Continuous time state space modeling with an application to high-frequency road traffic data

Siem Jan Koopman[a,b,c], Jacques J.F. Commandeur[a,d]
Frits D. Bijleveld[a,d], Sunčica Vujić[e,f]

[a] *Vrije Universiteit Amsterdam, The Netherlands*
[b] *CREATES, Aarhus University, Denmark*
[c] *Tinbergen Institute, Amsterdam, The Netherlands*
[d] *SWOV Institute for Road Safety Research, The Hague, The Netherlands*
[e] *University of Antwerp, Belgium*
[f] *University of Bath, UK*

July 29, 2017

Abstract

We review Kalman filter and related smoothing methods for the continuous time state space model. The attractive property of continuous time state space models is that time gaps between consecutive observations in a time series are allowed to vary throughout the process. We discuss some essential details of the continuous time state space methodology and review the similarities and the differences between the continuous time and discrete time approaches. An application in the modelling of road traffic data is presented in order to illustrate the relevance of continuous time state space modeling in practice.

Key words: Kalman filter, state space methods, continuous time models, dynamic models, road traffic speed data.

1   Introduction

We present a practical treatment of continuous time state space modeling. The main features of the analysis are highlighted and explored in some generality. We further present and discuss the main results of an empirical study related to road safety analysis. This application of the continuous time methodology in time series analysis shows how it can be used in practice.

A time series is a set of observations which are sequentially ordered over time. In a discrete time state space analysis the time series observations are assumed to be equally spaced in time. Although missing data may give rise to different time gaps between available

observations in discrete time series also, these time gaps are then always multiples of the time unit specific to the time series at hand (e.g., a year for annual data, a month for monthly data, etc.). In this chapter, on the other hand, we concentrate on the continuous time state space model and some of its special cases. In continuous state space models the time gaps between consecutive observations of a series are typically allowed to vary freely from one pair of consecutive observations to the next. The expositions in this chapter rely mostly on the textbooks by Harvey (1989) and by Durbin and Koopman (2012). For an introduction to state space time series analysis, we refer to Commandeur and Koopman (2007).

The literature on continuous time modelling in statistics and econometrics is extensive. It is beyond the scope of this paper to present a full review of this literature. A major and key reference to continuous time models in econometrics is the review of Bergstrom (1984) where various results on parameter estimation for dynamic structural models in continuous time are provided. The many benefits of continuous time modelling are also illustrated. In the statistics literature, there is a considerable focus on smoothing methods that are formulated in continuous time. For example, a standard treatment using the continuous time approach is developed by Green and Silverman (1994). But also the earlier contributions of Wahba (1978) and Silverman (1985) have been of key importance in the development of signal extraction and spline-smoothing in continuous time. The connections between spline-smoothing and continuous time state space analysis are first established in the work of Wecker and Ansley (1983).

In this review chapter we provide a detailed account of a continuous time state space approach to time series analysis. The outline of this chapter is as follows. In Section 2 we formulate the general continuous time state space model and discuss two well-known special cases. Section 3 discusses the estimation of the unobserved states together with the unknown model parameters. Finally, in Section 4 we apply the methodology to an empirical example consisting of road traffic speed data.

2   A continuous time modelling framework

Let t_τ denote the time-point at which observation τ in the series was measured, $\tau = 1, 2, \ldots, T$. Note that τ is an integer denoting the *number* of the observation in the time series, while t_τ is the *time* at which this observation was made. Thus, unlike τ, t_τ can be any non-negative number, for example, 10 years, 200 days, 300.405 milliseconds, etc. The only requirement is that $t_1 < t_2 < t_3 < \cdots < t_T$. The general linear Gaussian state space model for the T-dimensional observation sequence $y_1, \ldots, y_T$ is given by

$$y_\tau = Z_\tau \alpha_\tau + \varepsilon_\tau, \qquad \varepsilon_\tau \sim \text{NID}(0, H_\tau), \qquad\qquad\qquad (1)$$

$$\alpha_{\tau+1} = T_\tau \alpha_\tau + R_\tau \eta_\tau, \qquad \eta_\tau \sim \text{NID}(0, Q_\tau), \qquad \tau = 1, \ldots, T, \qquad (2)$$

where α_τ is the state vector, ε_τ and η_τ are disturbance vectors and the system matrices Z_τ, T_τ, R_τ, H_τ and Q_τ are fixed and known. A selection of the elements of the system matrices may depend on an unknown parameter vector. Equation (1) is referred as the *observation* or *measurement equation*, while equation (2) is called the *state* or *transition equation*. The $p \times 1$ observation vector y_τ contains the p observations at time-point t_τ and the $m \times 1$ state vector α_τ is unobserved. The $p \times 1$ irregular vector ε_τ has zero mean and $p \times p$ variance matrix H_τ.

