Stochastic Spline Functions with Unequal Time Steps

Stephen P Smith, August 2018

Abstract. A piece-wise quadratic spline is introduced as a time series coming with unequal time steps, and where the second derivative of the spline at the junction points is impacted by random Brownian motion. A measurement error is also introduced, and this changes the spline into semi-parametric regression. This makes a total of two dispersion parameters to be estimated by a proposed REML analysis that unitizes the K-matrix. The spline itself only has three location effects that are treated as fixed, and must be estimated. A proposed prediction of a future observation beyond the spline's end point is presented, coming with a prediction error variance.

1. Introduction

This paper deals with time series and stochastic spline functions that are applicable with unequal time steps. The first order time series offers computational ease when time durations are unequal (e.g., Smith 1995). A second order, or higher order, time series can be formulated with unequal time steps (Jones and Ackerson 1990). There are different approaches for treating unequal time steps, including increasing the number of equally spaced time steps and applying interpolation as an approximation. The treatment of unequally spaced time steps might be initially conceived as complicated compared to more conventional approaches that use equal time steps. Nevertheless, it is necessary to move over to a data analysis that fully accommodates continuous time; e.g., see Erdogan et al., (2005), or variants of Kalman filtering for continuous time in Grewal and Andrews (1993).

The piece-wise continuous quadratic spline (e.g., see Behforooz 1988) is reintroduced in Section 2, but with the adaptations to include measurement error and a second derivative that's impacted by Brownian motion. This leaves the spline as a non-stationary time series. The associated K-matrix for the stochastic spline, known to be symmetric and indefinite, is described in Section 3. The estimation of dispersion parameters by restricted maximum likelihood (REML) is described in Section 4, and a Bayesian forecast is developed in Section 5. The spline turned into a semi-parametric regression is presented in Section 6.

2. Piece-wise Quadratic Spline Function

Define the observation equations, given by:

(1) $x(t)= u(t)+e_t \quad , t=t_0, t_1, ...,t_n$

Where e_t~IID $N(0,\sigma^2)$. This can be viewed as a time series, where $t=t_0, t_1, ...,t_n$, involving n+1 observations. There are n unequal durations, each denoted by $\Delta_k=t_k-t_{k-1}$, where k=1,2,...,n. A piece-wise quadratic function can also be employed to represent u(t):

(2) $$u(t) = \beta_{1k} + \beta_{2k}t + \beta_{3k}t^2, \quad if\ t \in [t_{k-1}, t_k]$$

Each of n intervals will have three coefficients to estimate from data. The function u(t) is further restricted to be continuous and first-derivative continuous, such that:

(3) $$\begin{gathered} \beta_{1k-1} + \beta_{2k-1}t_k + \beta_{3k-1}t_k^2 = u(t_k)\ . \\ \beta_{2k-1} + 2\beta_{3k-1}t_k = \beta_{2k} + 2\beta_{3k}t_k \quad , \qquad k = 2,3,\ldots,n \end{gathered}$$

The second derivatives at the connections, t_k, k=2,3, ..., n, are still permitted to change, but those changes will be restricted by a stochastic amount:

(4) $$2\beta_{3k-1} - 2\beta_{3k} = \varepsilon_k \quad , k = 2,3,\ldots,n$$

where the ϵ_k are IID $N[0, \Delta_k\sigma^2_\epsilon]$, depicting Brownian motion due to a change Δ_k in time.

In this model there are three location parameters to estimate, $\beta_{1\,1}$, $\beta_{2\,1}$ and $\beta_{3\,1}$, and two dispersion parameters, σ^2, and σ^2_ϵ, which is very feasible for modest sized data sets. A Bayesian prior can also be introduced for the three location parameters, and this will restrict the fit even more. By dropping the error term in (1), i.e., setting σ^2=0, the only random variation that is left comes from (4), which is suitable for spline fitting or situations for detecting hits on a well defined track coming with little measurement error.

To fit a time-dependent curve on a two-dimensional spatial surface, or higher, note that equations (1), (2), (3) and (4), become vector equations with dimension two, or higher. The distribution of ϵ_k is now multivariate normal, with mean vector null and variance matrix Δ_k**Q**., where **Q** is a positive definite matrix.

