Pruned Skewed Kalman Filter and Smoother: With Applications to the Yield Curve and Asymmetric Monetary Policy Shocks – Online Appendix – (Not For Publication)

Gaygysyz Guljanov[a], Willi Mutschler[b,], Mark Trede[a]

[a] *Center for Quantitative Economics, University of Münster.*

[b] *School of Business and Economics, University of Tübingen.*

Abstract

This online appendix contains detailed results on the Monte-Carlo study of the paper and derives the smoothing step of the skewed Kalman filter in general.

1. A Monte Carlo Study

In this section we conduct a thorough Monte Carlo study to evaluate the performance of the *Pruned Skewed Kalman Filter and Smoother* in terms of accuracy and speed. To this end, we consider three different state-space models as data-generating processes (DGP). DGP (1) is a univariate model given by the following parametrization:

$$G = 0.8, \qquad F = 10, \qquad \mu_\varepsilon = 1, \qquad \Sigma_\varepsilon = 0.01, \qquad \mu_\eta = 0.3, \qquad \Sigma_\eta = 0.64, \qquad \nu_\eta = 0, \qquad \lambda_\eta = -0.89 \tag{DGP (1)}$$

DGP (2) is a multivariate model with four state and three observable variables and randomly drawn parameters:

Email addresses: gaygysyz.guljanov@wiwi.uni-muenster.de (Gaygysyz Guljanov), willi@mutschler.eu (Willi Mutschler), mark.trede@uni-muenster.de (Mark Trede)

$$G = \begin{pmatrix} 0.5488 & 0.1738 & -0.2949 & 0.1534 \\ -0.2864 & 0.1060 & 0.3628 & 0.3334 \\ -0.3898 & -0.0252 & 0.5339 & 0.3163 \\ 0.2389 & 0.1958 & -0.0027 & 0.5519 \end{pmatrix} \qquad F = \begin{pmatrix} -0.7196 & 0.8221 & 0.4602 & -0.6412 \\ -2.0887 & -0.8201 & -1.2380 & 0.3937 \\ 0.6347 & -0.5109 & 0.8476 & 0.6819 \end{pmatrix}$$

$$\Sigma_\varepsilon = \begin{pmatrix} 0.0108 & -0.0276 & -0.0314 \\ -0.0276 & 0.1129 & -0.0025 \\ -0.0314 & -0.0025 & 0.2889 \end{pmatrix} \cdot 10^{-6} \qquad \mu_\varepsilon = \begin{pmatrix} 0.8565 \\ -0.3010 \\ -0.82705 \end{pmatrix}$$

$$\Sigma_\eta = \begin{pmatrix} 0.0013 & -0.0111 & 0.0116 & -0.0089 \\ -0.0111 & 0.1009 & -0.2301 & 0.1014 \\ 0.0116 & -0.2301 & 3.3198 & -1.0618 \\ -0.0089 & 0.1014 & -1.0618 & 1.0830 \end{pmatrix} \qquad \mu_\eta = \begin{pmatrix} 0.3455 \\ -1.8613 \\ 0.7765 \\ -0.5964 \end{pmatrix} \qquad \nu_\eta = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad \lambda_\eta = 0.89 \qquad \text{DGP (2)}$$

Note that in DGP (1) and DGP (2) we introduce the auxiliary hyperparameter $\lambda_\eta \in]-1;1[$ to re-parametrize the skewness parameters according to the following relationships:

$$\Gamma_\eta = \lambda_\eta {\Sigma_\eta}^{-1/2}, \qquad \Delta_\eta = (1-\lambda_\eta^2)\mathbf{I}_{n_\eta}$$

In this case, we can simplify $\Delta_\eta + {\Gamma_\eta}' \Sigma_\eta \Gamma_\eta$ to the identity matrix I_{n_η} such that the unconditional expectation vector and the covariance matrix of η_t can be calculated in closed-form (Flecher et al., 2009):

$$E[\eta_t] = \mu_\eta + \left(\sqrt{\frac{2}{\pi}}\lambda_\eta {\Sigma_\eta}^{1/2}\right)\mathbf{1}_{n_\eta}, \qquad V[\eta_t] = \Sigma_\eta\left(1 - \frac{2}{\pi}\lambda_\eta^2\right) \tag{1}$$

Arellano-Valle & Azzalini (2008) and Käärik et al. (2015) provide related discussions on the usefulness of this re-parameterization for the skew-normal distribution.[1] Lastly, DGP (3) is a multivariate model with three state variables that are all observable and do not use this re-parametrization:

$$G = \begin{pmatrix} 0.9969 & 0.1256 & -0.4803 \\ -0.8221 & 0.0386 & 0.6687 \\ 0.5605 & 0.6397 & -0.4333 \end{pmatrix} \quad F = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \Sigma_\varepsilon = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot 10^{-4} \quad \mu_\varepsilon = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Sigma_\eta = \begin{pmatrix} 0.64 & 0 & 0 \\ 0 & 0.36 & 0 \\ 0 & 0 & 0.49 \end{pmatrix} \quad \Gamma_\eta = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -6 \end{pmatrix} \quad \Delta_\eta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mu_\eta = \begin{pmatrix} 0.3 \\ -0.1 \\ 0.2 \end{pmatrix} \quad \nu_\eta = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{DGP (3)}$$

Note that DGP (3) is particularly of interest to our empirical application as it implies that $\eta_{1,t}$ is right-skewed, $\eta_{2,t}$ is symmetric (Gaussian), and $\eta_{3,t}$ is left-skewed.

The initial state distribution is set to a normal one with an initial covariance matrix with 10 on the di-

[1] This is without loss of generality. We mainly use this to quickly compute $E[\eta_t]$ and $V[\eta_t]$ as we use these values as input parameters for the *Gaussian Kalman Filter*. In fact, our replication codes contain functions to compute the unconditional mean and the covariance matrix for any parameterization of the multivariate CSN distribution. Moreover, in our empirical application we do not use this re-parametrization.

agonal, $x_{0|0} \sim CSN(0, 10I_{n_x}, 0, 0, I_{n_x}) = N(0, 10I_{n_x})$, following a suggestion of Harvey & Phillips (1979). To compute the likelihood function, we make use of the standard predictive decomposition based on the conditional distribution of y_t given y_{t-1}:

$$y_t|y_{t-1} \sim \; CSN(\hat{y}_{t|t-1},\; \Omega_{t|t-1},\; K_{t-1}^{Skewed},\; \nu_{t|t-1},\; \Delta_{t|t-1} + (\Gamma_{t|t-1} - K_{t-1}^{Skewed}F)\Sigma_{t|t-1}\Gamma'_{t|t-1})$$

where $\hat{y}_{t|t-1} = F\mu_{t|t-1} + \mu_\epsilon$ is the predicted value and $\Omega_{t|t-1} = F\Sigma_{t|t-1}F' + \Sigma_\epsilon$ the prediction-error covariance matrix of the Gaussian Kalman filter.

1.1. State estimation

For forecasting, it is helpful to condense the filtered distribution $x_{t|t}$ into a point estimator. Since the CSN distribution is asymmetric, the expectation $E[x_{t|t}]$ is one potential, but not necessarily the best point estimator. Let $L[\tilde{x}_t, x_t]$ denote the loss function for a point estimator $\tilde{x}_t$ if the true value is x_t. Depending on the loss function, different point estimators will minimize the expected loss. Of course, if the loss function is quadratic, i.e. $L_2[\tilde{x}_t, x_t] = (\tilde{x}_t - x_t)^2$, the expected loss is minimal if $\tilde{x}_t = E[x_{t|t}]$. If the loss function is $L_1[\tilde{x}_t, x_t] = |\tilde{x}_t - x_t|$, the best point estimator is the median of $x_{t|t}$. And the asymmetric loss function

$$L_q[\tilde{x}_t, x_t] = \begin{cases} a|\tilde{x}_t - x_t| & \text{for } x_t > \tilde{x}_t \\ b|\tilde{x}_t - x_t| & \text{for } x_t \leq \tilde{x}_t \end{cases}$$

results in the $a/(a+b)$-quantile of $x_{t|t}$ as point estimator. A similar discussion applies to the smoothed states $x_{t|T}$.

We start by simulating $R = 2400$ sample paths for x_t and y_t of different length $T = \{40, 80, 110\}$ (plus a burn-in of 100 periods). The shocks η_t are drawn from the CSN distribution, whereas the measurement errors ε_t are drawn from the normal distribution. We compute the expected losses to assess how well the *Pruned Skewed Kalman Filter and Smoother* estimate the unobserved state variables in comparison to the conventional *Skewed* or *Gaussian Kalman Filters and Smoothers*. That is, for each sample $r = 1, ..., R$ the loss is computed as

$$\text{Loss}^{(r)} := \sum_{t=20}^{T} L[\tilde{x}_t^{(r)}, x_t^{(r)}]$$

where in the univariate case L is any of the three loss functions L_1, L_2 and L_3 (with $a = 1$ and $b = 4$) under consideration, while in the multivariate case we focus only on L_2 as multivariate versions of L_1 and L_3 are not readily available. Note that in order to avoid too large an impact of the initial distribution $x_{0|0}$, the losses are calculated after a burn-in phase of 20 periods. The expected loss is then estimated by averaging

over all replications

$$\text{Expected Loss} = \frac{1}{R}\sum_{r=1}^{R} \text{Loss}^{(r)}.$$

Tables 1 and 2 report the *Expected Loss* and the 5th and 95th percentiles of $Loss^{(r)}$ of our Monte-Carlo simulation exercise for the different variants of both filters and smoothers. Three things are worth pointing out. First, the *Skewed Kalman Filter and Smoother* are superior to the *Gaussian Kalman Filter and Smoother* in all cases. Even though the better performance is rather small in the univariate case, it becomes really measurable in the multivariate case. This is not surprising, since the closed skew-normal distribution deviates only mildly from symmetry and normality (Liseo & Parisi, 2013) and the conventional Kalman filter and smoother are naturally optimal in its domain, i.e. when data is very close to normal. Nevertheless, in the more general case, the conventional Kalman filter and smoother simply neglect the skewed behavior; while the *Skewed Kalman Filter* embeds normality as a special case. Second, our pruning algorithm is very accurate and numerically almost equivalent to the non-pruned *Skewed Kalman Filter* (up to the twelfth digit in the univariate case and up to the 5th digit in the multivariate case). Third, the pruning threshold does not matter measurably in the univariate case and makes only a small numerical difference in multivariate settings. Clearly, the closer the tolerance is to 0, i.e. to the non-pruned filter and smoother, the more accurate we estimate the states. However, as we have argued above the non-pruned version of the filter and smoother is only feasible in the univariate case, while in multivariate settings we manage to deal with the numerical challenges for very small sample sizes only. Our pruning algorithm, on the other hand, is able to overcome this problem. Even with very low tolerance thresholds we are able to compute the filtering and smoothing steps without running into the curse of increasing skewness dimensionality. We conclude that overall both the *Pruned Skewed Kalman Filter and Smoother* perform well in terms of accuracy. However, there is a trade-off between accuracy and speed, which we analyze next.

