A Randomized Missing Data Approach to Robust Filtering with Applications to Economics and Finance *

Dobrislav Dobrev[1], Derek Hansen[2], and Paweł J. Szerszeń[3]

[1,3]Federal Reserve Board of Governors

[2]University of Michigan

June 30, 2019

PRELIMINARY AND INCOMPLETE

Abstract

We put forward a simple new approach to robust filtering of state-space models, motivated by the idea that the inclusion of only a small fraction of available highly precise measurements can still extract most of the attainable efficiency gains for filtering latent states, estimating model parameters, and producing out-of-sample forecasts. The new class of particle filters we develop aims to achieve a degree of robustness to outliers and model misspecification by purposely randomizing the subset of utilized highly precise but possibly misspecified or outlier contaminated data measurements, while treating the rest as if missing. The arising robustness-efficiency trade-off is controlled by varying the fraction of randomly utilized measurements or the incurred relative efficiency loss from such randomized utilization of the available measurements. As an empirical illustration, we consider popular state space models for inflation and equity returns with stochastic volatility and document favorable performance of our robust particle filter and density forecasts on both simulated and real data. More generally, our randomization approach makes it easy to robustly incorporate highly informative but possibly contaminated modern "big data" streams for improved state-space filtering and forecasting.

Keywords: state space models, particle filtering, robustness, outliers, misspecification, stochastic volatility, realized volatility, return density forecasting, inflation forecasting

*Emails: dobrislav.p.dobrev@frb.gov, dereklh@umich.edu and pawel.j.szerszen@frb.gov. We are grateful to Ian Dew Becker, Edward Herbst, Nicholas Polson, Tatevik Sekhposyan, Neil Shephard and Jonathan Wright for very helpful discussions and comments. We also thank conference participants at the 2018 2nd Workshop on Financial Econometrics and Empirical Modeling of Financial Markets, Kiel Institute for the World Economy, the 2018 NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics (SBIES), Stanford University, 2018 24th International Conference Computing in Economics and Finance, Universitá Cattolica, 2018 EEA-ESEM Congress, University of Cologne, 2018 NBER-NSF Time Series Conference, UCSD, 2018 12th International Conference on Computational and Financial Econometrics, University of Pisa, and the 2018 Paris Financial Management Conference, IPAG Business School. We also thank seminar participants at the Federal Reserve Board of Governors, Northwestern University, the U.S. Census Bureau, and the University of Michigan. This article represents the views of the authors, and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or other members of its staff.

1. Introduction

State-space models play an important role in fields as diverse as engineering, medicine, economics, and finance. It is well known that the successful use of state-space models in empirical work, particularly in forecasting applications, can be quite sensitive to even a small degree of model misspecification, which can be induced, for example, by the presence of outliers in collected measurement data. As such, the growing data availability and processing capabilities has motivated the development of novel methods for filtering of state-space models with specific focus on robustness to potentially complex data imperfections. One approach imposes heavy-tailed distributional assumptions as in Durbin and Koopman (2000) or Harvey and Luati (2014). Another approach adds thresholding or an outlier detection step as in Calvet, Czellar, and Ronchetti (2015), Crevits and Croux (2017), Maìz, Molanes-Lòpez, Mìguez, and Djuriè (2012), among others.

In this paper we propose a simple new approach to robust filtering of state-space models. It builds on intuitive insights from the well-established literature on handling missing data measurements in state space models. In particular, we note that in many cases only a small fraction of available highly precise measurements can extract much of the attainable efficiency gains for filtering latent states, estimating model parameters, and producing out-of-sample forecasts. Thus, there is often only marginal incremental value from including all highly precise measurements. This opens the possibility to exclude subsets of the available measurements to achieve a degree of robustness without much efficiency loss.

Motivated by this important property of state-space models, we develop a new class of particle filters that achieve a degree of robustness to outliers and model misspecification by purposely randomizing the subset of included highly precise but possibly somewhat misspecified or outlier contaminated data measurements while treating the rest as if missing. The arising robustness-efficiency trade-off in our robust filtering framework can thus be controlled either by varying the fraction of randomly included measurements, or by varying the target loss of efficiency with respect to the infeasible case of fully utilizing all available measurements in the absence of misspecification.

First, as a motivating example, we revisit the case of exogenously missing measurements

to illustrate how even a fairly sparse subset of highly precise measurements may often suffice to attain considerable efficiency gains. Next, we lay down a fully Bayesian framework for probabilistic inclusion of collected data measurements for the sake of improving robustness without incurring much of an efficiency loss. In its simplest form, our approach to robust particle filtering assumes a fixed unconditional probability that the measurement would be treated as missing. In its more generalized forms, our approach can infer *a posteriori* the probability that an outlier is present through specifying the distribution of distorted measurements.

On the theory side, we build upon a widely used and applicable state space model (SSM) specification. We assume that there are two types of observations available. One that is not very precise but always reliable, and another one that is highly precise but is more prone to substantial measurement distortions inducing model misspecification. The high precision of these observations can be accommodated by using adapted sampling methods such as the Auxiliary Particle Filter from Pitt and Shephard (1999). However, in this setting incorporating all observations distorts the signal about latent states. As we do not directly observe which observations are unreliable, we introduce a new latent variable in the system that governs the informativeness of data measurements by probabilistically designating each observation as either informative or as missing. We then demonstrate how this randomization approach can robustify the filtering, and hence the model-implied out-of-sample forecasts, to the possible presence of data imperfections.

As such, the missing data randomization approach we propose is akin to but different from classical resampling methods for robust forecasting such as bootstrap aggregating (bagging) originally developed by Breiman (1996). First, most of the existing literature on bagging has dealt with cross-sectional settings rather than time series models. Second, a handful of studies that have considered the benefits of bootstrapping and bagging in a time-series context have done so either by resampling the full-sample residuals to assess estimation uncertainty such as Stoffer and Wall (1991) or by resampling blocks of dependent observations to robustify forecasts as in Inoue and Kilian (????) and Bergmeir, Hyndman, and Benítez (2016). By contrast, our approach probabilistically induces randomly missing individual data measurements as a way to generate a mixture distribution that guards against possible misspecification. This makes it possible to alleviate ex-ante any major tensions between the model and the full data sample to improve out-of-sample

forecasting performance.

On the empirical side, we consider popular state space models for inflation and equity returns with stochastic volatility and document favorable performance of our robust particle filter and density forecasts both on simulated and real data. In our first empirical illustration we apply our robustification framework to the classical unobserved components (UC) model for estimating inflation trends. It has been well documented by Stock and Watson (2007) and Stock and Watson (2016), among others, that the forecasting performance of the UC model is hindered by the presence of time-variation in the precision of inflation rate measurements and additional measurement distortions due to outliers. We therefore study a robustly filtered version of the classical UC model based on our missing data randomization approach vis-à-vis the unobserved components/stochastic volatility outlier-adjustment (UCSVO) model that was put forward by Stock and Watson (2016) as another way to minimize the detrimental impact of outliers by way of subjecting them to particular distributional assumptions. We find that our robustly filtered UC model can offer some gains in terms of out-of-sample forecasting performance, perhaps thanks to avoiding any stringent parametric assumptions on the outliers as in the UCSVO model. The favorable forecasting results we obtain for our robustly filtered UC model thus further underscore the findings in Stock and Watson (2007) and Stock and Watson (2016) regarding the need for time-series inflation forecasts to address the presence of potentially complex inflation measurement imperfections.

Our second empirical illustration considers filtering stochastic volatility (SV) models in the presence of realized volatility (RV) measures. The filtering of SV models in state space form has traditionally been based solely on daily return measurements. However, it has long been noted for example by Polson and Roberts (1994) and Barndorff-Nielsen and Shephard (2002) that more precise RV measurements of the latent volatility states can greatly improve filtering efficiency. In practice, this requires accounting also for potential finite sample biases and outliers in the RV measures due to imperfectly suppressed microstructure noise and jumps encountered in high-frequency intraday returns entering the RV measures. One strain of such filtering approaches considered by Takahashi, Omori, and Watanabe (2009), Koopman and Scharth (2013), Creel and Kristensen (2015) among others is to model the variance of the RV measurements as homoscedastic with variance estimated from the data rather than explicitly given by the so called realized quarticity (RQ) measures as

implied by the feasible asymptotic theory for RV measures. The obtained estimates of the variance of the RV measurements by such filtering approaches tend to be much bigger than the average RQ-implied variances in accordance with the underlying asymptotic theory, thereby pointing to the presence of outliers or a degree of model misspecification. Another approach considered by Dobrev and Szerszen (2010) is to directly specify the variance of the RV measurements to be driven by RQ measures as implied by the underlying asymptotic theory for sufficiently large intraday return samples. However, the presence of outliers or model misspecification then tends to lead to substantially different parameter estimates and poor out-of-sample forecast performance with statistically significant differences in comparison to the baseline model without RV measurements. As such, none of the existing approaches for filtering SV models in the presence of RV measures appear to directly address the inherent need for robust filtering. By contrast, our new filtering approach based on intentional probabilistically reduced use of the RV measurements allows us to impose RQ-driven variances of the RV measurements in line with the underlying asymptotic theory, while at the same time attaining meaningful improvement in out-of-sample multi-horizon return density forecasts in comparison to the baseline model without RV measurements. In this sense, our robust filtering approach is able to sufficiently alleviate distortions due to data imperfections or model misspecification when estimating SV models in the presence of RV measures. More generally, also in many other data-rich settings our randomization device can make it easy to robustly incorporate highly informative but possibly contaminated modern "big data" streams for improved state-space filtering.

The rest of the paper is organized as follows. Section 2 illustrates the principles motivating our robust filtering approach through a simple motivating example. Section 3 lays down the setup and theoretical framework for our robust particle filter. Section 4 provides empirical illustrations on simulated and real data. Section 5 draws a summary and conclusions.

2. Motivating Example: Missing Data and Filtering Efficiency

We motivate our randomized missing data approach to robust filtering by noting that in many standard linear state space models the efficiency of the Kalman filter need not drop all that fast

as the proportion of missing data measurements goes up. To give an illustrative example of this important property, we consider the canonical case of a single autoregressive latent state x_t filtered from two different available measurements $y_t^{(1)}$ and $y_t^{(2)}$ as specified by the following simplified version of a linear state space model:

$$x_{t+1} = \kappa x_t + \sigma_\eta \eta_t \tag{1}$$

$$y_t^{(1)} = x_t + \sigma_\epsilon^{(1)} \epsilon_t^{(1)} \tag{2}$$

$$y_t^{(2)} = x_t + \sigma_\epsilon^{(2)} \epsilon_t^{(2)} \tag{3}$$

where the error terms η_t, $\epsilon_t^{(1)}$ and $\epsilon_t^{(2)}$ are i.i.d. $N(0,1)$. We further restrict attention to the case when the latent state x_t is highly persistent and the first measurement $y_t^{(1)}$ is always available (e.g. cheap to obtain) but is characterized by a fairly weak signal-to-noise ratio $\sigma_\eta/\sigma_\epsilon^{(1)}$, while the second measurement $y_t^{(2)}$ can be missing at random most of the time (e.g. costly to obtain) but is otherwise characterized by a relatively strong signal-to-noise ratio $\sigma_\eta/\sigma_\epsilon^{(2)}$.

For numerical illustration of the impact of missing measurements $y_t^{(2)}$ on the efficiency of the Kalman filter in this toy-model example, we further set $\kappa = 0.99$, $\sigma_\eta = 0.1$, $\sigma_\epsilon^{(1)} = 2$, and $\sigma_\epsilon^{(2)} = 0.2$. The persistence level and signal-to-noise ratios are chosen to roughly mimic the ones encountered when filtering stochastic volatility (aka x_t) from daily absolute returns (aka $y_t^{(1)}$) and realized volatility measures (aka $y_t^{(2)}$), while abstracting from non-linearities that we tackle as part of the general framework for robust particle filtering that we lay down in section 3 below.

For illustrative purposes, we resort to the standard Kalman filter directly applicable in this setting and further restrict attention to the case when the first type of measurements $y_t^{(1)}$ are not contaminated by outliers, while the second type of measurements $y_t^{(2)}$ can be prone to outliers or misspecified in other ways. This again mimics the case of filtering stochastic volatility models based on daily return measurements deemed to always be reliable in conjunction with realized volatility measurements that can be distorted by various market microstructure effects and so called "mini flash crashes" (large but short-lived intraday price swings) that need not necessarily be consistent with the volatility dynamics at daily and longer horizons.

We first note that in the absence of outliers in $y_t^{(2)}$ as illustrated on Figure 2 the filtered state variance does not drop by much even when only 10% of $y_t^{(2)}$ are observed (with the rest 90% missing at random) compared to the filtered state variance when all $y_t^{(2)}$ are observed. For this to happen, it suffices that the filtered state variance in between consecutive sparse observations of the high-precision measurements $y_t^{(2)}$ does not deteriorate too fast towards the filtered state variance based only on the much less precise measurements $y_t^{(1)}$, the way this holds for the filtered state variances depicted on Figure 2. In this sense, the availability of even a small fraction of the more precise measurements $y_t^{(2)}$ suffices to attain substantial efficiency gains in comparison to the case of relying only on the much less precise measurements $y_t^{(1)}$.

There are two important implications to consider. On the one hand, in applications where the more precise measurements $y_t^{(2)}$ might be costly to obtain, there appears to be clear scope for cost-benefit analysis when deciding whether to collect all of them or only a suitably small random fraction of them. On the other hand, in applications where the more precise measurements $y_t^{(2)}$ can be easy to obtain but might be subject to outliers, there appears to be clear scope for robustifying the filter by purposely limiting and randomizing the use of $y_t^{(2)}$. Our focus in what follows is on developing the latter as a randomized missing data approach to robust filtering of state space models, while relegating the former for future research.

To illustrate this idea, we extend our motivating example by adding large outliers of size ten standard deviations to 1% of the measurements $y_t^{(2)}$. We then set the fraction $\beta = 10\%$ of the measurements that will be used and run the Kalman filter for many different uniform random draws of a fraction β of the measurements $y_t^{(2)}$, while treating the rest as missing. Averaging the filtered states across the different draws drastically reduces the impact of outliers on the Kalman filter as illustrated on Figure 3.

Combining all of the above suggests that such a randomized missing data approach to filtering state space models can help greatly limit the import of outlier-prone measurements without much of an efficiency loss. Table 1 sheds further light on this by tabulating parameter RMSE for different proportions $\beta = 0\%, 1\%, 2\%, 5\%, 10\%, 20\%, 30\%, 40\%, 50\%, 60\%, 70\%, 80\%, 90\%, 100\%$ of retained measurements $y_t^{(2)}$ with the corner case $\beta = 0\%$ corresponding to using only $y_t^{(1)}$ and not using any of $y_t^{(2)}$ and the corner case $\beta = 100\%$ corresponding to using $y_t^{(1)}$ and all of $y_t^{(2)}$. In particular, the

reported RMSE values in Table 1 confirm that there is indeed a pretty minimal loss of efficiency when using relatively small fractions β of the more precise measurements $y_t^{(2)}$ in the absence of outliers, while in the presence of outliers there can be enormous efficiency gains from substantial missing data randomization over $y_t^{(2)}$ (e.g. $\beta = 10\%$) compared to the two corner cases of using only $y_t^{(1)}$ (i.e. $\beta = 0\%$) or using $y_t^{(1)}$ and all $y_t^{(2)}$ (i.e. $\beta = 100\%$).

To sum up, this motivating example suggests that inducing randomized missing observations can be pursued successfully with the Kalman filter to achieve robust filtering of state space models without much of an efficiency loss. In the next section we show how the same idea can be exploited much more generally and computationally efficiently in the case of particle filtering by developing a more formal randomized missing data approach to robust particle filtering of state space models.

3. Robust Filtering via Missing Data Randomization

3.1 Setup and notation

Our state space model has two sets of parameters: a set of one or more parameters controlling the randomization of missing data with $\beta \in [0, 1]$ defining the target fraction of non-missing highly precise observations, and a set Θ that includes all other SSM parameters. $\theta \in \Theta$ is identified by the in-sample data and is estimated using the SMC^2 method of Chopin, Jacob, and Papaspiliopoulos (2013). The value of β is not directly identified by the data and is meant to be specified by the econometrician by controlling the robustness-efficiency trade-off resulting from inducing a fraction $1 - \beta$ of randomized missing observations. The choice $\beta = 0$ corresponds to not using any highly precise measurements whatsoever, while the choice $\beta = 1$ corresponds to using all of them without inducing any missing data randomization. Such corner values of β, however, would be suboptimal for out-of-sample forecasts as long as some sufficiently sparse subsets of the highly-precise measurements are not contaminated with the distortions encountered when using the full set of measurements. Therefore, we choose a set of values of β on a suitable grid on the unit interval and find corresponding parameter estimates $\Theta(\beta)$. Our goal is to obtain a range of non-corner values of β for which the estimated SSM model produces better out-of-sample forecasts than those for the corner values $\beta = 0$

and $\beta = 1$, thereby characterizing the robustness-efficiency trade-off. The out-of-sample criterion is based only on the ability to forecast the reliable part of the data. This enables a direct comparison to the model that only uses a single, reliable, but noisy measurement equation.

