Missing Data Substitution for Enhanced Robust Filtering and Forecasting in Linear State-Space Models

Dobrislav Dobrev* Paweł J. Szerszeń*

January 20, 2025

Abstract

Replacing faulty measurements with missing values can suppress outlier-induced distortions in state-space inference. We therefore put forward two complementary methods for enhanced outlier-robust filtering and forecasting: supervised missing data substitution (MD) upon exceeding a Huber threshold, and unsupervised missing data substitution via exogenous randomization (RMDX).

Our supervised method, MD, is designed to improve performance of existing Huber-based linear filters known to lose optimality when outliers of the same sign are clustered in time rather than arriving independently. The unsupervised method, RMDX, further aims to suppress smaller outliers whose size may fall below the Huber detection threshold. To this end, RMDX averages filtered or forecasted targets based on measurement series with randomly induced subsets of missing data at an exogenously set randomization rate. This gives rise to regularization and bias-variance trade-off as a function of the missing data randomization rate, which can be set optimally using standard cross-validation techniques.

We validate through Monte Carlo simulations that both methods for missing data substitution can significantly improve robust filtering, especially when combined together. As further empirical validation, we document consistently attractive performance in linear models for forecasting inflation trends prone to clustering of measurement outliers.

Keywords: Kalman filter, outliers, Huberization, missing data, randomization.

JEL classification: C15; C22; C53; E37.

*Board of Governors of the Federal Reserve System, 20th St. and Constitution Ave. NW, Washington, DC 20551. Emails: dobrislav.p.dobrev@frb.gov and pawel.j.szerszen@frb.gov. 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

In many applications involving state-space models the optimality of existing robust filters may not hold in practice when imposed assumptions on the structure of measurement errors are violated. Thus, even if state-space models are an important workhorse in many natural and social sciences, the filtering and forecasting performance of available methods can still be improved upon in the presence of incorrectly specified measurement outliers. This is the case both for robust filtering approaches based on heavy-tailed distributional assumptions as in Durbin and Koopman (2000) or Harvey and Luati (2014) and for more recent alternatives adding Huber-based thresholding or outlier detection as in Calvet et al. (2015), Crevits and Croux (2018), Maìz et al. (2012), among others.

In this paper, we build on the concept that replacing faulty measurements with missing data reduces filtering distortions due to outliers. We thus put forward two complementary methods for enhanced robust filtering and forecasting: supervised missing data substitution (MD) upon exceeding a Huber threshold, and unsupervised missing data substitution via exogenous randomization (RMDX).

Our supervised method for missing data substitution, MD, is specifically designed to improve the performance of existing Huber-based linear filters whose optimality gets violated when outliers of the same sign are clustered in time instead of arriving independently. More specifically, we formulate the MD-RobKF filter as an enhancement of the RobKF filter of (Calvet et al., 2015) based on supervised missing data substitution in lieu of truncation upon exceeding a Huber threshold. MD-RobKF performs similarly to RobKF when the latter is optimal in the root-mean square error (RMSE) sense but MD-RobKF improves on RobKF when its optimality condition is not satisfied, e.g. in case outliers of the same sign are clustered in time. This is because missing data substitution in MD-RobKF eliminates outliers instead of truncating them as done in the RobKF filter. In doing so, our MD-RobKF filter

naturally suppresses the accumulation of filtering errors in the presence of patches of outliers of the same sign (or clusters of highly correlated outliers), thereby improving performance relative to the RobKF filter, which can accumulate consecutive filtering error terms despite trimming them in size to the Huber truncation threshold.

Our unsupervised method for missing data substitution via exogenous randomization, RMDX, is designed to further suppress smaller outliers whose size might fall below the Huber detection threshold. To accomplish this, given a filter F, we formulate the enhanced filter RMDX-F with exogenously randomized missing data substitution by taking the average of filtered or forecasted targets based on the filter F for measurement series with randomly induced subsets of missing values at an exogenously set randomization rate. This gives rise to a bias-variance trade-off controlled solely by the missing data randomization rate. As a result, the randomization rate plays the role of a regularization parameter and can be optimally set using standard cross-validation techniques under RMSE or other loss functions of interest.

From this standpoint, RMDX can improve the performance of any filter in the presence of outliers violating its optimality conditions. We thus consider the following three RMDX-enhanced filters:

1. RMDX-KF enhancing the standard Kalman filter KF;
2. RMDX-RobKF enhancing the robust RobKF filter;
3. RMDX-MD-RobKF enhancing our supervised MD-RobKF filter.

The last one of these should be expected to outperform the rest as it combines supervised missing data substitution in MD-RobKF to suppress outliers above the Huber threshold and unsupervised missing data substitution in RMDX to further suppress outliers below the Huber threshold.

Conceptually, our RMDX method can be viewed as a state space time-series extension of bootstrap aggregation (bagging), originally developed by Breiman (1996). As a key

distinction from bagging, RMDX preserves time-series dependence by retaining the original time index of each observation and randomly drawing only induced missing values in each re-sampled measurement series. This connection to bagging gives some further insight into the sources of efficiency gains offered by RMDX.

At a more intuitive level, RMDX exploits the fact that a significant portion of the information needed for filtering latent states, estimating model parameters, and generating out-of-sample forecasts can often be extracted from a relatively small subset of the available measurements. This is especially the case when the latent process is highly persistent, which limits information loss when replacing subsets of data points with missing values. Thus, randomizing over the induced missing data points can achieve robustness to outliers and model misspecification without much efficiency loss.

On the theory side, we formally derive the arising bias-variance trade-off in RMDX as a function of the missing data randomization rate. Our key result is that the bias term decreases, while the variance term increases as the proportion of measurements not substituted by missing values shrinks. This establishes the existence of an optimal randomization rate for filters violating RMSE-optimality in the presence of outliers and justifies grid search for the optimal randomization rate via standard cross-validation techniques.

To validate these findings, we conduct Monte Carlo simulations demonstrating that both our supervised and unsupervised methods for missing data substitution can offer substantial improvements in robust filtering, particularly in the case where outliers arrive in clusters rather than independently. In the case of *iid* outliers - when optimality of the RobKF filter holds - we find that our enhanced MD-RobKF filter does not outperform RobKF by a significant margin as should be expected. By contrast, in the case of clustered outliers of the same sign - when RobKF optimality is violated - we find that MD-RobKF significantly outperforms the RobKF filter. Similarly, we document that the RMDX-enhanced

filters do not outperform significantly RobKF when outliers arrive independently. However, when outliers of the same sign arrive in patches, our combined RMDX-MD-RobKF filter significantly outperforms the RobKF and RMDX-RobKF as well as the KF and RMDX-KF filters for a wide range of outlier sizes and missing data randomization rates, as dictated by our theory results for the arising bias-variance trade-off. The efficiency gains for the combined RMDX-MD-RobKF filter relative to MD-RobKF are most pronounced in the presence of outliers falling below the Huber truncation threshold that remain undetected by MD-RobKF but are being suppressed at least to some extent through randomized missing data substitution in the combined RMDX-MD-RobKF filter.

For further empirical validation on real data, we consider state-space models for extracting inflation trends and document favorable performance of our supervised and unsupervised missing data substitution enhancements of robust filters and the resulting out-of-sample forecasts. We consider three alternative state-space models: the standard unobserved components (UC) model with an inflation trend following a random walk, the UC model with autoregressive (AR) inflation trend, and the UC model with AR inflation trend having a mean fixed to the long-run inflation target of 2% (ARMF). 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 clustered outliers in inflation measurements. For this reason, the considered setting can serve as a natural real-world testing platform to evaluate the effectiveness of the proposed methods for improved robust filtering.

We demonstrate that our RMDX-MD-RobKF filter employing both supervised and unsupervised missing data substitution meaningfully improves out-of-sample performance for all considered models and is preferred across all considered forecast horizons, with the longest horizons experiencing the largest performance gains. We find strongest support for the ARMF version of the UC model, where we fix the inflation mean to the long-run inflation target of 2% and therefore reduce the impact of parameter uncertainty. We further document

that our RMDX-MD-RobKF filter performs especially well for the less parsimonious AR and ARMF specification for inflation dynamics where overfitting outliers can lead to more severe distortions. The predictive accuracy achieved using our RMDX-MD-RobKF filter for these models compares well also to that of the well-established UCSVO model of Stock and Watson (2016) at short forecast horizons and improves even further when forecasting inflation trends over longer horizons. We conclude that the proposed methods for missing data substitution to enhance outlier-robust filtering and forecasting are easy to implement and most effective when combined together.

2. Robust Filtering in The Presence of Outliers

In this section we lay down our missing data substitution framework for robust filtering as an enhancement to the Huberization approach to robust filtering by Calvet et al. (2015). We propose two complementary methods: supervised missing data substitution (MD) upon exceeding a Huber threshold, and unsupervised missing data substitution via exogenous randomization (RMDX). As a key theory result, we establish that the exogenously set missing data randomization rate in RMDX acts as a regularization parameter controlling the arising bias-variance trade-off. This implies the existence of an optimal randomization rate for filters violating RMSE-optimality in the presence of outliers and justifies grid search for the optimal randomization rate via standard cross-validation techniques.

2.1 Problem formulation

We consider a Gaussian state-space model with state density $g_\theta(x_t|x_{t-1},\theta)$ and observation density $f_\theta(y_t|x_t,\theta)$. The available measurements are contaminated by additive outliers (AO) as follows:

$$y_t = y_t^\star + \eta\, u_t \ , \tag{1}$$

where $y_t^\star \sim f_\theta(\cdot|x_t, \theta)$ follows the true observation density, while the AO term is characterized by the disturbance u_t scaled by magnitude $\eta \in \mathbb{R}$.

Although the standard Kalman filter (KF) is optimal in this setting in the absence of AO (i.e. $\eta = 0$), the presence of AO (i.e. $\eta \neq 0$) generally invalidates the use of KF for optimal inference of the unobserved states $X_t = (x_1, \cdots, x_t)$, model parameters θ, or any associated function of interest $h(X_t, \theta|Y_t = (y_1, ..., y_t))$. This leads to the need to develop robust filtering alternatives to the standard KF. In what follows, we restrict attention to root mean squared error (RMSE) optimal filtering, noting that our results extend also to other applicable generalizations of the arising bias-variance trade-off under other loss functions.

