
Time-Varying Trends in the Hourly Ozone Time Series for Wake county

Zekun Xu

Department of Statistics
NC State University
Raleigh, NC 27695
zxu13@ncsu.edu

Ye Liu

Department of Statistics
NC State University
Raleigh, NC 27695
yliu87@ncsu.edu

Abstract

Both ozone (O_3) and nitrogen dioxide (NO_2) are major air pollutants in the urban atmosphere, whose public health effects have raised increasing concerns. In this report, we propose the use of a dynamic linear model to analyze the trends in the hourly O_3 data for Wake county of North Carolina in 2016. Furthermore, we find out that there is a significant interaction between the O_3 series and the NO_2 series after adjusting for the cyclic daily component.

1 Introduction

Ozone (O_3) and nitrogen dioxide (NO_2) are two of the most ubiquitous oxidants found in the ambient air, whose adverse effects on public health have been well studied. As was summarized in Miles et al. (2015), the stratospheric ozone is produced and destroyed by catalytic cycles involving nitrogen, hydrogen and halogen radicals by the photolysis of molecular oxygen at shorter UV wavelength. In particular, the interaction between O_3 and NO_2 has potential effects on public health. McKee and Rodriguez (1993) pointed out that the health effects of O_3 include lung function impairment, symptoms (e.g. cough), aggravation of asthma, pulmonary inflammation, increased susceptibility to respiratory infection, and lung structure damage (e.g. centriacinar lesions). On the other hand, NOx mainly impacts on respiratory conditions causing inflammation of the airways at high levels, where a long-term exposure can decrease lung function, increase the risk of respiratory conditions and increases the response to allergens. Thus, it is of scientific importance to study the trends in the O_3 series and estimate its interaction with the NO_2 series, which is exactly the goal of this report.

Thompson et al. (2001) provided a comprehensive literature review on the variety of statistical methodologies that have been proposed to study the spatio-temporal O_3 trends. In this report, we use the univariate dynamic linear model approach to estimate the trends in the hourly O_3 series and its interaction with the NO_2 series for Wake county of North Carolina in 2016. The outline of the report is as follows. Section 2 introduces the data set and the exploratory data analysis. Section 3 sets up the modeling framework for the proposed dynamic linear model. Section 4 presents the results and interpretations.

2 Data

The 2016 hourly time series data for O_3 and NO_2 are publicly available from the website of United States Environmental Protection Agency (US EPA) at the link: https://aqs.epa.gov/aqswb/airdata/download_files.html#Raw.

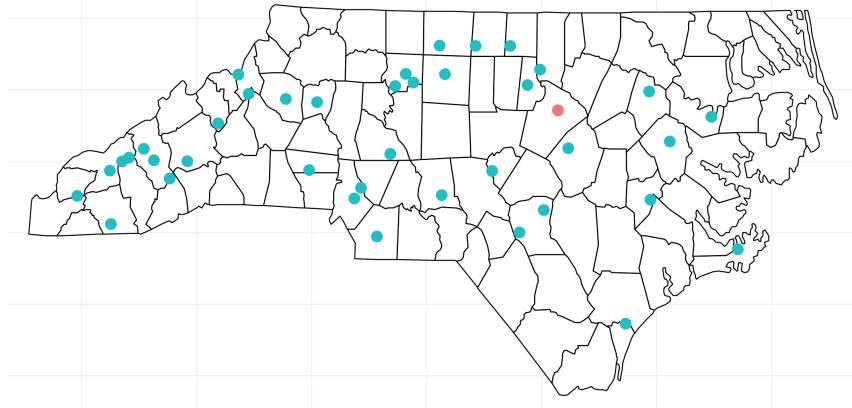


Figure 1: There are 40 monitoring sites for criteria gases in North Carolina, where the dot in red corresponds to the monitoring site in Wake county.

In Figure 1, each blue dot represent an individual monitoring site with hourly O_3 and NO_2 measurements in North Carolina. The red dot stands for the site located in Wake county, whose hourly O_3 and NO_2 data will be used in the analysis. There are 8224 non-missing observations out of the total 8784 hours of data for the site in Wake county in 2016.

Thompson et al. (2001) reported a summary of the transformations that have been used by different researchers who analyzed ozone series. In this paper, we choose to use the square root transformation on the raw data like in Carroll et al. (1997) and Reynolds et al. (1999). The reason for applying data transformation is that the normality assumption is a necessary condition for the dynamic linear model to be valid in the latter analysis.