Formally, in our setup we consider a general state space model with, possibly unknown, parameters $\theta \in \Theta$, and latent Markov process x_t with an initial state density $\mu_\theta(x_1)$

$$y_t^{(1)}|x_t, \theta \sim f_1(.|x_t, \theta) \tag{4}$$

$$y_t^{(2)}|x_t, \theta \sim f_2(.|x_t, \theta) \tag{5}$$

$$y_t^{(j)} \sim f_j(.|Y^{t-1}) \tag{6}$$

$$x_t|x_{t-1}, \theta \sim g(.|x_{t-1}, \theta) \tag{7}$$

$$C_t \sim Bernoulli(\beta) \tag{8}$$

$$\hat{y}_t^{(2)} = \begin{cases} y_t^{(2)} \text{ if } C_t = 1 \\ y_t^{(j)} \text{ if } C_t = 0 \end{cases} \tag{9}$$

where T is the number of observations, $t = 1, 2, \ldots T$, $x_t \in \mathbb{R}^{d_x}$, $g(x_t|x_{t-1})$ is the state transition kernel, $y_t^{(1)} \in \mathbb{R}^{d_{y1}}$, $y_t^{(2)} \in \mathbb{R}^{d_{y2}}$, and $y_t^{(j)} \in \mathbb{R}^{d_{y2}}$. We assume that, conditional on x_t and θ, $y_t^{(1)}$, $y_t^{(2)}$ and x_t are independent.[1] Most importantly, f_1 and f_2 are observation kernels through which the signal about the latent states x_t and the parameters θ is received, while f_j is the observation kernel that is uninformative about the latent states and that does not depend on the set of model parameters θ. The indicator variable C_t is not directly observed, and controls when the observation $\hat{y}_t$ is (and when it is not) informative. The observations are divided into two groups depending on the level of precision: a low-precision set $y_t^{(1)}$ observed at all times t, and a high-precision set $y_t^{(2)}$ randomly incorporated only when an indicator C_t takes unit value, and treated as missing when $C_t = 0$. The property of low precision of $y_t^{(1)}$ is defined by $var(y_t^{(1)}|x_t) \gg var(y_t^{(2)}|x_t)$. Note that when $C_t = 0$, $y_t^{(2)}$ is a latent variable and $y_t^{(j)}$ is observed but does not provide any relevant information. Thus, if $C_t = 0$, the observed value $\hat{y}_t^{(2)}$ is completely uninformative. In Figure 1, we illustrate the block-diagram model of the dependencies in our missing data randomization SSM specification.

[1] This condition is further relaxed in our finance application by assuming dependence between conditional distributions of x_t and $y_t^{(1)}$ i.e., the leverage effect.

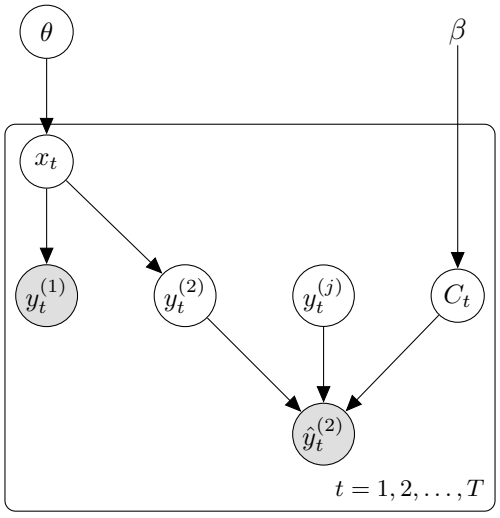

Figure 1: SSM model with possibly corrupted/uninformative second measurement

3.2 SMC - based estimation method

3.2.1. Particle Filter

Our primary goal is to infer the posterior distribution $P(\theta, x^t | y^{(1),t}, \hat{y}^{(2),t}, \alpha)$. For brevity we use the notation $Y^t := (y^{(1),t}, \hat{y}^{(2),t})$ to denote all observed variables up to and including t. We define $\beta_t := P(C_t = 1 | Y^{t-1}, \beta)$. In the simplest version of our model we assume that β_t is constant in time, i.e., $\beta_t = \beta$ for all t. The posterior probability $\hat{\beta}_t := P(C_t = 1 | \hat{y}_t^{(2)}, \hat{Y}^{t-1}, \alpha, \beta)$ can be derived by the application of Bayes' rule.

$$\begin{aligned}
\hat{\beta}_t := P(C_t = 1 | \hat{y}_t^{(2)}, Y^{t-1}, \alpha, \beta) &= \frac{\beta_t P(\hat{y}_t^{(2)} | C_t = 1, Y^{t-1}, \alpha)}{\beta_t P(\hat{y}_t^{(2)} | C_t = 1, Y^{t-1}, \alpha) + (1 - \beta_t) P(\hat{y}_t^{(2)} | C_t = 0, Y^{t-1}, \alpha)} \\
&= \frac{\beta_t f_2(\hat{y}_t^{(2)} | Y^{t-1}, \alpha)}{\beta_t f_2(\hat{y}_t^{(2)} | Y^{t-1}, \alpha) + (1 - \beta_t) f_j(\hat{y}_t^{(2)} | Y^{t-1}, \alpha)} \\
f_2(\hat{y}_t^{(2)} | Y^{t-1}, \alpha, \beta) &= \iint_{x_t, \theta} f_2(\hat{y}_t^{(2)} | x_t, \theta) dP(x_t, \theta | Y^{t-1}, \alpha, \beta)
\end{aligned} \tag{10}$$

$$P(\theta, x^t|Y^t, \alpha, \beta) = \frac{f_1(y_t^{(1)}|x_t, \theta)}{f_1(y_t^{(1)}|\hat{y}_t^{(2)}, Y^{t-1}, \alpha)}(\hat{\beta}_t \frac{f_2(\hat{y}_t^{(2)}|x_t, \theta)}{f_2(\hat{y}_t^{(2)}|Y^{t-1}, \alpha)} + (1 - \hat{\beta}_t))P(\theta, x^t|Y^{t-1}) \tag{11}$$

To apply the SMC^2 method of Chopin, Jacob, and Papaspiliopoulos (2013), we derive the filtering procedure to estimate the latent states conditional on θ, rather than directly sample from the joint distribution in equation (10). Specifically, given a sample from $P(x^{t-1}|Y^{t-1}, \theta)$, we show how to draw from $P(x^t|Y^t, \theta)$. By conditioning on θ, we introduce an additional source of information about the latent indicators C_t, that requires a slight modification of equation (10) by the subsequent application of Bayes' theorem:

$$\begin{aligned}\hat{\beta}_{t,\theta} := P(C_t = 1|\hat{y}_t^{(2)}, Y^{t-1}, \theta, \beta) &= \frac{\hat{\beta}_t P(\theta|\hat{y}_t^{(2)}, C_t = 1, Y^{t-1})}{\hat{\beta}_t P(\theta|\hat{y}_t^{(2)}, C_t = 1, Y^{t-1}) + (1 - \hat{\beta}_t)P(\theta|\hat{y}_t^{(2)}, C_t = 0, Y^{t-1})} \\ &= \frac{\hat{\beta}_t f_2(\hat{y}_t^{(2)}|\theta, Y^{t-1})}{\hat{\beta}_t f_2(\hat{y}_t^{(2)}|\theta, Y^{t-1}) + (1 - \hat{\beta}_t) f_2(\hat{y}_t^{(2)}|Y^{t-1})}\end{aligned} \tag{12}$$

This yields a mixture analogous to the formula for the joint posterior distribution in equation (11):

$$P(x^t|Y^t, \theta, \beta) = \frac{f_1(y_t^{(1)}|x_t, \theta)}{P(y_t^{(1)}|\hat{y}_t^{(2)}, Y^{t-1}, \theta)}(\hat{\beta}_{t,\theta} \frac{f_2(\hat{y}_t^{(2)}|x_t, \theta)}{f_2(\hat{y}_t^{(2)}|Y^{t-1}, \theta)} + (1 - \hat{\beta}_{t,\theta}))P(x^t|Y^{t-1}, \theta) \tag{13}$$

The above can be further simplified substituting for $\hat{\beta}_{t,\theta}$ from equation (12), which yields the conditional posterior distribution for x^t:

$$\begin{aligned}P(x^t|Y^t, \theta, \beta) &= \frac{f_1(y_t^{(1)}|x_t, \theta)}{P(y_t^{(1)}|\hat{y}_t^{(2)}, Y^{t-1}, \theta)} \frac{\hat{\beta}_t f_2(\hat{y}_t^{(2)}|x^t, \theta, Y^{t-1}) + (1 - \hat{\beta}_t) f_2(\hat{y}_t^{(2)}|Y^{t-1})}{\hat{\beta}_t f_2(\hat{y}_t^{(2)}|\theta, Y^{t-1}) + (1 - \hat{\beta}_t) f_2(\hat{y}_t^{(2)}|Y^{t-1})} P(x^t|Y^{t-1}, \theta) \\ &\propto f_1(y_t^{(1)}|x_t, \theta)\hat{\beta}_t(f_2(\hat{y}_t^{(2)}|x^t, \theta, Y^{t-1}) + (1 - \hat{\beta}_t) f_2(\hat{y}_t^{(2)}|Y^{t-1}))g(x_t|x_{t-1}, \theta)P(x^{t-1}|Y^{t-1}, \theta)\end{aligned} \tag{14}$$

Thus, the sequential estimation of latent states can be performed by propogating forward with $g(x_t|x_{t-1},\theta)$ and reweighting with:

$$w_{x^t,\theta,t} := f_1(y_t^{(1)}|x_t,\theta)\hat{\beta}_t(f_2(\hat{y}_t^{(2)}|x^t,\theta,Y^{t-1}) + (1-\hat{\beta}_t)f_2(\hat{y}_t^{(2)}|Y^{t-1})) \tag{15}$$

Compared to the standard bootstrap filter of Gordon, Salmond, and Smith (1993), this procedure only requires the additional step of calculating the predictive density $f_2(\hat{y}_t^{(2)}|Y^{t-1})$. This can be estimated by integrating over the joint distribution of $(\theta, x^{t-1})|Y^{t-1}$ as shown in equation (10).

3.3 Estimation of Model Parameters: SMC^2 Method

To estimate the posterior distribution of model parameters (θ) from the observations Y^T, we use the SMC^2 algorithm from Chopin, Jacob, and Papaspiliopoulos (2013). This algorithm performs Bayesian inference by sampling from the posterior distributions $P(\theta|Y^t,\beta)$ sequentially for $t = 1,2,...,T$.

To begin, we integrate out the latent states (x^t) from the posterior distribution in equation (11).

$$\begin{aligned} P(\theta|Y^t,\alpha,\beta) &= \frac{f_1(y_t^{(1)}|\theta)}{f_1(y_t^{(1)}|\hat{y}_t^{(2)},Y^{t-1},\alpha)}(\hat{\beta}_t\frac{f_2(\hat{y}_t^{(2)}|\theta)}{f_2(\hat{y}_t^{(2)}|Y^{t-1},\alpha)} + (1-\hat{\beta}_t))P(\theta|Y^{t-1}) \\ &\propto f_1(y_t^{(1)}|\theta)(\hat{\beta}_t f_2(\hat{y}_t^{(2)}|\theta) + (1-\hat{\beta}_t)f_2(\hat{y}_t^{(2)}|Y^{t-1},\alpha))P(\theta|Y^{t-1},\alpha,\beta) \end{aligned} \tag{16}$$

Let $w_{\theta,t} := f_1(y_t^{(1)}|\theta)(\hat{\beta}_t f_2(\hat{y}_t^{(2)}|\theta) + (1-\hat{\beta}_t)f_2(\hat{y}_t^{(2)}|Y^{t-1},\alpha))$. We can then target $P(\theta|Y^t)$ sequentially via importance sampling, using the fact that $P(\theta|Y^t) \propto w_{\theta,t}P(\theta|Y^{t-1}) \propto \big(\prod_{i=1}^t w_{\theta,i}\big)P(\theta|\alpha)$. In general, $w_{\theta,t}$ is not analytically tractable. However, we can utilize the importance weights derived in equation (15) to obtain an unbiased estimate, denoted as $\tilde{w}_{\theta,t}$.

$$
\begin{aligned}
w_{\theta,t} &= \int_{x^t} f_1(y_t^{(1)}|x_t,\theta)(\hat{\beta}_t(f_2(\hat{y}_t^{(2)}|x^t,\theta,Y^{t-1}) + (1-\hat{\beta}_t)f_2(\hat{y}_t^{(2)}|Y^{t-1}))dP(x^t|\theta,Y^{t-1}) \\
&= \int_{x^t} w_{x^t,\theta,t} dP(x^t|\theta,Y^{t-1}) \\
\tilde{w}_{\theta,t} &= \sum_{j=1}^{N_x} w_{x_j^t,\theta,t}
\end{aligned}
\tag{17}
$$

where the above sum is calculated over N_x paths of $x_j^t \sim P(.|\theta, Y^{t-1})$ drawn using the particle filter in Section 3.2.1.

To avoid sample degeneration, SMC^2 augments the importance sampling above with a particle Markov Chain Monte Carlo (PMCMC) kernel. Starting with a sample of $\theta^m \sim P(.|Y^t, \alpha, \beta)$, the weights $\tilde{w}_{\theta,t+1}, \tilde{w}_{\theta,t+2}, \ldots$ are calculated until the implied effective sample size (ESS) of θ^m drops below a pre-specified threshold at some $t' > t$. The PMCMC kernel presented in Andrieu, Doucet, and Holenstein (2010) is then used to rejuvenate the sample of θ. Since this kernel targets $P(\theta|Y^{t'}, \alpha, \beta)$ via an "exact approximation", the weights are reset after rejuvenation ($\prod_{i=1}^{t'} w_{\theta,i} = 1$). The combination of these two steps allow sampling from the filtered marginal distribution of $P(\theta|Y^t, \beta)$ at each t in an online manner. Moreover, since the particle filter targets $P(x^t|Y^t, \theta^m, \beta)$ for each θ^m, the algorithm ultimately yields a sample from the joint distribution $P(\theta, x^t|y_{1:t}^{(1)}, \hat{y}_{1:t}^{(2)}, \beta)$.

3.4 Selection of alternate distribution

As shown in (10) and (11), the choice of f_j plays a role in determining the proportion of inclusion for the highly precise measurement $y_t^{(2)}$. It turns out that defining β as a $(\hat{y}, Y^{t-1}, \alpha, \beta)$-measurable function is equivalent to choosing a density for f_j. For example, specifying that $\hat{\beta}_t = \beta_t$ yields the "blind" randomization from the motivating example in Section 2. This assumption is equivalent to setting the density $f_j(\hat{y}_t^{(2)}|Y^{t-1}, \alpha)$ to the predictive distribution $f_2(\hat{y}_t^{(2)}|Y^{t-1}, \alpha)$. This is shown by applying Bayes' rule:

$$\begin{aligned} \beta_t &= \frac{\beta_t f_2(\hat{y}_t^{(2)}|\hat{Y}^{t-1},\alpha)}{\beta_t f_2(\hat{y}_t^{(2)}|\hat{Y}^{t-1},\alpha) + (1-\beta_t) f_j(\hat{y}_t^{(2)}|\hat{Y}^{t-1},\alpha)} \\ f_j(\hat{y}_t^{(2)}|Y^{t-1},\alpha) &= f_2(\hat{y}_t^{(2)}|Y^{t-1},\alpha) \end{aligned} \tag{18}$$

In this case, there is no unconditional learning about whether the data was contaminated on a particular day from concurrent information. To suppress outliers, a more diffuse f_j will weight against observations in the tails. For example, within the context of forecasting, one possible candidate for f_j is $f_j(\hat{y}_t^{(2)}|Y^{t-1}) = f_2(\hat{y}_t^{(2)}|Y^{t-h})$, where h is some horizon. Then we would have:

$$\begin{aligned} \hat{\beta}_t &= \frac{\beta_t f_2(\hat{y}_t^{(2)}|\hat{Y}^{t-1},\alpha,\beta)}{\beta_t f_2(\hat{y}_t^{(2)}|\hat{Y}^{t-1},\alpha,\beta) + (1-\beta_t) f_2(\hat{y}_t^{(2)}|Y^{t-h},\alpha,\beta)} \\ &= \frac{\beta_t P(\hat{Y}_{t-h+1:t-1}|\hat{y}_t^{(2)}, C_t = 1, \hat{Y}^{t-h},\alpha,\beta)}{\beta_t P(\hat{Y}_{t-h+1:t-1}|\hat{y}_t^{(2)}, C_t = 1, \hat{Y}^{t-h},\alpha,\beta) + (1-\beta_t) P(\hat{Y}_{t-h+1:t-1}|\hat{Y}^{t-h},\alpha,\beta)} \end{aligned} \tag{19}$$

In this way, we determine $\hat{\beta}_t$ using the likelihood of our previous h observations given $\hat{y}_t^{(2)}$ is a valid observation.

