Received: 19 January 2018 | Accepted: 11 February 2018

DOI: 10.1002/for.2524

RESEARCH ARTICLE

WILEY

Predicting crypto-currencies using sparse non-Gaussian state space models

Christian Hotz-Behofsits[1] | Florian Huber[2] | Thomas Otto Zörner[2]

[1]Department of Marketing, Vienna University of Economics and Business, Vienna, Austria

[2]Department of Economics, Vienna University of Economics and Business, Vienna, Austria

Correspondence
Florian Huber, Department of Economics, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria.
Email: fhuber@wu.ac.at

Funding information
Czech Science Foundation, Grant/Award Number: 17-14263S

Abstract

In this paper we forecast daily returns of crypto-currencies using a wide variety of different econometric models. To capture salient features commonly observed in financial time series like rapid changes in the conditional variance, non-normality of the measurement errors and sharply increasing trends, we develop a time-varying parameter VAR with *t*-distributed measurement errors and stochastic volatility. To control for overparametrization, we rely on the Bayesian literature on shrinkage priors, which enables us to shrink coefficients associated with irrelevant predictors and/or perform model specification in a flexible manner. Using around one year of daily data, we perform a real-time forecasting exercise and investigate whether any of the proposed models is able to outperform the naive random walk benchmark. To assess the economic relevance of the forecasting gains produced by the proposed models we, moreover, run a simple trading exercise.

KEYWORDS

Bitcoin, density forecasting, stochastic volatility, *t*-distributed errors

1 | INTRODUCTION

In the present paper we develop a non-Gaussian state-space model to predict the price of three crypto-currencies. Taking a Bayesian stance enables us to introduce shrinkage into the modeling framework, effectively controlling for model and specification uncertainty within the general class of state-space models. To control for potential outliers we propose a time-varying parameter vector autoregressive (VAR) model (Cogley & Sargent, 2005; Primiceri, 2005) with heavy-tailed innovations,[1] as well as a stochastic volatility specification of the error variances. Since the literature on robust determinants of price movements in crypto-currencies is relatively sparse (for an example, see Cheah & Fry, 2015), we apply Bayesian shrinkage priors to decide whether using information from a set of potential predictors improves predictive accuracy.

The recent price dynamics of various crypto-currencies point towards a set of empirical key features that an appropriate modeling strategy should accommodate. First, conditional heteroskedasticity appears to be an important regularity commonly observed (Chu, Chan, Nadarajah, & Osterrieder, 2017). This implies that volatility is changing over time in a persistent manner. If this feature is neglected, predictive densities are either too wide (during tranquil times) or too narrow (in the presence of tail events, i.e., pronounced movements in the price of a

[1]For a recent exposition on how to introduce flexible error distributions into VAR models with stochastic volatility, see Chiu, Mumtaz, and Pintér (2017).

given asset).[2] Second, the conditional mean of the process is changing. This implies that, within a standard regression framework, the relationship between an asset price and a set of exogenous covariates is time varying. In the case of various crypto-currencies this could be due to changes in the degree of adoption of institutional and/or private investors, regulatory changes, issuance of additional crypto-currencies or general technological shifts (Böhme, Christin, Edelman, & Moore, 2015). Thus it might be necessary to allow for such shifts by means of time-varying regression coefficients. Third, and finally, varios crypto-currencies display a rather strong degree of co-movment with each other (see Urquhart, 2017). In our paper, we consider Bitcoin, Ethereum, and Litecoin—three popular choices. All three tend to be strongly correlated with each other, implying that a successful econometric framework should incorporate this information.

The goal of this paper is to systematically assess how different empirically relevant forecasting models perform when used to predict daily changes in the price of Bitcoin, Ethereum, and Litecoin. The models considered include a wide range of univariate and multivariate models that are flexible along several dimensions. We consider VARs that feature drifting parameters as well as time-varying error variances. To cope with the curse of dimensionality we introduce recent shrinkage priors (see Feldkircher, Huber, & Kastner, 2017) and a flexible specification for the law of motion of the regression parameters (Huber, Kastner, & Feldkircher, 2017). In addition, we introduce a heavy-tailed measurement error distribution to capture potential outlying observations (see, among others, Carlin, Polson, & Stoffer, 1992; Geweke & Tanizaki, 2001).

We jointly forecast the three crypto-currencies considered by using daily data from October 2016 to October 2017, with the last 160 days being used as a hold-out period. In a forecasting comparison, we find that time-varying parameter VARs with some form of shrinkage perform well, beating univariate benchmarks like the AR(1) model with stochastic volatility (SV) as well as a random walk with SV. Constant-parameter VARs tend to be inferior to their counterparts that feature time-varying parameters, but still prove to be relevant competitors. Especially during days which are characterized by large price changes, controlling for heteroskedasticity in combination with a flexible error variance–covariance structure pays off in terms of predictive accuracy. These findings are generally corroborated by considering probability integral transforms, showing that more flexible models lead to better calibrated predictive distributions. Moreover, a trading exercise provides a comparable picture. Models that perform well in terms of predictive likelihoods also tend to do well when used to generate trading signals.

The remainder of this paper is structured as follows. Section 2 provides an overview of the data as well as empirical key features of the three crypto-currencies considered. Moreover, this section details how the additional explanatory variables are constructed. Section 3 introduces the econometric framework adopted, providing a brief discussion of the model as well as the Bayesian prior setup and posterior simulation. Section 4 presents the empirical forecasting exercise, while Section 5 focuses on applying the proposed models to perform portfolio allocation tasks. Finally, the last section summarizes and concludes the paper.

2 | EMPIRICAL KEY FEATURES

In this section we first identify important empirical key features of crypto-currencies and then propose a set of covariates that aim to explain the low-to-medium frequency behavior of the underlying price changes.

For the present paper, we focus on the daily change in the log price of Bitcoin, Ethereum, and Litecoin. To explain movements in the price of the three crypto-currencies considered, we include information on equity prices (measured through the log returns of the S&P 500 index), the relative number of search queries for each respective crypto-currency from Google trends, the number of English Wikipedia page views as well as the difference between the weekly cumulative price trend from common mining hardware and similar, but mining-unsuitable, GPU-related products to capture the effect of supply-side factors.

The data spans the period from 26 November 2016 to 3 October 2017, yielding a panel of 316 daily observations. Bitcoin, Ethereum, and Litecoin closing prices are taken from a popular crypto-currency meta-platform.[3] They originate from major crypto exchanges and are averaged according to their daily trading volume. Furthermore, alternative financial investments are represented by the S&P 500 indices daily closing prices. Additionally, demand-side predictors like the relative number of worldwide search operations from Google Trends and the number of Wikipedia page views (in English) are used. Because large-scale crypto-currency mining impacts supply and prices for the required equipment at the same time, hardware price trends are utilized to express changes in supply. To capture these effects, we gather GPU prices from Amazon's bestseller lists and extract the price trend of common mining hardware. We construct this predictor

[2]Controlling for heteroskedasticity in macroeconomic and financial data proves to be an important task when it comes to prediction; see Clark (2011); Clark and Ravazzolo (2015); Huber and Feldkircher (2017).

[3]For more information, see coinmarketcap.com.

by computing the difference between the weekly cumulative price trend from common mining hardware (e.g., AMD Radeon RX 480 graphic cards) and similar, but GPU-related products that are unsuitable for mining activities (e.g., an AMD Radeon R5 230 graphics card).

To provide additional information on the recent behavior of crypto-currencies, Figure 1 presents the log returns (left panel) as well as the squared log returns (right panel) for all three currencies under scrutiny. At least two features are worth emphasizing. First, notice that in the first part of the sample (i.e., the end of 2016 and the beginning of 2017), price changes have been comparatively small. This can be seen in both panels of the figure and for Bitcoins and Litecoins. For Ethereum, the pattern is slightly different, but we still observe a general increase in variation during the second part of 2017.

