Working Paper Series —— 13/2023

A Sparse Kalman Filter: A Non-Recursive Approach

Michal Andrle, Jan Brůha

Czech National Bank —— Working Paper Series —— 13/2023

CZECH NATIONAL BANK

Reviewed by: Alex Gibberd (Lancaster University)
František Brázdik (Czech National Bank)
Project Coordinator: Volha Audzei

A Sparse Kalman Filter: A Non-Recursive Approach

Michal Andrle and Jan Brůha *

Abstract

We propose an algorithm to estimate unobserved states and shocks in a state-space model under sparsity constraints. Many economic models have a linear state-space form – for example, linearized DSGE models, VARs, time-varying VARs, and dynamic factor models. Under the conventional Kalman filter, which is essentially a recursive OLS algorithm, all estimated shocks are non-zero. However, the true shocks are often zero for multiple periods, and non-zero estimates are due to noisy data or ill-conditioning of the model. We show applications where sparsity is the natural solution. Sparsity of filtered shocks is achieved by applying an elastic-net penalty to the least-squares problem and improves statistical efficiency. The algorithm can be adapted for non-convex penalties and for estimates robust to outliers.

Abstrakt

Navrhujeme algoritmus pro odhad nepozorovaných stavů a šoků ve stavovém modelu při omezení na řídkost řešení. Mnohé ekonomické modely mají lineární stavovou formu – např. linearizované modely DSGE, VAR, časově proměnlivé VAR a dynamické faktorové modely. Při použití konvenčního Kalmanova filtru, který je v zásadě rekurzivním OLS algoritmem, jsou všechny odhadované šoky nenulové. Skutečné šoky však jsou často po větší počet období nulové a nenulové odhady jsou důsledkem šumu v datech nebo špatné podmíněnosti modelu. Ukazujeme aplikace, kde je přirozeným řešením řídkost. Řídkost filtrovaných šoků je dosažena aplikací penalizace pomocí metody elastické sítě na problém nejmenších čtverců a zlepšuje statistickou efektivnost. Algoritmus lze upravit pro nekonvexní penalizaci a pro odhady robustní vůči extrémním hodnotám.

JEL Codes: C32, C52, C53.
Keywords: Kalman filter, regularization, sparsity.

*Michal Andrle, International Monetary Fund, 700 19th Street, N.W., Washington, D.C. 20431; mandrle@imf.org.
Jan Brůha, Czech National Bank, Na Příkopě, 28, 115 03 Praha, Czech Republic; jan.bruha@cnb.cz

We thank Volha Audzei, Frantisek Brazdik, Alex Giberd, and seminar participants at the Czech National Bank. We also thank the participants of CFE 2021 for comments. The authors are solely responsible for all errors and omissions. The views expressed herein are those of the authors and should not be attributed to the Czech National Bank, or the International Monetary Fund, its Executive Board, or its management.

1. Introduction

Linear state-space models make up a large class of interesting economic models, such as linearized DSGE models, VARs, trend-cyclical VARs, time-varying VARs, and dynamic factor models. These models relate unobserved shocks or disturbances to observed variables through a set of linear dynamic equations. Under the assumption of linearity, the task of estimating the shocks is a least-squares problem (Whittle, 1983). The well-known Kalman filter is an ingenious algorithm for solving this least-squares problem recursively. The unobserved shocks can therefore be estimated without the Kalman filter using least-squares projection (Kollmann, 2013). The outcomes of the Kalman filter and least-squares projection are equivalent.

Although the fact that the Kalman filter can be replaced by just one – albeit possibly huge – linear regression has been noticed, it seems that the possible advantages of the regression formulation have not been fully realized.

Many structural economic models are terribly ill-conditioned – they are overly sensitive to small changes in data. For example, DSGE models usually add a structural (or "structural") shock each time a new observable is added to avoid stochastic singularity.[1] Hence, the number of unknowns (i.e., shocks) grows at the same pace as the number of observations. In addition, many shocks may have similar impulse responses, which leads to poor shock identification. In such cases, a relatively minor change in the data inputs can lead to a substantial change in the estimated path of the structural shocks and in the interpretation of the data. Supply and demand shocks can alternate quickly and offset each other, and are often strongly cross-correlated. This is equivalent to multicollinearity in the linear regression problem.[2]

Traditional linear least squares can perform poorly when the regressors are correlated and/or the number of unknowns is large relative to the number of observations. In such cases, it is widely recognized that "regularization" of the ill-conditioned problem leads to more reliable outcomes.[3]

We illustrate that regularization techniques can be used to improve estimates of unobserved shocks in state-space models, and we show how to do so. There are many ways to regularize the least-squares problem. One of the most common approaches is to add a suitable penalty term to the least-squares objective function. Depending on the form of the penalty term, the regularized solution may be sparse: only some of the unknowns are non-zero.

There are cases where one *wants* to have a sparse solution based on economic reasoning, beyond efficiency and stability concerns. For example, if one of the unobserved variables is related to a policy target, such as keeping long-term inflation expectations anchored to the central bank's target, it may be the case that such a variable rarely changes. It would change if the official target were to change or, in our case, if monetary policy were to lose credibility. Using the historical data, we

[1] We call a state-space system stochastically singular if the implied matrix of the multivariate spectrum of the observed variables is rank deficient for almost all frequencies where it is defined. This typically happens if the number of shocks to the state equations is lower than the number of observable equations and there is no measurement noise. A notorious example is real business cycle models, where one shock (a productivity shock) drives all the macroeconomic variables (consumption, investment, hours). NK DSGE modelers tend to avoid this by applying specific shocks to each structural equation.

[2] Andrle (2014) describes this phenomenon for the case of structural models.

[3] For example, Theobald (1974) shows that for *any arbitrary* linear regression problem there is a penalty parameter such that the resulting ridge regression achieves a smaller RMSE than OLS; see also van Wieringen (2020).

may then want to see only a few shocks to such a variable. But that is something the Kalman filter cannot do, being simply a least-squares projection. All shocks will be non-zero at all times.

The regression formulation of the filtering problem is immensely powerful. It is possible to use any efficient algorithm that regularizes the least-squares solution and that possibly also achieves sparsity, such as Lasso. It is also straightforward to consider various extensions, such as non-convex penalty functions and alternatives to the least-squares objective to achieve robustness to outliers. The penalty is dictated by the goal of the analysis.

We organize the rest of the paper as follows. The next section 2 introduces the general idea behind the sparse Kalman filter, and section 3 illustrates it on a simple example. Section 4 discusses two extensions: non-convex penalties and robustness. In the last section, we conclude. Appendices contain additional materials, including a small Monte Carlo study that illustrates the usefulness of sparsity for statistical efficiency.

2. Basic Principles

Throughout the paper, we use the terms "filtered" and "smoothed" shocks interchangeably, so that shocks are estimated using the full history of observations.

Let us consider a general linear state-space model:

$$x_t = A_t x_{t-1} + K_t \varepsilon_t \tag{1}$$

$$y_t = C_t x_t + \Omega_t \nu_t, \tag{2}$$

where x_t are unobserved states, ε_t are shocks to the states, y_t are observed variables, and ν_t are measurement errors. In the following, we will assume that the shocks and measurement errors have zero mean and unit variance.[4] By allowing the observation vector y_t to change dimension over time and making the matrices C_t time-varying, we can accommodate arbitrary patterns of missing data, asynchronous time releases, mixed-frequency data, and expert judgment (priors).