2.2 Robust KF via Huberization

Under the additional restriction of zero mean and locally bounded AO disturbances u_t, Calvet et al. (2015) establish the RMSE optimality of the robust KF (henceforth RobKF) using Huberization of the Kalman updates. Estimation of the filtering distribution $x_t|Y_t$ for $Y_t = \{y_1, y_2, ..., y_t\}$ based on the RobKF modification of Kalman's algorithm is as follows. Given $\bar{x}_t = \mathbb{E}[x_t|Y_{t-1}]$, $P_t = \mathbb{V}[x_t|Y_t]$, Kalman gain K_t and innovation ϵ_t at time t, the standard Kalman update for the state distribution $x_t|Y_t \sim \mathcal{N}(\hat{x}_t, P_t)$ would be $\hat{x}_t = \bar{x}_t + K_t\,\epsilon_t$. RobKF bounds the impact of AO entering y_t on ϵ_t by Huberizing the prediction error:

$$\hat{x}_t^{\text{RobKF}} = \bar{x}_t + K_t\,\epsilon_t\,min\left(1\,,\,\frac{\kappa}{\parallel K_t\,\epsilon_t \parallel}\right) \tag{2}$$

The optimal value of the Huberization constant κ depends on the model parameters. However, RMSE optimality of RobKF need not hold in real-world settings with AO disturbances u_t that are not zero mean and locally bounded. This includes the important case of patches of outliers with the same sign (or clusters of highly correlated outliers) known to pose a particular challenge in many applications. Not only the dependence structure, but

also the distribution of u_t is generally unknown in practice and can vary over time, reducing filtering robustness and effectiveness of both Huberization approaches like Calvet et al. (2015) and heavy-tailed modeling approaches in the spirit of Harvey and Luati (2014).

All of this suggests that existing robust filters can still be further enhanced in many real-world applications. Here, we propose two simple methods for enhanced robust filtering, building on the concept that replacing faulty measurements with missing data can help reduce remaining filtering distortions due to outliers.

2.3 Robust KF via Supervised Missing Data Substitution

Our supervised method for missing data substitution is designed to improve the performance of RobKF in cases when its optimality gets violated. We thus formulate the supervised MD filter (MD-RobKF) by missing data substitution in lieu of truncation upon exceeding the Huber threshold in RobKF by modifying equation (2) above as follows:

$$\hat{x}_t^{\text{MDRobKF}} = \bar{x}_t + K_t\,\epsilon_t\,\mathbb{I}\left[1 \;\leq\; \frac{\kappa}{\parallel K_t\,\epsilon_t \parallel}\right] \tag{3}$$

In this way, filtering errors due to outliers above the Huber threshold ($\parallel K_t\epsilon_t \parallel > \kappa$) get eliminated instead of truncated. This prevents accumulation of errors in the same direction in the presence of patches of outliers of the same sign (or clusters of correlated outliers) as in the case of the standard RobKF filter in equation (2). Consequently, supervised missing data substitution significantly improves the filter's ability to overcome patches or clusters of correlated outliers. As we document below, our MD-RobKF filter can therefore outperform the standard RobKF filter of Calvet et al. (2015) in such settings by a wide margin.

It should be noted that other outlier detection methods can be employed for robust filtering via supervised missing data substitution in a similar fashion. However, all techniques for outlier detection share one common drawback: they cannot identify smaller outliers that

do not exceed the detection threshold. This motivates our unsupervised approach to missing data substitution based on exogenous randomization aiming to improve robustness also to smaller outliers falling below viable detection thresholds.

2.4 Robust Filtering via Unsupervised Randomization of Missing Data

Our unsupervised missing data substitution method, RMDX, is based on averaging filtering and forecasting targets for randomly induced subsets of missing data in the measurement series at an exogenously set randomization rate. We start by augmenting any given state-space model to induce exogenously randomized missing data as follows:

$$x_0 \sim \mu_\theta(\cdot) \tag{4}$$

$$x_t|x_{t-1}, \theta \sim g_\theta(\cdot|x_{t-1}, \theta) \tag{5}$$

$$c_t \sim P(.|y_1, ..., y_T, \beta) = P(.|\beta) \tag{6}$$

$$y_t|x_t, \theta \sim \begin{cases} f_\theta(\cdot|x_t) \text{ if } c_t = 1 \\ \text{missing if } c_t = 0 \end{cases} \tag{7}$$

As standard, $x_t \in \mathbb{R}^{d_x}$ is a latent Markov process parameterized by $\theta \in \Theta$ with an initial distribution μ_θ (equation (4)) and a transition kernel g_θ (equation (5)). The observation $y_t \in \mathbb{R}^{d_y}$ contains information about the latent states x_t through the kernel f_θ (equation (7)). However, unlike a standard state-space model, we further impose that y_t is only observed a fraction of the time and governed by random infusion of missing data via the exogenously drawn indicator c_t as specified by equations (6) and (7).

We then consider a filter F and enhance its robustness to outliers with randomized missing data substitution to obtain filter RMDX-F. Let $x_t^F|Y_t, C_t$ denote the F-filtered

distribution of x_t given an exogenous draw $C_t = \{c_1, c_2, ..., c_t\}$ inducing missing data points in the observations $Y_t = \{y_1, y_2, ..., y_t\}$ for the states $X_t = \{x_1, x_2, ..., x_t\}$ as specified by equations (6) and (7). Under the AO contamination structure given by equation (1) this F-filtered distribution would be distorted only for draws of C_t such that the subsets of time indices $\{t : c_t = 1\}$ and $\{t : u_t \neq 0\}$ have a non-empty intersection. As a key theory result below we show that, under mild regularity conditions, a simple combinatorial argument implies that the share of draws subject to such distortion would shrink as missing data gets induced at a higher rate. Therefore, constructing the filter RMDX-F by averaging targets of interest over the full set of F-filtered distributions $\{x_t^F|Y_t, C_t, \beta\}$ for all possible draws $\{C_t|\beta\}$ can lead to bias reduction at the expense of higher variance. This gives rise to regularization and bias-variance trade-off as a function of the randomization rate β, which can be set optimally using standard cross-validation techniques.

More formally, our RMDX enhancement of filter F through unsupervised missing data substitution can thus be defined as follows:

Definition 1. *RMDX-F filter extension of filter F*

Given filter F and function $h(X_T^F, \theta^F|Y_t)$ of filtered states X_T and model parameters θ, define the RMDX-F filtered estimate of $h(X_T, \theta|Y_t, \beta)$ for $T \geq t$ as

$$\bar{h}^{RMDX\text{-}F}(X_T, \theta|Y_t, \beta) := \sum_{C_t^i \in \{0,1\}^t} h(X_T^F, \theta^F|Y_t, C_t^i) P(C_t^i|\beta) \quad , \tag{8}$$

where $X_T^F, \theta^F|Y_t, C_t^i$ stand for the F-filtered states X_T and model parameters θ given measurements Y_t and indicators C_t^i inducing missing data points at randomization rate β as dictated by equations (6) and (7).

In essence, RMDX aims to enhance robustness to outliers of F-filtered estimates for any function of interest $h(X_T, \theta)$ of latent states and model parameters by averaging over all respective estimates $h(X_T^F, \theta^F|Y_t, C_t^i)$ based on F-filtered X_t^F and θ_t^F given Y_t, $T \geq t$,

and each possible draw of the induced missing data indicator C_t^i under $P(\cdot|\beta)$. In terms of implementation, the choice of probability weights $P(C_t|\beta)$ for the missing data indicators C_t is flexible. The most basic scheme one could use is sampling each $C_t \sim Bernoulli(\beta)$ independently. However, especially for low β, this design assigns significant probability to using a low number of observations, which can lead to identification and numerical instability issues due to an insufficient number of observations. Therefore, as a more practical approach we recommend fixing the number of observations at $[\beta t]$, where $[x]$ is the nearest integer to x. This leads to the following distribution for C_t:

$$P(C_t) = \begin{cases} \binom{t}{[\beta t]}^{-1} \text{ if } |C_t| = [\beta t] \\ 0 \text{ otherwise} \end{cases} \tag{9}$$

In practice, since the sample space of C_t is growing exponentially with t, it is infeasible to calculate the expectation in equation (8) across all possible C_t. Therefore, in applications, we replace it with an average over Monte Carlo samples from $P(.|\beta)$. While this introduces some stochastic noise to the estimation, in practice we find that this noise is small provided a large enough number of indicator draws.

This specific formulation of our RMDX method represents a natural state space time-series extension of bagging, originally developed by Breiman (1996) for *iid* data. Time-series dependence is fully preserved by RMDX thanks to retaining the original time index of each observation and randomly drawing only induced missing values in each re-sampled measurement series, exploiting the ability of state-space models to handle missing data. As such, the RMDX construction of alternative measurement series by inducing randomly missing data points plays a similar role in a time-series setting as the drawing of random subsamples in *iid* settings originally considered by Breiman (1996). To the best of our knowledge, the benefits of this particular avenue for extending bagging to a time-series setting have not been explored in such context.

2.5 Bias Variance Decomposition Theory

We show that RMDX introduces a bias-variance trade-off controlled by the missing data randomization rate β. Our key result is that the bias term decreases, while the variance term increases as the proportion of measurements not substituted by missing values shrinks. This establishes the existence of an optimal randomization rate for filters violating RMSE-optimality in the presence of outliers and justifies grid search for the optimal randomization rate via standard cross-validation techniques. Our theory builds on the following generic assumption on the prevalence and nature of outliers contaminating the available measurements as the sample size t gets large.

Assumption 1. *A fraction $m \in [0, 1)$ of the observations are corrupted by outliers, with $M \subset \{1, 2, ..., t\}$ denoting the subset of time indices of the observations contaminated by outliers and $M' = \{1, 2, ..., t\} \setminus M$ denoting the subset of time indices of all other uncontaminated observations.*

As a key result, we first establish bias reduction by RMDX under mild regularity conditions ensuring that outliers are not prevalent and cannot cause explosive filtering biases.