3.5 Optimality of the randomization parameter β

In this Section, we discuss the optimal choice of the parameter β. To setup the notation, we define a model M_{β,f_j} that follows the above specifiation, making explicit its dependence on the parameter β and the distribution of the junk states f_j. When fitting the model below, we assume that there is only a single measurement equation corresponding to $y_t^{(2)}$ that might be contaminated, while the measurement equation corresponding to $y_t^{(1)}$ has no information about the latent states. We also fit our model with the additional assumption that the observation kernel $f_j(\hat{y}_t^{(2)}) = f_2(\hat{y}_t^{(2)}|Y^{t-1})$, denoting this model as M_β. Both assumptions are made without loss of generality, in part because we allow for the true model that generates the data, denoted $D = M_{\hat{\beta},f_j}$, to follow an arbitrary f_j.

As an objective, we want to maximize the following expectation for the data-generating process D:

$$\begin{aligned}
&\mathbb{E}_{\hat{y}^{t+1}}(\log(P(\hat{y}_{t+1}|\hat{y}^t, M_\beta))|D) = \\
&\quad = \int_{\hat{y}^{t+1}} \log\Big(\int_{(x^t,\theta)} f_2(\hat{y}_{t+1}|x^t,\theta) \int_{C^t} dP(C^t|\beta) \frac{f_2(\hat{y}_{\{i:C_i^t=1\}}|x^t,\theta)}{F_2(\hat{y}_{\{i:C_i^t=1\}})} dP(x^t,\theta)\Big) dP(\hat{y}^{t+1}|D) \\
&\quad = \int_{\hat{y}^{t+1}} \log\Big(\sum_{C^t} P(C^t|\beta) \int_{(x^t,\theta)} f_2(\hat{y}_{t+1}|x^t,\theta) \frac{f_2(\hat{y}_{\{i:C_i^t=1\}}|x^t,\theta)}{F_2(\hat{y}_{\{i:C_i^t=1\}})} dP(x^t,\theta)\Big) dP(\hat{y}^{t+1}|D) \quad (20)
\end{aligned}$$

where $\{i : C_t^t = 1\}$ denotes the set of included observations as specified by the path of C^t. Below we present our main theory results:

Theorem 1 (Coherence of the optimal β in the special case $D = M_{\hat{\beta}}$)**.** *Suppose that $D = M_{\hat{\beta}}$ with $\hat{\beta} \in [0, 1]$. Then the critical point, $\bar{\beta}$, of the expectation of the log-forecasting density (20) is $\bar{\beta} = \hat{\beta}$.*

Proof. See the Appendix. □

We also verify that the critical point is a unique maximum:

Theorem 2 (Existence of an optimal β in the general case $D = M_{\hat{\beta},f_j}$)**.** *Suppose that $D = M_{\hat{\beta},f_j}$ with $\hat{\beta} \in [0, 1]$ and general f_j. Then the expectation of the log-forecasting density (20) with respect to β has a global maximum on $[0, 1]$.*

Proof. See the Appendix. □

The above results are important in light of the quality of the density forecasts formulated using our model specification. Theorem 1 states that, in the ideal case where $D = M_\beta$, the optimal choice of the parameter β coincides with the true value of the randomization parameter $\hat{\beta}$. Moreover, Theorem 2 guarentees that there is a global maximum $\bar{\beta} \in [0, 1]$, even in the case where the data-generating process has an arbitrary, possibly unknown distribution governing the junk states. However, in this general case, it is not necessarily true that $\bar{\beta} = \hat{\beta}$, and the optimality of $\bar{\beta}$ is determined by the characteristics of the true f_j.

4. Empirical Illustrations

4.1 Real data application: inflation forecasting

Market participants and central bankers alike traditionally pay considerable attention to data-driven assessments of long-run inflation expectations. A large part of the literature on inflation forecasting has considered alternative econometric approaches to estimating inflation trends based on time series modelling of officially released price index data.[2] Stock and Watson (2007) and Stock and Watson (2016) provide compelling evidence for time-variation in the precision of inflation rate measurements as well as the presence of additional measurement distortions due to outliers. Inspired by these findings, we consider the ability of our robust filtering approach to successfully guard against the impact of inflation measurement imperfections without the need to explicitly model them for the purposes of improved forecasting of long-run inflation trends. In particular, we apply our robustification framework to the classical unobserved components (UC) representation of the IMA(1,1) "naïve" benchmark model for inflation forecasting scrutinized also by Stock and Watson (2007). We then compare in terms of out-of-sample forecasting performance our robustly filtered version of the classical UC model vis-à-vis the unobserved components/stochastic volatility outlier-adjustment (UCSVO) model, which was put forward by Stock and Watson (2016) as a way to minimize the detrimental impact of outliers by way of subjecting them to particular distributional assumptions. For illustrative purposes we therefore follow Stock and Watson (2016) and model the same aggregate PCE price index quarterly data series (PCE-all) from 1960Q1 to 2015Q2 with an identical forecast evaluation period set to 1990Q1-2015Q2.[3]

[2] Another large strand of the literature aims to exploit the predictive content of collected survey-based measures of inflation. For a recent literature survey on inflation forecasting see for example Faust and Wright (2013).

[3] We thank Stock and Watson (2016) for providing as an online supplement the data and program codes necessary for replicating their results.

The UC model specification we consider is given by the following set of equations:

$$y_t|x_t, \theta \sim f_2(.|x_t) = N(x_t, \sigma_\eta^2) \tag{5'}$$

$$y_t^{(j)} \sim f_j(.) = f_2(\hat{y}_t|Y^{t-1}) \tag{6'}$$

$$x_t|x_{t-1}, \theta \sim g(x_t|x_{t-1}) = N(x_{t-1}, \sigma_\varepsilon^2) \tag{7'}$$

$$C_t \sim Bernoulli(\beta), \tag{8'}$$

where y_t denotes the PCE inflation rate measurement at time t, x_t is the latent trend component of inflation, and C_t is the Bernoulli indicator variable inducing missing data randomization as a key feature of our robust filtering framework specialized to the classical UC model. We compare 4, 8 and 12 quarter ahead forecasts of average inflation based on our robustly filtered UC model specification against forecasts based on the UCSVO model specification directly taken from Stock and Watson (2016) and document that missing data randomization applied to the basic UC model with a suitably chosen parameter β can produce UC model forecasts that perform on par or better than the UCSVO model forecasts favored by Stock and Watson (2016).

It is important to note that the UC model has only one observation equation with possibly contaminated data.[4] Hence, the choice of the parameter β must be determined using the forecasting distribution of the data that might be prone to contamination, which we model by assuming its lack of information about the latent inflation states defined in equation (6'). The favorable forecasting results we obtain for our robustly filtered UC model thus further underscore the findings in Stock and Watson (2007) and Stock and Watson (2016) regarding the need for time-series inflation forecasts to address the presence of potentially complex inflation measurement imperfections. While our robustly filtered UC model appears to be able to offer some gains in terms of forecasting performance (perhaps thanks to avoiding any stringent parametric assumptions on the outliers as in UCSVO), our results can also be viewed as largely corroborating the remarkably good forecasting performance of the UCSVO model.

In Table 2, we present the parameter estimates for the robustly filtered UC model with our missing data randomization approach for varying values of the parameter β and using the full

[4] The model with a single observation equation can be viewed as a model specified in equations (4) - (8) but with arbitrarily high measurement noise of the first measurement equation (4).

sample from 1960Q1 to 2015Q2 for estimation of the model. It is evident that both the observation and state equation variance increases with the increase of β. This is explained by the requirement of the UC model to fit increasingly more and larger outliers present in the data. The increase is more pronounced for the parameter σ_η that controls the precision of the observation equation. To better illustrate the changes in the estimates of these parameters as we change the randomization parameter β we also present the plots with their estimates in Figure 4. As it is evident from Figure 4, the low and medium values of β do not have a significant impact on the estimates of σ_ε, while the estimates of σ_η grow steadily with the values of β. Since higher values of β naturally imply that more outliers are to be fitted by the model we should expect the variances of both the observation and state equations to increase with this parameter. However, this effect is stronger for the observation equation.

Next, we turn our attention to the forecasting performance of our robustly filtered UC model with missing data randomization versus the UCSVO model. We use the Amisano and Giacomini (2007) test statistics, $\widehat{WLR}$, to compare the forecasting performance of the two models. The $\widehat{WLR}$ statistics are computed with the comparison to the model with $\beta = 1$ that uses all data with no randomization, which corresponds to the standard UC specification in Stock and Watson (2007). Table 3 presents results for the UC model estimated recursively using the sample from 1960Q1 to 2005Q2 to generate forecast densities evaluated at 1-, 4-, 8-, and 12-quarter-ahead horizons starting from 1990Q1. For brevity, we document the results for three different choices of the parameter β of 0.15, 0.25 and 0.9. It is evident that all specifications with data randomization perform better than the benchmark UC model producing positive t-statistics with statistically significant improvements for all horizons for the specifications with lower values of β of 0.15 and 0.25.

Following our results in Theorems 1 and 2, an optimal β exists for any given forecast horizon. In Figures 5 and 6, we present the results discussing in more detail the optimality of β for different choices of the forecasting horizons of 1, 4, 8, and 12 quarters ahead. In order to avoid look-ahead bias, we design our analysis such that it does not use any future observations to determine model parameters, states and the randomization parameter β when forecasts are formulated using expanding windows. Figure 5 presents the results for the full sample spanning the full period from 1960Q1 to 2015Q2 with the mean (across time) forecasting log-densities for each

point on the considered grid of β(s) and each forecast horizon. As expected, the log-densities take higher values at shorter horizons. The optimal β(s) that produce the best forecast performance for each horizon, i.e., the peak values for each curve, take values that allow for significant missing data randomization, and with the peak values of β decreasing with forecast horizons. Hence, the forecasts made for longer horizons prefer a higher degree of missing data randomization naturally alleviating the presence of outliers.

In Figure 6 we present only the derived maxima for each considered forecast horizon derived similar to those in Figure 5, but with varying terminal day of the sample available for forecasting, and with the last point on each curve coinciding with the maxima in Figure 5. As it is visible from Figure 6, the optimal β(s) tend to take higher values early in the sample when the sample size is small and decrease with the amount of available data as forecasts are formulated using the expanding windows. The optimal β(s) also tend to take smaller values as the forecast horizon increases.

In Figures 7 and 8 we further present a comparison of the forecasting performance of our robustly filtered UC model with missing data randomization to the standard UC model ($\beta = 1$), and to the UCSVO model of Stock and Watson (2016). The latent states, parameters and β are estimated from 1960Q1 to 2015Q2 in an on-line fashion, and the results for the mean forecasting densities (Figure 7) and mean squared forecast errors (MSFE) (Figure 8) are formed from 1990Q1 onward. The analysis is performed for each forecast horizon with optimal β determined in a feasible fashion similar to Figure 6. Hence, for each model, the curves correspond to the average (across time) path of evaluated forecasting densities along the path of β(s) derived as in Figure 6, i.e., on each day and for each model the posterior distribution of parameters and the optimal values of beta are updated over time as the new information arrives. As it is visible in Figure 7, all UC models that allow for data randomization are capable of outperforming the standard UC model at all forecasting horizons irrespective of what the chosen criterion for the optimal β is (ranging from maximizing the performance of forecast densities from 1 up to 12 quarters ahead). The MSFEs in Figure 8 clearly show similar pattern to the one documented in Figure 7 but also allow for a direct comparison with the results obtained for the UCSVO model in Stock and Watson (2016). All models allowing for missing data randomization dominate the classical UC model and perform

better than the UCSVO model (with the exception of the model targeting a 1-quarter ahead forecast objective). Summing up, our results indicate that our robust filtering approach via missing data randomization can significantly improve the forecasting performance of the classical UC model by naturally safeguarding against overfitting outliers as a new viable alternative to the UCSVO model.

4.2 Simulation Results: Nonlinear SSM

In this Section, we present simulation-based results for a stochastic volatility model defined by the following observation and state equations:

$$y_t^{(1)}|x_t, \theta \sim f_1(.|x_t) = N(0, exp(\frac{x_t}{2})) \tag{4"}$$

$$y_t^{(2)}|x_t, \theta \sim f_2(.|x_t) = N(x_t, \sigma_{2,t}^2) \tag{5"}$$

$$y_t^{(j)} \sim f_j(.) = f_2(.|x_t) \; s.t. \; |y_t^{(j)} - x_t| > 5\sigma_{2,t} \tag{6"}$$

$$x_t|x_{t-1}, \theta \sim g(x_t|x_{t-1}) = N(\kappa_x\theta_x + (1-\kappa_x)x_{t-1}, \sigma_x^2) \tag{7"}$$

$$C_t \sim Bernoulli(\beta). \tag{8"}$$

and with the rest of the model assumptions following the specification in equations (4) - (9). We assume that the contaminated data are sampled from a truncated normal distribution with the same mean as in equation (5"), but the error term is at least 5 standard deviations in either of the tails.[5] As it is standard in the asset pricing literature, we assume that normal innovations in equations (4") and (7") are correlated with correlation ρ_x, the so called leverage effect. Without loss of generality, the variance $\sigma_{2,t}^2$ is set at the constant implied by the asymptotic variance factor given by 2.96/78 for 5-minute returns and the MedRV variance measure from Andersen, Dobrev, and Schaumburg (2012).[6] Observations $y_t^{(1)}$ are interpreted as low-frequency asset returns, and observations $\hat{y}_t^{(2)}$ are high-frequency volatility measures obtained using intra-day returns. Hence, the first observation equation is called the return equation, and the second is called the volatility

[5] Our specification of the contaminated observations ensures that there is very limited information content about latent states contained by these data points. In the next version of the paper the analysis will be extended to allow for other contamination specifications.

[6] Alternative jump-robust variance measures include most notably the bipower and multipower variation measures introduced by Barndorff-Nielsen and Shephard (2004), Barndorff-Nielsen and Shephard (2005b), and Barndorff-Nielsen, Shephard, and Winkel (2006). For more details see Section 2.2 of Dobrev and Szerszen (2010).

measurement equation. The state equation is a standard AR(1) process with mean reversion rate κ_x, mean θ_x, and volatility of volatility parameter σ_x. If $C_t = 0$ for all t, there is only a single return equation in the model, so the model falls into the class of stochastic volatility models that has been widely applied in the finance and asset pricing literature.[7] On the other hand, the model with additional volatility measurement equation, $C_t = 1$ for all t, has also been studied in the more recent strain of this literature, see for example Creel and Kristensen (2015), Koopman and Scharth (2013), Dobrev and Szerszen (2010) and Takahashi, Omori, and Watanabe (2009). We consider a more general specification allowing for unreliable volatility measurement observations with equation (8").

Table 4 presents estimation results for the model specification with slow mean reversion (high persistence) of $k_x = 0.0132$. The slow mean reversion is consistent with the S&P 500 index return data. All other model parameters are also set at a level reflecting the range of values typically found for S&P 500 index returns. To generate the data, we draw a sample of $T = 1260$ observations (i.e., 5 years of daily data) and then perform the estimation discussed in Section 3.3 using the particle filter from Section 3.2.1. The estimations are performed at various levels of parameter β of 0, 0.01, 0.1, 0.9, and 1. In the table the first three columns present the posterior medians and associated high credibility (95%) intervals (HCIs) for the model with perfectly reliable volatility measurements. The middle three columns present the estimation results in a similar setting and with a rate of 10% of volatility measurement observations that are unreliable and contaminated. The last three columns present the results for the case of very highly contaminated data with 90% of volatility measurement observations that are unreliable.

For the models that use uncontaminated data shown in the left three columns, the estimation results show a reasonable level of estimation precision for all values of β. As found in Dobrev and Szerszen (2010), we can also expect the model with $\beta = 1$ to outperform other models as it allows for all precise measurements to be fully utilized. Dobrev and Szerszen (2010) document that the parameters ρ_x and σ_x show the biggest efficiency gains, which also translates into better estimates of other coefficients, including the mean reversion coefficient κ_x, which naturally requires a long sample, i.e., a large T, to be precisely estimated.

[7] See Andersen, Benzoni, and Lund (2002), Eraker, Johannes, and Polson (2003), Jacquier, Polson, and Rossi (2002), Jacquier, Polson, and Rossi (2004) and Jones (2003), among many others.

In contrast to the fully reliable data, the model estimated with a high proportion of unreliable $\hat{y}_t^{(2)}$ observations shows very large biases for high levels of β 0.9 and 1, and very good performance at lower values of β. For the presented levels of contamination of 10% and 90%, and a level of β higher than the proportion of reliable data, the unreliable observations contribute to the overestimation of the volatility of volatility parameter σ_x. The parameter ρ_x is also estimated with a bias that reduces the leverage effect. These findings are consistent with the models fitting the outliers produced by the unreliable volatility measurements $\hat{y}_t^{(2)}$. The biases are mostly eliminated for the lower values of β of 0, 0.01 and 0.1 for the model with 90% contamination, and for $\beta = 0, 0.01, 0.1, 0.9$ for the model with 10% contamination.[8] The positive bias in σ_x can be explained by inspecting the posterior quantiles and mean of the latent volatility states x_t shown in Figures 9 and 10. Since the higher values of β allow the distorted $\hat{y}_t^{(2)}$ observations to have a bigger impact on the volatility estimates, they contribute to the much higher volatility of volatility estimates. An interesting feature of the model is mixing between the high-frequency information content and the daily-data content for the non-extreme values of the parameter β, clearly visible in the posterior variance of the latent volatility states presented in Figure 11. The spikes in the variance of latent volatility states indicate a high discrepancy between both sources of information.