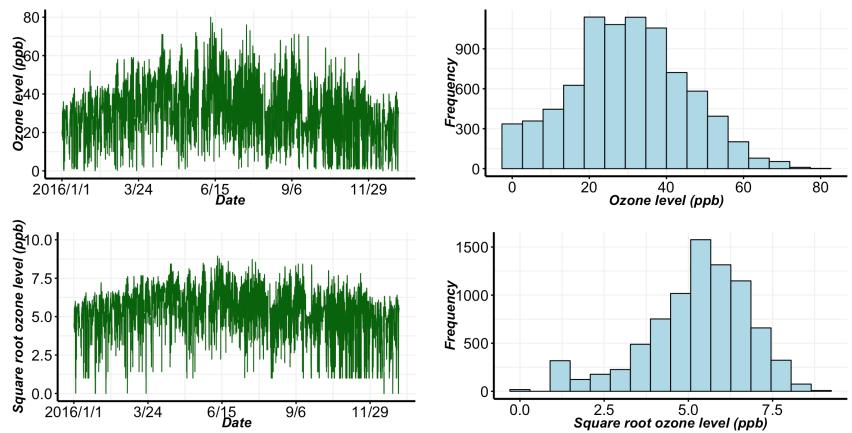


Figure 2: Hourly O_3 time series before (top) and after (bottom) the square root transformation. There are 560 missing values in the hourly O_3 for Wake county in 2016.

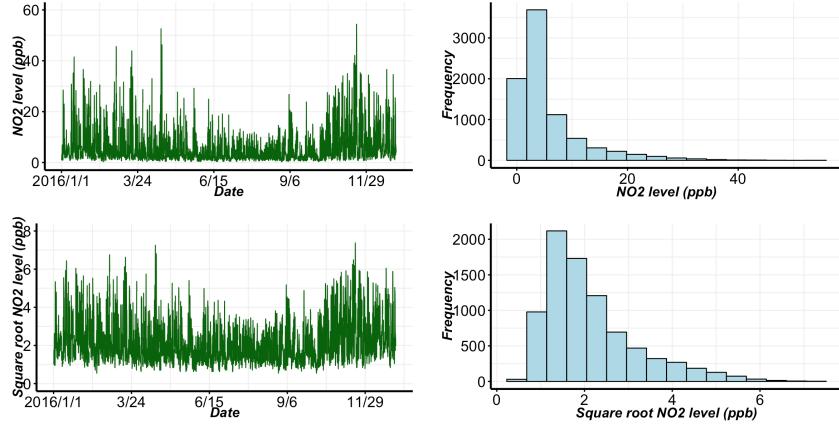


Figure 3: Hourly NO_2 time series before (top) and after (bottom) the square root transformation. There are 523 missing values in the hourly NO_2 for Wake county in 2016.

In Figure 2, the top row shows the time series and the histogram of the raw O_3 level for Wake county in 2016, while the bottom row shows the same time series and histogram after applying the square root transformation on the raw O_3 data. As we can see from the histogram, the distribution of O_3 level looks less skewed after the square root transformation. The similar story happens in Figure 3, which displays the time series and histogram of the NO_2 level for Wake county in 2016 before and after the square root transformation. Although not perfect, the square root transformation does fix the skewness in the distribution to some extent.

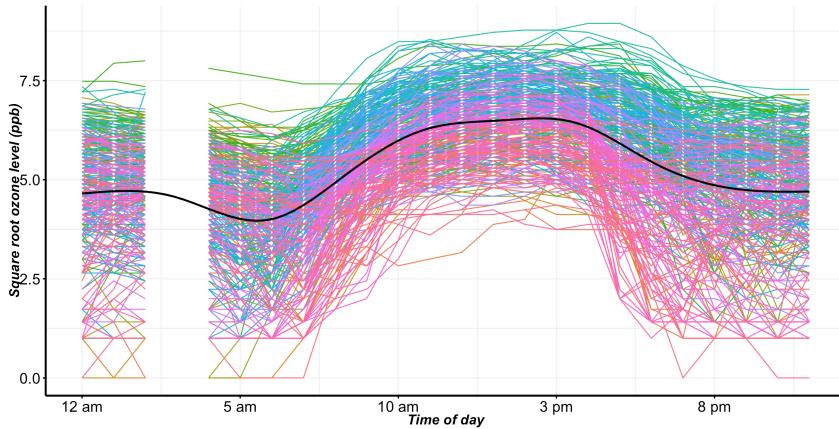


Figure 4: Each colored curve represents the hourly trajectory for the square root of O_3 measurements within a single day, where the thick black curve stands for the mean trend. The O_3 measurements are systematically missing at 3 am each day.

