Nonlinear Expectations: Making Sense of Professional Forecasts

APPENDIX

Julio Ortiz

Boston University

Contents

Appendix A Proofs

Proof of Proposition 1

We have the following expressions for forecast errors and revisions, respectively:

$$\mathbf{FE^i} = \mathbf{CA}(\mathbf{I} - \kappa\mathbf{C})(\mathbf{s_t} - \mathbf{s^i_{t|t-1}}) + \mathbf{CBw_{t+1}} + \mathbf{Dv^i_{t+1}} - \mathbf{CA}\kappa\mathbf{Dv^i_t}$$

$$\mathbf{FR^i} = \mathbf{CA}\kappa\mathbf{Dv^i_t} + \mathbf{CA}\kappa\mathbf{C}(\mathbf{s_t} - \mathbf{s^i_{t|t-1}})$$

Then,

(a) $\beta_1 \propto \text{Cov}(FE^i, FR^i) = \mathbf{CA(I - \kappa C)\Psi(CAKC)^\top - CAK(Dv^i_t v^i_t D)(CA\kappa)^\top}$

where $\mathbf{\Psi}$ denotes the state estimation error variance. This becomes

$$\begin{aligned} \beta_1 \quad &\propto \quad CA(I - \kappa C)\Psi(CA\kappa C)^\top - CAK(Dv^i_t v^i_t D)(CA\kappa)^\top \\ &= CA\Big\{(I - \kappa C)\Psi C - \kappa D v^i_t v^i_t D\Big\}(CA\kappa)^\top \\ \beta_1 \quad &= 0 \end{aligned}$$

because the term in brackets is zero by the definition of the Kalman gain.

(b) Denoting $\overline{FE}$ and $\overline{FR}$ as the cross-sectional mean of the forecast error and revision, respectively, we have

$$\alpha_1 \propto \text{Cov}(\overline{\mathbf{FE}}, \overline{\mathbf{FR}}) = CA(I - \kappa C)\overline{\Psi}(CA\kappa C)^\top$$

The variance of the average revision is $\text{Var}(\overline{FR}) = CA\kappa C\Psi(CA\kappa C)^\top$ Thus, we have

$$\alpha_1 = CA(I - \kappa C)(CA\kappa C)^{-1}$$

Proof of Proposition 2

Recall that

$$\alpha_1 = [(\widehat{\mathbf{z}}_{t+h|t} - \widehat{\mathbf{z}}_{t+h|t-1})^\top(\widehat{\mathbf{z}}_{t+h|t} - \widehat{\mathbf{z}}_{t+h|t-1})]^{-1}(\widehat{\mathbf{z}}_{t+h|t} - \widehat{\mathbf{z}}_{t+h|t-1})^\top(\mathbf{z}_{t+h} - \widehat{\mathbf{z}}^i_{t+h|t})$$

and

$$\beta_1 = [(\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1})^\top (\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1})]^{-1} (\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1})^\top (\mathbf{z}_{t+h} - \widehat{\mathbf{z}}^i_{t+h|t})$$

To prove the proposition, I will show that the covariance between consensus errors and revisions is weakly greater than that for pooled errors and revisions. I will then show that the variance of the consensus revision is weakly smaller than the variance of the pooled variance.

We can express the covariance between errors and revisions as

$$\begin{aligned}\mathbb{C}(\mathbf{z}_{t+h} - \widehat{\mathbf{z}}^i_{t+h|t}, \widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}) &= \int\int (\mathbf{z}_{t+h} - \widehat{\mathbf{z}}^i_{t+h|t})(\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}) didt \\ &- \int\int (\mathbf{z}_{t+h} - \widehat{\mathbf{z}}^i_{t+h|t}) didt - \int\int (\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}) didt\end{aligned}$$

and at the consensus level

$$\begin{aligned}\mathbb{C}(\mathbf{z}_{t+h} - \widehat{\mathbf{z}}_{t+h|t}, \widehat{\mathbf{z}}_{t+h|t} - \widehat{\mathbf{z}}_{t+h|t-1}) &= \int \left(\mathbf{z}_{t+h} - \int \widehat{\mathbf{z}}^i_{t+h|t} di\right)\left(\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}] di\right) dt \\ &- \int \left(\mathbf{z}_{t+h} - \int \widehat{\mathbf{z}}^i_{t+h|t} di\right) dt - \int\int (\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}_{t+h|t-1}) didt\end{aligned}$$

