gretl working papers

Measures of variance for smoothed disturbances in linear state-space models: a clarification

Allin Cottrell
Wake Forest University
cottrell@wfu.edu

Riccardo (Jack) Lucchetti
Università Politecnica delle Marche
r.lucchetti@univpm.it

Matteo Pelagatti
Università di Milano – Bicocca
matteo.pelagatti@unimib.it

working paper #3

July 29, 2016

Abstract

We clarify a point regarding the appropriate measure(s) of the variance of smoothed disturbances in the context of linear state-space models. This involves explaining how two different concepts, which are sometimes given the same name in the literature, relate to each other. We also describe the behavior of several common software packages is in this regard.

Keywords: State-space models, Disturbance smoother, Auxiliary residuals.

JEL codes: C32, C63

1 Introduction and motivation

In this note we aim to clarify a point regarding the appropriate measure(s) of the variance of estimated (smoothed) disturbances in the context of linear state-space models. The "disturbance smoother" that yields such estimates was foreshadowed in Harvey and Koopman (1992) and first systematically developed in Koopman (1993). Further exposition and examples of use can be found in Koopman, Shephard and Doornik (1999), Durbin and Koopman (2012) and Commandeur, Koopman and Ooms (2011).

What stands in need of clarification? Not the backward-recursive mechanism for producing the smoothed disturbances as such. That is clearly and consistently explained in the sources just mentioned. It is a computationally efficient method that not only gives optimal estimates of the disturbances in the state equation (and, if applicable, the observation equation) of a state-space model, but also, with little extra computation, yields a smoothed estimate of the unobserved state (though not its variance). The clarification just concerns the appropriate variance measure for the inferred disturbances. Our point is that there are two such measures, each one suitable for a particular analytical purpose, but the literature is not as clear as it might be on the distinction between the two, and their respective uses.

To motivate our discussion, suppose a practitioner were to compare results from different software packages when applying the Local Level model, as defined in Harvey (1989), to the famous Nile flow data set.[1] When comparing smoothed disturbances from different packages, our practitioner might see a picture similar to Figure 1: the two plots are very similar—except for a striking difference in scale on the y-axis. This difference would be far from inconsequential if one wanted to use the smoothed residuals for their primary purpose, namely diagnostic testing. Users of Eviews would tend to conclude in favor of a structural break around 1900 ("00" on the x-axis), where the auxiliary residual exceeds 3.0 in absolute value,[2] but this conclusion would not be shared at all by users of the KFAS package for R.

Faced with this sort of discrepancy it would be natural to suspect a software bug. However, that's not the problem here: it is more subtle, and involves the unfortunate fact that two different definitions of variance, which are sometimes given the same name in the literature, are being used.

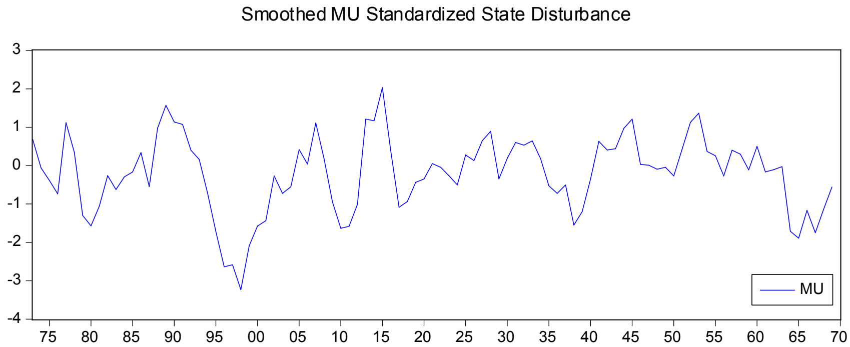
In explicating this issue we concentrate on two references: the SsfPack documentation (Koopman, Shephard and Doornik, 1999) and the extended discussion in Durbin and Koopman (2012). That's in part because we think these sources are the most likely to be accessed by practitioners, and also because the apparent tension between their respective expositions of the disturbance variance may be a source of puzzlement.

Section 2 sets up notation and compares the respective measures of variance of smoothed disturbances in the two references just mentioned. Section 3 confirms our interpretation of these measures by means of simulation and section 4 provides an illustration of what they are respectively good for. Section 5 draws a parallel with the familiar case of Ordinary Least Squares residuals, and section 6 gives an account of what is produced under the name of variance of smoothed disturbances by various software packages. An appendix gives an exposition of the disturbance smoother as a case of linear projection, thereby providing a more rigorous basis for statements made in the main text.

[1] A whole issue of the *Journal of Statistical Software* was devoted, a few years ago, to just this problem; see Commandeur *et al.* (2011).

[2] As would users of STAMP, which produces essentially the same plot.

