State Space Estimation: from Kalman Filter Back to Least Squares

Miroslav Plašil[1] | *Prague University of Economics and Business, Prague, Czech Republic*

Received 19.1.2023 (revision received 28.2.2023), **Accepted** (reviewed) 4.3.2023, **Published** 16.6.2023

Abstract

This note reviews a direct least squares estimation of a state space model and highlights its advantages over the standard Kalman filter in some applications. Although there is a close relationship between these two concepts, dual understanding of the estimation problem seems to be little appreciated by the mainstream econometric literature as well as applied researchers. Due to computational and theoretical advancements, the least squares estimation of a state space model has become a viable alternative in many fields, showing great potential in solving otherwise difficult problems. This note gathers and discusses some possible applications to illustrate the point and contribute to their wider use in practice.

Keywords	***DOI***	***JEL code***
Multi-objective least squares, State Space model, Kalman Filter	*https://doi.org/10.54694/stat.2023.3*	*C10, C51, C63*

INTRODUCTION

State space modelling has become the most popular approach to describing the behavior of dynamic systems. The state space representation of a model is very flexible and can subsume a wide class of models, including structural time series models, time-varying parameter models, dynamic factor models and many others. The range of applications varies from medicine, sociology and economics to national statistics. For instance, advanced temporal disaggregation methods used by many statistical institutes can be cast into a state space framework (Proietti, 2006).

Estimation of the unknown model quantities (or states) most often relies on the well-known Kalman filter recursions, which offer a computationally efficient solution to the underlying estimation problem (Kalman, 1960; Kalman and Bucy, 1961). However, one can argue that Kalman filter is just a logical extension of Gauss' original ideas on parameter estimation (Sorenson, 1970) and an elegant way to solve large – at the time infeasible – least squares problems (see Kollmann, 2013, or Andrle, 2014, among others).

The equivalence between Kalman filtering and 'large' least squares was well understood in the old days. Yet, it seems to have vanished from the radar screens of many applied researchers and statistical

[1] Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: plasil@vse.cz. Also the Czech National Bank, Prague, Czech Republic.

practitioners nowadays. This is quite unfortunate as alternative views at the estimation problem may equip researchers with an array of additional tools and help solve otherwise difficult problems. Advances in both computational power and theory make it now possible to benefit from these tools.

This note aims to return the equivalence between Kalman filtering and 'direct' least squares estimation of a state space model back into the spotlight and reviews why the latter can be of practical usefulness. It does not necessarily present any new results but offers another perspective on how to approach the estimation problem in question while trying to justify some of its benefits.

Unlike more traditional approaches that link the Kalman filter and the least squares estimator via the concept of the likelihood function (Chan and Jeliazkov, 2009; Aravkin et al., 2017), recursive least squares estimation (Pollock, 1999), or penalized least squares regression (Gomes, 1999), this note draws on the formulation of the multi-objective least squares problem and its neat solution. Although there is a close connection between this formulation and above-mentioned concepts (penalized least squares, in particular), the explicit reference to the multi-objective least squares is new in the literature to my utmost knowledge. It can be of some methodological interest, but above all this is of huge practical importance. Potential benefits are at least twofold: first, ordinary least squares and its extensions belong to the standard toolkit of many researchers. Reducing state space estimation to a common least squares problem thus invites to an active use of the least squares toolkit in the domain of state space estimation and modelling. The full scope of possibilities is yet to be fully appreciated by future research, but some of the state-of-the-art applications are discussed below to illustrate practical merits. Second, data augmentation approach used for solving multi-objective least squares enjoys a great flexibility in incorporating additional (cross) restrictions on the states which can greatly discipline high-level properties of the whole model. Data augmentation approach also provides a clear link to other statistical concepts, such as celebrated mixed estimation of Theil and Goldberger (1961) or more recent ridge regression approach to time-varying parameter estimation (Goulet Coulombe, 2020). Last but not the least, data augmentation approach enables to use standard regression routines available in common statistical environments (R, Matlab, Julia, Stata and others) to estimate state space models, thus avoiding a need for specific software or packages. More advanced users can obtain state estimates by using simple matrix multiplications that allow incorporating many practical extensions with ease.