Proposition 1. *Given Assumption 1 and $t \leq T$, let $r = m \cdot t = |M|$ be the number of outliers among the observations Y_t for time indices in set M and $k = \beta t = |C_t^i|$ be the number of retained observations by each RMDX draw of the indicator path C_t^i, $i = 1, 2, ..., \binom{t}{k}$ with $P(C_t^i|\beta)$ given by equation (9) and corresponding F-filtered states and model parameters $X_T^F, \theta^F|Y_t, C_t^i$. Further assume that $r = o(t)$ and the biases associated with the function $h(X_T^F, \theta^F|Y_t, C_t^i)$ for each indicator path C_t^i, $i = 1, 2, ..., \binom{t}{k}$ are uniformly bounded: $\left|\mathbb{E}_t h(X_T^F, \theta^F|Y_t, C_t^i) - \mathbb{E}_t h(x_T, \theta)\right| < B$ as t gets large. Then the RMDX-F filtered and forecasted function values $\bar{h}^{RMDX\text{-}F}(X_T, \theta|Y_t, \beta) = \frac{1}{\binom{t}{k}} \sum_i h(X_T^F, \theta^F|Y_t, C_t^i)$ would be*

asymptotically unbiased as $\beta = \frac{k}{t}$ *becomes small when* t *gets large:*

$$\mathbb{E}_t \bar{h}^{RMDX\text{-}F}(X_T, \theta | Y_t, \beta) - \mathbb{E}_t h(X_T, \theta) \longrightarrow 0 \quad as\ \beta = \frac{k}{t} \to 0\ with\ t \to \infty \tag{10}$$

Proof. See Appendix A □

The obtained results in Propositions 1 establish that the RMDX framework offers a way to reduce prediction biases by lowering the randomization rate ($\beta < 1$). On the flip side, akin to sub-sampling, restricting data utilization always comes at the expense of larger variance relative to utilizing the full sample ($\beta = 1$). This leads to our main theory result that the RMDX framework introduces a bias-variance trade-off controlled by the randomization rate β as detailed by Proposition 2 below.

Proposition 2. *Bias-Variance Decomposition*

$$\mathbb{E}_t\Big(\bar{h}^{RMDX\text{-}F}(X_T, \theta|\beta) - h(X_T, \theta)\Big)^2 = \underbrace{\mathbb{E}_t\Big(\bar{h}^{RMDX\text{-}F}(X_T, \theta|\beta) - \mathbb{E}_t \bar{h}^{RMDX\text{-}F}(X_T, \theta|\beta)\Big)^2}_{Reducible\ variance\ term\ \downarrow\ as\ \beta\ \uparrow} +$$

$$+ \underbrace{\Big(\mathbb{E}_t\big(\bar{h}^{RMDX\text{-}F}(X_T, \theta|\beta)\big) - \mathbb{E}_t h(X_T, \theta)\Big)^2}_{Reducible\ bias\ term\ \downarrow\ as\ \beta\ \downarrow} + \underbrace{\mathbb{E}_t\Big(\mathbb{E}_t h(X_T, \theta) - h(X_T, \theta)\Big)^2}_{Irreducible\ variance\ term}$$

where $t \leq T$*, and all expectations reflect the distribution of measurements and latent states given information up to time* t*.*

Proof. See Appendix B □

The arising bias-variance trade-off establishes the existence of an optimal randomization rate for filters violating RMSE-optimality in the presence of outliers and justifies grid search for the optimal randomization rate via standard cross-validation techniques.

Corollary 1. *Existence of optimal missing data randomization rate*

For any filter F *that is not RMSE optimal in the presence of measurement outliers there*

exists an optimal $\tilde{\beta} \in (0, 1]$ *minimizing RMSE of the RMDX-F filter as a function of the randomization rate* β*.*

To broaden the scope of these insights beyond the popular special case of a square loss function, it is useful to note that the bias-variance decomposition in Proposition 2 readily extends also to a wide range of other loss functions. First, James and Hastie (1997) and James (2003) have shown how to obtain a straightforward generalization of the bias-variance decomposition in Proposition 2 for any symmetric loss function. Second, Heskes (1998), Wu and Vos (2012) and Vos and Wu (2015) have further obtained an analogous bias-variance decomposition valid for any kind of error measure that can be derived from a Kullback-Leibler divergence or loglikelihood stemming from the underlying probability model. Therefore, the above insights regarding the RMDX-induced bias-variance tradeoff as a function of the randomization rate β readily extend for any such loss functions.

3. Performance Comparisons on Simulated Data

To validate our theory results and demonstrate the attainable enhancements in robust filtering via missing data substitution, we conduct finite-sample performance comparisons in controlled Monte Carlo experiments. We consider the following three filters and their RMDX-enhanced counterparts: (1) RMDX-KF enhancing the standard Kalman filter KF; (2) RMDX-RobKF enhancing the robust RobKF filter; (3) RMDX-MD-RobKF enhancing our supervised MD-RobKF filter.

Our simulations are based on the setup studied in Calvet et al. (2015) under *i.i.d.* AO contamination structure with varying contamination magnitudes η in accordance with equation 1. However, apart from *i.i.d* outliers when RobKF optimality holds we also consider the empirically relevant case of clustered arrivals of outliers of the same sign (patches of one-sided AO contamination) violating the necessary condition for RMSE optimality of

RobKF. We document substantial improvements in robust filtering when employing our methods for missing data substitution, particularly in the case where outliers arrive in clusters rather than independently. We discuss the data generating process and contamination structures in Section 3.1 and then compare filtering performance in Section 3.2.

3.1 Monte Carlo Setup

3.1.1. State Space Model with Additive Outliers

Following Calvet et al. (2015) we consider the following linear Gaussian state-space model

$$y_t^1 = 0.1x_t^1 - 0.1x_t^2 + \nu_t^1 \qquad y_t^2 = 0.1x_t^1 + 0.1x_t^2 + \nu_t^2$$
$$x_t^1 = 0.9x_{t-1}^1 + \omega_{t-1}^1 \qquad x_t^2 = 0.9x_{t-1}^2 + \omega_{t-1}^2$$

where $y_t = (y_t^1, y_t^2) \in R^2$, $x_t = (x_t^1, x_t^2) \in R^2$, $\nu_t \sim N(0, I_2)$, $\omega_t \sim N(0, I_2)$, and $E(\nu_t \omega_t') = 0$.

We consider two types of AO-based contamination structures in accordance with equation (1). First, we consider continuous contamination used in Calvet et al. (2015) with *i.i.d.* arrivals at 5% frequency. Conditional on contamination present in the data, the contamination term is given by a product of contamination coefficient $\eta \in R$ and $u_t \in R^2$, $\eta \cdot u_t$, where u_t is sampled uniformly from a ball of radius $||y_t^* - \mu_t^*||$, where μ_t^* is a filtered mean of states at time t formulated with uncontaminated observations $y_1^*, ..., y_t^*$, i.e., $\mu_t^* = E(x_t | y_1^*, ..., y_t^*)$ that can be easily obtained with an application of the Kalman filter upon observing the true measurements. The value of contamination coefficient controls the size of the arriving outliers. We assume that $\eta \in [-40, 40]$, which allows to study a wide range of AO contamination levels.[1]

In our second contamination scenario, we consider patch-based arrivals of outliers of the

[1] The values studied in Calvet et al. (2015) range from −25 to 25.

same sign but keeping the definition of the outlier size the same as in the first setup. More specifically, we assume that contaminated observations have frequency of 5% and are clustered (patched) on blocks of N_c observations. The blocks are assumed to be interleaved by the same constant number of uncontaminated observations. The length of contamination clusters N_c determines the extent of the patch-based contamination, with higher values of N_c indicating longer uninterrupted spans of unreliable information that can distort an insufficiently robust filter.

In both scenarios, we set the sample length to $10,000$.[2] This translates into about 500 contaminated observations for the *i.i.d.* contaminated scenario. For the second scenario we pick 10 blocks of $N_c = 50$ observations each for the patch-based contamination.

3.1.2. Robust Kalman Filters

We consider six filter specifications. The standard KF, the RobKF and the supervised MD-RobKF filters are discussed in Section 2. We contrast each filter's performance to the proposed extensions augmenting all three filters with unsupervised missing data substitution based on RMDX. The unsupervised RMDX versions of KF, RobKF and MD-RobKF filters are denoted as RMDX-KF, RMDX-RobKF and RMDX-MD-RobKF, respectively. For the RobKF-based filters we set $\kappa = 3.08$ as the optimal Huberization threshold for this model from Calvet et al. (2015). For the RMDX augmented filters, we specify the randomization paramater $\beta(\eta)$ to take the values that minimize the root mean squared error (RMSE) for the true states for an assumed contamination magnitude η.

Our filter comparisons are based on evaluating RMSEs for the filtered states and the failure rates for 90% prediction bands for the states. As documented in Calvet et al. (2015), the RobKF filter clearly dominates the KF filter, yielding lower RMSEs and failure rates closer to the theoretical 10%. We explore the scope for further improvement via RMDX.

[2] This sample length has been found by Calvet et al. (2015) to be sufficient to produce satisfactory precision when comparing filtering performance for this model.