1.2. Computing time

The performance of the *Pruned Skewed Kalman Filter* should also be evaluated in relation to its computing time. Table 3 reports the time in **seconds** required to compute 1000 evaluations of the log-likelihood function of univariate DGP (1) and multivariate DGP (2) for different sample sizes. We can see that as the number of observation periods grows, it takes more time to evaluate the likelihood function in all cases. The curse of increasing dimensionality inherent in the non-pruned *Skewed Kalman Filter* becomes apparent. Even though we are able to evaluate the Gaussian cdfs of increasing dimension in the univariate case, this comes at a cost as the computational time increases exponentially. In the multivariate case, the calculations are still feasible in principle for small sample sizes, but explode relatively quickly for medium to large sample sizes or require an unreasonably huge amount of computational time and memory. This becomes even more

DGP	T	Loss	Gaussian Kalman Filter	Pruned Skewed Kalman Filter no pruning	tol=1e-6	tol=1e-4	tol=1e-2
(1)	40	L_1	0.166535064 [0.1236;0.2142]	0.166527075 [0.1235;0.2144]	0.166527075 [0.1235;0.2144]	0.166527075 [0.1235;0.2144]	0.166527075 [0.1235;0.2144]
(1)	40	L_2	0.002080850 [0.0012;0.0033]	0.002080707 [0.0012;0.0033]	0.002080707 [0.0012;0.0033]	0.002080707 [0.0012;0.0033]	0.002080707 [0.0012;0.0033]
(1)	40	L_a	0.293173611 [0.2168;0.3852]	0.293160294 [0.2167;0.3852]	0.293160294 [0.2167;0.3852]	0.293160294 [0.2167;0.3852]	0.293160288 [0.2167;0.3852]
(1)	80	L_1	0.486149061 [0.4151;0.5639]	0.486122323 [0.4151;0.5638]	0.486122323 [0.4151;0.5638]	0.486122323 [0.4151;0.5638]	0.486122327 [0.4151;0.5638]
(1)	80	L_2	0.006087964 [0.0045;0.0080]	0.006087485 [0.0045;0.0080]	0.006087485 [0.0045;0.0080]	0.006087485 [0.0045;0.0080]	0.006087485 [0.0045;0.0080]
(1)	80	L_a	0.853449643 [0.7129;1.0017]	0.853414297 [0.7127;1.0008]	0.853414297 [0.7127;1.0008]	0.853414297 [0.7127;1.0008]	0.853414287 [0.7127;1.0008]
(1)	110	L_1	0.724620237 [0.6368;0.8186]	0.724598686 [0.6367;0.8185]	0.724598686 [0.6367;0.8185]	0.724598686 [0.6367;0.8185]	0.724598685 [0.6367;0.8185]
(1)	110	L_2	0.009073018 [0.0071;0.0113]	0.009072577 [0.0071;0.0113]	0.009072577 [0.0071;0.0113]	0.009072577 [0.0071;0.0113]	0.009072577 [0.0071;0.0113]
(1)	110	L_a	1.272405929 [1.1026;1.4533]	1.272363081 [1.1024;1.4536]	1.272363081 [1.1024;1.4536]	1.272363081 [1.1024;1.4536]	1.272363058 [1.1024;1.4536]
(2)	40	L_2	4.23932054 [2.1343;6.9488]	4.17299006 [2.1172;6.9381]	4.17299000 [2.1172;6.9381]	4.17298994 [2.1173;6.9382]	4.17450665 [2.0989;6.9805]
(2)	80	L_2	12.30937668 [8.4001;17.0700]	-	12.11085307 [8.3181;16.9039]	12.11085498 [8.3181;16.9048]	12.11665912 [8.3003;16.9054]
(2)	110	L_2	18.39547677 [13.4673;24.0186]	-	18.10271658 [13.1829;23.6744]	18.10272323 [13.1834;23.6743]	18.11208988 [13.2441;23.6814]

Table 1: Expected losses for filtered states, lower is better. 5th and 95th percentiles in square brackets.

severe if we increase the dimension of the state-space system matrices which is rather likely for real data applications.

The proposed *Pruned Skewed Kalman Filter* does not suffer from this and performs convincingly well. It is only slightly affected by a growing sample size; relatively speaking, it behaves very similar to the conventional Kalman filter in this regard. That is, the relative time increase between a sample size of 50 and 250 is approximately 3.93 for the univariate *Gaussian Kalman Filter*, whereas for the *Pruned Skewed Kalman Filter* we get a factor of approximately 4.15 for a very tight pruning threshold of 1e-6, and 4.25 for a very rough tolerance of 1e-2. In absolute terms, using 1e-2 as the tolerance level is 1.5 times faster than using 1e-6 as the tolerance level. In multivariate settings, the results are similar. The average time needed to compute the likelihood once is at least twice as fast when using a pruning threshold of 1e-2 compared to 1e-6. Combined with the results of the previous section, we conclude that a threshold of 1e-2 to 1e-4 seems to be a good compromise between accuracy and speed for multivariate models. For univariate models, one can easily lower this to a very small threshold such as 1e-6. In a nutshell: the lower the more accurate, the higher the faster.

Nevertheless, we do not want to hide the obvious fact that the *Gaussian Kalman Filter* is clearly the

DGP	T	Loss	Gaussian Kalman Smoother	Pruned Skewed Kalman Smoother no pruning	tol=1e-6	tol=1e-4	tol=1e-2
(1)	40	L_1	0.166539615 [0.1239;0.2143]	0.166528591 [0.1236;0.2145]	0.166528591 [0.1236;0.2145]	0.166528596 [0.1236;0.2145]	0.166528596 [0.1236;0.2145]
(1)	40	L_2	0.002080868 [0.0012;0.0033]	0.002080555 [0.0012;0.0033]	0.002080555 [0.0012;0.0033]	0.002080555 [0.0012;0.0033]	0.002080555 [0.0012;0.0033]
(1)	40	L_a	0.293186237 [0.2162;0.3848]	0.293159674 [0.2163;0.3847]	0.293159674 [0.2163;0.3847]	0.293159669 [0.2163;0.3847]	0.293159669 [0.2163;0.3847]
(1)	80	L_1	0.486116200 [0.4153;0.5645]	0.486083757 [0.4156;0.5644]	0.486083757 [0.4156;0.5644]	0.486083761 [0.4156;0.5644]	0.486083761 [0.4156;0.5644]
(1)	80	L_2	0.006087517 [0.0045;0.0080]	0.006086645 [0.0045;0.0080]	0.006086645 [0.0045;0.0080]	0.006086645 [0.0045;0.0080]	0.006086645 [0.0045;0.0080]
(1)	80	L_a	0.853428859 [0.7133;1.0015]	0.853347301 [0.7125;1.0010]	0.853347301 [0.7125;1.0010]	0.853347286 [0.7125;1.0010]	0.853347286 [0.7125;1.0010]
(1)	110	L_1	0.724563642 [0.6363;0.8186]	0.724528917 [0.6357;0.8184]	0.724528917 [0.6357;0.8184]	0.724528915 [0.6357;0.8184]	0.724528915 [0.6357;0.8184]
(1)	110	L_2	0.009072162 [0.0070;0.0113]	0.009071158 [0.0070;0.0113]	0.009071158 [0.0070;0.0113]	0.009071158 [0.0070;0.0113]	0.009071158 [0.0070;0.0113]
(1)	110	L_a	1.272365795 [1.1024;1.4536]	1.272252412 [1.1017;1.4534]	1.272252412 [1.1017;1.4534]	1.272252388 [1.1017;1.4534]	1.272252388 [1.1017;1.4534]
(2)	40	L_2	0.37817718 [0.1077;1.0473]	0.37728799 [0.1100;1.0227]	0.37728800 [0.1100;1.0227]	0.37728833 [0.1100;1.0226]	0.37740658 [0.1099;1.0301]
(2)	80	L_2	0.66853761 [0.3625;1.3612]	-	0.66267846 [0.3611;1.3307]	0.66267825 [0.3611;1.3306]	0.66275740 [0.3612;1.3280]
(2)	110	L_2	0.88605162 [0.5568;1.5668]	-	0.87956632 [0.5538;1.5548]	0.87956592 [0.5538;1.5549]	0.87960797 [0.5550;1.5550]

Table 2: Expected losses for smoothed states, lower is better. 5th and 95th percentiles in square brackets.

speed champion: it is roughly ten times faster than our proposed algorithm, but we are on the order of (neglectable) **milliseconds** here. Other approaches to evaluate the likelihood, such as Sequential Monte Carlo, are typically much slower by a factor of several hundred or thousand. Moreover, our implementation of both the conventional as well as the *Pruned Skewed Kalman Filters* are very textbook-like to fix ideas and highlight the underlying intuition. There is still much room for performance gains in the codes by e.g. avoiding inverses, adapting a steady-state filter and using Chandrasekhar recursions. We experimented with several such changes to the code and are able to cut the computational time at least by a half.[2] The *Pruned Skewed Kalman Filter* is therefore a very attractive and comparatively fast addition to the filtering toolkit of researchers who deal with skewed data and distributions. In the next section, we explore the finite sample properties of quasi-maximum likelihood estimators for the skewness parameters.

1.3. Accuracy of parameter estimation

In the last simulation exercise, we generate $R = 1000$ datasets from the multivariate DGP (3) with different sample sizes $T = \{100, 150, 200\}$. We then estimate the underlying parameters of the distribution of η_t, i.e.

[2]Granted our implementation of the *Gaussian Kalman Filter* can also be made faster. That's why we report the results for non-optimized, textbook-style codes.

DGP	T	Gaussian Kalman Filter	Pruned Skewed Kalman Filter no pruning	1e-6	1e-4	1e-2
(1)	50	0.2555 [0.18;0.39]	26.4004 [19.85;42.68]	5.7352 [4.26;8.55]	5.4255 [4.15;8.35]	3.7236 [2.82;5.74]
(1)	100	0.4236 [0.34;0.63]	178.4764 [164.04;272.46]	9.3932 [8.56;14.91]	9.2007 [8.30;14.70]	6.2735 [5.61;10.14]
(1)	150	0.5988 [0.51;0.92]	764.5933 [730.60;860.85]	13.8686 [13.08;14.44]	13.5295 [12.72;14.12]	9.2433 [8.61;9.72]
(1)	200	0.7869 [0.68;1.28]	2276.7823 [2208.36;2374.31]	18.7564 [17.65;19.54]	18.3690 [17.25;19.01]	12.5310 [11.65;13.20]
(1)	250	1.0029 [0.87;1.65]	5407.6977 [5292.97;5522.95]	23.8373 [22.29;24.89]	23.4432 [21.77;24.56]	15.8381 [14.71;16.57]
(2)	50	0.8439 [0.71;1.46]	554.8769 [518.17;660.16]	16.9193 [15.82;17.94]	12.8243 [11.90;15.25]	8.3821 [7.74;9.32]
(2)	100	1.6507 [1.42;2.99]	10902.8252 [9388.04;13659.95]	36.6296 [32.88;42.71]	26.8729 [24.38;30.29]	17.0615 [15.77;18.32]
(2)	150	3.3166 [2.65;5.69]	-	56.8434 [49.07;89.69]	35.2654 [30.71;55.59]	26.1939 [22.86;41.05]
(2)	200	4.6808 [3.50;7.50]	-	80.1512 [65.44;118.95]	49.6175 [40.84;74.17]	36.6851 [30.44;54.95]
(2)	250	5.2042 [4.35;9.10]	-	90.9946 [81.35;143.74]	55.7294 [50.84;88.30]	41.4539 [37.92;64.54]

Table 3: Time in seconds to compute 1000 evaluations of the log-likelihood function on AMD EPYC 7402P (24 cores, 96GB RAM). 5th and 95th percentiles in square brackets.

μ_η, $log(diag(\Sigma_\eta))$ and $diag(\Gamma_\eta)$, while fixing all other parameters at their true values.[3] Note that DGP (3) implies that $\eta_{1,t}$ is right-skewed, $\eta_{2,t}$ is symmetric (Gaussian), and $\eta_{3,t}$ is left-skewed.

Inspired by Atkinson et al. (2019), we measure parameter accuracy by reporting not only the average, 5th and 95th percentile of our estimates in the Monte Carlo sample, but also the normalized root-mean square-error (NRMSE) for each estimated parameter. That is, for some parameter j and Kalman filter variant f the error is the difference between the parameter estimate $\hat{\theta}_{j,f,r}$ for dataset r and the true parameter value θ_j:

$$NRMSE_f^j = \frac{1}{\theta_j}\sqrt{\frac{1}{R}\sum_{r=1}^{R}(\hat{\theta}_{j,f,r} - \theta_j)^2}$$

In other words, we normalize the RMSE by the true value θ_j to remove differences in the scales of the parameters.

