How useful are time-varying parameter models for forecasting economic growth in CESEE?

Martin Feldkircher, Nico Hauzenberger[1]

Empirical evidence has shown that a prerequisite for generating reliable macroeconomic forecasts is either the inclusion of a large information set or modeling time variation in the models' parameters and volatilities. In this paper we examine these claims in a comparative manner, forecasting GDP growth for six CESEE economies. We use Bayesian techniques and evaluate the models based on both the accuracy of their point forecasts as well as the degree of uncertainty surrounding these predictions. Our results indicate that forecasts from a fully-fledged time-varying parameter model tend to outperform those from its constant parameter competitors. Adding more information, e.g. from other countries, by contrast, does not improve forecast performance significantly for most of the countries under study. Last, we analyze whether it pays to forecast GDP growth indirectly by summing up forecasts of GDP components. This approach yields competitive forecasts, yet it preserves an economic interpretation of the underlying drivers for the economic growth forecasts, which is of crucial importance from a practitioner's view.

JEL classification: C11, C32, C53, E17
Keywords: forecasting, CESEE, time-varying parameter, aggregate GDP forecast

> *"Those who have knowledge, don't predict. Those who predict, don't have knowledge."*
> *Lao Tzu*

Forecasting economic growth for Central, Eastern and Southeastern European (CESEE) countries is of key interest to individuals, firms and banks that have a stake in these economies. Also, due to the forward-looking element of monetary policy, macroeconomic forecasting has always been a core research field for central bankers. Today, a great number of forecasting models are applied at central banks on a regular basis. They range from large-scale models (e.g. the models used in the Banca d'Italia) and dynamic stochastic general equilibrium models (DSGE, e.g. used in the Bank of England) to structural or semi-structural time series models, such as the OeNB's FORCEE model to forecast economic growth in CESEE economies, with the latter yielding reliable forecasts as has been demonstrated in Crespo Cuaresma et al. (2009) and Slačík et al. (2014). However, in the aftermath of the global financial crisis, most quantitative models used by central banks came in for heavy criticisms (Hendry and Muellbauer, 2018). Since then, policymakers have been seeking flexible, yet economically consistent, forecasting models. These models should be able to adapt quickly to changes in the economic environment, which sometimes happen more gradually, sometimes abruptly. The challenge for a researcher is that flexibility can be achieved in different ways (Carriero et al., 2016). One way to ensure the model is capable of adapting quickly is to include a rich information set. Given that most CESEE economies use an export-driven

[1] *Oesterreichische Nationalbank, Foreign Research Division, martin.feldkircher@oenb.at, and Vienna University of Economics and Business, niko.hauzenberger@wu.ac.at. Opinions expressed by the authors of studies do not necessarily reflect the official viewpoint of the Oesterreichische Nationalbank (OeNB) or of the Eurosystem. The authors would like to thank Peter Backé, Florian Huber, Julia Wörz, Michael Pfarrhofer and two anonymous referees for helpful comments and valuable suggestions.*

growth model, a more complete modeling of the external sector could prove particularly useful. Another way of introducing flexibility is to use econometrically more sophisticated models that allow parameters to drift over time.

In this paper we examine forecasts derived from a range of Bayesian vector-autoregressive (BVAR) models for six non-euro area EU Member States from the CESEE region: Bulgaria, Croatia, the Czech Republic, Hungary, Poland and Romania. BVARs seem to be particularly suited for forecasting GDP growth for CESEE economies since the time series available for these economies are rather short (Brázdik and Franta, 2017). The models we examine vary in the degree to which they can adapt to changes in economic conditions and in the amount of information on foreign economic conditions they include. Our main research question is whether time-varying parameter models can improve forecast performance over more simple, linear-in-parameters models for the CESEE region. Since these economies underwent boom-bust cycles and structural breaks during the estimation period, time-varying parameter models might prove especially useful for forecasting CESEE growth, which so far has not been investigated systematically for the region. Following Crespo Cuaresma et al. (2009), we model the components of GDP jointly and compute forecasts either directly or by aggregating forecasts of GDP components. For the latter we propose two approaches: first, simply summing up GDP components' forecasts accounting for their relative shares in overall GDP and second, optimizing the shares/weights of the components based on how well the model can predict them. For all models we compute predictive densities to evaluate their forecasting performance. By this we ensure that models that yield both an accurate point forecast and a small degree of uncertainty surrounding the prediction are rewarded.

Our results are as follows: First we find evidence for forecast improvements achieved by the proposed time-varying parameter model over constant parameter models and univariate benchmark models. However, the specification of the time-varying parameter model is such that time variation in the parameters is kept relatively tight. Our results show that setting the respective prior too loose results in overfitting and in turn poor forecast performance. Second, including a large information set – namely variables from all countries in the region – does not improve forecast performance. An exception to this is Hungary, for which this "region-wide" model yields the best forecast at both the one-quarter and four-quarters forecast horizon. Last, weighted forecasts of GDP components are competitive with direct time-series forecasts of GDP growth. This finding is important since it shows that not only does the proposed forecast method yield sound predictions but it can also be used in an institutional forecasting process, e.g. in a central bank, where the focus is not only on the point forecast but also on growth drivers.

The rest of the paper is structured as follows: The next section reviews the literature, and section 3 introduces the data. Section 4 describes the econometric framework, and section 5 discusses different ways of forecast aggregation. In section 6 we discuss the results, and section 7 concludes the paper.

1 Review of the literature

In the aftermath of the global financial crisis, the economic profession started to develop new models that should yield more reliable forecasts. The consensus of this literature is that forecasting with vector-autoregressive (VAR) models can be improved by exploiting large information sets and accounting for changes in the relationships of the macroeconomic variables by modeling time variation in their volatilities (Carriero et al., 2016). The first claim – the more information included, the better the forecast – has been empirically verified by several studies using different econometric techniques and data sets (see, among others, Bańbura, et al., 2010; Carriero et al., 2011; Koop, 2013; Carriero et al., 2015). Also, for CESEE forecasting, it has proven useful to include a large information set. For example, Franta et al. (2016) show that including a rich set of high-frequency information in a mixed-frequency vector autoregression outperforms official CNB inflation forecasts. Other applications that use large information sets cover the area of nowcasting. Kunovac and Špalat (2014) use over 40, and Armeanu et al. (2017) use 80 high-frequency indicators to successfully nowcast Romanian and Croatian GDP, respectively. For an excellent review of this literature, consider Riedl and Wörz (2018).

The second key feature of a useful forecasting model – namely accounting for time variation using more sophisticated models – can be technically implemented in different ways. In its simplest form, time variation can be captured by allowing the volatility of the residual part of the model to vary over time (stochastic volatility). Such a model would yield precise inference during times in which volatility is low (i.e., the part of variation that is left unexplained by the model), while credible intervals are inflated during turbulent times. The bounds surrounding the forecast would thus vary over time, which allows gauging the reliability of a current forecast at hand – a feature that is absent in a purely linear model. In fact, the literature has shown that accounting for time variation in variances significantly improves forecasts (Cogley and Sargent, 2005; Clark and Ravazzolo, 2015; Carriero et al., 2016; Chan and Eisenstat, 2018).

In a fully-fledged time-varying parameter model, not only residual variances but also the parameters that reflect the economic relationships would be allowed to vary over time. Such a model could be particularly useful when dealing with macroeconomic data of economies that have undergone structural changes or pronounced boom-bust phases, e.g., CESEE economies. Here, the claim that more information and accounting for time-variation improves forecasting should be modified. Huber et al. (2018) indicate a trade-off between the size of the information set and the flexibility of the model: time-varying parameter models are particularly useful for small-scale models, where moving coefficients can account for missing information, while in a richer data information environment it suffices to account for stochastic volatility. In general, the numerosity of parameters to estimate in time-varying parameter models is huge, and these models usually suffer from issues related to overfitting. This holds also true for small-scale applications. Hence, it is crucial to put some regularization/shrinkage on the coefficients when estimating time-varying parameter models (Bitto and Frühwirth-Schnatter, 2018; Belmonte et al., 2014; Eisenstat et al., 2016). Can these models then improve forecasting? There is a lot of empirical evidence that demonstrates the usefulness of time-varying parameter models, albeit most studies use U.S. data (Cogley and Sargent, 2005; Primiceri, 2005; D'Agostino et al., 2013; Aastveit et al., 2017). For CESEE economies, only Ravnik (2014) examines the usefulness of time-varying parameter VAR models. He shows that short-term forecasts for Croatian GDP can be significantly improved using a Bayesian time-varying parameter VAR relative to simple benchmark models as well as fixed parameter VARs.