(a) Eviews



(b) R/KFAS

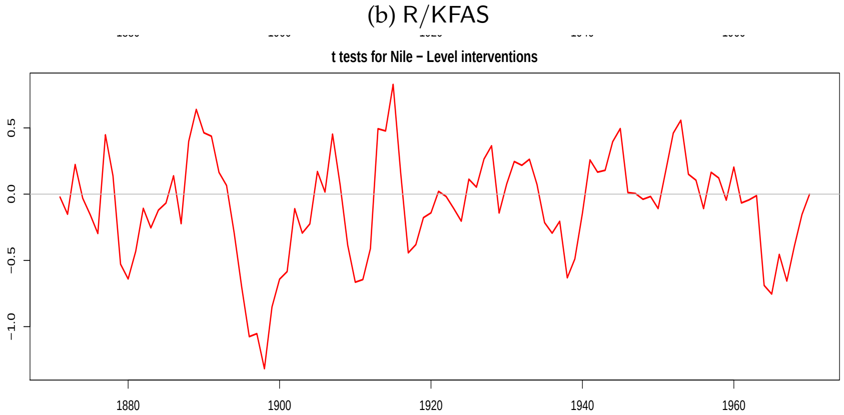


Figure 1: Standardized residuals from two software packages, pertaining to the level component of the Local Level model on the Nile data

2 Two measures of variance

We begin by establishing notation. We represent the linear state-space model by the following two equations, for the unobserved state and the observable outcome respectively:

$$\alpha_{t+1} = T_t\alpha_t + v_t \qquad v_t \sim \text{NID}(0, Q_t) \tag{1}$$

$$y_t = Z_t\alpha_t + w_t \qquad w_t \sim \text{NID}(0, R_t) \tag{2}$$

More general specifications are possible, of course, but given the question we have in view we lose no relevant generality by working with the above representation.

In Koopman, Shephard and Doornik (1999) (hereinafter KSD), the disturbances in the state and observation equations are represented as, respectively, $H_t\varepsilon_t$ and $G_t\varepsilon_t$ where $\varepsilon_t \sim \text{NID}(0, I)$, thereby allowing for dependence between the two equations if $H_tG_t' \neq 0$.

For present purposes we confine ourselves to the case of mutual independence of the disturbances, $H_tG_t' = 0$, and for this case KSD write the smoothed disturbances and their variances as follows (Koopman *et al.*, 1999, pages 124–5):

$$\begin{aligned}
\mathbb{E}(H_t\varepsilon_t|Y_n) &= H_tH_t'r_t \\
\mathbb{V}(H_t\varepsilon_t|Y_n) &= H_tH_t'N_tH_tH_t' \\
\mathbb{E}(G_t\varepsilon_t|Y_n) &= G_tG_t'e_t \\
\mathbb{V}(G_t\varepsilon_t|Y_n) &= G_tG_t'D_tG_tG_t'
\end{aligned}$$

Using the notation specified above we may rewrite these KSD results as

$$\begin{aligned}
\mathbb{E}(v_t|Y_n) &= Q_tr_t \\
\mathbb{V}(v_t|Y_n) &= Q_tN_tQ_t \\
\mathbb{E}(w_t|Y_n) &= R_te_t \\
\mathbb{V}(w_t|Y_n) &= R_tD_tR_t
\end{aligned}$$

The ancillary quantities e_t, r_t, N_t and D_t are defined on pages 130–131 of KSD; their precise definition need not concern us here since it is not in question (but see the Appendix for further discussion). The expression "Y_n" means $\{y_1, \ldots, y_n\}$, indicating conditionality on all n observations.

In Durbin and Koopman (2012) (hereinafter DK), the disturbances in the state and observation equations are represented as, respectively, $R_t\eta_t$—where $\eta_t \sim \text{NID}(0, Q_t)$ and R_t is a selection matrix—and $\varepsilon_t \sim \text{NID}(0, H_t)$.

Using this notation they write the smoothed disturbances and their variances as[3]

$$\begin{aligned}
\mathbb{E}(\eta_t|Y_n) &= Q_tR_t'r_t \\
\mathbb{V}(\eta_t|Y_n) &= Q_t - Q_tR_t'N_tR_tQ_t \\
\mathbb{E}(\varepsilon_t|Y_n) &= H_te_t \\
\mathbb{V}(\varepsilon_t|Y_n) &= H_t - H_tD_tH_t
\end{aligned}$$

(See pages 95–96 of DK.)

Under our representation of the state-space model, DK's "R_t" becomes an identity matrix, and renaming of the other vectors and matrices consistently with our chosen notation gives

$$\begin{aligned}
\mathbb{E}(v_t|Y_n) &= Q_tr_t \\
\mathbb{V}(v_t|Y_n) &= Q_t - Q_tN_tQ_t \\
\mathbb{E}(w_t|Y_n) &= R_te_t \\
\mathbb{V}(w_t|Y_n) &= R_t - R_tD_tR_t
\end{aligned}$$

Therefore, converted to a common notation, we have the results shown in Table 1. The expectations agree (obviously) but the variances are quite different.