As discussed in Section 3.1, the parameter β is not identified by the in-sample data. However, we can find the level of β that produces optimal out-of-sample forecasts. We use the test of Amisano and Giacomini (2007) for comparing the out-of-sample forecast performance of two models. Under the null hypothesis of equal performance, the test statistic, $\widehat{WLR}$, has a standard normal distribution. We choose the model using only noisy $y_t^{(1)}$ observations ($\beta = 0$) as a baseline specification to which we compare other model specifications ($\beta = 0.01, 0.1, 0.9, 1$). Positive test statistics indicate better performance of the competing model, and negative values indicate better performance of the baseline model.

The Amisano and Giacomini (2007) test statistics, $\widehat{WLR}$, are presented in Table 5 for five forecasting horizons of 1, 2, 5, 10 and 22 days. We first turn our attention to the estimation with uncontaminated and fully reliable additional measurements $\hat{y}_t^{(2)}$. As it is evident from the Table, all models that use additional information content of $\hat{y}_t^{(2)}$ observations ($\beta > 0$) outperform the baseline

[8] In the next version of the paper we plan to show to what extent the biases are reduced by running a high number of independent replications of the data.

model that uses only the much noisier measurement equation $y_t^{(1)}$ ($\beta = 0$). In contrast, the models with high values of β (0.9, 1) perform much worse than the baseline model when the data is highly contaminated (90% contamination), producing negative test statistics.

Most importantly, even in the case of highly contaminated 90% proportion of the data, the model with $\beta = 0.01, 0.1$ that incorporates information from the additional measurement equation performs well at shorter forecast horizons assuming highly persistent volatility states. For the model with only 10% of contaminated data, the improvements over the baseline model are found for all considered forecasting horizons ranging from 1 day, up to 22 days (1 month) and for the non-extreme values of β of 0.1 and 0.9. This can be explained by a better ability to filter highly persistent latent states even if a smaller proportion of highly informative observations is reliable.

4.3 Real data application: S&P 500 return density forecasting

Next we consider a real data application of the above stochastic volatility model defined in equations (4") - (8") to S&P 500 return density forecasting. Following up on the convention set up in the previous Section, the daily open-to-close S&P 500 index return observations are denoted as $y_t^{(1)}$, and high-frequency based daily open-to-close volatility measure observations are denoted as $\hat{y}_t^{(2)}$. In addition, assuming 78 5-minute intraday returns within each day, we also directly observe $\sigma_{2,t}^2 = \frac{2.96\widehat{IQ}_t}{78\widehat{IV}_t^2}$.[9] We split the analysis and consider three sample periods each covering approximately 7 years of data: the first covering the data from 1998-01-02 to 2004-11-11, the second covering the period from 2004-11-12 to 2011-06-07, and the third covering the most recent data from 2011-06-08 to 2017-12-29. Our goal is to show that the results hold for all data sets irrespective of the studied sample periods.

Similar to Section 4.2, we first present the parameter estimates. In Table 6 the posterior medians are reported for each of the model parameters for different values of the parameter β. The small values of κ_x indicate that volatility is a highly persistent process. At lower values of β (0, 0.01 and 0.1), we find strong evidence for a leverage effect with negative values of the estimates of ρ_x ranging from -0.84 to -0.71 for the sample starting in 1998, from -0.37 to -0.29 for the sample

[9] For further details see Section 2.2 of Dobrev and Szerszen (2010).

starting in 2011, and with the estimates for the sample starting in 2004 taking intermediate values ranging from -0.5 to -0.3. In contrast to the findings for small values of β, however, we find vastly different and lower (in absolute value) estimates of the leverage effect ρ_x and higher volatility of volatility σ_x estimates for higher values of the parameter β of 0.9 and 1. These results are similar to the simulation results with unreliable and contaminated observations in Section 4.2 suggesting that the high frequency measures, despite providing a very precise information about the latent volatility states, might be prone to instabilities driving a wedge between a model estimated using only daily returns ($\beta = 0$), and the model with high frequency volatility measures and high values of β. The simulation results in Section 4.2 also suggest the upward bias in the estimates of the parameter σ_x. Our results for the S&P 500 data confirm this finding, as the estimates of volatility of volatility σ_x are much higher for higher values of the parameter β. This finding is consistent with the filtered estimates of the latent volatility presented in Figures 12 - 14 for each of the studied subsamples. In the figures, as the values of parameter β increase, we can see more pronounced outliers in the posterior means of the volatility estimates and higher volatility of volatility, due to possible high-frequency data imperfections, and much smaller variance of the posterior distribution due to very precise $\hat{y}_t^{(2)}$ measurements. Hence the model with high values of β must fit characteristics of volatility that are not necessarily consistent with the daily data, corresponding to the estimation with $\beta = 0$, that shows less pronounced peaks and troughs.

An important feature of our model is that it allows for mixing between using and shutting down the information content of the possibly contaminated measurements $\hat{y}_t^{(2)}$ on the posterior distributions of parameters and states. Hence, it does not simply exclude the observation, but allows for learning about the consistency of the additional, high-frequency based observation equation, with the daily return data $y_t^{(1)}$. The posterior estimates of the C_t mixing variable are presented in Figures 15 - 17. It is evident that the posterior means are close to the unconditional levels of the assumed values of the parameter β, with occasional spikes and troughs. The most notable spikes of the mixing variable can be found during the financial crisis of late 2007 and 2008, and hence the model learns about rapid changes in volatility estimates during turbulent market events. A closer inspection of the posterior distributions of the indicator C_t in Figures 15 - 17 and volatility x_t in Figures 12 - 14 indicates that the spikes and troughs seen in the filtered distribution of C_t often coincide with

the spikes of the variance of the filtered volatility estimates for the non-extreme values of β, i.e., when the model is mixing between the inclusion or rejection of the high-frequency data. Hence, the model is sensitive to the discrepancy between the signals about the latent volatility coming from the daily data or the high frequency data.

The out-of-sample forecasting results are presented in Table 7. Similar to Table 5, the results are presented for each assumed value of the parameter β and the Amisano and Giacomini (2007) test statistics, $\widehat{WLR}$, are computed for the comparison against the baseline model that uses only daily return data with $\beta = 0$. Our results for the S&P500 returns in Table 7 closely resemble the results in Table 5 for the unreliable and contaminated $\hat{y}_t^{(2)}$ observations. The models with higher values of β, 0.9 and 1, perform equally well or worse in the out-of-sample forecasts than the baseline model using only daily return data ($\beta = 0$) at longer forecasting horizons of 5 days and more, and offer a slight improvement only at 1-day forecast horizon. The better performance at 1-day horizon for the model using all high-frequency data can be explained by the forecast naturally relying on the most-recent (although contaminated) volatility estimates, rather than model (biased) parameters that determine the model's performance at longer horizons. Most importantly our proposed model is able to improve upon the forecasts based on daily data assuming a lower value of $\beta = 0.1$, which corresponds to utilizing only 10% of the high-frequency volatility measures, which performs better than the model with the lowest value of $\beta = 0.01$ and proves the importance of the high frequency measurement equation. The improvements are most pronounced at shorter forecast horizons of 1 to 10 days, although the improvements can also be seen at the longest time horizon of 22 days for the most recent samples. The value of $\beta = 0.1$ enforces a large probability of shutting down the information content of the high-frequency volatility measures, $\hat{y}_t^{(2)}$, but still allows for the model to learn from the high frequency data when volatility changes sharply during the financial crisis and the post-crisis period. Hence, the model can robustify against high-frequency based volatility measure imperfections and outliers, while still allowing to learn about latent volatility states that translates into the observed improvements in out-of-sample density forecasts.

5. Conclusions

We put forward a simple new approach to robust filtering of state-space models, motivated by the idea that the inclusion of only a small fraction of available highly precise measurements can still extract most of the attainable efficiency gains for filtering latent states, estimating model parameters, and producing out-of-sample forecasts. The new class of particle filters we develop aims to achieve a degree of robustness to outliers and model misspecification by purposely randomizing the subset of utilized highly precise but possibly misspecified or outlier contaminated data measurements, while treating the rest as if missing. The arising robustness-efficiency trade-off is controlled by varying the fraction of randomly utilized measurements or the incurred relative efficiency loss from such randomized utilization of the available measurements. As an empirical illustration, we consider popular state space models for inflation and equity returns with stochastic volatility and document favorable performance of our robust particle filter and density forecasts on both simulated and real data. More generally, our randomization approach makes it easy to robustly incorporate highly informative but possibly contaminated modern "big data" streams for improved state-space filtering and forecasting.

A. Full online derivation of posterior distribution

In this section, we show how to derive the distribution $P(\theta, x^t|\hat{Y}^t, \alpha, \beta)$, using equations (6), (8), and (9). First, we consider the full joint posterior of all unknown variables.

$$P(x^t, \theta, y_t^{(2)}, C_t, y_t^{(j)}|\hat{y}_t^{(2)}, Y^{t-1}, \alpha, \beta) = P(x^t, \theta|y_t^{(2)}, \hat{Y}^{t-1}, \alpha)P(y_t^{(2)}, y_t^{(j)}, C_t|\hat{y}_t^{(2)}, \hat{Y}^{t-1}, \alpha, \beta) \quad (21)$$

The joint distribution of x^t, θ only depends on $y_t^{(2)}$, so C_t and $y_t^{(j)}$ can be integrated out to simplify the derivations. We consider the joint distribution of variables controlled by β:

$$P(y_t^{(2)}, y_t^{(j)}, C_t|\hat{y}_t^{(2)}, \hat{Y}^{t-1}, \alpha, \beta) = P(y_t^{(2)}|\hat{y}_t^{(2)}, C_t, \hat{Y}^{t-1})P(y_t^{(j)}|\hat{y}_t^{(2)}, C_t, \hat{Y}^{t-1})P(C_t|\hat{y}_t^{(2)}, \hat{Y}^{t-1}) \quad (22)$$

Per our setup,

$$P(y_t^{(2)}|\hat{y}_t^{(2)}, C_t, Y^{t-1}) = \begin{cases} \delta_{\hat{y}_t^{(2)}}(y_t^{(2)}) \text{ if } C_t = 1 \\ P(y_t^{(2)}|\hat{Y}^{t-1}) \text{ if } C_t = 0 \end{cases} \quad (23)$$

$$P(y_t^{(j)}|\hat{y}_t^{(2)}, C_t, Y^{t-1}) = \begin{cases} P(y_t^{(j)}|\hat{Y}^{t-1}) \text{ if } C_t = 1 \\ \delta_{\hat{y}_t^{(2)}}(y_t^{(j)}) \text{ if } C_t = 0 \end{cases} \quad (24)$$

As before in equation (10), we define $\hat{\beta}_t := P(C_t = 1|\hat{y}_t^{(2)}, Y^{t-1}) = \frac{\beta_t f_2(\hat{y}_t^{(2)}|Y^{t-1})}{\beta_t f_2(\hat{y}_t^{(2)}|Y^{t-1}) + (1-\beta_t)}$.

$$
\begin{aligned}
P(y_t^{(2)}, y_t^{(j)}, C_t = 1 | \hat{y}_t^{(2)}, Y^{t-1}, \alpha, \beta) &= \hat{\beta}_t \delta_{\hat{y}_t^{(2)}}(y_t^{(2)}) P(y_t^{(j)} | Y^{t-1}) \\
P(y_t^{(2)}, y_t^{(j)}, C_t = 0 | \hat{y}_t^{(2)}, Y^{t-1}, \alpha, \beta) &= (1 - \hat{\beta}_t) P(y_t^{(2)} | Y^{t-1}) \delta_{\hat{y}_t^{(2)}}(y_t^{(j)}) \\
P(y_t^{(2)}, y_t^{(j)} | \hat{y}_t^{(2)}, Y^{t-1}, \alpha, \beta) &= \hat{\beta}_t \delta_{\hat{y}_t^{(2)}}(y_t^{(2)}) P(y_t^{(j)} | Y^{t-1}) + (1 - \hat{\beta}_t) P(y_t^{(2)} | Y^{t-1}) \delta_{\hat{y}_t^{(2)}}(y_t^{(j)}) \\
P(y_t^{(2)} | \hat{y}_t^{(2)}, Y^{t-1}, \alpha, \beta) &= \hat{\beta}_t \delta_{\hat{y}_t^{(2)}}(y_t^{(2)}) + (1 - \hat{\beta}_t) P(y_t^{(2)} | Y^{t-1})
\end{aligned}
\tag{25}
$$

This gives us the following marginal posterior:

$$
P(x^t, \theta, y_t^{(2)} | \hat{y}_t^{(2)}, Y^{t-1}, \alpha, \beta) = P(x^t, \theta | y_t^{(2)}, Y^{t-1}, \alpha)(\hat{\beta}_t \delta_{\hat{y}_t^{(2)}}(y_t^{(2)}) + (1 - \hat{\beta}_t) P(y_t^{(2)} | Y^{t-1})) \tag{26}
$$

Next, we need to derive the importance weights to sample $x^t, \theta | \hat{y}_t^{(2)}, Y^{t-1}$ from $x^{t-1}, \theta | Y^{t-1}$.

$$
P(x^t, \theta | y_t^{(2)}, Y^{t-1}, \alpha) = \frac{f_2(y_t^{(2)} | x^t, \theta, \hat{Y}^{t-1})}{f_2(y_t^{(2)} | \hat{Y}^{t-1})} P(x^t, \theta | Y^{t-1}, \alpha) \tag{27}
$$

Putting (27) inside of (28), we get:

$$
\begin{aligned}
P(x^t, \theta, y_t^{(2)} | \hat{y}_t^{(2)}, Y^{t-1}, \alpha, \beta) &= P(x^t, \theta | Y^{t-1}, \alpha) \frac{f_2(y_t^{(2)} | x^t, \theta, \hat{Y}^{t-1})}{f_2(y_t^{(2)} | \hat{Y}^{t-1})} (\hat{\beta}_t \delta_{\hat{y}_t^{(2)}}(y_t^{(2)}) + (1 - \hat{\beta}_t) f_2(y_t^{(2)} | \hat{Y}^{t-1})) \\
&= P(x^t, \theta | Y^{t-1}, \alpha) (\hat{\beta}_t \frac{f_2(y_t^{(2)} | x^t, \theta, \hat{Y}^{t-1})}{f_2(y_t^{(2)} | \hat{Y}^{t-1})} \delta_{\hat{y}_t^{(2)}}(y_t^{(2)}) + (1 - \hat{\beta}_t) f_2(y_t^{(2)} | x^t, \theta, \hat{Y}^{t-1}))
\end{aligned}
\tag{28}
$$

$$
\begin{aligned}
P(x^t,\theta|\hat{y}_t^{(2)},Y^{t-1},\alpha,\beta) &= P(x^t,\theta|Y^{t-1},\alpha)\int_{y_t^{(2)}}(\hat{\beta}_t\frac{f_2(y_t^{(2)}|x^t,\theta,\hat{Y}^{t-1})}{f_2(y_t^{(2)}|\hat{Y}^{t-1})}\delta_{\hat{y}_t^{(2)}}(y_t^{(2)})+(1-\hat{\beta}_t)f_2(y_t^{(2)}|x^t,\theta,\hat{Y}^{t-1}))dy_t^{(2)} \\
&= P(x^t,\theta|Y^{t-1},\alpha)(\hat{\beta}_t\frac{f_2(\hat{y}_t^{(2)}|x^t,\theta,\hat{Y}^{t-1})}{f_2(\hat{y}_t^{(2)}|\hat{Y}^{t-1})}+(1-\hat{\beta}_t))
\end{aligned}
\tag{29}
$$

Finally, equation (11) is found by a simple application of Bayes' rule via the first observation:

$$
\begin{aligned}
P(x^t,\theta|\hat{Y}^t,\alpha,\beta) &= P(x^t,\theta|y_t^{(1)},\hat{y}_t^{(2)},\hat{Y}^{t-1},\alpha,\beta) \\
&= \frac{f_1(y_t^{(1)}|x^t,\theta)}{f_1(y_t^{(1)}|\hat{y}_t^{(2)},\hat{Y}^{t-1})}P(x^t,\theta|\hat{y}_t^{(2)},Y^{t-1},\alpha,\beta) \\
&= \frac{f_1(y_t^{(1)}|x^t,\theta)}{f_1(y_t^{(1)}|\hat{y}_t^{(2)},\hat{Y}^{t-1})}(\hat{\beta}_t\frac{f_2(\hat{y}_t^{(2)}|x^t,\theta,\hat{Y}^{t-1})}{f_2(\hat{y}_t^{(2)}|\hat{Y}^{t-1})}+(1-\hat{\beta}_t))P(x^t,\theta|Y^{t-1},\alpha)
\end{aligned}
\tag{30}
$$

B. Proof of Theorem 1 and Theorem 2

We first derive some properties of $P(C^t|\beta)$. Let $\sum C^t$ represent the number of "successes" in the path of C^t.

$$
\begin{aligned}
P(C^t) &= \beta^{\sum C^t}(1-\beta)^{\sum C^t} \\
\frac{\delta}{\delta\beta}P(C^t) &= \beta^{\sum C^t-1}(1-\beta)^{t-\sum C^t-1}(\sum C^t-\beta t) \\
\sum_{C^t}\frac{\delta}{\delta\beta}P(C^t|\beta) &= 0 \\
\frac{\delta^2}{\delta\beta^2}P(C^t|\beta) &= \beta^{\sum C^t-2}(1-\beta)^{t-\sum C^t-2}(\beta^2(t-1)(t)+\beta(-2\sum C^t(t-1))+(\sum C^t)^2) \\
P(C^t)\frac{\delta^2}{\delta\beta^2}P(C^t|\beta)-(\frac{\delta}{\delta\beta}P(C^t|\beta))^2 &= \beta^{2\sum C^t-2}(1-\beta)^{2t-2\sum C^t-2}(\sum C^t(2\beta-1)-\beta^2T)
\end{aligned}
\tag{31}
$$

Since $\sum C^t \le t$ and $2\beta-1<\beta^2$, it follows that last term in (31) is not positive. Moreover,

we find that the equation $\Sigma C(2\beta-1)-\beta^2 T=0$ has no real roots, so for $\beta\in(0,1)$, the last term in (31) is strictly negative.

