Trend analysis using state-space models

Sonia Mazzi *Data Science Campus - Office for National Statistics - UK*

Trend analysis, the analysis of past and present information on the evolution of a process in order to make predictions about its future states, is crucial in areas such as finance and business. This paper proposes a model, a local-linear state-space model (LLM-SSM), for trend analysis based on extracting the relevant quantities to indicate level and rate of change from past data. The theory and technical details underpinning signal extraction and forecasting using the LLM-SSM are presented as well as illustrations with real-life data on the analysis of emerging technologies.

Keywords: Local Linear Model, State-Space Model, Kalman Filter, Diffuse Kalman Filter, smoothing filter

1 Introduction

Given data on a variable of interest collected at sequential time points, or time series data, trend analysis attempts to predict future levels of the variable using historical ones. Of importance in trend analysis is the detection of upwards and downwards trends as well as trend reversals. For example, in finance the general direction the market is taking during a specified period of time is of importance, so much that markets are named according to current prevailing trend type: if the trend is upward, it is a "bull market" and investors will try to profit until a trend reversal occurs, a "bull-to bear-market", and avoid losses when the trend is downwards, a "bear market".

An important area where trend analysis is also crucial is emergence studies. Emergence refers to a certain event happening with increased frequency, from undetectable to noticeable growth, and further evolutions into declining or static states as time evolves. For example, the analysis of technological emergence offers valuable insights to those determining research and development priorities, are in charge of portfolio management, or carrying out technology opportunity analysis and managing innovation. See, for example, Burmaogluab, Sartenaer and Porter (2019). Likewise there are many other areas where emergence studies play a crucial role: epidemiology, social studies, economics, etc.

To aid trend analysis, trend extraction from the data is useful, however not enough. Extracting the trend of data may provide a visual indication of upwards/downwards periods but, in terms of automatic trend tendency and reversal detection, trend extraction alone is not enough.

Let us idealise a trend as a time-evolving continuous smooth path in two dimensions, one being the variable of interest and the other one time. At a given time point we consider the level in the variable of interest and the instantaneous rate of change, or first derivative, of the level. The first derivative is precious in trend analysis as from its sign it can be determined whether the trend is increasing (first derivative is positive) or decreasing (first derivative is negative). The trend of the first derivative may announce a trend-reversal: a positive first derivative with a decreasing trend indicates growth which is diminishing as time evolves, a negative derivative with a positive trend indicates that, although the trend of the variable of interest is downwards, the decrements are diminishing.

In trend analysis imposing a global deterministic trend to observational data will many times fail to describe the underlying trend behaviour. In emergence studies, at the beginning of the considered time period, the trend could be lying near zero to start growing at different paces and then either stay constant at a level or begin to decline. A purely stochastic model, such as an ARIMA model, although somewhat more flexible, is useful for obtaining estimates of the level and for predicting the future level, provided the model fits the data well. Although locally a deterministic model could hold, due to different factors the local behaviours change as time evolves. Also, to be taken into account, is the fact that data available for trend analysis is most of the times observational and once data is collected outside a fully controlled environment it is subject to many imponderable influences, mostly not of main interest but which need to be taken into account at the time of analysis.

In this paper we propose a model for trend analysis and prediction which is stochastic in nature, which allows for the imponderables to be taken into account, and which aims to model the trend behaviour and its instantaneous rate of change locally requiring only that the changes in local behaviour are gradual. The model is robust in that it doesn't follow non-permanent local changes. Also, the paradigm for trend analysis described here is an all-in-one package: it is useful for trend and instantaneous rate of change extraction, model validation and prediction.

In Section 2 we introduce the trend analysis paradigm, the local linear trend model (LLM-SSM). Section 3 casts the LLM-SSM into state-space form. This feature of the LLM-SSM allows to use all the existing algorithms for signal extraction, validation and prediction in the context of a state-space model. Section 4 states the diffuse Kalman filter (DKF) for the LLM-SSM. The DKF provides all the necessary quantities for prediction, Section 5, and smoothing, Section 6.

