

Trend analysis using state-space models with an application to emergence studies (DRAFT)

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. The results are used to create an emergence index.

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

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, Sartenauer and Porter \(2019\)](#). In technological emergence, the [Gartner hype cycle](#), a standard adoption model for emerging technologies, is practically an institution in high tech. In this model, trends in expectations on technologies are identified and named: emergence, “The Technology Trigger”, steep upwards trend; excessive enthusiasm, “The Peak of Inflated Expectations”, a change point; excessive disappointment, “The Trough of Disillusionment”, steep downwards trend; gradual, practical adoption, “The Slope of Enlightenment”, moderate upwards trend; and mainstream adoption, “The Plateau of Productivity”, a static or stable state.

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. In Section 7 we propose a way to quantify emergence based on the LLM-SSM.

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 $\epsilon_1, \dots, \epsilon_n$ 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 \(1991\)](#).

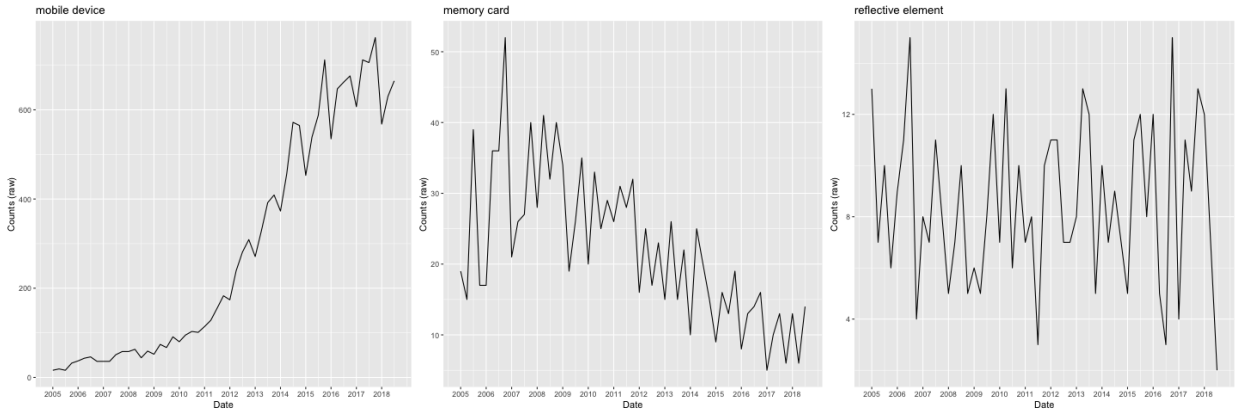
We say that y is pure-trend data if, besides from white noise, the only source of variation of y is trend. Note that we do not wish to explain the data generating process and we do not wish to explain what makes y vary. We wish to extract a smooth signal indicating the local trend and predict its behaviour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



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

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

where $v = \{v_i\}$ and $\eta = \{\eta_i\}$ are independent white noise processes with mean zero and variance σ_v^2 and σ_η^2 respectively. ϵ is independent of v and η .

The full model is

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

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, \dots, u_n are i.i.d with $\text{Var}(u_i) = \sigma^2 I_3$, where I_n denotes the $n \times n$ identity matrix, and

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

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

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

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 $(q - s^t S^{-1} s)/n$ and the log-likelihood (maximised with respect to σ^2 and α_1) is $-\frac{1}{2} [n \log(\hat{\sigma}^2) + \sum_{i=1}^n \log(|D_i|)]$.

Hyperparameter estimation

The parameters $\delta, \sigma_\epsilon, \sigma_\eta, \sigma_v$ 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 $\sigma_v^* = \sigma_v/\sigma_\epsilon$, $\sigma_\eta^* = \sigma_\eta/\sigma_\epsilon$ and δ . σ_v^* 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_v^* , 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 $y = \{y_1, \dots, y_n\}$, 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, $\alpha_{n+k+1} = T\alpha_{n+k} + Hu_{n+k} = \dots = T^k\alpha_{n+1} + H\sum_{j=1}^k T^{j-1}u_{n+k+1-j}$. 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, $E(\alpha_{n+1}|y)$.

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

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

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

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 $E(\mu_t|y)$, $t = 1, \dots, n$. 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 $E(\alpha_{t+1}|y_1, \dots, y_t)$. We use the smoothing filter, a backwards recursion, to obtain the smoothed values of the series.