By simple but rather tedious substitutions, we can write the observations as a function of the shocks and the initial conditions x_0:

$$y_t = C_t \left(\prod_{i=1}^{t} A_{t-i} \right) x_0 + C_t \left[\sum_{i=1}^{t-1} \left(\prod_{j=0}^{t-i} A_{t-j-1} \right) K_i \varepsilon_i \right] + C_t K_t \varepsilon_t + \Omega_t \nu_t. \tag{3}$$

[4] This can always be achieved by appropriately scaling matrices K_t and Ω_t

By stacking equations (3), we obtain the following system:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}}_{\equiv \mathbf{Y}} = \underbrace{\left[\begin{array}{cccc|cccc} C_1K_1 & 0 & \dots & 0 & \Omega_1 & 0 & \dots & 0 \\ C_2A_2K_1 & C_2K_2 & \dots & 0 & 0 & \Omega_2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ C_T\left(\prod_{j=0}^{T-2} A_{T-j}\right)K_1 & C_T\left(\prod_{j=0}^{T-3} A_{T-j}\right)K_2 & \dots & C_TK_T & 0 & 0 & \dots & \Omega_T \end{array}\right]}_{\equiv \mathbf{A}} \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \\ \hline \nu_1 \\ \nu_2 \\ \vdots \\ \nu_T \end{bmatrix}}_{\equiv \mathbf{E}} + \underbrace{\begin{bmatrix} C_1A_1 \\ C_2A_2A_1 \\ \vdots \\ C_T\prod_{i=0}^{T-1} A_{T-i} \end{bmatrix}}_{\equiv \mathbf{B}} x_0. \tag{4}$$

2.1 Known Initial Conditions

If the initial conditions for the state variables, x_0, are known (or imposed), we can subtract them from the vector of observations to get a system of linear equations:

$$\mathbf{Y} - \mathbf{B}x_0 = \mathbf{A} \times \mathbf{E},$$

which can be solved for unknown shocks $\mathbf{E}$ by the least-squares method:

$$\min_{\mathbf{E}} (\mathbf{Y} - \mathbf{B}x_0 - \mathbf{A}\mathbf{E})^T (\mathbf{Y} - \mathbf{B}x_0 - \mathbf{A}\mathbf{E}) + \wp(\mathbf{E}), \tag{5}$$

where $\wp(\mathbf{E})$ is a suitable penalty term for regularizing the solution. Once the shocks $\mathbf{E}$ are estimated, it is easy to find the unobserved states by iterating on (4) or the state-space form.

As a useful starting point, let us consider an L^1 norm on the shocks to the state equations:

$$\wp(\mathbf{E}) = \lambda \sum_{t=1}^{T} ||\varepsilon_t||_1, \tag{6}$$

while the measurement errors are not penalized.[5] This penalty achieves sparsity, which, as we have argued, can be beneficial from the point of view of statistical efficiency and sometimes also of economic reasoning.

Problem (5) and (6) is basically a **Lasso regression**, which is often used in machine learning, and there are efficient algorithms to solve it. For the computation experiments in our paper, we use the ADMM algorithm (see, for example, chapter 3.5 of Fan et al., 2020).[6]

[5] We can think of them as residuals unexplained by the model.

[6] For many interesting economic applications, we can expect matrix $\mathbf{A}$ in (5) to be ill-conditioned. Coordinate descend algorithms may be rather slow in such a situation. Andrle (2014) uses the singular value decomposition of matrix $\mathbf{A}$ to demonstrate why and how ill-conditioned systems often arise in structural macroeconomic models.

2.2 Unknown Initial Conditions

The initial conditions are rarely known. We can estimate them together with the shocks. In a model with multiple states, the effects of some initial conditions may be similar to the effects of the shocks in the initial periods. That would result in collinearity and the initial conditions might then be imprecisely estimated.[7] Regularized estimation can be helpful in such a case. For the initial conditions, we do not, however, propose a Lasso-type penalty.

In fact, the researcher usually has at least a rough idea of what the initial conditions might be. Let $\tilde{x}_0$ denote such an idea. Then, we propose to estimate the shocks and initial conditions as a solution to the problem:

$$\min_{\mathbf{E},\xi_0}(\mathbf{Y}-\mathbf{B}\tilde{x}_0-\mathbf{B}\xi_0\mathbf{A}-\mathbf{E})^T(\mathbf{Y}-\mathbf{B}\tilde{x}_0-\mathbf{B}\xi_0-\mathbf{A}\mathbf{E})+\wp(\mathbf{E})+(\tilde{x}_0-\xi_0)^T\mathscr{W}(\tilde{x}_0-\xi_0), \tag{7}$$

for a suitable positive semidefinite matrix $\mathscr{W}$. (7) is an **elastic-net regression** and the initial conditions can be found using the trivial substitution $x_0=\tilde{x}_0+\xi_0$.

Matrix $\mathscr{W}$ can reflect the desired degree of confidence in the a priori knowledge about the initial conditions. Our experience shows that even modest regularization of the initial conditions can greatly improve the stability of the filtration.

2.3 Residua of Filtration

Unless the underlying system is stochastically singular, the conventional Kalman filter decomposes all the movements in the observation variables y_t, i.e., there is nothing unexplained. If the number of states is lower than the number of observables (as, for example, in dynamic factor models), measurement noise can be added to prevent the system from being stochastically singular. This is an analogous situation to OLS. If the number of linearly independent explanatory variables is equal to or greater than the number of observations, there are no residua: the fit is perfect.

On the other hand, the sparse filter may not explain all the movements in the observable variables, regardless of the number of shocks. This is much the same as in regression models with lasso-type penalties: even in systems with more explanatory variables than observations,[8] some variance in the explained variable remains unexplained. From the statistical point of view, this is exactly what guards the model from being overfitted.

3. Application: A Trend-Cycle Decomposition of Euro Area Inflation

In this part of the paper, we apply the proposed approach to a simple univariate model for a decomposition of the euro area quarterly inflation rate. No doubt there are more elaborate models for inflation. Nevertheless, the simplicity of the model serves well to illustrate the ideas proposed in this paper.

[7] Researchers working with non-trivial state-space models have probably noticed many times that the initial conditions sometimes exhibit high sensitivity to data changes and that the estimated shocks display erratic behavior in the initial periods, offsetting each other. This is precisely a symptom of this collinearity problem.

[8] Such regressions naturally arise, for instance, in genetic studies, when an expansion in the number of observations (for example, the number of people whose genes are being investigated) is accompanied by an expansion in the number of explanatory variables (the number of different sequences of genes).

We decompose inflation π_t into three components:

$$\pi_t = \bar{\pi}_t + \hat{\pi}_t + \dot{\pi}_t. \tag{8}$$

Here, $\bar{\pi}_t$ is the trend component modeled as a random walk, $\hat{\pi}_t$ is the cyclical component modeled as a stationary AR(2) process, and $\dot{\pi}_t$ is the high-frequency component modeled as an invertible MA(1) process. Andrle and Brůha (2017) use a similar decomposition of inflation, and the cyclical component co-moves well with the cyclical component of output and unemployment.

The state equation reads as:

$$\begin{bmatrix} \bar{\pi}_t \\ \hat{\pi}_t \\ \hat{\pi}_{t-1} \\ \dot{\pi}_t \\ \dot{\pi}_{t-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \alpha_1 & \alpha_2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \bar{\pi}_{t-1} \\ \hat{\pi}_{t-1} \\ \hat{\pi}_{t-2} \\ \dot{\pi}_{t-1} \\ \dot{\pi}_{t-2} \end{bmatrix} + \begin{bmatrix} \bar{\sigma} & 0 & 0 \\ 0 & \hat{\sigma} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \dot{\sigma} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{\varepsilon}_t \\ \hat{\varepsilon}_t \\ \dot{\varepsilon}_t \end{bmatrix}, \tag{9}$$

where $\bar{\varepsilon}_t$ is the shock to the trend component, $\hat{\varepsilon}_t$ is the shock to the cyclical component, $\dot{\varepsilon}_t$ is the shock to the high-frequency component, $\bar{\sigma}$, $\hat{\sigma}$, and $\dot{\sigma}$ are the corresponding standard deviations, and α_1 and α_2 are the coefficients of the AR(2) process that captures the cyclical dynamics.

The observation equation is as follows:

$$\pi_t = \begin{bmatrix} 1 & 1 & 0 & 1 & \vartheta \end{bmatrix} \begin{bmatrix} \bar{\pi}_t \\ \hat{\pi}_t \\ \hat{\pi}_{t-1} \\ \dot{\pi}_t \\ \dot{\pi}_{t-1} \end{bmatrix}, \tag{10}$$

where ϑ is the parameter of the MA process that captures the dynamics of the high-frequency component.