2 The local linear trend state space model

Suppose that we have $y = \{y_i\}_{i=1}^{n}$ and $y_i = \mu_i + \epsilon_i$, where μ_i is a stochastic process and ϵ_i's are independent and identically distributed with mean zero and constant variance. This is a signal-plus-noise model. $\mu = \{\mu_i\}$ is the signal, which is unobserved and the feature that we wish to extract from $y = \{y_i\}$, the variable of interest. The process $\epsilon = \{\epsilon_i\}$ is the noise, which has no significant stochastic structure, such as autocorrelation, and has expected value zero. We assume that $\epsilon_1, \ldots, \epsilon_n$ are independent and identically distributed (i.i.d.) random variables with mean 0 and variance σ^2.

The stochastic specification of μ_i allows to model y in different ways. We will use a particular stochastic model for μ, suitable for trend analysis. This model will then be cast into state-space form which will allow us to use all the algorithms for prediction and smoothing.

For general details on state-space models and signal-extraction algorithms see de Jong (1991).

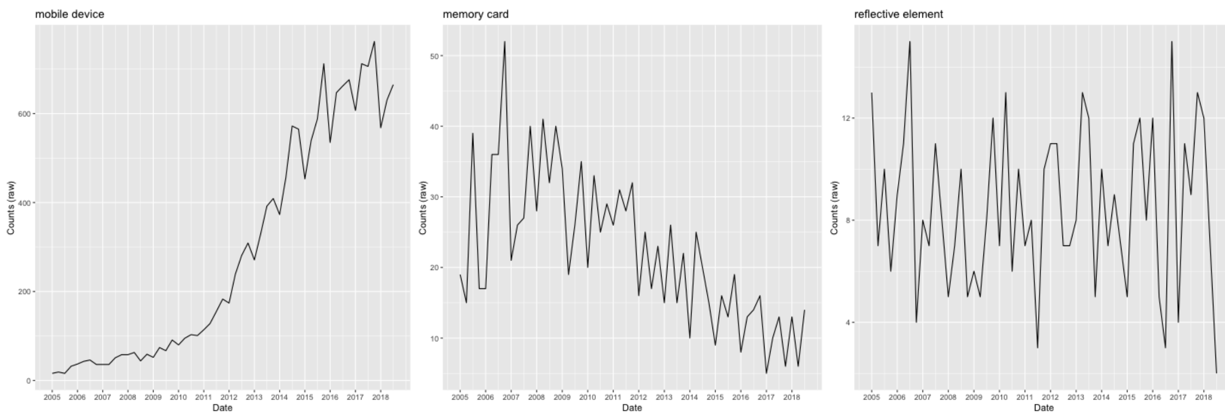
We say that y is pure-trend data if, besides from white noise, the only source of variation of y is trend. Note that we do not wish to explain the data generating process and we do not wish to explain what makes y vary. We wish to extract a smooth signal indicating the local trend and predict its behaviour in terms of trend.

EXAMPLE. *The emerging technologies time series data*.

Indicators of technological emergence can address varied subjects: breakthrough science, novel technology, commercial innovation, etc.

One such indicator can be based on extracting "relevant" terms from patent applications, and assessing the counts of the occurrence of such terms at different time points. The relevance of a term is quantified by using natural language processing criteria, namely TF-IDF, which is a way to weigh the frequency with which a term appears in a set of documents. For a given period of time, this process creates time series of counts for each relevant term.

The time plots below are for quarterly counts of nine terms between 2005 and the third quarter of 2018 extracted from an open database of US patent applications, https://bulkdata.uspto.gov/. The plots below aim to illustrate that the shape of the general trend of each series can be quite varied. In what follows we will use the data to further illustrate the quantification of emergence.



In the LLM-SSM it is assumed that $y_i = \mu_i + \epsilon_i$, $\mu_{i+1} = \mu_i + d_i + \nu_i$, and $d_{i+1} = \delta d_i + \eta_i$, where $\{\epsilon_i\}$, $\{\nu_i\}$, $\{\eta_i\}$ are mutually independent white noise processes with variances σ^2, σ_ν^2 and σ_η^2.

The rationale for the model stems from initially assuming that the signal or trend, μ_i, follows a fully deterministic specification: $\mu_i = a + d\,i$, with a and d constant. Whilst this specification would be clearly unsatisfactory for most time series, if we assume that $\mu = \{\mu_i\}$ is "smooth", μ could be described locally, rather than globally, by a linear model (Taylor's theorem). The idea is to let a and d vary with time to allow μ_i to adapt to the evolution of the series.