3. Building the K-matrix

The K-matrix described by Smith (2001) is symmetric and indefinite. Its utility rest in part on how easy its is to build by plugging in the model specifications directly as simple primitives. In particular, it is not necessary to form differences, or differences of differences, to form linear equations that are typically employed in spline calculation. Then the permuted K-matrix can be directly subjected to matrix factorization, leading to REML, estimation and prediction (Section 4), and on to Bayesian forecasting (Section

5).

The K-matrix for the model specification listed in Section 2, with σ^2=0, is presented below.

$$\mathbf{K} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{X} & \mathbf{y} \\ \mathbf{0} & \mathbf{W} & \mathbf{Z} & \mathbf{0} \\ \mathbf{X}^T & \mathbf{Z}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{y}^T & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

where: **0** represents null matrices or vectors of appropriate order; and the following assignments apply.

$$\mathbf{X}_{3n+1 \times 3n+3} = \begin{bmatrix}
1 & t_0 & t_0^2 & & & & & & & & & & & & & & & & \\
1 & t_1 & t_1^2 & & & & & & & & & & & & & & & & \\
 & 1 & 2t_1 & & -1 & -2t_1 & & & & & & & & & & & & & \\
 & & & 1 & t_1 & t_1^2 & & & & & & & & & & & & & \\
 & & & 1 & t_2 & t_2^2 & & & & & & & & & & & & & \\
 & & & & 1 & 2t_2 & & -1 & -2t_2 & & & & & & & & & & \\
 & & & & & & & & & \bullet & & & & & & & & & \\
 & & & & & & & & & & \bullet & & & & & & & & \\
 & & & & & & & & & & & \bullet & & & & & & & \\
 & & & & & & & & & & & & \bullet & & & & & & \\
 & & & & & & & & & & & & & 1 & t_{n-1} & t_{n-1}^2 & & & \\
 & & & & & & & & & & & & & 1 & t_n & t_n^2 & & & \\
 & & & & & & & & & & & & & & 1 & 2t_n & & -1 & -2t_n \\
 & & & & & & & & & & & & & & & & 1 & t_n & t_n^2
\end{bmatrix}$$

$$\mathbf{Z}_{n \times 3n+3} = \begin{bmatrix}
0 & 0 & 2 & 0 & 0 & -2 & & & & & & & & & \\
 & & & 0 & 0 & 2 & 0 & 0 & -2 & & & & & & \\
 & & & & & & \bullet & & & & & & & & \\
 & & & & & & & \bullet & & & & & & & \\
 & & & & & & & & \bullet & & & & & & \\
 & & & & & & & & & 0 & 0 & 2 & 0 & 0 & -2
\end{bmatrix}$$

$$\mathbf{y}_{3n+1\times 1}=\begin{bmatrix} x(t_0)\\ x(t_1)\\ 0\\ x(t_1)\\ x(t_2)\\ 0\\ \bullet\\ \bullet\\ \bullet\\ x(t_{n-1})\\ x(t_n)\\ 0\\ x(t_n) \end{bmatrix}$$

Lastly, **W** is a diagonal n by n matrix, with k-th diagonal $w_k=\Delta_k\sigma^2_\epsilon$.

4. REML and Estimation

Because there is the matrix **W** in **K**, it is possible to find a permutation matrix **P**, such that Ж=**PKP**^{T} can be factored as Ж=**LDL**^{T}, where **L** is lower triangular and **D** is diagonal with diagonals 1 and -1. Had **W** equaled null, this particular factorization (being a Cholesky decomposition) would unavailable, and an alternative factorization (e.g., Ashcraft, Grimes and Lewis 1998; Bunch and Parlett 1971) would be needed (for estimation only) making **L** unit lower triangular and admitting to a block diagonal structure for **D** consisting of some additional 2 by 2 blocks. The matrix Ж, as a suitable permutation of **K** that leaves the last row and column in the last position, is presented below.