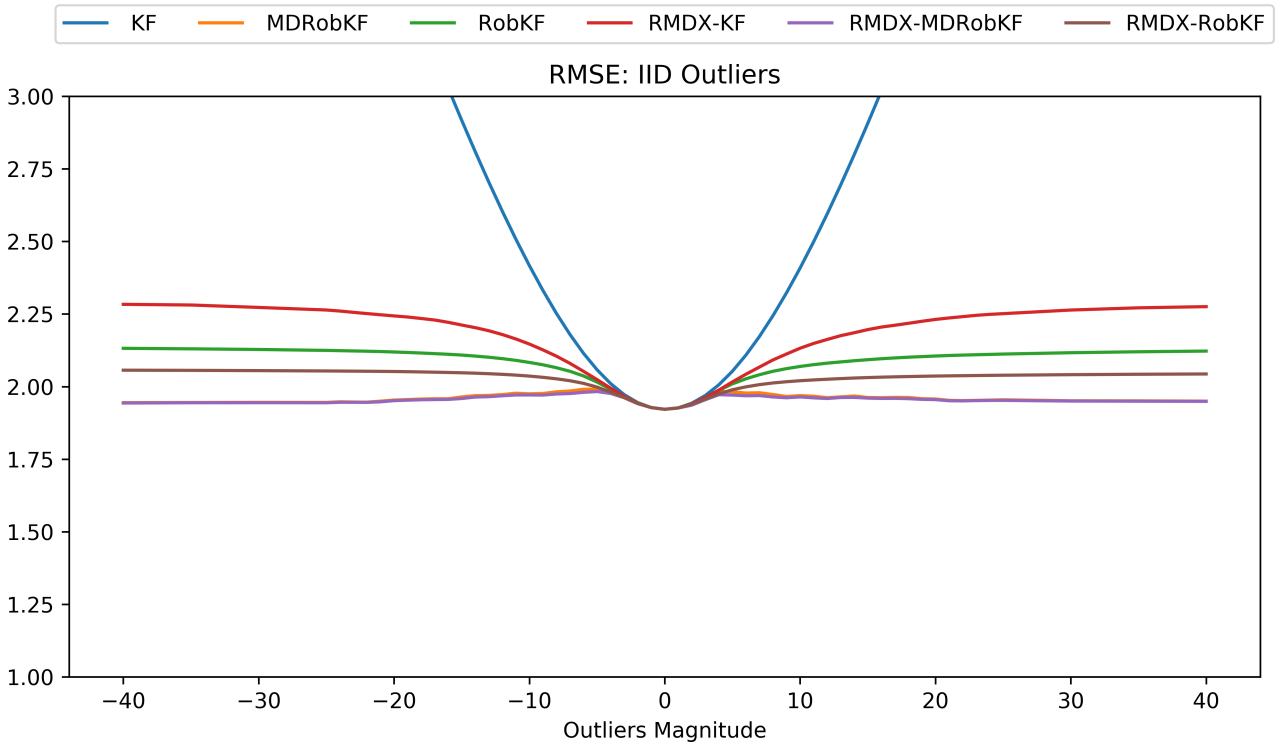
3.2 Efficiency and Robustness Comparisons

3.2.1. The case of *i.i.d.* - contaminated measurements

The RMSEs attained by each filter for a range of contamination levels η are presented in the top panel of Figure 1 and Panel A of Table 1. Calvet et al. (2015) show RMSE-optimality of the RobKF filter in the case of *iid* contamination. The attained RMSE values of the RobKF filter clearly dominate the standard KF filter at any contamination level with a reduction of up to 65% of the RMSE for the extreme size of contamination with $|\eta| = 40$. Our proposed supervised MD-RobKF filter still improves on the RobKF filter showing the lowest RMSEs among filters not augmented by RMDX. The unsupervised RMDX method further improves the KF, RobKF and MD-RobKF filters, with the higher improvements for the standard KF filter, followed by RobKF filter and only minimal gains for the supervised MD-RobKF filter. The improvements for the MD-RobKF filter although minor, are largest in the middle-range of contamination sizes. Overall, the performance of the supervised MD-RobKF filter dominates other filters, including the Huberized RobKF filter thanks to removing outliers exceeding the Huber detection threshold. Enhancing the filters using our unsupervised RMDX method yields extra gains thanks to further suppressing filtering distortions by smaller outliers falling below the Huber detection threshold.

In the bottom-panel of Figure 1 and in Panel B of Table 1 we present failure rates computed for every contamination level at 10% theoretical coverage based on 90% prediction bands. Overall, the results for failure rates are similar to those for RMSE performance. It is evident that both the KF and RobKF filters are outperformed by the MD-RobKF filter. The RMDX-augmented filters show further improvements and perform best. The RMDX method offers smallest gains for the already very well performing MD-RobKF filter.

Figure 1: RMSE and failure rates as a function of contamination size for i.i.d. contaminated observations. The values are computed for the KF, RobKF, MD-RobKF, RMDX-KF, RMDX-RobKF, RMDX-MD-RobKF filters with β minimizing each filter's RMSE for a given contamination size.



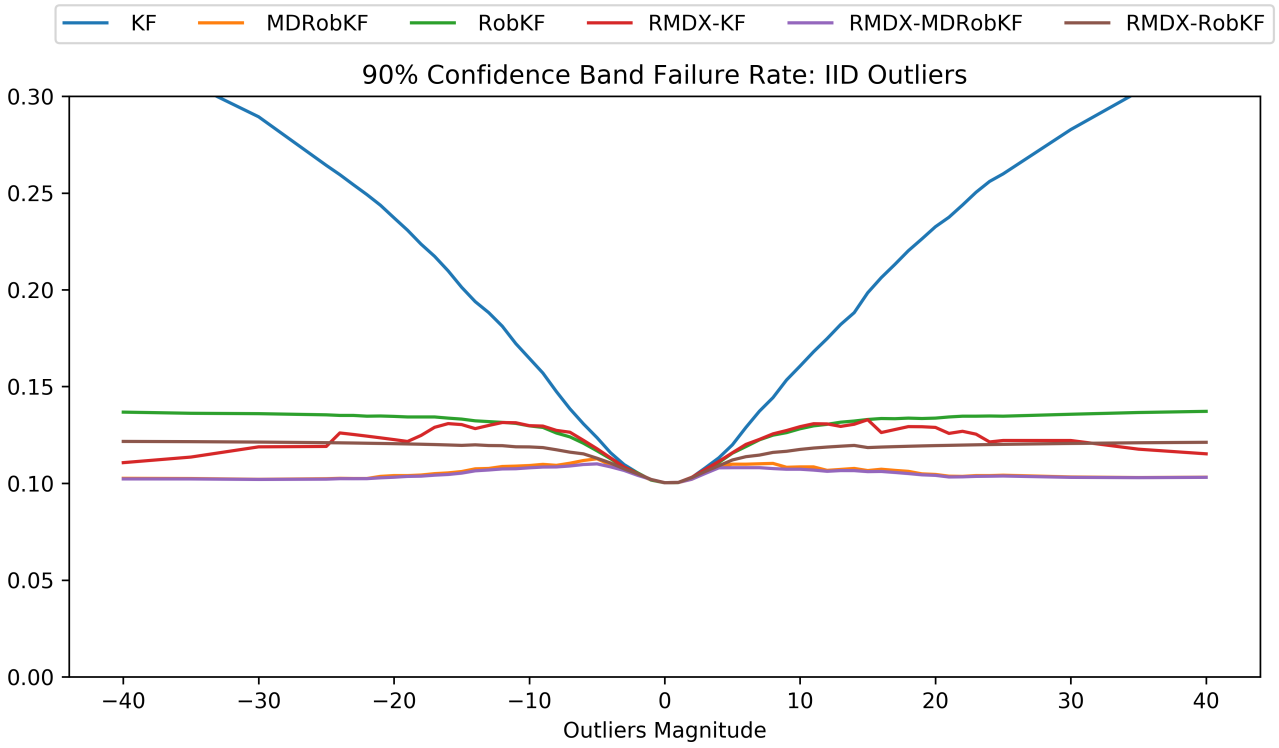
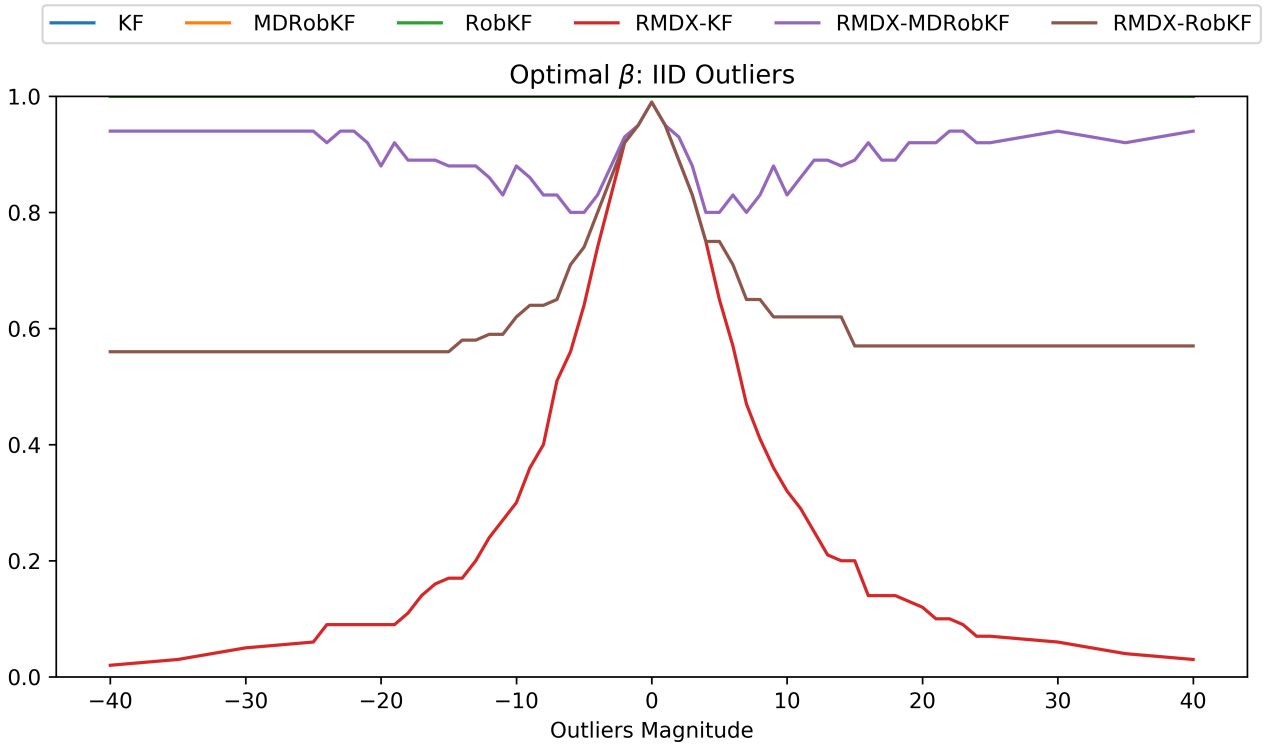


Table 1: Filter-optimized RMSEs and failure rates at 10% level for i.i.d.-contaminated data.