The Smoothing Filter (SF) is the recursion

$$\begin{aligned} N_{i-1} &= Z^t D_i^{-1} E_i + L_i^t N_i, \\ R_{i-1} &= Z^t D_i^{-1} Z + L_i^t R_i L_i, \end{aligned}$$

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

Then,

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

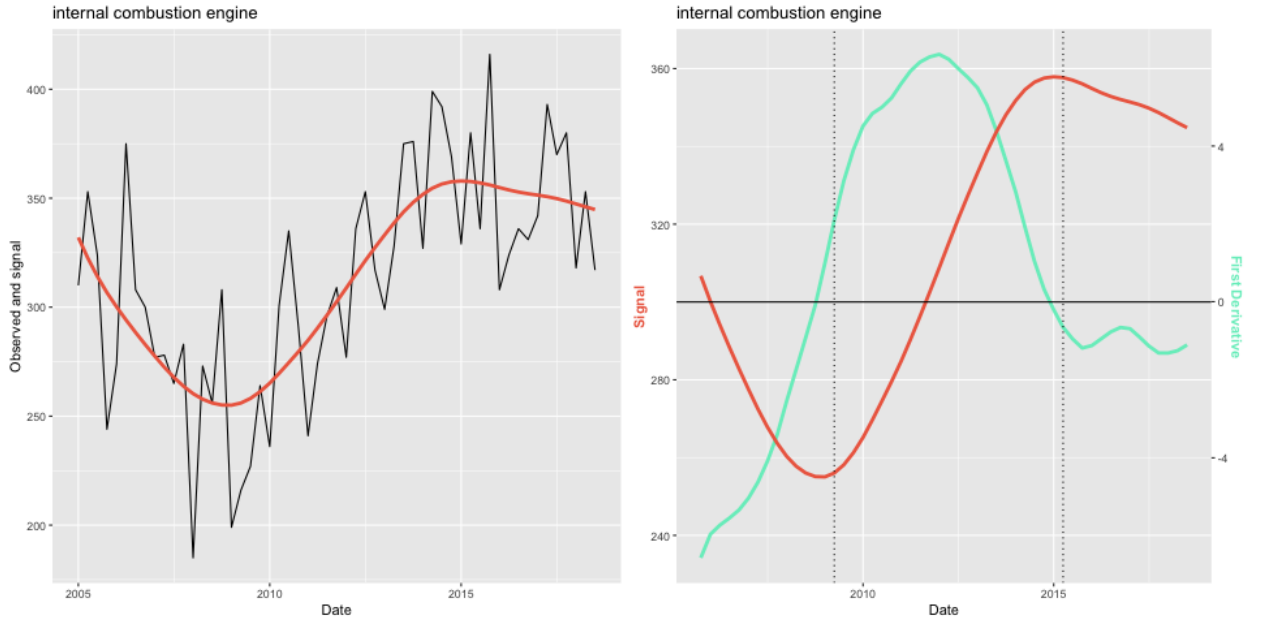
where $N_{i-1, \gamma}$ denotes all but the last column of $A_i + P_i N_{i-1}$.

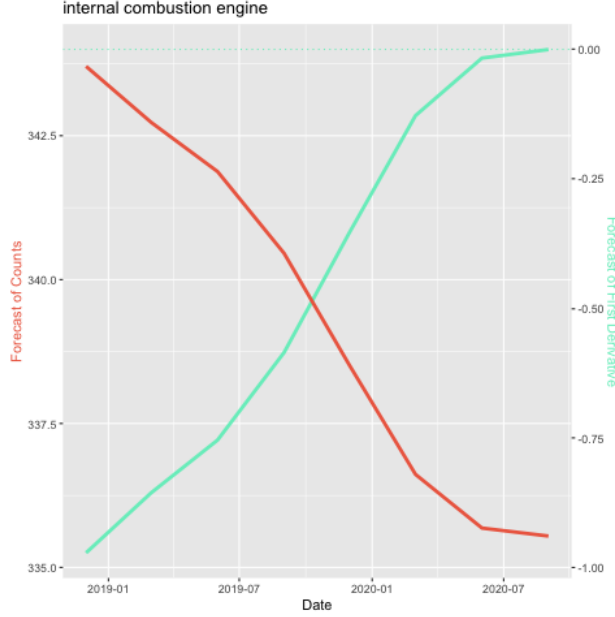
The state vector is $\alpha_i = (\mu_i \ d_i)^t$. The second entry of α_i can be interpreted as we usually interpret the first derivative. Negative values indicate a decline in the levels of μ , positive values of d_i indicate incremental values of μ_i , values of d_i near zero may indicate a change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





7 Constructing an emergence index

In this section we explain how the level or trend and the first derivative of a pure trend series can be combined and used to quantify and compare emergence.