Second, the degree of co-movement between the three currencies increased markedly in 2017, where most major peaks and troughs coincide. This carries over to the squared returns, where we find that especially the sharp increase in volatility in September 2017 was common to all three currencies considered.

These two empirical regularities suggest that the proposed model should be able to capture co-movement between Bitcoin, Litecoin, and Ethereum prices as well as changes in the first moment of the sampling density. Moreover, the right panel indicates that large shocks appear to be quite common, calling for a flexible error distribution that allows for heteroskedasticity.

In order to provide further information on the amount of co-movement in our dataset, Figure 2 shows a heatmap of the lower Cholesky factor of the empirical correlation matrix of the nine time series included. The upper part of the figure reveals that all three assets display a pronounced degree of co-movement. This indicates that each individual time series might carry important information on the behavior of the remaining two time series, pointing towards the necessity to control for this empirical regularity. For the remaining factors we do find nonzero correlation but these correlations appear to be rather muted. Nevertheless, we conjecture that the set of fundamentals above should be a reasonable starting point to explain movements in the price of crypto-currencies.

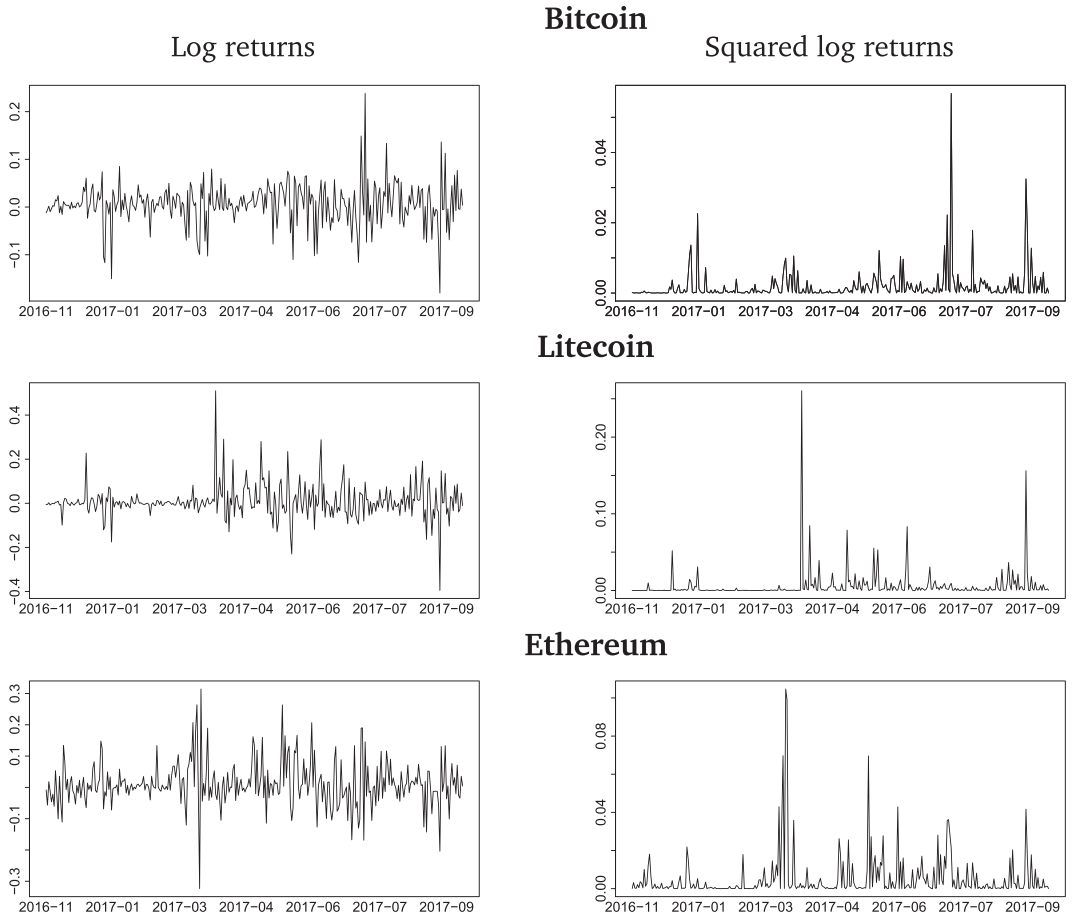

FIGURE 1 Data overview: logarithmic returns and squared logarithmic returns

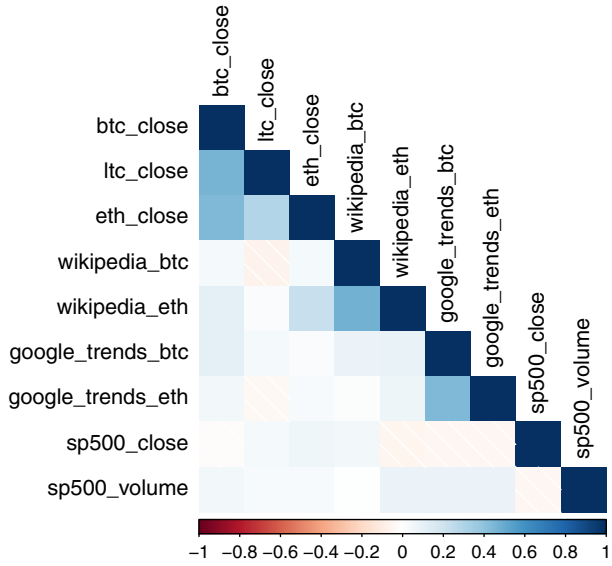

FIGURE 2 Lower Cholesky factor of the empirical correlation matrix of the dataset used [Colour figure can be viewed at wileyonlinelibrary.com]

3 | ECONOMETRIC FRAMEWORK

3.1 | A multivariate state-space model

To capture the empirical features of the three crypto-currencies, a flexible econometric model is needed. We assume that the three crypto-currencies as well as the additional covariates are stored in an m-dimensional vector $\{\boldsymbol{y}_t\}_{t=1}^{T}$ that follows a VAR(p) model with time-varying coefficients:

$$\boldsymbol{y}_t = \boldsymbol{A}_{1t}\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_{pt}\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{1}$$

with $\boldsymbol{A}_{jt}$ (for j=1,...,p) being a set of $m\times m$-dimensional coefficient matrices and $\boldsymbol{\varepsilon}_t$ is a multivariate vector of reduced form shocks with a time-varying variance–covariance matrix:

$$\boldsymbol{\Sigma}_t = \boldsymbol{U}_t\boldsymbol{H}_t\boldsymbol{U}_t'. \tag{2}$$

Hereby we let $\boldsymbol{U}_t$ be a lower unitriangular matrix with diag=$\boldsymbol{\iota}_m$ and $\boldsymbol{\iota}_m$ being an m-dimensional vector of ones. Moreover, $\boldsymbol{H}_t$ is a diagonal matrix with typical diagonal element $[\boldsymbol{H}_t]_{jj} = e^{h_{jt}}$. The logarithmic volatilities are assumed to follow an AR(1) process:

$$h_{jt} = \mu_j + \rho_j(h_{jt-1} - \mu_j) + \varsigma_j\nu_{jt}, \quad \nu_{jt}\sim\mathcal{N}(0,1), \tag{3}$$

where μ_j denotes the unconditional mean of the log-volatility process, while ρ_j and ς_j are the persistence and variance parameters, respectively.

Following Carriero, Clark, and Marcellino (2015) and Feldkircher et al. (2017), we rewrite Equation 1 as follows:

$$\tilde{\boldsymbol{U}}_t\boldsymbol{\varepsilon}_t = \boldsymbol{\eta}_t, \tag{4}$$

where $\tilde{\boldsymbol{U}}_t = \boldsymbol{U}_t^{-1}$ and $\boldsymbol{\eta}_t$ is a vector of orthogonal shocks with a time-varying variance–covariance matrix.