2 Data

In this section we briefly describe the data we use to forecast GDP growth for Bulgaria, Croatia, the Czech Republic, Hungary, Poland and Romania. Following Crespo Cuaresma et al. (2009), we collect quarterly data on real GDP (*gdp*), its components (i.e., gross fixed capital formation (*inv*), private consumption (*c*), public consumption (*g*), imports (*m*) and exports (*x*)), nominal exchange rates vis-à-vis the euro (*e*), consumer prices (π), short-term interest rates (*i*), wages (*wg*) and private credit (*pc*). National data are augmented by euro area data on short-term interest rates (i_{EA}) and GDP (gdp_{EA}). All data except interest rates are in logarithms, seasonally adjusted and transformed to satisfy stationarity by first differencing. Note that by first differencing, long-run relationships are not taken into account, which could lead to more imprecise forecasts over the longer term. The forecasting gains from accounting for cointegration are, however, modest (Carriero et al., 2015), and time-varying parameter models are typically estimated with stationary data. Exchange rates are not included in the models for Croatia and Bulgaria since both countries – to a different extent – pursue a fixed exchange rate regime with the euro as the anchor currency.

Depending on the country, data are either available for the period from Q1 1995 to Q3 2017 (Czech Republic, Hungary, Poland and Romania) or from Q1 2000 to Q3 2017 (Bulgaria and Croatia).

3 Econometric framework

In this section we describe the setting we use to forecast output growth. For a typical country *c*, we estimate variants of the following VAR model:

$$\begin{pmatrix} gdp_{ct} \\ c_{ct} \\ inv_{ct} \\ g_{ct} \\ x_{ct} \\ m_{ct} \\ wg_{ct} \\ \pi_{ct} \\ i_{ct} \\ pc_{ct} \\ e_{ct} \end{pmatrix} = A_{c1,t} \begin{pmatrix} gdp_{c,t-1} \\ c_{c,t-1} \\ inv_{c,t-1} \\ g_{c,t-1} \\ x_{c,t-1} \\ m_{c,t-1} \\ wg_{c,t-1} \\ \pi_{c,t-1} \\ i_{c,t-1} \\ pc_{c,t-1} \\ e_{c,t-1} \end{pmatrix} + \cdots + A_{cp,t} \begin{pmatrix} gdp_{c,t-p} \\ c_{c,t-p} \\ inv_{c,t-p} \\ g_{c,t-p} \\ x_{c,t-p} \\ m_{c,t-p} \\ wg_{c,t-p} \\ \pi_{c,t-p} \\ i_{c,t-p} \\ pc_{c,t-p} \\ e_{c,t-p} \end{pmatrix} + B_{ct} \begin{pmatrix} gdp_{EAt} \\ i_{EAt} \\ \vdots \\ gdp_{EA,t-p+1} \\ i_{EA,t-p+1} \\ 1 \end{pmatrix} + \varepsilon_{ct} \qquad (1)$$

We jointly model GDP growth, its components and additional key macroeconomic variables, such as wage growth, consumer price inflation, short-term interest rates, private credit growth and changes in the exchange rate vis-à-vis the euro. The exogenous euro area variables are internal projections from the ECB and hence do not have to be predicted endogenously within the model. These data are available over the forecast horizon, assuming exogenous variables are given a priori.[2]

[2] *More precisely, we use confidential quarterly forecasts of the ECB's Broad Macroeconomic Projection Exercise (BMPE) conducted by Eurosystem staff. The forecasts are available twice a year, in March and September, which coincides with the timing of the OeNB' forecast exercise for the CESEE economies. For this study, rather than using forecast vintages, we have used forecasts from September 2018 for the whole estimation and forecast evaluation period. This is consistent with the macro data, which also stem from the last available vintage.*

Jointly modeling GDP components and overall GDP can aggravate multicollinearity issues that typically plague VAR models. Since we do not cover stock changes and statistical discrepancy, our model is not perfectly collinear, though. Remaining collinearity will be treated by using shrinkage priors and focusing on density forecasts that punish models that suffer from forecast uncertainty caused by overfitting.

More compactly, the model for $t=1,\cdots,T$ can be written in matrix form as follows:

$$y_{ct} = A_{c1,t}y_{c,t-1} + \cdots + A_{cp,t}y_{c,t-p} + B_{ct}X_{ct} + \varepsilon_{ct}, \tag{2}$$

with y_{ct} denoting the M-dimensional vector of endogenous variables, X_{ct} the N-dimensional vector of exogenous regressors, $A_{cj,t}(j=1,\cdots,p)$ denote M x M potentially time-varying coefficient matrices, B_{ct} a M x N potentially time-varying coefficient matrix corresponding to exogenous variables, including an intercept term as well. The constant parameter VAR model arises as a special case of equation (2) with $A_{cj,t} = A_{cj}(j=1,\cdots,p)$ and $B_{ct} = B_c$ for all t.

For both variants, constant and time-varying parameter models, we assume that the errors $\mathcal{E}_{ct}$ are multivariate Gaussian with zero mean and a variance-covariance matrix Σ_{ct} that can be factorized as

$$\Sigma_{ct} = L_c H_{ct} L_c{}'. \tag{3}$$

Here L_c is an M x M lower triangular matrix with ones on the diagonal and H_{ct} denotes an M-dimensional diagonal matrix with time-varying elements $e^{h_{ic,t}}$, for $i=1,\cdots,M$ (Cogley and Sargent, 2005; Huber and Feldkircher, 2017). As emphasized in the introduction, stochastic volatility is an important feature of a successful forecasting model. The time-varying (logarithm of) volatilities, $h_{ic,t}$, are assumed to follow an AR-(1) process (Jacquier et al., 1994; Kim et al., 1998; Kastner and Frühwirth-Schnatter, 2014). Specifically,

$$h_{ic,t} = \mu_{ic} + \varphi_{ic}\left(h_{ic,t-1} - \mu_{ic}\right) + \epsilon_{ic,t}, \tag{4}$$

with μ_{ic} denoting the unconditional mean of the log volatility, φ_{ic} the persistence parameter with $|\varphi_{ic}|<1$ and $\epsilon_{ic,t}$ the error term, which is Gaussian with mean zero and variance ω_{ic}^2.

3.1 Threshold time-varying parameter BVAR with stochastic volatility (TTVP-SV)

Using the set-up described above, we examine the predictive performance of three multivariate model classes and two univariate benchmark models. To begin with, we introduce the most flexible specification, which is the threshold time-varying parameter model with stochastic volatility. For that purpose, it proves to be convenient to collect all coefficient matrices $A_{cj,t}$, $(j=1,\ldots,p)$ and B_{ct} in a matrix C_{ct} and in addition define $c_{ct} = \text{vec}(C_{ct})$. In the following, the i^{th} element of the full coefficient vector c_{ct} evolves according to a random walk,

$$c_{ic,t} = c_{ic,t-1} + \eta_{ic,t}. \tag{5}$$

The way the model handles time variation in the coefficients deserves some explanation. Huber et al. (2018) propose letting parameters drift depending on the size of previous coefficient movements. More precisely, for each coefficient of the model, a threshold γ_{ic} is estimated. In case an estimated coefficient movement at time t, gauged by the absolute change between period t and t–1, is sufficiently large (i.e. surpasses the estimated threshold), the coefficient is deemed moving. In case the threshold is not surpassed, the coefficient is pushed toward the value for period t–1. Formally, this is achieved by specifying the shocks to coefficients as a mixture of two Gaussians:

$$\eta_{ic,t} \sim \delta_{ic,t} N\left(0, \sigma_{ic,1}^2\right) + \left(1 - \delta_{ic,t}\right) N\left(0, \sigma_{ic,0}^2\right) \quad (6)$$

with $\sigma^2_{ic,1} >> \sigma^2_{ic,0}$. δ_i denotes a binary indicator being one if the absolute change of the coefficient is larger than the estimated threshold value γ_{ic}, and zero otherwise (Huber et al., 2018). The high variance state ($\sigma^2_{ic,1}$) translates into time-variation of coefficients without an additional constraint, whereas the low variance state ($\sigma^2_{ic,0}$) implies that the coefficient in period t is tightly centered on the coefficient of the previous period t–1 and thus approximately held constant over time. Therefore, a crucial hyperparameter specified by the researcher a priori is $\sigma^2_{ic,0} = \xi$, with ξ being a scaling factor that governs the minimum level of time variation on coefficient movements.[3] We examine five variations of the TTVP-SV model, ranging from a very loose prior (TTVP-SV ξ=1e-04) to a very tight prior (TTVP-SV ξ=1e-08).