Table 4 shows the parameter estimates by used Kalman filter variant (first column header) and for the three different sample sizes (second column header). Each cell includes the average value (first row), the 5th and 95th percentile in square brackets (second row), and the NRMSE in curly brackets (third row). Overall the

[3] We log-transform the variance to avoid the non-negativity constraint during the estimation procedure. The reported estimates are re-transformed.

		Pruned Skewed KF (1e-6)			*Pruned Skewed KF (1e-2)*			*Gaussian KF*		
Param	Truth	100	150	200	100	150	200	100	150	200
$[\mu_\eta]_1$	0.30	0.302 [0.19;0.43] {0.261}	0.299 [0.21;0.40] {0.193}	0.296 [0.22;0.38] {0.173}	0.302 [0.19;0.43] {0.260}	0.299 [0.21;0.40] {0.193}	0.297 [0.22;0.39] {0.173}	0.922 [0.83;1.01] {2.081}	0.924 [0.86;1.00] {2.083}	0.921 [0.86;0.98] {2.075}
$[\mu_\eta]_2$	-0.10	−0.101 [−0.20;−0.00] {−0.610}	−0.099 [−0.18;−0.02] {−0.480}	−0.100 [−0.17;−0.03] {−0.415}	−0.101 [−0.20;−0.00] {−0.615}	−0.099 [−0.18;−0.02] {−0.484}	−0.100 [−0.17;−0.03] {−0.418}	−0.101 [−0.20;−0.00] {−0.615}	−0.099 [−0.18;−0.02] {−0.484}	−0.100 [−0.17;−0.03] {−0.418}
$[\mu_\eta]_3$	0.20	0.195 [0.08;0.29] {0.321}	0.200 [0.11;0.27] {0.245}	0.198 [0.13;0.26] {0.214}	0.195 [0.08;0.29] {0.321}	0.200 [0.11;0.27] {0.245}	0.198 [0.13;0.26] {0.214}	−0.345 [−0.42;−0.28] {2.734}	−0.343 [−0.40;−0.28] {2.723}	−0.344 [−0.39;−0.29] {2.723}
$[\Sigma_\eta]_{11}$	0.64	0.654 [0.44;0.87] {0.203}	0.658 [0.50;0.84] {0.168}	0.656 [0.51;0.81] {0.148}	0.654 [0.44;0.87] {0.203}	0.658 [0.50;0.84] {0.168}	0.656 [0.51;0.81] {0.148}	0.260 [0.19;0.34] {0.597}	0.262 [0.21;0.32] {0.593}	0.261 [0.21;0.31] {0.594}
$[\Sigma_\eta]_{22}$	0.36	0.351 [0.27;0.44] {0.141}	0.353 [0.29;0.42] {0.117}	0.354 [0.30;0.41] {0.101}	0.351 [0.27;0.44] {0.141}	0.353 [0.29;0.42] {0.117}	0.354 [0.30;0.41] {0.101}	0.351 [0.27;0.44] {0.141}	0.353 [0.29;0.42] {0.117}	0.354 [0.30;0.41] {0.101}
$[\Sigma_\eta]_{33}$	0.49	0.491 [0.33;0.64] {0.197}	0.494 [0.37;0.62] {0.160}	0.489 [0.39;0.60] {0.133}	0.491 [0.33;0.64] {0.197}	0.494 [0.37;0.62] {0.160}	0.489 [0.39;0.60] {0.133}	0.193 [0.14;0.25] {0.609}	0.195 [0.15;0.24] {0.605}	0.193 [0.16;0.23] {0.607}
$[\Gamma_\eta]_{11}$	5.00	6.178 [3.31;11.47] {0.811}	5.712 [3.56;9.22] {0.513}	5.631 [3.78;8.63] {0.395}	6.173 [3.31;11.43] {0.804}	5.712 [3.56;9.21] {0.513}	5.630 [3.78;8.62] {0.394}			
$[\Gamma_\eta]_{22}$	0.00	−0.001 [−0.01;0.00] {}	−0.001 [−0.01;0.00] {}	−0.000 [−0.00;0.00] {}	−0.001 [−0.01;0.00] {}	−0.001 [−0.01;0.00] {}	−0.001 [−0.01;0.00] {}			
$[\Gamma_\eta]_{33}$	-6.00	−7.538 [−15.35;−3.76] {−0.741}	−6.849 [−11.09;−4.21] {−0.476}	−6.581 [−9.82;−4.30] {−0.376}	−7.535 [−15.36;−3.77] {−0.739}	−6.848 [−11.08;−4.21] {−0.474}	−6.581 [−9.82;−4.30] {−0.377}			

Table 4: Distribution of parameter estimates. Cells contain average, [5,95] percentiles and {NRMSE}.

estimates using the *Pruned Skewed Kalman Filter* are convincingly good for both a very low and a rather large pruning threshold. Most mass is centered around the true value and the distribution becomes narrower with larger sample sizes. The *Pruned Skewed Kalman Filter* successfully uncovers the skewed distribution of $\eta_{1,t}$ and $\eta_{3,t}$, but also Gaussianity of $\eta_{2,t}$. Note that the Gaussian Kalman filter completely misses the skewed distribution of η_t; which is evident in heavily biased values of μ_η and Σ_η. However, this bias is in fact misleading, because when using the Gaussian Kalman filter, μ_η and Σ_η are actually estimates for $E[\eta_t]$ and $V[\eta_t]$, which in our exercise are equal to $[0.9192; -0.1000; -0.3433]$ and $diag([0.2565; 0.3600; 0.1948])$, respectively. So the Gaussian Kalman filter still remains a powerful tool, if one is only concerned about estimating the mean and variance of the process. Of course those two estimators are biased due to the overlooked skewed distribution. In contrast, the *Pruned Skewed Kalman Filter* is more general as it nests Gaussianity as a special case.

2. Skewed Kalman Smoother

In this section we derive the smoothing step of the skewed Kalman filter in general.

2.1. Introduction and notations

Derivation of the smoothing step is a tedious task, also in terms of notations. In order to carry out this task as neatly as possible, we will proceed as follows. We first set notations and abbreviations. Afterwards, in a separate section, we show some useful derivations which we use later on. Then the derivation of the smoothing step proceeds with finding smoothed variables of periods $T-1$, $T-2$, $T-3$ and $T-4$, recursively. Here T denotes the last time period. From what we learn by deriving the above four periods manually, we devise general formulas for any time period. The more compact formulas for any time period is then given in the last section.

Now, let us start with with notations and abbreviations.

$$
\begin{aligned}
J_t &= \Sigma_{t|t} G' \Sigma_{t+1|t}^{-1} \\
K_t &= \Sigma_{t+1|t} F' (F \Sigma_{t+1|t} F' + \Sigma_\varepsilon)^{-1} (y_{t+1} - F \mu_{t+1|t}).
\end{aligned}
$$

For any t, we know the prediction step distribution (see Rezaie & Eidsvik (2014))

$$
x_{t+1|t} \sim CSN_{p,q_t+q_\eta}(\mu_{t+1|t}, \Sigma_{t+1|t}, \Gamma_{t+1|t}, \nu_{t+1|t}, \Delta_{t+1|t})
$$

where

$$
\begin{aligned}
\mu_{t+1|t} &= G\mu_{t|t} + \mu_\eta \\
\Sigma_{t+1|t} &= G\Sigma_{t|t}G' + \Sigma_\eta \\
\Gamma_{t+1|t} &= \begin{pmatrix} \Gamma_{t|t}J_t \\ \Gamma_\eta\Sigma_\eta\Sigma_{t+1|t}^{-1} \end{pmatrix} \\
\nu_{t+1|t} &= \begin{pmatrix} \nu_{t|t} \\ \nu_\eta \end{pmatrix} \\
\Delta_{t+1|t} &= \begin{pmatrix} \Delta_{t|t} + \Gamma_{t|t}\Sigma_{t|t}\Gamma'_{t|t} - \Gamma_{t|t}J_tG\Sigma_{t|t}\Gamma'_{t|t} & \Gamma_{t|t}J_t\Sigma_\eta\Gamma'_\eta \\ \Gamma_\eta\Sigma_\eta J'_t\Gamma'_{t|t} & \Delta_\eta + \Gamma_\eta\Sigma_\eta\Gamma'_\eta - \Gamma_\eta\Sigma_\eta\Sigma_{t+1|t}^{-1}\Sigma_\eta\Gamma'_\eta \end{pmatrix}
\end{aligned}
$$

and the update step distribution (see Rezaie & Eidsvik (2014))

$$
x_{t+1|t+1} \sim CSN_{p,q_t}(\mu_{t+1|t+1}, \Sigma_{t+1|t+1}, \Gamma_{t+1|t+1}, \nu_{t+1|t+1}, \Delta_{t+1|t+1})
$$

where

$$
\begin{aligned}
\mu_{t+1|t+1} &= \mu_{t+1|t} + K_t \\
\Sigma_{t+1|t+1} &= \Sigma_{t+1|t} - \Sigma_{t+1|t}F'(F\Sigma_{t+1|t}F' + \Sigma_\varepsilon)^{-1}F\Sigma_{t+1|t} \\
\Gamma_{t+1|t+1} &= \Gamma_{t+1|t} \\
\nu_{t+1|t+1} &= \nu_{t+1|t} - \Gamma_{t+1|t}K_t \\
\Delta_{t+1|t+1} &= \Delta_{t+1|t}.
\end{aligned}
$$

2.2. Preliminary theorems and lemmas

Theorem 1.

$$
x_t|x_{t+1}, D_t \sim CSN(\mu_{dt}, \Sigma_{dt}, \Gamma_{dt}, \nu_{dt}, \Delta_{dt})
$$

with

$$
\begin{aligned}
\mu_{dt} &= \mu_{t|t} + \Sigma_{t|t} G' \Sigma_{t+1|t}^{-1} (x_{t+1} - G\mu_{t|t} - \mu_\eta) \\
\Sigma_{dt} &= \Sigma_{t|t} - \Sigma_{t|t} G' \Sigma_{t+1|t}^{-1} G \Sigma_{t|t} \\
\Gamma_{dt} &= \begin{pmatrix} \Gamma_{t|t} \\ -\Gamma_\eta G \end{pmatrix} \\
\nu_{dt} &= \begin{pmatrix} \nu_{t|t} \\ \nu_\eta \end{pmatrix} - \begin{pmatrix} \Gamma_{t|t} \Sigma_{t|t} G' \Sigma_{t+1|t}^{-1} \\ \Gamma_\eta - \Gamma_\eta G \Sigma_{t|t} G' \Sigma_{t+1|t}^{-1} \end{pmatrix} (x_{t+1} - G\mu_{t|t} - \mu_\eta) \\
&= \begin{pmatrix} \nu_{t|t} - \Gamma_{t|t} \Sigma_{t|t} G' \Sigma_{t+1|t}^{-1} (x_{t+1} - G\mu_{t|t} - \mu_\eta) \\ \nu_\eta - \Gamma_\eta + \Gamma_\eta G \Sigma_{t|t} G' \Sigma_{t+1|t}^{-1} (x_{t+1} - G\mu_{t|t} - \mu_\eta) \end{pmatrix} \\
\Delta_{dt} &= \Delta_c \\
&= \begin{pmatrix} \Delta_{t|t} & 0 \\ 0 & \Delta_\eta \end{pmatrix}.
\end{aligned}
$$