We arrive at the following expression for the derivative of the expectation in equation (20) with respect to β:

$$\begin{aligned}\frac{\delta}{\delta\beta}\mathbb{E}_{\hat{y}^{t+1}}(\log(P(\hat{y}_{t+1}|\hat{y}^t,M_\beta))|D) &= \int_{\hat{y}^{t+1}}\frac{\delta}{\delta\beta}\log\Big(\sum_{C^t}P(C^t|\beta)\int_{(x^t,\theta)}f_2(\hat{y}_{t+1}|x^t,\theta)\frac{f_2(\hat{y}_{\{i:C_i^t=1\}}|x^t,\theta)}{F_2(\hat{y}_{\{i:C_i^t=1\}})}dP(x^t,\theta)\Big)dP(\hat{y}^{t+1}|D)\\ &= \int_{\hat{y}^{t+1}}\frac{\sum_{C^t}\frac{\delta}{\delta\beta}P(C^t|\beta)\int_{(x^t,\theta)}f_2(\hat{y}_{t+1}|x^t,\theta)\frac{f_2(\hat{y}_{\{i:C_i^t=1\}}|x^t,\theta)}{F_2(\hat{y}_{\{i:C_i^t=1\}})}dP(x^t,\theta)}{\sum_{C^t}P(C^t|\beta)\int_{(x^t,\theta)}f_2(\hat{y}_{t+1}|x^t,\theta)\frac{f_2(\hat{y}_{\{i:C_i^t=1\}}|x^t,\theta)}{F_2(\hat{y}_{\{i:C_i^t=1\}})}dP(x^t,\theta)}dP(\hat{y}^{t+1}|D)\end{aligned} \tag{32}$$

It is important to note that part of the density of the DGP is the same as the denominator of the derivative of the expectation of the forecasting log-density.

$$P(\hat{y}^{t+1}|M_{\hat{\beta}})=\sum_{C^t}P(C^t|\hat{\beta})\int_{(x^t,\theta)}f_2(\hat{y}_{t+1}|x^t,\theta)\frac{f_2(\hat{y}_{\{i:C_i^t=1\}}|x^t,\theta)}{F_2(\hat{y}_{\{i:C_i^t=1\}})}dP(x^t,\theta)P(\hat{y}^t|M_{\hat{\beta}}) \tag{33}$$

Consequently, substituting into (32):

$$\begin{aligned}\frac{\delta}{\delta\beta}\mathbb{E}_{\hat{y}^{t+1}}(\log(P(\hat{y}_{t+1}|\hat{y}^t,M_\beta))|D)(\hat{\beta}) &= \int_{\hat{y}^{t+1}}\sum_{C^t}\frac{\delta}{\delta\hat{\beta}}P(C^t|\hat{\beta})\int_{(x^t,\theta)}f_2(\hat{y}_{t+1}|x^t,\theta)\frac{f_2(\hat{y}_{\{i:C_i^t=1\}}|x^t,\theta)}{F_2(\hat{y}_{\{i:C_i^t=1\}})}dP(x^t,\theta)dP(\hat{y}^t|M_{\hat{\beta}})\\ &= \sum_{C^t}\frac{\delta}{\delta\hat{\beta}}P(C^t|\hat{\beta})\ =0,\end{aligned} \tag{34}$$

which is our first result.

Our second result in the general case $D=M_{\hat{\beta},f_j}$ follows by continuity of the expectation of the log-forecasting density in equation (20) as a function of β on the compact interval $[0,1]$.

References

Amisano, G., and R. Giacomini, 2007, "Comparing density forecasts via weighted likelihood ratio tests," *Journal of Business & Economic Statistics*, 25(2), 177–190.

Andersen, T. G., L. Benzoni, and J. Lund, 2002, "An empirical investigation of continuous-time equity return models," *The Journal of Finance*, 57(3), 1239–1284.

Andersen, T. G., D. Dobrev, and E. Schaumburg, 2012, "Jump-robust volatility estimation using nearest neighbor truncation," *Journal of Econometrics*, 169(1), 75–93.

——— , 2014, "A robust neighborhood truncation approach to estimation of integrated quarticity," *Econometric Theory*, 30(1), 3–59.

Andrieu, C., A. Doucet, and R. Holenstein, 2010, "Particle markov chain monte carlo methods," *Journal of the Royal Statistical Society*, 72(3), 269–342.

Barndorff-Nielsen, O. E., and N. Shephard, 2002, "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2), 253–280.

——— , 2004, "Power and bipower variation with stochastic volatility and jumps," *Journal of financial econometrics*, 2(1), 1–37.

——— , 2005a, "How accurate is the asymptotic approximation to the distribution of realized variance," *Identification and inference for econometric models. A Festschrift in honour of TJ Rothenberg*, pp. 306–311.

——— , 2005b, "Variation, jumps, market frictions and high frequency data in financial econometrics," in *Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress*. Cambridge University Press.

Barndorff-Nielsen, O. E., N. Shephard, and M. Winkel, 2006, "Limit theorems for multipower variation in the presence of jumps," *Stochastic processes and their applications*, 116(5), 796–806.

Bergmeir, C., R. Hyndman, and J. Benítez, 2016, "Bagging exponential smoothing methods using STL decomposition and BoxCox transformation," *International Journal of Forecasting*, 32(2), 303 – 312.

Breiman, L., 1996, "Bagging Predictors," *Machine Learning*, 24(2), 123–140.

Calvet, L. E., V. Czellar, and E. Ronchetti, 2015, "Robust filtering," *Journal of the American Statistical Association*, 110(512), 1591–1606.

Chopin, N., P. E. Jacob, and O. Papaspiliopoulos, 2013, "SMC2: an efficient algorithm for sequential analysis of state space models," *Journal of the Royal Statistical Society*, 75(3), 397–426.

Creel, M., and D. Kristensen, 2015, "ABC of SV: Limited information likelihood inference in stochastic volatility jump-diffusion models," *Journal of Empirical Finance*, 31, 85–108.

Crevits, R., and C. Croux, 2017, "Robust estimation of linear state space models," .

Dobrev, D., and P. Szerszen, 2010, "The information content of high-frequency data for estimating equity return models and forecasting risk," *FEDS Working Paper*.

Durbin, J., and S. J. Koopman, 2000, "Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1), 3–56.

Eraker, B., M. Johannes, and N. Polson, 2003, "The impact of jumps in volatility and returns," *The Journal of Finance*, 58(3), 1269–1300.

Faust, J., and J. H. Wright, 2013, "Forecasting Inflation," Elsevier, vol. 2, chap. Chapter 1, pp. 2–56.

Gordon, N. J., D. J. Salmond, and A. F. Smith, 1993, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, pp. 107–113. IET.

Harvey, A., and A. Luati, 2014, "Filtering With Heavy Tails," *Journal of the American Statistical Association*, 109(507), 1112–1122.

Harvey, A., E. Ruiz, and N. Shephard, 1994, "Multivariate stochastic variance models," *The Review of Economic Studies*, 61(2), 247–264.

Inoue, A., and L. Kilian, ????, *Journal of the American Statistical Association*(482), 511–522.

Jacquier, E., N. G. Polson, and P. E. Rossi, 2002, "Bayesian analysis of stochastic volatility models," *Journal of Business & Economic Statistics*, 20(1), 69–87.

——— , 2004, "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors," *Journal of Econometrics*, 122(1), 185–212.

Jones, C. S., 2003, "The dynamics of stochastic volatility: evidence from underlying and options markets," *Journal of econometrics*, 116(1), 181–224.

Koopman, S. J., and M. Scharth, 2013, "The analysis of stochastic volatility in the presence of daily realized measures," *Journal of Financial Econometrics*, 11(1), 76–115.

Liu, J., and M. West, 2001, "Combined parameter and state estimation in simulation-based filtering," in *Sequential Monte Carlo methods in practice*. pp. 197–223.

Maìz, C., E. Molanes-Lòpez, J. Mìguez, and P. Djuriè, 2012, "A particle filtering scheme for processing time series corrupted by outliers," *IEEE Transactions on Signal Processing*, 60(9), 4611–4627.

Pitt, M. K., and N. Shephard, 1999, "Filtering via simulation: Auxiliary particle filters," *Journal of the American statistical association*, 94(446), 590–599.

Polson, N. G., and G. O. Roberts, 1994, "Bayes factors for discrete observations from diffusion processes," *Biometrika*, 81(1), 11–26.

Shephard, N., 1996, "Statistical aspects of ARCH and stochastic volatility," *Monographs on Statistics and Applied Probability*, 65, 1–68.

Stock, J. H., and M. W. Watson, 2007, "Why has US inflation become harder to forecast?," *Journal of Money, Credit and banking*, 39(s1), 3–33.

——— , 2016, "Core inflation and trend inflation," *Review of Economics and Statistics*, 98(4), 770–784.

Stoffer, D. S., and K. D. Wall, 1991, "Bootstrapping State-Space Models: Gaussian Maximum Likelihood Estimation and the Kalman Filter," *Journal of the American Statistical Association*, 86(416), 1024–1033.

Takahashi, M., Y. Omori, and T. Watanabe, 2009, "Estimating stochastic volatility models using daily returns and realized volatility simultaneously," *Computational Statistics & Data Analysis*, 53(6), 2404–2426.

Figure 2: Kalman Filter without Outliers: Filtered State Variance.

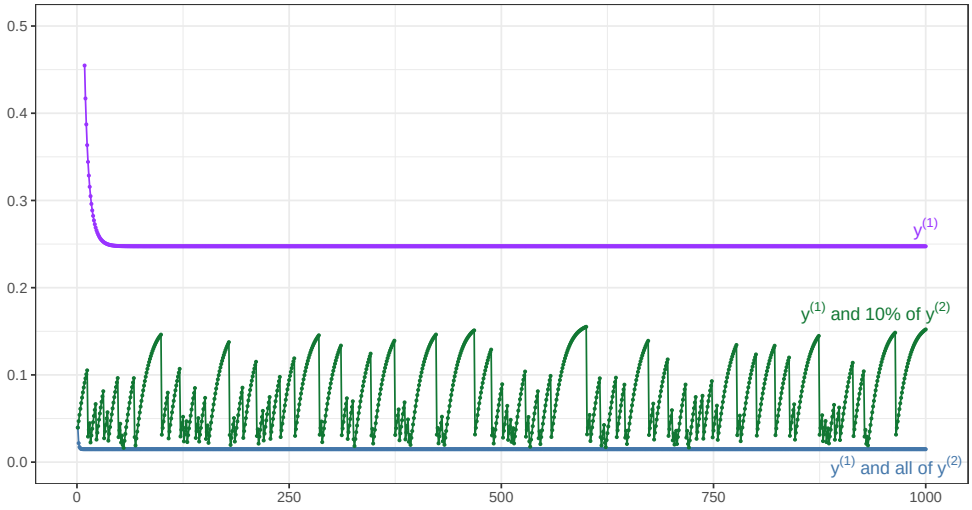

Figure 3: Kalman Filter with Outliers: Filtered States

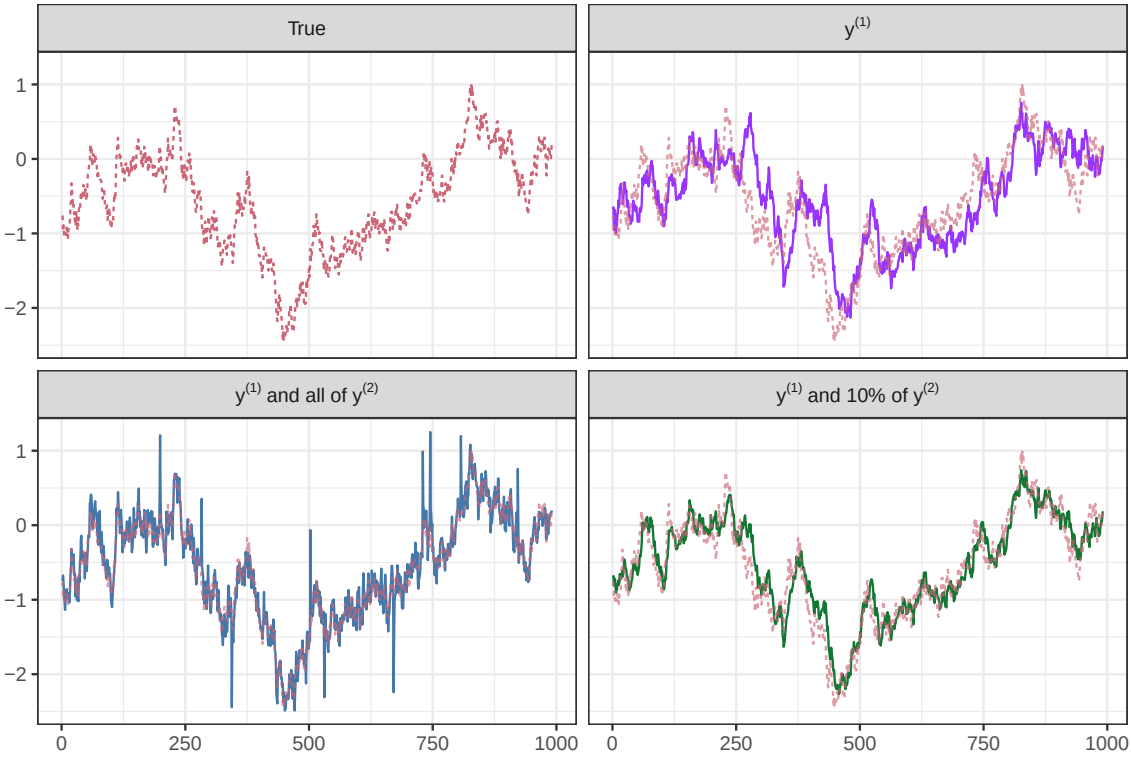

Figure 4: Posterior distributions of UC parameters with different values of β estimated on quarterly inflation data from 1960Q1 to 2015Q2. From lightest to darkest, the shaded regions correspond to the middle 50%, 90% and 98% of the distribution.

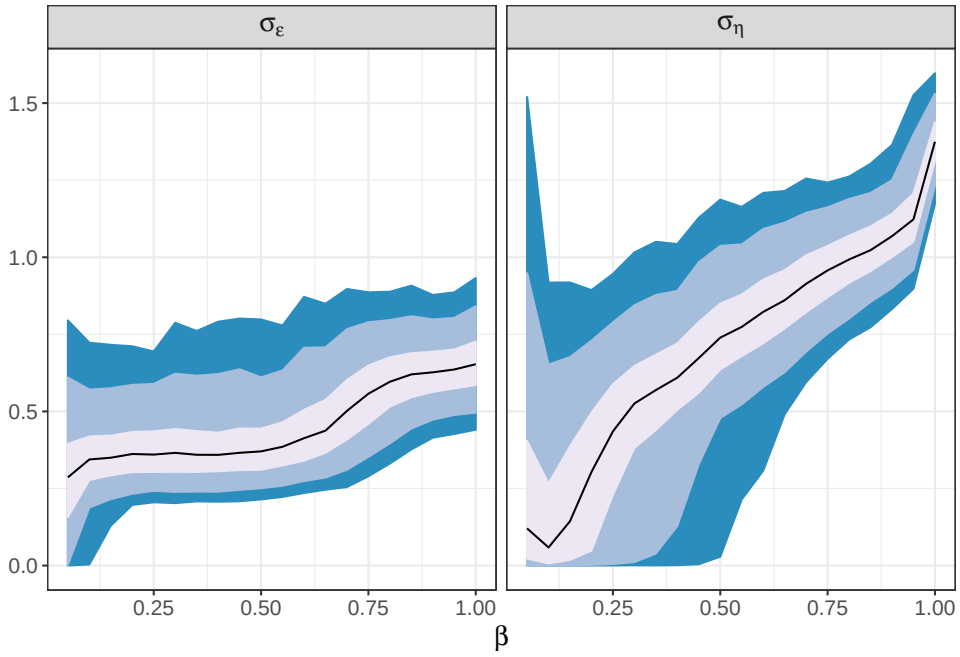

Figure 5: Mean log-density over different forecast horizons as a function of β estimated on quarterly inflation data from 1960Q1 to 2015Q2.