1 A PRIMER ON THE LEAST SQUARES ESTIMATION AND STATE SPACE MODELLING

1.1 Least squares

For the sake of completeness and to set up the notation, it is helpful first to review the basics of the least squares estimation theory. Suppose that the data and parameters are related according to the linear (regression) model:

$$y = Xb + e, \tag{1}$$

where e are the measurement errors, y and X are formed by the observed data and b is a $q \times 1$ vector of parameters to be estimated. Good estimates, $\hat{b}$, should make the errors 'small' in some sense, i.e. they should make them close to zero as possible, $y - X\hat{b} \approx 0$.

The least squares meet this objective by minimizing the sum of squared errors, which is equivalent to minimizing squared Euclidean norm of the vector of errors, $||.||^2$:

$$\arg\min_b f = \arg\min_b \|Xb - y\|^2 = \arg\min_b \|e\|^2 = \left(e_1^2 + e_2^2 + \ldots + e_T^2\right). \tag{2}$$

The minimization problem (2) is known to have (provided the matrix X has linearly independent columns) a closed-form solution (Boyd and Vandenberghe, 2018):

$$\hat{b} = (X'X)^{-1}X'y\,. \tag{3}$$

No specific distributional assumptions are needed to obtain the solution in (3) but these might be necessary for statistical inference. Under the standard assumptions of the ordinary linear regression model and Gaussian errors, the estimator $\hat{b}$ is normally distributed with mean b and variance $\sigma^2(X'X)^{-1}$:

$$\hat{b} = N(b,\sigma^2(X'X)^{-1}), \tag{4}$$

where σ^2 is the variance of the error term.

In some applications, the researchers may wish to work with several least squares objectives $f_1, f_2,\dots, f_k$ associated with the same vector of parameters b, all of which should be minimized simultaneously. In this case, a standard solution for finding the values of the unknown parameter vector is to use a weighted sum of the objectives (also referred to as linear scalarization of the multi-objective optimization problem in some areas):

$$f = \lambda_1 f_1 + \ldots + \lambda_k f_k = \lambda_1\|X_1 b - y_1\|^2 + \ldots + \lambda_k\|X_k b - y_k\|^2, \tag{5}$$

where $\lambda_1,\ldots,\lambda_k$ are positive constants representing the weights attached to individual objectives. The higher is the value of λ_i, the stronger is our desire for f_i to be minimized. In other words: constants λ_i determine a relative trade-off between individual objectives. Since scaling all the weights by any positive number does not change the solution of (5), it is possible to set λ_1 to 1 with no loss of generality.

The ingenious thing about multiple-objective least squares is that the objective (5) can be minimized using data augmentation approach (also known as *stacking*). It relies on stacking the matrices X_i and the corresponding vectors y_i below one another and solving a single (but bigger) least squares problem on the newly formed 'data set' (Boyd and Vandenberghe, 2018):

$$\underset{b}{\arg\min}\left\|\begin{pmatrix}\sqrt{\lambda_1}X_1\\ \vdots \\ \sqrt{\lambda_k}X_k\end{pmatrix}b - \begin{pmatrix}\sqrt{\lambda_1}y_1\\ \vdots \\ \sqrt{\lambda_k}y_k\end{pmatrix}\right\|^2 = \underset{b}{\arg\min}\|\tilde{X}b - \tilde{y}\|^2, \tag{6}$$

where $\tilde{X}$ and $\tilde{y}$ represent a stacked matrix and a stacked vector, respectively. One can therefore still make use of the Formulas (3) and (4) to derive a solution to (5). To illustrate the underlying machinery, let us take an example of the ridge regression (Hoerl and Kennard, 1970). It was originally developed as a tool to deal with highly correlated independent variables but is also heavily used as a regularization tool in many machine learning applications (Hastie, 2020). Ridge regression seeks to minimize the following objective function:

$$f = \|Xb - y\|^2 + \lambda\|b\|^2 = \|Xb - y\|^2 + \lambda\|I_q b - 0_q\|^2,$$

where I_q denotes identity matrix of order q and 0_q is a vector of q zeroes. Ridge regression modifies the objective of sum of squared errors associated with the traditional linear regression by adding a sum of squares of individual regression coefficients. The second objective serves as a penalty, which shrinks the coefficients towards zero. The trade-off between the goodness of fit and the size of coefficients is regulated through the hyperparameter λ.