Note that the ith equation (for $i>1$) of this system can be written as

$$y_{it} = \boldsymbol{A}_{i\bullet,t}\boldsymbol{x}_t - \sum_{j=1}^{i-1}\tilde{u}_{ij}\varepsilon_{jt} + \eta_{it}. \tag{5}$$

We let $\boldsymbol{x}_t=(\boldsymbol{y}'_{t-1},...,\boldsymbol{y}'_{t-p})'$ be the stacked vector of covariates and $\boldsymbol{A}_t=[\boldsymbol{A}_{1t},...,\boldsymbol{A}_{pt}]$ is the $m\times mp$ matrix of stacked coefficients with $\boldsymbol{A}_{i\bullet,t}$ selecting the ith row of the matrix concerned. Equation 5 is a standard linear regression model with heteroskedastic innovations and the (negative) of the reduced-form shocks of the preceding $i-1$ equations as additional regressors. In the case of i=1, Equation 5 reduces to a simple univariate regression with $\boldsymbol{x}_t$ as covariates. It proves to be convenient to rewrite Equation 5 as follows:

$$y_{it} = \boldsymbol{\beta}_{it}'\boldsymbol{z}_{it} + \eta_{it}, \tag{6}$$

where $\boldsymbol{\beta}_{it} = (\boldsymbol{A}_{i\bullet}, \tilde{u}_{i1}, ..., \tilde{u}_{ii-1})'$ is a $k_i = mp + (i-1)$-dimensional vector of regression coefficients and $\boldsymbol{z}_{it}=[\boldsymbol{x}'_t, -\varepsilon_{1t}, ..., -\varepsilon_{i-1,t}]'$. One important implication of Equation 6 is that the covariance parameters are effectively estimated in one step alongside the VAR coefficients.

We assume that $\boldsymbol{\beta}_{it}$ evolves according to a random walk process:

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_{it-1} + \boldsymbol{e}_{it}. \tag{7}$$

The shocks to the states $\boldsymbol{e}_{it}\sim\mathcal{N}(\boldsymbol{0},\Theta_i)$ follow a Gaussian distribution with diagonal variance–covariance matrix $\Theta_i = \text{diag}(\vartheta_{i1},...,\vartheta_{ik_i})$. To facilitate variable selection/shrinkage we follow Frühwirth-Schnatter and Wagner (2010), Belmonte, Koop, and Korobilis (2014), and Bitto and Frühwirth-Schnatter (2016), and rewrite the model given by Equations 6 and 7 as follows:

$$y_{it} = \boldsymbol{\beta}_{i0}'\boldsymbol{z}_{it} + \tilde{\boldsymbol{\beta}}_{it}'\sqrt{\Theta_i}\boldsymbol{z}_{it} + \eta_{it}, \tag{8}$$

$$\tilde{\boldsymbol{\beta}}_{it} = \tilde{\boldsymbol{\beta}}_{it-1} + \boldsymbol{\xi}_{it}, \quad \boldsymbol{\xi}_{it}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_{k_i}), \tag{9}$$

$$\tilde{\boldsymbol{\beta}}_{i0} = \boldsymbol{0}. \tag{10}$$

The matrix $\sqrt{\boldsymbol{\Theta}_i}$ is a matrix square root such that $\boldsymbol{\Theta}_i = \sqrt{\boldsymbol{\Theta}_i}\sqrt{\boldsymbol{\Theta}_i}$ with typical element $\sqrt{\vartheta_{ij}}$ and $\tilde{\beta}'_{ij,t}$ the jth

element of $\tilde{\boldsymbol{\beta}}_{it}'$ reads $(\beta_{ij,t}-\beta_{ij,0})/\sqrt{\vartheta_{ij}}$. This parametrization, labeled the noncentered parametrization, implies that the state innovation variances are moved into the observation equation (see Equation 8) and treated as standard regression coefficients. Thus, if $\sqrt{\vartheta_{ij}}=0$, the coefficient associated with the jth element in $\boldsymbol{z}_{it}$ is constant over time.

Up to this point we have remained silent on the distributional assumptions on the measurement errors. In what follows we depart from the literature on TVP-VARs and assume that the measurement errors are heavy tailed and follow a t-distribution. This choice is based on evidence in the literature (Gallant, Hsieh, & Tauchen, 1997; Geweke, 1994; Jacquier, Polson, & Rossi, 2004) which calls for heavy-tailed distributions when used to model daily financial market data. As can be seen in Figure 1, we also observe multiple outlying observations for all three crypto-currencies under consideration.

Since the assumption of non-Gaussian errors would render typical estimation methods like the Kalman filter infeasible, we follow Harrison and Stevens (1976), West (1987), and Gordon and Smith (1990), and use a scale mixture of Gaussians to approximate the t-distribution:

$$\eta_{it}|h_{it}\sim t_{\nu_i}(0,e^{h_{it}})\Leftrightarrow\eta_{it}|h_{it},\phi_{it}\sim\mathcal{N}(0,\phi_{it}e^{h_{it}}), \tag{11}$$

$$\phi_{it}|\nu_i\sim\mathcal{G}^{-1}(\nu_i/2,\nu_i/2). \tag{12}$$

Note that the degree-of-freedom parameter ν_i is equation specific, implying that the excess kurtosis of the underlying error distribution is allowed to change across equations, a feature that might be important given the different time series involved. The latent process ϕ_{it} simply serves to rescale the Gaussian distribution in the case of large shocks. Note that if $\phi_{it}=1$ for all i,t we obtain the standard time-varying parameter VAR as in Primiceri (2005).

3.2 | Prior specification

The prior setup adopted closely follows Feldkircher et al. (2017). More specifically, we use a normal-gamma (NG) shrinkage prior on the elements of $\boldsymbol{\beta}_{i0}$ and $\sqrt{\Omega_i}$.

The NG prior comprises a Gaussian prior on the coefficients alongside a set of local and global shrinkage parameters for the first mp elements of $\boldsymbol{\beta}_{i0}$ and $\mathrm{diag}(\sqrt{\Omega_i})$:

$$\beta_{ij,0}|\tau^2_{\beta,ij}\sim\mathcal{N}(0,\tau^2_{\beta,ij}), \tag{13}$$

$$\sqrt{\vartheta_{ij}}|\tau^2_{\vartheta,ij}\sim\mathcal{N}(0,\tau^2_{\vartheta,ij}), \tag{14}$$

for $i=1,\ldots,m$ and $j=1,\ldots,mp$. Here we let $\tau^2_{s,ij}$(for $s\in\{\beta,\vartheta\}$) denote local shrinkage parameters with

$$\tau^2_{s,ij}|\lambda_L\sim\mathcal{G}\left(\kappa,\frac{\kappa\lambda_L}{2}\right), \tag{15}$$

where κ is a hyperparameter specified by the researcher and λ_L is a global shrinkage parameter that is lag specific, that is, applied to the elements in $\boldsymbol{\beta}_{i0}$ and $\sqrt{\Omega_i}$ associated with the Lth lag of $\boldsymbol{y}_t$, and constructed as follows:

$$\lambda_L=\prod_{l=1}^{L}\pi_l,\quad \pi_l\sim\mathcal{G}(c_0,d_0). \tag{16}$$

This implies that if $\pi_l>1$, the prior introduces more shrinkage with increasing lag orders. The degree of overall shrinkage is controlled through the hyperparameters c_0 and d_0.

Note that this specification pools the parameters that control the amount of time variation as well as the time-invariant regression parameters. This captures the notion that if a variable is not included initially, the probability of having a time-varying coefficient also decreases (by increasing the lag-specific shrinkage parameter λ_L).