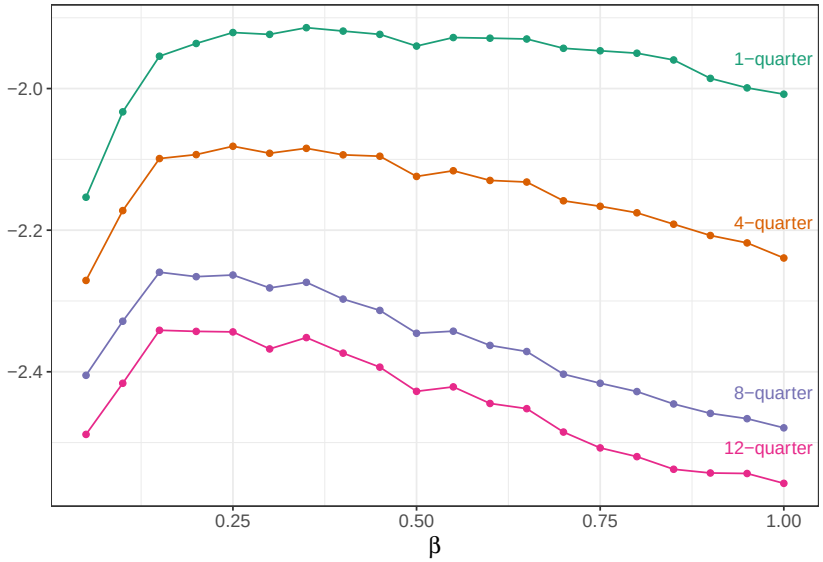

Figure 6: Optimal β from 1960Q1 to 2015Q2 as determined by maximizing the log-density of forecasts over 1-,4-,8- and 12-quarter horizons with all information available at that point.

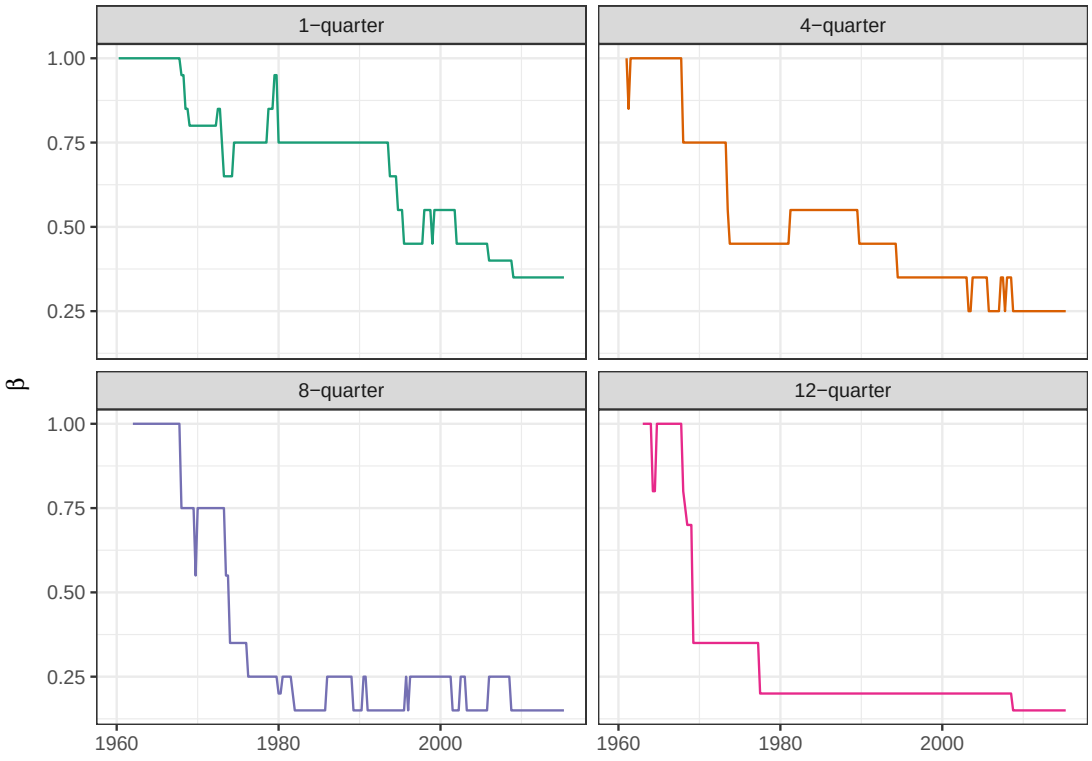

Figure 7: Mean forecasting log-density over different horizons (higher values are better). Latent states, parameters, and β are estimated from 1960Q1 to 2015Q2, and the mean-log forecast density is calculated with forecasts formulated in 1990Q1 onward. Each solid curve corresponds to the horizon used to select β as in Figure 6. The UC model with $\beta = 1$ is plotted for reference.

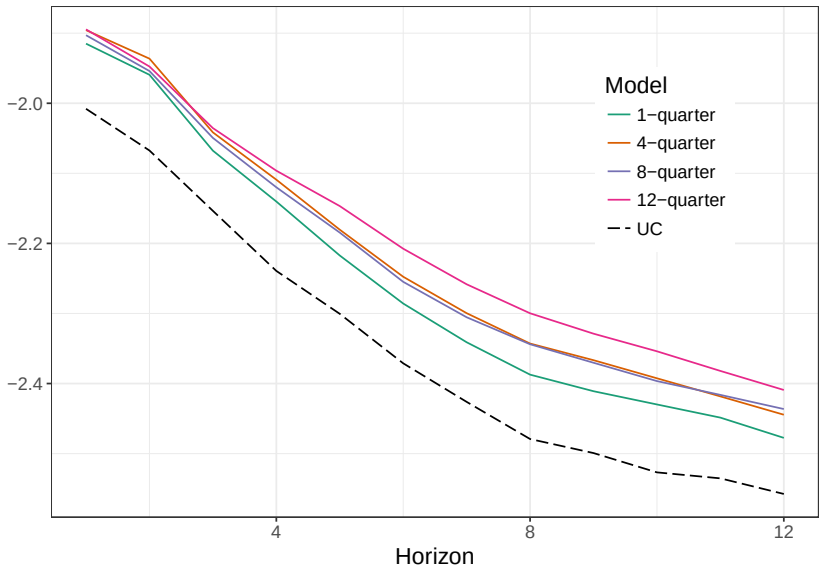

Figure 8: Mean Squared Forecasting Error (MSFE) over different horizons (lower values are better). Latent states, parameters, and β are estimated from 1960Q1 to 2015Q2, and the MSFE is calculated with forecasts formulated in 1990Q1 onward. Each solid curve corresponds to the horizon used to select β as in Figure 6. MSFEs from the UC model with $\beta = 1$ and the UCSVO model from Stock and Watson (2016) are plotted for reference.

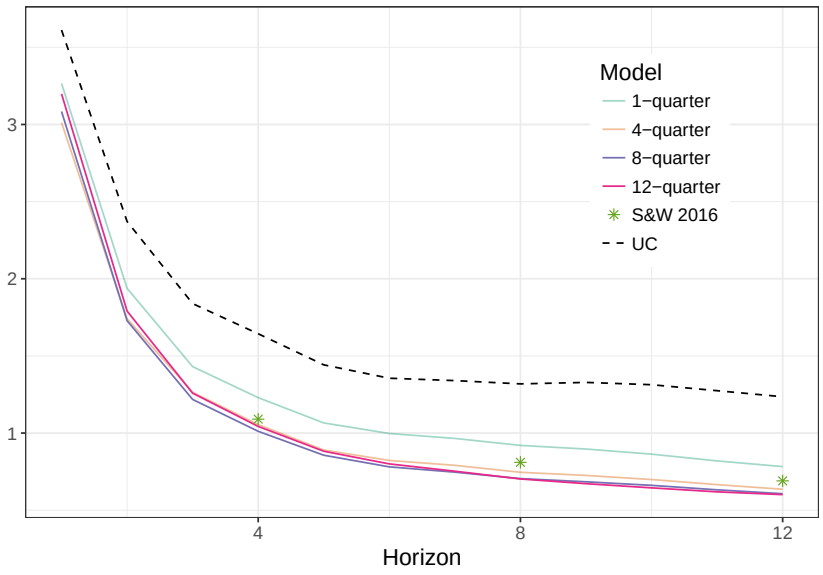

Figure 9: Filtered posterior quantiles of x_t, simulated with $\kappa_x = 0.0132$. From lightest to darkest, the shaded areas represent the range of the middle 50%, 90%, and 98% of the distribution.

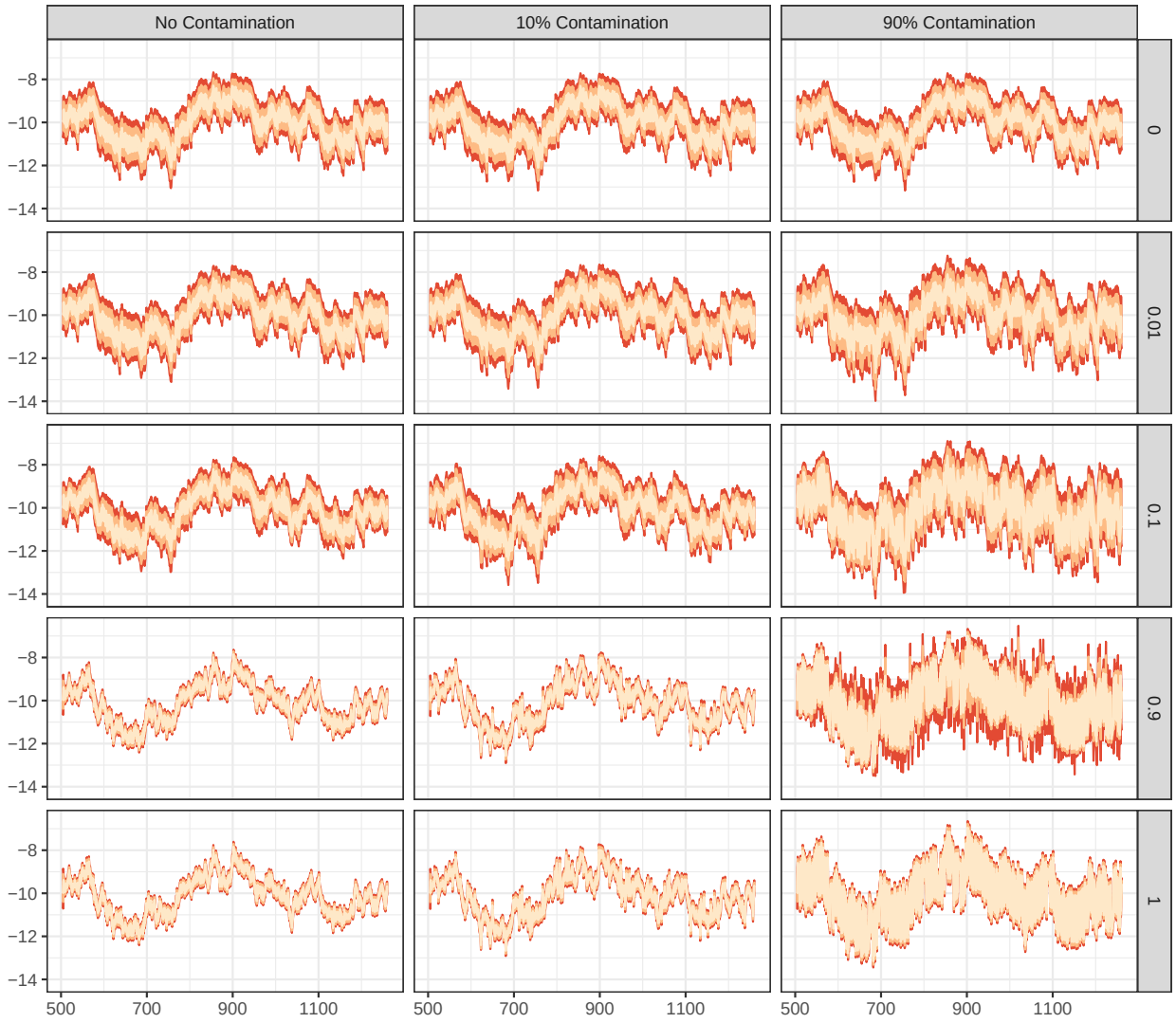

Figure 10: Filtered posterior mean of x_t, simulated with $\kappa_x = 0.0132$.

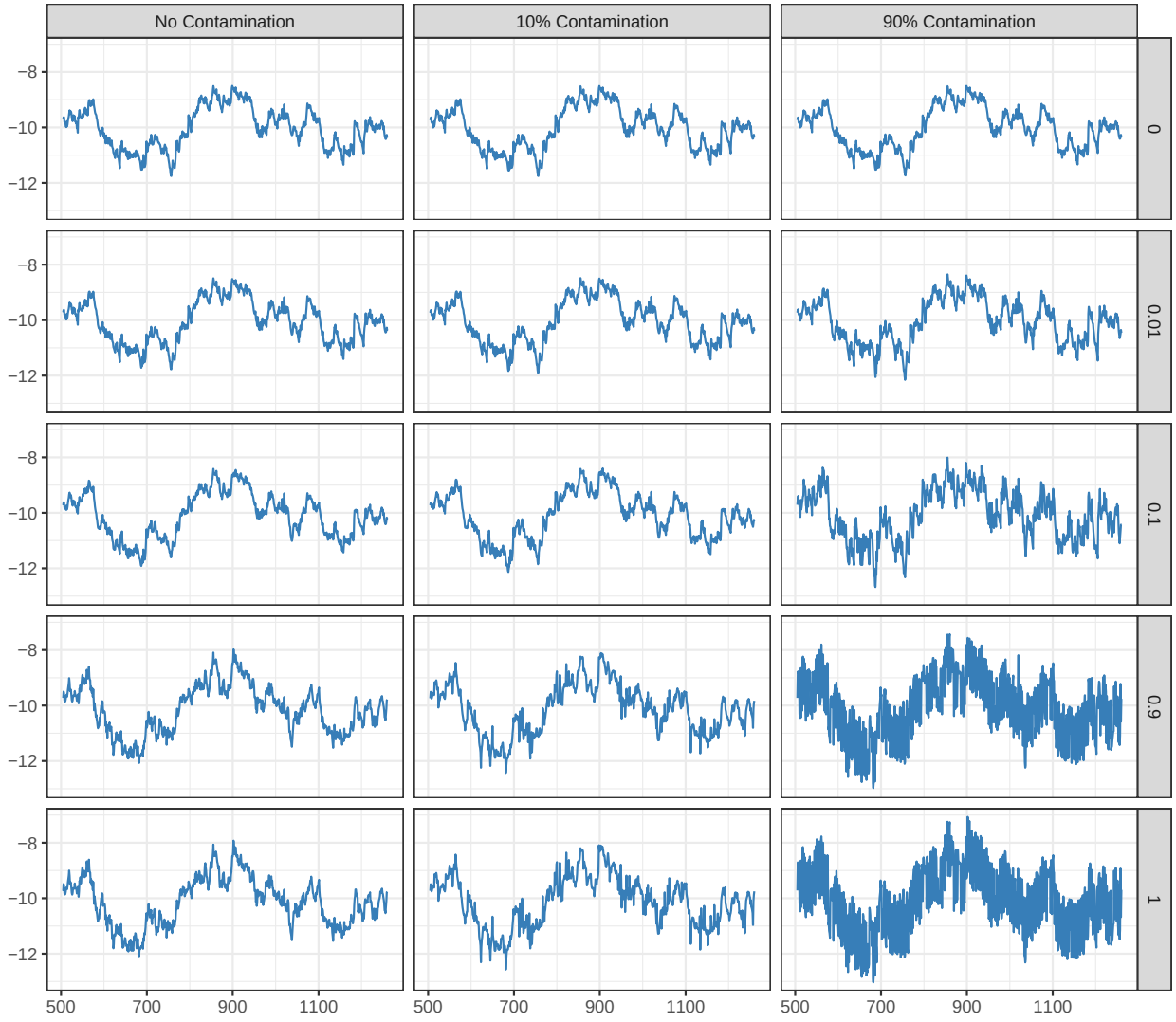

Figure 11: Filtered posterior variance of x_t, simulated with $\kappa_x = 0.0132$.