Ridge regression naturally fits into the multi-objective least squares framework with $\lambda_1 = 1,\ \lambda_2 = \lambda,\ X_1 = X, y_1 = y, X_2 = I_q, y_2 = 0_q$. Stacking the matrices according to (6) and feeding them into the solution (3) one can obtain a closed-form expression for the estimates, $\hat{b}$:

$$\hat{b} = \left(\begin{pmatrix} X' & \sqrt{\lambda} I_q \end{pmatrix} \begin{pmatrix} X \\ \sqrt{\lambda} I_q \end{pmatrix} \right)^{-1} \begin{pmatrix} X' & \sqrt{\lambda} I_q \end{pmatrix} \begin{pmatrix} y \\ \sqrt{\lambda} 0_q \end{pmatrix} = \left(X'X + \lambda I_q \right)^{-1} X'y, \tag{7}$$

which is the canonical formula of ridge regression.

To close this section, we note that positive constants λ_i are usually scalars in the multi-objective least squares setting, but the least squares framework is flexible enough to account for a more complex system of weights where each observation or variable in the objective function can potentially conform to a different set of weights. If this is deemed necessary for the analysis, scalars λ_i in (6) can be replaced by the arbitrary square matrices W_i allowing for the implementation of fully general weighting schemes. In technical terms, this is a standard weighted least squares problem commonly encountered in practice. The choice of λ_i and W_i can be problem-dependent or they can sometimes be inferred from data (e.g. via cross validation, see Goulet Coulombe, 2020). In many applications, however, the weights are put equal to the reciprocal of the error variances (see below).

1.2 State space model

More than anything else, the state space form is an elegant and fully general representation of dynamic models associated with the underlying likelihood-based estimation procedure – the Kalman filter. Below we consider a simple linear state space model, which consists of two equations:

$$y_t = G_t \alpha_t + \varepsilon_t, \tag{8}$$

$$\alpha_t = F_t \alpha_{t-1} + \nu_t, \tag{9}$$

where an $n \times 1$ vector of observations, y_t, $t = 1,2, \dots, T$, depends on a $q \times 1$ vector of unknown states, α_t, which follow a Markovian process. System matrices G_t and F_t define the structure of the model and are assumed to be known. Interpretation of the states and system matrices is problem-dependent, but the states typically represent some unobserved latent variables (trends, cycles, factors) or time-varying regression parameters. In addition to defining the model structure, the matrix G_t can also contain observations of independent variables.

Observation equation (8) bears close similarities to the linear regression model. But unlike the traditional regression model, the states (quantities to be estimated) are not assumed to be constant and evolve over time. Transition equation (9) restricts their dynamics to follow an autoregressive process of order one.[2] The specification of the state space model is completed by specifying distributional assumptions on the error terms ε_t and ν_t. In the classical setting, they are assumed to be independently and identically distributed and Gaussian:

$$\begin{pmatrix} \varepsilon_t \\ \nu_t \end{pmatrix} \sim N\left(0, \begin{pmatrix} \Omega_{11} & 0 \\ 0 & \Omega_{22} \end{pmatrix} \right). \tag{10}$$

[2] A random walk specification can be used in many applications. Note that we also allow for autoregressive processes of higher order since they can always be expressed as a large AR(1) model.

Formulas (8)–(10) form the basis for Kalman filter estimation, a recursive procedure for producing optimal estimates in the mean-square error sense. To spare space, we do not present Kalman filter formulas here, nor their derivation.[3] However, a few notes are in order.

First, estimates of the states can come in two flavors: *i*) estimates based on the information available up to time *t*, commonly referred to as the Kalman filter, and *ii*) estimates based on the whole data sample (i.e. on information up to time *T*), commonly referred to as the Kalman smoother. The direct 'large' least squares estimation of the states described below is equivalent to the Kalman smoother rather than the Kalman filter as it uses the whole data set. In theory, obtaining a series of filtered estimates via 'large' least squares would be possible but often impractical since one would always need to compute solution to a large least squares problem and store the estimate of the latest state once new observations have been added to the sample.