3.2 Constant parameter BVAR

Next, we consider constant parameter Bayesian VAR models with stochastic volatility that allow handling large information sets (see, for example, Bańbura et al., 2010; Carriero et al., 2011; Koop, 2013; Carriero et al., 2015). The specifications we examine cover the well-known Minnesota prior put forth by Doan et al. (1984) and Litterman (1986). We include two versions, one with stochastic volatility (Minnesota-SV) and one assuming homoscedastic variances (Minnesota). As a workhorse of central banks' forecasters, the Minnesota prior assumes a random walk a priori for log-transformed time series and a white noise process for log-differenced endogenous variables. In a classic deterministic fashion, shrinkage is introduced by downweighting more distant lags and lags of other endogenous variables more heavily, compared to own lags. In particular, the first own lags are expected to be essential drivers of a persistent economic time series. We also use a prior that is particularly useful for handling large data sets, namely the normal gamma (NG-SV) generalized to the VAR case by Huber and Feldkircher (2017).[4] This prior belongs to the family of global-local shrinkage priors and proves particularly useful when pushing coefficients strongly toward zero, which is necessary to handle large-scale models. The advantage of the normal-gamma prior arises since the prior distribution is characterized by heavier tails, which ensures that coefficients are allowed to be non-zero when supported by the data, although the overall degree of shrinkage is high (Griffin and Brown, 2010).

[3] *See Huber et al. (2018) for more details.*

[4] *For the TTVP-SV, model variable selection is addressed by using a normal-gamma prior on the initial state of coefficients at $t = 0$.*

3.2.1 Multi-country BVAR

Last, we modify equation (1) by stacking all country-specific VARs to yield a constant parameter multi-country model with stochastic volatility. This "regional" set-up constitutes a (very) large-scale VAR model and is estimated in order to investigate whether modeling cross-country spillovers pays off. Here we opt for estimating all countries jointly, which is in contrast to other multi-country models, such as global VARs (see, for example, Crespo Cuaresma et al., 2016). These estimate separate country models, which are linked in a second step by a measure of economic connectivity, such as the extent of bilateral trade. Estimating all countries jointly assesses cross-country dependencies empirically without the help of further assumptions/ measures of connectivity. Since the model constitutes a large VAR, we opt for the normal-gamma prior with a specification that implies a high degree of shrinkage (Multi-NG-SV).

3.3 Univariate competitors

The set of competing models is completed by two univariate models: an autoregressive model of order 1 (AR1-SV) and a random walk (RW-SV). Moreover, the AR1-SV model is linear in parameters. In order to obtain legitimate benchmark models, we also allow for stochastic volatility, since this feature commonly yields large gains for density forecasts. The prior distribution for the autoregressive coefficient is weakly informative. For both univariate specifications we also impose time-varying variances. That is, the logarithm of volatilities is defined as AR-(1) process as in equation (4).[5]

For all models we use Bayesian estimation methods. We employ a Markov chain Monte Carlo (MCMC) algorithm for all proposed models enabling inference of the joint posterior distribution. We use 5,000 draws for obtaining the predictive densities after a burn-in phase of 3,000 draws. For the estimation of time-varying volatilities, we exploit the R package stochvol (Kastner, 2016).

4 Forecast aggregation

Once we have found a promising forecasting model, the question arises how to conduct the forecast. In theory, given the forecasting model fits the data well, aggregating forecasts from sub-components should boost forecast performance. In a recent contribution and in the context of inflation forecasts, Bermingham and D'Agostino (2014) indeed find that aggregating forecasts from CPI subcomponents can improve forecast performance. In practice, there is a range of pitfalls for forecast aggregation of output or inflation, though, since the predictive accuracy depends on two (potentially countervailing) effects, namely the predictive accuracy of all components and the cancel-out effects of components' forecast errors. Moreover, Lütkepohl (2011a) and Lütkepohl (2011b) highlight potential problems when aggregating time series with time-varying weights. It is therefore not surprising that some studies such as Hubrich (2005), Hendry and Hubrich (2006) and Hendry and Hubrich (2011) point to mixed evidence regarding the superiority of forecast aggregation over using direct forecasts.

In the following, we use two approaches to yield GDP growth forecasts from subcomponents. The first one is a simple weighted aggregation of GDP components'

[5] *For further details on prior specifications, see the appendix.*

forecasts, where the weights correspond to realized components' shares in overall GDP. For the second approach, weights are optimized based on their historical forecast performance (Geweke and Amisano, 2011).

We first focus on simple aggregation based on realized GDP shares. Here the h-step ahead forecast conditional on information in period t can be decomposed as follows:

$$gdp_{t+h|t} = w_{t+h} Z_{t+h|t} + \vartheta_{t+h}, \qquad (7)$$

with $w_{t+h} = (w_{C,t+h}, w_{I,t+h}, w_{G,t+h}, w_{X,t+h}, -w_{M,t+h})$ being the vector of weights assigned to the components vector $Z_{t+h|t} = (c_{t+h|t}, i_{t+h|t}, g_{t+h|t}, x_{t+h|t}, m_{t+h|t})'$. ϑ_{t+h} accounts for inventory investments and statistical discrepancies (see, for example, Marcellino et al., 2003; Ravazzolo and Vahey, 2014). We treat ϑ_{t+h} as an unforecastable white noise process, centered on zero.

The simple "bottom-up" approach boils down to weighting each component's forecast by its share in overall output in the current period t. That is,

$$w_{z,t+h} = \frac{z_t}{GDP_t} \qquad for\ z_t = \{C_t, I_t, G_t, X_t, M_t\} \qquad (8)$$

where upper-case letters denote the corresponding levels of the variables and $GDP_t = C_t + I_t + G_t + X_t - M_t$. In this case, we keep the corresponding weights fixed over the h-step ahead periods to the value of period t, which is assumed to be known. Note that this approach yields an economically consistent forecast, in the sense that GDP components' realized contributions sum up to overall growth. However, as noted by Lütkepohl (2011a) and Brüggemann and Lütkepohl (2013), actual figures of output and components may not be available contemporaneously and are, more generally, subject to revisions.

As a second approach to forecast aggregation we propose considering components' forecasts as a portfolio of predictions, which must be optimally weighted with respect to a loss function (Timmermann, 2006; Geweke and Amisano, 2011; Ravazzolo and Vahey, 2014). Geweke and Amisano (2011) provide a framework that maximizes the historical forecast performance to yield optimized weights. These weights are then used to sum up the predictive densities of the GDP components' forecasts. In other words, this procedure ensures that inaccurate forecasts of components are down-weighted and those that can be predicted more successfully are up-weighted. Berg and Henzel (2015) successfully apply these methods for a set with different models when forecasting euro area output and inflation. Ravazzolo and Vahey (2014) combine forecasts of disaggregate time series to forecast U.S. personal consumption expenditures.

Here, we follow Geweke and Amisano (2011) and evaluate the historical log predictive score of aggregate output growth obtained via combining expenditure-side forecasts. That is, we maximize forecast weights for the components based on their respective historical performance, which is evaluated for the combined GDP growth forecast. This is in contrast to Geweke and Amisano (2011), who choose weights maximizing historical performance for each component.[6]

[6] *The difference to the approach of Geweke and Amisano (2011) is that we do not combine forecasts of different models for a single quantity of interest but combine forecasts of components for an aggregate (see, for example, Timmermann, 2006; Ravazzolo and Vahey, 2014).*

Therefore, the optimal weights vector is chosen according to

$$w^*_{t+h} = \max_{w_{t+h}} \frac{1}{t-(T_0+1)} \sum_{\tau=T_0+1}^{t} log\, p(gdp_\tau^{e.p.}|I_{\tau-1}, w_{t+h}) \quad (9)$$

with $I_{\tau-1}$ denoting the historical information set containing all parameters and latent quantities estimated for this period. The superscript *e.p.* denotes the ex post (realized) value of output growth and T_0 indicates the start of the hold-out sample. Optimization is carried out over a grid of possible weights, where we define the bounds of the grid based on the ex ante (at period *t*) realized value of a components' GDP share. That is, we restrict the possible optimized weights to a neighborhood of the historically realized weights. Note that while simple aggregation ensures that the overall GDP growth forecast is consistent in an economic sense, by optimizing weights we lose this property but probably yield more accurate forecasts overall.