KSD make no mention of what we're calling the DK variance. DK, on the other hand, use the KSD variance at some points; for example, their Figure 2.8 shows auxiliary residuals calculated as smoothed disturbances divided by the square roots of the KSD variance. So what's going on?

Having no desire to keep the reader is suspense, we proceed straight to the resolution of the puzzle: the formulae in KSD and DK are both perfectly valid, they just refer to different

[3] DK write u_t for what is called e_t in KSD; here we follow KSD and write e_t.

	KSD	DK
$\mathbb{E}(v_t\|Y_n)$	$Q_t r_t$	$Q_t r_t$
$\mathbb{V}(v_t\|Y_n)$	$Q_t N_t Q_t$	$Q_t - Q_t N_t Q_t$
$\mathbb{E}(w_t\|Y_n)$	$R_t e_t$	$R_t e_t$
$\mathbb{V}(w_t\|Y_n)$	$R_t D_t R_t$	$R_t - R_t D_t R_t$

Table 1: Comparison of KSD and DK formulae

concepts (while, unfortunately, using the same symbols for these different concepts). The variance of a random vector x is of course

$$\mathbb{E}\left[\left(x - \mathbb{E}(x)\right)\left(x - \mathbb{E}(x)\right)'\right]$$

The ambiguity relates to the appropriate definition of $\mathbb{E}(x)$. In KSD this is the unconditional expectation of the disturbance, namely zero. So their formulae in fact produce

$$\mathbb{V}_{\text{KSD}}(\hat{v}_t|Y_n) = \mathbb{E}(\hat{v}_t\hat{v}_t')$$
$$\mathbb{V}_{\text{KSD}}(\hat{w}_t|Y_n) = \mathbb{E}(\hat{w}_t\hat{w}_t'),$$

that is, the mean squared deviation of the inferred disturbances from zero.

In DK, on the other hand, the expectation of $\hat{v}_t$ is taken to be the actual v_t (and similarly for w_t), so we have

$$\mathbb{V}_{\text{DK}}(\hat{v}_t|Y_n) = \mathbb{E}[(\hat{v}_t - v_t)(\hat{v}_t - v_t)']$$
$$\mathbb{V}_{\text{DK}}(\hat{w}_t|Y_n) = \mathbb{E}[(\hat{w}_t - w_t)(\hat{w}_t - w_t)'],$$

that is, the mean squared deviation of the inferred disturbances from their true values.

The ambiguity can also be resolved by considering that $\hat{v}_t$ and $\hat{w}_t$ are unbiased and efficient predictors of the unobservable quantities v_t and w_t.[4] The above argument can then be summarized by saying that $\mathbb{V}_{\text{KSD}}$ is the variance of the *predictor function*, whereas $\mathbb{V}_{\text{DK}}$ is the variance of the *prediction error*.

3 Simulation results

If our interpretations are correct, this has implications for the results from simulation of a linear state-space model with given parameters. For the sake of simplicity, let the state and the observable both be scalars. Then if we run K simulations, indexed by i, we should find that

1. $K^{-1}\sum_{i=1}^{K}\hat{v}_{t,i}^2$ is a good approximation to $\mathbb{V}_{\text{KSD}}(\hat{v}_t)$; and

2. $K^{-1}\sum_{i=1}^{K}(\hat{v}_{t,i} - v_{t,i}^*)^2$ is a good approximation to $\mathbb{V}_{\text{DK}}(\hat{v}_t)$,

where in this context $v_{t,i}^*$ indicates the (known) artificial disturbance to the state at observation t on replication i and $\hat{v}_{t,i}$ indicates its estimate obtained via the disturbance smoother. Strictly analogous relations should hold with regard to the disturbance in the observation equation.

[4] Absolute efficiency depends on the Gaussianity assumption, but this could be dropped if one is willing to adopt a wider notion of efficiency.

We checked this implication using the famous Nile flow data, which have been much analysed in the state-space modeling literature. The SsfPack distribution, for example, includes a sample program `ssfnile.ox` which carries out maximum likelihood estimation of the Local Level model on these data. We used the disturbance smoother to obtain analytical results for the disturbance variances under this model, then ran $K = 50000$ replications in which we simulated data using the ML parameter estimates along with pseudo-random normal disturbances and calculated the respective sample variances.

The results are shown in Figure 2. It is clear that the simulated variance series do indeed fluctuate around the analytical values with which they should be associated according to the reasoning above.