The second thing to observe is that due to its recursive nature, the Kalman filter/smoother needs to be initialized with some starting values to work. This consists in formulating a prior distribution of the initial state $\alpha_1 \sim N(x_0, D)$. The prior represents a best guess about the state and its covariance matrix before any observation has been collected. Strictly speaking, the least squares estimation procedure described below does not require the initialization step, but it might be helpful if there exist strong expert views on the proper starting values of the states. To maintain the closest link possible to the estimates produced by the Kalman smoother, we account for the starting values and associated uncertainty in the estimation procedure.

1.3 Estimating the state space model with the large least squares

To relate the state space model to a least squares problem it is useful to rewrite the Formulas (8)–(10) into a matrix form. Using $y = \left(y_1', y_2', \ldots, y_T'\right)'$, $\alpha = \left(\alpha_1', \alpha_2', \ldots, \alpha_T'\right)'$

$$G = \begin{bmatrix} G_1 & & \\ & \ddots & \\ & & G_T \end{bmatrix}; \; H = \begin{bmatrix} I_q & & & & \\ -F_2 & I_q & & & \\ & -F_3 & I_q & & \\ & & \ddots & \ddots & \\ & & & -F_T & I_q \end{bmatrix},$$

$$\varepsilon \sim N\left(0, I_T \otimes \Omega_{11}\right),$$

$$v \sim N\left(0, S\right),$$

where $\otimes$ denotes Kronecker product and

$$S = \begin{bmatrix} D & & & & \\ & \Omega_{22} & & & \\ & & \Omega_{22} & & \\ & & & \ddots & \\ & & & & \Omega_{22} \end{bmatrix},$$

[3] The formulas are readily available elsewhere including a handful of online sources. For the textbook treatment, see Kim and Nelson (1999), for example.

the state space model (8)–(10) can be compactly rewritten as (see also Chan and Jeliazkov, 2009):

$$y = G\alpha + \varepsilon\,, \tag{11}$$

$$H\alpha = \nu\,. \tag{12}$$

Formulas (11) and (12) bear the structure of the standard linear regression model, and indeed, one can treat them as such. Least squares estimation of unknown states proceeds with minimizing the sum of squared errors in both equations (i.e. $y - G\hat{\alpha} \approx 0$ and $0 - H\hat{\alpha} \approx 0$). Since there does not generally exist a unique solution that would simultaneously minimize both objectives related to the same vector of states, there will be a trade-off between the two (regulated by their relative weights, λ_i):

$$\underset{\alpha}{\arg\min}\, f = \underset{\alpha}{\arg\min}\, \lambda_1 \|G\alpha - y\|^2 + \lambda_2 \|H\alpha - 0\|^2. \tag{13}$$

Due to the structure of the objective (13), its minimization can be treated as a multi-objective least squares problem and solved using a data augmentation approach (6). Inserting the stacked matrix into (3) while making no distributional assumptions, the solution to (13) is equal to:

$$\hat{\alpha} = \left(\left(\sqrt{\lambda_1} G' \sqrt{\lambda_2} H' \right) \begin{pmatrix} \sqrt{\lambda_1} G \\ \sqrt{\lambda_2} H \end{pmatrix} \right)^{-1} \left(\sqrt{\lambda_1} G' \quad \sqrt{\lambda_2} H' \right) \begin{pmatrix} \sqrt{\lambda_1} y \\ \sqrt{\lambda_2} 0 \end{pmatrix} = \left(\lambda_1 G' G + \lambda_2 H' H \right)^{-1} \lambda_1 G' y\,. \tag{14}$$

Formula (14) with the hyperparameters λ_1 and λ_2 left unspecified provides a general result that does not automatically coincide with the Kalman smoother as the latter uses a specific weighting scheme. However, it is worth noting that (14) is equivalent to the flexible least squares (FLS) estimator of Kalaba and Tesfatsion (1989), which was designed to handle time-varying parameter regression problems. Derivation in (14) thus establishes a clear formal relationship between the Kalman smoother and the FLS estimator. Kalman smoother simply represents a specific case of a more general FLS estimator – a thing that seems to be missed by the existing literature.