5 Results

The merits of the proposed models are evaluated with a pseudo out-of-sample forecasting exercise by comparing log predictive likelihood scores (Geweke and Amisano, 2010). We also provide a detailed analysis of the components' point forecast to identify the main sources of forecast errors and potential canceling-out effects when combining forecasts.

For the evaluation of one-step and four-step ahead predictions we keep a hold-out sample of size *H* from Q1 2010 to Q4 2017 for all countries, except Hungary. For Hungary, we start from Q1 2011 since for the early part of the hold-out sample, forecasts of most models showed an explosive behavior.[7]

Moreover, we use the first out-of-sample period Q1 2010 (or Q1 2011) for the initial optimization of weights. For all models with time-varying parameters and/or stochastic volatility, coefficient estimates are kept constant at the value corresponding to the last observation in the estimation sample when constructing the forecast.

In the following, we plug in the realized values of the hold-out sample in the predictive density for calculating the log predictive likelihood. Hence larger values indicate a better forecasting performance. Note that LPS scores have to be interpreted relative to a benchmark model, which we choose as a simple random walk model with stochastic volatility (RW-SV). Hence a direct comparison of LPS scores across countries is not meaningful.

In table 1 we report cumulative "pseudo" log predictive scores over the hold-out sample.[8] For both the one-step and four-steps ahead forecast horizon, we evaluate the predictive performance of the direct forecast for GDP growth (GDP direct), the composite forecast (GDP w(t-1)) and the composite forecast using optimized weights of the GDP's components (GDP w(opt.dir)). Predictive densities for the aggregate forecasts are evaluated using realized GDP growth. For completeness, we evaluate the joint predictive density of output growth and the expenditure

[7] *This might be related to the comparatively strong downturn in economic growth in 2009 in Hungary.*

[8] *See, for example, Kastner (2018), showing that the name pseudo arises from the fact that we approximate the predictive densities with a Gaussian distribution. From a practitioner's point of view, this strategy makes it easy to calculate both the joint predictive likelihood and marginal predictive likelihood for a subset of variables of interest (in this case, output growth and the expenditure components).*

components of GDP (joint). This serves as an overall measure of forecast performance for the variables of interest.[9] The figures for the best performing model in each column are in bold.

First, we want to answer the question whether GDP growth forecasts can be improved using a time-varying parameter model. The simplest form to capture time variation is by allowing error variances to vary over time, and its importance for forecasting has been demonstrated in a range of recent empirical contributions. Our results corroborate these findings, which can be seen by the inferior performance of the Minnesota prior with homoscedastic variance (Minnesota). Forecasts from this model are frequently outperformed even by simple univariate benchmark models with stochastic volatility. More interestingly, looking at models that accommodate time variation also in drifting parameters, we find variants of the threshold model outperforming constant parameter BVARs in Bulgaria, Croatia and mostly so in the Czech Republic. This holds true for both forecast horizons and regardless of whether the direct or composite forecasts are considered. In the Czech Republic and considering four-steps ahead forecasts, a BVAR that is linear in parameters and using optimized weights to aggregate the forecast yields a nearly identical performance though. For the remaining countries it turns out that the TTVP-SV model yields superior composite GDP forecasts for both forecast horizons and regardless of how the single component forecasts are aggregated. Constant parameter BVARs, however, excel when forecast performance is assessed using the direct GDP growth forecast. Also, in these instances, improvements over the TTVP-SV forecasts are modest. As a special case, Hungary is the only country where the regional model (Multi-NG-SV) shows a competitive forecast performance. Looking at direct GDP growth forecasts for the one- and four-steps ahead forecast horizon, this model even outperforms its competitors, indicating that cross-country linkages play an important role for forecasting Hungarian GDP growth.

Second, we draw attention to the question whether an aggregation of GDP components' forecasts can improve forecast performance compared to directly forecasting the GDP series. For most economies and at the one-step ahead forecast horizon, direct forecasts are slightly more accurate in terms of LPS scores. A reason for this could be that the predictive uncertainty of the components' forecasts aggregates when summing up the forecasts. Another reason could be that we do not model stock changes implicitly, assuming that their contribution is zero over the forecasting horizon. The finding that direct forecasts outperform aggregate forecasts is, however, not a general pattern. More specifically, composite forecasts in Croatia, the Czech Republic and Poland yield higher LPS scores than direct forecasts at the four-steps ahead forecast horizon. Only in Bulgaria do we find evidence that direct forecasts excel at both forecast horizons and by a great margin. An explanation for this could be the historically high contribution of statistical discrepancy in overall GDP growth, which is not captured by the components.

Third, our results allow us to examine the usefulness of specifying a regional multi-country model that takes into account the degree of economic integration among the countries under review. For most of the economies, the foreign sector plays a crucial role, and attempts have been made to better model the external sector

[9] *The LPS score of the joint predictive density is typically not identical to summing up the LPS scores of the marginal distributions of each variable of interest, since the latter would neglect cross-variable dependence.*

Table 1

Cumulative log predictive scores

	One-step ahead				Four-steps ahead			
	GDP direct	GDP w(t–1)	GDP w (opt. dir.)	Joint	GDP direct	GDP w(t–1)	GDP w (opt. dir.)	Joint
BG								
TTVP-SV BVAR								
ξ=1e-04	137.90	**127.33**	**135.30**	708.07	99.30	93.17	98.79	512.99
ξ=1e-05	149.88	118.43	129.06	699.91	109.00	95.38	102.30	512.59
ξ=1e-06	163.65	118.72	129.04	720.21	**114.54**	**97.36**	**104.49**	519.43
ξ=1e-07	158.54	118.60	128.94	715.22	112.52	97.33	103.87	517.40
ξ=1e-08	157.05	118.20	127.91	713.53	110.59	96.75	102.90	514.41
Constant parameter BVAR with SV								
Minesota-SV	163.05	112.25	124.69	**724.86**	111.99	85.98	94.46	**531.78**
NG-SV	162.09	116.64	126.14	711.48	108.62	92.63	97.17	508.61
Multi-NG-SV	147.90	110.91	122.52	689.87	107.60	84.89	93.40	501.19
Constant parameter BVAR no SV								
Minnesota	99.34	75.56	85.02	499.50	78.50	54.91	63.30	369.66
Univariate competitors with SV								
AR1-SV	136.78	94.03	105.31	660.71	103.15	64.49	74.35	484.44
RW-SV	137.83	101.51	113.28	668.07	101.26	70.17	80.73	468.01
CZ								
TTVP-SV BVAR								
ξ=1e-04	129.63	118.64	127.31	732.19	81.80	75.24	81.98	459.89
ξ=1e-05	145.95	133.04	141.76	814.19	100.97	98.98	102.68	587.15
ξ=1e-06	148.49	138.04	**147.37**	828.05	96.86	108.96	111.64	598.33
ξ=1e-07	150.26	137.90	146.74	830.47	107.20	**109.09**	111.66	614.77
ξ=1e-08	**151.28**	**138.39**	146.14	829.71	**110.13**	109.06	111.28	**618.99**
Constant parameter BVAR with SV								
Minnesota-SV	146.73	131.21	140.88	**831.48**	102.49	96.75	103.02	605.86
NG-SV	149.63	135.66	144.57	829.07	107.22	106.83	**111.68**	601.16
Multi-NG-SV	146.97	121.62	134.33	789.78	100.47	86.15	95.66	543.35
Constant parameter BVAR no SV								
Minnesota	103.99	81.89	92.45	559.65	80.43	62.03	70.88	429.88
Univariate competitors with SV								
AR1-SV	138.78	92.57	104.03	707.71	104.26	67.39	77.17	523.32
RW-SV	141.13	106.93	119.74	743.91	99.23	74.83	85.61	519.78
HR								
TTVP-SV BVAR								
ξ=1e-04	129.12	128.99	132.93	727.89	91.1	90.74	94.18	517.57
ξ=1e-05	**145.12**	136.79	**141.83**	777.24	106.82	106.01	107.69	600.36
ξ=1e-06	145.04	138.38	141.25	**783.45**	105.43	107.03	**109.25**	610.58
ξ=1e-07	143.15	**138.72**	139.84	780.19	105.82	108.06	109.04	615.62
ξ=1e-08	143.13	**138.72**	140.53	781.51	**107.1**	**109.41**	**109.25**	**620.25**
Constant parameter BVAR with SV								
Minnesota-SV	142.91	123.5	131.69	772.58	103.7	95.67	99.78	596.95
NG-SV	142.84	131.2	136.82	770.73	106.92	105.39	106.97	594.25
Multi-NG-SV	137.16	124.08	131.23	746.43	99.85	94.39	99.07	549.82
Constant parameter BVAR no SV								
Minnesota	96.65	79.58	86.85	507.22	73.94	60.59	66.71	378.22
Univariate competitors with SV								
AR1-SV	138.91	109.44	116.62	714.29	103.75	80.03	86.03	526.82
RW-SV	141.54	117.98	125.58	729.11	103.96	85.62	92.67	525.11