Figure 4 displays the hourly trajectory for the O_3 measurements each day on the square-root scale. There is a systematic missing pattern that happens at 3 am everyday. As can be seen from the plot, the O_3 level appears higher on average from around 9 am to 4 pm than during the night. This periodic daily effect is quite striking such that it is necessary to incorporate this "seasonal" component in the modeling framework.

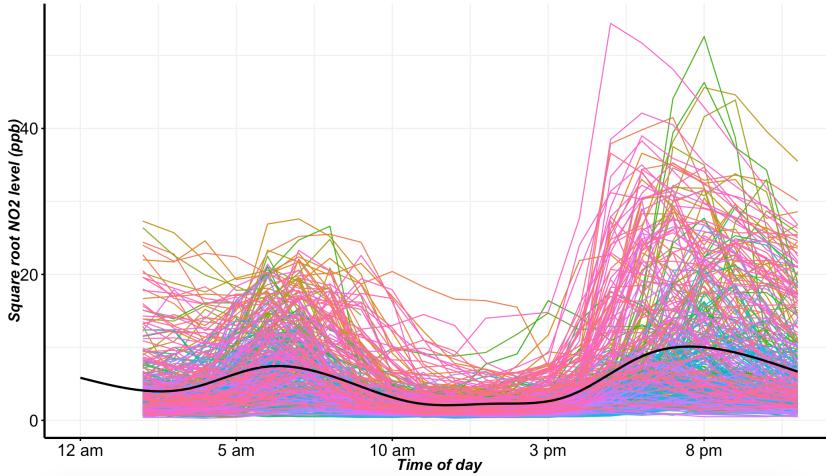


Figure 5: Each colored curve represents the hourly trajectory for the square root of NO_2 measurements within a single day, where the thick black curve stands for the mean trend. The NO_2 measurements are systematically missing at 1 am each day.

Figure 5 displays the hourly trajectory for the NO_2 measurements each day on the square-root scale. There is a systematic missing pattern that happens at 1 am everyday. As can be seen from the plot, the NO_2 level appears lower on average from around 9 am to 4 pm than during the night, which is opposite to the hourly trend for O_3 observed in Figure 4. Therefore, we would expect that there may exist a negative association between O_3 and NO_2 .

3 Methodology

The dynamic linear model (also called the linear state-space model) approach is utilized to analyze the time-varying trends in the hourly O_3 series, especially its interaction with the NO_2 series. This approach accounts for the non-stationarity in the time series. The basic structure of our modeling framework resembles Huerta et al. (2004) and Laine et al. (2014). Denote by y_t the square-root transformed O_3 level (ppb) at time t ($t = 1, \dots, T$), which is the response variable in the univariate dynamic linear model. Denote z_t to be the square-root transformed NO_2 level (ppb) at time t , which is considered an exogenous input. Let μ_t be the local mean level at time t . Further, let $\alpha_t = [\alpha_{1t}, \alpha_{2t}]$ be the cyclic component, i.e. the daily cycle with a period of 24 in the hourly O_3 series. This periodic daily effect is modeled using one harmonic via the standard Fourier representation.

To quantify the association between the O_3 and the NO_2 series, two competing dynamic linear models are considered. In the first model, the effect of NO_2 on O_3 is assumed to be constant over time as follows,

$$\begin{aligned} y_t &= \mathbf{A}_t \mathbf{x}_t + \beta z_t + v_t, & v_t &\sim N(0, R) \\ \mathbf{x}_t &= \Phi \mathbf{x}_{t-1} + \mathbf{w}_t, & w_t &\sim N(0, Q), \end{aligned} \tag{1}$$

where $\mathbf{A}_t \equiv [1, 1, 0]$, $\mathbf{x}_t = [\mu_t, \alpha_{1t}, \alpha_{2t}]$, $\Phi \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos 2\pi/24 & \sin 2\pi/24 \\ 0 & -\sin 2\pi/24 & \cos 2\pi/24 \end{bmatrix}$.