$$\begin{bmatrix} \mathbf{A}_0 & \mathbf{B}_1 & & & & & & & \mathbf{y}_0 \\ \mathbf{B}_1^{T} & \mathbf{A}_1 & \mathbf{B}_2 & & & & & & \mathbf{y}_1 \\ & \mathbf{B}_2^{T} & \mathbf{A}_2 & & & & & & \mathbf{y}_2 \\ & & & \bullet & & & & & \\ & & & & \bullet & & & & \\ & & & & & \bullet & & & \\ & & & & & & \mathbf{A}_{n-1} & \mathbf{B}_n & \mathbf{y}_{n-1} \\ & & & & & & \mathbf{B}_n^{T} & \mathbf{A}_n & \mathbf{y}_n \\ \mathbf{y}_0^{T} & \mathbf{y}_1^{T} & \mathbf{y}_2^{T} & & & & \mathbf{y}_{n-1}^{T} & \mathbf{y}_n^{T} & \end{bmatrix}$$

where:

$$\mathbf{A}_k = \begin{cases} \begin{bmatrix} w_{k+1} & 2 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 2t_{k+1} & 0 & t_{k+1}^2 & 0 & t_k^2 \\ 0 & 2t_{k+1} & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & t_{k+1} & 0 & t_k \\ 0 & t_{k+1}^2 & 0 & t_{k+1} & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & t_k^2 & 0 & t_k & 0 & 1 & 0 \end{bmatrix}_{7 \times 7} & if\ k < n \\ \begin{bmatrix} 0 & t_n^2 & 0 & 0 \\ t_n^2 & 0 & t_n & 1 \\ 0 & t_n & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}_{4 \times 4} & if\ k = n \end{cases}$$

$$\mathbf{B}_k = \begin{cases} \begin{bmatrix} 0 & -2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2t_k & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{7\times 7} & \text{if } k < n \\ \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -2t_n & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}_{7\times 4} & \text{if } k = n \end{cases}$$

$$\mathbf{y}_k = \begin{cases} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ x(t_{k+1}) \\ 0 \\ x(t_k) \end{bmatrix} & \text{if } k < n \\ \begin{bmatrix} x(t_n) \\ 0 \\ 0 \\ 0 \end{bmatrix} & \text{if } k = n \end{cases}$$

Because is Ж banded, the Cholesky decomposition only requires linear computing time and storage. Moreover, the non-sparse structure in Ж is the same for each $\mathbf{B}_k$, $1\leq k\leq n$, that maps into it, and it’s the same for each $\mathbf{A}_k$, $1\leq k<n$, that maps into it, resulting in further savings. Having computed the Cholesky decomposition **L**, where $\text{Ж}_{N\times N}=\mathbf{LDL}^T$, Smith (2001b) gives the log-likelihood (Log-L) as:

$$Log-L=-\frac{1}{2}\sum_{k<N}\log(L_{kk})\quad -\frac{1}{2}L_{NN}^2$$

Derivatives are available (Smith 2001b; Smith, Nikolic and 2012) for the purpose of maximizing Log-L, thus deriving the REML estimate σ^2_ϵ. The function, Log-L, can also be maximized using derivative-free methods, and this may be preferable if stability issues are encountered.

Alternatives to REML are available for non-parametric regression involving the estimation of smoothness parameters using cross-validation or a generalization of cross-validation.

Its remarkable that this approach even works, given that a spline provides a perfect fit to the data. The chi-square statistic, represented by L_{NN}^2, will generally tend to zero for perfect fits. However, this is not expected in the current example. Note that the current example is already much different than typical problems given that data points are registered twice in the vector **y**; with the exception of $x(t_0)$.