The $p \times m$ matrix Z_τ links the observation vector y_τ with the unobservable state vector α_τ and may consist of regression variables. The $m \times m$ transition matrix T_τ in (2) determines the dynamic evolution of the state vector. The $r \times 1$ disturbance vector η_τ for the state vector update has zero mean and $r \times r$ variance matrix Q_τ. The observation and state disturbances ε_τ and η_τ are assumed to be serially independent and independent of each other at all time points. In many standard cases, $r = m$ and matrix R_τ is the identity matrix I_m. In other cases, matrix R_τ is a $m \times r$ selection matrix with $r < m$. Although matrix R_τ can be specified freely, it is often composed of a selection from the first r columns of the identity matrix I_m. It further implies that often we can treat the matrix R_τ as a constant matrix that does not vary with τ. Similarly, all system matrices are assumed to be (deterministically) varying with τ but in many cases of practical interest, most system matrices are fixed for all τ.

The initial state vector α_1 is assumed to be generated as

$$\alpha_1 \sim \text{NID}(a_1, P_1),$$

independently of the observation and state disturbances ε_τ and η_τ, where initial mean a_1 and initial variance P_1 can be treated as given and known in almost all stationary processes for the state vector. For nonstationary processes and regression effects in the state vector, the associated elements in the initial mean a_1 can be treated as unknown and need to be estimated. For an extensive discussion of *initialisation* in state space analysis, we refer to Durbin and Koopman (2012, Chapter 5).

2.1   Local level and local linear trend models

By appropriate choices of the vectors α_τ, ϵ_τ and η_τ, and of the matrices Z_τ, T_τ, H_τ, R_τ and Q_τ, a wide range of different continuous time state space models can be derived from (1) and (2). Here we focus on the continuous time equivalents of the discrete local level and local linear trend models. Other model formulations can be considered as well since our state space framework allows for many different linear dynamic specifications that are commonly used in time series analysis. However, the arguments for continuous time formulations are similar and therefore our treatment below remains relatively general.

Let $\delta_\tau = t_\tau - t_{\tau-1}$ denote the amount of time elapsed between two consecutive observations τ and $\tau - 1$. Also defining

$$\alpha_\tau = \mu_\tau, \quad \eta_\tau = \xi_\tau, \quad Z_\tau = T_\tau = R_\tau = 1, \quad H_\tau = \sigma_\varepsilon^2, \quad Q_\tau = \delta_\tau \sigma_\xi^2,$$

(all variables are scalars) for $\tau = 1, \ldots, T$, model (1)-(2) reduces to the univariate continuous local level model as given by

$$
\begin{aligned}
y_\tau &= \mu_\tau + \varepsilon_\tau, & \varepsilon_\tau &\sim \text{NID}(0, \sigma_\varepsilon^2), \\
\mu_{\tau+1} &= \mu_\tau + \xi_\tau, & \xi_\tau &\sim \text{NID}(0, \delta_\tau \sigma_\xi^2),
\end{aligned}
\tag{3}
$$

for $\tau = 1, \ldots, T$. Note that (3) reduces to the discrete local level model when the observations are equally spaced, i.e., when $\delta_\tau = t_\tau - t_{\tau-1} = 1$, say, for all $\tau = 1, \ldots, T$.

The local level model can be regarded as the most basic version of a state space model. It is intuitive as it can be interpreted as a model representation for y_τ that is generated by the normal distribution with a time-varying mean μ_τ and a fixed variance σ_ε^2. The continuous

time formulation only applies to the dynamic process of the time-varying mean. The local level model also provides a statistical specification for the exponentially weighted moving average (EWMA) forecasting method that is very popular amongst professional practitioners. The forecast function of the local level model is equivalent to the EWMA but the state space treatment also provides statistical standard errors to the point forecasts, see the discussion below. A full discussion and treatment of the local level model is provided by Harvey (1989) and Durbin and Koopman (2012, Chapter 2).