Source: Authors' calculations.

Note: Log predictive scores, cumulative over the hold-out sample. The left-hand part of the table refers to the one-step ahead forecast horizon, the right-hand part of the table refers to the four-steps ahead forecast horizon. "GDP direct" refers to a model's direct GDP growth forecast, "GDP w(t-1)" and "GDP w(opt.dir)" refers to GDP forecasts obtained by aggregating forecasts of GDP components as described in the main text. "Joint" refers to LPS of the joint predictive density for the variables of interest, namely GDP growth and growth of its expenditure components. The figures that refer to the best model are in bold.

Table 1 continued

Cumulative log predictive scores

	One-step ahead				Four-steps ahead			
	GDP direct	GDP w(t–1)	GDP w (opt. dir.)	Joint	GDP direct	GDP w(t–1)	GDP w (opt. dir.)	Joint
HU								
TTVP-SV BVAR								
ξ=1e-04	107.83	99.67	106.66	598.9	72.7	67.03	73.03	401.76
ξ=1e-05	125.59	109.61	114.47	648.49	87.69	82.56	84.18	457.31
ξ=1e-06	127.29	113.83	120.15	654.51	85.71	84.51	85.17	441.93
ξ=1e-07	121.46	114.66	121.11	645.45	80.13	85.1	84.61	434.1
ξ=1e-08	118.41	**115.37**	**122.19**	640.79	78.4	**85.63**	**85.27**	437.84
Constant parameter BVAR with SV								
Minnesota-SV	122.49	101.16	109.46	654.14	85.42	71.75	78.88	473.48
NG-SV	126.01	107.4	114.46	**655.98**	87.27	81.92	84.32	457.46
Multi-NG-SV	**128.47**	100.63	109.01	619.69	**89.94**	72.13	79.07	**458.36**
Constant parameter BVAR no SV								
Minnesota	87.56	57.7	67.86	465.66	68.14	42.63	51.1	343.42
Univariate competitors with SV								
AR1-SV	125.1	89.69	100.68	625.61	83.97	63.62	72.96	426.33
RW-SV	128.26	91.04	101.87	625.49	86.77	62.5	71.76	409.17
PL								
TTVP-SV BVAR								
ξ=1e-04	129.7	124.9	132.14	708.82	82.65	81.49	86.53	438.45
ξ=1e-05	147.09	132.81	144.88	780.95	108.12	106.65	112.45	551.48
ξ=1e-06	151.05	134.02	**145.72**	812.38	109.82	112.62	117.35	550.08
ξ=1e-07	150.18	**134.26**	145.21	811.54	112.56	**113.89**	**120.27**	578.83
ξ=1e-08	150.06	134.08	144.23	809.72	112.06	113.84	120.17	582.79
Constant parameter BVAR with SV								
Minnesota-SV	**152.86**	127.87	139.99	**823.24**	**115.63**	104.4	111.76	**585.71**
NG-SV	150.45	132.75	144.77	813.09	110.99	107.07	114.7	567.13
Multi-NG-SV	145.73	128.05	138.35	791.56	107.79	99.87	108.18	553.67
Constant parameter BVAR no SV								
Minnesota	105.41	79.57	91.21	548.87	83.35	63.8	73.53	411.62
Univariate competitors with SV								
AR1-SV	136.65	102.68	113.38	724.81	104.93	81.64	90.61	535.03
RW-SV	144.97	109.49	121.93	737.84	105.15	77.86	89.35	512.56
RO								
TTVP-SV BVAR								
ξ=1e-04	124.91	120.46	126.27	652.26	85.93	88.49	92.71	462.16
ξ=1e-05	130.11	**123.75**	**129.03**	669.99	97.77	**95.72**	**99.23**	501.21
ξ=1e-06	131.27	123.5	126.66	670.27	99.5	94.49	97.88	517.63
ξ=1e-07	129.91	122.93	126.05	670.83	99.18	93.74	96.91	521.9
ξ=1e-08	130.21	122.73	126.12	669.91	100.33	94.36	97.74	**522.68**
Constant parameter BVAR with SV								
Minnesota-SV	**133.26**	119.06	125.48	**672.72**	101.71	90.68	96.38	507.61
NG-SV	129	116.45	120.45	661.83	**102.07**	93.46	96.87	506.31
Multi-NG-SV	127.7	111.81	117.16	646.59	94.73	86.39	90.1	483.95
Constant parameter BVAR no SV								
Minnesota	95.78	85.88	90.64	502.68	75.01	63.32	67.77	369.82
Univariate competitors with SV								
AR1-SV	127.71	102.93	110.87	611.41	96.44	76.7	84.28	466.65
RW-SV	127.66	101.64	110	606.09	93.15	71.04	78.35	421.23

Source: Authors' calculations.

Note: Log predictive scores, cumulative over the hold-out sample. The left-hand part of the table refers to the one-step ahead forecast horizon, the right-hand part of the table refers to the four-steps ahead forecast horizon. "GDP direct" refers to a model's direct GDP growth forecast, "GDP w(t-1)" and "GDP w(opt.dir)" refers to GDP forecasts obtained by aggregating forecasts of GDP components as described in the main text. "Joint" refers to LPS of the joint predictive density for the variables of interest, namely GDP growth and growth of its expenditure components. The figures that refer to the best model are in bold.

(see e.g. Stoevsky, 2009, for Bulgaria, or Kolasa and Rubaszek, 2018, in the context of a DSGE framework). Note that by including euro area variables, the baseline models already control for developments in the most important export markets for the six economies. Slačík et al. (2014), however, take into account trade and economic links among the CESEE countries by including trade-weighted GDP in addition to euro area variables and show that this can further improve forecast performance. We go beyond the specification proposed in Slačík et al. (2014) by modeling all variables from all economies jointly, controling for euro area developments by including euro area GDP and interest rates as additional exogeneous regressors. Looking at the results, we do not find compelling evidence that forecasts improve if we directly take into account economic linkages among the countries. An exception to this is Hungary, for which a regional specification yields the best forecast at both forecast horizons.

So far, the evaluation of forecasts was based on the overall performance over the hold-out sample. It could be argued that certain models/model classes perform better during volatile times, while others perform better during normal times. In the extreme case, the excellent predictive performance of the TTVP-SV models could stem from a few data points such as turning points which models linear in parameters are not able to capture appropriately. To examine this in more detail, we have examined in a robustness exercise the performance of the different models for each time point in the hold-out sample. The results are available from the authors upon request. Briefly, we do not find evidence of time-specific swings in performance. In other words, the models that had a superior track record over the whole hold-out sample tend to perform equally well over the hold-out sample. This holds true for both forecast horizons, the joint density of the GDP components and the marginal GDP forecasts.

Summing up, we find that for most economies the TTVP-SV model tends to excel at both the one-step and four-steps ahead forecast horizon. Our specification of the Minnesota prior, however, turns out to be a tough competitor, and the forecast performance of both model classes is relatively close. Hence it is not surprising that from the different specifications of the TTVP-SV model, the one that uses a tight prior (ξ=10^{-8}) on coefficient movements tends to do a good job for all countries. For some countries, forecasts from multivariate models are competitive only at the end of the hold-out sample. These tend to be countries with shorter time series. For the remaining economies, the models that perform well do so equally over the hold-out sample, not showing any large swings in performance.

