the two versions. These differences can be explained by the fact that the λ's differ across the two options (multi- versus univariate t-errors), mainly because of differences during the financial crisis; see Fig. 12.

The COVID-19 observations lead to a decrease in the Phillips curve slope by about 80% for real GDP growth; see Panel (b) (columns 'Gaussian' and 'Fat-tailed, SgEq'). This strong parameter revision in the univariate fat-tailed model (but also in the Gaussian model) occurs because the COVID-19 observations are not heavily discounted, since the λ_{HICPx} remains rather flat. In contrast, when β is estimated under multivariate t-distributed errors (column 'Fat-tailed, System'), the slope coefficient changes considerably less, by about 30%. In addition, when estimating this equation with a known λ from the VAR (column 'Gaussian(λ_{VAR})') or the real GDP growth equation (column 'Gaussian(λ_{RGDP})'), the slope coefficients also change by a similar amount as compared to the system estimation, whereas when plugging in λ_{HICPx}, they change substantially (column 'Gaussian(λ_{HICPx})').

Table 5
Simple Phillips curve coefficients with real GDP growth (qoq, sa).

Coeff.	Gaussian		Fat-tailed		Gaussian(λ_{VAR})		Gaussian(λ_{RGDP})		Gaussian(λ_{HICPx})	
	System	SgEq	System	SgEq	System	SgEq	System	SgEq	System	SgEq
α	0.041	0.042	0.029	0.034	0.029	0.029	0.040	0.041	0.038	0.038
ρ	0.945	0.945	0.955	0.951	0.955	0.955	0.951	0.951	0.950	0.950
β	0.031	0.030	0.029	0.028	0.029	0.029	0.029	0.028	0.027	0.027

(a) Simple Phillips curve estimated until 2019:Q4

Coeff.	Gaussian		Fat-tailed		Gaussian(λ_{VAR})		Gaussian(λ_{RGDP})		Gaussian(λ_{HICPx})	
	System	SgEq	System	SgEq	System	SgEq	System	SgEq	System	SgEq
α	0.088	0.090	0.055	0.073	0.056	0.054	0.070	0.069	0.074	0.073
ρ	0.943	0.942	0.952	0.951	0.951	0.952	0.947	0.948	0.950	0.950
β	0.006	0.006	0.020	0.006	0.020	0.021	0.020	0.020	0.005	0.005

(b) Simple Phillips curve estimated until 2021:Q2

Note: Posterior mean of Phillips curve coefficients based on various modelling assumptions. The first top row denotes the class of the error distribution—i.e. Gaussian is normal errors, Fat-tailed is t-errors, and Gaussian($\cdot$) is normal errors with exogenous time-varying variance—and the second row describes whether the equation was estimated as a system (System) or as single equation (SgEq).

6. Conclusions

The parameters of Gaussian VAR models, frequently employed to analyse inflation by central bankers and other practitioners, are strongly affected by the COVID-19 observations and this has a bearing on the inflation forecast path.

We showed that relaxing the assumption of normal errors and allowing them to follow a multivariate t-distribution appear to tackle the parameter instability problem. In this way, tail events are accommodated by the residuals, and this diminishes the impact on the parameters. We also showed that it is the joint multivariate and fat-tailed error modelling that manages to robustify models against extreme observations. In a single-equation setup such as a Phillips curve, the fat-tailed error solution is less effective.

Within a standard Gaussian Bayesian VAR, the choice of prior matters, with more informative specifications or a higher degree of prior shrinkage (more suitable for larger BVARs) acting in favour of stabilising the results after the COVID-19 shock. Both a fat-tailed BVAR and a large Gaussian BVAR with more prior shrinkage ensure stable unconditional forecasts. Nevertheless, as past data do not record an event similar to COVID-19, adding relevant off-model information to the purely model-based forecast is crucial for producing more trustworthy forecasts.

In a real-time forecast exercise, we found that the fat-tailed BVAR is notably superior to its Gaussian counterpart. Also, including off-model information, such as that in the ECB's Survey of Professional Forecasters, brings sizeable gains to forecasting inflation during the pandemic. Both the considered hard and soft conditional inflation forecasts would have been superior to unconditional ones in a standard Gaussian BVAR. Compared to these, the BVAR with fat-tailed errors loses part, but not all of its comparative advantage.