By defining

$$\alpha_\tau = \begin{pmatrix} \mu_\tau \\ \nu_\tau \end{pmatrix}, \quad \eta_\tau = \begin{pmatrix} \xi_\tau \\ \zeta_\tau \end{pmatrix}, \quad T_\tau = \begin{bmatrix} 1 & \delta_\tau \\ 0 & 1 \end{bmatrix}, \quad Z_\tau = \begin{pmatrix} 1 & 0 \end{pmatrix},$$

$$H_\tau = \sigma_\varepsilon^2, \quad \mathrm{Var}(\eta_\tau) = Q_\tau = \delta_\tau \begin{bmatrix} \sigma_\xi^2 + \frac{1}{3}\delta_\tau^2\sigma_\zeta^2 & \frac{1}{2}\delta_\tau\sigma_\zeta^2 \\ \frac{1}{2}\delta_\tau\sigma_\zeta^2 & \sigma_\zeta^2 \end{bmatrix}, \quad \text{and} \quad R_\tau = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

the scalar notation of (1) and (2) leads to

$$
\begin{aligned}
y_\tau &= \mu_\tau + \varepsilon_\tau, & \varepsilon_\tau &\sim \mathrm{NID}(0, \sigma_\varepsilon^2), \\
\mu_{\tau+1} &= \mu_\tau + \delta_\tau\nu_\tau + \xi_\tau, & & \\
\nu_{\tau+1} &= \nu_\tau + \zeta_\tau, & &
\end{aligned}
\tag{4}
$$

for $\tau = 1, \ldots, T$, and we obtain the univariate continuous local linear trend model. Unlike in the discrete local linear trend model, we see that the disturbances of the level and the slope component are correlated through the off-diagonal elements $\frac{1}{2}\delta_\tau^2\sigma_\zeta^2$ in matrix Q_τ in the continuous local linear trend model. However, as mentioned by Harvey (1989, p.487) "this difference is unlikely to be of any great importance".

The treatment above for the local linear trend model has many connections with the statistical literature on spline smoothing. Reviews of methods related to spline smoothing are given in Silverman (1985), Wahba (1990) and Green and Silverman (1994, Chapter 2). Some of these connections with the approach given above and the more traditional methods are given by Wahba (1990) but are also discussed in Wecker and Ansley (1983). These connections are also highlighted in Durbin and Koopman (2012, Chapter 3).

2.2 Multivariate continuous time state space models

The treatments as set out for univariate time series series above can be easily extended to multivariate time series. This is one of the advantages of the state space approach since multivariate spline smoothing methods are not widespread.

In case we let y_τ denote a $p \times 1$ vector of observations, a multivariate local linear trend model can be applied to the p time series simultaneously:

$$
\begin{aligned}
y_\tau &= \mu_\tau + \varepsilon_\tau, & \varepsilon_\tau &\sim \mathrm{NID}(0, \Sigma_\varepsilon), \\
\mu_{\tau+1} &= \mu_\tau + \eta_\tau, & \eta_\tau &\sim \mathrm{NID}(0, \Sigma_\eta),
\end{aligned}
\tag{5}
$$

for $\tau = 1, \ldots, T$, where μ_τ, ε_τ, and η_τ are $p \times 1$ vectors, Σ_ε is a $p \times p$ variance matrix, and

$$\Sigma_\eta = \delta_\tau \begin{bmatrix} \Sigma_\xi + \frac{1}{3}\delta_\tau^2\Sigma_\zeta & \frac{1}{2}\delta_\tau\Sigma_\zeta \\ \frac{1}{2}\delta_\tau\Sigma_\zeta & \Sigma_\zeta \end{bmatrix}$$

is a $2p \times 2p$ matrix, Σ_ξ and Σ_ζ being the $p \times p$ variance matrices of the level and the slope disturbances, respectively.

3 State space methods for continuous time models

The model formulations as discussed above are all special cases of the general linear Gaussian state space model. We can therefore rely on the associated methods for signal extraction, parameter estimation and forecasting. The most important and well-known method for this class of state space models is the Kalman filter that allows the (predictive and filtered) estimation of the unobserved state vector α_τ when the system matrices have given values. It also enables the computation of the log-likelihood function of the model, for a given parameter vector, via the prediction error decomposition. It allows the maximisation of the log-likelihood function with respect to the parameter vector, in order to obtain its maximum likelihood estimate. On the basis of these parameter estimates, signal extraction and forecasting can take place. We next provide more details of this central part of the state space methodology.