If the errors are assumed to be Gaussian, the weights λ_1, λ_2 can be replaced by the weighting matrices W_1 and W_2. In particular, it can be shown that optimal estimates are obtained by putting the weights equal to the inverse of the error variance in the measurement and the state equation, respectively, i.e. $W_1 = I_T \otimes \Omega_{11}^{-1}$ and $W_2 = S^{-1}$ (Aravkin et al., 2021). Plugging these weights into (14) one obtains:

$$\hat{\alpha} = \left(G' \left(I_T \otimes \Omega_{11}^{-1} \right) G + H' S^{-1} H \right)^{-1} G' \left(I_T \otimes \Omega_{11}^{-1} \right) y. \tag{15}$$

The expression (15) describing the solution to the weighted least squares problem provides exactly the same estimates of the unknown states as the Kalman smoother (see also Chan and Jeliazkov, 2009) and establishes formal equivalence between the two. The (normally distributed) estimator also has the same variance:[4]

$$Var\left(\hat{\alpha}\right) = \left(G' \left(I_T \otimes \Omega_{11}^{-1} \right) G + H' S^{-1} H \right)^{-1}. \tag{16}$$

[4] In multi-objective least squares framework, the variance of the estimator (16) can be obtained directly by applying Formula (4) on the stacked matrix $\tilde{X} = \begin{pmatrix} \sqrt{\lambda_1} G \\ \sqrt{\lambda_2} H \end{pmatrix}$ with the weights $\lambda_1 = I_T \otimes \Omega_{11}^{-1}$ and $\lambda_2 = S^{-1}$.

2 PRACTICAL USEFULNESS OF THE LEAST SQUARES APPROACH

Making a case for the least squares view on the state space model estimation should take nothing away from the Kalman filter brilliance. It still ranks among the most frequently used algorithms in statistical computing while giving rise to a myriad of useful extensions (e.g. to non-linear systems). There might be situations, however, in which the least squares approach can be preferable or even more natural to use. Since their formal review seems to be missing in the current literature, we try, at least, to point to some of them to provide practitioners with a brief outline of potential benefits and tools. Interest readers are, however, encouraged to dive in deeper into the references provided below.

2.1 Speed and stability of computations

The very motivation of Kalman for designing his celebrated recursive algorithm was to avoid a need to solve large-scale least squares problems. Their solution requires inverting (potentially very) large matrices in (15) which can be computationally expensive and oftentimes unstable. Against this backdrop, it might be surprising to argue in favor of the least squares approach. While it might be difficult to beat the Kalman filter in online applications, when it is necessary to update the solution every time new data arrive, this is not necessarily the case in static problems where all data are already available to the analyst.

An important observation to make is that the matrix to be inverted in (15) has a very specific structure: it is a symmetric positive definite block tridiagonal matrix (Aravkin et al., 2021). Recalling that matrix inversion has a close relation to solving a system of linear equations (Golub and Van Loan, 1996), the computation of (15) can be understood as a solution to symmetric block tridiagonal (SBT) system of linear equations.

Solving SBT systems has a long tradition (independent of the Kalman filter literature and largely unknown to applied econometricians) with many fast algorithms having been proposed and used in practice. Aravkin et al. (2021) showed that some of the popular recursive algorithms for solving SBT systems are actually equivalent to traditional Kalman filter recursions applied to state space estimation.[5] New methods for solving SBT systems are still being developed, which creates a potential for further speed-ups of the estimation process. Importantly, the solution to SBT systems also sheds some additional light on the stability guarantees of the Kalman filter algorithm and can serve as a basis for more flexible statistical modelling frameworks (see Aravkin et al., 2017, and Aravkin et al., 2021).

A whole range of new possibilities for the least squares approach to speed up the computations can also be found in Bayesian state space modeling and posterior sampling. Given that other parameters in the model are kept fixed, it is straightforward to show that the conditional posterior distribution of the states is Gaussian and its mean and variance coincide with (15) and (16), respectively (Chan and Jeliazkov, 2009). To obtain samples from the full posterior distribution of states, one can simply draw a random sample from the multivariate normal distribution fully specified by (15) and (16). This approach – unlike many alternative approaches – avoids the need to sample states recursively (Carter and Cohn, 1994). Using a Cholesky decomposition to sidestep the matrix inversion in (15), Chan and Jeliazkov (2009) report 20–40 % faster run times than the traditional posterior sampling of the states. The speed gains can grow even further with an efficient SBT solver.[6]

---

[5] Taking the opposite perspective, robust algorithms to solve SBT systems can potentially be understood as new (faster and more stable) implementations of the Kalman filter recursions.