Proof. Our starting point is the equation

$$
\begin{pmatrix} x_{t|t} \\ x_{t+1|t} \end{pmatrix} = \begin{pmatrix} I \\ G \end{pmatrix} x_{t|t} + \begin{pmatrix} 0 \\ \eta_t \end{pmatrix}.
$$

which is equivalent to

$$
\begin{pmatrix} x_{t|t} \\ x_{t+1|t} \end{pmatrix} = \underbrace{\begin{pmatrix} I & 0 \\ G & I \end{pmatrix}}_{\equiv A} \begin{pmatrix} x_{t|t} \\ \eta_t \end{pmatrix}.
$$

Firstly, let us find the joint distribution of

$$
\begin{pmatrix} x_{t|t} \\ \eta_t \end{pmatrix} \sim \mathrm{CSN}(\mu_j,\ \Sigma_j,\ \Gamma_j,\ \nu_j,\ \Delta_j)
$$

where,

$$\mu_j = \begin{bmatrix} \mu_{t|t} \\ \mu_\eta \end{bmatrix}$$
$$\Sigma_j = \begin{bmatrix} \Sigma_{t|t} & 0 \\ 0 & \Sigma_\eta \end{bmatrix}$$
$$\Gamma_j = \begin{bmatrix} \Gamma_{t|t} & 0 \\ 0 & \Gamma_\eta \end{bmatrix}$$
$$\nu_j = \begin{bmatrix} \nu_{t|t} \\ \nu_\eta \end{bmatrix}$$
$$\Delta_j = \begin{bmatrix} \Delta_{t|t} & 0 \\ 0 & \Delta_\eta \end{bmatrix}$$

Then, take the linear transformation of the above joint distribution with matrix A.

$$A \begin{pmatrix} x_{t|t} \\ \eta_t \end{pmatrix} \sim \text{CSN}(\mu_A,\ \Sigma_A,\ \Gamma_A,\ \nu_A,\ \Delta_A)$$

where,

$$\mu_A = A\mu_j = \begin{bmatrix} \mu_{t|t} \\ G\mu_{t|t} + \mu_\eta \end{bmatrix}$$
$$\Sigma_A = A\Sigma_j A' = \begin{bmatrix} \Sigma_{t|t} & \Sigma_{t|t}G' \\ G\Sigma_{t|t} & G\Sigma_{t|t}G' + \Sigma_\eta \end{bmatrix}$$
$$\Gamma_A = \Gamma_j A^{-1} = \begin{bmatrix} \Gamma_{t|t} & 0 \\ 0 & \Gamma_\eta \end{bmatrix} \begin{bmatrix} I & 0 \\ -G & I \end{bmatrix} = \begin{bmatrix} \Gamma_{t|t} & 0 \\ -\Gamma_\eta G & \Gamma_\eta \end{bmatrix}$$
$$\nu_A = \nu_j = \begin{bmatrix} \nu_{t|t} \\ \nu_\eta \end{bmatrix}$$
$$\Delta_A = \Delta_j = \begin{bmatrix} \Delta_{t|t} & 0 \\ 0 & \Delta_\eta \end{bmatrix}$$

Using the rules for conditional distributions (property 5 of the paper) we obtain

$$x_t | x_{t+1}, D_t \sim CSN(\mu_{dt}, \Sigma_{dt}, \Gamma_{dt}, \nu_{dt}, \Delta_{dt})$$

$\square$

Using the notations above and some simplifications, we can rewrite out theorem as follows:

$$x_t|x_{t+1}, D_t \sim CSN(\mu_{dt}, \Sigma_{dt}, \Gamma_{dt}, \nu_{dt}, \Delta_{dt})$$

with

$$\begin{aligned}
\mu_{dt} &= \mu_{t|t} + J_t(x_{t+1} - \mu_{t+1|t}) \\
\Sigma_{dt} &= \Sigma_{t|t} - J_t G \Sigma_{t|t} \\
\Gamma_{dt} &= \begin{pmatrix} \Gamma_{t|t} \\ -\Gamma_\eta G \end{pmatrix} \\
\nu_{dt} &= \nu_{t+1|t} - \Gamma_{t+1|t}(x_{t+1} - \mu_{t+1|t}) \\
\Delta_{dt} &= \begin{pmatrix} \Delta_{t|t} & 0 \\ 0 & \Delta_\eta \end{pmatrix}.
\end{aligned}$$

Theorem 2. *The following equality holds,*

$$\Phi(\Gamma_{t+1|t+1}(x_{t+1} - \mu_{t+1|t+1}); \nu_{t+1|t+1}, \Delta_{t+1|t+1}) = \Phi(0; \nu_{dt}, \Delta_{dt} + \Gamma_{dt}\Sigma_{dt}\Gamma'_{dt})$$

or, equivalently,

$$\Phi(0; \nu_{t+1|t+1} - \Gamma_{t+1|t+1}(x_{t+1} - \mu_{t+1|t+1}), \Delta_{t+1|t+1}) = \Phi(0; \nu_{dt}, \Delta_{dt} + \Gamma_{dt}\Sigma_{dt}\Gamma'_{dt}).$$

Proof. We rewrite each expression in more basic terms. For the left hand side we obtain

$$\begin{aligned}
\Gamma_{t+1|t+1}(x_{t+1} - \mu_{t+1|t+1}) &= \Gamma_{t+1|t}(x_{t+1} - \mu_{t+1|t} - \Sigma_{t+1|t}F'(F\Sigma_{t+1|t}F' + \Sigma_\varepsilon)^{-1}(y_{t+1} - F\mu_{t+1|t})) \\
\nu_{t+1|t+1} &= \nu_{t+1|t} - \Gamma_{t+1|t}\Sigma_{t+1|t}F'(F\Sigma_{t+1|t}F' + \Sigma_\varepsilon)^{-1}(y_{t+1} - F\mu_{t+1|t}) \\
\Delta_{t+1|t+1} &= \begin{pmatrix} \Delta_{t|t} + \Gamma_{t|t}\Sigma_{t|t}\Gamma'_{t|t} - \Gamma_{t|t}\Sigma_{t|t}G'\Sigma^{-1}_{t+1|t}G\Sigma_{t|t}\Gamma'_{t|t} & -\Gamma_{t|t}\Sigma_{t|t}G'\Sigma^{-1}_{t+1|t}\Sigma_\eta\Gamma'_\eta \\ (-\Gamma_{t|t}\Sigma_{t|t}G'\Sigma^{-1}_{t+1|t}\Sigma_\eta\Gamma'_\eta)' & \Delta_\eta + \Gamma_\eta\Sigma_\eta\Gamma'_\eta - \Gamma_\eta\Sigma_\eta\Sigma^{-1}_{t+1|t}\Sigma_\eta\Gamma'_\eta \end{pmatrix}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\nu_{t+1|t+1} - \Gamma_{t+1|t+1}(x_{t+1} - \mu_{t+1|t+1}) &= \nu_{t+1|t} - \Gamma_{t+1|t}(x_{t+1} - \mu_{t+1|t}) \\
&= \begin{pmatrix} \nu_{t|t} \\ \nu_\eta \end{pmatrix} - \begin{pmatrix} \Gamma_{t|t}\Sigma_{t|t}G'\Sigma^{-1}_{t+1|t} \\ \Gamma_\eta \Sigma_\eta \Sigma^{-1}_{t+1|t} \end{pmatrix} (x_{t+1} - \mu_{t+1|t}) \\
&= \begin{pmatrix} \nu_{t|t} - \Gamma_{t|t}\Sigma_{t|t}G'\Sigma^{-1}_{t+1|t}(x_{t+1} - \mu_{t+1|t}) \\ \nu_\eta - \Gamma_\eta \Sigma_\eta \Sigma^{-1}_{t+1|t}(x_{t+1} - \mu_{t+1|t}) \end{pmatrix}
\end{aligned}$$

and for the right hand side we get

$$\begin{aligned}
\nu_{dt} &= \begin{pmatrix} \nu_{t|t} - \Gamma_{t|t}\Sigma_{t|t}G'\Sigma^{-1}_{t+1|t}(x_{t+1} - G\mu_{t|t} - \mu_\eta) \\ \nu_\eta - \Gamma_\eta + \Gamma_\eta G\Sigma_{t|t}G'\Sigma^{-1}_{t+1|t}(x_{t+1} - G\mu_{t|t} - \mu_\eta) \end{pmatrix} \\
\Delta_{dt} + \Gamma_{dt}\Sigma_{dt}\Gamma'_{dt} &= \begin{pmatrix} \Delta_{t|t} & 0 \\ 0 & \Delta_\eta \end{pmatrix} + \begin{pmatrix} \Gamma_{t|t} \\ -\Gamma_\eta G \end{pmatrix} (\Sigma_{t|t} - \Sigma_{t|t}G'\Sigma^{-1}_{t+1|t}G\Sigma_{t|t}) \begin{pmatrix} \Gamma_{t|t} \\ -\Gamma_\eta G \end{pmatrix}'
\end{aligned}$$

Closer inspection shows that both expressions are the same.

□

Theorem 3.

$$\nu^{top}_{T|T} = \nu_{t+1|t} - \Gamma_{t+1|t}(\mu_{t+1|T} - \mu_{t+1|t}) \tag{2}$$

where $\nu^{top}_{T|T}$ are the top $t+2$ elements of $\nu_{T|T}$ such that the dimensions fit.

Proof. Note that

$$K_t = \mu_{t+1|t+1} - \mu_{t+1|t}$$

<u>For period $T-1$</u>: It can be easily seen that this equation holds:

$$\nu_{T|T} = \nu_{T|T-1} - \Gamma_{T|T-1}(\mu_{T|T} - \mu_{T|T-1})$$

<u>For period $T-2$</u>: We will start with equation above

$$\begin{aligned}
\nu^{top}_{T|T} &= \nu_{T|T-1} - \Gamma_{T|T-1}(\mu_{T|T} - \mu_{T|T-1}) \\
&= \begin{pmatrix} \nu_{T-1|T-1} \\ \nu_\eta \end{pmatrix} - \begin{pmatrix} \Gamma_{T-1|T-1} J_{T-1} \\ \Gamma_\eta \Sigma_\eta \Sigma^{-1}_{T|T-1} \end{pmatrix} (\mu_{T|T} - \mu_{T|T-1})
\end{aligned}$$

We can now take the first row and the corresponding $\nu^{top}_{T|T}$:

$$\begin{aligned}
\nu^{top}_{T|T} &= \nu_{T-1|T-1} - \Gamma_{T-1|T-1} J_{T-1}(\mu_{T|T} - \mu_{T|T-1}) \\
&= \nu_{T-1|T-2} - \Gamma_{T-1|T-2} K_{T-2} - \Gamma_{T-1|T-2} J_{T-1}(\mu_{T|T} - \mu_{T|T-1}) \\
&= \nu_{T-1|T-2} - \Gamma_{T-1|T-2} \Big[K_{T-2} + J_{T-1}(\mu_{T|T} - \mu_{T|T-1}) \Big] \\
&= \nu_{T-1|T-2} - \Gamma_{T-1|T-2} \Big[\mu_{T-1|T-1} - \mu_{T-1|T-2} + J_{T-1}(\mu_{T|T} - \mu_{T|T-1}) \Big] \\
&= \nu_{T-1|T-2} - \Gamma_{T-1|T-2} \Big[\mu_{T-1|T} - \mu_{T-1|T-2} \Big]
\end{aligned}$$

<u>For period $T-3$</u>: We will start with equation above

$$\begin{aligned}
\nu^{top}_{T|T} &= \nu_{T-1|T-2} - \Gamma_{T-1|T-2}(\mu_{T-1|T} - \mu_{T-1|T-2}) \\
&= \begin{pmatrix} \nu_{T-2|T-2} \\ \nu_\eta \end{pmatrix} - \begin{pmatrix} \Gamma_{T-2|T-2} J_{T-2} \\ \Gamma_\eta \Sigma_\eta \Sigma^{-1}_{T-1|T-2} \end{pmatrix} (\mu_{T-1|T} - \mu_{T-1|T-2})
\end{aligned}$$