For the covariance parameters indexed by $j=mp+1,\ldots,k_i$ the prior is specified analogously to Equations 13 and 14 but with λ_L replaced by ϱ. This choice implies that all covariance parameters as well as the corresponding process innovation variances are pushed to zero simultaneously. For ϱ we again use a Gamma distributed prior:

$$\varrho\sim\mathcal{G}(a_0,b_0), \tag{17}$$

with a_0, b_0 being hyperparameters.

This prior specification has the convenient property that the parameters λ_L and ϱ introduce prior dependence, pooling information across different coefficient types (i.e., regression coefficients and process innovation variances), introducing strong *global* shrinkage on all coefficients concerned. By contrast, the introduction of the local scaling parameters $\tau_{s,ij}$ serves to provide flexibility in the presence of strong overall shrinkage introduced by λ_L and ϱ. Thus, even if the aforementioned global scaling parameters are large (i.e., heavy shrinkage is introduced in the model), the local scalings provide sufficient flexibility to drag away posterior mass from zero and allowing for nonzero coefficients. The role of the hyperparameter κ is to control the tail behavior of the prior. If κ is small (close to zero), the prior places more mass on zero but the tails of the marginal prior obtained after integrating over the local scales become thicker (see Griffin & Brown, 2010, for a discussion).

For the parameters of the log-volatility equation (Equation 3) we follow Kastner and Frühwirth-Schnatter

(2014) and Kastner (2015a) and use a normally distributed prior on $\mu_j \sim \mathcal{N}(0, 10^2)$, a Beta prior on $\frac{\rho_j + 1}{2} \sim \mathcal{B}(25, 5)$ and a Gamma prior on $\varsigma_j \sim \mathcal{G}(1/2, 1/2)$. In addition, we specify a uniform prior on $v_i \sim \mathcal{U}(2, 20)$, effectively ruling out the limiting case of a Gaussian distribution if v_i becomes excessively large.

3.3 | Full conditional posterior simulation

Estimation of the model is carried out using Markov chain Monte Carlo (MCMC) techniques. Our MCMC algorithm consists of the following blocks:

1. Conditional on the remaining parameters/states in the model, simulate the full history of $\{\tilde{\boldsymbol{\beta}}_{it}\}_{t=1}^{T}$ using a forward-filtering backward sampling algorithm (Carter & Kohn, 1994; Frühwirth-Schnatter, 1994) on an equation-by-equation basis.
2. The full history of the log-volatility process as well as the parameters of Equation 3 are obtained by relying on the algorithm proposed in Kastner and Frühwirth-Schnatter (2014) and implemented in the R package stochvol (Kastner, 2015b).
3. The time-invariant components $\boldsymbol{\beta}_{i0}$ as well as $\theta_i = \text{diag}(\boldsymbol{\Theta}_i)$ are simulated from a multivariate Gaussian posterior that takes a standard form (see Feldkircher et al., 2017).
4. The sequence of local scaling parameters is simulated from a generalized inverted Gaussian (GIG) distributed posterior distribution given by

$$\tau_{\beta,ij} | \bullet \sim \mathcal{GIG}(\kappa - 1/2, \beta_{ij,0}^2, \kappa\lambda_L), \tag{18}$$

$$\tau_{\vartheta,ij} | \bullet \sim \mathcal{GIG}(\kappa - 1/2, \vartheta_{ij,0}^2, \kappa\lambda_L) \tag{19}$$

   for $j \in \mathcal{A}_L$. The posterior distribution for the scalings associated with the covariance parameters is similar, with λ_L replaced by ϱ.
5. We obtain draws from the posterior of the lag-specific shrinkage parameter associated with the lth lag by combining the likelihood $\prod_{i=1}^{m} \prod_{j \in \mathcal{A}_l} p(\tau_{\beta,ij}^2, \tau_{\vartheta,ij}^2 | \pi_l, \lambda_{l-1})$ with the prior on π_l. The resulting posterior distribution is a Gamma distribution:

$$\pi_l | \bullet \sim \mathcal{G}\left(c_0 + \kappa R, d_0 + \lambda_{l-1} \frac{\kappa}{2} \sum_{i=1}^{m} \sum_{j \in \mathcal{A}_l} (\tau_{\beta,ij}^2 + \tau_{\vartheta,ij}^2)\right), \tag{20}$$

   with the $\bullet$ indicating the conditioning on everything else, $R = 2pm^2$ and $\lambda_0 = 1$. The set $\mathcal{A}_l$ selects all coefficients associated with the lth lag of $\boldsymbol{y}_t$.

   Similarly, the conditional posterior of ϱ is given by

$$\varrho | \bullet \sim \mathcal{G}\left(a_0 + \kappa v, b_0 + \frac{\kappa}{2} \sum_{i=1}^{m} \sum_{j=mp+1}^{k_i} (\tau_{\beta,ij}^2 + \tau_{\vartheta,ij}^2)\right), \tag{21}$$

   where $v = m(m-1)$ denotes the number of covariance parameters in addition to the number of process variances for the corresponding parameters.
6. The full history of $\{\phi_{it}\}_{t=1}^{T}$ is obtained by independently simulating from an inverted Gamma distribution (see Kastner, (2015c)):

$$\phi_{it} | \bullet \sim \mathcal{G}^{-1}\left(\frac{v_i + 1}{2}, \frac{v_i + \eta_{it}^2 e^{-h_{it}}}{2}\right), \tag{22}$$

   for $t = 1, \ldots, T$.
7. To simulate the degrees of freedoms v_i, we perform an independent Metropolis–Hastings (MH) step described in Kastner (2015c).

This algorithm is repeated a large number of times with the first N^{burn} observations being discarded as burn-in.[4] Note that the equation-by-equation algorithm yields significant computational gains relative to competing estimation algorithms that rely on full-system estimation of the VAR model.

4 | FORECASTING RESULTS

4.1 | Model specification and design of the forecasting exercise

In this section, we briefly describe model specification and the design of the forecasting exercise. The prior setup for our benchmark specification (henceforth labeled the t-TVP NG) model closely follows the existing literature on NG shrinkage priors (Bitto & Frühwirth-Schnatter, 2016; Feldkircher et al. 2017; Griffin & Brown, 2010; Huber & Feldkircher, 2017). More specifically, we set $\kappa = 0.1$, $c_0 = 1.5$ and $c_1 = 1$ to center the prior on π_l above unity while $a_0 = b_0 = 0.01$. The choice for κ implies that we place a large amount of prior mass on zero while at the same time allowing for relatively thick tails. Our choice for the Gamma prior on ϱ introduces heavy shrinkage on the covariance parameters as well as the corresponding process standard deviations.

For all models (i.e., the competitors introduced in the next subsection) we consider, as well as the proposed model, we include a single lag of the endogenous variables. Higher lag orders are generally possible but, given

[4] In the empirical application we use 30,000 overall iterations, with the first 15,000 being discarded as burn-in.

the high dimension of the state space and the increased computational complexity, we stick to one lag. In addition, experimenting with slightly higher lag orders leads to models that are relatively unstable during several points in time in our estimation sample.

The design of our forecasting exercise is the following. We start with an initial estimation period that spans the period between the end of November 2016 (22 November) to the end of April 2017 (26 April). The remaining 160 days are used as a hold-out period. After obtaining the one-step-ahead predictive density for 27 April 2017, we consequently expand the estimation sample by a single day until the end of the sample is reached. This yields a sequence of 160 one-day-ahead predictive densities.

To assess the predictive fit of our model we use the log-predictive likelihood (LPL), motivated in Geweke and Amisano (2010), for example, and the root mean square forecast error (RMSE). Using LPLs enables us to assess not only how well the model fits in terms of point predictions but also how well higher moments of the predictive density are captured. In addition, to assess model calibration we use univariate probability integral transforms (Amisano & Geweke, 2017; Clark, 2011; Diebold, Gunther, & Tay, 1998).