In a similar way as in discrete state space models, in continuous state space models for given values of all system matrices – and for known initial conditions a_1 and P_1 – the state vector can be estimated in three different ways, yielding what are known as the *filtered*, the *predicted*, and the *smoothed* state vector. Depending on the types of state estimates required in the analysis, the estimates of the state vector can be obtained by performing one or two passes through the observed time series:

1. a *forward* pass, from $t = 1, \ldots, n$, using a recursive algorithm known as the *Kalman filter* enables the computation of filtered and predicted states and prediction errors, including their variances; from the prediction error and their variances, we can compute the log-likelihood function of the given continuous state space model;

2. a *backward* pass, from $t = n, \ldots, 1$, using all filtered and associated variables from the Kalman filter and using recursive algorithms known as *state and disturbance smoothers* enables the computation of smoothed estimates of states and disturbances; it requires the storage of the Kalman filter variables.

In continuous time state space models the standard Kalman, state and disturbance smoothing filters can be used; see Durbin and Koopman (2012, Chapter 4) for technical details. A specific difference of substance between discrete and continuous time models is that the variance matrix Q_τ in equation (2) of the state space formulation of the model (containing the variances of the state disturbances) is typically time-invariant in the discrete case while it becomes a time-varying matrix for continuous time state space models.

We have discussed these continuous time state space methods above as if the disturbance variances are given and known. In practice, of course, these parameters are unknown, and they have to be estimated. Just as in the discrete time series situation, the parameter estimates are obtained via maximum likelihood methods which are discussed in Durbin and Koopman (2012, Chapter 7). It requires an optimization algorithm and for this purpose quasi-Newton methods are typically used. Each time new parameter values are proposed by the search-for-the-maximum algorithm, the Kalman filter is used to compute the log-likelihood function. In many applications it is found that the maximum is found quickly and the estimation process does not take much computing time.

4   An application in road safety and traffic control

We consider our continuous time modelling approach to a full day of measurements of the speed of passing motor vehicles on a fixed location in the right lane of a Dutch motorway, starting at midnight and ending at midnight of the following day. For our analysis of this interesting and important time series for road safety studies, we have considered the continuous time models and methods as set out in the previous sections. All computations are implemented in the `OxMetrics` object-oriented programming environment of Doornik (2013) together with the `SsfPack` library of state space routines of Koopman, Shephard, and Doornik (2008). Initial analyses are carried out by means of the discrete time versions of our models using the `STAMP` software of Koopman, Harvey, Doornik, and Shephard (2007).[1]

There is a total of $25,539$ passages in this series meaning that we also have $25,539$ observations. The time of each passage is measured as the number of milliseconds elapsed since the start of the measurements, and the difference between the time of the last and the first observation of the series, i.e., $t_T - t_1$, is $86,396$ milliseconds which indeed corresponds to a full $86,396/60^2 = 24$ hours. The average time lapse between consecutive observations in the series is 3.383 milliseconds with a minimum of 0.038 milliseconds and a maximum of 1417.4 milliseconds. The variance of the time lapses δ_τ is 230.170.

From the perspective of road safety it is of interest to analyse passages of cars at different speed levels. In our analyses we consider two groups of speed levels: slow passages with a speed of less than 100 km/hour (but faster than 75 km/hour as we discard very slow passages which may be due to measurement failings); and fast passages with a speed of higher than 120 km/hour. These two different groups constitute a total of $14,435$ passages ($9,010$ slow and $5,425$ fast passages).

The analyses of the two series are based on the continuous time local linear trend model (4). To enforce a smoother evolving signal in this highly noisy time series of speed passages, we restrict the variance of the level component to be zero. The remaining variances are estimated by the method of maximum likelihood (ML). We yield the following estimation results. At convergence of the ML process, the parameter estimates for the variance of the slope disturbances and the measurement errors are respectively given by $\sigma_\zeta^2 = 1.061 \times 10^{-6}$ and $\sigma_\varepsilon^2 = 31.966$ for the slow passages and $\sigma_\zeta^2 = 5.434 \times 10^{-6}$ and $\sigma_\varepsilon^2 = 43.295$ for the fast passages.