We can now take the first row and the corresponding $\nu^{top}_{T|T}$:

$$\begin{aligned}
\nu^{top}_{T|T} &= \nu_{T-2|T-2} - \Gamma_{T-2|T-2} J_{T-1}(\mu_{T-1|T} - \mu_{T-1|T-2}) \\
&= \nu_{T-2|T-3} - \Gamma_{T-2|T-3} K_{T-3} - \Gamma_{T-2|T-3} J_{T-2}(\mu_{T-1|T} - \mu_{T-1|T-2}) \\
&= \nu_{T-2|T-3} - \Gamma_{T-2|T-3} \Big[K_{T-3} + J_{T-2}(\mu_{T-1|T} - \mu_{T-1|T-2}) \Big] \\
&= \nu_{T-2|T-3} - \Gamma_{T-2|T-3} \Big[\mu_{T-2|T-2} - \mu_{T-2|T-3} + J_{T-2}(\mu_{T-1|T} - \mu_{T-1|T-2}) \Big] \\
&= \nu_{T-2|T-3} - \Gamma_{T-2|T-3} \Big[\mu_{T-2|T} - \mu_{T-2|T-3} \Big]
\end{aligned}$$

<u>For period (any) t</u>: We will assume the following holds for $t+1$:

$$\nu^{top}_{T|T} = \nu_{t+2|t+1} - \Gamma_{t+2|t+1}(\mu_{t+2|T} - \mu_{t+2|t+1})$$

We will start with equation above

$$\begin{aligned}
\nu^{top}_{T|T} &= \nu_{t+2|t+1} - \Gamma_{t+2|t+1}(\mu_{t+2|T} - \mu_{t+2|t+1}) \\
&= \begin{pmatrix} \nu_{t+1|t+1} \\ \nu_{\eta} \end{pmatrix} - \begin{pmatrix} \Gamma_{t+1|t+1} J_{t+1} \\ \Gamma_{\eta} \Sigma_{\eta} \Sigma^{-1}_{t+2|t+1} \end{pmatrix} (\mu_{t+2|T} - \mu_{t+2|t+1})
\end{aligned}$$

We can now take the first row and the corresponding $\nu^{top}_{T|T}$:

$$\begin{aligned}
\nu^{top}_{T|T} &= \nu_{t+1|t+1} - \Gamma_{t+1|t+1} J_{t+1}(\mu_{t+2|T} - \mu_{t+2|t+1}) \\
&= \nu_{t+1|t} - \Gamma_{t+1|t} K_t - \Gamma_{t+1|t} J_{t+1}(\mu_{t+2|T} - \mu_{t+2|t+1}) \\
&= \nu_{t+1|t} - \Gamma_{t+1|t} \Big[K_t + J_{t+1}(\mu_{t+2|T} - \mu_{t+2|t+1}) \Big] \\
&= \nu_{t+1|t} - \Gamma_{t+1|t} \Big[\mu_{t+1|t+1} - \mu_{t+1|t} + J_{t+1}(\mu_{t+2|T} - \mu_{t+2|t+1}) \Big] \\
&= \nu_{t+1|t} - \Gamma_{t+1|t} \Big[\mu_{t+1|T} - \mu_{t+1|t} \Big]
\end{aligned}$$

□

Lemma 1. *For any t, the following holds,*

$$\Phi\Big[\Gamma_{dt}(x_t-\mu_{dt});\nu_{dt},\Delta_{dt}\Big]=\Phi\Bigg[\begin{pmatrix}\Gamma_{t|t} & 0\\ -\Gamma_\eta G & \Gamma_\eta\end{pmatrix}\left(\begin{pmatrix}x_t\\ x_{t+1}\end{pmatrix}-\begin{pmatrix}\mu_{t|T}\\ \mu_{t+1|T}\end{pmatrix}\right);$$
$$\nu_{t+1|t}-\Gamma_{t+1|t}(\mu_{t+1|T}-\mu_{t+1|t}),$$
$$\begin{pmatrix}\Delta_{t|t} & 0\\ 0 & \Delta_\eta\end{pmatrix}\Bigg] \tag{3}$$

where,

$$\mu_{t|T}\equiv\mu_{t|t}+\Sigma_{t|t}G'\Sigma_{t+1|t}^{-1}(\mu_{t+1|T}-\mu_{t+1|t})$$

which will coincide with first parameter of the distribution of $X_{t|T}$.

Proof. Note that $\Gamma_\eta GJ_t-\Gamma_\eta=-\Gamma_\eta\Sigma_\eta\Sigma_{t+1|t}^{-1}$.

$$\begin{aligned}
&\Phi\Big[\Gamma_{dt}(x_t-\mu_{dt});\nu_{dt},\Delta_{dt}\Big]\\
&=\Phi\left[\begin{pmatrix}\Gamma_{t|t}\\ -\Gamma_\eta G\end{pmatrix}(x_t-\mu_{t|t})-\begin{pmatrix}\Gamma_{t|t}J_t\\ -\Gamma_\eta GJ_t\end{pmatrix}(x_{t+1}-\mu_{t+1|t});\nu_{t+1|t}-\begin{pmatrix}\Gamma_{t|t}J_t\\ \Gamma_\eta\Sigma_\eta\Sigma_{t+1|t}^{-1}\end{pmatrix}(x_{t+1}-\mu_{t+1|t}),\begin{pmatrix}\Delta_{t|t} & 0\\ 0 & \Delta_\eta\end{pmatrix}\right]\\
&=\Phi\Big[\Gamma_{t|t}(x_t-\mu_{t|t});\nu_{t|t},\Delta_{t|t}\Big]\times\Phi\Big[-\Gamma_\eta G(x_t-\mu_{t|t})+\Gamma_\eta(x_{t+1}-\mu_{t+1|t});\nu_\eta,\Delta_\eta\Big]
\end{aligned}$$

Let us expand these cdfs

$$
\begin{aligned}
&\Phi\Big[\Gamma_{t|t}(x_t-\mu_{t|t});\nu_{t|t},\Delta_{t|t}\Big]\times\Phi\Big[-\Gamma_\eta G(x_t-\mu_{t|t})+\Gamma_\eta(x_{t+1}-\mu_{t+1|t});\nu_\eta,\Delta_\eta\Big]\\
&=\Phi\Big[\Gamma_{t|t}(x_t-\mu_{t|t}-J_t(\mu_{t+1|T}-\mu_{t+1|t}));\ \nu_{t|t}-\Gamma_{t|t}J_t(\mu_{t+1|T}-\mu_{t+1|t}),\ \Delta_{t|t}\Big]\\
&\qquad\times\Phi\Big[-\Gamma_\eta G(x_t-\mu_{t|t}-J_t(\mu_{t+1|T}-\mu_{t+1|t})+J_t(\mu_{t+1|T}-\mu_{t+1|t}))+\Gamma_\eta(x_{t+1}-\mu_{t+1|T}+\mu_{t+1|T}-\mu_{t+1|t});\ \nu_\eta,\ \Delta_\eta\Big]\\
&=\Phi\Big[\Gamma_{t|t}(x_t-\mu_{t|T});\ \nu_{t|t}-\Gamma_{t|t}J_t(\mu_{t+1|T}-\mu_{t+1|t}),\ \Delta_{t|t}\Big]\\
&\qquad\times\Phi\Big[-\Gamma_\eta G(x_t-\mu_{t|T})+\Gamma_\eta(x_{t+1}-\mu_{t+1|T})-\Gamma_\eta GJ_t(\mu_{t+1|T}-\mu_{t+1|t})+\Gamma_\eta(\mu_{t+1|T}-\mu_{t+1|t});\ \nu_\eta,\ \Delta_\eta\Big]\\
&=\Phi\Big[\Gamma_{t|t}(x_t-\mu_{t|T});\ \nu_{t|t}-\Gamma_{t|t}J_t(\mu_{t+1|T}-\mu_{t+1|t}),\ \Delta_{t|t}\Big]\\
&\qquad\times\Phi\Big[-\Gamma_\eta G(x_t-\mu_{t|T})+\Gamma_\eta(x_{t+1}-\mu_{t+1|T});\ \nu_\eta+(\Gamma_\eta GJ_t-\Gamma_\eta)(\mu_{t+1|T}-\mu_{t+1|t}),\ \Delta_\eta\Big]\\
&=\Phi\left[\begin{pmatrix}\Gamma_{t|t} & 0\\ -\Gamma_\eta G & \Gamma_\eta\end{pmatrix}\left(\begin{pmatrix}x_t\\ x_{t+1}\end{pmatrix}-\begin{pmatrix}\mu_{t|T}\\ \mu_{t+1|T}\end{pmatrix}\right);\ \nu_{t+1|t}-\Gamma_{t+1|t}(\mu_{t+1|T}-\mu_{t+1|t}),\ \begin{pmatrix}\Delta_{t|t} & 0\\ 0 & \Delta_\eta\end{pmatrix}\right]
\end{aligned}
$$

□

Lemma 2. *As to the product of two normal pdfs, we obtain for any t*

$$
\begin{aligned}
\phi\Big[x_t;\ \mu_{dt},\ \Sigma_{dt}\Big]\times\phi\Big[x_{t+1};\ \mu_{t+1|T},\ \Sigma_{t+1|T}\Big]=\phi\Bigg[&\begin{pmatrix}x_t\\ x_{t+1}\end{pmatrix};\\
&\begin{pmatrix}\mu_{t|t}+J_t(\mu_{t+1|T}-\mu_{t+1|t})\\ \mu_{t+1|T}\end{pmatrix},\\
&\begin{pmatrix}\Sigma_{t|t}+J_t(\Sigma_{t+1|T}-\Sigma_{t+1|t})J_t' & J_t\Sigma_{t+1|T}\\ \Sigma_{t+1|T}J_t' & \Sigma_{t+1|T}\end{pmatrix}\Bigg]
\end{aligned}
\tag{4}
$$

where,

$$
\begin{aligned}
\mu_{t|T}&\equiv\mu_{t|t}+\Sigma_{t|t}G'\Sigma_{t+1|t}^{-1}(\mu_{t+1|T}-\mu_{t+1|t})\\
\Sigma_{t|T}&\equiv\Sigma_{t|t}+\Sigma_{t|t}G'\Sigma_{t+1|t}^{-1}(\Sigma_{t+1|T}-\Sigma_{t+1|t})\Sigma_{t+1|t}^{-1}G\Sigma_{t|t}
\end{aligned}
$$

which will coincide with the first two parameters of the distribution of $X_{t|T}$.

Proof. Recall the conditioning theorem for Normal distribution. Let $X\sim N_p(\psi,\Omega)$ be partitioned into X_1

of length p_1 and X_2 of length p_2, such that $X = (X_1', X_2')'$. The parameters are partitioned accordingly,

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}, \quad \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix},$$

Then,

$$X_{1|2} = (X_1 | X_2 = x_2) \sim N_{p_1}(\psi_{1|2}, \Omega_{1|2})$$

with $\psi_{1|2} = \psi_1 + \Omega_{12}\Omega_{22}^{-1}(x_2 - \psi_2)$ and, $\Omega_{1|2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$.