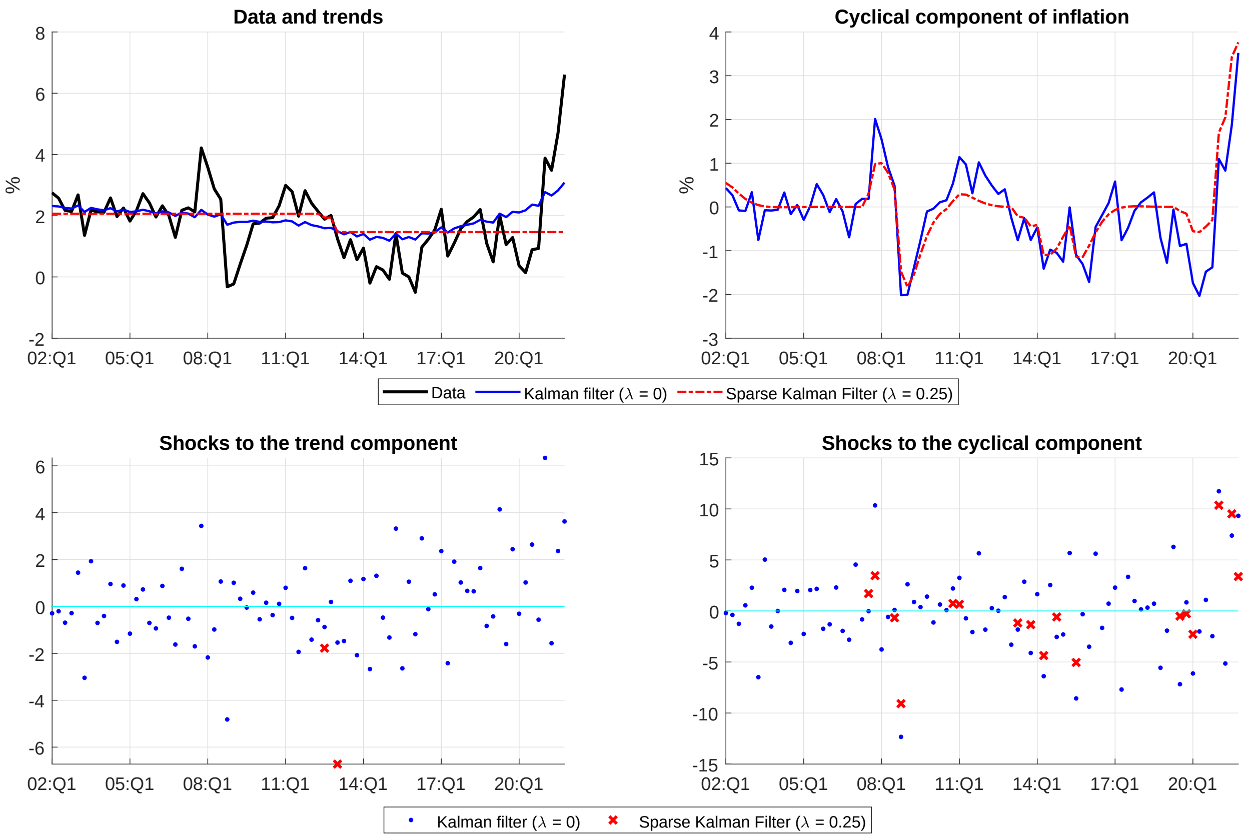
The model is calibrated to quarterly euro area data from 2000Q1 to 2021Q4. The resulting parameter values are as follows: $\alpha_1 = 1.14$, $\alpha_2 = -0.37$, $\vartheta = -0.24$, $\bar{\sigma} = 0.0704$, $\hat{\sigma} = 0.1810$, and $\dot{\sigma} = 0.045$.[9]

The effects of the sparsity imposed by $\lambda > 0$ in (6) on the estimation of the trend, the cycle, and the shocks is displayed in Figure 1. The upper left chart displays the data, the conventional Kalman filter trend, and the trend implied by the sparse filter with $\lambda = 0.25$. The upper right chart then compares the cycles implied by the two filters. The lower charts display the non-zero shocks to the trend component (lower left) and the cyclical component (lower right) from the two filters.

The conventional Kalman filter identifies non-zero shocks to all components in all periods. This is the implication of the least-squares nature of the Kalman filter. It is also apparent that many of those shocks offset each other: this is true not only within equations, but also across equations. As a result, the filtered trend generated by the conventional Kalman filter is an incessantly varying series, with many small shocks moving it up and down. The cyclical part is also incessantly affected by small shocks, some of them canceling each other out in subsequent periods.

[9] It is common to specify the cyclical components as an AR(2) process. Andrle and Plasil (2017) show how using *system priors* to estimate the parameters helps guarantee stationarity, concentration of the variance at business-cycle frequencies, and an absence of secondary cycles.

Figure 1: Filtration and Sparse Filtration of Euro Area Inflation (q/q in %)



On the other hand, the sparse filter identifies just a few shocks to the trend and a small number of shocks to the cyclical part of the model. As a result, the inflation trend is piece-wise constant. Before 2014, this trend seems firmly anchored at 2%. In 2012 and 2013, there are two negative shocks to the trend that together cause the trend to drop by 0.6 p.p. The timing of this drop corresponds to outside evidence of a fall in long-run inflation expectations around this time (Busetti et al., 2014; Corsello et al., 2021).

The sparse filter also identifies fewer shocks to the cycle. There are positive shocks prior to 2008 and negative cyclical shocks in 2009 corresponding to the cooling of the euro area economy after the outbreak of the Great Recession. The negative cyclical shocks to inflation in our simple setup proxy the shocks stemming from the decline in output in the euro area, the decline in global output, and the associated decline in commodity prices. Of course, a univariate model cannot separate demand from supply shocks. Further, there are negative cyclical shocks during the euro area sovereign debt crisis and negative shocks around the outbreak of the covid-19 pandemic, while positive inflation shocks appear later during the pandemic.

How should one choose the penalty parameter λ? This is an important issue. In practical empirical work with penalized regression, cross-validation is a popular approach. It can be used here as well. One can run the filter with various missing observations,[10] use a smoother to compute those missing observations, and then compute statistics such as the mean average error or the root mean square error to use for cross-validation. As we are in the time series context, one way of finding a suitable

[10] It is extremely easy to run the sparse filter with missing observations. It is sufficient just to delete the rows in (4) that correspond to the missing observations.

value of λ is based on the prediction errors $\varepsilon_T^h(\lambda) \equiv y_{T+h|T}(\lambda) - y_{T+h}$ for various T and h, where $y_{T+h|T}$ is the prediction of y_{T+h} based on the sample ending at T using the filter with regularization parameter λ.[11] The choice of approach is dictated by the intended application.

The reader may ask about the possibility of **estimating** the parameters of the state-space model under the proposed approach. Our approach does not assume any functional form of the residua, hence we are silent on likelihood-based methods. Nevertheless, one can use likelihood-free techniques (such as EM algorithms or matching moments) without any difficulty under sparse and/or robust filtering.

To demonstrate these approaches on our example, the following Table 1 displays the mean absolute error for various values of λ (including $\lambda = 0$, which corresponds to the standard Kalman filter) and various validation schemes. The column CV presents the mean absolute error for 99 repetitions of the cross-validation exercise: for each repetition, 20% of the observations are randomly deleted and the filters must extrapolate them. The remaining columns display the mean average forecast error $\frac{1}{40}\sum_{t=41}^{80}\left|\varepsilon_t^h\right|$ for $h = 1, \ldots 4$. As in machine learning applications, penalization can have a beneficial effect on predictive performance.