Siegel’s (1965) equations are available for estimating, or predicting, all the coefficients that define the piece-wise continuous quadratic functions, by solving for **b** in the following equations:

$$(5)\qquad \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{X} \\ \mathbf{0} & \mathbf{W} & \mathbf{Z} \\ \mathbf{X}^T & \mathbf{Z}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

It is noted that the coefficient matrix and right-hand side of (5) are already sub-matrices of **K**. Moreover, as the last row and column of **K** are left un-permuted in Ж, the right-hand side is represented in Ж and gets over-written during the Cholesky factorization: $\text{Ж} \leftarrow \mathbf{L}=(\mathbf{D}^{-1}\mathbf{L}^{-1}\text{Ж})^T$. Therefore, solving (5) is already half-way complete. What remains is to extract a lower triangular matrix Ĺ from **L** by striking the last row, and to extract a column vector ŕ from **L** by transposing the last row of **L** and removing the last element representing L_{NN}. The solution to (5) is found by applying backward substitution to the upper triangular system, $Ĺ^T$ś= ŕ, where λ_1, λ_2 and **b** will all be located in ś as determined by **P**. The mapping of ś (and ŕ) to effects in the model is given by (6).

(6) $$\dot{\mathbf{s}} = \begin{bmatrix} \mathbf{s}_0 \\ \mathbf{s}_1 \\ \mathbf{s}_2 \\ \bullet \\ \bullet \\ \bullet \\ \mathbf{s}_{n-1} \\ \mathbf{s}_n \end{bmatrix}$$

where the following mapping holds,

$$\mathbf{s}_k \leftarrow \begin{cases} \begin{bmatrix} \lambda_{1k} \\ \beta_{3k} \\ \lambda_{2k} \\ \beta_{2k} \\ \lambda_{3k} \\ \beta_{1k} \\ \lambda_{46} \end{bmatrix} & if \;\; k < n \\ \begin{bmatrix} \beta_{3n} \\ \lambda_{5n} \\ \beta_{2n} \\ \beta_{1n} \end{bmatrix} & if \;\; k = n \end{cases}$$

5. Bayesian Forecast

By implication, flat non-informative priors were assigned to the three location parameters in the model presented in Section 2. With σ^2_ϵ estimated by REML, and plugged back in the model providing a prior distribution on second derivatives of u(t), i.e., by equation (4), what results in an empirical Bayes approach to making forecast predictions of future values of $u(t_{n+1})=x(t_{n+1})$, where $t_{n+1}>t_n$.

Define $\Delta_{n+1}=t_{n+1}-t_n$, and assume that Δ_{n+1} is not too large to invalidate the model. Even though the spine results in a perfect fit for all the data points, i.e., u(t)=x(t) where $t=t_0$, t_1, ..., t_n, it is not enough to limit consideration to the statistical errors found in estimating $\beta_{1\ n}$, $\beta_{2\ n}$ and $\beta_{3\ n}$ that comes in the best prediction available with:

$$u(t_{n+1})=\beta_{1\ n} + \beta_{2\ n}\ t_{n+1} + \beta_{3\ n}\ t_{n+1}{}^2$$

Its not enough that the adjacent quadratic functions were made to agree at the points t_1, t_2, ..., t_{n-1}. The statistical errors impacting the best prediction are important, but consideration must also be given to the forecasting errors that come directly from (4). That error is given by

$$2\ \beta_{3\ n} - 2\ \beta_{3\ n+1} = \epsilon_{n+1},$$

where $x(t_{n+1})$ is the future observation that is being predicted, and ϵ_{n+1} is distributed as $N[0,\ \Delta_k\sigma^2_\epsilon]$. Its necessary to see how this error impacts on the ideal prediction, $u(T)=\beta_{1\ n+1} + \beta_{2\ n+1}\ t_{n+1} + \beta_{3\ n+1}\ t_{n+1}{}^2$, that now involves parameters that were not estimated. That impact is derived from equations (3) and (4), first by building the following equations, and then solving them.

$$\begin{bmatrix} 1 & \Delta_{n+1} & \Delta_{n+1}^2 \\ & 1 & 2\Delta_{n+1} \\ & & 2 \end{bmatrix} \begin{bmatrix} \beta_{1n} - \beta_{1n+1} \\ \beta_{2n} - \beta_{2n+1} \\ \beta_{3n} - \beta_{3n+1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \varepsilon_{n+1} \end{bmatrix}$$