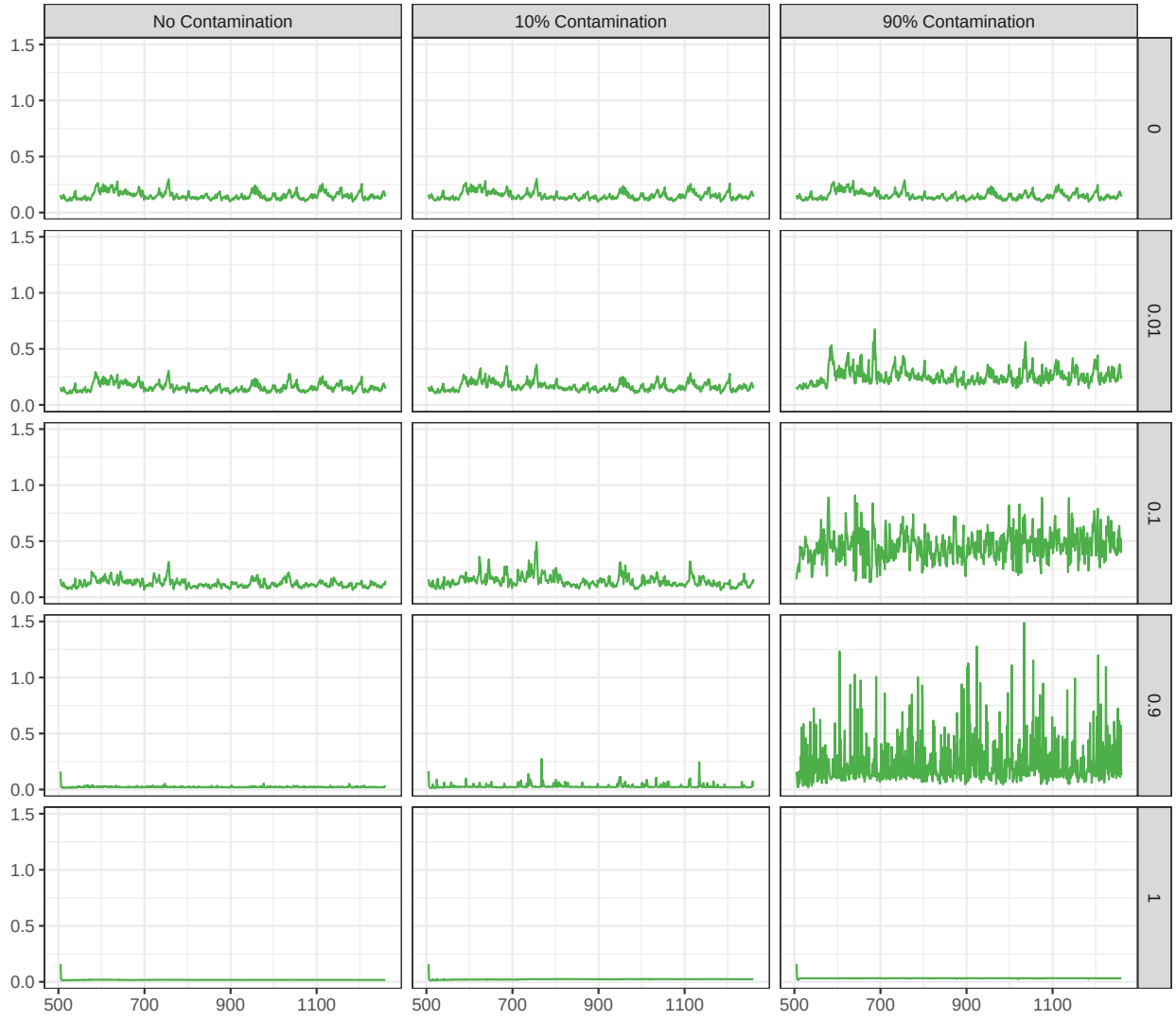

Figure 12: Posterior filtered distribution, mean, and variance of x_t from S&P-500 returns (1999-01-28 to 2004-11-11).

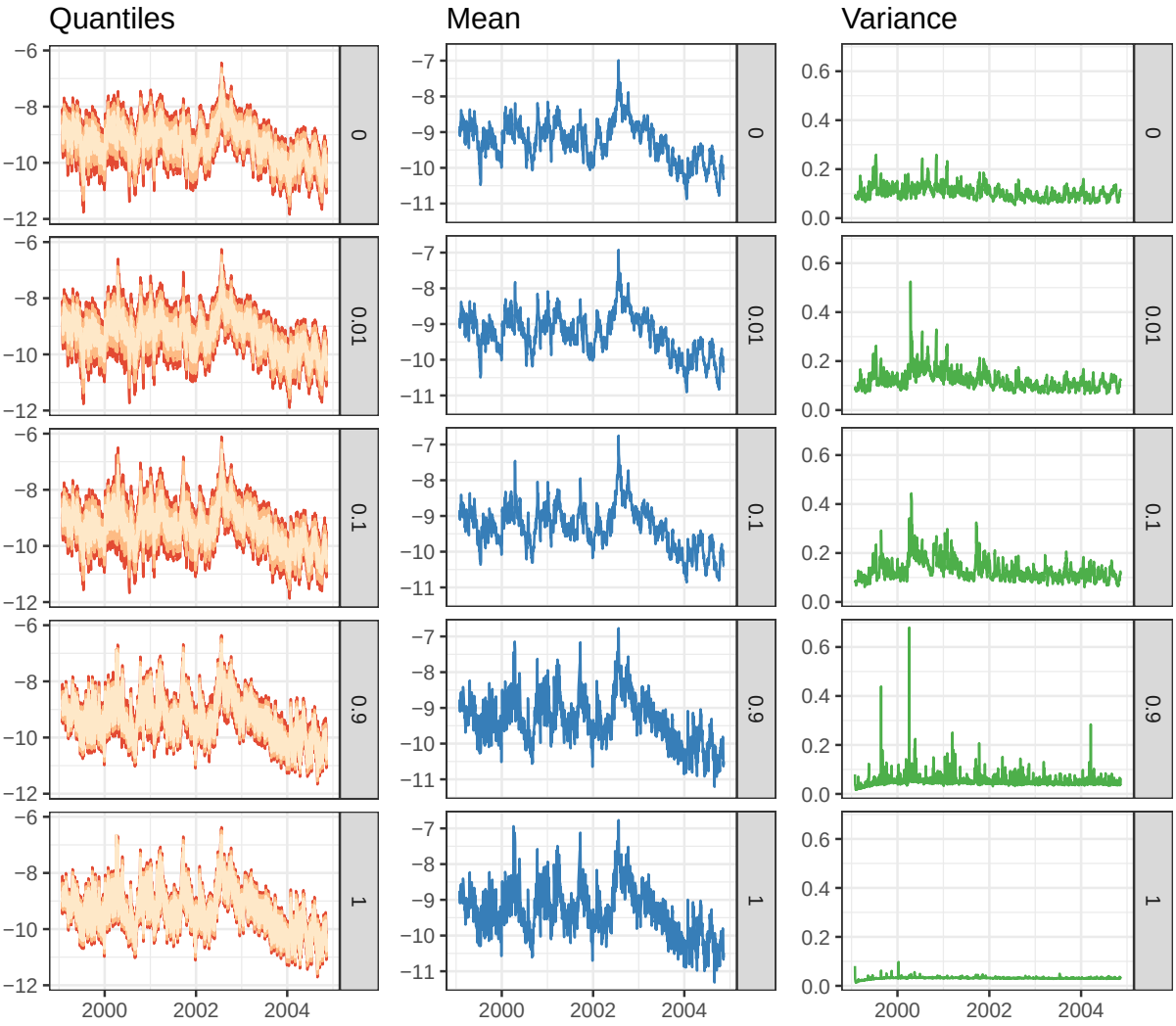

Figure 13: Posterior filtered distribution, mean, and variance of x_t from S&P-500 returns (2005-11-14 to 2011-06-07).

From lightest to darkest, the shaded areas represent the range of the middle 50%, 90%, and 98% of the distribution.

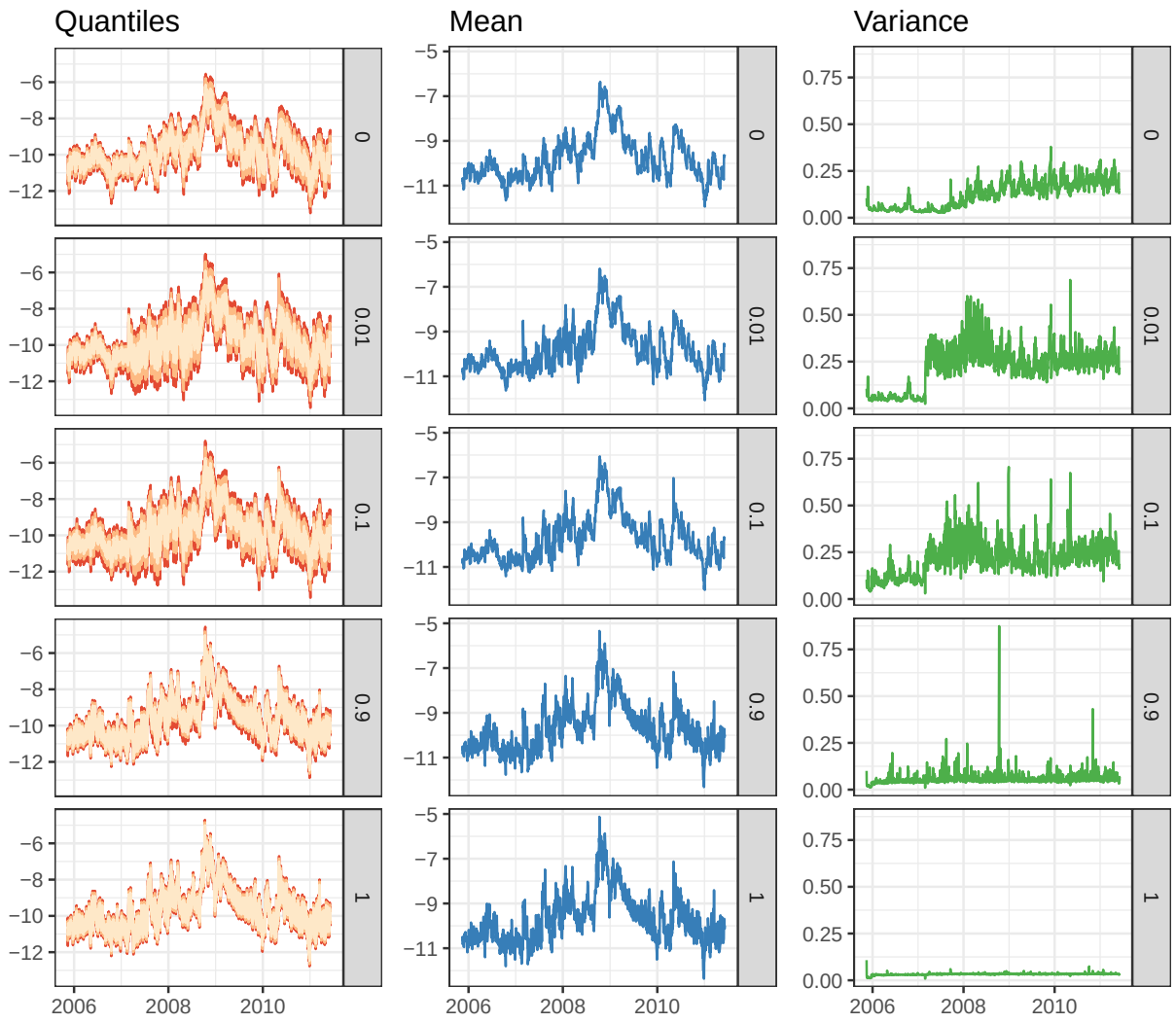

Figure 14: Posterior filtered distribution, mean, and variance of x_t from S&P-500 returns (2012-06-08 to 2017-12-29).

From lightest to darkest, the shaded areas represent the range of the middle 50%, 90%, and 98% of the distribution.

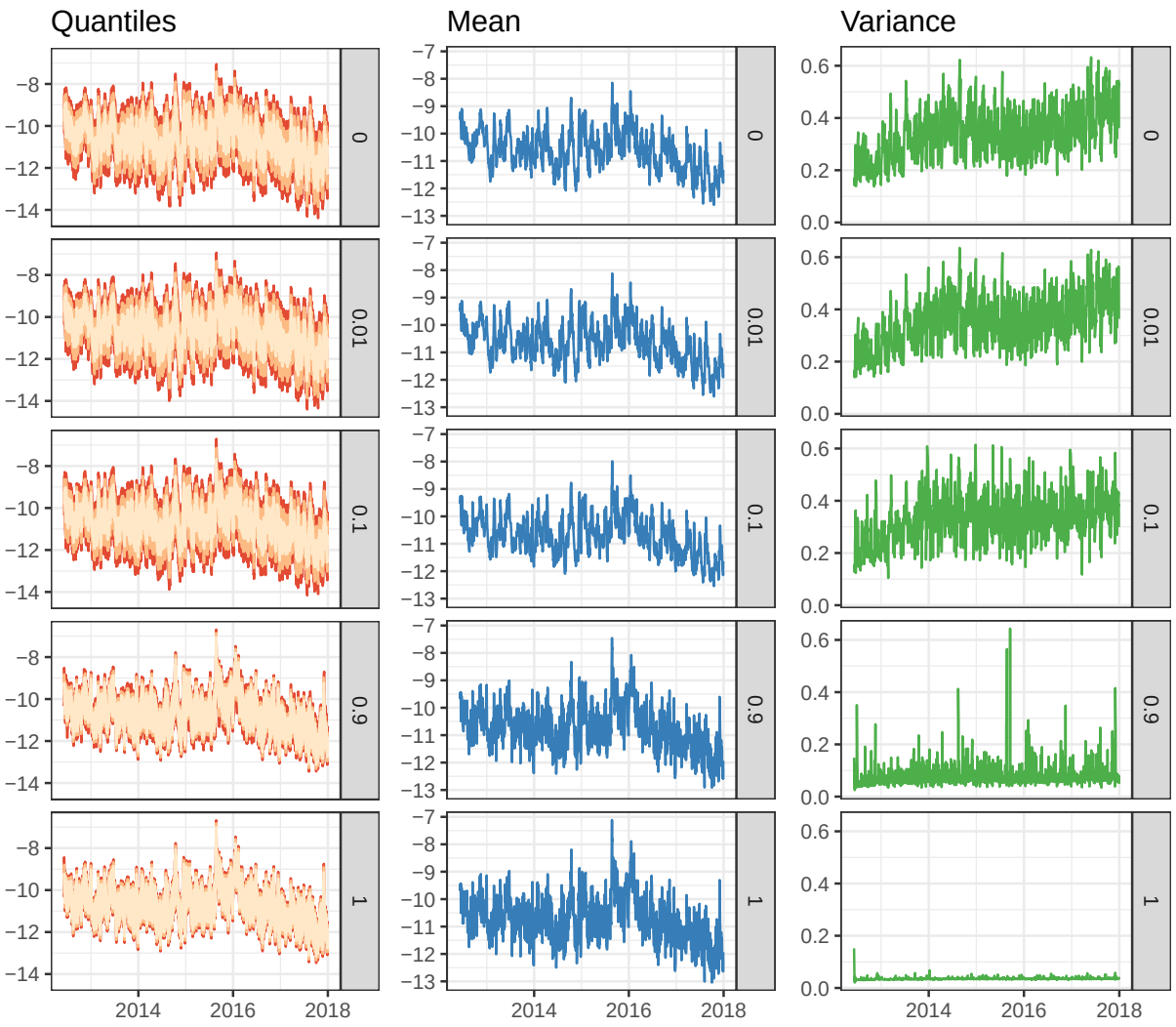

Figure 15: Posterior filtered mean of C_t from S&P-500 returns (1999-01-28 to 2004-11-11) for each β.

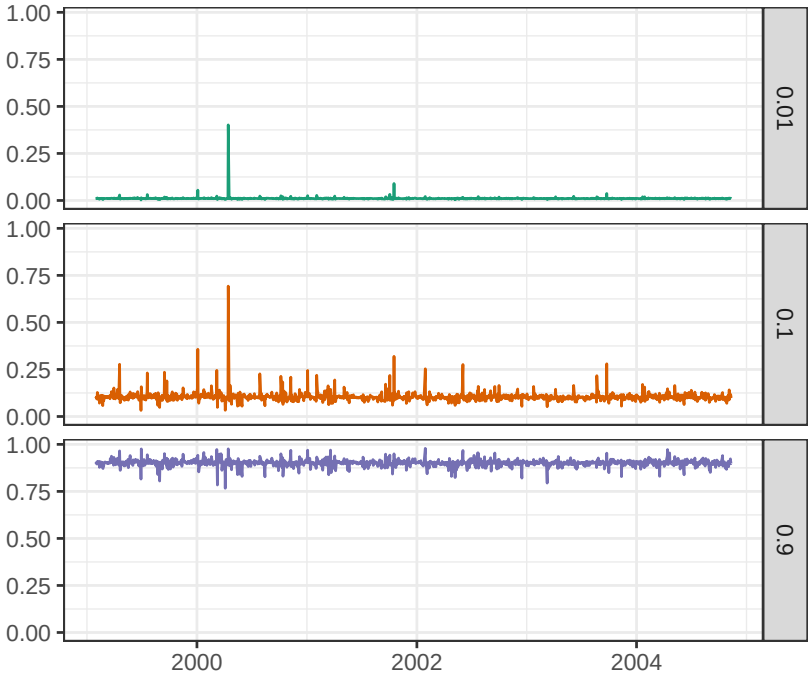

Figure 16: Posterior filtered mean of C_t from S&P-500 returns (2005-11-14 to 2011-06-07) for each β.