Table 1: Mean Absolute Error for Various Penalty Parameters

		Forecasting Horizon			
Penalty parameter	CV	1	2	3	4
$\lambda = 0$	2.9862	1.1074	1.1140	1.0875	1.0668
$\lambda = 0.05$	2.8005	1.0629	1.0972	1.1087	1.1382
$\lambda = 0.10$	2.7041	1.0622	1.1063	1.1405	1.1775
$\lambda = 0.25$	2.8597	1.0349	1.0827	1.1160	1.1354
$\lambda = 0.50$	3.4076	1.0849	1.1140	1.1319	1.1415
$\lambda = 0.75$	4.0028	1.1075	1.1203	1.1275	1.1310

4. Extensions: Robustness and Alternative Penalties

The sparse filter can be extended in many dimensions. Below, we give examples of alternative regularization schemes and show how to make the filter robust to outliers.

4.1 Alternative Regularization Penalties

L^1 (6) set some of the shocks to zero and also made the non-zero ones shrink towards zero. This shrinkage towards zero is a general property of convex penalties L^q for $q \geq 1$.

It may sometimes be preferable to set some of the shocks to zero but not shrink the remaining ones. In that case, one has to use non-convex penalties. One of the most commonly used is the Smoothed Clipped Absolute Deviation (SCAD) penalty proposed by Fan and Li (2001). It is defined

[11] It is also very easy to compute this prediction once the Kalman filter or the sparse filter has been use to estimate the shocks in the state equations.

as follows:

$$\mathscr{C}_{\lambda,\alpha}(x) = \begin{cases} \lambda|x| & \text{if} \quad |x| \leq \lambda \\ -\frac{x^2+2\alpha\lambda+\lambda^2}{2(\alpha-1)} & \text{if} \quad \lambda < |x| \leq \alpha\lambda \\ 0.5(\alpha+1)\lambda^2 & \text{if} \quad |x| > \alpha\lambda \end{cases}. \tag{11}$$

The penalty depends on two parameters $\lambda > 0$ and $\alpha > 1$. The first determines the degree of shrinkage for small values, while the second determines the region of no penalization.

Solving the optimization problem with the SCAD penalty is more difficult than solving it with the Lasso penalty. This is because (11) is not convex. Fortunately, the optimization problem with the SCAD penalty can be approximated using iterated weighted Lasso problems (see, for example, chapter 2.8.6 of Bühlmann and Geer, 2011) as follows:

$$\mathbf{E}^{[k]} = \min_{\mathbf{E}}(\mathbf{Y} - \mathbf{B}x_0 - \mathbf{AE})^T(\mathbf{Y} - \mathbf{B}x_0 - \mathbf{AE}) + \sum_{t=1}^{T}\sum_{j} w_{j,t}^{[k]}|\varepsilon_{j,t}|, \tag{12}$$

where the weights in the kth iteration are given as the derivative of (11):

$$w_{j,t}^{[k]} = \mathscr{C}'_{\lambda,\alpha}\left(\varepsilon_{j,t}^{[k-1]}\right) = \begin{cases} \lambda & \text{if} \quad |\varepsilon_{j,t}^{[k-1]}| \leq \lambda \\ \frac{\alpha\lambda - |\varepsilon_{j,t}^{[k-1]}|}{\alpha-1} & \text{if} \quad \lambda < |\varepsilon_{j,t}^{[k-1]}| \leq \alpha\lambda \\ 0 & \text{if} \quad |\varepsilon_{j,t}^{[k-1]}| > \alpha\lambda \end{cases}, \tag{13}$$

where $\varepsilon_{j,t}^{[k-1]}$ is the filtered jth shock at time t obtained in the $(k-1)$th iteration. In our experiments, we use the results from the Lasso penalty as the starting point for the iteration exercise: $\varepsilon_{j,t}^{[0]} = \varepsilon_{j,t}^{Lasso}$.

4.2 Robustness to Outliers

Another useful extension is to make the filter robust to outliers. As the Kalman filter solves the least-squares problem, it is sensitive to unusual observations – to outliers. It may be good to have a filter robust to such observations. The regression formulation helps us here too.

In the regression context, a popular approach to robust regression is the method of iteratively reweighted least squares. The method is based on repeated running of weighted least squares, where observations with large residuals receive small or zero weights (Street et al., 1988). This can be directly applied here. To achieve robustness, the filter is computed by iteration on the following problem:

$$\min_{\mathbf{E},\xi_0}(\mathbf{Y} - \mathbf{B}\tilde{x}_0 - \mathbf{B}\xi_0\mathbf{A} - \mathbf{E})^T\mathbb{W}(\mathbf{Y} - \mathbf{B}\tilde{x}_0 - \mathbf{B}\xi_0 - \mathbf{AE}) + \wp(\mathbf{E}) + (\tilde{x}_0 - \xi_0)^T\mathscr{W}(\tilde{x}_0 - \xi_0), \tag{14}$$

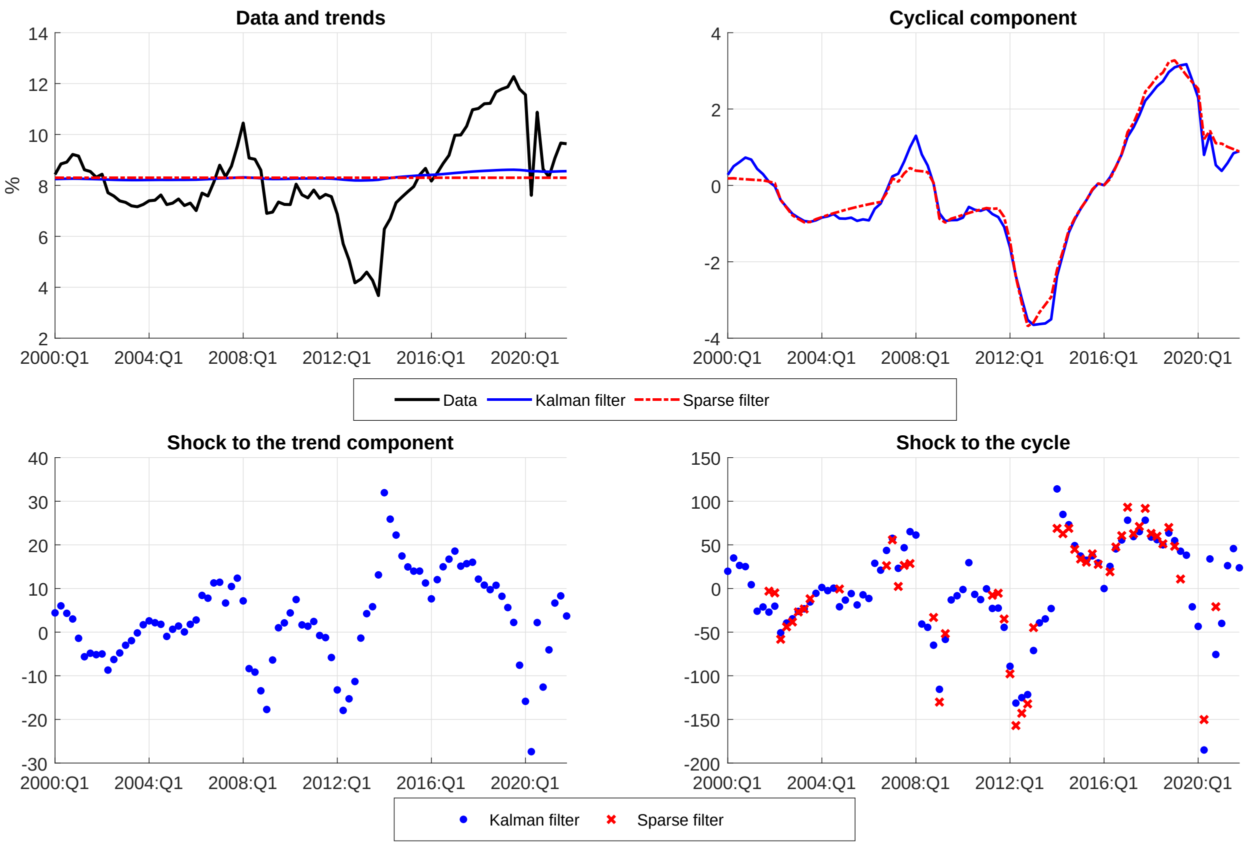
where $\mathbb{W}$ is the weighted matrix. In the initial iteration, $\mathbb{W}$ is equal to the identity matrix. In later iterations, it is a diagonal matrix, with the diagonal entries being the weights inversely related to the magnitude of the residua. This idea can be applied both to the conventional Kalman filter and to the sparse Kalman filter with various penalties.