We wish to show that the second equation is weakly greater than the first. One can note immediately that the second and third terms of both equations are equal (given the linearity of the expectations operator), and so they cancel out. The resulting inequality that we wish to verify is

$$\int \left(\mathbf{z}_{t+h} - \int \widehat{\mathbf{z}}^i_{t+h|t} di\right)\left(\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}] di\right) dt \geq \int\int (\mathbf{z}_{t+h} - \widehat{\mathbf{z}}^i_{t+h|t})(\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}) didt$$

By distributing the revision into the error on either side of the inequality, we can express each side as the sum of two terms. The first of these will drop out as we will have

$$\int \mathbf{z}_{t+h}\left(\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}] di\right) dt$$

on the LHS and

$$\int \int \mathbf{z}_{t+h}(\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}) di dt$$

on the RHS. Again, due to the linearity of the expectations operator, these terms cancel out. The remaining inequality is therefore

$$\begin{aligned}
-\int \left(\int \widehat{\mathbf{z}}^i_{t+h|t} di \right) \left(\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}] di \right) dt &\geq -\int \int \widehat{\mathbf{z}}^i_{t+h|t}(\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}) di dt \\
\int \left(\int \widehat{\mathbf{z}}^i_{t+h|t} di \right) \left(\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}] di \right) dt &\leq \int \int \widehat{\mathbf{z}}^i_{t+h|t}(\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}) di dt \\
\int \left(\int \widehat{\mathbf{z}}^i_{t+h|t} di \right)^2 dt - \int \left(\int \widehat{\mathbf{z}}^i_{t+h|t} di \right) \left(\int \widehat{\mathbf{z}}^i_{t+h|t-1} di \right) dt &\leq \int \int \widehat{\mathbf{z}}^{i2}_{t+h|t} di dt - \int \int \widehat{\mathbf{z}}^i_{t+h|t} \widehat{\mathbf{z}}^i_{t+h|t-1} di dt \\
\int \int \widehat{\mathbf{z}}^i_{t+h|t} \widehat{\mathbf{z}}^i_{t+h|t-1} di dt - \int \left(\int \widehat{\mathbf{z}}^i_{t+h|t} di \right) \left(\int \widehat{\mathbf{z}}^i_{t+h|t-1} di \right) dt &\leq \int \int \widehat{\mathbf{z}}^{i2}_{t+h|t} di dt - \int \left(\int \widehat{\mathbf{z}}^i_{t+h|t} di \right)^2 dt \\
\int \left[\int \widehat{\mathbf{z}}^i_{t+h|t} \widehat{\mathbf{z}}^i_{t+h|t-1} di - \left(\int \widehat{\mathbf{z}}^i_{t+h|t} di \right) \left(\int \widehat{\mathbf{z}}^i_{t+h|t-1} di \right) \right] dt &\leq \int \left[\int \widehat{\mathbf{z}}^{i2}_{t+h|t} di - \left(\int \widehat{\mathbf{z}}^i_{t+h|t} di \right)^2 \right] dt
\end{aligned}$$

which is true since the terms in hard brackets on the RHS is the cross-sectional variance of the forecast whereas the term in hard brackets on the LHS is a cross-sectional covariance. Hence, the covariance of the consensus errors with consensus revisions is weakly greater than the covariance of individual-level pooled errors and revisions.

Finally, I show that the variance of the consensus revision is weakly smaller than the variance of the pooled revision. This is simpler to verify. Note that

$$\mathbb{V}(\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}) = \int \int (\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1})^2 di dt - \left(\int \int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}] di dt \right)^2$$

and

$$\mathbb{V}(\widehat{\mathbf{z}}_{t+h|t} - \widehat{\mathbf{z}}_{t+h|t-1}) = \int \left(\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}] di \right)^2 dt - \left(\int \left[\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}] di \right] dt \right)^2$$

Once again, the second term in each of the above revision variance equations will cancel out. The resulting condition that we wish to verify is

$$\int \left(\int [\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1}] di \right)^2 dt \leq \int \int (\widehat{\mathbf{z}}^i_{t+h|t} - \widehat{\mathbf{z}}^i_{t+h|t-1})^2 di dt$$

which holds by Jensen's inequality.