The recorded speed levels of the passages for the slow and the fast groups are presented in Figure 1 and Figure 3, respectively, together with their estimated trend components which are also presented separately in Figure 2 and Figure 4, respectively. We learn from these graphs that the number of passages of motor vehicles on the motorway diminishes during the night. It is especially observable for the fast speed passages, between roughly $8,000$ milliseconds after midnight (i.e., around half past three in the morning) and $20,000$ milliseconds after midnight (i.e., around six o'clock in the morning), that the number of passages is clearly much smaller. At the same time we can conclude that the speed of the fast group increases somewhat as it is quiet on the motorway during these night hours. In contrast, somewhat later in the night and up to the early morning hours, the speed of the slow group diminishes to clearly lower speed levels which is possibly due to a relatively intensified presence of more heavy trucks that generally drive slower and on the right lane of the road.

---

[1]The programming code of our analyses is available upon request to the authors.

We have shown in our analysis that the unequal time lapses between consecutive vehicles can be handled effectively using our continuous time trend model. In this particular application that is highly relevant for road safety studies, there are several directions in which the results of our analysis can be improved. Diagnostic tests on the one-step ahead prediction errors indicate that neither the assumption of independence nor the assumption of normality of the residuals is satisfied: the Box-Ljung statistic for independence has values for $Q(10)$ that are too high; their values should be smaller than 16.95 in order to be non-significant at the usual 5% level. Also, the values for the Bowman-Shenton test for normality are too high; their values should be smaller than 5.99 to be non-significant.

Although our reported initial findings are highly interesting, the continuous time trend model appears to be somewhat away from a correct model specification for the analysis of traffic speed data. Further research can be conducted in order to obtain a more satisfactory model that is capable of capturing the remaining autocorrelation and non-normality of the data. However, this research falls outside the scope of our current review on continuous time state space modeling.

5 Conclusions

We have discussed the basic principles of a model-based continuous time approach using the state space methodology. The methodology is especially designed for the analysis of irregularly spaced data. We have highlighted the potential of this approach in an illustration of high-frequency intra-daily time series of speed measures from vehicles that pass a certain point at a motorway.

References

Bergstrom, A. R. (1984). Gaussian estimation of structural parameters in higher order continuous time dynamic models. In Z. Griliches and M. Intriligator (Eds.), *The Handbook of Econometrics, Volume 2*, pp. 1145–1212. North-Holland.

Commandeur, J. and S. J. Koopman (2007). *An Introduction to State Space Time Series Analysis*. Oxford: Oxford University Press.

Doornik, J. A. (2013). *Object-oriented matrix programming using Ox 7.00*. London: Timberlake Consultants Press.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods* (2nd ed.). Oxford: Oxford University Press.

Green, P. and B. W. Silverman (1994). *Nonparameteric Regression and Generalized Linear Models: A Roughness Penalty Approach*. London: Chapman & Hall.

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

Koopman, S. J., A. C. Harvey, J. A. Doornik, and N. Shephard (2007). *Stamp 8.0: Structural Time Series Analyser, Modeller and Predictor*. London: Timberlake Consultants.

Koopman, S. J., N. Shephard, and J. A. Doornik (2008). *Statistical Algorithms for Models in State Space Form: SsfPack 3.0*. London: Timberlake Consultants.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society, Series B 47*, 1–52.

Wahba, G. (1978). Improper priors, spline smoothing, and the problems of guarding against model errors in regression. *Journal of the Royal Statistical Society, Series B 40*, 364–372.

Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia: SIAM.

Wecker, W. E. and C. F. Ansley (1983). The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association 78*, 81–89.