4.3 Application: Robust Trend-Cyclical Decomposition of the Job Finding Rate

In this part of the paper, we apply the robust (and sparse) Kalman filter to the time series of the job finding rate of the registered unemployed in the Czech Republic. This time series is a nice cyclical indicator and hence is of interest to policymakers. It exhibits model cyclical behavior, with the exception of the first two quarters of 2020, when it shows a down and up movement apparently not related to the underlying trend or cycle.

We illustrate robustness using a simple model to decompose this time series into its trend, cycle, and high-frequency components. The trend is modeled as a random walk, the cycle is a stationary AR process, and the high-frequency noise is an iid shock. Figure 2 displays the Kalman and sparse Kalman decomposition using this model. The figure is organized similarly to Figure 1. As in the previous example, the sparse filter needs far fewer shocks to explain the trend and the cycle. For sparse filtering, the trend is basically the long-run mean.

Figure 2: Filtration and Sparse Filtration of the Czech Job Finding Rate



Both filters, however, are affected by the erratic behavior of the series at the beginning of the pandemic in 2020. The conventional Kalman filter attributes the one-off fall in the job finding rate both to the trend and to the cycle. We can see large negative shocks to both components. The sparse filter attributes it to the cyclical component only. However, it is quite unlikely that this movement is of a trend or a cyclical nature. It is in fact an unusual observation probably caused by the extraordinary situation at the outbreak of covid-19, when both the government and private agents were just starting to learn how to live with the pandemic.

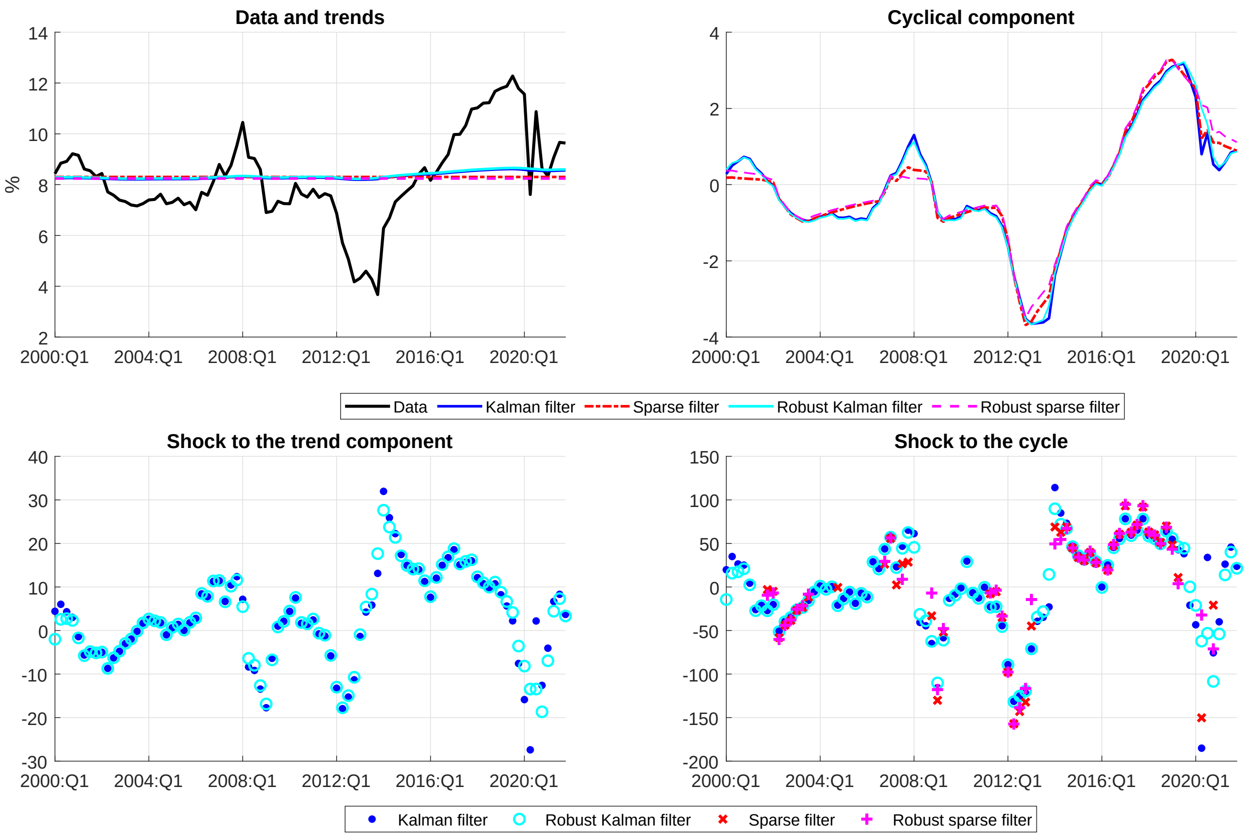
In such a situation, we want to use a *robust* version of these filters. For this experiment, we use a biquadratic weight function that puts weights on each observation as follows:

$$w_i \propto \max\left(0, 1 - \left((8m)^{-1} r_i\right)^2\right)^2,$$

where r_i is the residuum of the ith observation and m is the mean absolute deviation of all the residua.

Figure 3 shows the effect of the robustness on both the conventional Kalman filter and on its sparse variant. The shocks from the robust Kalman filter correspond to the shocks from the standard Kalman filter, except for early 2020. The robust Kalman filter does not propagate the outlier to the trend or to the cycle. The robust sparse filter finds no shock to the trend and hence the trend is constant, as in the case of the non-robust version of the sparse filter. However, the robust sparse filter – unlike the non-robust one – does not propagate the outlier to the cycle, either. Except for this episode, the robust and non-robust sparse shocks are very similar to each other.

Figure 3: Robust Filtration of the Czech Job Finding Rate



5. Conclusion

In this paper, we showed that the non-recursive, regression formulation of the Kalman filter can be useful, as it enables the application of wide range of modern machine learning methods that are used to achieve robustness and sparsity. This is often helpful for many real-world economic problems from both the statistical point of view (more precise estimates of shocks) and the economic point of view (interpretation can be simpler). One can go beyond the applications in this paper, for example, by using non-convex penalties instead of the lasso-type penalty.

While the idea of sparse filtering is not new, it has not been applied to many economic problems. The trend filtering by Tibshirani (2014) is an exciting approach for filtering univariate data, and a fast and precise algorithm was proposed by Ramdas and Tibshirani (2016). However, extending trend filtering to the multivariate setting is not very straightforward. The approach described in this paper can easily be applied to multivariate problems.

We also suggested a robust extension to the Kalman filter. One of the earliest methods for achieving robustness was based on the wonderful idea of the Gaussian sum filter (Alspach and Sorenson, 1972). This filter assumes that the noise in the observation equation is a mixture of two Gaussian distributions: the first is one with reasonable standard errors and applies to most observations, and the second is one with large standard errors and contains model outliers. However, running the Gaussian sum filter involves numerical integration, which is usually done by performing simulations and/or pruning past states. Numerical integration along with various approximations in the state update step is a typical feature of filters with heavy-tailed distributions in the measurement equation. Iterative least-squares reweighing can be a way of avoiding this.

To conclude, we propose that our regression formulation of the Kalman filter makes it possible to reap the benefits of the sparsity tools suggested by the modern literature on machine learning and high-dimensional statistics.

References

ALSPACH, D. AND H. SORENSON (1972): "Nonlinear Bayesian Estimation using Gaussian Sum Approximations." *IEEE Transactions on Automatic Control,* 17(4):439–448.

ANDRLE, M. (2014): "Estimating Structural Shocks with DSGE Models." Manuscript available at http://michalandrle.weebly.com/uploads/1/3/9/2/13921270/opfilter.pdf.