Appendix B Empirics

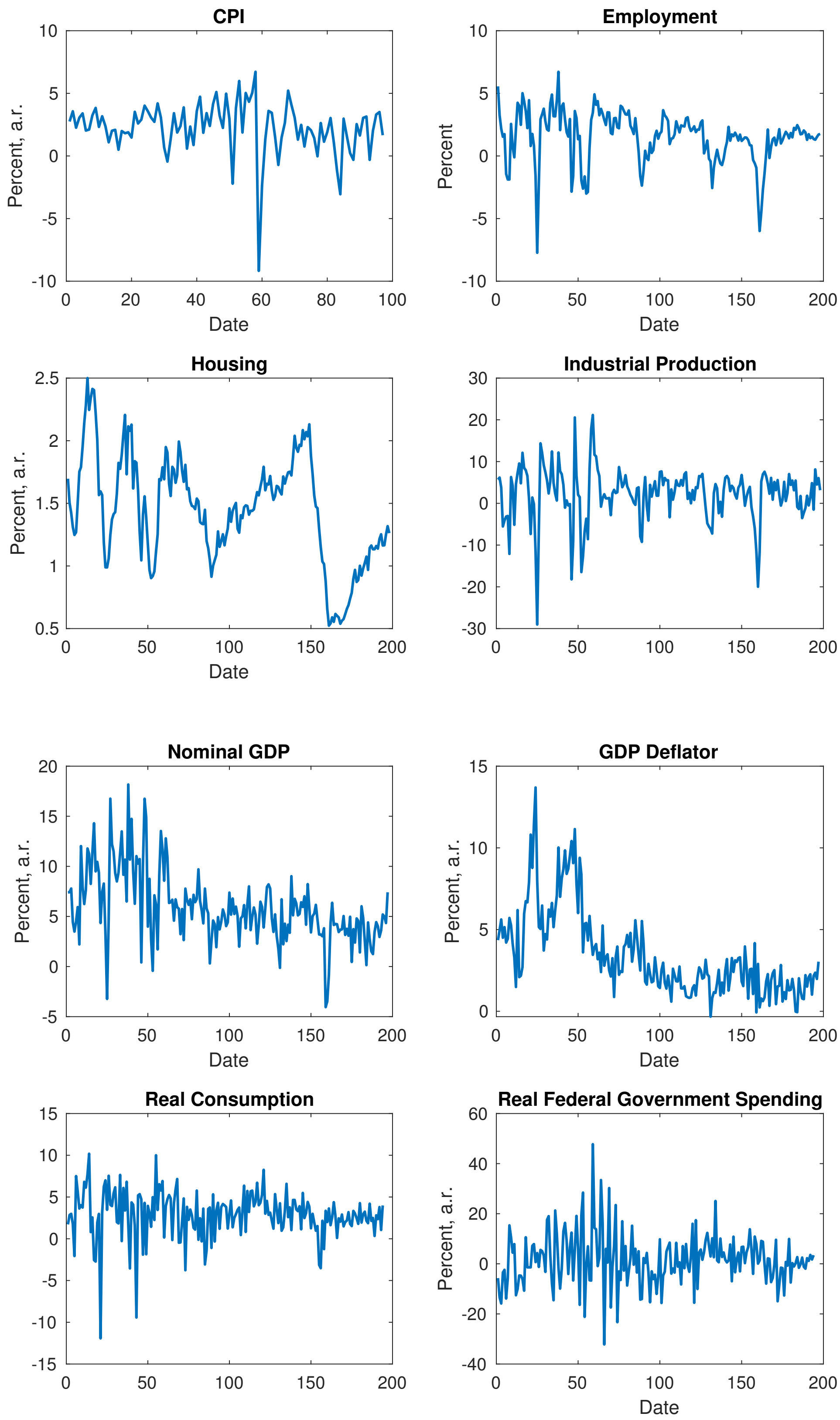
B.1 SPF: Variable Descriptions

While the paper focuses on inflation forecasts based on the GDP deflator, in this subsection I report additional results that make use of several other variables. Before presenting these results, I provide the variable descriptions below:

- NGDP–Quarterly nominal GDP growth forecast (seasonally adjusted, annual rate). Prior to 1992, these are forecasts for nominal GNP.
- RGDP–Quarterly real GDP growth forecast (seasonally adjusted, annual rate).
- PGDP–Quarterly GDP price index growth forecast (seasonally adjusted, annual rate). From 1992 - 1995, GDP implicit deflator is used, and prior to 1992, GNP implicit deflator.
- UNEMP–Forecasts for the quarterly average unemployment rate (seasonally adjusted, average of underlying monthly levels).
- EMP–Quarterly average growth of nonfarm payroll employment (seasonally adjusted, average of underlying monthly levels).
- RNRESIN–Quarterly growth forecast of real nonresidential fixed investment. Also known as business fixed investment (seasonally adjusted, annual rate).
- RRESINV–Quarterly growth forecast of real residential fixed investment (seasonally adjusted, annual rate).
- TBILL–Quarterly forecast of average three-month Treasury bill rate (percentage points, average of underlying daily levels).
- HOUSING– Quarterly growth forecast of average housing starts (seasonally adjusted, annual rate, average of underlying monthly levels).
- CPI–Quarterly forecasts of the headline CPI inflation rate (percentage points, seasonally adjusted, annual rate). Quarterly forecasts are annualized q/q percent changes of quarterly average price index level (average of underlying monthly levels).

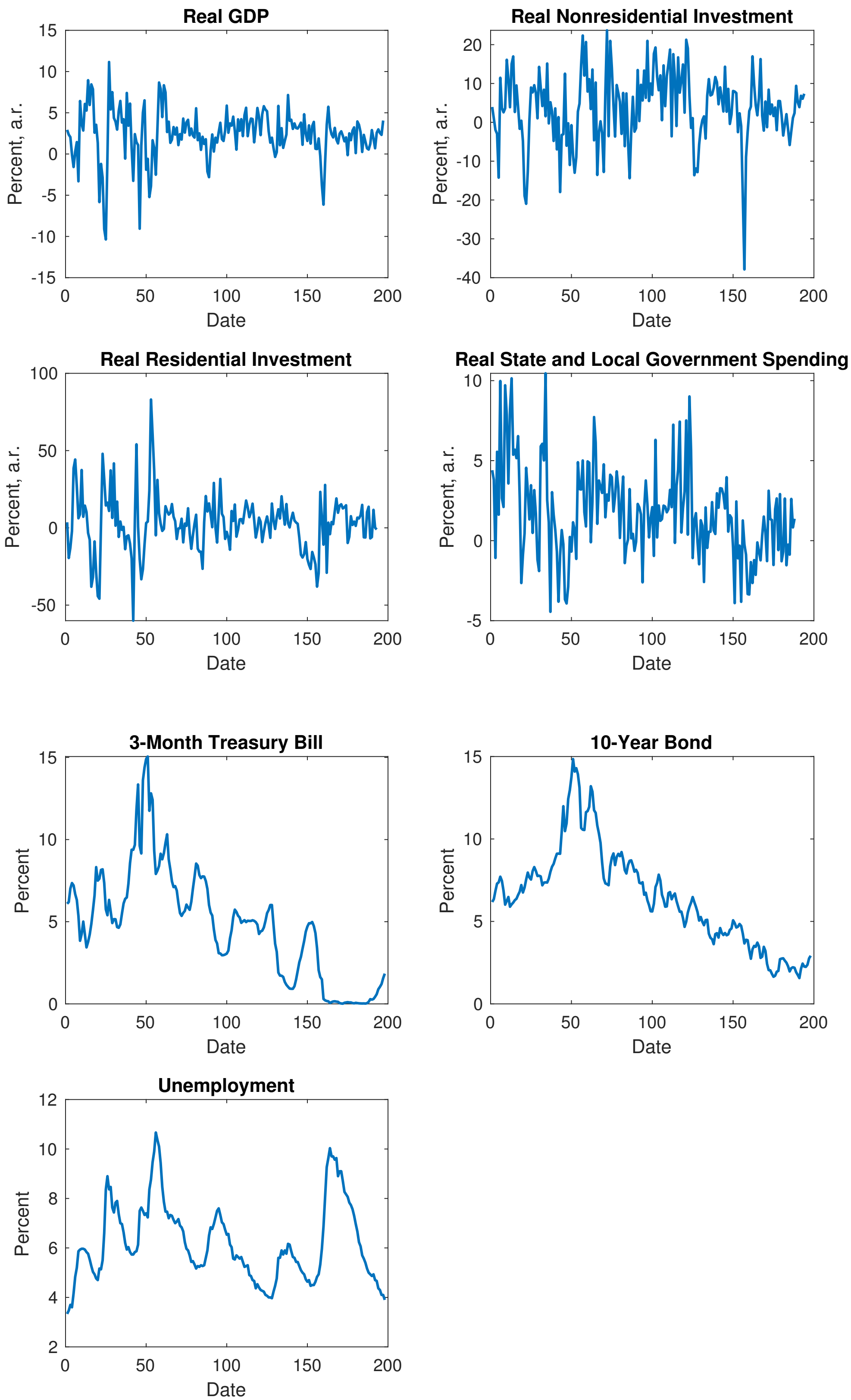
- RCONSUM – Quarterly growth forecast of real personal consumption expenditures (seasonally adjusted, annual rate).
- RFEDGOV –Quarterly growth forecast of real federal government consumption and gross investment (seasonally adjusted, annual rate).
- INDPROD – Quarterly forecasters of level of the index of industrial production, seasonally adjusted (quarterly forecasts are for quarterly average of underlying monthly levels).
- TBOND–Quarterly average 10-year Treasury bond rate (percentage points, average of the underlying daily levels). the underlying daily levels
- RSLGOV–Quarterly growth forecast of real state and local government consumption and gross investment (seasonally adjusted, annual rate).