Notwithstanding our documented forecast gains over a standard Gaussian BVAR, also in our best models, in the first phase of the pandemic, inflation tended to surprise on the downside, and subsequently it turned out to be notably higher than foreseen. The puzzling behaviour of inflation during the pandemic is also reflected in the marked revisions to the short-term inflation expectations of professional forecasters, who were surprised by inflation, first on the downside (in the first year of the pandemic) and then on the upside (in the second pandemic year).

As a lesson for practitioners, we suggest that before deriving conclusions based on existing models, one has to check and be aware of how the COVID-19 observations affect the parameter estimates and the implied forecasts. Also, adding meaningful off-model information to the model, such as that in the Survey of Professional Forecasters, brings forecast gains when forecasting inflation in turbulent times such as the pandemic. As new observations pile up, the distortion of traditionally estimated parameters diminishes. Yet, a simple model with Gaussian and homoscedastic errors cannot account for the extreme observations in COVID-19 times.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. The proposed econometric solution for estimating BVARs after March 2020. Errors with a t-distribution and dummy observation prior

Bayesian VAR with multivariate t-distributed errors

Let y_t be an $n \times 1$ vector of variables that is observed over the periods $t = 1, \ldots, T$. Consider the following generic VAR(p) model with independent multivariate t-errors:

$$y_t = a_0 + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \epsilon_t, \quad \epsilon_t \sim t(0, \Sigma, \nu) \quad (1)$$

where a_0 is an $n \times 1$ vector of intercepts and $A_1, \ldots, A_p$ are $n \times n$ coefficient matrices. Let $x'_t = (1, y'_{t-1}, \ldots, y'_{t-p})$ be a $1 \times k$ vector of an intercept and lags with $k = 1 + np$, and let $A = (a_0, A_1, \ldots, A_p)'$ be a $k \times n$ matrix. The error term ϵ_t follows an independent multivariate t-distribution with

covariance matrix Σ of dimension $n \times n$ and degrees of freedom ν.

The multivariate t-distribution of the error term can be represented by a scale mixture of normal distributions; see Geweke (1993) and Ni and Sun (2005):

$$y_t' = x_t'A + \epsilon_t, \quad \epsilon_t|\lambda_t \sim N(0, \lambda_t \Sigma) \tag{2}$$

where λ_t is a latent state variable that is inverse-gamma distributed, i.e. $\lambda_t \sim IG(\nu/2, \nu/2)$.

Next, we stack the observation over $t = 1, \ldots, T$, which yields

$$Y = XA + E, \quad E \sim MN(0, \Sigma \otimes \Lambda) \tag{3}$$

where Y, X, and E are respectively of dimensions $T \times n$, $T \times k$, and $T \times n$. MN denotes the matric-variate normal distribution, $\otimes$ is the Kronecker product, and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_T)$ is of dimension $T \times T$.

We assume that the priors for $p(A, \Sigma)$, $p(\Lambda)$, and $p(\nu)$ are independent. For (A, Σ) we adopt a standard normal-inverse-Wishart prior:

$$\Sigma \sim IW(S_0, \nu_0), \quad A|\Sigma \sim MN(A_0, \Sigma \otimes V_A) \tag{4}$$

For Λ, we assume that λ_t is independently inverse-gamma distributed:

$$\lambda_t \sim IG(\nu/2, \nu/2) \tag{5}$$

and for ν we follow (Chan & Hsiao, 2014) and assume a uniform prior:

$$\nu \sim U(2, \bar{\nu}) \tag{6}$$

where $\bar{\nu} = 50$ is sufficiently large to approximate normal errors.

Because of the prior structure, the posterior draws can be obtained by sequentially sampling from: (1) $p(A, \Sigma|Y, \Lambda)$; (2) $p(\Lambda|Y, A, \Sigma)$ and (3) $p(\nu|Y, A, \Sigma, \Lambda)$. As detailed in Chan (2020), natural conjugacy of the prior $p(A, \Sigma)$ survives even with a general covariance structure in Λ. Therefore, one can derive conditional posterior quantities for $p(A, \Sigma|Y, \Lambda)$ that are still normal-inverse-Wishart distributed. The conditional posterior is given by

$$(A|Y, \Sigma, \Lambda) \sim MN(\hat{A}, \Sigma \otimes K_A^{-1}) \tag{7}$$

$$(\Sigma|Y, \Lambda) \sim IW(\hat{S}, \nu_0 + T) \tag{8}$$

where

$$K_A = V_A^{-1} + X'\Lambda^{-1}X$$
$$\hat{A} = K_A^{-1}(V_A^{-1}A_0 + X'\Lambda^{-1}Y)$$
$$\hat{S} = S_0 + A_0'V_A^{-1}A_0 + Y'\Lambda^{-1}Y - \hat{A}'K_A^{-1}\hat{A}.$$