ANDRLE, M. (2013): "Understanding DSGE Filters in Forecasting and Policy Analysis." IMF Working Papers 2013/098, International Monetary Fund.

ANDRLE, M. AND J. BRŮHA (2017): "Forecasting and Policy Analysis with Trend-Cycle Bayesian VARs." Manuscript available at https://michalandrle.weebly.com/uploads/1/3/9/2/13921270/tc_ vars.pdf.

ANDRLE, M. AND M. PLASIL (2017): "System Priors for Econometric Time Series." Working Papers 2017/01, Czech National Bank.

BÜHLMANN, P. AND S. A. V. D. GEER (2011): *Statistics for High-Dimensional Data: Methods, Theory and Applications.* Springer.

BUSETTI, F., G. FERRERO, A. GERALI, AND A. LOCARNO (2014): "Deflationary Shocks and De-anchoring of Inflation Expectations." Questioni di Economia e Finanza (Occasional Papers) 252, Bank of Italy, Economic Research and International Relations Area.

CORSELLO, F., S. NERI, AND A. TAGLIABRACCI (2021): "Anchored or De-anchored? That is the Question." *European Journal of Political Economy,* 69(C).

FAN, J. AND R. LI (2001): "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties." *Journal of the American Statistical Association,* 96(456):1348–1360.

FAN, J., R. LI, C.-H. ZHANG, AND H. ZOU (2020): *Statistical Foundations of Data Science.* CRC Press.

KOLLMANN, R. (2013): "Estimating the State Vector of Linearized DSGE Models without the Kalman Filter." *Economics Letters,* 120(1):65–66.

KOOPMAN, S. J. AND A. HARVEY (2003): "Computing Observation Weights for Signal Extraction and Filtering." *Journal of Economic Dynamics and Control,* 27:1317–1333.

RAMDAS, A. AND R. J. TIBSHIRANI (2016): "Fast and Flexible ADMM Algorithms for Trend Filtering." *Journal of Computational and Graphical Statistics,* 25(3):839–858.

STREET, J., R. CARROLL, AND D. RUPPERT (1988): "A Note on Computing Robust Regression Estimates Via Iteratively Reweighted Least Squares." *The American Statistician,* 42(2): 152–54.

THEOBALD, C. (1974): "Generalizations of Mean Square Error Applied to Ridge Regression." *Journal of the Royal Statistical Society. Series B (Methodological),* 36(1):103–106.

TIBSHIRANI, R. J. (2014): "Adaptive Piecewise Polynomial Estimation via Trend Filtering." *The Annals of Statistics,* 42(1).

VAN WIERINGEN, W. N. (2020): "Lecture Notes on Ridge Regression."

WHITTLE, P. (1983): *Prediction and Regulation by Linear Least-Square Methods.* University Of Minnesota Press.

Appendix A: Additional Results

A.1 Filter Stability

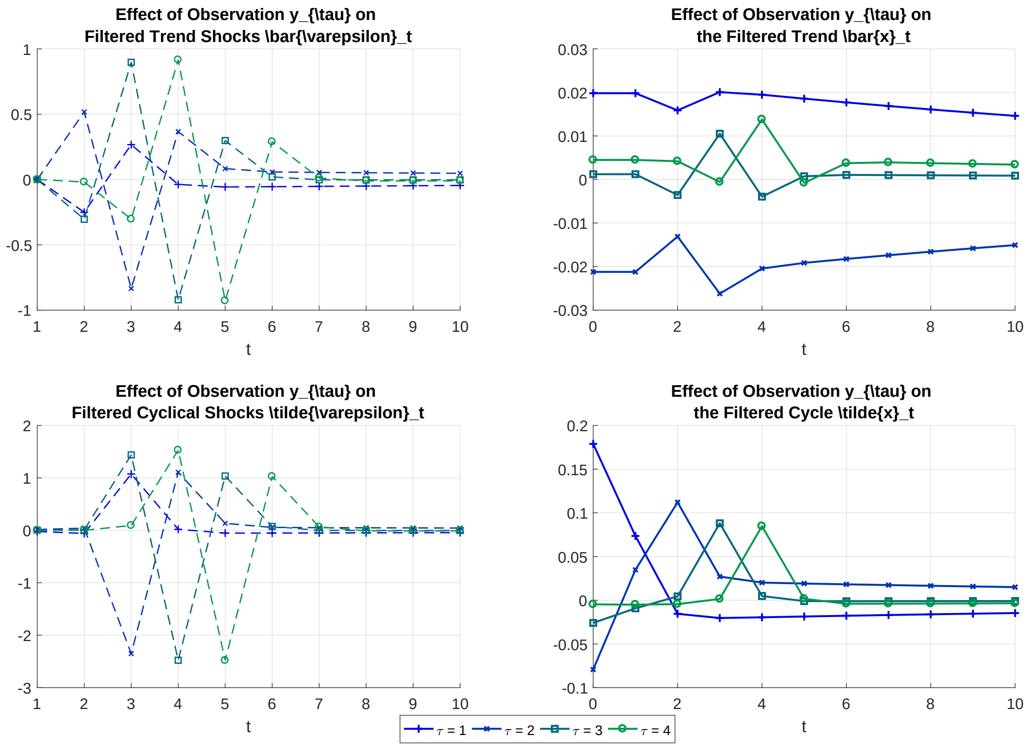
The following figures show the partial effects of a change in observation y_τ on the filtered shocks and states. We consider what would happen to the filtration if y_τ was increased by 0.1 p.p. at $\tau = 1 \dots 4$. The filtered shocks ε_t (and hence states x_t) are functions of the observations:

$$\varepsilon_t = f_t(y_1, y_2 \dots y_T).$$

For the standard Kalman filter, functions f_t are obviously linear[12] (but still time-varying), but for the sparse Kalman filter with the Lasso penalty, they are non-linear.[13]

Figure A1 reports $\Delta_\tau \varepsilon_t = f_t(y_1, y_2 \dots y_\tau + 0.1, \dots y_T) - f_t(y_1, y_2 \dots y_\tau, \dots y_T)$ for the standard Kalman filter for the shocks to the trend $\bar{\varepsilon}_t$ (the upper left chart) and the shocks to the cycle $\tilde{\varepsilon}_t$ (the lower left chart). The right-hand charts then show the effect of the change in observations on the filtered states – this effect is a combination of the effects on the shocks and those on the initial conditions.

Figure A1: Effect of a Change in Observations on Shocks and States: Standard Kalman Filter



The figure reveals the roaming-like behavior of the impacts of the observations on the filtration, especially for early observations. For example, a positive increase in observations in period $\tau = 1$ increases the filtered shock at time $t = 1$ and reduces it at time $t = 2$ and the overall impact of the trend estimation is positive. On the other hand, a positive increase in the observations in period

[12] Koopman and Harvey (2003); Andrle (2013)