4.2 | Competing models

Our set of competing models ranges from univariate benchmark models that feature SV to a wide set of multivariate benchmark models. The first set of models considered are a random walk (RW-SV) and the AR(1) model (henceforth labeled AR-SV), both estimated with SV. We use noninformative priors on the AR(1) regression coefficient and the same prior setup for the log-volatility equation, as discussed in the previous section. These two models serve to illustrate whether a multivariate modeling approach pays off and, in addition, whether allowing for structural changes in the underlying regression parameters improves predictive capabilities.

In addition, we consider a set of nested multivariate benchmark models. To quantify the accuracy gains of time-varying parameter specifications, we estimate three constant parameter VARs with SV. The first VAR uses the prior setup described above but with $\sqrt{\vartheta}_{ij} = 0$ for all i, j. The second model is a nonconjugate Minnesota VAR with asymmetric shrinkage across equations. To select the hyperparameters we follow Giannone, Lenza, and Primiceri (2015) and place hyperpriors on all hyperparameters and estimate them using a random walk MH step. The last VAR we consider is a model that features a stochastic search variable selection (SSVS) prior specified as in George, Sun, and Ni (2008). This implies that a two-component Gaussian prior is used, with the Gaussians differing in terms of their prior variance. One component features a large prior variance (labeled the slab distribution), which introduces relatively little prior information, whereas the second component has a prior variance close to zero (the spike component) that strongly forces the posterior of the respective coefficient to zero. We set the hyperparameters (i.e., the prior standard deviations) for the slab distribution by using the OLS standard deviation times a constant (10 in our case), while the prior standard deviation on the spike component is set equal to 0.1 times the OLS standard deviation.

Moreover, we include two time-varying parameter models with SV and Gaussian measurement errors. The first TVP-VAR considered (labeled TVP) is based on an uninformative prior (obtained by setting the prior variances to unity for both the initial states as well as the process standard deviations). The next benchmark model (called TVP NG) is our proposed specification with an NG prior but with Gaussian errors (i.e., $\phi_{it} = 1$ for all i,t). This choice serves to assess whether additional flexibility on the measurement errors is needed.

Finally, the last model considered is the most flexible specification in terms of the law of motion of the latent states. This model, labeled the threshold TVP-VAR (labeled TTVP) is based on Huber et al. (2017) and captures the notion that parameter movements are only allowed if they are sufficiently large. To achieve this, a threshold specification for the process variances is adopted. This specification depends on a latent indicator that, in turn, is driven by the absolute size of parameter changes. Thus, if the change in a given regression parameter is large (i.e., exceeds a certain threshold we estimate), we use a large variance in Equation 7. By contrast, if the change is small the process variance is set to a small constant that is close to zero. The prior specification adopted here closely follows the benchmark specification outlined in Huber et al. (2017) and we refer to the original paper for additional details.

4.3 | Out-of-sample forecasting performance

We start by considering the forecasting performance in terms of log predictive likelihoods (LPS). Table 1 displays the LPS as well as the RMSEs for the competing models. The first column shows the joint LPS for the three crypto-currencies considered, while the next three columns display the marginal LPS for a given crypto-currency. The final three columns show the RMSEs.

Considering the joint LPS indicates that, across models, the t-TVP NG specification outperforms the remaining models. This points towards the necessity to

TABLE 1 Joint and marginal log predictive likelihoods for all models considered (left panel) and root mean square forecast errors (right panel). For the joint log predictive likelihood we integrate out the effect of the other variables included in $\boldsymbol{y}_t$ and focus exclusively on the predictive performance for the three crypto-currencies

		Log predictive score			Root mean square error		
	JointLPS	Bitcoin	Litecoin	Ethereum	Bitcoin	Litecoin	Ethereum
TTVP	621.023	286.360	134.231	153.201	0.050	0.084	0.078
TVP	451.631	187.474	106.946	97.300	0.074	0.133	0.134
TVP NG	632.410	286.134	144.629	159.562	0.050	0.083	0.079
t-TVP NG	643.873	277.679	161.768	166.988	0.050	0.084	0.078
Minn-VAR	577.779	283.399	123.580	153.274	0.051	0.085	0.078
NG-VAR	592.391	286.483	130.194	148.553	0.051	0.084	0.078
SSVS	586.083	286.255	122.346	153.081	0.051	0.084	0.078
RW-SV	483.952	240.751	131.410	112.487	0.073	0.112	0.114
AR-SV	598.936	280.487	158.899	159.725	0.051	0.085	0.078

allow for both a flexible error distribution as well as time-varying parameters with appropriate shrinkage priors. Especially when compared to the constant-parameter VAR models, all three TVP-VAR specifications with some form of shrinkage yield pronounced accuracy gains. Note also that the AR(1) model with SV proves to be a tough competitor relative to the set of Bayesian VARs.

The necessity of introducing shrinkage in the TVP-VAR framework can be seen by comparing the joint forecasting performance of the TVP model with the remaining TVP-VARs considered. Note that in our medium-scale model a TVP-VAR with relatively little shrinkage leads to overfitting issues, which in turn are detrimental to forecasting performance.

Zooming into the results for the three crypto-currencies, we generally observe that models performing well in terms of the joint LPS also do well on average. One interesting exception is our proposed t-TVP NG specification. While the performance gains for Litecoin and Ethereum appear to be substantial vis-à-vis the competing models, we find that Bitcoin predictions appear to be inferior relative to the TTVP and the TVP NG specifications. If the researcher is interested in predicting the price of Bitcoin, the two best-performing models are the TTVP specification and the Bayesian VAR with a normal-gamma shrinkage prior. Interestingly, note that the comparatively weaker joint performance of the BVAR models stems from weaker Litecoin and Ethereum predictions, whereas Bitcoin predictions appear to be rather precise.

Considering point forecasting, performance generally corroborates the findings for density forecasts. Here we again observe that models which yield precise predictive densities also work well when only point predictions are considered. Note, however, that the differences in terms of RMSE between multivariate models and the univariate AR(1) model are negligible. This somewhat highlights that forecasting gains in terms of predictive likelihoods stem from higher moments of the predictive density like the predictive variance (in terms of the marginal log scores) or a more appropriate modeling strategy for the predictive variance–covariance structure.

Next, we investigate whether differences in forecasting performance appear to be time varying. Figure 3 shows the log predictive Bayes factors relative to the random walk with SV. Comparing the model performances over time points towards a pronounced degree of heterogeneity over time. Results for Bitcoin (see panel (a)) show that the two best-performing models are the TTVP and the TVP NG specifications. While the former yields a slightly better performance over time, the latter proves to be the best-performing model during the first part of the hold-out period. For the remaining models we find only relatively little time variation in their predictive performance. Considering the results for Litecoin (see panel (b)) we find pronounced movements in relative forecasting accuracy. More specifically, we find that while forecasting performance appears to be homogeneous during the first months of the hold-out period, from May 2017 onward the t-TVP NG specification starts to perform extraordinarily well, improving upon all competitors by large margins.

Finally, panels (c) and (d) show the performance for Ethereum as well as the overall performance over time. Here we generally find results that are comparable with the findings described above. Note that the overall log predictive likelihood displays a pattern similar to that of the marginal LPS for the remaining crypto-currencies.