In the second model, we assume that the effect of NO_2 on O_3 varies over time as follows,

$$\begin{aligned} y_t &= \mathbf{A}_t \mathbf{x}_t + v_t, & v_t &\sim N(0, R) \\ \mathbf{x}_t &= \Phi \mathbf{x}_{t-1} + \mathbf{w}_t, & w_t &\sim N(0, Q), \end{aligned} \tag{2}$$

where $\mathbf{A}_t \equiv [1, 1, 0, z_t]$, $\mathbf{x}_t = [\mu_t, \alpha_{1t}, \alpha_{2t}, \beta_t]$, $\Phi \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\pi/24 & \sin 2\pi/24 & 0 \\ 0 & -\sin 2\pi/24 & \cos 2\pi/24 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

The model that minimizes the one-step-ahead mean-square forecast error (MSFE) will be selected as the final model, which is a similar scheme as in Baíbura et al. (2010). More specifically, we split the data into 2 parts. The first 75% of the data is used as the training set, where the performance is tested on the rest 25% of the data (validation set).

To obtain the maximum likelihood estimates for the parameters in the dynamic linear model, we choose to use the EM algorithm (Dempster et al. 1977), which naturally handles the missing values in the data. The detailed fitting procedure based on Chapter 6 in Shumway and Stoffer (2006) is illustrated in the Appendices.

4 Results

The maximum likelihood estimate for β in Model (1) is -0.3118, with a 95% confidence interval of (-0.3227, -0.2881), which indicates that there is a significant negative association between the NO_2 and O_3 series. Since the MSFE from Model (1) is 0.4112, which is greater than that of 0.2651 from Model (2), we choose Model (2) to be the final model. This means that the interaction between O_3 and NO_2 is better characterized by a time-varying trend rather than a constant effect.

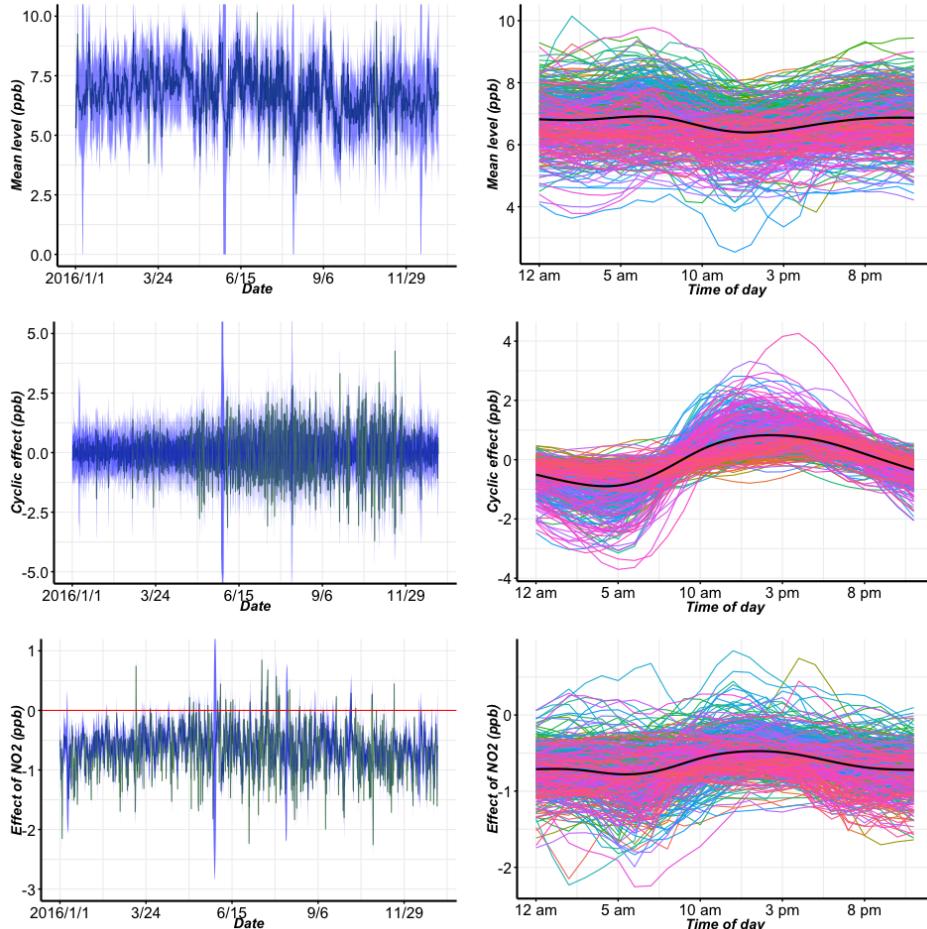


Figure 6: The three plots on the left represent the estimated mean (green) and 95% confidence interval (blue) for the O_3 level on the square-root scale (top), the daily component (middle), and the effect of NO_2 on the square-root scale (bottom). Each colored line in the three plots on the right display the estimated mean square-root O_3 level on the square-root scale (top), the daily component (middle), and the effect of NO_2 on the square-root scale (bottom) in a single day.