The posterior distribution for $p(\Lambda|Y, A, \Sigma)$ can be easily derived, as the $\lambda_1, \ldots, \lambda_T$ are conditionally independent given the data and (A, Σ). Therefore, one can show that the posterior of λ_t is given by

$$(\lambda_t|Y, A, \Sigma) \sim IG\left(\frac{\nu+n}{2}, \frac{1}{2}\left(\nu + (y_t' - x_t'A)\Sigma^{-1}(y_t' - x_t'A)'\right)\right) \tag{9}$$

Next, the posterior distribution for $p(\nu|Y, \Lambda, A, \Sigma)$ conditional on Λ is independent from the data and (A, Σ).

As shown by Chan and Hsiao (2014), the conditional posterior for ν is given by

$$(\nu|\Lambda) \propto p(\Lambda|\nu)p(\nu)$$
$$\propto \frac{(\nu/2)^{T\nu/2}}{\Gamma(\nu/2)^T}\left(\Pi_{t=1}^T \lambda_t\right)^{-\frac{\nu}{2}+1} \exp\left(-\frac{\nu}{2}\sum_{t=1}^T \lambda_t^{-1}\right) \tag{10}$$

for $2 < \nu < \bar{\nu}$, and 0 otherwise. This density is non-standard and we use the independence-chain Metropolis–Hastings algorithm of Chan and Hsiao (2014) to sample ν.

To summarise, we can sequentially sample the parameters $(A^{(s)}, \Sigma^{(s)}, \Lambda^{(s)}, \nu^{(s)})$ using a Gibbs sampler with a Metropolis–Hasting step:

1. Draw $\Sigma^{(s)}$ from an $IW(\hat{S}, T + \nu_0)$ distribution
2. Draw $A^{(s)}$ from an $MN(\hat{A}, \Sigma \otimes K_A^{-1})$
3. Draw $\lambda_t^{(s)}$ from an $IG\left(\frac{\nu+n}{2}, \frac{1}{2}\left(\nu + (y_t' - x_t'A)\Sigma^{-1}(y_t' - x_t'A)'\right)\right)$ for $t = 1, \ldots, T$
4. Draw $\nu^{(s)}$ from a proposal distribution and accept/reject using a Metropolis–Hasting step

Inference in the BVAR model with t-errors. To conduct inference in the BVAR model with t-errors, it is convenient to work with the normal mixture representation of (1):

$$y_t = a_0 + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \lambda_t^{\frac{1}{2}} Cu_t, \quad u_t \sim N(0, I),$$
$$\lambda_t \sim IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \tag{11}$$

where $CC' = \Sigma$. Thus, conditional on λ_t, the BVAR with t-errors can be represented as a standard Gaussian VAR with independent but not identically distributed errors. Specifically, the error term exhibits idiosyncratic heteroscedasticity, which is controlled by λ_t.

To compute impulse response functions and unconditional and conditional forecasts from (11), we follow the exposition of Jarociński (2010) and derive impulse response coefficients as well as the projection matrices H for the variables and R for the residuals of the BVAR model with t-errors. We start by simplifying the derivation and define the VAR in terms $z_t = y_t - \mu$—with deviations of y_t from the central level μ defined as:

$$\mu = (I - (A_1 + \cdots + A_p))^{-1} a_0$$

The VAR in the companion form is

$$\tilde{z}_t = F\tilde{z}_{t-1} + G\tilde{u}_t \tag{12}$$

where

$$F = \begin{pmatrix} A_1 & \cdots & A_p \\ I & & 0 \end{pmatrix}, G = \begin{pmatrix} \lambda_t^{\frac{1}{2}}C & 0 \\ 0 & 0 \end{pmatrix}, \tilde{z}_t = \begin{pmatrix} z_t \\ \vdots \\ z_{t-p+1} \end{pmatrix},$$