If $\mu_i = a + d\,i$ then $\mu_{i+1} = \mu_i + d$.

To allow for flexibility in the model we could establish that

$$\begin{aligned} \mu_{i+1} &= \mu_i + d_i + \nu_i, \\ d_{i+1} &= \delta d_i + \eta_i. \end{aligned}$$

where $\nu = \{\nu_i\}$ and $\eta = \{\eta_i\}$ are independent white noise processes with mean zero and variance σ_ν^2 and σ_η^2 respectively. ϵ is independent of ν and η.

The full model is

$$\begin{aligned} y_i &= \mu_i + \epsilon_i, \\ \mu_{i+1} &= \mu_i + d_i + \nu_i, \\ d_{i+1} &= \delta d_i + \eta_i. \end{aligned} \tag{1}$$

3 State-space form of the LLM-SSM

To cast the model into state-space form we write

$$\begin{aligned} y_i &= Z\alpha_i + Gu_i, \\ \alpha_{i+1} &= T\alpha_i + Hu_i. \end{aligned}$$

where $u_0, u_1, \ldots, u_n$ are i.i.d with $Var(u_i) = \sigma^2 I_3$, where I_n denotes the $n \times n$ identity matrix, and

$$\alpha_i = \begin{pmatrix} \mu_i \\ d_i \end{pmatrix}, \; Z = (1\ 0), \; T = \begin{pmatrix} 1 & 1 \\ 0 & \delta \end{pmatrix}, \; u_i = \begin{pmatrix} \epsilon_i \\ \eta_i \\ \nu_i \end{pmatrix}, \; G = (1\ 0\ 0), \; H = \begin{pmatrix} 0 & \sigma_\nu/\sigma & 0 \\ 0 & 0 & \sigma_\eta/\sigma \end{pmatrix},$$

with $\alpha_1 = W_0\beta$, $\beta = b + B\gamma$, $\gamma \sim (c, \sigma^2 C)$ and in this case $W_0 = B = I_2$, $b = 0_2$, where 0_n denotes a vector of zeroes of length n.

If we think of μ_i as $\mu_i = a_i + d_i t_i$, α_1 represents $(a_1 + d_1 t_1 \quad d_1)^t$. v^t means the transpose of v.

The Diffuse Kalman Filter (DKF) is a generalisation of the Kalman Filter which allows to deal with initial conditions in the same paradigm that estimation, signal extraction and prediction take place. See de Jong (1991).

4 The Diffuse Kalman Filter

The DKF in this case is the recursion

$$\left.\begin{aligned} E_i &= (0, y_i) - ZA_i, \\ D_i &= ZP_iZ^t + GG^t, \\ K_i &= TP_iZ^tD_i^{-1}, \\ A_{i+1} &= TA_i + K_iE_i, \\ P_{i+1} &= (T - K_iZ)P_iT^t + HH^t, \\ Q_{i+1} &= Q_i + E_i^tD_i^{-1}E_i, \end{aligned}\right. \qquad i = 1, \ldots, n,$$

with starting conditions $A_1 = (-I_2, 0_2)$, $P_1 = 0$, $Q_1 = 0$.

After the DKF is run, obtain Q_{n+1} and partition it

$$Q_{n+1} = \begin{pmatrix} S & s \\ s^t & q \end{pmatrix},$$

so that if Q_{n+1} has c columns and r rows, S is a $(q-1) \times (c-1)$ matrix.

The MLE of α_1 is $S^{-1}s$, with covariance matrix $\sigma^2 S^{-1}$, the MLE of σ^2 is $(q - s^tS^{-1}s)/n$ and the log-likelihood (maximised with respect to σ^2 and α_1) is $-\frac{1}{2}\left[n\log(\hat{\sigma}^2) + \sum_{i=1}^n \log(|D_i|)\right]$.