In Figure 6, the 2 plots in the top panel display the estimated mean O_3 level on the square-root scale, where we can see the mean series is detrended and looks like a white noise process. The 2 plots in the middle panel display the cyclic daily effect, where the effect seems to increase during the day and decrease during the night. The more interesting pattern is shown in the bottom panel, which is the time-varying effect of NO_2 on O_3 . There are 7824 (89%) of the 8784 hours in 2016 where the estimated upper bound of the 95% confidence interval is smaller than 0. Thus, there is a negative association between the O_3 level and NO_2 for most of the time for Wake county in 2016, which agrees with what we observe in the exploratory data analysis. This negative association between NO_2 and O_3 makes sense since NO_2 serves as a catalyst in the decomposition of ozone in the atmosphere. Moreover, such an interaction between the two series can be used to forecast the hourly O_3 level, which may be useful to alert the public to the potential health effects.

References

- Baíbura, M., D. Giannone, and L. Reichlin (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics* 25(1), 71–92.
- Carroll, R., R. Chen, E. George, T. Li, H. Newton, H. Schmiediche, and N. Wang (1997). Ozone exposure and population density in harris county, texas. *Journal of the American Statistical Association* 92(438), 392–404.
- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1–38.
- Huerta, G., B. Sansó, and J. R. Stroud (2004). A spatiotemporal model for mexico city ozone levels. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 53(2), 231–248.
- Laine, M., N. Latva-Pukkila, and E. Kyrölä (2014). Analysing time-varying trends in stratospheric ozone time series using the state space approach. *Atmospheric Chemistry and Physics* 14(18), 9707–9725.
- McKee, D. J. and R. M. Rodriguez (1993). Health effects associated with ozone and nitrogen dioxide exposure. *Water, Air, and Soil Pollution* 67(1-2), 11–35.
- Miles, G., R. Siddans, B. Kerridge, B. Latter, and N. Richards (2015). Tropospheric ozone and ozone profiles retrieved from gome-2 and their validation. *Atmospheric Measurement Techniques* 8(1).
- Reynolds, J., D. Caccia, P. Sampson, and P. Guttorp (1999). Meteorological adjustment of chicago, illinois, regional surface ozone observations with investigation of trends. *The National Research Center for Statistics and the Environment, The University of Washington, Technical Reports Series, Technical Report* (25).
- Shumway, R. H. and D. S. Stoffer (2006). *Time series analysis and its applications: with R examples*. Springer Science & Business Media.
- Thompson, M. L., J. Reynolds, L. H. Cox, P. Guttorp, and P. D. Sampson (2001). A review of statistical methods for the meteorological adjustment of tropospheric ozone. *Atmospheric environment* 35(3), 617–630.

Appendix I: Kalman filter and smoother for dynamic linear model

Assume that the dynamic linear model is of the following form,

$$\begin{aligned} \mathbf{y}_t &= \mathbf{A}_t \mathbf{x}_t + \mathbf{B} z_t + \mathbf{v}_t, & \mathbf{v}_t &\sim N(\mathbf{0}, \mathbf{R}) \\ \mathbf{x}_t &= \Phi \mathbf{x}_{t-1} + \Gamma \mathbf{u}_t + \mathbf{w}_t, & \mathbf{w}_t &\sim N(\mathbf{0}, \mathbf{Q}). \end{aligned} \tag{3}$$

Define the conditional expectation $\mathbf{x}_t^s = \mathbb{E}(\mathbf{x}_t | \mathbf{Y}_s)$, where $\mathbf{Y}_s = [\mathbf{y}_1, \dots, \mathbf{y}_s]$. Define the conditional covariance $\mathbf{P}_{t_1, t_2}^s = \mathbb{E}[(\mathbf{x}_{t_1} - \mathbf{x}_{t_1}^s)(\mathbf{x}_{t_2} - \mathbf{x}_{t_2}^s)^T | \mathbf{Y}_s]$. If $t_1 = t_2 = t$, we write the conditional variance as \mathbf{P}_t^s . Then for Model (3) with the initial conditions $\mathbf{x}_0^0 = \boldsymbol{\mu}_0$, $\mathbf{P}_0^0 = \boldsymbol{\Sigma}_0$, for $t = 1, \dots, n$, we have the prediction equations such that