$y = \{y_i\}_{i=1}^n$ represents one time series of quarterly counts of the term during a given time period. We assume that y follows the LLM-SSM (1). $\mu = \{\mu_i\}$ represents the signal, level or trend and σ_ϵ is the noise standard deviation. Also, $d = \{d_i\}_{i=1}^n$ represents the rates of change, or first derivatives, of the signal at each time point.

Positive values of d_i indicate that the signal is increasing, negative values of d_i indicate that the signal is decreasing. If $d_i = 0$ then that indicates a point of maximum, minimum or perhaps an inflection point of the signal.

7.1 The index

During a time period, say consecutive times j_1, \dots, j_m , we define

$$E_1 = \sum_{k=1}^m d_{j_k}$$

as a measure of emergence during that given time period. Positive values of E_1 indicate that growth was dominant in the time period. Negative values of E_1 indicate that decline was dominant in the time period. Values of E_1 near zero, positive or negative, could arise in many ways. One way is if, for example, the d_{j_k} 's are mostly very small and so the series is not changing level greatly. Or, it could be that there are approximately an equal number of positive and negative d 's, and so there was growth and decline in the same period.

Now, how do we compare total emergence of different terms? The derivative measures the rate of change of a function. Given a smooth function f and a point x , the derivative of f at x , the instantaneous rate of change of f at x , is

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

So, we see that if f takes on larger values than g , f' will be generally larger than g' (an exception is if f and g differ by a constant). This means that E_1 can take on large values simply because the values y_i are high.

For this reason, and since we would like to compare emergence for different series y , we define another measure of emergence, which we propose to adopt as an emergence index, which considers the rate of growth relative to the value of the signal, or the ratio of the first derivative and the signal. This is equivalent to what in financial econometrics is called the rate of returns (in continuous time) or simply the returns. Define E_2 to be

$$E_2 = \sum_{k=1}^m \frac{d_{j_k}}{\mu_{j_k}}.$$

where μ_i denotes the signal at time i and d_i denotes the derivative of the signal at time i . We call the measure E_2 *net growth*.

Note that E_1 and E_2 have the same sign.

Normalising the emergence index

It should be noted that E_1 and E_2 may increase or decrease simply because the considered time span is enlarged. For this reason, and to allow a comparison of emergence for the same series during different time periods, we will consider the measures \bar{E}_1 and \bar{E}_2

$$\bar{E}_1 = \frac{E_1}{m}, \quad \text{and} \quad \bar{E}_2 = \frac{E_2}{m},$$

where m is the number of time periods used to compute E_1 and/or E_2 . Using \bar{E}_1 or \bar{E}_2 doesn't change emergence rankings produced for different series during the same time period. Another advantage of using \bar{E}_2 instead of E_2 is that the absolute value of \bar{E}_2 is generally less than 1. This simplifies comparison across series.

Computing E_2

Another important issue with the computation of E_2 is the fact that if some of the counts in y are zero or positive but small, this may inflate the value of E_2 . For this reason, we propose that, for computational purposes, E_2 is defined as

$$E_2 = \sum_k \xi_{j_k}$$

where

$$\xi_{j_k} = d_{j_k} / \mu_{j_k}, \quad \text{if } \mu_{j_k} > 3, \quad \text{and} \quad \xi_{j_k} = 0, \quad \text{if } \mu_{j_k} \leq 3.$$

This doesn't affect the quantification of emergence as we assume that a series with very low counts, fluctuating between 0 and 3, can be deemed to have negligible emergence. This last assumption must therefore be verified for the particular application being considered.