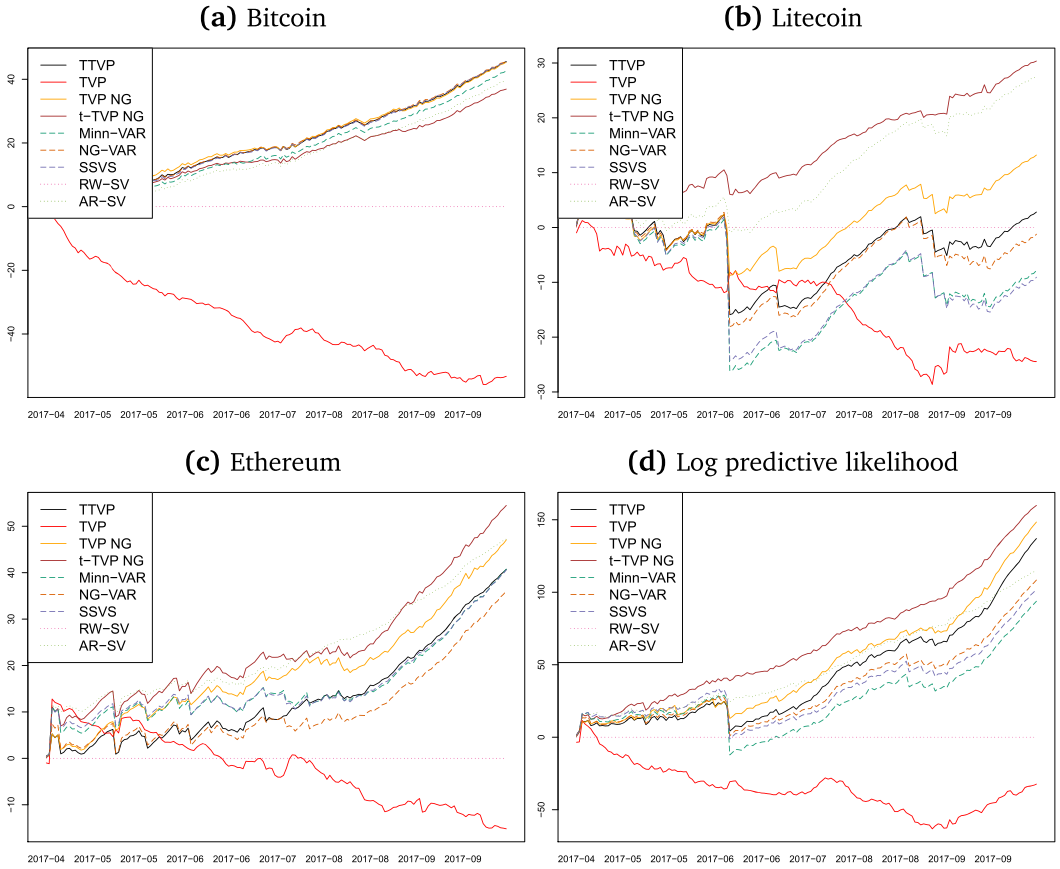

FIGURE 3 Log predictive Bayes factors relative to the TVP-VAR over time: (a) Bitcoin; (b) Litecoin; (c) Ethereum; (d) log predictive likelihood [Colour figure can be viewed at wileyonlinelibrary.com]

However, compared to panel (a) we observe that the t-TVP specification also excels in terms of joint density predictions. The main difference is that the superior performance of the t-TVP NG model in terms of predicting Litecoin prices lifts the log predictive Bayes factor above those obtained for all competing models.

4.4 | Model evaluation using probability integral transforms

Following Diebold et al. (1998), Clark (2011), and Amisano and Geweke (2017), if a given model $\mathcal{M}_i$ is correctly specified one can show that

$$z_{jt,i} = \Phi^{-1}(F_y(y_{jt}|\boldsymbol{y}_{1:t-1}, \mathcal{M}_i)) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \qquad (23)$$

for $t = t_0,...,T$ and $j = 1,...,m$ and t_0 indicating the first observation of the hold-out period (i.e., 22 November). Hereby we let Φ^{-1} denote the inverse distribution function of the standard normal distribution and $F_y(y_{jt}|\boldsymbol{y}_{1:t-1}, \mathcal{M}_i)$ denotes the cumulative distribution function associated with the underlying predictive distribution of model i. If the model is correctly specified the sequence of normalized forecast errors $\{z_{jt}\}_{t=t_0}^{T}$ is independent and identically standard normally distributed.

Figure 4(a–c) shows the normalized forecast errors across models and for all three crypto-currencies considered, while Table 2 provides statistical tests that aim to support our visual assessment of Figure 4. In the case of Bitcoin and Litecoin, we find that the mean appears to be close to zero. This finding is corroborated by the first column in Table 2, which displays the empirical mean obtained by regressing $z_{jt,i}$ on a constant, with p-values in parentheses. Note that for Ethereum we find the normalized forecast errors of the majority of models to be centered above zero. The two exceptions are the TVP NG specification and the Minnesota prior VAR. Considering again panel (c) reveals that these deviations from zero are mainly driven by the failure to capture the conditional mean during the beginning of the hold-out period.

Considering the variances reveals that in the case of Bitcoin the variances of the normalized errors are all well below unity, indicating that the estimated predictive

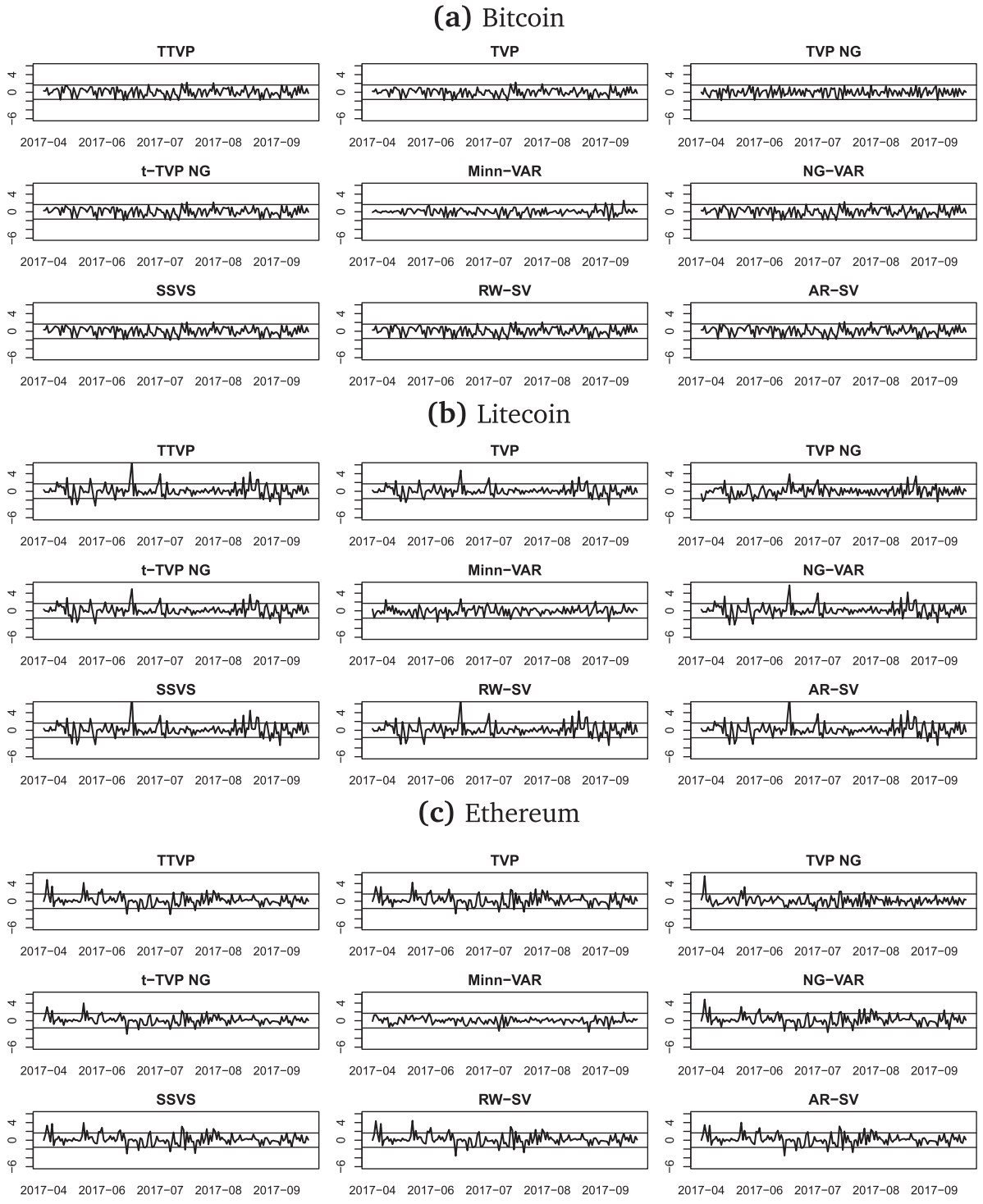

FIGURE 4 Normalized forecast errors across models and crypto-currencies: (a) Bitcoin; (b) Litecoin; (c) Ethereum. Errors are obtained by applying the inverse cumulative distribution function of a normal distribution to the PIT of the one-step-ahead forecast errors