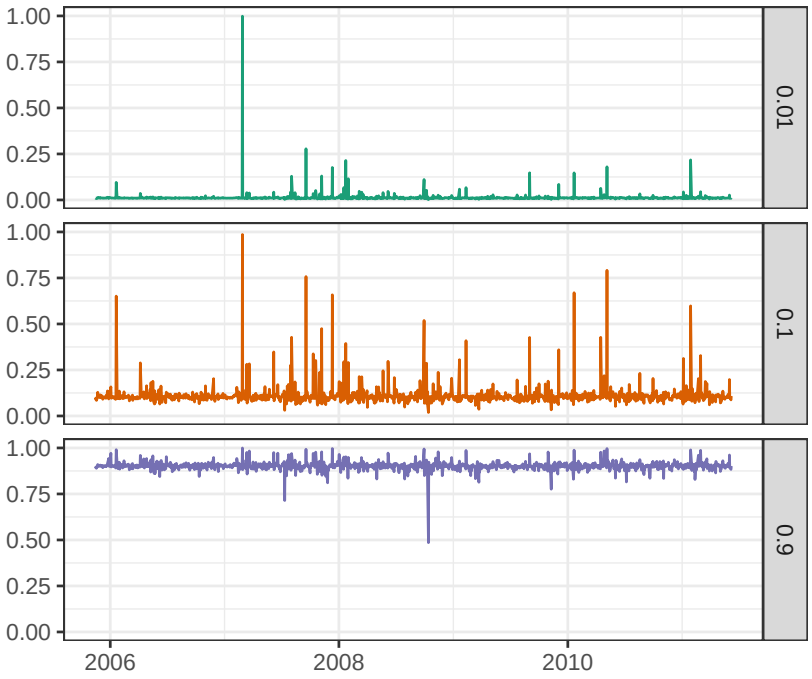

Figure 17: Posterior filtered mean of C_t from S&P-500 returns (2012-06-08 to 2017-12-29) for each β.

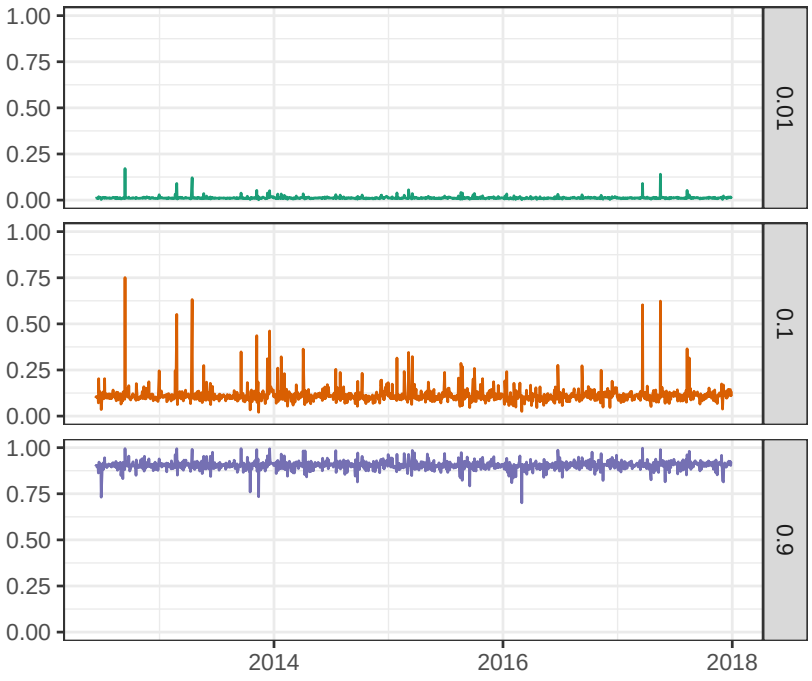

Table 1: Parameter RMSE for Kalman Filter with Missing Data Randomization.

	Parameter RMSE			
	Without Outliers		With Outliers	
β	κ	σ_η	κ	σ_η
0.00	**0.121**	**0.083**	**0.121**	**0.083**
0.01	0.073	0.039	0.076	0.044
0.02	0.038	0.023	0.044	0.032
0.05	0.008	0.012	0.018	0.032
0.10	**0.007**	**0.009**	**0.016**	**0.037**
0.20	0.006	0.008	0.020	0.049
0.30	0.006	0.007	0.023	0.056
0.40	0.006	0.007	0.025	0.063
0.50	0.006	0.007	0.027	0.069
0.60	0.006	0.007	0.029	0.075
0.70	0.006	0.006	0.032	0.080
0.80	0.006	0.006	0.035	0.084
0.90	0.006	0.006	0.037	0.090
1.00	**0.006**	**0.006**	**0.041**	**0.093**

Table 2: UC model parameters estimated with different values of β from 1960Q1 to 2015Q2.

β	Posterior Percentile	σ_ϵ	σ_η
0.15	50%	0.350	0.144
	2.5%	0.191	0.000
	97.5%	0.641	0.781
0.25	50%	0.360	0.435
	2.5%	0.222	0.002
	97.5%	0.645	0.870
0.90	50%	0.627	1.068
	2.5%	0.445	0.868
	97.5%	0.841	1.290
1.00	50%	0.653	1.375
	2.5%	0.467	1.205
	97.5%	0.891	1.564

Table 3: **Comparison of WLR forecasting performance for different** β. The UC model was run from 1960Q1 to 2005Q2, with the forecast density evaluated at 1-,4-,8-, and 12-quarter-ahead horizons starting in 1990Q1. The Weighted Likelihood Ratio (WLR) test from Amisano and Giacomini (2007) was used to compare the forecasts of different values of β versus $\beta = 1$.

β	h	$\widehat{\text{WLR}}$	$\hat{\sigma}_{\text{WLR}}$	t	P
0.15	1	0.2981	1.75	1.71	0.95
	4	0.2661	0.86	3.08	0.99
	8	0.2593	0.45	5.56	0.99
	12	0.2688	0.33	7.85	0.99
0.25	1	0.2748	1.52	1.82	0.96
	4	0.2507	0.73	3.39	0.99
	8	0.2514	0.47	5.15	0.99
	12	0.2688	0.37	6.93	0.99
0.90	1	0.0403	0.26	1.58	0.94
	4	0.0222	0.16	1.36	0.91
	8	0.0130	0.12	1.03	0.84
	12	0.0170	0.19	0.84	0.80

Table 4: Percentiles of posterior distribution of parameters with 0%, 10%, and 90% of days contaminated. Estimates are from data simulated for $T = 1260$ with $\kappa_x = 0.0132$, $\rho_x = -0.67$, and $\sigma_x = .16$.

		Contamination								
		0%			**10%**			**90%**		
β	Posterior Percentile	κ_x	ρ_x	σ_x	κ_x	ρ_x	σ_x	κ_x	ρ_x	σ_x
0.00	50%	0.01	-0.66	0.15	0.02	-0.64	0.15	0.01	-0.63	0.15
	2.5%	0.01	-0.77	0.11	0.01	-0.76	0.12	0.01	-0.76	0.12
	97.5%	0.03	-0.50	0.19	0.03	-0.49	0.19	0.03	-0.48	0.19
0.01	50%	0.02	-0.65	0.16	0.02	-0.64	0.16	0.03	-0.49	0.24
	2.5%	0.01	-0.76	0.12	0.01	-0.75	0.12	0.02	-0.64	0.19
	97.5%	0.03	-0.50	0.20	0.03	-0.51	0.20	0.05	-0.34	0.32
0.10	50%	0.01	-0.64	0.15	0.01	-0.65	0.15	0.07	-0.30	0.39
	2.5%	0.01	-0.74	0.12	0.01	-0.75	0.13	0.04	-0.44	0.32
	97.5%	0.02	-0.54	0.18	0.02	-0.53	0.19	0.11	-0.18	0.50
0.90	50%	0.02	-0.58	0.17	0.02	-0.57	0.17	0.39	-0.10	1.02
	2.5%	0.01	-0.64	0.16	0.01	-0.61	0.15	0.39	-0.16	0.99
	97.5%	0.02	-0.51	0.18	0.02	-0.45	0.18	0.47	-0.07	1.05
1.00	50%	0.02	-0.56	0.17	0.05	-0.36	0.27	0.48	-0.12	1.06
	2.5%	0.01	-0.62	0.16	0.03	-0.37	0.26	0.43	-0.17	1.01
	97.5%	0.03	-0.52	0.19	0.05	-0.31	0.29	0.52	-0.06	1.09

Table 5: **Forecasting Weighted Likelihood Ratio (WLR) ($\kappa_x = 0.0132$).** Forecast density was evaluated at 1,2,5,10, and 22-day horizons. The Weighted Likelihood Ratio (WLR) test from Amisano and Giacomini (2007) was used to compare the forecasts of different values of β versus $\beta = 0$.

		Contamination											
		0%				**10%**				**90%**			
β	h	$\widehat{\text{WLR}}$	$\hat{\sigma}_{\text{WLR}}$	t	P	$\widehat{\text{WLR}}$	$\hat{\sigma}_{\text{WLR}}$	t	P	$\widehat{\text{WLR}}$	$\hat{\sigma}_{\text{WLR}}$	t	P
0.01	1	0.005	0.03	3.69	0.99	0.004	0.04	2.97	0.99	-0.000	0.09	-0.11	0.45
	2	0.004	0.03	3.60	0.99	0.004	0.03	2.85	0.99	-0.000	0.09	-0.06	0.47
	5	0.004	0.03	3.63	0.99	0.003	0.03	2.53	0.99	-0.002	0.08	-0.78	0.21
	10	0.003	0.03	2.74	0.99	0.001	0.03	1.29	0.90	-0.004	0.08	-1.22	0.11
	22	0.001	0.03	0.97	0.83	-0.001	0.03	-0.57	0.28	-0.010	0.10	-2.79	0.00
0.10	1	0.020	0.13	4.15	0.99	0.019	0.14	3.78	0.99	-0.009	0.21	-1.16	0.12
	2	0.019	0.14	3.79	0.99	0.017	0.14	3.42	0.99	-0.008	0.22	-1.05	0.14
	5	0.018	0.13	3.79	0.99	0.016	0.14	3.27	0.99	-0.010	0.19	-1.42	0.07
	10	0.013	0.12	2.96	0.99	0.010	0.12	2.40	0.99	-0.009	0.17	-1.47	0.07
	22	0.010	0.13	2.13	0.98	0.007	0.13	1.49	0.93	-0.021	0.19	-2.96	0.00
0.90	1	0.032	0.26	3.40	0.99	0.034	0.27	3.50	0.99	-0.043	0.28	-4.20	0.00
	2	0.027	0.27	2.72	0.99	0.027	0.27	2.76	0.99	-0.064	0.31	-5.69	0.00
	5	0.029	0.24	3.31	0.99	0.028	0.25	3.13	0.99	-0.085	0.38	-6.19	0.00
	10	0.014	0.20	1.92	0.97	0.013	0.22	1.66	0.95	-0.078	0.44	-4.93	0.00
	22	0.014	0.18	2.08	0.98	0.012	0.19	1.77	0.96	-0.060	0.39	-4.22	0.00
1.00	1	0.033	0.26	3.42	0.99	0.036	0.27	3.67	0.99	-0.039	0.30	-3.56	0.00
	2	0.027	0.27	2.71	0.99	0.027	0.26	2.88	0.99	-0.068	0.32	-5.77	0.00
	5	0.029	0.24	3.26	0.99	0.029	0.23	3.46	0.99	-0.084	0.40	-5.75	0.00
	10	0.014	0.20	1.84	0.96	0.011	0.20	1.46	0.92	-0.078	0.44	-4.89	0.00
	22	0.013	0.18	2.02	0.97	-0.001	0.19	-0.18	0.43	-0.058	0.40	-4.02	0.00

Table 6: Percentiles of posterior distribution of parameters estimated for three samples: 1998-01-02 to 2004-11-11, 2004-11-12 to 2011-06-07, and 2011-06-08 to 2017-12-29.

		Sample Start Date								
		1998-01-02			**2004-11-12**			**2011-06-08**		
β	Posterior Percentile	κ_x	ρ_x	σ_x	κ_x	ρ_x	σ_x	κ_x	ρ_x	σ_x
0.00	50%	0.01	-0.84	0.13	0.02	-0.50	0.18	0.08	-0.37	0.38
	2.5%	0.01	-0.90	0.11	0.01	-0.61	0.15	0.05	-0.46	0.32
	97.5%	0.02	-0.75	0.15	0.03	-0.37	0.22	0.10	-0.27	0.45
0.01	50%	0.01	-0.82	0.13	0.03	-0.38	0.23	0.08	-0.34	0.39
	2.5%	0.01	-0.87	0.11	0.02	-0.48	0.19	0.06	-0.44	0.33
	97.5%	0.02	-0.71	0.16	0.03	-0.25	0.28	0.11	-0.24	0.46
0.10	50%	0.01	-0.71	0.13	0.03	-0.30	0.23	0.07	-0.29	0.37
	2.5%	0.01	-0.80	0.11	0.02	-0.39	0.18	0.05	-0.38	0.32
	97.5%	0.02	-0.59	0.16	0.03	-0.15	0.27	0.09	-0.20	0.41
0.90	50%	0.09	-0.29	0.29	0.07	-0.10	0.38	0.16	-0.11	0.50
	2.5%	0.07	-0.33	0.28	0.05	-0.25	0.36	0.13	-0.15	0.48
	97.5%	0.11	-0.25	0.32	0.08	-0.03	0.40	0.19	-0.06	0.53
1.00	50%	0.13	-0.22	0.36	0.08	-0.12	0.42	0.20	-0.06	0.55
	2.5%	0.11	-0.28	0.34	0.07	-0.12	0.41	0.17	-0.11	0.53
	97.5%	0.15	-0.21	0.38	0.08	-0.11	0.42	0.21	-0.04	0.57

Table 7: **Forecasting WLR results for SP-500 returns**. The time period of the forecasts were 1999-01-28 to 2004-11-11, 2005-11-14 to 2011-06-07, and 2012-06-08 to 2017-12-29. A minimum of 1 year of data was used for forecasting, with the estimation of parameters beginning at 1998-01-02, 2004-11-12, and 2011-06-08, respectively. Forecast density was evaluated at 1,2,5,10, and 22-day horizons. The Weighted Likelihood Ratio (WLR) test from Amisano and Giacomini (2007) was used to compare the forecasts of different values of β versus $\beta = 0$.

		Sample Start Date											
		1999-01-28				**2005-11-14**				**2012-06-08**			
β	Horizon	$\widehat{\text{WLR}}$	$\hat{\sigma}_{\text{WLR}}$	t	$P(t>0)$	$\widehat{\text{WLR}}$	$\hat{\sigma}_{\text{WLR}}$	t	$P(t>0)$	$\widehat{\text{WLR}}$	$\hat{\sigma}_{\text{WLR}}$	t	$P(t>0)$
0.01	1	0.0009	0.03	1.20	0.88	0.0019	0.15	0.47	0.68	0.0025	0.02	4.51	0.99
	2	0.0009	0.03	1.33	0.90	0.0007	0.16	0.16	0.56	0.0021	0.02	3.85	0.99
	5	0.0004	0.03	0.52	0.69	-0.0053	0.21	-0.95	0.17	0.0021	0.02	4.21	0.99
	10	0.0004	0.03	0.52	0.70	-0.0081	0.28	-1.09	0.13	0.0016	0.02	3.54	0.99
	22	-0.0005	0.02	-0.87	0.19	-0.0007	0.29	-0.08	0.46	0.0013	0.02	2.95	0.99
0.10	1	0.0054	0.10	2.11	0.98	0.0202	0.19	3.90	0.99	0.0131	0.09	5.46	0.99
	2	0.0046	0.09	1.95	0.97	0.0180	0.18	3.63	0.99	0.0107	0.08	4.70	0.99
	5	0.0034	0.08	1.62	0.94	0.0154	0.19	3.00	0.99	0.0113	0.07	5.81	0.99
	10	0.0027	0.08	1.25	0.89	0.0135	0.24	2.06	0.98	0.0092	0.07	5.23	0.99
	22	-0.0032	0.10	-1.21	0.11	0.0191	0.32	2.23	0.98	0.0070	0.06	4.31	0.99
0.90	1	0.0063	0.21	1.12	0.86	0.0323	0.30	4.02	0.99	0.0246	0.25	3.69	0.99
	2	-0.0031	0.18	-0.65	0.25	0.0140	0.29	1.81	0.96	0.0089	0.22	1.50	0.93
	5	-0.0123	0.20	-2.30	0.01	0.0038	0.27	0.53	0.70	0.0062	0.19	1.24	0.89
	10	-0.0270	0.31	-3.23	0.00	-0.0118	0.37	-1.18	0.11	-0.0000	0.19	-0.00	0.49
	22	-0.0298	0.32	-3.45	0.00	-0.0158	0.62	-0.94	0.17	0.0028	0.14	0.73	0.76
1.00	1	0.0047	0.22	0.81	0.79	0.0314	0.30	3.90	0.99	0.0235	0.26	3.35	0.99
	2	-0.0063	0.19	-1.26	0.10	0.0080	0.29	1.04	0.85	0.0067	0.23	1.11	0.86
	5	-0.0197	0.24	-3.07	0.00	-0.0064	0.33	-0.73	0.23	0.0014	0.20	0.25	0.59
	10	-0.0346	0.36	-3.58	0.00	-0.0248	0.48	-1.91	0.02	-0.0041	0.21	-0.73	0.23
	22	-0.0325	0.34	-3.57	0.00	-0.0260	0.78	-1.23	0.10	0.0023	0.15	0.56	0.71