EXAMPLE. *Emerging technologies data (cont.). Comparing E_1 and E_2 .*

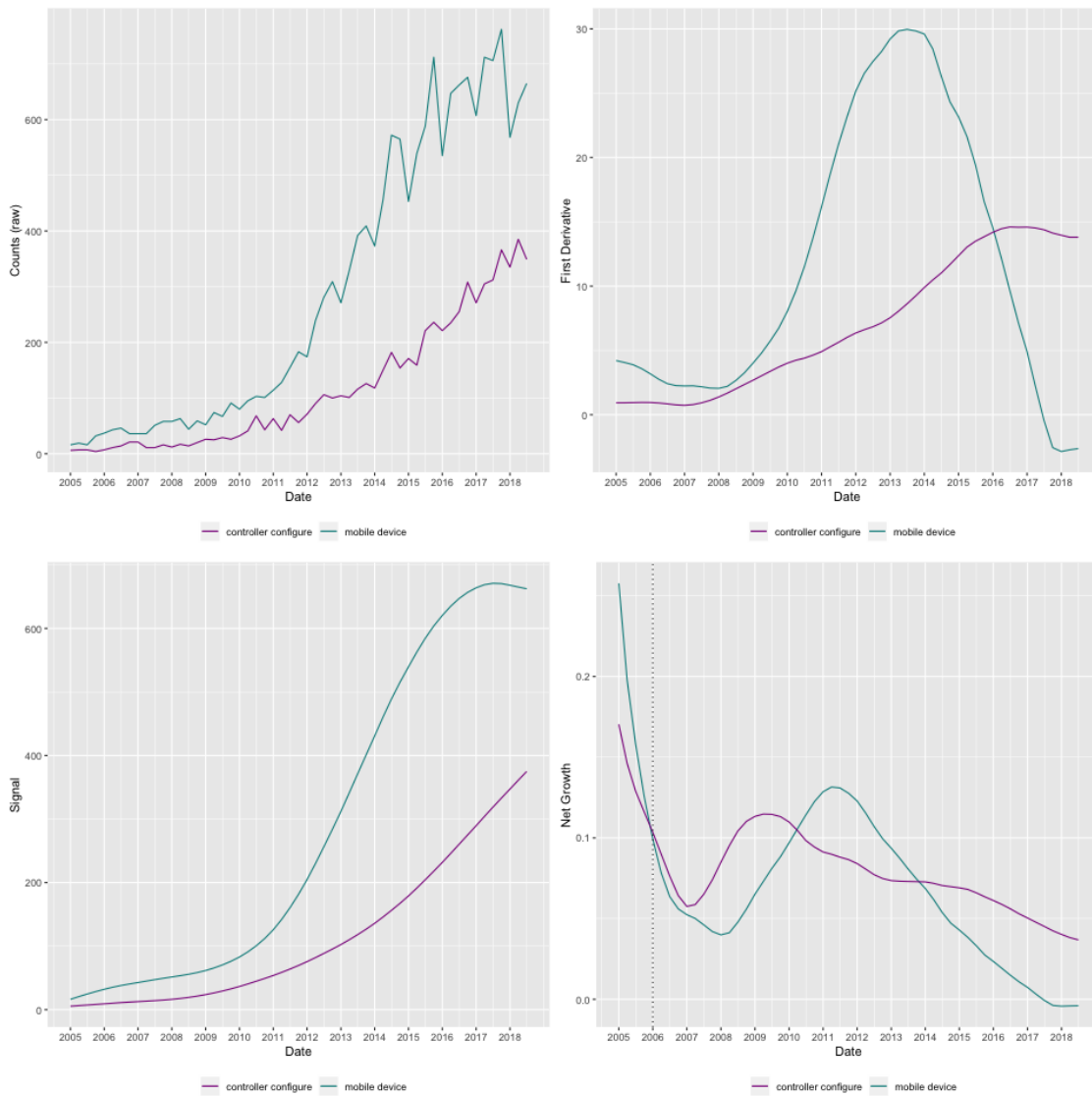
Consider the terms “mobile device” and “controller configure”. For the time period 1st quarter 2005 - 3rd quarter 2018 the raw counts of “mobile device” are higher than those of “controller configure”. See the top left panel of the plots below. The bottom left panel shows the extracted trends or smooth signals for both series of counts. Both series appear to be emergent and, to the naked eye, “mobile device” seems to be “more emergent” than “controller device”. The top right panel shows the extracted first derivatives for both series and again we see the dominance of “mobile device”. When we consider net growth for both series at each time point (the first derivative divided by the signal) we see now that the dominance of “mobile device” is not so clear. In the first considered year, 2005, both series show an elevated net growth, which is common for early stages of emergence.

When we add net growth at each time point, whether we include the first year or not, it becomes clear that although “mobile device” has larger E_1 , $\sum d_i$, “controller configure” has a longer, more steady and persistent growth over the time period than “mobile device”. So, it seems right that “controller configure” should have a larger emergence index than “mobile device”, as reflected by E_2 , $\sum d_i / \mu_i$.

Our assertions above are summarised and quantified in the following table. Note how E_2 allows to focus on the whole time period or in a sub-period.