Figure 1: **Slow speed passages:** speed measures (in km/hour) of slow passages of motor vehicles during a full day on a fixed location in the right lane of a Dutch motorway (in tiny dots) together with the smoothed estimated trend component (solid line) from the continuous time local linear trend model. The horizontal x-axis represents the time index measured in seconds of a full day starting at midnight 0:00 hours.
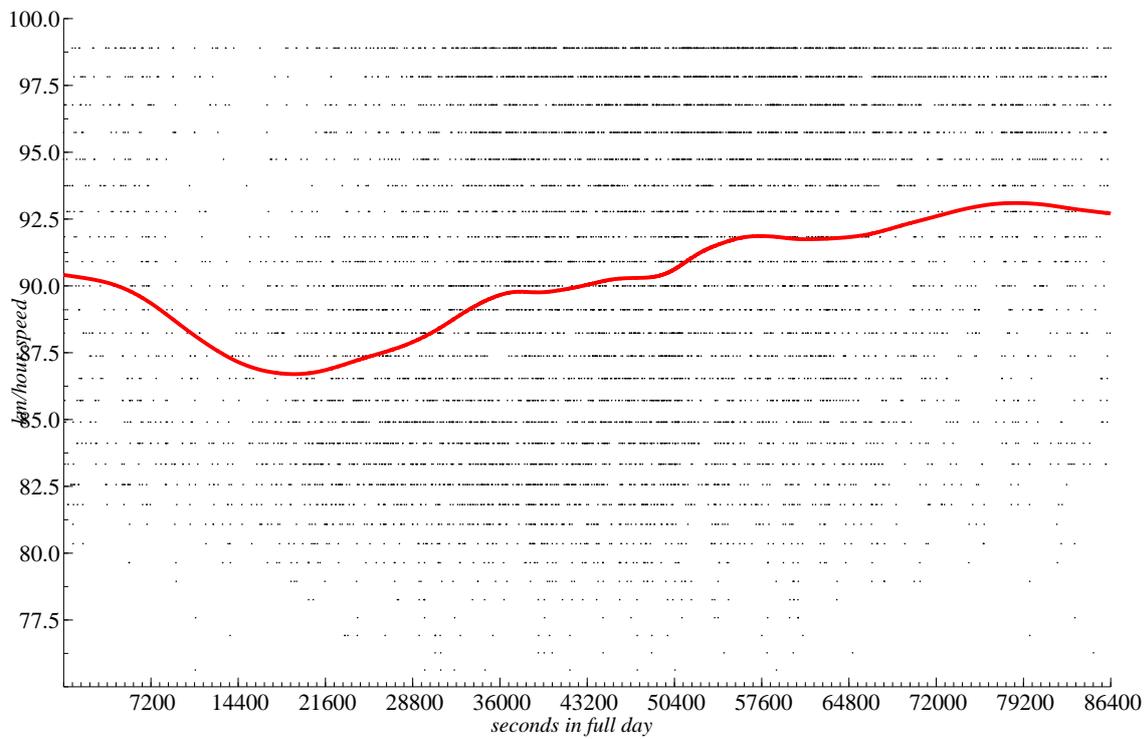
Figure 2: **Slow speed trend:** the smoothed estimated trend component from the continuous time local linear trend model. The horizontal x-axis represents the time index measured in seconds of a full day starting at midnight 0:00 hours.