	Contamination coefficient								
Filter	-40	-20	-10	-5	0	5	10	20	40
Panel A: RMSEs									
KF	6.150	3.499	2.417	2.058	1.922	2.053	2.408	3.487	6.136
RobKF	2.132	2.119	2.083	2.015	1.922	2.009	2.069	2.105	2.122
MD-RobKF	1.945	1.954	1.975	1.991	1.922	1.982	1.969	1.957	1.950
RMDX-KF	2.289	2.246	2.149	2.025	1.922	2.016	2.131	2.230	2.280
RMDX-RobKF	2.059	2.052	2.036	1.999	1.922	1.989	2.020	2.037	2.045
RMDX-MD-RobKF	1.944	1.952	1.971	1.982	1.922	1.971	1.964	1.955	1.949
Panel B: Failure Rates									
KF	0.324	0.237	0.165	0.124	0.100	0.120	0.161	0.233	0.323
RobKF	0.137	0.135	0.130	0.117	0.100	0.116	0.128	0.134	0.137
MD-RobKF	0.103	0.104	0.109	0.113	0.100	0.110	0.109	0.105	0.103
RMDX-KF	0.106	0.125	0.130	0.119	0.100	0.115	0.128	0.125	0.123
RMDX-RobKF	0.123	0.122	0.118	0.113	0.100	0.112	0.117	0.120	0.122
RMDX-MD-RobKF	0.102	0.103	0.108	0.110	0.100	0.108	0.108	0.104	0.103

The optimal randomization rates β minimizing each filter's RMSE depicted in Figure 2 suggest that the optimal values are markedly lower for the RMDX-augmented KF and RobKF filters and the optimal values decrease with the level of contamination. In contrast, the supervised MD-RobKF filter does not benefit to the same extent when augmenting it with missing data randomization using RMDX as outliers below the Huber detection threshold are less likely to cause large filtering distortions in the i.i.d case and its lowest values of optimal β of around 0.8 coincide with the highest improvements in RMSEs at mid-sized levels of contamination in Panel A of Table 1.

Figure 2: Opitmal β minimizing each filter's RMSE as a function of contamination size for i.i.d. contaminated observations. The values are computed for the RMDX-KF, RMDX-RobKF and RMDX-MD-RobKF filters.



3.2.2. The case of patch-contaminated measurements

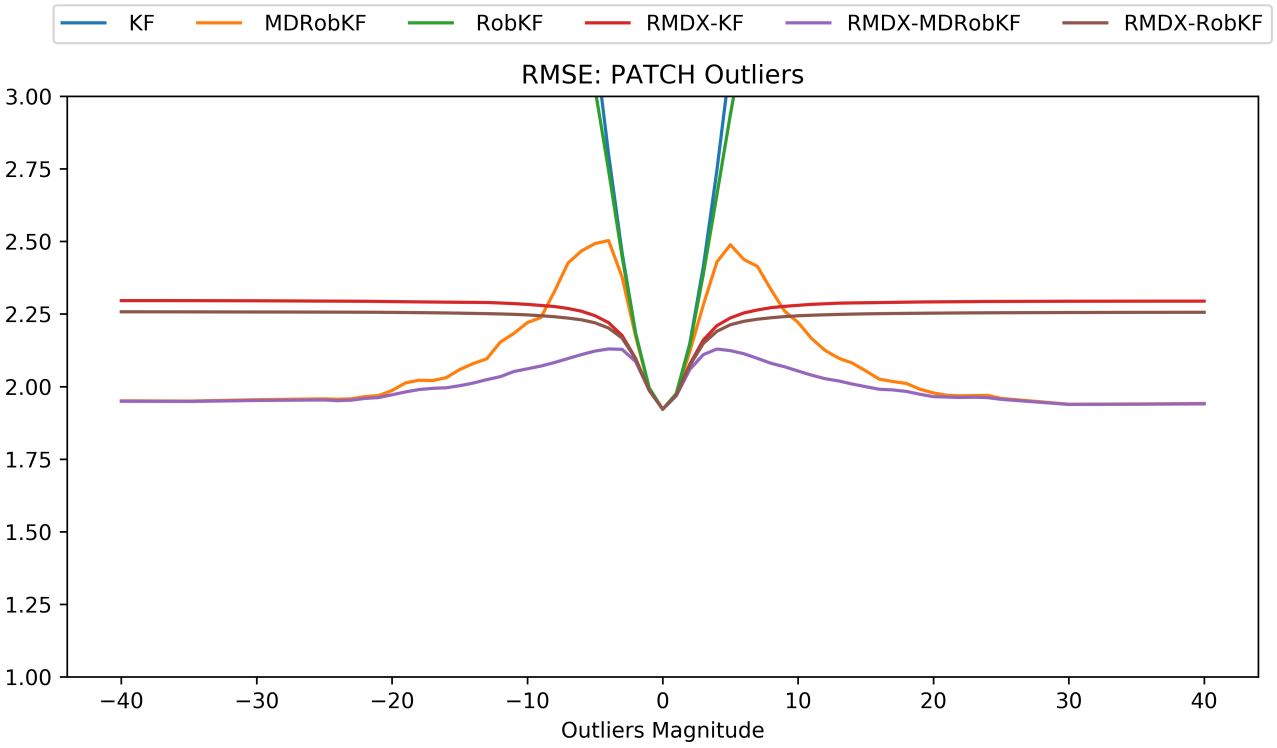
In the case of patch-contaminated measurements, the necessary condition for RMSE-based optimality of the RobKF filter is violated. As a result, both the KF and RobKF filters have subpar RMSE performance as can be seen in the top panel of Figure 3 and in Panel A of Table 2. The performance of the supervised MD-RobKF filter is superior to other non-RMDX augmented filters and even in the most challenging scenario of middle-sized outliers is about 30% worse than under no contamination. The unsupervised RMDX method improves performance of all filters. The RMSE reduction increases with contamination size for both the KF and Rob-KF filters. The RMDX-MD-RobKF version of the supervised MD-RobKF filter clearly outperforms all other filters. As seen in Figure 3, the RMDX-MD-RobKF filter

significantly improves over the MD-RobKF filter especially in the range of mid-sized outliers and shows only minimal performance gains for large-sized outliers where performance of the MD-RobKF filter is already close to optimal.

When it comes to the observed failure rates shown in the bottom panel of Figure 3 and reported in Panel B of Table 2, the filters perform similarly to their RMSE performance. The KF and the RobKF filters perform the worst and are clearly dominated by the MD-RobKF filter and RMDX-augmented filters. Similar to the case of *iid* contamination, the biggest gains attributed to RMDX-augmentation of the supervised MD-RobKF filter can be observed for small and middle-range contamination levels.

An inspection of the optimal randomization rates β minimizing each filter's RMSE, presented in Figure 4, shows that the optimal values are markedly lower for the RMDX-augmented KF and RobKF filters when compared to the case of *iid* outliers, with optimal values of β decreasing with the magnitude of contamination going up. Among these filters, the RobKF filter shows the most drastic drop in optimal β when compared to those in the *iid* case for all contamination levels. The more extreme randomization of missing data helps in somewhat alleviating the lack of robustness of the RobKF filter when outliers are clustered. The drop in the optimal missing data randomization level for the RMDX-MD-RobKF filter is not as extreme and is more significant for the most challenging region of mid-sized outliers with only marginal benefit from randomization for larger-sized outliers.

Figure 3: RMSE and failure rates as a function of contamination size for patch-contaminated observations. The values are computed for the KF, RobKF, MD-RobKF, RMDX-KF, RMDX-RobKF and RMDX-MD-RobKF filters with β minimizing each filter's RMSE for a given contamination size.



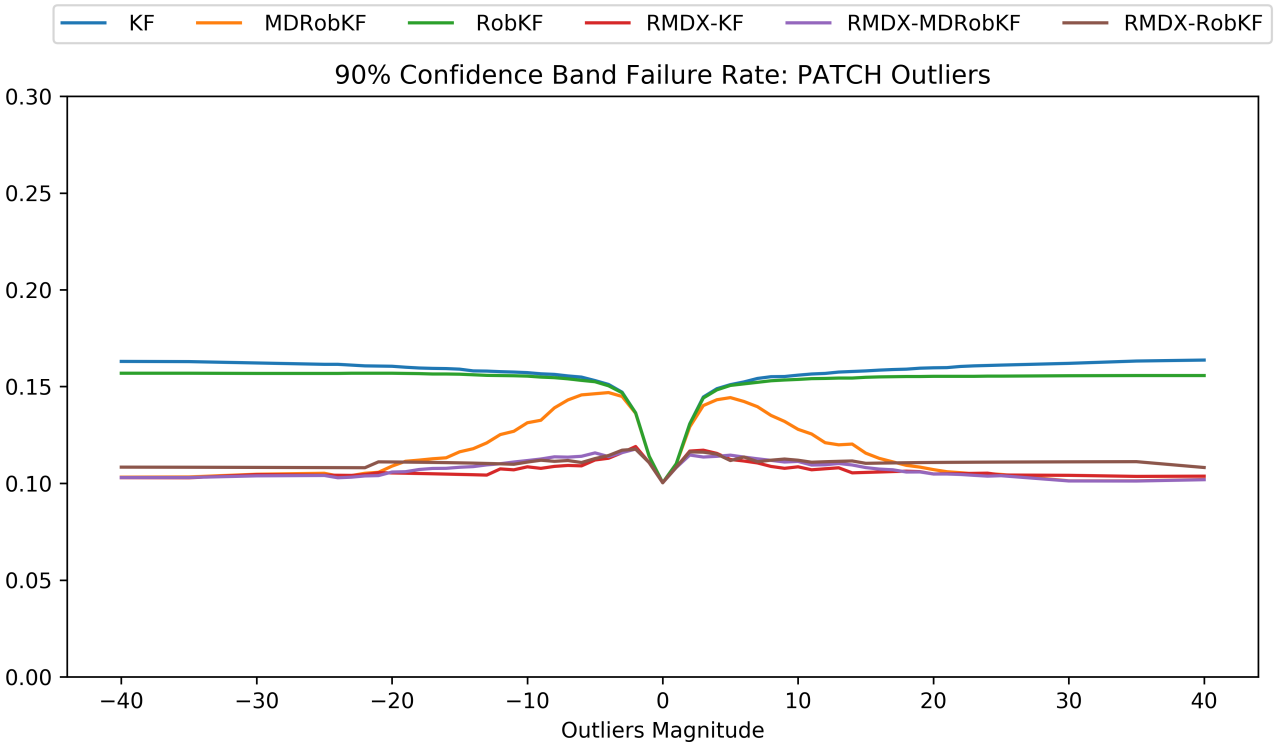
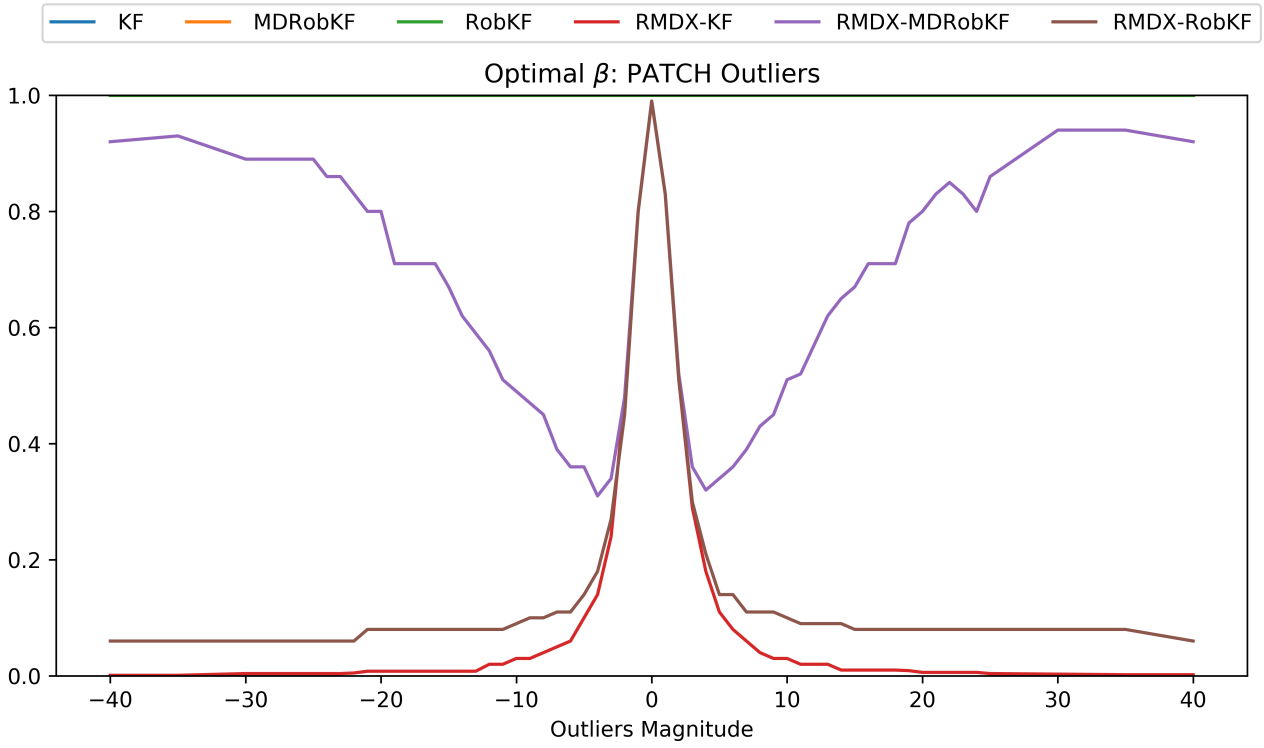