Now, let us write left hand side of equation 4 in more basic terms and add/subtract some of its expressions,

$$\begin{aligned}
&\phi\Big[x_t;\ \mu_{t|t} + J_t(x_{t+1} - \mu_{t+1|t}),\ \Sigma_{t|t} - J_t G \Sigma_{t|t}\Big] \times \phi\Big[x_{t+1};\ \mu_{t+1|T},\ \Sigma_{t+1|T}\Big] \\
&= \phi\Big[x_t; \\
&\quad \mu_{t|t} + J_t \Sigma_{t+1|t}\Sigma_{t+1|t}^{-1}(x_{t+1} - \mu_{t+1|T} + \mu_{t+1|T} - \mu_{t+1|t}), \\
&\quad \Sigma_{t|t} - J_t \Sigma_{t+1|t} \underbrace{\Sigma_{t+1|t}^{-1} G \Sigma_{t|t}}_{=J_t'} + J_t \Sigma_{t+1|T}\Sigma_{t+1|T}^{-1}\Sigma_{t+1|T}J_t' - \underbrace{J_t\Sigma_{t+1|T}}_{\text{analogous to } \Omega_{12}} \Sigma_{t+1|T}^{-1}\Sigma_{t+1|T}J_t'\Big] \\
&\times \phi\Big[x_{t+1};\ \mu_{t+1|T},\ \Sigma_{t+1|T}\Big] \\
&= \phi\Big[x_t; \\
&\quad \mu_{t|t} + J_t(\mu_{t+1|T} - \mu_{t+1|t}) + \underbrace{J_t\Sigma_{t+1|t}}_{\text{analogous to } \Omega_{12}} \underbrace{\Sigma_{t+1|t}^{-1}}_{\Omega_{22}^{-1}}(x_{t+1} - \mu_{t+1|T}), \\
&\quad \Sigma_{t|t} - J_t\Sigma_{t+1|t}J_t' + J_t\Sigma_{t+1|T}J_t' - J_t\Sigma_{t+1|T}\Sigma_{t+1|T}^{-1}\Sigma_{t+1|T}J_t'\Big] \\
&\times \phi\Big[x_{t+1};\ \mu_{t+1|T},\ \Sigma_{t+1|T}\Big]
\end{aligned}$$

At this point, it is trivial to get the equation of lemma, if we use the reverse of the conditioning theorem for Normal distribution.

□

Lemma 3. *The equality below holds,*

$$\Delta_{dt} + \Gamma_{dt}\Sigma_{dt}\Gamma_{dt}' = \Delta_{t+1|t+1}. \tag{5}$$

Proof. We will write $\Delta_{dt} + \Gamma_{dt}\Sigma_{dt}\Gamma'_{dt}$ in more basic terms and evaluate:

$$\Delta_{dt} + \Gamma_{dt}\Sigma_{dt}\Gamma'_{dt} = \begin{pmatrix} \Delta_{t|t} & 0 \\ 0 & \Delta_\eta \end{pmatrix} + \begin{pmatrix} \Gamma_{t|t} \\ -\Gamma_\eta G \end{pmatrix} (\Sigma_{t|t} - J_t G \Sigma_{t|t}) \begin{pmatrix} \Gamma'_{t|t} & -G'\Gamma'_\eta \end{pmatrix}$$

$$= \begin{pmatrix} \Delta_{t|t} & 0 \\ 0 & \Delta_\eta \end{pmatrix} + \begin{pmatrix} \Gamma_{t|t}(\Sigma_{t|t} - J_t G\Sigma_{t|t})\Gamma'_{t|t} & -\Gamma_{t|t}(\Sigma_{t|t} - J_t G\Sigma_{t|t})G'\Gamma'_\eta \\ -\Gamma_\eta G(\Sigma_{t|t} - J_t G\Sigma_{t|t})\Gamma'_{t|t} & \Gamma_\eta G(\Sigma_{t|t} - J_t G\Sigma_{t|t})G'\Gamma'_\eta \end{pmatrix}$$

$$= \begin{pmatrix} \Delta_{t|t} + \Gamma_{t|t}\Sigma_{t|t}\Gamma'_{t|t} - \Gamma_{t|t}J_t G\Sigma_{t|t}\Gamma'_{t|t} & -\Gamma_{t|t}\Sigma_{t|t}G'\Gamma'_\eta + \Gamma_{t|t}\Sigma_{t|t}G'\Sigma^{-1}_{t+1|t} \underbrace{G\Sigma_{t|t}G'}_{=\Sigma_{t+1|t}-\Sigma_\eta} \Gamma'_\eta \\ -\Gamma_\eta G\Sigma_{t|t}\Gamma'_{t|t} + \Gamma_\eta \underbrace{G\Sigma_{t|t}G'}_{=\Sigma_{t+1|t}-\Sigma_\eta} \Sigma^{-1}_{t+1|t} G\Sigma_{t|t}\Gamma'_{t|t} & \Delta_\eta + \Gamma_\eta \underbrace{G\Sigma_{t|t}G'}_{=\Sigma_{t+1|t}-\Sigma_\eta} \Gamma'_\eta - \Gamma_\eta \underbrace{G\Sigma_{t|t}G'}_{=\Sigma_{t+1|t}-\Sigma_\eta} \Sigma^{-1}_{t+1|t} \underbrace{G\Sigma_{t|t}G'}_{=\Sigma_{t+1|t}-\Sigma_\eta} \Gamma'_\eta \end{pmatrix}$$

$$= \begin{pmatrix} \Delta_{t|t} + \Gamma_{t|t}\Sigma_{t|t}\Gamma'_{t|t} - \Gamma_{t|t}J_t G\Sigma_{t|t}\Gamma'_{t|t} & \Gamma_{t|t}J_t\Sigma_\eta\Gamma'_\eta \\ \Gamma_\eta\Sigma_\eta J'_t\Gamma'_{t|t} & \Delta_\eta + \Gamma_\eta\Sigma_\eta\Gamma'_\eta - \Gamma_\eta\Sigma_\eta\Sigma^{-1}_{t+1|t}\Sigma_\eta\Gamma'_\eta \end{pmatrix}$$

$$= \Delta_{t+1|t}$$

$$= \Delta_{t+1|t+1}$$

□

2.3. Smoothing formulas for period T-1

In the final period, $x_{T|T}$ is both the filtered and the smoothed distribution. This is where the backward recursion kicks in.

First, consider the penultimate period $T-1$. The conditional density of $x_{T-1}|x_T, D_{T-1}$ is (see theorem 1)

$$\frac{\Phi(\Gamma_{dT-1}(x_{T-1} - \mu_{dT-1}); \nu_{dT-1}, \Delta_{dT-1})}{\Phi(0; \nu_{dT-1}, \Delta_{dT-1} + \Gamma_{dT-1}\Sigma_{dT-1}\Gamma'_{dT-1})}\phi(x_{T-1}; \mu_{dT-1}, \Sigma_{dT-1})$$

where both μ_{dT-1} and ν_{dT-1} are functions of x_T. To find the distribution of $x_{T-1}|D_T$ we average the density of $x_{T-1}|x_T, D_{T-1}$ over $x_T|D_T$,

$$\int \frac{\Phi(\Gamma_{dT-1}(x_{T-1} - \mu_{dT-1}); \nu_{dT-1}, \Delta_{dT-1})}{\Phi(0; \nu_{dT-1}, \Delta_{dT-1} + \Gamma_{dT-1}\Sigma_{dT-1}\Gamma'_{dT-1})} \frac{\Phi(\Gamma_{T|T}(x_T - \mu_{T|T}); \nu_{T|T}, \Delta_{T|T})}{\Phi(0; \nu_{T|T}, \Delta_{T|T} + \Gamma_{T|T}\Sigma_{T|T}\Gamma'_{T|T})}$$
$$\times \phi(x_{T-1}; \mu_{dT-1}, \Sigma_{dT-1})\phi(x_T; \mu_{T|T}, \Sigma_{T|T})dx_T.$$

Due to theorem 2, the top right cdf and the bottom left cdf are identical,

$$\Phi(\Gamma_{T|T}(x_T - \mu_{T|T}); \nu_{T|T}, \Delta_{T|T}) = \Phi(0; \nu_{dT-1}, \Delta_{dT-1} + \Gamma_{dT-1}\Sigma_{dT-1}\Gamma'_{dT-1}).$$

Having cancelled the cdfs, the integral simplifies to

$$\int \frac{\Phi(\Gamma_{dT-1}(x_{T-1}-\mu_{dT-1});\nu_{dT-1},\Delta_{dT-1})}{\Phi(0;\nu_{T|T},\Delta_{T|T}+\Gamma_{T|T}\Sigma_{T|T}\Gamma'_{T|T})} \times \phi(x_{T-1};\mu_{dT-1},\Sigma_{dT-1})\phi(x_T;\mu_{T|T},\Sigma_{T|T})dx_T.$$

We first look at the joint distribution of x_{T-1} and x_T (given D_T) and then marginalize. Let μ_{jT-1}, Σ_{jT-1}, Γ_{jT-1}, ν_{jT-1}, Δ_{jT-1} denote the parameters of the joint distribution. Obviously, (after using lemma 2)

$$\mu_{jT-1} = \begin{pmatrix} \mu_{T-1|T-1} + J_{T-1}(\mu_{T|T} - \mu_{T|T-1}) \\ \mu_{T|T} \end{pmatrix}$$

$$\Sigma_{jT-1} = \begin{pmatrix} \Sigma_{T-1|T-1} + J_{T-1}(\Sigma_{T|T} - \Sigma_{T|T-1})J'_{T-1} & J_{T-1}\Sigma_{T|T} \\ \Sigma_{T|T}J'_{T-1} & \Sigma_{T|T} \end{pmatrix}$$

Less obvious, but still relatively straightforward (use the lemma 1 and the lemma 3),

$$\Gamma_{jT-1} = \begin{pmatrix} \Gamma_{T-1|T-1} & 0 \\ -\Gamma_\eta G & \Gamma_\eta \end{pmatrix}$$

$$\nu_{jT-1} = \nu_{T|T}$$

$$\Delta_{jT-1} = \begin{pmatrix} \Delta_{T-1|T-1} & 0 \\ 0 & \Delta_\eta \end{pmatrix}.$$

The marginal distribution of $x_{T-1}|D_T$ is (obtained using lemma 2.3.1 of Genton (2004))

$$\mu_{T-1|T} = \mu_{T-1|T-1} + J_{T-1}(\mu_{T|T} - \mu_{T|T-1})$$

$$\Sigma_{T-1|T} = \Sigma_{T-1|T-1} + J_{T-1}(\Sigma_{T|T} - \Sigma_{T|T-1})J'_{T-1}$$

$$\Gamma_{T-1|T} = \begin{pmatrix} \Gamma_{T-1|T-1} \\ N_{T-1} \end{pmatrix}$$

$$\nu_{T-1|T} = \nu_{T|T}$$

$$\Delta_{T-1|T} = \begin{pmatrix} \Delta_{T-1|T-1} & 0 \\ 0 & \tilde{\Delta}_{T-1} \end{pmatrix}$$

where

$$N_{T-1} = -\Gamma_\eta G + \Gamma_\eta M_{T-1}$$
$$M_{T-1} = \Sigma_{T|T} J'_{T-1} \Sigma^{-1}_{T-1|T}$$
$$L_{T-1} = \Sigma_{T|T} - M_{T-1} \Sigma_{T-1|T} M'_{T-1}$$

and

$$\tilde{\Delta}_{T-1} = \Delta_\eta + \Gamma_\eta L_{T-1} \Gamma'_\eta.$$

2.4. Smoothing formulas for period T-2

Now consider period $T-2$. The conditional density of $x_{T-2}|x_{T-1}, D_{T-2}$ is

$$\frac{\Phi(\Gamma_{dT-2}(x_{T-2} - \mu_{dT-2}); \nu_{dT-2}, \Delta_{dT-2})}{\Phi(0; \nu_{dT-2}, \Delta_{dT-2} + \Gamma_{dT-2}\Sigma_{dT-2}\Gamma'_{dT-2})} \phi(x_{T-2}; \mu_{dT-2}, \Sigma_{dT-2})$$

where μ_{dT-2} and ν_{dT-2} are functions of x_{T-1}. To find the distribution of $x_{T-2}|D_T$ we again derive the joint distribution of x_{T-2} and x_{T-1} given D_T,

$$\frac{\Phi(\Gamma_{dT-2}(x_{T-2} - \mu_{dT-2}); \nu_{dT-2}, \Delta_{dT-2})}{\Phi(0; \nu_{dT-2}, \Delta_{dT-2} + \Gamma_{dT-2}\Sigma_{dT-2}\Gamma'_{dT-2})} \frac{\Phi(\Gamma_{T-1|T}(x_{T-1} - \mu_{T-1|T}); \nu_{T-1|T}, \Delta_{T-1|T})}{\Phi(0; \nu_{T-1|T}, \Delta_{T-1|T} + \Gamma_{T-1|T}\Sigma_{T-1|T}\Gamma'_{T-1|T})}$$
$$\times \phi(x_{T-2}; \mu_{dT-2}, \Sigma_{dT-2}) \phi(x_{T-1}; \mu_{T-1|T}, \Sigma_{T-1|T}).$$