Hyperparameter estimation

The parameters $\delta, \sigma_\epsilon, \sigma_\eta, \sigma_\nu$ need to be estimated. We can do this via maximum likelihood. Note that the DKF provides an estimate of σ_ϵ. In the DKF the unknown parameters, to be estimated, appear as $\sigma_\nu^* = \sigma_\nu/\sigma_\epsilon$, $\sigma_\eta^* = \sigma_\eta/\sigma_\epsilon$ and δ. σ_ν^* and σ_η^* are "signal-to-noise ratio" parameters. Since we are only interested in smooth signals, we restrict the parameter space to search for σ_ν^*, and also for σ_η^*, to be the interval $[0, 0.5]$. For the same reason, the search for a smooth signal, the parameter δ is restricted to the interval $[0.85, 1]$.

5 Prediction/forecasting using a state-space model

We know that

$$\begin{aligned} y_i &= Z\alpha_i + Gu_i, \\ \alpha_{i+1} &= T\alpha_i + Hu_i. \end{aligned}$$

Then $y_{n+k+1} = Z\alpha_{n+k+1} + Gu_{n+k+1}$, $k \geq 0$. If $y = \{y_1, \ldots, y_n\}$, the predictor of y_{n+k+1} given y is

$$E(y_{n+k+1}|y) = ZE(\alpha_{n+k+1}|y),$$

since $E(u_{n+k+1}|y) = 0$ if $k \geq 0$.

Now, $\alpha_{n+k+1} = T\alpha_{n+k} + Hu_{n+k} = \cdots = T^k\alpha_{n+1} + H\sum_{j=1}^{k} T^{j-1}u_{n+k+1-j}$. This can be easily shown by induction.

Therefore, $E(\alpha_{n+k+1}|y) = T^kE(\alpha_{n+1}|y)$ and so the predictor of y_{n+k+1} is

$$E(y_{n+k+1}|y) = ZE(\alpha_{n+k+1}|y) = ZT^kE(\alpha_{n+1}|y).$$

In order to obtain any prediction of the series we just need the one-step ahead predictor of the state vector.

The DKF gives us all what is needed to compute the one step-ahead predictor of the state vector, $E(\alpha_{n+1}|y)$.

$$\hat{\alpha}_{n+1} = E(\alpha_{n+1}|y) = A_{n+1}(-S^{-1}s;\, 1),$$
$$mse(\hat{\alpha}_{n+1}) = \sigma^2[P_{n+1} + MS^{-1}M^t],$$

where M denotes all but the last column of A_{n+1}.

So, the predictor of y_{n+k+1} given y and its mean squared error are

$$\hat{y}_{n+k+1} = ZT^k\hat{\alpha}_{n+1},$$
$$mse(\hat{y}_{n+k+1}) = ZT^k\hat{\alpha}_{n+1}(T^k)^tZ^t.$$

6 Signal extraction

In the context of a pure trend model we wish to extract the smooth signal. That is we would like to predict $E(\mu_t|y)$, $t = 1, \ldots, n$. Since $\mu_t = Z\alpha_t$, $E(\mu_t|y) = ZE(\alpha_t|y)$, it suffices to predict $E(\alpha_t|y)$.

The DKF gives us the predictions of $E(\alpha_{t+1}|y_1, \ldots, y_t)$. We use the smoothing filter, a backwards recursion, to obtain the smoothed valued of the series.

The Smoothing Filter (SF) is the recursion

$$N_{i-1} = Z^tD_i^{-1}E_i + L_i^tN_i,$$
$$R_{i-1} = Z^tD_i^{-1}Z + L_i^tR_iL_i,$$

with $N_n = 0$, $R_n = 0$, $L_i = T - K_iZ$, $i = 1, \ldots, n$, and all other quantities as in the DKF.