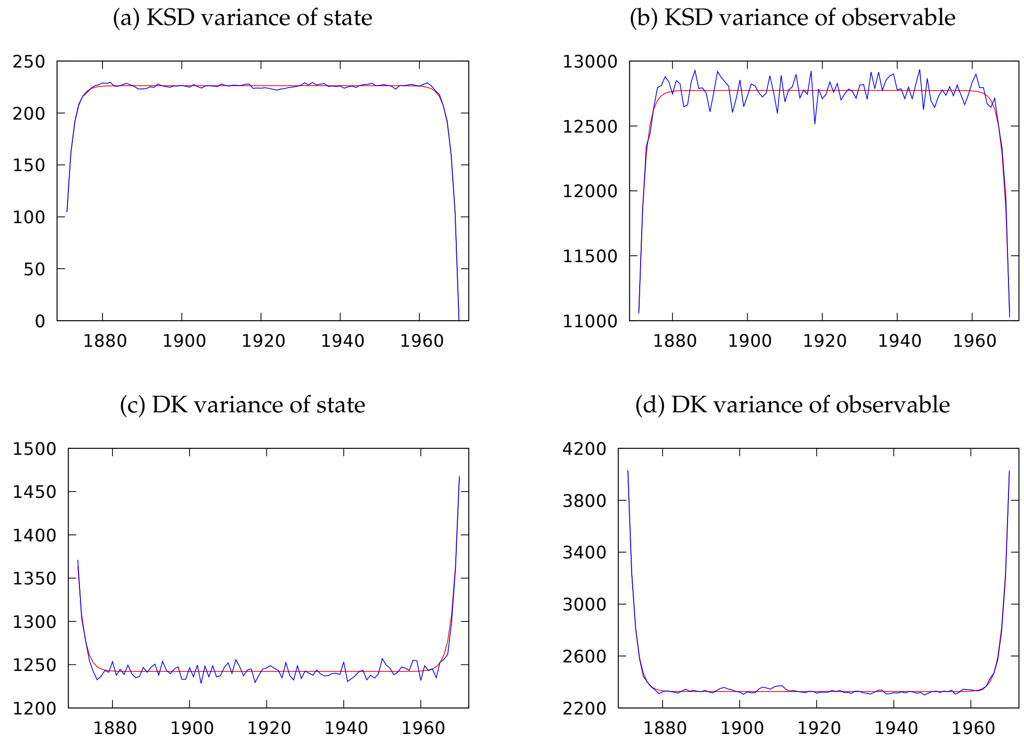


Figure 2: Analytical variances (red) and simulated variances (blue)

It is also clear from the shapes of the KSD and DK versions of the variance that the former is at minimum at the beginning and end of the time series, while the latter reaches its maximum at the same places.[5] Recall that the disturbance smoother is a predictor of the unobservable disturbance v_t (w_t) and that at the beginning and at the end of the time series the information that can be used to predict that disturbance is minimal, since half of the neighboring observations, the most relevant, are missing. If you had no relevant observations y_t, the best guess for v_t (w_t) would be its unconditional expectation, namely 0: the less information, the closer the prediction of v_t (w_t) to 0. This implies that the variance of v_t (w_t) around 0 is smallest at the beginning and at the end of the time series. However, since at these times $\hat{v}_t$ ($\hat{w}_t$) is shrunk towards zero, this bias makes its MSE as predictor of

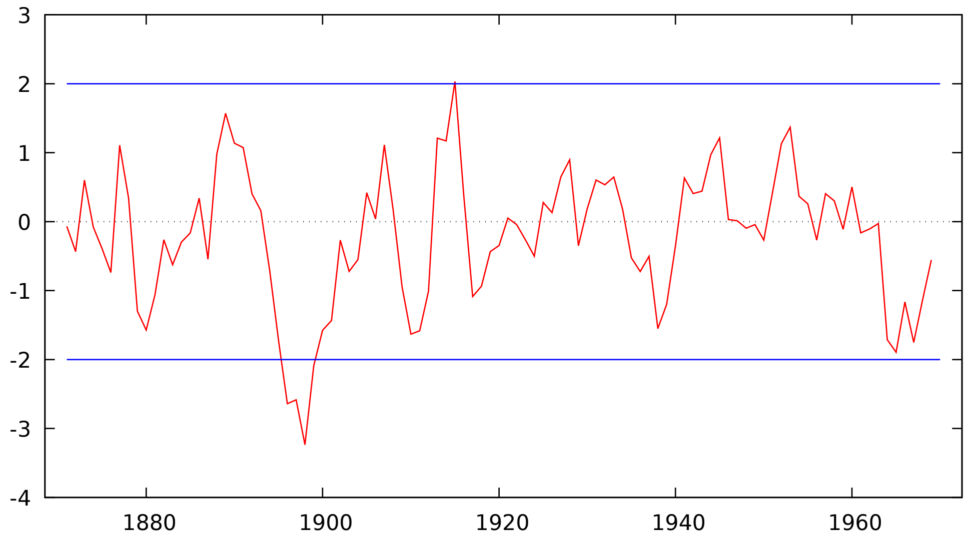
[5]Durbin and Koopman note, of their variance measure as computed on the Nile data, "the extent that these conditional variances are larger at the beginning and end of the sample" (DK, page 25), but they don't offer an explanation of this effect.

v_t (and w_t) higher. This is also coherent with the intuition that predictions based on less information are generally worse.