$$\tilde{u}_t = \begin{pmatrix} u_t \\ \vdots \\ 0 \end{pmatrix}.$$

Take $\tilde{z}_{S+h}$ and, using (12), recursively substitute $\tilde{z}_{S+h-1}$, $\ldots, \tilde{z}_{S+1}$. In the first N rows we obtain z_{S+h} expressed in

terms of data up to S and subsequent errors

$$z_{S+h} = F_{(1..N,.)}^h \tilde{z}_S + \lambda_{S+h}^{\frac{1}{2}} C u_{S+h} + \Phi_1 \lambda_{S+h-1}^{\frac{1}{2}} C u_{S+h-1}$$
$$+ \cdots + \Phi_{h-1} \lambda_{S+1}^{\frac{1}{2}} C u_{S+1} \tag{13}$$

where Φ_j is the upper-left $n \times n$ block of F^j (F to the power j). $\Phi_j C \lambda_t^{\frac{1}{2}}$ is the matrix of orthogonalised impulse response after j periods. $F_{(1..N,.)}^h$ is the matrix composed of the first N rows of F^h. The stacked vector of $z_{S+1}...z_T$ can be written as

$$\begin{pmatrix} z_{S+1} \\ \vdots \\ z_T \end{pmatrix} = \begin{pmatrix} F_{(1..N,.)} \\ \vdots \\ F_{(1..N,.)}^{T-S} \end{pmatrix} \tilde{z}_S$$
$$+ \begin{pmatrix} \lambda_{S+1}^{\frac{1}{2}} C & 0 & 0 \\ \Phi_1 \lambda_{S+1}^{\frac{1}{2}} C & \lambda_{S+2}^{\frac{1}{2}} C & 0 \\ \Phi_{T-S+1} \lambda_{T-S+1}^{\frac{1}{2}} C & ...\Phi_1 \lambda_{T-1}^{\frac{1}{2}} C & \lambda_T^{\frac{1}{2}} C \end{pmatrix}$$
$$\times \begin{pmatrix} u_{S+1} \\ \vdots \\ u_T \end{pmatrix} \tag{14}$$

or shortened as

$$z = H \tilde{z}_S + Ru \tag{15}$$

The above derivations are very similar to those of standard stationary VARs. The only difference is that the forecast errors are now conditionally heteroscedastic on the sequence of $\lambda_{S+1}, \ldots, \lambda_{T-S+1}$, which follows directly from the normal mixture representation of the multivariate t-distribution.

Forecast evaluation metrics

Root mean squared forecast error (RMSFE): Lower values indicate that the forecast is close to the actual value.

$$\text{RMSFE}(y_{i,t+h}^o) = \sqrt{\frac{\sum_{t=t_0}^{T-h} \left(y_{i,t+h}^o - E(y_{i,t+h}|Y_{1:t})\right)^2}{T - h - t_0 + 1}}$$

where $E(y_{i,t+h}|Y_{1:t})$ is the predictive mean forecast.

Average log predictive likelihood (ALPL): Higher values indicate that actual observations are more likely under the predictive density.

$$\text{ALPL}(y_{i,t+h}^o) = \frac{1}{T - h - t_0 + 1} \sum_{t=t_0}^{T-h} \log p(y_{i,t+h}$$
$$= y_{i,t+h}^o | Y_{1,t})$$

where $p(y_{i,t+h}|Y_{1:t})$ is the predictive density.

Average continuous ranked probability score (ACRPS):. Lower values indicate that actual observations are more likely under the predictive distribution.

$$\text{ACRPS}(y_{i,t+h}^o) = \frac{1}{T - h - t_0 + 1} \sum_{t=t_0}^{T-h} \text{CRPS}_{i,t}(y_{i,t+h}^o)$$

where $\text{CRPS}_{i,t}(y_{t+h}^o) = \int_{-\infty}^{\infty} (F(z) - 1\{y_{i,t+h}^o \leq z\})^2 dz = E_f |Y_{i,t+h} - y_{i,t+h}^o| - 0.5 E_f |\tilde{Y}_{i,t+h} - Y_{i,t+h}'|$, and f denotes the cumulative distribution function associated with the predictive density f, $1\{y_{t+h}^o \leq z\}$ denotes an indicator function taking the value 1 if $y_{t+h}^o \leq z$ and 0 otherwise, and Y_{t+h} and Y_{t+h}' are independent random draws from the posterior predictive density.