According to theorem 3 and lemma 1, for $t = T-2$, the top left cdf can be rewritten as

$$\Phi(\Gamma_{dT-2}(x_{T-2} - \mu_{dT-2}); \nu_{dT-2}, \Delta_{dT-2})$$
$$= \Phi\left(\begin{pmatrix} \Gamma_{T-2|T-2} & 0 \\ -\Gamma_\eta G & \Gamma_\eta \end{pmatrix} \left(\begin{pmatrix} x_{T-2} \\ x_{T-1} \end{pmatrix} - \begin{pmatrix} \mu_{T-2|T} \\ \mu_{T-1|T} \end{pmatrix}\right); \nu^{top}_{T|T}, \begin{pmatrix} \Delta_{T-2|T-2} & 0 \\ 0 & \Delta_\eta \end{pmatrix}\right)$$

The top right cdf can be factorized as

$$\Phi(\Gamma_{T-1|T}(x_{T-1} - \mu_{T-1|T}); \nu_{T-1|T}, \Delta_{T-1|T}) = \Phi\left(\begin{pmatrix} \Gamma_{T-1|T-1} \\ N_{T-1} \end{pmatrix} (x_{T-1} - \mu_{T-1|T}); \nu_{T|T}, \begin{pmatrix} \Delta_{T-1|T-1} & 0 \\ 0 & \tilde{\Delta}_{T-1} \end{pmatrix}\right)$$
$$= \Phi(\Gamma_{T-1|T-1}(x_{T-1} - \mu_{T-1|T}); \nu^{top}_{T|T}, \Delta_{T-1|T-1}) \quad (6)$$
$$\times \Phi(N_{T-1}(x_{T-1} - \mu_{T-1|T}); \nu^{btm}_{T|T}, \tilde{\Delta}_{T-1}) \quad (7)$$

where $\nu_{T|T}$ is suitably partitioned into $\nu^{top}_{T|T}$ and $\nu^{btm}_{T|T}$.

Next, turn to the bottom left cdf. Using theorem 2 and theorem 3 it can be shown that

$$
\begin{aligned}
&\Phi(0; \nu_{dT-2}, \Delta_{dT-2} + \Gamma_{dT-2}\Sigma_{dT-2}\Gamma'_{dT-2}) \\
&= \Phi(\Gamma_{T-1|T-1}(x_{T-1} - \mu_{T-1|T-1}); \nu_{T-1|T-1}, \Delta_{T-1|T-1}) \\
&= \Phi(\Gamma_{T-1|T-1}(x_{T-1} - \mu_{T-1|T} + \mu_{T-1|T} - \mu_{T-1|T-1}); \nu_{T-1|T-1}, \Delta_{T-1|T-1}) \\
&= \Phi(\Gamma_{T-1|T-1}(x_{T-1} - \mu_{T-1|T}); \nu_{T-1|T-1} - \Gamma_{T-1|T-1}(\mu_{T-1|T} - \mu_{T-1|T-1}), \Delta_{T-1|T-1})
\end{aligned}
$$

and

$$
\begin{aligned}
&\nu_{T-1|T-1} - \Gamma_{T-1|T-1}(\mu_{T-1|T} - \mu_{T-1|T-1}) \\
&= \nu_{T-1|T-2} - \Gamma_{T-1|T-2}K_{T-2} - \Gamma_{T-1|T-2}(\mu_{T-1|T} - \mu_{T-1|T-2} - K_{T-2}) \\
&= \nu_{T-1|T-2} - \Gamma_{T-1|T-2}(\mu_{T-1|T} - \mu_{T-1|T-2}) \\
&= \nu^{top}_{T|T}
\end{aligned}
$$

As a result, we have

$$
\begin{aligned}
&\Phi(0; \nu_{dT-2}, \Delta_{dT-2} + \Gamma_{dT-2}\Sigma_{dT-2}\Gamma'_{dT-2}) \\
&= \Phi(\Gamma_{T-1|T-1}(x_{T-1} - \mu_{T-1|T}); \nu^{top}_{T|T}, \Delta_{T-1|T-1})
\end{aligned}
$$

Therefore, the first part of equation 6 and bottom left cdf cancel each other out.

The remaining part of the top right cdf can be merged into the top left cdf. This results in the numerator

$$
\Phi\left(\begin{pmatrix} \Gamma_{T-2|T-2} & 0 \\ -\Gamma_\eta G & \Gamma_\eta \\ 0 & N_{T-1} \end{pmatrix} \begin{pmatrix} x_{T-2} - \mu_{T-2|T} \\ x_{T-1} - \mu_{T-1|T} \end{pmatrix}; \nu_{T|T}, \begin{pmatrix} \Delta_{T-2|T-2} & 0 & 0 \\ 0 & \Delta_\eta & 0 \\ 0 & 0 & \tilde{\Delta}_{T-1} \end{pmatrix}\right)
$$

Let μ_{jT-2} etc denote the parameters of the joint distribution,

$$
\begin{aligned}
\mu_{jT-2} &= \begin{pmatrix} \mu_{T-2|T-2} + J_{T-2}(\mu_{T-1|T} - \mu_{T-1|T-2}) \\ \mu_{T-1|T} \end{pmatrix} \\
\Sigma_{jT-2} &= \begin{pmatrix} \Sigma_{T-2|T-2} + J_{T-2}(\Sigma_{T-1|T} - \Sigma_{T-1|T-2})J'_{T-2} & J_{T-2}\Sigma_{T-1|T} \\ \Sigma_{T-1|T}J'_{T-2} & \Sigma_{T-1|T} \end{pmatrix} \\
\Gamma_{jT-2} &= \begin{pmatrix} \Gamma_{T-2|T-2} & 0 \\ -\Gamma_\eta G & \Gamma_\eta \\ 0 & N_{T-1} \end{pmatrix} \\
\nu_{jT-2} &= \nu_{T|T} \\
\Delta_{jT-2} &= \begin{pmatrix} \Delta_{T-2|T-2} & 0 & 0 \\ 0 & \Delta_\eta & 0 \\ 0 & 0 & \tilde{\Delta}_{T-1} \end{pmatrix}.
\end{aligned}
$$

The marginal distribution of $x_{T-2}|D_T$ is

$$
\begin{aligned}
\mu_{T-2|T} &= \mu_{T-2|T-2} + J_{T-2}(\mu_{T-1|T} - \mu_{T-1|T-2}) \\
\Sigma_{T-2|T} &= \Sigma_{T-2|T-2} + J_{T-2}(\Sigma_{T-1|T} - \Sigma_{T-1|T-2})J'_{T-2} \\
\Gamma_{T-2|T} &= \begin{pmatrix} \Gamma_{T-2|T-2} \\ N_{T-2} \\ N_{T-1}M_{T-2} \end{pmatrix} \\
\nu_{T-2|T} &= \nu_{T|T} \\
\Delta_{T-2|T} &= \begin{pmatrix} \Delta_{T-2|T-2} & 0 \\ 0 & \tilde{\Delta}_{T-2} \end{pmatrix}.
\end{aligned}
$$

with

$$
\begin{aligned}
N_{T-2} &= -\Gamma_\eta G + \Gamma_\eta M_{T-2} \\
M_{T-2} &= \Sigma_{T-1|T}J'_{T-2}\Sigma^{-1}_{T-2|T} \\
L_{T-2} &= \Sigma_{T-1|T} - M_{T-2}\Sigma_{T-2|T}M'_{T-2}
\end{aligned}
$$

and

$$
\tilde{\Delta}_{T-2} = \begin{pmatrix} \Delta_\eta & 0 \\ 0 & \tilde{\Delta}_{T-1} \end{pmatrix} + \begin{pmatrix} \Gamma_\eta \\ N_{T-1} \end{pmatrix} L_{T-2} \begin{pmatrix} \Gamma_\eta \\ N_{T-1} \end{pmatrix}'.
$$

2.5. Smoothing formulas for period T-3

Consider period $T-3$. The conditional density of $x_{T-3}|x_{T-2}, D_{T-3}$ is

$$\frac{\Phi(\Gamma_{dT-3}(x_{T-3}-\mu_{dT-3});\nu_{dT-3},\Delta_{dT-3})}{\Phi(0;\nu_{dT-3},\Delta_{dT-3}+\Gamma_{dT-3}\Sigma_{dT-3}\Gamma'_{dT-3})}\phi(x_{T-3};\mu_{dT-3},\Sigma_{dT-3})$$

where μ_{dT-3} and ν_{dT-3} are functions of x_{T-2}. To find the distribution of $x_{T-3}|D_T$ we again derive the joint distribution of x_{T-3} and x_{T-2} given D_T,

$$\frac{\Phi(\Gamma_{dT-3}(x_{T-3}-\mu_{dT-3});\nu_{dT-3},\Delta_{dT-3})}{\Phi(0;\nu_{dT-3},\Delta_{dT-3}+\Gamma_{dT-3}\Sigma_{dT-3}\Gamma'_{dT-3})}\frac{\Phi(\Gamma_{T-2|T}(x_{T-2}-\mu_{T-2|T});\nu_{T-2|T},\Delta_{T-2|T})}{\Phi(0;\nu_{T-2|T},\Delta_{T-2|T}+\Gamma_{T-2|T}\Sigma_{T-2|T}\Gamma'_{T-2|T})}$$
$$\times\phi(x_{T-3};\mu_{dT-3},\Sigma_{dT-3})\phi(x_{T-2};\mu_{T-2|T},\Sigma_{T-2|T}).$$

According to theorem 3 and lemma 1 for $t=T-3$, the top left cdf can be rewritten as

$$\begin{aligned}&\Phi(\Gamma_{dT-3}(x_{T-3}-\mu_{dT-3});\nu_{dT-3},\Delta_{dT-3})\\&=\Phi\left(\begin{pmatrix}\Gamma_{T-3|T-3} & 0\\ -\Gamma_\eta G & \Gamma_\eta\end{pmatrix}\left(\begin{pmatrix}x_{T-3}\\ x_{T-2}\end{pmatrix}-\begin{pmatrix}\mu_{T-3|T}\\ \mu_{T-2|T}\end{pmatrix}\right);\nu^{top}_{T|T},\begin{pmatrix}\Delta_{T-3|T-3} & 0\\ 0 & \Delta_\eta\end{pmatrix}\right)\end{aligned}$$

The top right cdf can be factorized as

$$\begin{aligned}\Phi(\Gamma_{T-2|T}(x_{T-2}-\mu_{T-2|T});\nu_{T-2|T},\Delta_{T-2|T})&=\Phi\left(\begin{pmatrix}\Gamma_{T-2|T-2}\\ N_{T-2}\\ N_{T-1}M_{T-2}\end{pmatrix}(x_{T-2}-\mu_{T-2|T});\nu_{T|T},\begin{pmatrix}\Delta_{T-2|T-2} & 0\\ 0 & \tilde{\Delta}_{T-2}\end{pmatrix}\right)\\&=\Phi(\Gamma_{T-2|T-2}(x_{T-2}-\mu_{T-2|T});\nu^{top}_{T|T},\Delta_{T-2|T-2})\\&\quad\times\Phi\left(\begin{pmatrix}N_{T-2}\\ N_{T-1}M_{T-2}\end{pmatrix}(x_{T-2}-\mu_{T-2|T});\nu^{btm}_{T|T},\tilde{\Delta}_{T-2}\right)\end{aligned}$$

where $\nu_{T|T}$ is suitably partitioned into $\nu^{top}_{T|T}$ and $\nu^{btm}_{T|T}$.