5.1 Sources of forecast error

In this section we delve deeper to analyze sources of forecast performance for the aggregate GDP growth forecasts. For that purpose, we focus on the TTVP-SV model with a tight prior (ξ=10^{-8}) that showed a reasonable performance for all countries.

To assess forecast performance of the aggregate forecast in more detail, we focus on two measures of forecast performance: the average mean forecast error (MFE) and the average root mean of weighted square forecast error (RMWSFE). The MFE serves to gauge which components are over- or underestimated, indicates how much the predictions vary around the realizations and considers already the components' share/weight in total GDP (Júlio and Esperanca, 2012).

Table 2

Evaluation measures for point estimates

	w (t–1)						w (opt. dir)					
	MFE 1Q	MFE 4Q	RMWSFE 1Q	RMWSFE 4Q	wi 1Q	wi 4Q	MFE 1Q	MFE 4Q	RMWSFE 1Q	RMWSFE 4Q	wi 1Q	wi 4Q
BG												
GDP composite	0.1024	0.0876	0.3109	0.4780			0.0521	0.0555	0.3308	0.4106		
GDP direct	–0.0585	–0.0865	0.2476	0.4514			–0.0585	–0.0865	0.2476	0.4514		
c	0.0464	0.0819	0.2544	0.4269	64.2	64.1	0.0491	0.0920	0.2691	0.4797	67.9	72
g	0.0057	0.0195	0.0989	0.1249	16	16.1	0.0058	0.0210	0.1005	0.1348	16.3	17.4
inv	0.0285	0.1491	0.1537	0.4205	21.4	21.7	0.0230	0.1104	0.1238	0.3113	17.3	16.1
x	0.1163	0.1918	0.4989	0.6021	58.7	57.5	0.0738	0.1290	0.3164	0.4048	37.2	38.7
m	–0.0694	–0.3024	0.6452	0.7861	60.3	59.5	–0.0445	–0.2247	0.4138	0.5841	38.7	44.2
Error / discrepancy	–0.0272	–0.0061	0.3292	0.2653			–0.0272	–0.0061	0.3292	0.2653		
HR												
GDP composite	0.0313	0.0336	0.2073	0.3107			0.0026	–0.0584	0.2080	0.3156		
GDP direct	0.0431	0.0517	0.2154	0.3421			0.0431	0.0517	0.2154	0.3421		
c	0.0273	0.0206	0.1173	0.2045	57.9	58.1	0.0309	0.0239	0.1326	0.2374	65.5	67.4
g	–0.0151	–0.0359	0.0555	0.0692	20.5	20.5	–0.0167	–0.0455	0.0612	0.0879	22.5	26
inv	0.0289	0.0412	0.1396	0.2676	21	20.9	0.0218	0.0368	0.1051	0.2390	15.8	18.7
x	0.0137	0.1069	0.3168	0.3648	42	41	0.0099	0.0939	0.2285	0.3205	30.3	36
m	–0.0072	–0.0924	0.2681	0.4755	41.3	40.5	–0.0060	–0.1098	0.2212	0.5652	34.1	48.2
Error / discrepancy	–0.0173	–0.0084	0.2024	0.1909			–0.0173	–0.0084	0.2024	0.1909		
CZ												
GDP composite	–0.0137	0.0205	0.2114	0.3520			–0.0302	0.0359	0.1790	0.3546		
GDP direct	–0.0017	0.0791	0.1791	0.3707			–0.0017	0.0791	0.1791	0.3707		
c	0.0033	0.0206	0.0847	0.1532	48.4	48.4	0.0035	0.0202	0.0873	0.1500	49.9	47.4
g	0.0068	0.0263	0.0462	0.1032	19.9	20	0.0058	0.0247	0.0397	0.0970	17.1	18.8
inv	0.0009	0.0256	0.0919	0.3640	26.4	26.4	0.0009	0.0236	0.0910	0.3356	26.1	24.4
x	–0.0454	–0.0127	0.3972	0.4242	74.8	73.6	–0.0301	–0.0085	0.2640	0.2861	49.7	49.7
m	0.0756	–0.0002	0.4086	0.5143	69.5	68.5	0.0466	–0.0001	0.2518	0.3024	42.8	40.3
Error / discrepancy	–0.0512	–0.0327	0.2400	0.2138			–0.0512	–0.0327	0.2400	0.2138		
HU												
GDP composite	0.0614	0.0340	0.2614	0.4936			0.1033	0.0548	0.2317	0.4840		
GDP direct	–0.0182	–0.0253	0.2030	0.5486			–0.0182	–0.0253	0.2030	0.5486		
c	0.0055	–0.0079	0.0795	0.2639	50.5	50.5	0.0052	–0.0074	0.0753	0.2447	47.8	46.8
g	0.0162	0.0150	0.0899	0.1612	21.1	21.2	0.0148	0.0143	0.0818	0.1541	19.2	20.3
inv	–0.0247	–0.0432	0.2719	0.5043	20.4	20.4	–0.0178	–0.0368	0.1966	0.4297	14.7	17.4
x	–0.0487	–0.0684	0.4761	0.7337	89.2	87.9	–0.0353	–0.0746	0.3448	0.7999	64.6	95.8
m	0.0365	0.0806	0.5158	0.9328	81.1	80	0.0208	0.0809	0.2946	0.9363	46.3	80.3
Error / discrepancy	0.0743	0.0581	0.3235	0.3117			0.0743	0.0581	0.3235	0.3117		
PL												
GDP composite	–0.0924	0.0176	0.2893	0.2489			–0.0274	–0.0187	0.2259	0.2282		
GDP direct	0.0346	0.1106	0.1837	0.3355			0.0346	0.1106	0.1837	0.3355		
c	–0.0013	0.0075	0.0358	0.1513	60.5	60.6	–0.0012	0.0080	0.0339	0.1608	57.2	64.4
g	0.0141	0.0300	0.0546	0.0933	18.1	18.2	0.0128	0.0306	0.0496	0.0951	16.5	18.6
inv	0.0021	0.0743	0.1588	0.3633	20.6	20.7	0.0016	0.0506	0.1233	0.2474	16	14.1
x	–0.0713	–0.0402	0.2292	0.3092	44.8	43.8	–0.0562	–0.0282	0.1806	0.2171	35.3	30.7
m	0.0731	–0.0061	0.4045	0.4653	44.1	43.3	0.0415	–0.0039	0.2294	0.2989	25	27.8
Error / discrepancy	–0.1076	–0.0406	0.2993	0.2140			–0.1076	–0.0406	0.2993	0.2140		
RO												
GDP composite	–0.0546	–0.0816	0.4064	0.6672			–0.0183	–0.1933	0.3604	0.4127		
GDP direct	–0.0561	–0.0649	0.3265	0.5270			–0.0561	–0.0649	0.3265	0.5270		
c	0.0643	0.0843	0.2915	0.5583	63.7	63.2	0.0776	0.0880	0.3515	0.5834	76.8	66
g	0.0183	0.0060	0.1251	0.1363	14.8	15	0.0152	0.0085	0.1038	0.1931	12.3	21.3
inv	0.0032	0.0715	0.3062	0.5432	25.9	26.1	0.0024	0.0679	0.2368	0.5157	20	24.8
x	–0.1178	–0.1752	0.3388	0.4890	39.6	38.6	–0.1173	–0.2292	0.3374	0.6398	39.4	50.5
m	–0.0257	–0.0213	0.2913	0.5421	44	42.9	–0.0284	–0.0311	0.3215	0.7911	48.5	62.6
Error / discrepancy	–0.0064	–0.0505	0.4284	0.3840			–0.0064	–0.0505	0.4284	0.3840		

Source: Authors' calculations.

Note: Summary measures for TTVP model with ξ=10^(-8). MFE refers to the mean forecast error, RMWSFE to the root mean of weighted square forecast error as defined in the main text and wi to weights associated to each component. The left-hand part of the table shows composite forecasts calculated using realized components' weights (w (t-1)), the right-hand part shows composite forecasts using optimized weights (w (opt.dir)). The line "Error / discrepancy" is calculated as the difference between overall GDP and aggregated GDP.