The solution is:

$$\begin{bmatrix} \beta_{1n} - \beta_{1n+1} \\ \beta_{2n} - \beta_{2n+1} \\ \beta_{3n} - \beta_{3n+1} \end{bmatrix} = \begin{bmatrix} \dfrac{\varepsilon_{n+1}\Delta_T^2}{2} \\ -\varepsilon_{n+1}\Delta_T \\ \dfrac{\varepsilon_{n+1}}{2} \end{bmatrix}$$

These are now plugged into the ideal prediction, and when this is done the following is derived:

$$\beta_{1\ n+1} + \beta_{2\ n+1}\ t_{n+1} + \beta_{3\ n+1}\ t_{n+1}{}^2 = \beta_{1\ n} + \beta_{2\ n}\ t_{n+1} + \beta_{3\ n}\ t_{n+1}{}^2 - \tfrac{1}{2}\ \epsilon_{n+1}\ t_n{}^2$$

The statistical error now comes in two parts, a prediction error related to $\beta_{1\ n}$, $\beta_{2\ n}$, and β_{3n}, and a forecasting error related to ϵ_{n+1}. The variances are now computed for each separately, and then added together.

Consider again the lower triangular matrix Ĺ that was extracted from **L** in Section 4, and consider the column vector, ŕ, that was also extracted from **L**, and where the mapping given by (6) applies. But in the new application zero out all the elements of ŕ, except for those corresponding to the effects $\beta_{1\ n}$, $\beta_{2\ n}$, and $\beta_{3\ n}$. For those three elements assign the numbers 1, t_{n+1} and t_{n+1}^{2}, respectively. Now use forward substitution to solve for the vector ś in the upper triangular system, $Ĺ^{T}$ś= ŕ. Calculate the negative weighted sum of squares,

$$\sigma_{\beta}^{2} = -\sum_{i} d_{i} s_{i}^{2}$$

where s_i is the i-th element of ś and d_i is the i-th diagonal of **PDP**^{T}; i.e., if the i-th element of ś belongs to an effect in the model by the mapping (6) then d_i =-1, otherwise d_i =1.

The prediction of the future observation at time t_{n+1}, is: $\beta_{1\ n} + \beta_{2\ n}\ t_{n+1} + \beta_{3\ n}\ t_{n+1}^{2}$. The prediction error variance is:

$$\sigma_{\beta}^{2} + \frac{\Delta_{n+1}\sigma_{\varepsilon}^{2}\ t_{n}^{4}}{4}$$

With future observation in hand, a statistical evaluation can now be made to see if the new data point fits to the end of the spline. If its too far away, the observation may be part of the background noise, and does not belong with the prior observations. Alternatively, if the new observation hits on the spline, it can be added to the data set and the prior analysis can be updated to take the new observation into account. In this sense, the stochastic spline is a tool to discriminate future observations and it becomes an adaptive spline as more observations are added to the end. The Cholesky decomposition is easily updated, particularly if the bordering algorithm is used (Smith 2017).

6. Including Measurement Error

If the time series is too erratic, allowing for a measurement error in (1) will smooth out the estimate of u(t). This will turn the spline function into a semi-parametric regression, and the variance term σ^2 can be entered into the K-matrix in various places leading to its estimation by REML without any additional modification. If the time series is found too erratic, permitting the measurement error will make more reliable the forecast of a future observation.

With σ^2 included, the K-matrix becomes the following.

$$\mathbf{K} = \begin{bmatrix} \mathbf{V} & \mathbf{0} & \mathbf{X} & \mathbf{y} \\ \mathbf{0} & \mathbf{W} & \mathbf{Z} & \mathbf{0} \\ \mathbf{X}^T & \mathbf{Z}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{y}^T & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

Where $\mathbf{V} = \begin{bmatrix} \sigma^2 & & & & & & & & \\ & \sigma^2 & \sigma^2 & & & & & & \\ & \sigma^2 & \sigma^2 & & & & & & \\ & & & \sigma^2 & \sigma^2 & & & & \\ & & & \sigma^2 & \bullet & & & & \\ & & & & & \bullet & \sigma^2 & & \\ & & & & & \sigma^2 & \sigma^2 & & \\ & & & & & & & \sigma^2 & \sigma^2 \\ & & & & & & & \sigma^2 & \sigma^2 \end{bmatrix}$

That is, **V** is block diagonal with one 1×1 block and n 3×3 blocks along the diagonal. Even with σ^2 entered, **V** is still very singular because the 3×3 blocks are only rank 1.