4 What are the respective variances good for?

To clarify what the KSD and DK variance measures are good for, we again refer to the Local Level model of the Nile data. We focus on the state equation; that is, the random-walk component of the observed series. Figure 3 shows plots based on two calls to the disturbance smoother. (Here we just use the actual data, no simulation.)

(a) Auxiliary (standardized) residuals, state equation



(b) Estimated state disturbance with 90% confidence band

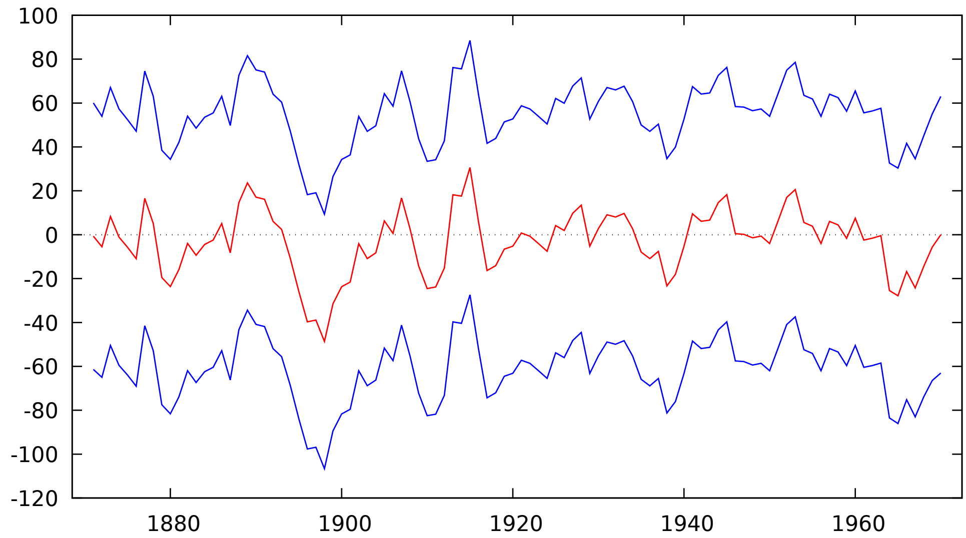


Figure 3: Nile data: auxiliary residuals (a) and $\hat{v}_t$ (b) from disturbance smoother

In the first call to the smoother we compute the KSD or sample variance of the disturbances, $E(\hat{v}_t \hat{v}_t')$, for each t. We then compute "auxiliary residuals" as $\hat{v}_t$ divided by the square root of the variance. These are shown in panel (a) of the Figure along with a band at ± 2 — for similar plots see Koopman *et al.* (1999), Pelagatti (2011). This plot suggests the presence of a structural break shortly prior to 1900, as several authors have observed. More generally, such plots may be useful for detecting outliers (level shifts, in the context of the Local Level model).

In the second smoother call we instead compute the DK variance $E[(\hat{v}_t - v_t)(\hat{v}_t - v_t)']$, the MSE of $\hat{v}_t$ considered as an estimator of v_t. And in panel (b) of the Figure we plot $\hat{v}_t$ along with a 90% confidence band (± 1.645 times the RMSE). This reveals that, given the sampling variance of $\hat{v}_t$, we're not really sure that *any* of the v_t values were truly different from zero, since the confidence band brackets zero at all observations. More generally, plots of this type may be useful for checking model specification: are we confident that such-and-such a posited disturbance is distinguishable from zero?

There's an apparent conflict in the case of the Nile data: some of the auxiliary residuals are "big enough" to suggest a structural break (panel a), yet it seems that none of them are strongly distinguishable from zero (panel b). The resolution is commonly reckoned to be that there was in fact a change in mean around 1900, but apart from that event there's little evidence for a non-zero σ_v^2. Or in other words the standard Local Level model is not really applicable to the data. While this diagnosis may be arrived at by other means, it's worth noting that it "suggests itself" when one cross-references the two sorts of plot.