We define the forecast error $f_{t+h|h}$ for aggregate output growth and the forecast error vector of the components $f^z_{t+h|h}$ as

$$f_{t+h|h} = \left(gdp_{t+h|h} - gdp^{e.p.}_{t+h|h}\right) \text{ and}$$

$$f^z_{t+h} = \begin{pmatrix} c_{t+h|t} - c^{e.p.}_{t+h} \\ inv_{t+h|t} - inv^{e.p.}_{t+h} \\ g_{t+h|t} - g^{e.p.}_{t+h} \\ x_{t+h|t} - x^{e.p.}_{t+h} \\ -(m_{t+h|t} - m^{e.p.}_{t+h}) \end{pmatrix}, (10)$$

Following Marcellino et al. (2003) and Júlio and Esperanca (2012), the RMWSFE and MFE for composite output growth are given by $MFE = H^{-1}\Sigma_{t\epsilon H} f_{t+h|t}$ and $RMSFE = \sqrt{H^{-1}\sum_{t\in H} f^2_{t+h|t}}$ and analogously for the components $MWFE_z = \bar{w}_{z,t+h}\, H^{-1}\sum_{t\in H} f_{z,t+h|t}$ and $RMWSFE_z = \bar{w}_{z,t+h}\sqrt{H^{-1}\sum_{t\in H} f^2_{z,t+h|t}}$ with $\bar{w}_{z,t+h} = H^{-1}\sum_{t\in H} w_{z,t+h|t}$.

In table 2, we summarize forecast errors for the different GDP components. The left-hand part of the table contains results using historical weights and the right-hand part optimized weights.

A few general patterns emerge from the data. First, for Bulgaria and Hungary, aggregating forecasts with historical weights leads to an overprediction, while the direct forecast tends to underestimate GDP growth for both horizons. The opposite is true for Poland. For Romania, both the direct and aggregate forecasts underpredict real GDP growth. Second, looking at the predictive accuracy of the components, we find that the forecasts of export and import growth are most inaccurate for all countries. Both components are driven to a large extent by macroeconomic conditions abroad, which are apparently not easily captured within the modelling framework. For Romania and Bulgaria, two countries whose growth model is strongly underpinned by domestic consumption, private consumption forecasts also tend to be relatively inaccurate. These observations hold true for both one-quarter and four-quarters ahead forecasts. While short-term predictions for investment growth tend to be quite accurate, in the longer term, these predictions get more inaccurate for half of the countries. Last, looking at the optimized weights, we can see that these mostly follow actual shares of GDP components. The relative predictive inaccuracy for export and import growth is mirrored in smaller, relative shares of these two components, though. This holds true for most economies.

6 Conclusions

In this paper we forecast CESEE GDP growth using a range of Bayesian vector autoregressive time series models that are either suitable to handle large information sets or are flexible enough to handle gradual as well as abrupt changes in parameters. In accordance with the FORCEE model, the prevalent forecasting model of the OeNB for forecasting GDP growth in CESEE economies (Crespo Cuaresma et al., 2009), we condition on external developments by augmenting the models with euro area variables and using external assumptions on their development over the forecast horizon. Opting for a Bayesian approach, we compute predictive densities to assess uncertainty surrounding forecasts in a statistically coherent way.

First and foremost, we ask whether a forecasting framework that can accommodate structural changes in the economic environment improves forecast quality. Our results show that it is of central importance to allow residual variances to change over time – a finding that is in line with the recent literature (see e.g. Clark and Ravazzolo, 2015). We then examine the forecasting performance of a fully-fledged time-varying parameter model with both residual variances and parameters changing over time. Our findings indicate that it is indeed this most flexible specification that tends to best forecast CESEE growth. There is one caveat, though: the prior that governs parameter time variation has to be set very tight. In other words, allowing for a bit of time variation improves forecast performance, but allowing for too much leads to overfitting and poor forecasts. This is also corroborated by looking at results of constant parameter BVARs that turn out to be strong competitors.

Second, we investigate whether it is better to forecast GDP growth directly or to construct forecasts of its components and then sum these component forecasts. We propose two ways to aggregate forecasts, one which uses historical (realized) shares of GDP components in overall GDP and one where weights are optimized based on historical forecast performance. Our results show that direct forecasts tend to yield the best forecast performance but not by a great margin. A researcher that needs to conduct an economically consistent forecast might thus still successfully use the models tested in this study. Looking at forecast accuracy of single GDP components, we find that investment growth in the longer term and export and import growth seem to be particularly hard to forecast throughout the region. This is against the background that the multivariate models we tested already control for developments in the euro area, the most important trading partner for CESEE economies. Estimating a "regional" model for all CESEE economies together turns out to be no viable option since this model yields – with the exception of Hungary – inferior forecasts.

Our study sets the path for further research relevant for central bank forecasting. For example, we did not cover the issue of how to bring expert judgment – information that is not contained in the data – into the forecasting process. This could be achieved by "tilting" the predictive density forecasts of growth components of interest to a future path that matches expectations of an informed country expert. Alternatively, it would be possible to combine density forecasts with survey expectations, a framework that has been proposed by Kociecki et al. (2011). Finally, in order to improve overall forecast accuracy, a more accurate modeling of export, import and investment growth is essential. This could be achieved by a more precise account of global economic conditions or by the inclusion of forward-looking measures of uncertainty or soft data on the business climate.

References

Aastveit, K. A., A. Carriero, T. E. Clark and M. Marcellino. 2017. Have standard VARs remained stable since the crisis? In: Journal of Applied Econometrics 32(5). 931–951.

Armeanu, D., J. V. Andrei, L. Lache and M. Panait. 2017. A multifactor approach to forecasting Romanian gross domestic product (GDP) in the short run. In: PLoS ONE 12(7): e0181379.

Bańbura, M., D. Giannone and L. Reichlin. 2010. Large Bayesian vector auto regressions. In: Journal of Applied Econometrics 25(1). 71–92.

Belmonte, M. A., G. Koop and D. Korobilis. 2014. Hierarchical shrinkage in time-varying parameter models. In: Journal of Forecasting 33(1). 80–94.

Berg, T. O. and S. R. Henzel. 2015. Point and density forecasts for the euro area using Bayesian VARs. In: International Journal of Forecasting. Vol. 31 Issue 4. 1067–1095.

Bermingham, C. and A. D'Agostino. 2014. Understanding and forecasting aggregate and disaggregate price dynamics. In: Empirical Economics 46(2). 765–788.

Bitto, A. and S. Frühwirth-Schnatter. 2018. Achieving shrinkage in a time-varying parameter model framework. In: Journal of Econometrics. Forthcoming. arXiv preprint arXiv:1611.01310.

Brázdik, F. and M. Franta. 2017. A BVAR model for forecasting of Czech inflation. CNB Working Paper No. 7/2017.

Brüggemann, R. and H. Lütkepohl. 2013. Forecasting contemporaneous aggregates with stochastic aggregation weights. In: International Journal of Forecasting 29(1). 60–68.

Carriero, A., T. E. Clark and M. Marcellino. 2015. Bayesian VARs: specification choices and forecast accuracy. In: Journal of Applied Econometrics 30(1). 46–73.

Carriero, A., T. E. Clark and M. Marcellino. 2016. Common drifting volatility in large Bayesian VARs. In: Journal of Business & Economic Statistics 34(3). 375–390.

Carriero, A., G. Kapetanios and M. Marcellino. 2011. Forecasting large datasets with Bayesian reduced rank multivariate models. In: Journal of Applied Econometrics 26(5). 735–761.

Chan, J. C. and E. Eisenstat. 2018. Bayesian model comparison for time-varying parameter VARs with stochastic volatility. In: Journal of Applied Econometrics 33(4). 509–532.

Clark, T. E. and F. Ravazzolo. 2015. Macroeconomic forecasting performance under alternative specifications of time-varying volatility. In: Journal of Applied Econometrics 30(4). 551–575.

Cogley, T. and T. J. Sargent. 2005. Drifts and volatilities: monetary policies and outcomes in the post WWII US. In: Review of Economic Dynamics 8(2). 262–302.

Crespo Cuaresma, J., M. Feldkircher and F. Huber. 2016. Forecasting with global vector autoregressive models: a Bayesian approach. In: Journal of Applied Econometrics 31(7). 1371–1391.