Table 2: Filter-optimized RMSEs and failure rates at 10% level for patch-contaminated data.

	Contamination coefficient								
Filter	-40	-20	-10	-5	0	5	10	20	40
Panel A: RMSEs									
KF	20.113	10.212	5.389	3.182	1.922	3.120	5.315	10.135	20.035
RobKF	5.355	4.851	4.012	3.026	1.922	2.937	3.959	4.818	5.332
MD-RobKF	1.951	1.986	2.220	2.493	1.922	2.488	2.221	1.978	1.942
RMDX-KF	2.323	2.295	2.287	2.244	1.922	2.237	2.287	2.293	2.318
RMDX-RobKF	2.261	2.260	2.248	2.226	1.922	2.216	2.243	2.257	2.258
RMDX-MD-RobKF	1.949	1.973	2.061	2.124	1.922	2.125	2.054	1.965	1.940
Panel B: Failure Rates									
KF	0.163	0.161	0.157	0.153	0.100	0.151	0.156	0.160	0.164
RobKF	0.157	0.157	0.155	0.153	0.100	0.151	0.154	0.155	0.156
MD-RobKF	0.103	0.109	0.131	0.146	0.100	0.144	0.128	0.107	0.102
RMDX-KF	0.109	0.106	0.104	0.112	0.100	0.111	0.104	0.107	0.109
RMDX-RobKF	0.107	0.107	0.112	0.108	0.100	0.118	0.112	0.114	0.107
RMDX-MD-RobKF	0.103	0.106	0.112	0.112	0.100	0.112	0.111	0.105	0.102

Figure 4: Opitmal β minimizing each filter's RMSE as a function of contamination size for patch-contaminated observations. The values are computed for the RMDX-KF, RMDX-RobKF and RMDX-MD-RobKF filters.



4. Empirical Illustration: Inflation forecasting

To illustrate the empirical performance of the robust supervised and unsupervised filters in filtering and forecasting we choose a well-known setting from the time-series literature on extracting inflation trends where the use of standard state-space models is known to suffer from the presence of clustered measurement outliers. This time-series setting offers an ideal real-world testing setup allowing us to compare the empirical performance of the proposed robust filters in forecasting when applied to standard state-space models.

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.[3] 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 persistent 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. What makes such real-data applications especially challenging are unknown model parameters that also need to be estimated. If parameters were known and set to true values, they would provide a natural anchor for the Huberized filters in limiting the impact of outliers on state estimation. Hence, if a filter is unable to perform well in the estimation of parameters, it is also likely to fail to detect outliers.

4.1 Models and Filters

We consider three model specifications: the standard unobserved components model (UC) with a random walk inflation trend, an unobserved components model with autoregressive

[3] For a literature survey on inflation forecasting see for example Faust and Wright (2013).

inflation trend and fixed mean at the long-run inflation target of 2% (ARMF), and an unobserved components model with autoregressive inflation trend (AR) with unknown mean. The first two models can be seen as constrained versions of the AR model:

$$x_t|x_{t-1},\theta \sim g(x_t|x_{t-1}) = N(\mu + \rho(x_{t-1}-\mu), \sigma_x^2) \tag{11}$$

$$y_t|x_t,\theta \sim f(y_t|x_t) = N(x_t, \sigma_y^2) \tag{12}$$

where $\rho = 1$ for the UC model, $\rho \in (-1,1)$ and $\mu = 2\%$ for the ARMF model, and $\rho \in (-1,1)$ for the AR model.[4,5] The model observations are given by y_t that is the annualized change in the log price-level (PCE) at quarter t. We estimate the above model with the same aggregate PCE price index quarterly data series (PCE-all) used in Stock and Watson (2016) from 1960Q1 to 2015Q1 with a forecast evaluation period set to 1990Q1-2015Q1, in order to directly compare with the more recent UCSVO model presented in that paper.[6]

We estimate all models using maximum likelihood estimation with expanding window using PCE inflation data from 1960Q1 to 2015Q1 and with no look-ahead, i.e., parameter estimates and state estimates at time t depend only on observations $Y_t = (y_1, y_2, \ldots, y_t)$. The estimation starts from 1979Q1. We consider all studied filters: the KF, the Rob-KF, the supervised MD-RobKF, and their extensions using our RMDX method for randomized missing data substitution. We re-estimate each model-filter pair given different values of $\beta \in \{0.05, 0.1, \ldots, 1\}$ (if a filter is RMDX-augmented) and/or different values of the Huber threshold $\kappa = (\kappa_1, \kappa_2, \kappa_3, \infty)$.[7]

[4] Note that the mean μ is not estimated in the UC model.

[5] We set the mean of the ARMF model at 2% which is an inflation target in the United States maintained by the Federal Reserve for many years. See Federal Reserve Board (2020): Why does the federal reserve aim for 2 percent inflation over time? https://www.federalreserve.gov/faqs/economy_14400.htm.

[6] We thank Stock and Watson (2016) for making publicly available the data and program codes necessary for replicating their results at the Harvard Dataverse repository: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GJWNZW.

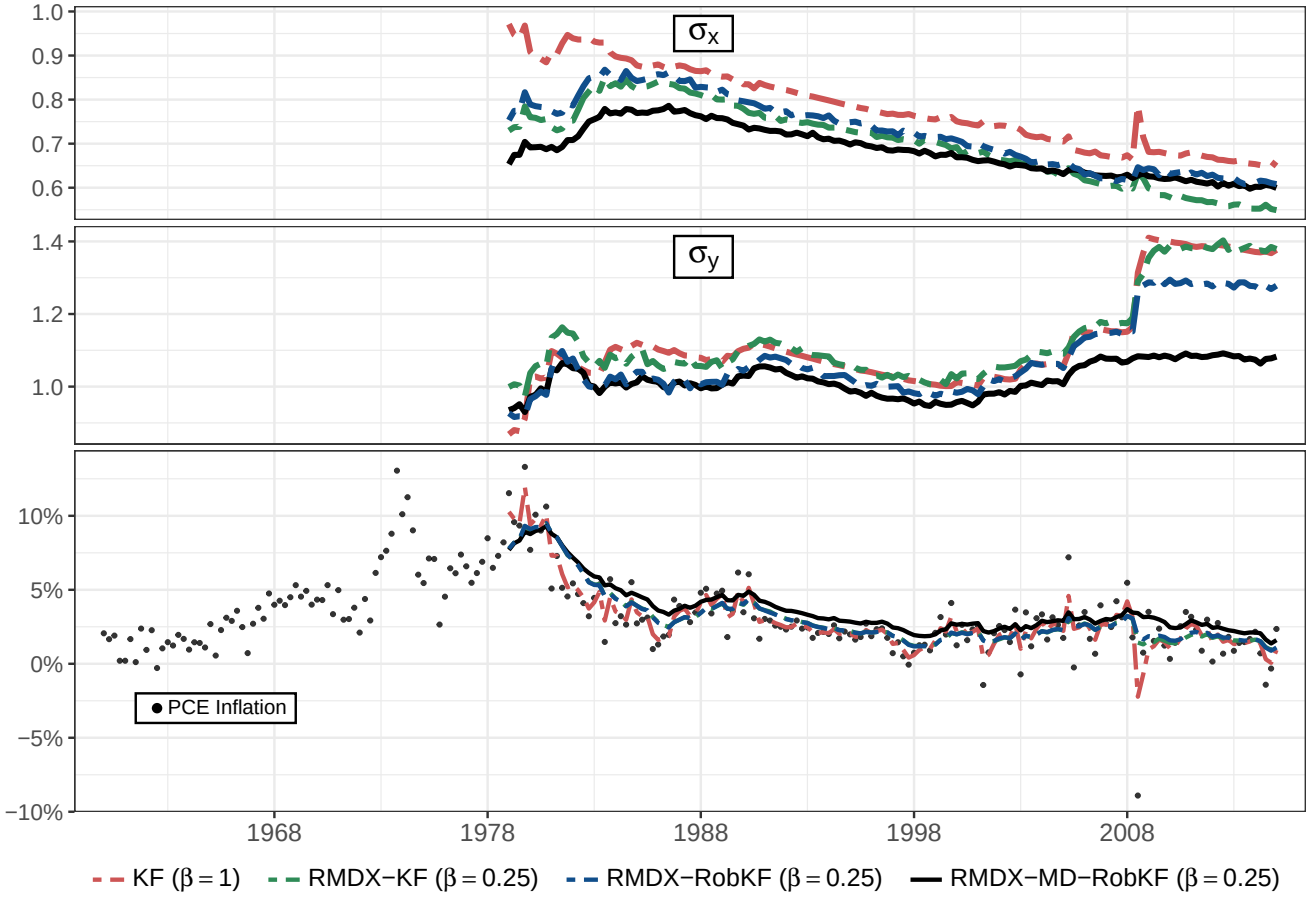
[7] In our application we set $\kappa_1 = 5.67$, $\kappa_2 = 7.63$, and $\kappa_3 = 11.34$, which coincide with Huberization at levels of 90%, 95%, and 99% for a choice of unconditional variance of x implied by $\sigma_x = 0.8$ and $\rho = 0.99$. We found points within such grid to produce viable estimation results. There is no Huberization with $\kappa = \infty$.