Heavy tails in single- and multiple-equation models

To understand the effect of univariate and multivariate t-distributed errors on the coefficients of a (multivariate) linear regression model, consider the VAR in (3) and rewrite it in seemingly unrelated regression form using the scale-mixture Gaussian representation of the t-distribution:

$$y = (I_n \otimes X)a + e, \quad e \sim N(0, \Sigma \otimes \Lambda) \tag{16}$$

where $y = \text{vec}(Y)$, $a = \text{vec}(A)$, $e = \text{vec}(E)$, I_n is an identity matrix of dimension n, and Λ is a diagonal covariance matrix. For simplicity it is instructive to focus on the conditional posterior mean of the coefficients under a diffuse (non-informative) prior.[32]

The conditional posterior mean of a is given by

$$\hat{a}_{GLS} = \left((I_n \otimes X)'(\Sigma^{-1} \otimes \Lambda^{-1})(I_n \otimes X)\right)^{-1}$$
$$\times (I_n \otimes X)'(\Sigma^{-1} \otimes \Lambda^{-1})y$$
$$= \left((\Sigma^{-1} \otimes X' \Lambda^{-1} X)\right)^{-1} (\Sigma^{-1} \otimes X' \Lambda^{-1})y$$
$$= \left(I_n \otimes (X' \Lambda^{-1} X)^{-1} X' \Lambda^{-1}\right) y$$
$$= \begin{pmatrix} (X' \Lambda^{-1} X)^{-1} X' \Lambda^{-1} y_1 \\ (X' \Lambda^{-1} X)^{-1} X' \Lambda^{-1} y_2 \\ ... \\ (X' \Lambda^{-1} X)^{-1} X' \Lambda^{-1} y_n \end{pmatrix}$$
$$\underbrace{=}_{\Lambda = I_T} \begin{pmatrix} (X'X)^{-1} X' y_1 \\ (X'X)^{-1} X' y_2 \\ ... \\ (X'X)^{-1} X' y_n \end{pmatrix} = \hat{\alpha}_{OLS}$$

where GLS and OLS denote the generalised and ordinary least squares estimates, respectively.

These formulas reveal two interesting properties. First, given Λ, the system estimation (System) is equivalent to single-equation estimation (SgEq). Second, given $\Lambda = I_T$, then $\hat{a}_{GLS} = \hat{a}_{OLS}$ (the error term is homoscedastic and independently Gaussian distributed). Note that under t-distributed errors $\Lambda \neq I_T$, and hence, $\hat{a}_{GLS} \neq \hat{a}_{OLS}$.

Because of the first property, if the system estimation does not yield the same estimate as an equation-by-equation estimation, then it must be the case that the implied volatility is different in the univariate setting as compared to the multivariate setting, i.e. $\Lambda^{SgEq} \neq \Lambda^{System}$.

To see this, suppose a and Σ are fixed. Using (16), the error is $\epsilon_t = y_t - (I_n \otimes x_t')a$. The conditional posterior of λ_t from a regression with univariate t-distributed errors

---

[32] Note that this posterior mean is equivalent to the generalised least squares estimate.

for variable $y_{i,t}$ is given by

$$\lambda_t^{SgEq} \sim IG\left(\frac{\nu+1}{2}, \frac{1}{2}\left(\nu + \frac{1}{\sigma_{ii}^2}\epsilon_{i,t}^2\right)\right)$$

whereas the conditional posterior of λ_t with multivariate t-distributed errors for y_t is given by

$$\lambda_t^{System} \sim IG\left(\frac{\nu+n}{2}, \frac{1}{2}\left(\nu + \epsilon_t'\Sigma^{-1}\epsilon_t\right)\right)$$

Thus, λ_t depends on different information sets (abstracting from distributional differences). Particularly, λ_t^{SgEq} depends only on the forecast error $\epsilon_{i,t}$ of the ith variable ($s_t^2 = \frac{1}{\sigma_{ii}^2}\epsilon_{i,t}^2$) in the univariate settings, while under the multivariate distribution, λ_t^{System} depends on the vector of forecast errors ϵ_t ($s_t^2 = \epsilon_t'\Sigma^{-1}\epsilon_t$, scalar).