The same row and column permutations presented in Section 4 are preferred because they permit a well behaved calculation of Log-L even with $\sigma^2 \to 0$. There remains small changes that are needed for the matrices $\mathbf{A}_k$ and $\mathbf{B}_k$, as presented below.

$$\mathbf{A}_k = \begin{cases} \begin{bmatrix} w_{k+1} & 2 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 2t_{k+1} & 0 & t_{k+1}^2 & 0 & t_k^2 \\ 0 & 2t_{k+1} & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & t_{k+1} & 0 & t_k \\ 0 & t_{k+1}^2 & 0 & t_{k+1} & \sigma^2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & t_k^2 & 0 & t_k & 0 & 1 & \sigma^2 \end{bmatrix}_{7\times 7} & \mathit{if}\ k < n \\ \begin{bmatrix} 0 & t_n^2 & 0 & 0 \\ t_n^2 & \sigma^2 & t_n & 1 \\ 0 & t_n & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}_{4\times 4} & \mathit{if}\ k = n \end{cases}$$

$$\mathbf{B}_k = \begin{cases} \begin{bmatrix} 0 & -2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2t_k & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{7\times 7} & \mathit{if}\ k < n \\ \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -2t_n & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}_{7\times 4} & \mathit{if}\ k = n \end{cases}$$

The prediction of a future observation at time t_{n+1} requires a separate calculation, but the formula remains: $\beta_{1\,n} + \beta_{2\,n}\ t_{n+1} + \beta_{3\,n}\ t_{n+1}^{2}$. However, because the future observation comes with its own measurement error, the prediction error variance is now:

$$\sigma^2 + \sigma_\beta^2 + \frac{\Delta_{n+1}\sigma_\varepsilon^2\, t_n^4}{4}$$

Reference

Ashcraft, C., R. G. Grimes, and J. G. Lewis, 1998, Accurate Symmetric Indefinite Linear Equation Solvers, *SIAM Journal on Matrix Analysis and Application*, 20, 513-561.

Behforooz, G., 1988, Quadratic Spline, *Applied Mathematics Letters*, 1 (2), 177-180.

Bunch, J.R., and B.N. Parlett, 1971, Direct Methods for Solving Symmetric Indefinite Systems of Linear Equations, *SIAM Journal on Numerical Analysis*, 8, 639–655.

Erdogan, E., S. Ma, A. Beygelzimer, I. Rish, 2005, Statistical Models for Unequally Spaced Time Series, *Proceeding of the 2005 SIAM International Conference of Data Mining*, Editors Kargupta, Srivastava, Kamath and Goodman 626-630.

Grewal, M.S., and A.P. Andrews, 1993, *Kalman Filtering: Theory and Practice*, Prentice Hall, New Jersey.

Jones, R. H., and L.M. Ackerson, 1990, Serial Correlation in Unequally Spaced Longitudinal Data, *Biometrika*, 77, 721-731.

Siegel, I.H., 1965, Deferment of Computation in the Method of Least Squares, *Mathematics of Computation*, 19 (90): 329-331.

Smith, J.R., M. Nikolic and S.P. Smith, 2012, Hunting the Higgs Boson using the Cholesky Decomposition of an Indefinite Matrix, memo, vixRa archived.

Smith, S.P., 1995, Differentiation of the Cholesky Algorithm, *Journal of Computational and Graphical Statistics*, 4 (2), 134-147.

Smith, S.P., 2001, Likelihood-Based Analysis of Linear State-Space Models Using the Cholesky Decomposition, *Journal of Computational and Graphical Statistics*, 10 (2): 350-369.

Smith, S.P., 2017, The Backward Differentiation of the Bordering Algorithm for an Indefinite Cholesky Factorization, memo, vixRa archived in Data Structures and Algorithms.