variance is generally too high. Put differently, this is an indication for a situation where too many actual observations fall in the center of the predictive distribution. This finding appears to be strongly supported by the second column of Table 2, which displays the estimated variance of the normalized forecast error obtained by regressing the squared error on a constant. For the t-TVP NG and TTVP specifications we find slightly higher variances. Our interpretation is that allowing for a flexible error specification, either by directly using non-Gaussian shocks in conjunction with stochastic volatility or by introducing more flexibility on the law of motion of

TABLE 2 Statistical results for the transformed forecast errors

	Mean (*p*-value)	Variance (*p*-value)	Persistence (*p*-value)
Bitcoin			
TTVP	0.060 (0.401)	0.821 (0.024)	−0.078 (0.329)
TVP	0.013 (0.838)	0.649 (0.000)	−0.085 (0.283)
TVP NG	0.004 (0.948)	0.683 (0.000)	−0.439 (0.000)
t-TVP NG	0.051 (0.466)	0.783 (0.005)	−0.060 (0.454)
Minn-VAR	0.007 (0.902)	0.490 (0.000)	−0.135 (0.089)
NG-VAR	0.022 (0.756)	0.809 (0.018)	−0.093 (0.243)
SSVS	0.058 (0.410)	0.799 (0.007)	−0.052 (0.513)
RW-SV	0.082 (0.246)	0.796 (0.007)	−0.058 (0.470)
AR-SV	0.098 (0.168)	0.804 (0.011)	−0.051 (0.518)
Litecoin			
TTVP	0.121 (0.255)	1.790 (0.030)	0.023 (0.772)
TVP	0.096 (0.254)	1.120 (0.544)	0.011 (0.891)
TVP NG	−0.009 (0.912)	1.154 (0.347)	−0.052 (0.516)
t-TVP NG	0.115 (0.202)	1.295 (0.187)	0.027 (0.731)
Minn-VAR	−0.049 (0.472)	0.732 (0.007)	−0.084 (0.292)
NG-VAR	0.114 (0.254)	1.596 (0.047)	0.008 (0.917)
SSVS	0.128 (0.253)	2.001 (0.031)	0.018 (0.821)
RW-SV	0.144 (0.188)	1.920 (0.018)	0.017 (0.831)
AR-SV	0.152 (0.177)	2.020 (0.021)	0.025 (0.756)
Ethereum			
TTVP	0.201 (0.025)	1.285 (0.212)	0.121 (0.127)
TVP	0.208 (0.026)	1.393 (0.047)	0.059 (0.461)
TVP NG	0.090 (0.250)	0.980 (0.93)	−0.026 (0.743)
t-TVP NG	0.148 (0.042)	0.848 (0.306)	0.072 (0.367)
Minn-VAR	0.023 (0.672)	0.478 (0.000)	0.047 (0.556)
NG-VAR	0.223 (0.014)	1.335 (0.107)	0.100 (0.207)
SSVS	0.188 (0.043)	1.393 (0.056)	0.075 (0.343)
RW-SV	0.194 (0.040)	1.429 (0.071)	0.065 (0.417)
AR-SV	0.176 (0.058)	1.380 (0.065)	0.052 (0.514)

the latent states, slightly helps to push the variances towards one.

For Litecoin and Ethereum, the variances appear to be closer to one for all TVP specifications except for the TTVP model (in the case of Litecoin). It is noteworthy that, especially for Litecoin, constant-parameter models with SV tend either to underestimate the predictive variance or fail to capture observations in the tail of the empirical distribution.

Finally, considering the persistence of $z_{jt,i}$ reveals that most models tend to produce normalized errors which display muted persistence levels. This is corroborated by the final column of Table 2, which shows the persistence parameter obtained by estimating AR(1) models in $z_{jt,i}$ along with its *p*-values.

5 | ECONOMIC PERFORMANCE CRITERIA: A SIMPLE TRADING EXERCISE

To assess which model excels in terms of economic performance criteria, we perform a trading exercise where each model is used to generate a set of optimal weights attached to each of the three crypto-currencies considered. Using the models discussed in the previous sections

as well as two additional investment strategies that are based on equal weights and a simple passive investment in Bitcoin allows us to infer whether constructing a trading strategy based on more sophisticated econometric models pays off in terms of generating superior returns.

We assume that investors adopt two strategies to find an optimal sequence of weights $\boldsymbol{w}_{it} = (w_{1i,t}, w_{2i,t}, w_{3i,t})'$. The first one is the standard minimum variance portfolio problem, which aims to allocate money between the three assets considered such that the portfolio variance is minimized. This implies that, for $t = t_0, \ldots, T$, the investor solves

$$\begin{aligned} \underset{\boldsymbol{w}_{it}}{\text{minimize}} \quad & \boldsymbol{w}_{it} \boldsymbol{P}_{i,t|t-1} \boldsymbol{w}_{it}' \\ \text{subject to} \quad & \mathbf{1}' \boldsymbol{w}_{it} = 1, \end{aligned} \tag{24}$$

where $\mathbf{1}$ is a three-dimensional vector of ones and $\boldsymbol{P}_{i,t|t-1}$ denotes the variance of model i's one-step-ahead predictive density.

The second strategy adds a specific portfolio target return to the optimization problem in Equation 24, that is:

$$\boldsymbol{w}_{it}' \boldsymbol{\mu}_{it|t-1} \geq r_t^*. \tag{25}$$

Here we let $\boldsymbol{\mu}_{it|t-1}$ denote the one-step-ahead predictive mean of model i and r_t^* is a potentially time-varying target return the investor wants to match. This strategy, called the target mean–variance portfolio, tries to minimize the overall portfolio variance while at the same time maintaining the desired return r_t^* (see Markowitz, 1952).

Table 3 shows annualized Sharpe ratios for the minimum variance portfolio strategy as well as for the target mean–variance portfolio strategy for $r_t^* = r^* \in \{\frac{0.10}{252}, \frac{0.15}{252}, \frac{0.30}{252}\}$. Considering the performance of the minimum variance portfolio (see first column in Table 3) shows that performance differences across models appear to be relatively small. This indicates that weights generated by the set of econometric models are similar and, when compared to the other strategies, more stable over time. Inspection of the weights (not shown) also suggests that this strategy yields weights that are seldom above one in absolute values (i.e., leveraged long/short positions). The single best-performing model is the no-shrinkage TVP specification, closely followed by the TVP NG model. Note that using simple equal weights also yields favorable risk/return ratios.

Considering the target mean–variance strategy for different target returns yields more heterogeneous model performances. The two best-performing models are the TTVP model and the constant-parameter VAR coupled with the SSVS prior. For the TVP VAR and the TVP NG model, we find that performance decreases when compared to the minimum variance portfolio strategy, while for the proposed t-TVP NG we observe increasing Sharpe ratios. Comparing different r^* yields no discernible differences, with most models that do well for modest target returns also performing well if target returns become more ambitious.