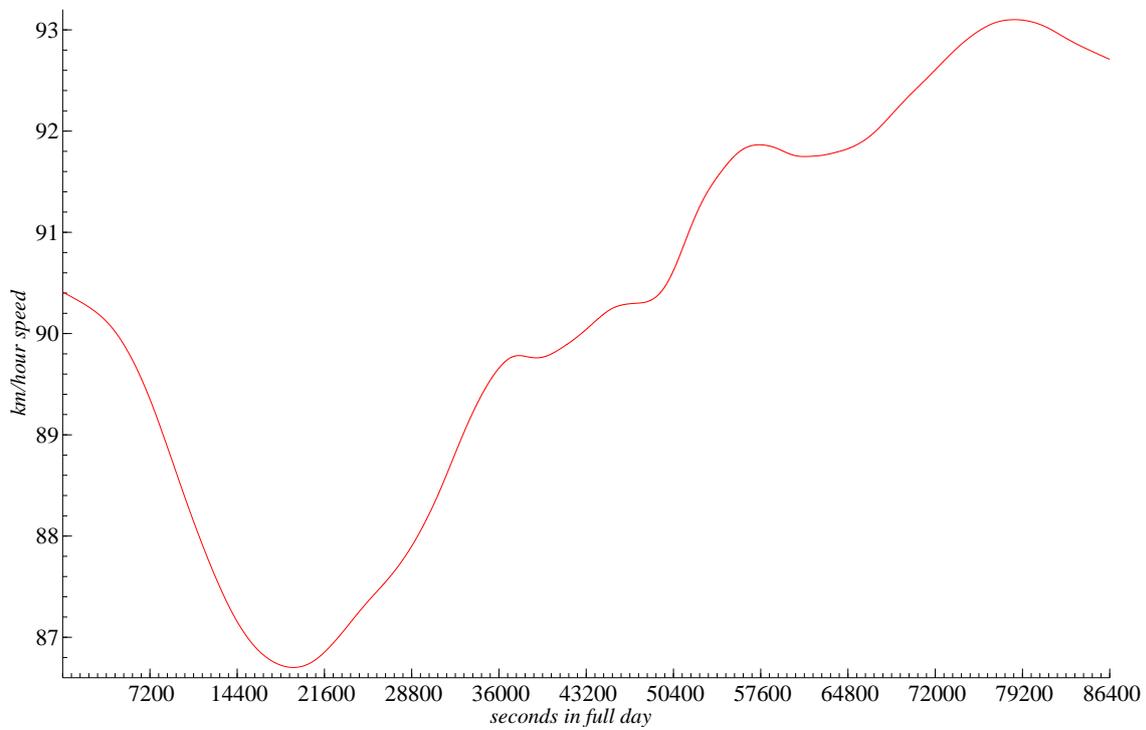
Figure 3: **Fast speed passages:** speed measures (in km/hour) of fast passages of motor vehicles during a full day on a fixed location in the right lane of a Dutch motorway (in tiny dots) together with the smoothed estimated trend component (solid line) from the continuous time local linear trend model. The horizontal x-axis represents the time index measured in seconds of a full day starting at midnight 0:00 hours.
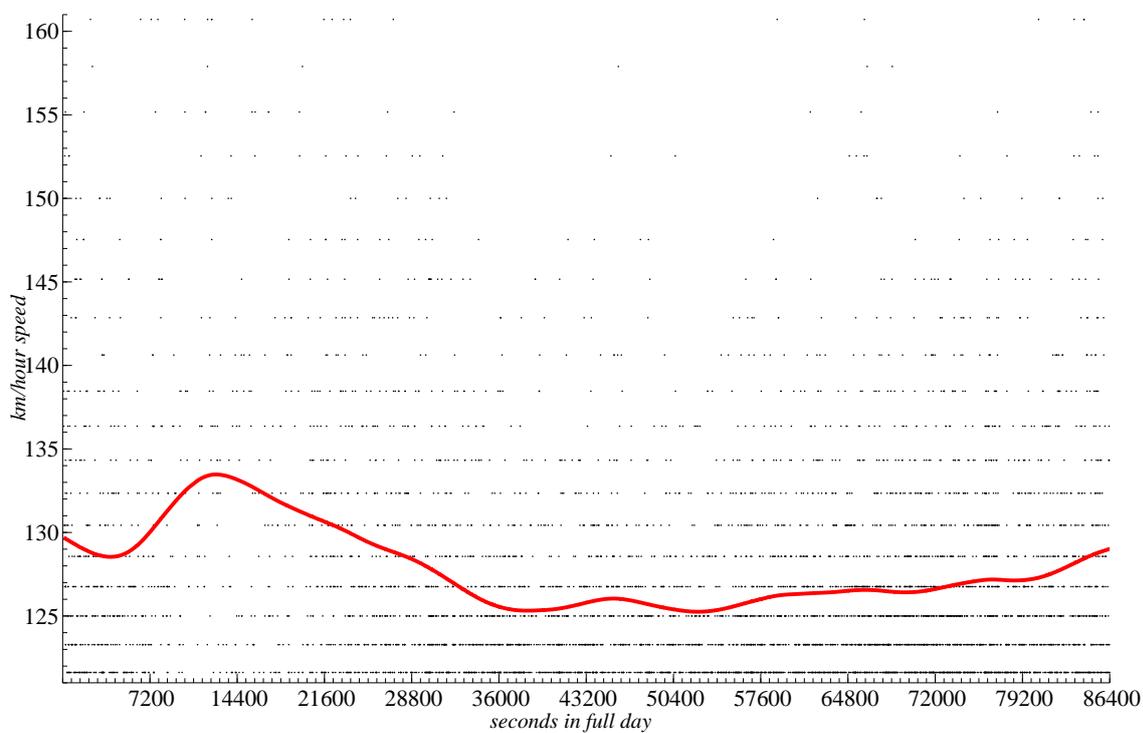
Figure 4: **Fast speed trend:** the smoothed estimated trend component from the continuous time local linear trend model. The horizontal x-axis represents the time index measured in seconds of a full day starting at midnight 0:00 hours.