Next, turn to the bottom left cdf. Using theorem 2 and theorem 3, it can be shown that as we did in the last section

$$\begin{aligned}&\Phi(0;\nu_{dT-3},\Delta_{dT-3}+\Gamma_{dT-3}\Sigma_{dT-3}\Gamma'_{dT-3})\\&=\Phi(\Gamma_{T-2|T-2}(x_{T-2}-\mu_{T-2|T});\nu^{top}_{T|T},\Delta_{T-2|T-2})\end{aligned}$$

Therefore, the first part of factorization of top right cdf and bottom left cdf cancel each other out. The

remaining part of the top right cdf can be merged into the top left cdf. This results in the numerator

$$\Phi\left(\begin{pmatrix}\Gamma_{T-3|T-3} & 0\\ -\Gamma_\eta G & \Gamma_\eta\\ 0 & N_{T-2}\\ 0 & N_{T-1}M_{T-2}\end{pmatrix}\begin{pmatrix}x_{T-3}-\mu_{T-3|T}\\ x_{T-2}-\mu_{T-2|T}\end{pmatrix};\nu_{T|T},\begin{pmatrix}\Delta_{T-3|T-3} & 0 & 0\\ 0 & \Delta_\eta & 0\\ 0 & 0 & \tilde{\Delta}_{T-2}\end{pmatrix}\right)$$

Let μ_{jT-3} etc denote the parameters of the joint distribution,

$$\begin{aligned}
\mu_{jT-3} &= \begin{pmatrix}\mu_{T-3|T-3}+J_{T-3}(\mu_{T-2|T}-\mu_{T-2|T-3})\\ \mu_{T-2|T}\end{pmatrix}\\
\Sigma_{jT-3} &= \begin{pmatrix}\Sigma_{T-3|T-3}+J_{T-3}(\Sigma_{T-2|T}-\Sigma_{T-2|T-3})J'_{T-3} & J_{T-3}\Sigma_{T-2|T}\\ \Sigma_{T-2|T}J'_{T-3} & \Sigma_{T-2|T}\end{pmatrix}\\
\Gamma_{jT-3} &= \begin{pmatrix}\Gamma_{T-3|T-3} & 0\\ -\Gamma_\eta G & \Gamma_\eta\\ 0 & N_{T-2}\\ 0 & N_{T-1}M_{T-2}\end{pmatrix}\\
\nu_{jT-3} &= \nu_{T|T}\\
\Delta_{jT-3} &= \begin{pmatrix}\Delta_{T-3|T-3} & 0 & 0\\ 0 & \Delta_\eta & 0\\ 0 & 0 & \tilde{\Delta}_{T-2}\end{pmatrix}.
\end{aligned}$$

The marginal distribution of $x_{T-3}|D_T$ is

$$\begin{aligned}
\mu_{T-3|T} &= \mu_{T-3|T-3}+J_{T-3}(\mu_{T-2|T}-\mu_{T-2|T-3})\\
\Sigma_{T-3|T} &= \Sigma_{T-3|T-3}+J_{T-3}(\Sigma_{T-2|T}-\Sigma_{T-2|T-3})J'_{T-3}\\
\Gamma_{T-3|T} &= \begin{pmatrix}\Gamma_{T-3|T-3}\\ N_{T-3}\\ N_{T-2}M_{T-3}\\ N_{T-1}M_{T-2}M_{T-3}\end{pmatrix}\\
\nu_{T-3|T} &= \nu_{T|T}\\
\Delta_{T-3|T} &= \begin{pmatrix}\Delta_{T-3|T-3} & 0\\ 0 & \tilde{\Delta}_{T-3}\end{pmatrix}.
\end{aligned}$$

with

$$\tilde{\Delta}_{T-3} = \begin{pmatrix} \Delta_\eta & 0 \\ 0 & \tilde{\Delta}_{T-2} \end{pmatrix} + \begin{pmatrix} \Gamma_\eta \\ N_{T-2} \\ N_{T-1}M_{T-2} \end{pmatrix} L_{T-3} \begin{pmatrix} \Gamma_\eta \\ N_{T-2} \\ N_{T-1}M_{T-2} \end{pmatrix}'$$

(note that the dimension of $\tilde{\Delta}_{T-2}$ is just large enough to make the sum fit).

2.6. Smoothing formulas for period T-4

Guess for CSN parameters of $x_{T-4}|D_T$:

$$\begin{aligned}
\mu_{T-4|T} &= \mu_{T-4|T-4} + J_{T-4}(\mu_{T-3|T} - \mu_{T-3|T-4}) \\
\Sigma_{T-4|T} &= \Sigma_{T-4|T-4} + J_{T-4}(\Sigma_{T-3|T} - \Sigma_{T-3|T-4})J'_{T-4} \\
\Gamma_{T-4|T} &= \begin{pmatrix} \Gamma_{T-4|T-4} \\ N_{T-4} \\ N_{T-3}M_{T-4} \\ N_{T-2}M_{T-3}M_{T-4} \\ N_{T-1}M_{T-2}M_{T-3}M_{T-4} \end{pmatrix} \\
\nu_{T-4|T} &= \nu_{T|T} \\
\Delta_{T-4|T} &= \begin{pmatrix} \Delta_{T-4|T-4} & 0 \\ 0 & \tilde{\Delta}_{T-4} \end{pmatrix}.
\end{aligned}$$

with

$$\tilde{\Delta}_{T-4} = \begin{pmatrix} \Delta_\eta & 0 \\ 0 & \tilde{\Delta}_{T-3} \end{pmatrix} + \begin{pmatrix} \Gamma_\eta \\ N_{T-3} \\ N_{T-2}M_{T-3} \\ N_{T-1}M_{T-2}M_{T-3} \end{pmatrix} L_{T-4} \begin{pmatrix} \Gamma_\eta \\ N_{T-3} \\ N_{T-2}M_{T-3} \\ N_{T-1}M_{T-2}M_{T-3} \end{pmatrix}'$$

2.7. Smoothing formulas for any time period

The CSN parameters for $x_t|D_T$ are

$$
\begin{aligned}
\mu_{t|T} &= \mu_{t|t} + J_t(\mu_{t+1|T} - \mu_{t+1|t}) \\
\Sigma_{t|T} &= \Sigma_{t|t} + J_t(\Sigma_{t+1|T} - \Sigma_{t+1|t})J_t' \\
\Gamma_{t|T} &= \begin{pmatrix} \Gamma_{t|t} \\ N_t \\ N_{t+1}M_t \\ N_{t+2}M_{t+1}M_t \\ \vdots \\ N_{T-1} \cdot \ldots \cdot M_{t+2}M_{t+1}M_t \end{pmatrix} \\
\nu_{t|T} &= \nu_{T|T} \\
\Delta_{t|T} &= \begin{pmatrix} \Delta_{t|t} & 0 \\ 0 & \tilde{\Delta}_t \end{pmatrix}.
\end{aligned}
$$

with

$$
\tilde{\Delta}_t = \begin{pmatrix} \Delta_\eta & 0 \\ 0 & \tilde{\Delta}_{t+1} \end{pmatrix} + \begin{pmatrix} \Gamma_\eta \\ N_{t+1} \\ N_{t+2}M_{t+1} \\ N_{T-1} \cdot \ldots \cdot M_{t+2}M_{t+1} \end{pmatrix} L_t \begin{pmatrix} \Gamma_\eta \\ N_{t+1} \\ N_{t+2}M_{t+1} \\ N_{T-1} \cdot \ldots \cdot M_{t+2}M_{t+1} \end{pmatrix}'
$$

with

$$
\begin{aligned}
J_t &= \Sigma_{t|t}G'\Sigma_{t+1|t}^{-1} \\
M_t &= \Sigma_{t+1|T}J_t'\Sigma_{t|T}^{-1} \\
N_t &= -\Gamma_\eta G + \Gamma_\eta M_t \\
L_t &= \Sigma_{t+1|T} - M_t\Sigma_{t|T}M_t'
\end{aligned}
$$

2.8. More compact formulas for any time period

We can write the formulas for general t more neatly via the following steps:

- Replace J_t with $\Sigma_{t|t}G'\Sigma_{t+1|t}^{-1}$.

- Replace L_t with $\Sigma_{t+1|T} - M_t \Sigma_{t|T} M_t'$
- Define $O_t \equiv \begin{bmatrix} N_t \\ O_{t+1} M_t \end{bmatrix}$ with $O_{T-1} \equiv N_{T-1}$ (Note that O_T is not defined)

After implementing the above steps, it is easy to see that we get:

$$
\begin{aligned}
\mu_{t|T} &= \mu_{t|t} + \Sigma_{t|t} G' \Sigma_{t+1|t}^{-1} (\mu_{t+1|T} - \mu_{t+1|t}) \\
\Sigma_{t|T} &= \Sigma_{t|t} + \Sigma_{t|t} G' \Sigma_{t+1|t}^{-1} (\Sigma_{t+1|T} - \Sigma_{t+1|t}) \Sigma_{t+1|t}^{-1} G \Sigma_{t|t} \\
\Gamma_{t|T} &= \begin{pmatrix} \Gamma_{t|t} \\ O_t \end{pmatrix} \\
\nu_{t|T} &= \nu_{T|T} \\
\Delta_{t|T} &= \begin{pmatrix} \Delta_{t|t} & 0 \\ 0 & \tilde{\Delta}_t \end{pmatrix}.
\end{aligned}
$$

with

$$
\tilde{\Delta}_t = \begin{pmatrix} \Delta_\eta & 0 \\ 0 & \tilde{\Delta}_{t+1} \end{pmatrix} + \begin{pmatrix} \Gamma_\eta \\ O_{t+1} \end{pmatrix} (\Sigma_{t+1|T} - M_t \Sigma_{t|T} M_t') \begin{pmatrix} \Gamma_\eta \\ O_{t+1} \end{pmatrix}'
$$

References

Arellano-Valle, R. B., & Azzalini, A. (2008). The centred parametrization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis*, *99*, 1362–1382. doi:10.1016/j.jmva.2008.01.020.

Atkinson, T., Richter, A., & Throckmorton, N. (2019). The zero lower bound and estimation accuracy. *Journal of Monetary Economics*, *115*, 249–264. doi:10.1016/j.jmoneco.2019.06.007.

Flecher, C., Naveau, P., & Allard, D. (2009). Estimating the closed skew-normal distribution parameters using weighted moments. *Statistics & Probability Letters*, *79*, 1977–1984. doi:10.1016/j.spl.2009.06.004.

Genton, M. G. (2004). *Skew-Elliptical Distributions and Their Applications - A Journey Beyond Normality*. S.l.: CRC PRESS.

Harvey, A. C., & Phillips, G. D. A. (1979). Maximum Likelihood Estimation of Regression Models with Autoregressive- Moving Average Disturbances. *Biometrika*, *66*, 49. doi:10.2307/2335241.

Käärik, M., Selart, A., & Käärik, E. (2015). On Parametrization of Multivariate Skew-Normal Distribution. *Communications in Statistics - Theory and Methods*, *44*, 1869–1885. doi:10.1080/03610926.2012.760277.

Liseo, B., & Parisi, A. (2013). Bayesian inference for the multivariate skew-normal model: A population Monte Carlo approach. *Computational Statistics & Data Analysis*, *63*, 125–138. doi:10.1016/j.csda.2013.02.007.

Rezaie, J., & Eidsvik, J. (2014). Kalman filter variants in the closed skew normal setting. *Computational Statistics & Data Analysis*, *75*, 1–14. doi:10.1016/j.csda.2014.01.014.