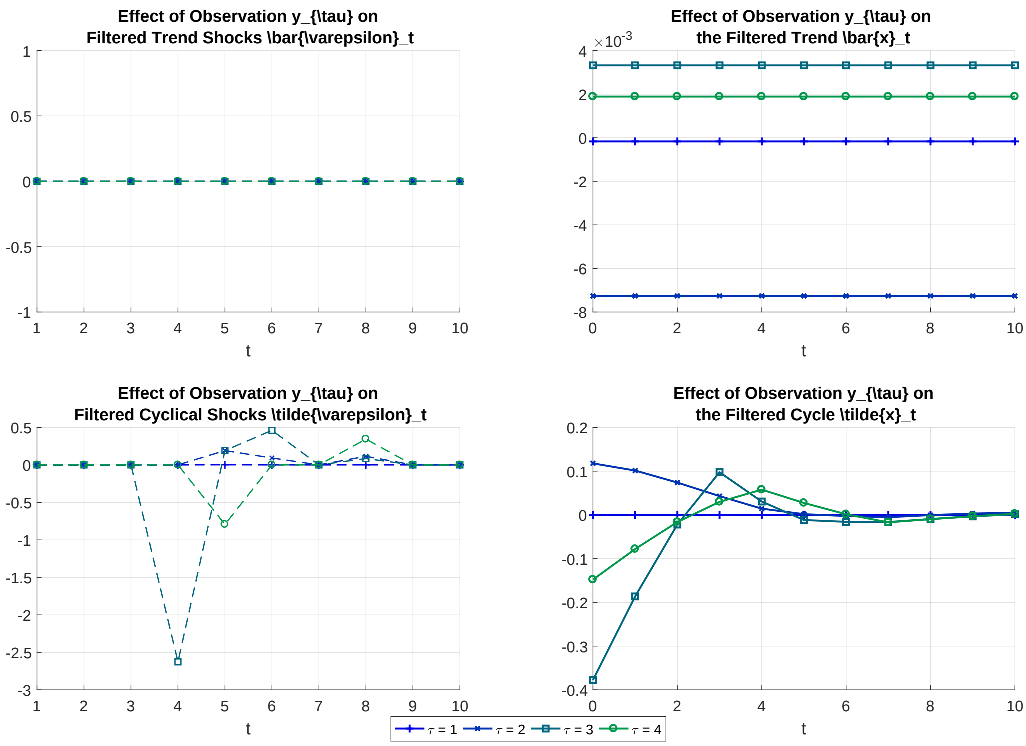
[13] If the ridge penalty were used instead of the Lasso penalty, linearity would be obtained, as in the case of the standard Kalman filter.

$\tau = 2$ increases the estimate of the filtered shock at times $t = 2$ and $t = 4$ and reduces it at time $t = 3$. The whole trajectory of the estimated trend component is (maybe surprisingly) reduced. This is mirrored by a monotonous increase in the estimated cycle (the lower right subchart), and the effect on the filtered shocks to the cycles is even more roaming (the lower left subchart).

Hence, the Kalman filter outcomes are rather sensitive to minor changes in the data. Note that because the functions that relate the shocks and states to the observations are linear for the standard Kalman filter, this behavior does not depend on the data: it would be present whatever data were fed into the filter. Why does this erratic behavior happen? Because of the collinearity of matrix **A** in (4). The effects of $\bar{\varepsilon}_1$ and $\bar{\varepsilon}_2$ on y_t are the same for $t \geq 2$.

Figure A2 is the analogy of Figure A1 for the sparse Kalman filter. Here, the linearity of functions f_t breaks down and depends on the data. The results in the figure are for the data used in section 3.

Figure A2: Effect of a Change in Observations on Shocks and States: Sparse Kalman Filter



Appendix B: Simulation Experiments

To highlight the performance of the sparse filter, we perform two simulation experiments.

In **the first experiment**, we use the simple unobserved component model for a univariate series $\{x_t\}_{t=1}^{T}$. The series is composed of a trend, a random walk, component $\bar{x}_t$, and a cyclical component $\widehat{x}_t$ that follows an AR(2) process:

$$
\begin{aligned}
x_t &= \bar{x}_t + \widehat{x}_t, \\
\bar{x}_t &= \bar{x}_{t-1} + \bar{\sigma}\bar{\varepsilon}_t, \\
\widehat{x}_t &= \alpha_1 \widehat{x}_{t-1} + \alpha_2 \widehat{x}_{t-2} + \widehat{\sigma}\widehat{\varepsilon}_t,
\end{aligned}
$$

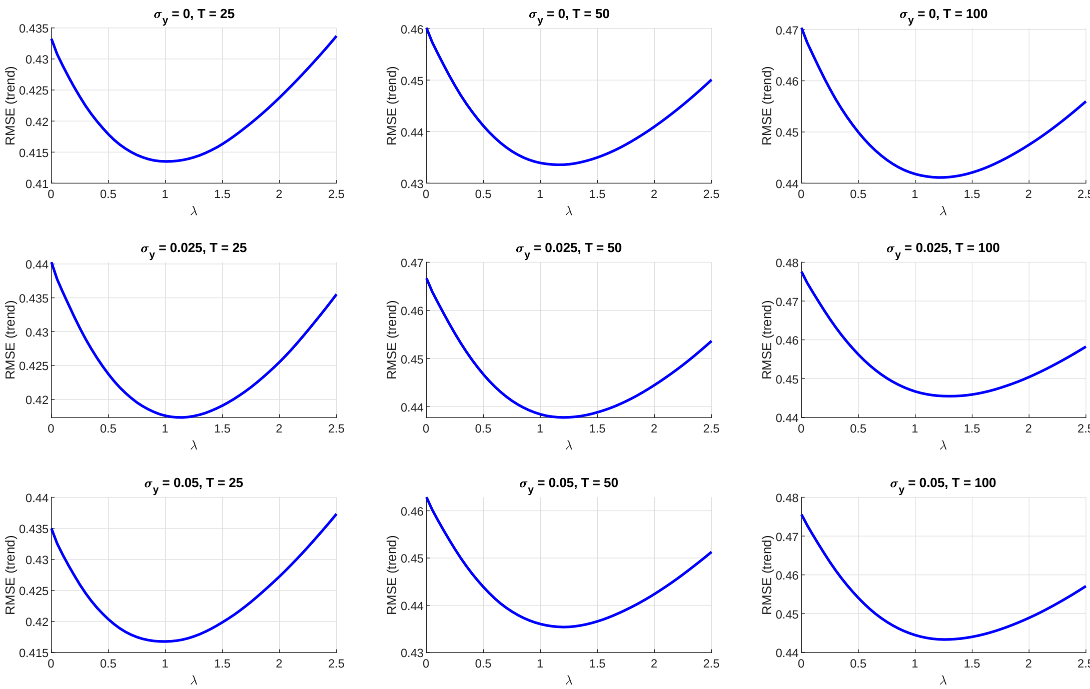
and one observes the series $\{y_t\}_{t=1}^{T}$:

$$y_t = x_t + \sigma_y \nu_t.$$

The shocks $\bar{\varepsilon}_t$, $\widehat{\varepsilon}_t$, and ν_t are uncorrelated iid zero-mean processes with standard deviations equal to 1. We set the following parameters for the experiment: $\alpha_1 = 1.4$, $\alpha_2 = -0.8$, $\bar{\sigma} = 0.2$, $\widehat{\sigma} = 0.5$. All the parameters are known, as are the initial conditions. In the experiment, we vary the length of the observed time series $T \in [20, 50, 100]$ and the observation noise $\sigma_y \in [0, 0.025, 0.05]$.

We repeat each setting of T and σ_y for 1,000 random draws and compute the RMSE of the filtration of the trend component $\bar{x}_t$ over a grid of regularization parameters $\lambda \in [0, 0.05, 0.1\, 0.15 \ldots 2.5]$. The mean RMSEs over the 1,000 replications as a function of λ are displayed in Figure B1.

Figure B1: RMSE as a Function of the Regularization Parameter (Unobserved Component Model)



Apparently, the lowest RMSEs are achieved for positive values of the regularization parameter, i.e., the shrinkage helps. Moreover, the value of the regularization parameter that achieves the lowest

RMSE is increasing in both the length of the observations T and the variance of the measurement noise σ_y. The latter effect is intuitive: higher variance of the measurement noise makes the trend more difficult to identify. Why is the value of the regularization parameter increasing in T? The fact is that for a time series of observations T, the number of states is $2T$. This means that the number of "unknowns" grows faster than the number of observations (in the same way that adding one more observation to a regression model adds two more regressors). That is why the regularization becomes more important.