For instance, suppose there is a large shock in the jth variable but not in the ith variable. Heuristically, the posterior of λ_t^{System} will be larger than λ_t^{SgEq} because there was no unusually large forecast error in the ith variable. As a consequence, when the parameters are updated in the next iteration of the Gibbs sampler, the data will be downweighed more strongly in the multivariate setup as opposed to the univariate setup. The estimated coefficients a are then different because $\lambda_t^{System} \neq \lambda_t^{SgEq}$.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2022.01.002.

References

Antolín-Díaz, J., Drechsel, T., & Petrella, I. (2021). Advances in nowcasting economic activity: Secular trends, large shocks and new data. Mimeo.

Ascari, G., Fagiolo, G., & Roventini, A. (2015). Fat-tail distributions and business-cycle models. *Macroeconomic Dynamics*, *19*, 465–476.

Baker, S., Bloom, N., & Davis, S. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, *131*, 1593–1636.

Bańbura, M., & Bobeica, E. (2020). Does the Phillips curve help to forecast euro area inflation?. *International Journal of Forecasting*, forthcoming.

Bańbura, M., Brenna, J., Paredes, F., & Ravazzolo, F. (2021). Combining bayesian vars with survey density forecasts: does it pay off?. *Working Paper 2543*, European Central Bank.

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, *25*, 71–92.

Baumann, U., Darracq Pariès, M., Westermann, T., Riggi, M., Bobeica, E., Meyler, A., Böninghausen, B., Fritzer, F., Trezzi, R., & Jonckheere, J. (2021). Inflation expectations and their role in eurosystem forecasting. *Occasional paper 264*, European Central Bank.

Bjerregård, M., Mller, J., & Madsen, H. (2021). An introduction to multivariate probabilistic forecast evaluation. *Energy and AI*, *4*, Article 100058.

Bobeica, E., & Jarociński, M. (2019). Missing disinflation and missing inflation: A VAR perspective. *International Journal of Central Banking*, *15*, 199–232.

Carriero, A., Clark, T., & Mertens, E. (2021). Addressing COVID-19 outliers in BVARs with stochastic volatility. *Working paper 21-02*, Federal Reserve Bank of Cleveland.

Chan, J. (2020). Large Bayesian VARs: A flexible Kronecker error covariance structure. *Journal of Business & Economic Statistics*, *38*, 68–79.

Chan, J., & Hsiao, C. (2014). Estimation of stochastic volatility models with heavy tails and serial dependence. In *Bayesian inference in the social sciences* (pp. 155–176). John Wiley & Sons Ltd, chap 6.

Chib, S., & Ramamurthy, S. (2014). DSGE models with Student-t errors. *Econometric Reviews*, *33*, 152–171.

Chiu, C.-W., Mumtaz, H., & Pinter, G. (2017). Forecasting with VAR models: Fat tails and stochastic volatility. *International Journal of Forecasting*, *33*, 1124–1143.

Christiano, L. (2007). Comment on 'on the fit of New Keynesian models'. *Journal of Business & Economic Statistics*, *25*, 143–151.

Clark, T., & Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, *30*, 551–575.

Cogley, T., & Sargent, T. (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, *8*, 262–302.

Cúrdia, V., Del Negro, M., & Greenwald, D. (2014). Rare shocks, great recessions. *Journal of Applied Econometrics*, *29*, 1031–1052.

de Vincent-Humphreys, R., Dimitrova, I., Falck, E., Henkel, L., & Meyler, A. (2019). Twenty years of the ECB Survey of Professional Forecasters. In *Economic Bulletin 1, Article 1*. European Central Bank.

Dieppe, A., van Roye, B., & Legrand, R. (2016). The BEAR toolbox. *Working paper 1934*, European Central Bank.

Doan, T., Litterman, R., & Sims, C. (1983). Forecasting and conditional projection using realistic prior distributions. *Working paper 1202*, National Bureau of Economic Research.

ECB (2019). Results of the third special questionnaire for participants in the ECB Survey of Professional Forecasters. *Report*, European Central Bank.

ECB (2020). The ECB Survey of Professional Forecasters- Second quarter of 2020. *Report*, European Central Bank.

Eser, F., Karadi, P., Lane, P., Moretti, L., & Osbat, C. (2020). The Phillips curve at the ECB. *Working paper 2400*, European Central Bank.