Let us return to the two contrasting plots shown in Figure 1. It can be seen that the one from Eviews is essentially the same as panel (a) of our Figure 3; Eviews is dividing the smoothed residuals by their KSD standard deviations. In the R/KFAS plot, the divisor is the (DK) RMSE. Since the latter is in this case a good deal larger, the y-axis scale is correspondingly smaller. The plot using KSD tells us (roughly speaking) that it's unlikely that the downspike around 1900 comes from the same statistical distribution as the other level disturbances. The KFAS "t tests" plot using DK should be seen as making the same point as we make in panel (b) of Figure 3, namely that—in the context of the standard Local Level model—we cannot reject the null hypothesis that none of the disturbances are non-zero.

In the space of possible graphical representations of the estimated disturbances there's one we have not yet considered, namely a counterpart to panel (b) of Figure 3 in which a band is drawn around the smoothed disturbances at plus-or-minus so many times the KSD standard deviation. Although we don't show it here, this is in fact what Eviews produces if the option to plot "Disturbance estimates" is selected. Such a plot should be used with caution: we usually think of a confidence interval as a band that brackets the true value of the quantity being estimated (in this case the true disturbance) with some specified probability, but this interpretation is valid only if the band is based on the (DK) RMSE.

5 An analogical explanation: OLS residuals

In the Appendix the reader can find the derivation of the variances and mean square errors of the smoothed disturbances. However, it may be helpful to note a parallel with the case of ordinary least squares (OLS) residuals. In a standard notation the OLS residual is, of course,

$$
\begin{aligned}
\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}} &= (\mathbf{X}\beta + \mathbf{u}) - \mathbf{X}\hat{\beta} = \\
&= (I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{u} \equiv \mathbf{M}_{\mathbf{X}}\mathbf{u}
\end{aligned}
$$

where $\mathbf{M}_{\mathbf{X}}$ is a symmetric, idempotent projection matrix.[6] Therefore

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = (\mathbf{M}_{\mathbf{X}}\mathbf{u})'(\mathbf{M}_{\mathbf{X}}\mathbf{u}) = \mathbf{u}'\mathbf{M}_{\mathbf{X}}{}'\mathbf{M}_{\mathbf{X}}\mathbf{u} = \mathbf{u}'\mathbf{M}_{\mathbf{X}}\mathbf{u}$$

And $\hat{\mathbf{u}}'\mathbf{u} = \mathbf{u}'\mathbf{M}_{\mathbf{X}}\mathbf{u}$, so $\hat{\mathbf{u}}'\hat{\mathbf{u}} = \hat{\mathbf{u}}'\mathbf{u}$. It follows that $E(\hat{\mathbf{u}}'\hat{\mathbf{u}}) = E(\hat{\mathbf{u}}'\mathbf{u})$, and since $E(\hat{\mathbf{u}}) = E(\mathbf{u}) = 0$ we get the result $\mathbb{V}(\hat{\mathbf{u}}) = \mathbb{C}\mathrm{ov}(\hat{\mathbf{u}}, \mathbf{u})$.

This is interesting because it means that

$$\mathbb{V}(\hat{\mathbf{u}} - \mathbf{u}) = \mathbb{V}(\hat{\mathbf{u}}) + \mathbb{V}(\mathbf{u}) - 2\mathbb{C}\mathrm{ov}(\hat{\mathbf{u}}, \mathbf{u}) = \mathbb{V}(\mathbf{u}) - \mathbb{V}(\hat{\mathbf{u}})$$

That is, the variance of the difference between $\hat{\mathbf{u}}$ and $\mathbf{u}$ is equal to the difference between the variance of $\mathbf{u}$ and the variance of $\hat{\mathbf{u}}$—a relationship that obviously does not hold for most pairs of random variables A and B but only on condition that $\mathbb{V}(A) = \mathbb{C}\mathrm{ov}(A, B)$. This is characteristic of the case where A is an efficient estimator of B: the estimator varies only "in the direction of" the estimand. (For more on this, see the Appendix.)

In this context, $\hat{\mathbf{u}}$ is clearly an unbiased and efficient predictor of $\mathbf{u}$, and looking back at Table 1 it is easy to draw an analogy to the case of linear state space disturbances. The (KSD) variance of $\hat{v}_t$ is $Q_t N_t Q_t$ and the (DK) variance of $(\hat{v}_t - v_t)$ is $Q_t - Q_t N_t Q_t$, where Q_t is the variance of v_t, so

$$\mathbb{V}(\hat{v}_t - v_t) = \mathbb{V}(v_t) - \mathbb{V}(\hat{v}_t)$$

which holds if and only if $\mathbb{V}(\hat{v}_t) = \mathbb{C}\mathrm{ov}(\hat{v}_t, v_t)$.

When we speak of "residual variance" in the context of OLS we usually mean $\mathbb{V}(\hat{\mathbf{u}})$, not $\mathbb{V}(\hat{\mathbf{u}} - \mathbf{u})$. We might take this as suggesting that the expression "the variance of the smoothed disturbance" should apply to the KSD formulation. There's nothing wrong with the DK formulation, but it might better be called the MSE of the smoothed disturbance.