The second experiment is centered on a dynamic factor model. We assume that a hidden factor f_t follows a stationary AR(2) model:

$$f_t = \alpha_1 f_{t-1} + \alpha_2 f_{t-2} + \sigma_f \varepsilon_t,$$

and one observes ten time series linked to the unobserved factor as follows:

$$y_{it} = \gamma_{i0} f_t + \gamma_{i1} f_{t-1} + \gamma_{i2} f_{t-2} + \xi_{it},$$

where xi_{it} are independent AR(1) processes:

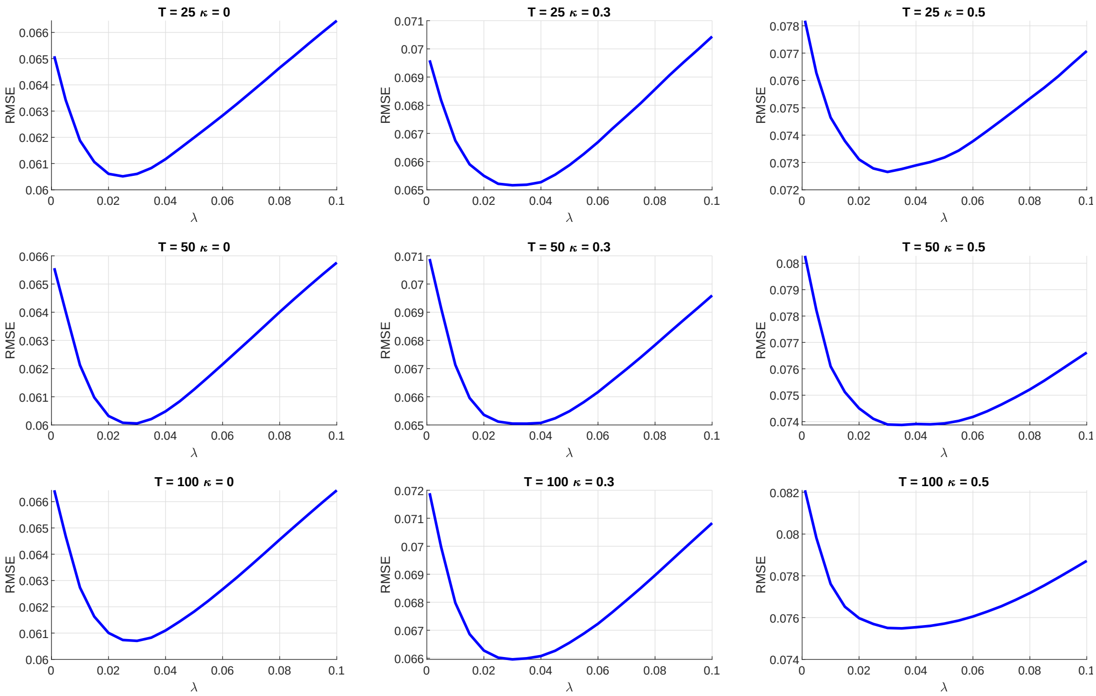
$$\xi_{it} = \kappa_i \xi_{it-1} + \sigma_i \nu_i.$$

The parameters of the experiment are set as follows: $\alpha_1 = 1.3$, $\alpha_2 = -0.8$, $\sigma_f = 1$. For each replication of the experiment, the loadings are drawn from the normal distribution $\gamma_{i0} \sim N(0,1)$, $\gamma_{i1} \sim N(0,0.5)$, $\gamma_{i2} \sim N(0,0.25)$. σ_i follows the inverse Gamma distribution: $\sigma_i^{-1} \sim \Gamma(5,1)$. The persistence of the idiosyncratic parts is drawn from the uniform distribution $\kappa_i \sim \mathscr{U}(\kappa, \kappa+0.2)$. In the experiment, we vary the length of the observations $T \in [20, 50, 100]$ and parameter $\kappa \in [0, 0.3, 0.5]$.

For each setting of T and κ, we perform 100 experiments and evaluate the RMSE for the common factor f_t for the regularization parameter λ over a grid. The mean RMSEs as a function of λ are displayed in Figure B2:

Again, for this application, we see that the value of the regularization parameter that achieves the minimum RMSE is non-zero. This value is increasing in both the number of observations T (the same effect as in the previous experiment) and in the value of κ, i.e., the persistence of the idiosyncratic parts of the series.

Figure B2: RMSE as a Function of the Regularization Parameter (Dynamic Factor Model)



CNB Working Paper Series (since 2022)

WP 13/2023	Michal Andrle Jan Brůha	*A sparse Kalman filter: A non-recursive approach*
WP 12/2023	Zuzana Gric Jan Janků Simona Malovaná	*What drives sectoral differences in currency derivate usage in a small open economy? Evidence from supervisory data*
WP 11/2023	Dominika Ehrenbergerová Simona Malovaná Caterina Mendicino	*How do climate policies affect holdings of green and brown firms' securities?*
WP 10/2023	Josef Bajzík	*Does shareholder activism have a long-lasting impact on company value? A meta-analysis*
WP 9/2023	Jan Brůha Hana Brůhová Foltýnová	*Long-term impacts of the COVID-19 pandemic on working from home and online shopping: Evidence from a Czech panel survey*
WP 8/2023	František Brázdik Karel Musil Stanislav Tvrz	*Implementing yield curve control measures into the CNB core forecasting model*
WP 7/2023	Alexis Derviz	*Foreign exchange implications of CBDCs and their integration via bridge coins*
WP 6/2023	Simona Malovaná Dominika Ehrenbergerová Zuzana Gric	*What do economists think about the green transition? Exploring the impact of environmental awareness*
WP 5/2023	Milan Szabo	*Cyclical investment behavior of investment funds: Its heterogeneity and drivers*
WP 4/2023	Monika Junicke Jakub Matějů Haroon Mumtaz Angeliki Theophilopoulou	*Distributional effects of monetary policy shocks on wage and hours worked: Evidence from the Czech labor market*
WP 3/2023	Simona Malovaná Jan Janků Martin Hodula	*Macroprudential policy and income inequality: The trade-off between crisis prevention and credit redistribution*
WP 2/2023	Michal Franta	*The Application of multiple-output quantile regression on the US financial cycle*
WP 1/2023	Martin Veselý	*Finding the optimal currency composition of foreign exchange reserves with a quantum computer*
WP 10/2022	Martin Hodula Milan Szabo Josef Bajzík	*Retail fund flows and performance: Insights from supervisory data*
WP 9/2022	Jiří Gregor Jan Janků Martin Melecký	*From central counter to local living: Pass-through of monetary policy to mortgage lending rates in districts*

WP 8/2022	Simona Malovaná Martin Hodula Zuzana Gric Josef Bajzík	*Borrower-based macroprudential measures and credit growth: How biased is the existing literature?*
WP 7/2022	Martin Časta	*How credit improves the exchange rate forecast*
WP 6/2022	Milan Szabo	*Meeting investor outflows in Czech bond and equity funds: Horizontal or vertical?*
WP 5/2022	Róbert Ambriško	*Nowcasting macroeconomic variables using high-frequency fiscal data*
WP 4/2022	Jaromír Baxa Jan Žáček	*Monetary policy and the financial cycle: International evidence*
WP 3/2022	Martin Hodula Milan Szabo Lukáš Pfeifer Martin Melecký	*Cooling the mortgage loan market: The effect of recommended borrower-based limits on new mortgage lending*
WP 2/2022	Martin Veselý	*Application of quantum computers in foreign exchange reserves management*
WP 1/2022	Vojtěch Molnár	*Price level targeting with imperfect rationality: A heuristic approach*

CNB Research and Policy Notes (since 2022)

RPN 2/2023	Eva Hromádková Ivana Kubicová Branislav Saxa	*How does interest rate pass-through change over time? Rolling windows and the role of the credit risk premium in the pricing of Czech loans*
RPN 1/2023	Tomáš Adam Aleš Michl Michal Škoda	*Balancing volatility and returns in the Czech National Bank Bank's foreign exchange portfolio*
RPN 2/2022	Jan Filáček Lucie Kokešová Matějková	*Disclosing dissent in monetary policy committees*
RPN 1/2022	Oxana Babecká Kucharčuková Jan Brůha Petr Král Martin Motl Jaromír Tonner	*Assessment of the nature of the pandemic shock: Implications for monetary policy*

CZECH NATIONAL BANK
Na Příkopě 28
115 03 Praha 1
Czech Republic

ECONOMIC RESEARCH DIVISION
Tel.: +420 224 412 321
Fax: +420 224 412 329
http://www.cnb.cz
e-mail: research@cnb.cz

ISSN 1803-7070