[6] Some of the solvers are freely available in common statistical environments. If the speed of computations is not the main issue, standard linear regression routines implemented in these environments should work fine for a lot of practitioners.

2.2 Implementation of expert views into estimation process

In many instances, modeling real-world phenomena unavoidably relies on the vast experience of researchers. In macroeconomics, for example, expert prior views are an integral part of any serious model-based macroeconomic evaluation. If used with caution, transparent and informed judgements can help discipline high-level properties of the model and can greatly enhance its practical usefulness. Such properties include steady-state values of the variables and the states, frequency-domain features of the model, its impulse-response functions or its real-time revision properties. One way of incorporating expert views into the model is to express these views as a set of linear stochastic restrictions:

$$r = R\alpha + \epsilon , \tag{17}$$

where r and R is a vector and a matrix, respectively, which define a nature of the restriction on the state dynamics and ϵ is a vector of random disturbances which regulates the tightness of the restriction.

Noting that additional expert information contained in (17) has a form of the linear regression model (1) with properly chosen *artificial* data, we can effortlessly incorporate the set of restrictions (17) into the estimation process using the data augmentation approach (6), i.e., by stacking the artificial data in (17) below those observed in (11) and (12) and derive an explicit solution similarly to that shown in (14).

This approach can be seen as an application of the mixed-estimation ideas of Theil and Goldberger (1961) to state space modeling and allows taking on board relevant information (expert knowledge) which can hardly be incorporated into the model through observed data or a change in the model structure. We note that the implementation of stochastic restrictions into Kalman filtering is also possible due to Doran (1992). However, the least squares view on restrictions can perhaps be more intuitive and easier to implement for some practitioners. The upside of the least squares view is also methodological since it directly reveals the intimate relationship between observed data, expert prior views (i.e., *artificial* or extraneous data stacked below observed sample) and structural modeling (stochastic restrictions in a form of artificial data imply cross-restrictions on the model states, thus tying down the structure of the model).

2.3 Vast stock of knowledge accumulated in the least squares domain

Over the years, statisticians have accumulated much knowledge related to the least squares methodology suggesting a lot of its useful extensions and computational shortcuts. At present, the full potential of these theoretical advances has not yet been systematically explored within the state space context. While the exhaustive exploration of this area is far beyond the scope of this note, I set forth at least two practical examples to demonstrate how the least squares theory might help solve otherwise difficult tasks.

First, an iteratively reweighted least squares algorithm (IRLS) can be employed to approximate the minimization of arbitrary vector norm, thus making it possible to incorporate any user-defined loss function (and therefore error distributions) into state space modeling.[7] Although there might exist tailor-made algorithms to handle specific loss functions, IRSL represents a universal and computationally feasible tool that can handle general cases.

For example, the least squares solution to the optimization problem (13) can be made equivalent to minimizing the L1 norm. This can be achieved by multiplying each observation by the weight

[7] In the Bayesian context, this can be used to elicit different types of prior distributions. Although from a computational perspective, some stability issues can, in theory, be associated with the iteratively reweighted least squares procedure, it usually shows good practical performance. Recently, some new convergence guarantees have been found for the iteratively reweighted least squares, see e.g. Kümmerle et al. (2021). Note however that there also might exist problem-specific computational algorithms that are faster and more stable than IRLS for some vector norms.

$\sqrt{1/|\hat{e}_i|}$ where $\hat{e}_i$ is a regression residual. It comes from the fact that $\sum|\hat{e}_i| = \sum\left(\sqrt{\frac{1}{|\hat{e}_i|}}\hat{e}_i\right)^2$. Since the size of the residuals is not known a priori, the solution is obtained iteratively by setting all the weights to 1 in the initial step and alternating between the weighted least squares estimation of the states and the calculation of residuals.