Figure B1: Real-Time Macroeconomic Time Series



Source: Survey of Professional Forecasters

Figure B2: Real-Time Macroeconomic Time Series



Source: Survey of Professional Forecasters

Appendix C Approximation Methods: Nonlinear Filters

C.1 Details on EKF

Consider the nonlinear model defined by (11) in the main text. The problem with invoking the standard Kalman filter is that obtaining $\mathbb{E}[F(\bar{\mathbf{s}}_{\mathbf{t}}|\mathcal{Z}_t^i)]$ is intractable. The extended Kalman filter proposes linearizing the nonlinear function $F(\cdot)$ around the optimal estimate (expanding $F(\cdot)$ about $\bar{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t}}$), to obtain

$$F(\bar{\mathbf{s}}_{\mathbf{t}}) = F(\bar{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t}}) + J(\bar{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t}})(\bar{\mathbf{s}}_{\mathbf{t}} - \bar{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t}}) + h.o.t. \tag{C.1}$$

where J is the Jacobian of $F(\cdot)$ and $h.o.t.$ refers to higher order terms. The Jacobian is defined as

$$J = \begin{bmatrix} \frac{\partial F_1}{\partial \bar{\mathbf{s}}_{\mathbf{1}}} & \cdots & \frac{\partial F_1}{\partial \bar{\mathbf{s}}_{\mathbf{n}}} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial \bar{\mathbf{s}}_{\mathbf{1}}} & \cdots & \frac{\partial F_n}{\partial \bar{\mathbf{s}}_{\mathbf{n}}} \end{bmatrix} \tag{C.2}$$

Then,

$$F(\bar{\mathbf{s}}_{\mathbf{t}}) \approx F(\bar{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t}}) + J(\bar{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t}})(\mathbf{z}^{\mathbf{i}}_{\mathbf{t}} - \mathbf{z}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t-1}}) \tag{C.3}$$

therefore

$$\mathbb{E}[F(\bar{\mathbf{s}}_{\mathbf{t}}|\mathcal{Z}_t^i] \approx F(\bar{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t}}) + J(\bar{\mathbf{s}}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t}})\mathbb{E}[(\mathbf{z}^{\mathbf{i}}_{\mathbf{t}} - \mathbf{z}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t-1}}|\mathcal{Z}_t^i)] \tag{C.4}$$

Hence, the forecast is given by

$$\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t-1}} = F(\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t-1}|\mathbf{t-1}}) \tag{C.5}$$

and the update is

$$\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t}} = \widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t-1}} + \widehat{\kappa}_t(\mathbf{z}^{\mathbf{i}}_{\mathbf{t}} - \mathbf{C}\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t-1}}) \tag{C.6}$$

where $\widehat{\kappa}_t = P_{t|t-1}J^\top(\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t-1}})\big[J(\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t-1}})P^i_{t|t-1}J^\top(\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t-1}})+\Sigma_v\big]^{-1}$ and $P^i_{t|t-1} = J(\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t-1}|\mathbf{t-1}})P_{t-1|t-1}J^\top(\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t-1}|\mathbf{t-1}})+$ Σ_w and $P^i_{t|t} = (I - \widehat{\kappa}_t J(\widehat{\bar{\mathbf{s}}}^{\mathbf{i}}_{\mathbf{t}|\mathbf{t-1}}))$