$$\mathbf{x}_t^{t-1} = \Phi \mathbf{x}_{t-1}^{t-1} + \Gamma \mathbf{u}_t \tag{4}$$

$$\mathbf{P}_t^{t-1} = \Phi \mathbf{P}_{t-1}^{t-1} \Phi^T + \mathbf{Q}, \tag{5}$$

along with the Kalman filtering equations as

$$\mathbf{x}_t^t = \mathbf{x}_t^{t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t^{t-1} - \mathbf{B} \mathbf{z}_t) \quad (6)$$

$$\mathbf{P}_t^t = (\mathbf{I} - \mathbf{K}_t \mathbf{A}_t) \mathbf{P}_t^{t-1} \quad (7)$$

$$\text{where } \mathbf{K}_t = \mathbf{P}_t^{t-1} \mathbf{A}_t^T (\mathbf{A}_t \mathbf{P}_t^{t-1} \mathbf{A}_t^T + \mathbf{R})^{-1} \text{ (Kalman gain).} \quad (8)$$

Thus, the prediction error (innovation) and its corresponding variance are

$$\boldsymbol{\epsilon}_t = \mathbf{y}_t - \mathbb{E}(\mathbf{y}_t | \mathbf{Y}_{t-1}) = \mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t^{t-1} - \mathbf{B} \mathbf{z}_t \quad (9)$$

$$\Sigma_t = \text{var}(\boldsymbol{\epsilon}_t) = \text{var} [\mathbf{A}_t(\mathbf{x}_t - \mathbf{x}_t^{t-1}) + \mathbf{v}_t] = \mathbf{A}_t \mathbf{P}_t^{t-1} \mathbf{A}_t^T + \mathbf{R}. \quad (10)$$

Suppose we partition $\mathbf{y}_t = [\mathbf{y}_t^{(1)}, \mathbf{y}_t^{(2)}]$ where $\mathbf{y}_t^{(1)}$ is fully observed and $\mathbf{y}_t^{(2)}$ is missing. Then in the Kalman filter for updating t, we substitute

$$\mathbf{y}_t = [\mathbf{y}_t^{(1)}, \mathbf{0}], \quad \mathbf{A}_t = [\mathbf{A}_t^{(1)}, \mathbf{0}], \quad \mathbf{B} = [\mathbf{B}^{(1)}, \mathbf{0}], \quad \mathbf{R}_t = \begin{bmatrix} \mathbf{R}^{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

Correspondingly, the prediction error and variance become

$$\boldsymbol{\epsilon}_t = [\boldsymbol{\epsilon}_t^{(1)}, \mathbf{0}], \quad \Sigma_t = \begin{bmatrix} \mathbf{A}_t^{(1)} \mathbf{P}_t^{t-1} \mathbf{A}_t^{(1)T} + \mathbf{R}^{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

For $t = n, n-1, \dots, 1$ with the initial conditions \mathbf{x}_n^n and \mathbf{P}_n^n obtained from equations (4) - (8), we have the Kalman smoothing equations as follows,

$$\mathbf{x}_{t-1}^n = \mathbf{x}_{t-1}^{t-1} + \mathbf{J}_{t-1}(\mathbf{x}_t^n - \mathbf{x}_t^{t-1}) \quad (11)$$

$$\mathbf{P}_{t-1}^n = \mathbf{P}_{t-1}^{t-1} + \mathbf{J}_{t-1}(\mathbf{P}_t^n - \mathbf{P}_t^{t-1}) \mathbf{J}_{t-1}^T \quad (12)$$

$$\text{where } \mathbf{J}_{t-1} = \mathbf{P}_{t-1}^{t-1} \Phi^T (\mathbf{P}_t^{t-1})^{-1}. \quad (13)$$

In addition, with $\mathbf{K}_n, \mathbf{J}_t$ ($t = 1, \dots, n$), \mathbf{P}_n^n , and the initial condition as

$$\mathbf{P}_{n,n-1}^n = (\mathbf{I} - \mathbf{K}_n \mathbf{A}_n) \Phi \mathbf{P}_{n-1}^{n-1} \quad (14)$$

we can derive the lag-one covariance smoother for $t = n, n-1, \dots, 2$ such that

$$\mathbf{P}_{t-1,t-2}^n = \mathbf{P}_{t-1}^{t-1} \mathbf{J}_{t-2}^T + \mathbf{J}_{t-1} (\mathbf{P}_{t,t-1}^n - \Phi \mathbf{P}_{t-1}^{t-1}) \mathbf{J}_{t-1}^T \quad (15)$$