Then,

$$\tilde{\alpha}_i = E(\alpha_i|y) = (A_i + P_iN_{i-1})(-S^{-1}s;\, 1)$$
$$mse(\tilde{\alpha}_i) = \sigma^2\left(P_i - P_iR_{i-1}P_i + N_{i-1,\gamma}S^{-1}N_{i-1,\gamma}^t\right),$$

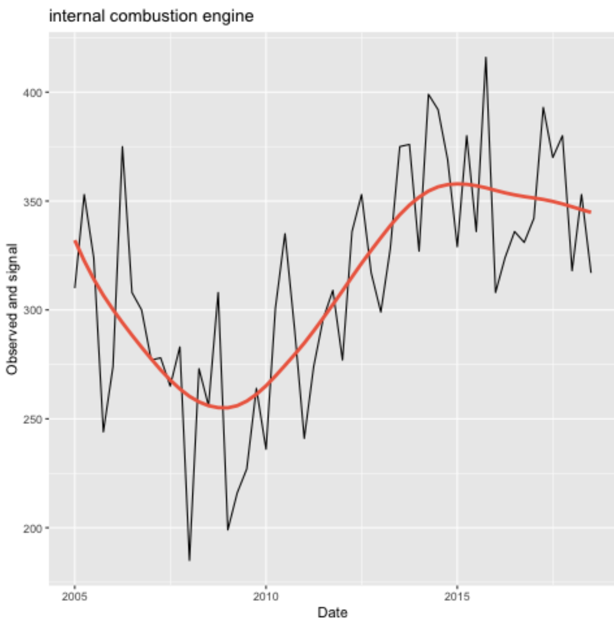
where $N_{i-1,\gamma}$ denotes all but the last column of $A_i + P_iN_{i-1}$.

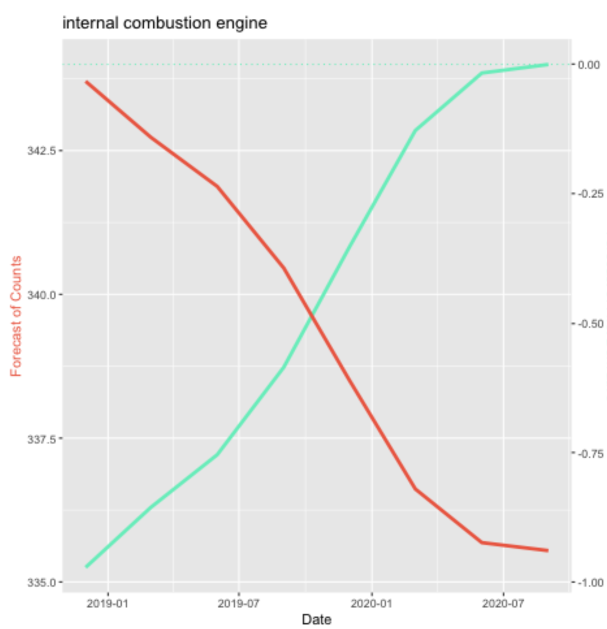
The state vector is $\alpha_i = (\mu_i \;\; d_i)^t$. The second entry of α_i can be interpreted as we usually interpret the first derivative. Negative values indicate a decline in the levels of μ, positive values of d_i indicate incremental values of μ_i, values of d_i near zero may indicate a change point in μ_i or simply very small change or constancy of the signal.

From the point of view of forecasting, it might be beneficial not just to forecast values of the series but also predict values of the first derivative to predict the general future trend: upwards, downwards, constant and/or approaching a change point.

EXAMPLE. *Emerging technologies data (cont.)*

In the case of emerging technologies data, we are not only interested in the local level of the series but also on how the local level changes. Therefore, the LLM-SSM, which provides a way to focus on these two features of a time series, seems appropriate for the study of the quantification of emergence of technological terms. On the left panel below is the time plot of quarterly counts of the term "internal combustion engine" between 2005 and the second quarter of 2018 together with the extracted signal, in red. On the right panel we plot the extracted signal (red) and first derivative (blue). We mark the time points at which the first derivative crosses zero as this marks the time from when the trend is increasing (first derivative positive) to decreasing (first derivative negative) or viceversa. The last panel has the forecasts, 1- to 8-quarters ahead of the last observation, of the level and first derivative. We see that although the trend is declining, it is expected that in 8 quarters there will be a trend reversal.





References

Burmaogluab, Serhat, Olivier Sartenaer and Alan Porter. 2019. "Conceptual definition of technology emergence: A long journey from philosophy of science to science policy." *Technology in Society* 59.

de Jong, Piet. 1991. "The diffuse Kalman filter." *The Annals of Statistics* 19(2):1073–1083.
URL: *https://projecteuclid.org/download/pdf_1/euclid.aos/1176348139*