6 Software packages

Several software packages offer the facility of computing the disturbance smoother. In light of our discussion above, it is important for users of such software to understand exactly what they are getting under the name of the "standard deviation" or "variance" of the smoothed disturbances. This section gives a brief account of five relevant packages.

SsfPack (`www.ssfpack.com`) is one of the best known packages for state-space modeling, written by Siem Jan Koopman, the creator of the algorithms discussed in this paper. It comprises a set of routines coded in C and made available as an Ox package. The disturbance smoother is computed by the function `SsfMomentEst`, which returns the diagonal of the (KSD) covariance matrix. This is the right quantity for computing t-ratios for detecting outliers and structural breaks. The (DK) MSE is not provided automatically but can be obtained by subtracting the variances returned by `SsfMomentEst` from the variances on the diagonal of the matrix Q_t. The function `KalmanSmo` returns the quantities r_t and e_t together with the diagonals of the respective covariance matrices N_t and D_t.

[6] See, for instance, Davidson and MacKinnon (2004, chapter 2).

Eviews (`www.eviews.com`) offers a simple way to specify and estimate models in state-space form using the `sspace` object. A method on this object lets you plot the "Std. disturbance estimates," obtained by dividing the estimates by their (KSD) standard deviations. These are the quantities that should be used to find structural breaks. You can also obtain a plot of the "Disturbance estimates," however the confidence intervals shown in this plot are based on the (KSD) variances and not on the (DK) MSE as shown in panel (b) of our Figure 3, which we would find more useful. Nonetheless, the KSD variances can be saved into the workspace and the MSE can then be obtained by subtracting these from the disturbance variances.

R/KFAS: There are several packages in R to estimate models in state-space form (Petris and Petrone, 2011); here we consider KFAS by Jouni Helske, which is easy to use, efficiently coded and implements the disturbance smoother. The function that does the smoothing is `KFS`, which returns $\hat{v}_t$ and $\hat{w}_t$, named `etahat` and `epshat`, and two arrays of MSE matrices (as DK) $Q_t - Q_t N_t Q_t$ and $R_t - R_t D_t R_t$, named `V_eta` and `V_eps`. In order to obtain the (KSD) variance, one should change the signs of these matrices and add the system covariance matrices Q_t and R_t, respectively.

SSM is a Matlab Toolbox written by Jyh-Ying Peng and John Aston (Peng and Aston, 2011) for estimating models in state-space form. The function that computes the disturbance smoothers and the relative variances is `[epshat etahat epsvarhat etavarhat] = disturbsmo(y, llm)`. Here `epshat` and `etahat` are $\hat{w}_t$ and $\hat{v}_t$ and the relative variances `epsvarhat` and `etavarhat` are of the KSD type.

Gretl: The original implementation of state-space modeling in gretl (Lucchetti, 2011) had limited support for the disturbance smoother, but versions 2016c and higher offer full support (along with a streamlined interface). The relevant function is `kdsmooth`: this makes available estimates of $\hat{v}_t$ and $\hat{w}_t$ in matrix form under the name `smdist` and their standard deviations under the name `smdisterr`. By default these standard deviations are of the KSD type, but the function has an optional Boolean parameter: if a non-zero argument is given, `smdisterr` instead holds DK-type RMSE values. Current documentation can be found in the *Gretl User's Guide* at `https://sourceforge.net/projects/gretl/files/manual/`.

To summarize, in most cases the variance measure produced natively by current state-space modeling software is KSD (and the production of DK-type MSEs is up to the user). The exceptions are R/KFAS, which produces DK values, and gretl, which offers an explicit choice between the two variants.

References

Commandeur, J. J. F., S. J. Koopman and M. Ooms (2011) 'Statistical software for state space methods', *Journal of Statistical Software* 41(1): 1–18.

Davidson, R. and J. G. MacKinnon (2004) *Econometric Theory and Methods*, New York: Oxford University Press.

Durbin, J. and S. J. Koopman (2012) *Time Series Analysis by State Space Methods*, Oxford: Oxford University Press, second edn.

Harvey, A. C. (1989) *Forecasting, structural time series models and the Kalman filter*, Cambridge: Cambridge University Press.

Harvey, A. C. and S. J. Koopman (1992) 'Diagnostic checking of unobserved-components time series models', *Journal of Business & Economic Statistics* 10(4): 377–389.

Koopman, S. J. (1993) 'Disturbance smoother for state space models', *Biometrika* 80: 117–126.

Koopman, S. J., N. Shephard and J. A. Doornik (1999) 'Statistical algorithms for models in state space using SsfPack 2.2', *Econometrics Journal* 2: 107–160.

Lucchetti, R. (2011) 'State space methods in gretl', *Journal of Statistical Software* 41(11): 1–22.