To better illustrate how model estimates differ across filters, the parameter estimates for the UC model estimated with representative types of filters are shown in Figure 5 (top two panels). It is evident that the highest values of volatilities σ_x and σ_y are estimated with the standard Kalman filter followed by the RMDX-KF filter that allows for missing data randomization.[8] The plots suggest that for the same level of missing data randomization β and Huberization κ, the parameter estimates can differ significantly across all studied filters. Overall, the lowest values of volatility estimates are produced by the RMDX-MD-RobKF filter consistent with its superior ability to suppress the impact of outliers also found in our simulation study. Most importantly, the measurement volatility estimates σ_y based on the RMDX-RobKF and the RMDX-MD-RobKF filters start to deviate significantly starting from the 2008 financial crisis suggesting a large number of clustered outliers above the Huber threshold that are accounted for by the supervised MD-RobKF filter.

To further illustrate how the choice of filters impact the estimation of latent states, we compare the filtered mean from the original UC model using the same pool of representative filters. As shown in Figure 5 (bottom panel), without applying missing data randomization the filtered mean given by the standard KF filter visually overfits the observed inflation in each quarter. On the other hand, the filtered mean produced by the RMDX-KF filter appears to follow a smoother path which better tracks the long-run trend of the process. Our proposed RMDX-MD-RobKF filter even further smooths the estimates over time with greatly reduced impact of visually clustered extreme observations towards the end of the sample in the post-2000 period. Hence, these results support our main finding in Section 3 that our preferred RMDX-MD-RobKF filter tends to be most effective in guarding against distortions by both *i.i.d.* and clustered outliers.

[8] We choose a sufficiently small value of $\beta = 0.25$ across all unsupervised filters augmented with our RMDX method to demonstrate filter performance with significant extent of missing data randomization. We also pick a value of $\kappa = 5.67$ that implies significant Huberization level.

Figure 5: Expanding window maximum likelihood parameter estimates of σ_x and σ_y for the Unobserved Components (UC) model of Stock & Watson (2007) using PCE inflation data from 1960Q1 to 2015Q1 (top two panels). The estimation starts at 1979Q1. The presented estimates are produced with the standard KF filter with no RMDX ($\beta = 1$, dashed red), the RMDX-KF filter ($\beta = 0.25$, dashed green), the RMDX-RobKF filter ($\beta = 0.25$, dashed blue) and the RMDX-MD-RobKF filter ($\beta = 0.25$, solid black). Filtered mean estimates are produced given MLE-estimated parameters (bottom panel). The plot shows filtered mean estimates of the standard KF filter with no RMDX ($\beta = 1$, dashed red), the RMDX-KF filter ($\beta = 0.25$, dashed green), the RMDX-RobKF filter ($\beta = 0.25$, dashed blue) and the RMDX-MD-RobKF filter ($\beta = 0.25$, solid black) applied to the UC model. The black dots represent the observed log-quarterly inflation.



4.2 Inflation Forecasting Performance

We compare the Mean Squared Forecasting Errors (MSFEs) of the UC, ARMF and AR models estimated with the standard KF, RobKF and MD-RobKF filters either non-augmented ($\beta = 1$), or augmented with RMDX, as discussed in the previous section.

Following Stock and Watson (2007) and Stock and Watson (2016), we set as a forecast target the average inflation 4-, 8-, and 12-quarters ahead and consider the forecast period starting from 1990Q1 up to 2015Q1, evaluating forecast performance based on the same MSFE criterion they use. 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, the randomization parameter β and Huberization constant κ when forecasts are formulated. We first estimate all models given each filter using expanding windows for all time periods t and for all considered values of β and κ. We then formulate optimal strategies for the optimal choice of the randomization rate β and Huber threshold κ given all observations Y_t available at time t and hence all associated recursive forecasts for each model, filter and forecast horizon. The first strategy minimizes mean squared forecast errors available up to time t, formulated recursively, of average inflation with horizon h, over all possible pairs of (β, κ). We denote such strategy as (Optimal, Optimal). The second strategy denoted as (Optimal, κ) minimizes the same mean squared forecast errors at each time t over β but given a value of κ. The third strategy that we denote as ($\beta = 1$, Optimal) minimizes the mean squared forecast errors over values of κ, given a value of $\beta = 1$, effectively producing optimal forecast strategy for each filter without RMDX enhancement. Given optimal strategies, we formulate out-of-sample forecasts at each time t for each model, filter and horizon using only information available up to time t.

In Table 3 we present the MSFEs calculated for each model, filter and optimal strategy for setting β and/or κ. It is evident that the standard KF filter performs worst across all models and horizons and is closely followed by the RobKF and MD-RobKF filters ($\beta = 1$). In general, RMDX-augmented filters produce significant improvements irrespective of the model and strategy used to choose the (β, κ) pair. As shown by the underlined MSFE values that denote the minimum MSFEs for each model and horizon, the RMDX-MD-RobKF filter is the best performing for the ARMF and AR models for the (Optimal, Optimal) strategy or $\kappa_1 = 5.67$, a significant size of the Huber threshold.

It is also worth noting that the ARMF model performs best across all forecast horizons. Since the mean of inflation is difficult to be estimated due to high persistence and short sample, fixing it may be more advantageous in providing model flexibility in comparison to the UC model. On the other hand, the AR model requires an estimation of the mean that translates into worse out-of-sample performance. Nevertheless, the RMDX-MD-RobKF filter contributes to the biggest relative improvements for the AR model, when compared to other filters, resulting in performance not far from the ARMF model. The performance gains from applying our RMDX method with randomized missing data substitution as well as our MD-RobKF filter with supervised missing data substitution increase with longer forecast horizons, as they depend more on the robust estimation of parameters and dynamics of the system.

For comparison, we consider also two existing benchmark models from the prior literature. The first approach, which we denote as UC-T, follows Harvey and Luati (2014) in replacing the Gaussian measurement in equation (12) with a scaled $t-$distribution. The second approach is the unobserved components model with stochastic volatility and outlier-adjustment (UCSVO) proposed by Stock and Watson (2016). The UCSVO model also minimizes the detrimental impact of outliers by subjecting them to particular distributional assumptions while also allowing for stochastic volatility. We report the performance of these models in the bottom two rows of Table 3. Our findings indicate that both the ARMF and AR models estimated with the preferred RMDX-MD-RobKF filter produce lower MSFEs than the UC-T model for all forecast horizons. The models also perform significantly better than the UCSVO model at the longer 8 and 12 quarter horizons, and the ARMF model performs comparably to the UCSVO benchmark for the shortest forecast horizon of 4 quarters also when using the RMDX-MD-RobKF filter.[9]

[9] It is important to note that the MSFE results for the UCSVO model reflect informative priors on the parameters governing the distribution of outliers, possibly affecting filtering and forecasting performance.

Table 3: MSFE comparison for different considered state space models for extracting inflation trends estimated and filtered with KF, RobKF and MD-RobKF filters either without RMDX ($\beta = 1$), or augmented with RMDX using β optimized (optimal) recursively to different forecasting horizons and using Huber threshold either preset (κ) or optimized (optimal). The considered forecast horizons are: four-quarter (Q4), eight-quarter (Q8) and twelve-quarter (Q12). The underlined values denote the minimum MSFE for each column. The last two rows report the MSFEs for the UC-T and UCSVO benchamrks described in the text.