Appendix II: Maximum likelihood estimation via EM algorithm

Using equations (9) and (10), the negative log likelihood for Model (3) can be specified as

$$-\log L(\Theta | \mathbf{Y}) = \frac{1}{2} \sum_{t=1}^n \log |\Sigma_t(\Theta)| + \frac{1}{2} \sum_{t=1}^n \boldsymbol{\epsilon}_t(\Theta)^T \Sigma_t(\Theta) \boldsymbol{\epsilon}_t(\Theta) \quad (16)$$

Accordingly, for the complete data,

$$\begin{aligned} -2 \log L(\Theta | \mathbf{X}, \mathbf{Y}) &= \log |\Sigma_0| + (\mathbf{x}_0 - \boldsymbol{\mu}_0)^T \Sigma_0^{-1} (\mathbf{x}_0 - \boldsymbol{\mu}_0) \\ &\quad + n \log |\mathbf{Q}| + \sum_{t=1}^n (\mathbf{x}_t - \Phi \mathbf{x}_{t-1} - \Gamma \mathbf{u}_t)^T \mathbf{Q}^{-1} (\mathbf{x}_t - \Phi \mathbf{x}_{t-1} - \Gamma \mathbf{u}_t) \\ &\quad + n \log |\mathbf{R}| + \sum_{t=1}^n (\mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t - \mathbf{B} \mathbf{z}_t)^T \mathbf{R}^{-1} (\mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t - \mathbf{B} \mathbf{z}_t). \end{aligned} \quad (17)$$

In the j^{th} iteration, the following quantity is computed in the E-step,

$$\begin{aligned} &\mathbb{E} \left[-2 \log L(\Theta | \mathbf{X}, \mathbf{Y}) | \mathbf{Y}, \Theta^{(j-1)} \right] \\ &= \log |\Sigma_0| + \text{tr} \left\{ \Sigma_0^{-1} [\mathbf{P}_0^n + (\mathbf{x}_0^n - \boldsymbol{\mu}_0)(\mathbf{x}_0^n - \boldsymbol{\mu}_0)^T] \right\} \\ &\quad + n \log |\mathbf{Q}| + \text{tr} \left\{ \mathbf{Q}^{-1} \sum_{t=1}^n \mathbb{E} [(\mathbf{x}_t - \Phi \mathbf{x}_{t-1} - \Gamma \mathbf{u}_t)(\mathbf{x}_t - \Phi \mathbf{x}_{t-1} - \Gamma \mathbf{u}_t)^T | \mathbf{Y}] \right\} \\ &\quad + n \log |\mathbf{R}| + \text{tr} \left\{ \mathbf{R}^{-1} \sum_{t=1}^n \mathbb{E} [(\mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t - \mathbf{B} \mathbf{z}_t)(\mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t - \mathbf{B} \mathbf{z}_t)^T | \mathbf{Y}] \right\} \end{aligned} \quad (18)$$

In the M-step, we maximize the equation (18) by setting the derivatives with respect to the parameters to 0. For example, setting the derivative with respect to Φ^T to zero would be done as follows,

$$\begin{aligned}
0 &= \frac{\partial \mathbb{E} [-2 \log L(\Theta | \mathbf{X}, \mathbf{Y}) | \mathbf{Y}, \Theta^{(j-1)}]}{\partial \Phi^T} \\
&= \frac{\partial \text{tr} \{ \mathbf{Q}^{-1} \sum_{t=1}^n \mathbb{E} [(\mathbf{x}_t - \Phi \mathbf{x}_{t-1} - \Gamma \mathbf{u}_t)(\mathbf{x}_t - \Phi \mathbf{x}_{t-1} - \Gamma \mathbf{u}_t)^T | \mathbf{Y}] \}}{\partial (\mathbf{x}_t - \Phi \mathbf{x}_{t-1} - \Gamma \mathbf{u}_t)^T} \frac{\partial (\mathbf{x}_t - \Phi \mathbf{x}_{t-1} - \Gamma \mathbf{u}_t)^T}{\partial \Phi^T} \\
&= \sum_{t=1}^n \mathbb{E} [\mathbf{x}_{t-1} (\mathbf{x}_t^T - \mathbf{x}_{t-1}^T \Phi^T - \mathbf{u}_t^T \Gamma^T) \mathbf{Q}^{-1}]
\end{aligned} \tag{19}$$