Fagiolo, G., Napoletano, M., & Roventini, A. (2008). Are output growth-rate distributions fat-tailed? Some evidence from OECD countries. *Journal of Applied Econometrics*, *23*, 639–669.

Ganics, G., & Odendahl, F. (2021). Bayesian VAR forecasts, survey information and structural change in the euro area. *International Journal of Forecasting*, *37*, 971–999.

Geweke, J. (1993). Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics*, *8*, 19–40.

Geweke, J., & Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, *26*, 216–230, Special Issue: Bayesian Forecasting in Economics.

Geweke, J., & Amisano, G. (2011). Hierarchical Markov normal mixture models with applications to financial asset returns. *Journal of Applied Econometrics*, *26*, 1–29.

Grothe, M., & Meyler, A. (2018). Inflation forecasts: Are market-based and survey-based measures informative?. *International Journal of Financial Research*, *9*, 171–188.

Guerrieri, V., Lorenzoni, G., & Werning, I. (2020). Macroeconomic implications of COVID-19: Can negative supply shocks cause demand shortages?. *Working paper 26918*, National Bureau of Economic Research.

Hartwig, B. (2020). Robust inference in time-varying structural VAR models: The DC-Cholesky multivariate stochastic volatility model. *Discussion paper 34/2020*, Deutsche Bundesbank.

Hartwig, B. (2021). Bayesian VARs and prior calibration in times of COVID-19. Mimeo.

Huber, F., Koop, G., Onorante, L., Pfarrhofer, M., & Schreiner, J. (2020). Nowcasting in a pandemic using non-parametric mixed frequency VARs. *Journal of Econometrics*.

Jacquier, E., Polson, N., & Rossi, P. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, *122*, 185–212.

Jarociński, M. (2010). Conditional forecasts and uncertainty about forecast revisions in vector autoregressions. *Economics Letters*, *108*, 257–259.

Jurado, K., Ludvigson, S., & Ng, S. (2015). Measuring uncertainty. *American Economic Review*, *105*, 1177–1216.

Krüger, F., Lerch, S., & Gneiting, T. (2021). Predictive inference based on Markov chain Monte Carlo output. *International Statistical Review*, *89*, 274–301.

Lenza, M., & Primiceri, G. (2020). How to estimate a VAR after March 2020. *Working paper 27771*, National Bureau of Economic Research.

Meyer, B., Prescott, B., & Sheng, X. (2021). The impact of the COVID-19 pandemic on business expectations. *International Journal of Forecasting*.

Mishkin, F. (2011). Monetary policy strategy: Lessons from the crisis. *Working paper 16755*, National Bureau of Economic Research.

Ng, S. (2021). Modeling macroeconomic variations after COVID-19. *Working paper 29060*, National Bureau of Economic Research.

Ni, S., & Sun, D. (2005). Bayesian estimates for vector autoregressive models. *Journal of Business & Economic Statistics*, *23*, 105–117.

Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, *72*, 821–852.

Primiceri, G., & Tambalotti, A. (2020). Macroeconomic forecasting in the time of COVID-19. In *Manuscript 2020*. Northwestern University.

Robertson, J., Tallman, E., & Whiteman, C. (2005). Forecasting using relative entropy. *Journal of Money, Credit and Banking*, *37*, 383–401.

Rossi, B. (2020). Forecasting in the presence of instabilities: How do we know whether models predict well and how to improve them. *Discussion paper 14472*, Center for Economic Policy Research.

Schorfheide, F., & Song, D. (2020). Real-time forecasting with a (standard) mixed-frequency VAR during a pandemic. *Working paper 20-26*, Federal Reserve Bank of Philadelphia..

Sims, C., & Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, *39*, 949–968.

Tallman, E., & Zaman, S. (2020). Combining survey long-run forecasts and nowcasts with BVAR forecasts using relative entropy. *International Journal of Forecasting*, *36*, 373–398.

Waggoner, D., & Zha, T. (1999). Conditional forecasts in dynamic multivariate models. *The Review of Economics and Statistics*, *81*, 639–651.

Yellen, J. (2015). Inflation dynamics and monetary policy. In *Speech at the Philip Gamble Memorial Lecture, University of Massachusetts, Amherst, Massachusetts, 24 2015*. Federal Reserve System.