Crespo Cuaresma, J., M. Feldkircher, T. Slačík and J. Wörz. 2009. Simple but effective: the OeNB's forecasting model for selected CESEE countries. In: Focus on European Economic Integration Q4/09. Vienna: OeNB. 84–95.

D'Agostino, A., L. Gambetti and D. Giannone. 2013. Macroeconomic forecasting and structural change. In: Journal of Applied Econometrics 28(1). 82–101.

Doan, T., R. B. Litterman and C. Sims. 1984. Forecasting and conditional projection using realistic prior distributions. In: Econometric Reviews 3(1). 1–100.

Eisenstat, E., J. C. Chan and R. W. Strachan. 2016. Stochastic model specification search for time-varying parameter VARs. In: Econometric Reviews 35(8-10). 1638–1665.

Franta, M., D. Havrlant and M. Rusnák. 2016. Forecasting Czech GDP using mixed-frequency data models. In: Journal of Business Cycle Research 12 (2). 165–185.

Geweke, J. and G. Amisano. 2010. Comparing and evaluating Bayesian predictive distributions of asset returns. In: International Journal of Forecasting 26 (2). 216–230

Geweke, J. and G. Amisano. 2011. Optimal prediction pools. In: Journal of Econometrics 164(1). 130–141.

Griffin, J. E. and P. J. Brown. 2010. Inference with normal-gamma prior distributions in regression problems. In: Bayesian Analysis 5(1). 171–188.

Hendry, D. F. and K. Hubrich. 2006. Forecasting economic aggregates by disaggregates. ECB Working Paper No. 589.

Hendry, D. F. and K. Hubrich. 2011. Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. In: Journal of Business & Economic Statistics 29(2). 216–227.

Hendry, D. F. and J. Muellbauer. 2018. The future of macroeconomics: macro theory and models at the Bank of England. In: Oxford Review of Economic Policy 34(1-2). 2g87–328.

Huber, F. and M. Feldkircher. 2017. Adaptive shrinkage in Bayesian vector autoregressive models. In: Journal of Business & Economic Statistics. 1–13.

Huber, F., G. Kastner and M. Feldkircher. 2018. Should I stay or should I go? A latent threshold approach to large-scale mixture innovation models. In: Journal of Applied Econometrics. Forthcoming.

Hubrich, K. 2005. Forecasting euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy? In: International Journal of Forecasting. Vol. 21. Issue 1. 119–136.

Jacquier, E., N. G. Polson and P. Rossi. 1994. Bayesian Analysis of Stochastic Volatility Models. In: Journal of Business & Economic Statistics 12(4). 371–89.

Júlio, P. and P. Esperanca. 2012. Evaluating the forecast quality of GDP components: An application to G7. GEE Papers 47.

Kastner, G. 2016. Dealing with stochastic volatility in time series using the R package stochvol. In: Journal of Statistical Software 69(5). 1–30.

Kastner, G. 2018. Sparse Bayesian time-varying covariance estimation in many dimensions. In: Journal of Econometrics. Forthcoming. arXiv preprint arXiv:1608.08468.

Kastner, G. and S. Frühwirth-Schnatter. 2014. Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. In: Computational Statistics & Data Analysis 76. 408–423.

Kim, S., N. Shephard and S. Chib. 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. In: The Review of Economic Studies 65(3). 361–393.

Kociecki, A., M. Kolasa and M. Rubaszek, 2011. Predictivistic Bayesian Forecasting System. National Bank of Poland Working Papers 87/2011.

Kolasa, M. and M. Rubaszek. 2018. Does the foreign sector help forecast domestic variables in DSGE models? National Bank of Poland Working Papers 282/2018.

Koop, G. 2013. Forecasting with medium and large Bayesian VARs. In: Journal of Applied Econometrics 28(2). 177–203.

Kunovac, D. and B. Špalat. 2014. Nowcasting GDP using available monthly indicators. Croatian National Bank Working Papers W-39 (October).

Litterman, R. B. 1986. Forecasting with Bayesian vector autoregressions – five years of experience. In: Journal of Business & Economic Statistics 4(1). 25–38.

Lütkepohl, H. 2011a. Forecasting aggregated time series variables. In: OECD Journal: Journal of Business Cycle Measurement and Analysis 2010(2). 1–26.

Lütkepohl, H. 2011b. Forecasting nonlinear aggregates and aggregates with time-varying weights. In: Jahrbücher für Nationalökonomie und Statistik 231(1). 107–133.

Marcellino, M., J. H. Stock and M. W. Watson. 2003. Macroeconomic forecasting in the euro area: Country specific versus area-wide information. In: European Economic Review 47(1). 1–18.

Primiceri, G. E. 2005. Time varying structural vector autoregressions and monetary policy. In: The Review of Economic Studies 72(3). 821–852.

Ravazzolo, F. and S. Vahey. 2014. Forecast densities for economic aggregates from disaggregate ensembles. In: Studies in Nonlinear Dynamics & Econometrics 18(4). 367–381.

Ravnik, R. 2014. Short-term forecasting of GDP under structural changes. Croatian National Bank, Working Papers W-40/2014.

Riedl, A. and J. Wörz. 2018. A simple approach to nowcasting GDP growth in CESEE economies. In: Focus on European Economic Integration Q4/18. Vienna: OeNB.

Sims, C. and T. Zha. 1998. Bayesian methods for dynamic multivariate models. In: International Economic Review 39(4). 949–968.

Slaćík, T., K. Steiner and J. Wörz. 2014. Can trade partners help better FORCEE the future? Impact of trade linkages on economic growth forecasts in selected CESEE countries. In: Focus on European Economic Integration Q1/14. Vienna: OeNB. 36–56.

Stoevsky, G. 2009. Econometric forecasting of Bulgaria's export and import flows. Bulgarian National Bank Discussion Papers DP/77/2009.

Timmermann, A. 2006. Forecast combinations. In: Handbook of economic forecasting 1. 135–196.

Appendix

Prior distributions

Since we employ Bayesian estimation methods, we need to impose suitable prior distributions on all parameters. As the majority of proposed models feature stochastic volatility, we follow Kastner and Frühwirth-Schnatter (2014) for the prior specification on parameters of equation (4). For the TTVP-SV model, we use an inverse gamma prior for $\sigma^2_{ic,1}$,

$$\sigma^2_{ic,1} \sim IG\left(\theta_{ic,0}, \theta_{ic,1}\right),$$

with $\theta_{ic,0} = \theta_{ic,1} = 0.01$ (Huber et al., 2018).
Moreover, we use shrinkage priors on the coefficients to avoid overfitting issues and improve forecasts of medium-scale (time-varying parameter) VAR. In constant parameter BVARs we impose either a normal-gamma shrinkage prior as put forward by Griffin and Brown (2010) and extended for VAR models by Huber and Feldkircher (2017) or a hierarchical Minnesota prior on the VAR coefficients. In TTVP-SV models we employ a normal-gamma prior on the time-invariant part of the coefficients.[10]

Following Sims and Zha (1998) and Feldkircher and Huber (2017), we specify a fully Bayesian Minnesota prior, integrating out the hyperparameters generally set by the researcher. The hierarchical priors already allow a great amount of flexibility, avoiding excessive shrinkage. We therefore integrate out three hyperparameters controlling the degree of shrinkage of a) own lags of endogenous variables, b) lags of other variables and c) the intercept and exogenous variables. When specifying a gamma prior on the two hyperparameters of a) and b), we obtain the marginal likelihood by specifying a Metropolis Hastings algorithm (Huber and Feldkircher, 2017). Moreover, when the hierarchical Minnesota prior is used, the elements of the lower triangular matrix L_c are centered on a Gaussian prior with relatively little information.

Following Huber and Feldkircher (2017), we specify a lag-wise normal-gamma prior on the coefficient matrices, imitating the Minnesota prior (Doan et al., 1984;

[10] *Note that a model linear in parameters is nested in the TTVP-SV specification, if the model is fully shrunk toward a constant parameter model.*

Litterman, 1986), implying increasing lag orders are shrunk toward zero to a higher degree. Additionally, we take advantage of the triangularization algorithms to treat the elements of L_c similar to VAR coefficients and thus also place a normal-gamma prior on these parameters, but with less shrinkage (Huber and Feldkircher, 2017).