Similarly, we can set the derivatives with respect to all the other parameters in the model to zero and solve the system of equations. Since \mathbf{Q} and \mathbf{R} are full-rank, then along with the quantities defined below,

$$S_{11} = \sum_{t=1}^n \mathbb{E} [\mathbf{x}_t \mathbf{x}_t^{nT} | \mathbf{Y}] = \sum_{t=1}^n (\mathbf{x}_t^n \mathbf{x}_t^{nT} + \mathbf{P}_t^n) \tag{20}$$

$$S_{10} = \sum_{t=1}^n (\mathbf{x}_t^n \mathbf{x}_{t-1}^{nT} + \mathbf{P}_{t,t-1}^n) \tag{21}$$

$$S_{00} = \sum_{t=1}^n (\mathbf{x}_{t-1}^n \mathbf{x}_{t-1}^{nT} + \mathbf{P}_{t-1}^n) \tag{22}$$

we can derive the parameter updates in the M-step as follows,

$$\boldsymbol{\mu}_0 = \mathbf{x}_0^n \tag{23}$$

$$\Sigma_0 = \mathbf{P}_0^n \tag{24}$$

$$\begin{aligned}
\Phi^T &= \left[S_{00} - \left(\sum \mathbf{u}_t \mathbf{x}_{t-1}^{nT} \right)^T \left(\sum \mathbf{u}_t \mathbf{u}_t^T \right)^{-1} \left(\sum \mathbf{u}_t \mathbf{x}_{t-1}^{nT} \right) \right]^{-1} \\
&\quad \left[S_{10}^T - \left(\sum \mathbf{u}_t \mathbf{x}_{t-1}^{nT} \right)^T \left(\sum \mathbf{u}_t \mathbf{u}_t^T \right)^{-1} \left(\sum \mathbf{u}_t \mathbf{x}_t^{nT} \right) \right]
\end{aligned} \tag{25}$$

$$\Gamma^T = \left(\sum \mathbf{u}_t \mathbf{u}_t^T \right)^{-1} \left[\left(\sum \mathbf{u}_t \mathbf{x}_t^{nT} \right) - \left(\sum \mathbf{u}_t \mathbf{x}_{t-1}^{nT} \right) \Phi^T \right] \tag{26}$$

$$\begin{aligned}
\mathbf{Q} &= S_{11} + \Phi S_{00}^T \Phi^T + \Gamma \left(\sum \mathbf{u}_t \mathbf{u}_t^T \right) \Gamma^T - S_{10} \Phi^T \\
&\quad - \left(\sum \mathbf{u}_t \mathbf{x}_t^{nT} \right)^T \Gamma^T - \Phi S_{10}^T + \Phi \left(\sum \mathbf{u}_t \mathbf{x}_{t-1}^{nT} \right)^T \Gamma^T \\
&\quad - \Gamma \left(\sum \mathbf{u}_t \mathbf{x}_t^{nT} \right) + \Gamma \left(\sum \mathbf{u}_t \mathbf{x}_{t-1}^{nT} \right) \Phi^T
\end{aligned} \tag{27}$$

$$\mathbf{B}^T = \left(\sum \mathbf{z}_t \mathbf{z}_t^T \right)^{-1} \left[\left(\sum \mathbf{z}_t \mathbf{y}_t^T \right) - \left(\sum \mathbf{z}_t \mathbf{x}_t^{nT} \mathbf{A}_t^T \right) \right] \tag{28}$$

$$\mathbf{R} = \sum \left[(\mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t^n - \mathbf{B} \mathbf{z}_t) (\mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t^n - \mathbf{B} \mathbf{z}_t)^T + \mathbf{A}_t \mathbf{P}_t^n \mathbf{A}_t^T \right] \tag{29}$$

If there are missing values as described in Appendix I, we substitute the parameters by their missing-value counterparts. Moreover, the update for \mathbf{R} becomes

$$\mathbf{R} = \sum \left\{ (\mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t^n - \mathbf{B} \mathbf{z}_t) (\mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t^n - \mathbf{B} \mathbf{z}_t)^T + \mathbf{A}_t \mathbf{P}_t^n \mathbf{A}_t^T + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \right\}$$

If a stochastic approximation EM (SAEM) algorithm is to be used, the fitting process will be divided into small batches, where the sufficient statistics for the quantities to be computed in the M-step will be recurrently updated.