Filter Specifications			4-Quarter			8-Quarter			12-Quarter		
Type	RMDX	κ	UC	ARMF	AR	UC	ARMF	AR	UC	ARMF	AR
KF	$\beta = 1$	-	1.651	1.515	1.658	1.257	1.011	1.345	1.163	0.831	1.366
KF	Optimal	-	1.398	1.240	1.559	1.024	0.733	1.350	0.866	0.509	1.364
RobKF	$\beta = 1$	Optimal	1.651	1.515	1.658	1.257	1.011	1.345	1.163	0.831	1.366
RobKF	Optimal	κ_1	1.364	1.210	1.550	0.990	0.679	1.294	0.842	0.488	1.356
RobKF	Optimal	κ_2	1.382	1.238	1.575	0.997	0.702	1.346	0.868	0.498	1.370
RobKF	Optimal	κ_3	1.403	1.239	1.560	1.029	0.732	1.323	0.865	0.489	1.359
RobKF	Optimal	Optimal	1.367	1.231	1.569	0.991	0.677	1.316	0.858	0.488	1.357
MD-RobKF	$\beta = 1$	Optimal	1.607	1.484	1.602	1.178	0.961	1.272	1.088	0.787	1.294
MD-RobKF	Optimal	κ_1	1.337	<u>1.131</u>	<u>1.175</u>	0.959	<u>0.526</u>	<u>0.647</u>	1.203	<u>0.404</u>	<u>0.480</u>
MD-RobKF	Optimal	κ_2	<u>1.315</u>	1.177	1.451	<u>0.924</u>	0.653	1.081	<u>0.798</u>	0.446	1.169
MD-RobKF	Optimal	κ_3	1.392	1.200	1.536	1.014	0.699	1.343	0.850	0.500	1.363
MD-RobKF	Optimal	Optimal	1.319	<u>1.131</u>	<u>1.175</u>	0.925	<u>0.526</u>	<u>0.647</u>	1.199	<u>0.404</u>	<u>0.480</u>

Benchmarks	4-Quarter	8-Quarter	12-Quarter
UC-T ($\beta = 1$)	1.41	1.13	1.01
UCSVO ($\beta = 1$)	1.09	0.81	0.69

5. Summary and Conclusions

In this paper we propose two complementary methods for enhanced robust filtering and forecasting: supervised missing data substitution (MD) upon exceeding a Huber threshold, and unsupervised missing data substitution via exogenous randomization (RMDX). We show that both missing data substitution methods improve outlier-robust filtering and foreacasting in state-space models, especially in empirically relevant cases where the optimality of existing robust filters may not hold due to violating assumptions on the outlier structure.

On the theory side, we design our supervised method, MD, to improve performance of existing Huber-based linear filters known to lose optimality when outliers of the same sign are clustered in time rather than arriving independently. The unsupervised method, RMDX, further aims to suppress smaller outliers whose size may fall below the Huber detection threshold. To this end, RMDX averages filtered or forecasted targets based on measurement series with randomly induced subsets of missing data at an exogenously set randomization rate. This leads to regularization and bias-variance trade-off as a function of the missing data randomization rate, which can be set optimally using standard cross-validation techniques.

In terms of empirical validation, we show that the proposed methods for missing data substitution are easy to implement and most effective when combined together, as documented by consistently favorable performance of our combined RMDX-MD-RobKF filter in controlled Monte Carlo experiments and a real-world application to extracting inflation trends known to suffer from the presence of clustered measurement outliers.

Looking forward, missing data substitution for enhanced robust filtering offers promising avenues for further exploration on both the theory and empirical side. Particularly intriguing in this regard is that it offers a state space time-series extension of bagging in the spirit of Breiman (1996) and extends also the rational inattention ideas of Sims (2003, 2011) based on purely statistical instead of economic loss underpinnings.

References

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

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

Crevits, R. and Croux, C. (2018). Robust estimation of linear state space models. *Communications in Statistics - Simulation and Computation*, 48:1–12.

Durbin, J. and Koopman, S. J. (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.

Faust, J. and Wright, J. H. (2013). Forecasting inflation. volume 2, chapter 1, pages 2–56. Elsevier.

Harvey, A. and Luati, A. (2014). Filtering with heavy tails. *Journal of the American Statistical Association*, 109(507):1112–1122.

Heskes, T. (1998). Bias/variance decompositions for likelihood-based estimators. *Neural Computation*, 10:1425–1433.

James, G. (2003). Variance and bias for general loss functions. *Machine Learning*, 51:115–135.

James, G. and Hastie, T. (1997). Generalizations of the bias/variance decomposition for prediction error. Technical report.

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

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50:665–690.

Sims, C. A. (2011). Rational inattention and monetary economics. In *Handbook of Monetary Economics*, volume 3A, chapter 4, pages 155–181. Elsevier.

Stock, J. H. and Watson, M. W. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39:3–33.

Stock, J. H. and Watson, M. W. (2016). Core inflation and trend inflation. *Review of Economics and Statistics*, 98(4):770–784.

Vos, P. and Wu, Q. (2015). Maximum likelihood estimators uniformly minimize distribution variance among distribution unbiased estimators in exponential families. *Bernoulli*, 21(4):2120–2138.

Wu, Q. and Vos, P. (2012). Decomposition of Kullback–Leibler risk and unbiasedness for parameter-free estimators. *Journal of Statistical Planning and Inference*, 142:1525–1536.

Acknowledgements

The authors thank Andrew Harvey, Alexei Onatskiy, Mark Salmon, Neil Shephard, Tatevik Sekhposyan, and Jonathan Wright for very helpful discussions and comments. The authors also thank conference participants at the Financial Econometrics Conference to mark Stephen Taylor's Retirement, Lancaster University, March 29-31, 2023 and the Financial Econometrics and Machine Learning (FinEML) Conference, Erasmus School of Economics, November 10-11, 2023, as well as seminar participants at the Federal Reserve Board of Governors, University of Cambridge, Heidelberg University, University of Luxembourg, University of Venice and University of Verona.

Appendix A Proof of Proposition 1

Proof. The indices of $i = 1, 2, ..., \binom{t}{k}$ of all indicator paths $\{C_t^i \in \{0,1\}^t \ : \ |\{s : c_s^i = 1\}| = k\}$ can be partitioned into two disjoint subsets $I_0 = \{i \ : \ c_s^i = 0 \text{ for all } s \in M\}$ and $I_1 = \{i \ : \ c_s^i = 1 \text{ for some } s \in M\}$ based respectively on whether the retained k observations by C_t^i contain outliers or not. Observe that $|I_0| = \binom{t-r}{k}$ and $|I_1| = \binom{t}{k} - \binom{t-r}{k}$. For brevity of notation, all F-filtered states and parameters estimated for indicator path C_t^i, $(X_T^F, \theta^F | Y_t, C_t^i)$, are denoted as (X_T^i, θ^i). This leads to the following expression for the bias of the RMDX filtered and forecasted states:

$$\begin{aligned}
\mathbb{E}_t\bar{h}(X_T, \theta|\beta) - \mathbb{E}_t h(X_T, \theta) &= \frac{1}{\binom{t}{k}} \sum_i \mathbb{E}_t h(X_T^i, \theta^i) - \mathbb{E}_t h(X_T, \theta) \\
&= \frac{1}{\binom{t}{k}} \sum_{i \in I_0} \mathbb{E}_t h(X_T^i, \theta^i) + \frac{1}{\binom{t}{k}} \sum_{i \in I_1} \mathbb{E}_t h(X_T^i, \theta^i) - \mathbb{E}_t h(X_T, \theta) \\
&= \Bigg(\underbrace{\frac{1}{\binom{t}{k}} \sum_{i \in I_0} \mathbb{E}_t h(X_T^i, \theta^i)}_{\text{No bias: } \binom{t-r}{k} \text{ terms}} - \ \frac{\binom{t-r}{k}}{\binom{t}{k}} \mathbb{E}_t h(X_T, \theta) \Bigg) \\
&\quad + \Bigg(\underbrace{\frac{1}{\binom{t}{k}} \sum_{i \in I_1} \mathbb{E}_t h(X_T^i, \theta^i)}_{\text{Bias: } \binom{t}{k} - \binom{t-r}{k} \text{ terms}} - \ \frac{\binom{t}{k} - \binom{t-r}{k}}{\binom{t}{k}} \mathbb{E}_t h(X_T, \theta) \Bigg)
\end{aligned}$$

The expression in the first bracket containing the no-bias terms would thus equal zero by construction. From this it follows:

$$\begin{aligned}
\left|\mathbb{E}_t\bar{h}(X_T, \theta|\beta) - \mathbb{E}_t h(X_T, \theta)\right| &= \left| \frac{1}{\binom{t}{k}} \sum_{i \in I_1} \mathbb{E}_t h(X_T^i, \theta^i) - \ \frac{\binom{t}{k} - \binom{t-r}{k}}{\binom{t}{k}} \mathbb{E}_t h(X_T, \theta) \right| \\
&= \frac{1}{\binom{t}{k}} \left| \underbrace{\sum_{i \in I_1} \left(\mathbb{E}_t h(X_T^i, \theta^i) - \mathbb{E}_t h(X_T, \theta) \right)}_{\binom{t}{k} - \binom{t-r}{k} \text{ terms}} \right| \leq \frac{\binom{t}{k} - \binom{t-r}{k}}{\binom{t}{k}} \cdot B \\
&= \left(1 \ - \ \frac{\binom{t-r}{k}}{\binom{t}{k}} \right) \cdot B \longrightarrow 0 \ \text{ as } \beta = \frac{k}{t} \to 0 \text{ with } t \to \infty \ ,
\end{aligned}$$

where the convergence to zero follows from

$$\frac{\binom{t-r}{k}}{\binom{t}{k}} = \frac{(t-r)(t-r-1)...(t-r-k+1)}{t(t-1)...(t-k+1)} \longrightarrow 1 \quad \text{as } r = o(t) \text{ with } t \to \infty$$

□

Remark. Similar reasoning implies that a degree of bias reduction can be attained also when $r = O(t)$ without anymore being able to guarantee that the bias would vanish asymptotically even if it can still be reduced.

Appendix B Proof of Proposition 2

Proof. The result follows directly from the following standard decomposition:

$$\begin{aligned}
&\mathbb{E}_t\Big(\bar{h}(X_T,\theta|\beta) - h(X_T,\theta)\Big)^2 = \\
&= \mathbb{E}_t\Big(\big(\bar{h}(X_T,\theta|\beta) - \mathbb{E}_t\bar{h}(X_T,\theta|\beta)\big) + \big(\mathbb{E}_t\bar{h}(X_T,\theta|\beta) - \mathbb{E}_t h(X_T,\theta)\big) + \big(\mathbb{E}_t h(X_T,\theta) - h(X_T,\theta)\big) \Big)^2 \\
&= \mathbb{E}_t\Big(\bar{h}(X_T,\theta|\beta) - \mathbb{E}_t\bar{h}(X_T,\theta)\Big)^2 + \Big(\mathbb{E}_t\big(\bar{h}(X_T,\theta|\beta)\big) - \mathbb{E}_t h(X_T,\theta)\Big)^2 + \mathbb{E}_t\Big(\mathbb{E}_t h(X_T,\theta) - h(X_T,\theta)\Big)^2
\end{aligned}$$

The first term reflects the reducible variance of RMDX predictions by increasing β towards its upper limit of 1 corresponding to full-sample inference. The second term reflects the reducible bias of RMDX predictions by decreasing β as established by Proposition 1. The last term reflects the DGP-implied irreducible variance of latent state predictions. □