Making use of a specific loss function and its properties can help solve many practical problems. For instance, imposing a L1 ('LASSO') penalty on the residuals in the observation equation (11) can result in the state space modeling robust to outliers, while the application of the L1 penalty to the transition equation (12) would force the signals to be piecewise linear. Moreover, it is straightforward to induce sparsity of time-varying signals (states) by considering an additional LASSO-type penalty on the sum of the absolute values of the states. This leads to a sparse LASSO-Kalman smoother analogous to the work of Angelosante et al. (2009).

Second, the least squares view on the state space estimation can be very beneficial within the area of the DSGE modeling – a workhorse modeling framework in modern macroeconomics (Christiano et al., 2018). Many DSGE models can be cast into the state space representation and its standard tools can be then applied. The need for the least squares machinery may arise in the domain of the stochastically singular DSGE models (i.e. models with more observables than economic shocks).[8] While no meaningful solution to such models can be found via the Kalman smoother, this represents a standard (under-determined) problem in the least squares domain, which is well-understood in linear algebra (see Andrle, 2014).

However, the role for the least squares does not end here. The estimation can also benefit from additional least squares restrictions on structural shocks in the DSGE model. True 'structural' or economic shocks in the DSGE models should be uncorrelated, but it is rarely true in empirical applications (Andrle, 2014). Using the multi-objective least squares framework (5) and penalizing solutions with correlated shocks, it is straightforward to retrieve structural shocks in line with the underlying economic theory. Moreover, the form of the solution (15) which directly expresses the states as a weighted average of observables can be useful for assessing frequency-domain properties of the underlying DSGE model as well as for the decomposition of unobserved quantities into observables (see e.g. Gomes, 2007). Such analyses are a vital part of the macroeconomic assessment in many central banks and international economic organizations.[9]

CONCLUSION

State space modeling and Kalman filtering belong to the standard toolkit of applied researchers in many scientific fields. The possibility of estimating unknown states as a large but standard linear regression problem has long been known but somewhat neglected in recent decades, partially due to huge success of the Kalman filter in many applications. However, computational advances made it possible to return the direct least squares estimation back into the spotlight and provide practitioners with a whole new range of possibilities. Least squares estimation of a state space model can lead to computational speed-ups in static applications and show practical improvements over traditional Kalman filtering while broadening our understanding of the latter.

[8] This is (or should be) a common situation in advanced macroeconomic models since the evolution of many economic variables is driven by only a few underlying economic forces (shocks), say a demand shock.

[9] In macroeconomics, the unknown states often represent important (directly unobserved) economic concepts such as an output gap. Decomposing contributions of observed variables to the estimation of unobserved states is thus very useful for policy and model assessment.

This note aimed at introducing the least squares estimation of a state space model to a wide audience of applied researchers. The ambition was to discuss new perspectives, demonstrate relationships between diverse statistical concepts and provide practitioners with a brief list of potential applications. The hope is that it can open the door to whole new promising avenues of research in the area of advanced state space modeling where the advantages of a (penalized) least squares solution can be thoroughly appreciated. Although this short note only covered a limited set of applications, the potential of using least squares methodology in the context of state space modelling is immense and yet to be fully realized in both theory and practice.

References

ANDRLE, M. (2014). *Estimating Structural Shocks with DSGE Models* [online]. <http://michalandrle.weebly.com/uploads/1/3/9/2/13921270/opfilter.pdf>.

ANGELOSANTE, D., ROUMELIOTIS, S. I., GIANNAKIS, G. B. (2009). Lasso-Kalman smoother for tracking sparse signals [online]. *Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, 181–185. <https://doi.org/10.1109/ACSSC.2009.5470133>.

ARAVKIN, A. I., BURKE, J. V., BELL, B. M., PILLONETTO, G. (2021). Algorithms for Block Tridiagonal Systems: Stability Results for Generalized Kalman Smoothing [online]. *IFAC-PapersOnLine*, 54(7): 821–826. <https://doi.org/10.1016/j.ifacol.2021.08.463>.

ARAVKIN, A. I, BURKE, J. V., LJUNG, L., LOZANO, A., PILLONETTO, G. (2017). Generalized Kalman smoothing: Modeling and algorithms [online]. *Automatica*, 86: 63–86. <https://doi.org/10.1016/j.automatica.2017.08.011>.

BOYD, S., VANDENBERGHE, L. (2018). *Introduction to Applied Linear Algebra – Vectors, Matrices, and Least Squares*. Cambridge University Press.