C.2 Details on UKF

The UKF makes use of the unscented transformation method which is a method for calculating moments of a random variable which undergoes a nonlinear transformation. Suppose that some random variable s (of dimension L) is propagates through a nonlinear function, $y = g(s)$. Assume that s has a mean of $\bar{s}$ and a covariance P_s. The moments of y can be computed by generating a sigma matrix χ which consists of $2L+1$ sigma vectors χ_i with associated weights W_i, according to the following:

$$\begin{aligned}
\mathcal{S}_0 &= \bar{\mathbf{s}} \\
\mathcal{S}_i &= \bar{\mathbf{s}} + \left(\sqrt{(L+\lambda)\mathbf{P_s}}\right)_i \quad i = 1, \ldots, L \\
\mathcal{S}_i &= \bar{\mathbf{s}} - \left(\sqrt{(L+\lambda)\mathbf{P_s}}\right)_{i-L} \quad i = L+1, \ldots, 2L
\end{aligned}$$

and

$$\begin{aligned}
W_0^{(m)} &= \frac{\lambda}{L+\lambda} \\
W_0^{(c)} &= \frac{\lambda}{L+\lambda} + (1 - \alpha^2 + \beta) \\
W_i^{(m)} &= W_i^{(c)} = \frac{1}{2(L+\lambda)} \quad i = 1, \ldots, 2L
\end{aligned}$$

where $\lambda = \alpha^2(L+\kappa) - L$ is a scaling parameter.

The sigma points are propagated through the nonlinear function $\mathcal{Y}_i = g(\mathcal{S}_i)$ for all i. The mean and covariance for $\mathbf{y}_i$ is then approximated using the weights as follows

$$\begin{aligned}
\bar{\mathbf{y}}_i &\approx \sum_{i=0}^{2L} W_i^{(m)} \mathcal{Y}_i \\
\mathbf{P_Y} &\approx \sum_{i=0}^{2L} W_i^{(c)} \left(\mathcal{Y}_i - \overline{\mathbf{y}_i}\right) \left(\mathcal{Y}_i - \overline{\mathbf{y}_i}\right)^\top
\end{aligned}$$

The UKF makes use of this procedure by defining the augmented state vector $\begin{bmatrix} \mathbf{s_t} & \mathbf{w_t} \end{bmatrix}^\top$

C.3 Details on Particle Filtering

Rather than using generalizations to the Kalman filter, one could instead take a Monte Carlo approach. In their seminal paper, Gordon et. al. (1993) propose the bootstrap filter which is a popular variant to the particle filter. In principle, this approach makes use to mass points (particles) to approximate the underlying filtering density, $p(s_t|Z_t^i)$. This is done by defining the set of particles and associated weights: $\chi = \{s^{(n)}, \omega^{(n)}\}_{n=1}^{N}$.

Importantly, the filter still follows a general predict-update algorithm. For each particle n, the forecaster propagates the estimate through the nonlinear system

$$s_t^{i,(n)} = F\big(s_{t-1}^{i(n)}, w_t\big)$$

and then updates the weight,[1]

$$\widetilde{\omega}_t^{i,(n)} = \omega_{t-1}^{i,(n)} \cdot p\big(z_t^i | s_t^{i,(n)}\big)$$

The forecaster then normalizes the weights

$$\omega_t^{i,(n)} = \frac{\widetilde{w}_t^{i,(n)}}{\sum_{n=1}^{N} \widetilde{\omega}_t^{i,(n)}}$$

so that they sum to one. Lastly, the nowcast of the state is computed as a weighted average of the particles

$$\widehat{s}_{t|t}^{i} = \sum_{n=1}^{N} s_t^{i,(n)} \cdot \omega_t^{i,(n)}.$$

One common issue with sequential importance sampling is that the sample of particles tends to degenerate as few particles are given most of the weight. As a result, I make use of the common sequential importance resampling scheme in which I resample the particles, each with a probability equal to its weight.

Forecast errors and revisions are analogous to the formulation with the Kalman filter generalizations. The only difference is that the particle filtered estimates are not formulated by making use of the Kalman filtering equations. Nonetheless, these estimates approximate the optimal forecast.

[1] The precise manner in which the weights are updated depends on choices for the importance distribution.