TABLE 3 Annualized Sharpe ratios of various competing investment strategies over the hold-out sample

		Target mean–variance		
	Min-variance	$r^* = \frac{0.10}{252}$	$\frac{0.15}{252}$	$\frac{0.30}{252}$
TTVP	2.379	2.900	2.923	2.978
TVP	2.579	2.015	2.019	2.031
TVP NG	2.510	2.069	2.053	1.995
t-TVP NG	2.365	2.452	2.465	2.498
Minn-VAR	2.066	−0.313	−0.243	0.004
NG-VAR	2.023	2.845	2.725	2.312
SSVS	1.997	2.942	2.948	2.943
RW-SV	2.040	1.399	1.415	1.464
AR-SV	2.201	2.390	2.407	2.453
Equal weights	2.528	2.528	2.528	2.528
Only BTC	2.419	2.419	2.419	2.419

Note. "Min-variance" refers to the minimum variance portfolio whereas "Target mean–variance" is the target mean–variance portfolio for different target returns. "Equal weights" refers to using $w_{jt} = 1/3$ for all j,t and only BTC sets the weight associated with Bitcoin equal to one.

Across strategies it is worth noting that performing a passive investment in Bitcoin only (i.e., setting the corresponding weight equal to one for all t) also works well but one could still improve upon that strategy by considering more flexible portfolio allocation strategies.

6 | CONCLUDING REMARKS

In this paper we perform a systematic comparison of univariate and multivariate time series models in terms of predicting one-day-ahead returns for three crypto-currencies, namely Bitcoin, Litecoin, and Ethereum. To match the pronounced degree of volatility observed in daily returns of crypto-currencies, we propose a medium-scale multivariate state-space model that features heavy-tailed measurement errors and stochastic volatility, a feature that turns out to be advantageous for density predictions. More generally, we find that it pays off to allow for time-varying parameters and a flexible error distribution only if suitable shrinkage priors are introduced. These priors introduce shrinkage to select the subset of time-varying coefficients in a flexible manner. To gauge the economic significance of our findings we also perform a trading exercise. The results show that models which perform well in forecasting also tend to work well when used to guide investment decisions.

ACKNOWLEDGMENTS

The authors thank Gregor Kastner, O. Skar, Belinda Haid, Jouchi Nakajima, and the participants of the 2017 NBP Workshop on Forecasting for helpful comments and suggestions. Financial support from the Czech Science Foundation, Grant 17-14263S, is gratefully acknowledged.

ORCID

Christian Hotz-Behofsits http://orcid.org/0000-0003-1921-1978
Florian Huber http://orcid.org/0000-0002-2896-7921
Thomas Otto Zörner http://orcid.org/0000-0001-7830-7801

REFERENCES

Amisano, G., & Geweke, J. (2017). Prediction using several macroeconomic models. *Review of Economics and Statistics*, *99*(5), 912–925.

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

Bitto, A., & Frühwirth-Schnatter, S. (2016). *Achieving shrinkage in a time-varying parameter model framework*. arXiv preprint arXiv:1611.01310.

Böhme, R., Christin, N., Edelman, B., & Moore, T. (2015). Bitcoin: Economics, technology, and governance. *Journal of Economic Perspectives*, *29*(2), 213–238.

Carlin, B. P., Polson, N. G., & Stoffer, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*, *87*(418), 493–500.

Carriero, A., Clark, T. E., & Marcellino, M. (2015). Large vector autoregressions with asymmetric priors (Working paper), School of Economics and Finance, Queen Mary University of London, London, UK.

Carter, C. K., & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, *81*(3), 541–553.

Cheah, E. T., & Fry, J. (2015). Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics Letters*, *130*, 32–36.

Chiu, C.-W., Mumtaz, H., & Pintér, G. (2017). Forecasting with VAR models: Fat tails and stochastic volatility. *International Journal of Forecasting*, *33*(4), 1124–1143.

Chu, J., Chan, S., Nadarajah, S., & Osterrieder, J. (2017). GARCH modelling of cryptocurrencies. *Journal of Risk and Financial Management*, *10*(4), 17.

Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business and Economic Statistics*, *29*(3), 327–341.

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

Cogley, T., & Sargent, T. J. (2005). Drift and volatilities: Monetary policies and outcomes in the post WWII U.S. *Review of Economic Dynamics*, *8*(2), 262–302.

Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, *39*(4), 863–883.

Feldkircher, M., Huber, F., & Kastner, G. (2017). *Sophisticated and small versus simple and sizeable: When does it pay off to introduce drifting coefficients in Bayesian VARs?* arXiv preprint arXiv:1711.00564.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models.*Journal of Time Series Analysis*, *15*(2), 183–202.

Frühwirth-Schnatter, S., & Wagner, H. (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics*, *154*(1), 85–100.

Gallant, A. R., Hsieh, D., & Tauchen, G. (1997). Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics*, *81*(1), 159–192.

George, E. I., Sun, D., & Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, *142*(1), 553–580.

Geweke, J. (1994). Bayesian analysis of stochastic volatility models: Comment. *Journal of Business and Economic Statistics*, *12*(4), 397–399.

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

Geweke, J., & Tanizaki, H. (2001). Bayesian estimation of state-space models using the Metropolis–Hastings algorithm within Gibbs sampling. *Computational Statistics and Data Analysis*, *37*(2), 151–170.

Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, *97*(2), 436–451.

Gordon, K., & Smith, A. (1990). Modeling and monitoring biomedical time series. *Journal of the American Statistical Association*, *85*(410), 328–337.

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

Harrison, P. J., & Stevens, C. F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society, Series B*, *38*(3), 205–247.

Huber, F., & Feldkircher, M. (2017). Adaptive shrinkage in Bayesian vector autoregressive models. *Journal of Business and Economic Statistics*, *34*(3), 1–13.

Huber, F., Kastner, G., & Feldkircher, M. (2017). *A new approach toward detecting structural breaks in vector autoregressive models*. arXiv preprint arXiv:607.04532v3.

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

Kastner, G. (2015a). Dealing with stochastic volatility in time series using the R package stochvol. *Journal of Statistical Software*. Retrieved from, http://cran.r-project.org/web/packages/stochvol/vignettes/article.pdf

Kastner, G. (2015b). Dealing with stochastic volatility in time series using the R package stochvol. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v069.i05

Kastner, G. (2015c). Heavy-tailed innovations in the R package stochvol. ePubWU Institutional Repository.

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

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, *7*(1), 77–91.

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

Urquhart, A. (2017). Price clustering in Bitcoin. *Economics Letters*, *159*, 145–148.

West, M. (1987). On scale mixtures of normal distributions. *Biometrika*, *74*(3), 646–648.

Christian Hotz-Behofsits is Teaching and Research Associate at the Vienna University of Economics and Business. In addition to monetization strategies for new media business models, his research interests include predictive models and applied machine learning.

Florian Huber is Assistant Professor at the Vienna University of Economics and Business. His research interests focus primarily on macroeconometric modeling and forecasting, time series analysis, and Bayesian econometrics. He has published in journals such as the Journal of Business and Economic Statistics, European Economic Review, Journal of Applied Econometrics, Journal of Economic Dynamics and Control, and the Oxford Bulletin of Economics and Statistics inter alia.

Thomas Zörner is working as a Research Associate at the University of Economics and Business on projects, funded by the Austrian National Bank (OeNB) about heterogeneous expectations in macroeconomic models and forecasting exchange rates. He holds a MSc in Economics, a BA in Anthropology and is currently enrolled at the PhD program in Economics at the Vienna University of Economics and Business. His main research focuses on credit and business cycles, macroeconometrics and non-linear models.

How to cite this article: Hotz-Behofsits C, Huber F, Zörner TO. Predicting crypto-currencies using sparse non-Gaussian state space models. *Journal of Forecasting*. 2018;1–14. https://doi.org/10.1002/for.2524