CHAN, J. C. C., JELIAZKOV, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *Mathematical Modelling and Numerical Optimisation*, 1(1/2): 101–120.

CARTER, C. K., KOHN, R. (1994). On Gibbs sampling for state space models [online]. *Biometrika*, 81(3): 541–553. <https://doi.org/10.1093/biomet/81.3.541>.

CHRISTIANO, L. J., EICHENBAUM, M. S., TRABANDT, M. (2018). On DSGE Models [online]. *Journal of Economic Perspectives*, 32(3): 113–40. <https://doi.org/10.1257/jep.32.3.113>.

DORAN, H. E. (1992). Constraining Kalman Filter and Smoothing Estimates to Satisfy Time-Varying Restrictions [online]. *The Review of Economics and Statistics*, MIT Press, 74(3): 568–572. <https://doi.org/10.2307/2109505>.

GOLUB, G. H., VAN LOAN, CH. F. (1996). *Matrix Computations*. Johns Hopkins University Press.

GOMES, V. (1999). Three Equivalent Methods for Filtering Finite Nonstationary Time Series [online]. *Journal of Business & Economic Statistics*, 17(1), 109–116. <https://doi.org/10.1080/07350015.1999.10524800>.

GOMES, V. (2007). Wiener–Kolmogorov Filtering and Smoothing for Multivariate Series With State–Space Structure [online]. *Journal of Time Series Analysis*, 28(3): 361–385. <https://doi.org/10.1111/j.1467-9892.2006.00514.x>.

GOULET COULOMBE, P. (2020). *Time-Varying Parameters as Ridge Regressions* [online]. *arXiv.org*. <https://doi.org/10.48550/arXiv.2009.00401>.

HASTIE, T. (2020). Ridge Regularization: an Essential Concept in Data Science [online]. *Technometrics*, 62(4), 426–433. <https://doi.org/10.1080/00401706.2020.1791959>.

HOERL, A. E., KENNARD R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems [online]. *Technometrics*, 12(1): 55–67. <https://doi.org/10.2307/1267351>.

KALABA, R., TESFATSION, L. (1989). Time-varying linear regression via flexible least squares [online]. *Computers and Mathematics with Applications*, 17(8–9): 1215–1245. <https://doi.org/10.1016/0898-1221(89)90091-6>.

KALMAN, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems [online]. *Journal of Basic Engineering*. 82: 35–45. <http://dx.doi.org/10.1115/1.3662552>.

KALMAN, R. E., BUCY, R. S. (1961). New Results in Linear Filtering and Prediction Theory [online]. *Journal of Basic Engineering*, 83(1): 95–108. <https://doi.org/10.1115/1.3658902>.

KIM, C-J., NELSON, C. R. (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications* [online]. The MIT Press. <https://doi.org/10.7551/mitpress/6444.001.0001>.

KOLLMANN, R. (2013). Estimating the state vector of linearized DSGE models without the Kalman filter [online]. *Economics Letters*, 120(1): 65–66. <https://doi.org/10.1016/j.econlet.2013.03.041>.

KÜMMERLE, C., VERDUN, C. M., STÖGER, D. (2021). Iteratively Reweighted Least Squares for Basis Pursuit with Global Linear Convergence Rate. In: BEYGELZIMER, A., DAUPHIN, Y., LIANG, P., VAUGHAN, J. W. (eds.) *Advances in Neural Information Processing Systems*.

POLLOCK, D. S. G. (1999). *Time-Series Analysis, Signal Processing and Dynamics*. London: The Academic Press.
PROIETTI, T. (2006). Temporal Disaggregation by State Space Methods: Dynamic Regression Methods Revisited [online]. *The Econometrics Journal*, 9(3): 357–372. <https://doi.org/10.1111/j.1368-423X.2006.00189.x>.
SORENSON, H. W. (1970). Least-squares estimation: from Gauss to Kalman [online]. *IEEE Spectrum*, 7 (7): 63–68. <https://doi.org/10.1109/MSPEC.1970.5213471>.
THEIL, H., GOLDBERGER, A. (1961). On Pure and Mixed Statistical Estimation in Economics [online]. *International Economic Review*, 2(1):65–78. <http://dx.doi.org/10.2307/2525589>.