Pelagatti, M. (2011) 'State space methods in Ox/SsfPack', *Journal of Statistical Software* 41(3): 1–25.

——— (2015) *Time Series Modelling with Unobserved Components*, Boca Raton: Chapman and Hall/CRC.

Peng, J.-Y. and J. Aston (2011) 'The State Space Models Toolbox for MATLAB', *Journal of Statistical Software* 41(6): 1–26.

Petris, G. and S. Petrone (2011) 'State Space Models in R', *Journal of Statistical Software* 41(6): 1–25.

Appendix: the disturbance smoother as projection

The disturbance smoother (DS) is an algorithm to compute a projection of the disturbances v_t and w_t onto the linear span of the observations $Y_n = \{y_1, \ldots, y_n\}$. In simpler words, the smoother provides the best linear combination of the observations Y_n (and a constant) for predicting the unobservable random quantities v_t and w_t. Here by "best" is meant that there is no other *linear* function of the observations whose mean square error (MSE) is smaller in predicting v_t and w_t.

If the normality assumptions in equations (1)–(2) hold, then the DS turns out to be also the best predictor of v_t and w_t, which corresponds to the conditional expectations $\mathbb{E}(v_t|Y_n)$ and $\mathbb{E}(w_t|Y_n)$, used in the text. In this case, there is no other (linear or non-linear) function of the observations that predicts v_t and w_t with smaller MSE.

To keep the derivation more general, let us assume that the DS provides just the linear projections, which we represent with the symbol $\mathbb{P}$:

$$\hat{v}_t = \mathbb{P}[v_t|Y_n], \qquad \hat{w}_t = \mathbb{P}[w_t|Y_n].$$

For a treatment of the theory of linear projections and their properties, see Pelagatti (2015, chapter 1), however here we will need just the following results. Let x be a random variable and y a random vector, then if $\hat{x} = \mathbb{P}[x|y]$ the projection of x onto the linear span of y, then it holds

$$\mathbb{E}(x - \hat{x}) = 0, \qquad \mathbb{E}(x - \hat{x})y = 0, \qquad \mathbb{E}(x - \hat{x})\hat{x} = 0 \qquad \mathbb{V}(\hat{x}) = \mathbb{C}\mathrm{ov}(\hat{x}, x).$$

The last two properties are direct consequences of the first two. Indeed, since $\hat{x}$ is a linear function of y and a constant, then by the first two properties it must be equal to zero. The last result is readily obtained: let us call μ the expectation of x, which by the first property is also the expectation of $\hat{x}$, then

$$\begin{aligned}
\mathbb{E}(\hat{x} - \mu)(x - \mu) &= \mathbb{E}(\hat{x} - \mu)(x - \hat{x} + \hat{x} - \mu) \\
&= \mathbb{E}(\hat{x} - \mu)(x - \hat{x}) + \mathbb{E}(\hat{x} - \mu)(\hat{x} - \mu) \\
&= 0 + \mathbb{V}(\hat{x}).
\end{aligned}$$

Now, KSD (pages 130–131) define the quantities r_t and e_t as linear functions of the observations Y_n with zero mean and covariance matrices, respectively, N_t and D_t. KSD prove that the linear projections of v_t and w_t onto the observations Y_n are given by $\hat{v}_t = Q_t r_t$ and $\hat{w}_t = R_t e_t$. It is immediate to see that the means of $\hat{v}_t$ and $\hat{w}_t$ are zero, while their variances are

$$\begin{aligned}
\mathbb{V}(\hat{v}_t) &= \mathbb{E}(\hat{v}_t \hat{v}_t') = \mathbb{E}(Q_t r_t r_t' Q_t) = Q_t N_t Q_t \\
\mathbb{V}(\hat{w}_t) &= \mathbb{E}(\hat{w}_t \hat{w}_t') = \mathbb{E}(R_t e_t e_t' R_t) = R_t D_t R_t.
\end{aligned}$$

Using the last two properties of the linear projections, the MSE of $\hat{v}_t$ and $\hat{w}_t$ as predictors of v_t and w_t are easily derived:

$$\begin{aligned}
\mathbb{E}[(v_t - \hat{v}_t)(v_t - \hat{v}_t)'] &= \mathbb{E}[(v_t - \hat{v}_t)v_t'] = \mathbb{V}(v_t) - \mathbb{C}\mathrm{ov}(\hat{v}_t, v_t) = Q_t - Q_t N_t Q_t \\
\mathbb{E}[(w_t - \hat{w}_t)(w_t - \hat{w}_t)'] &= \mathbb{E}[(w_t - \hat{w}_t)w_t'] = \mathbb{V}(w_t) - \mathbb{C}\mathrm{ov}(\hat{w}_t, w_t) = R_t - R_t D_t R_t.
\end{aligned}$$