Term	σ_ϵ	$E_1 = \sum_i d_i$	$E_2 = \sum_i d_i / \mu_i$	$E_2 = \sum_{i \geq 5} d_i / \mu_i$
mobile device	38.838	778.173	3.894	3.155
controller configure	13.567	511.429	4.455	3.893

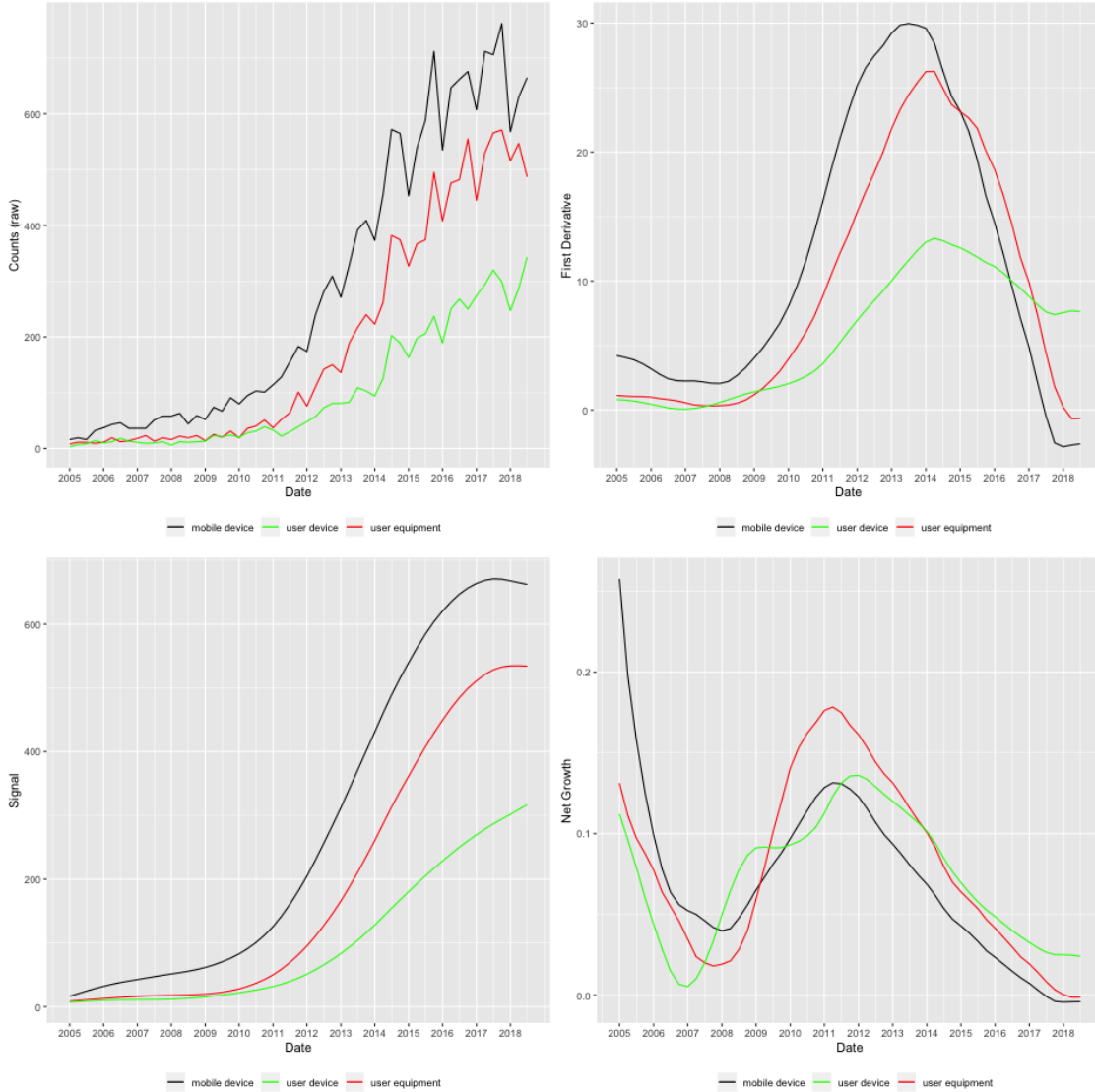
Term	$\bar{E}_1 = \frac{1}{55} \sum_i d_i$	$\bar{E}_2 = \frac{1}{55} \sum_i d_i / \mu_i$	$\bar{E}_2 = \frac{1}{51} \sum_{i \geq 5} d_i / \mu_i$
mobile device	14.149	0.071	0.062
controller configure	9.299	0.081	0.076



EXAMPLE. *Emerging technologies data (cont.). E_2 vs. \bar{E}_2*

We consider the series of counts for “mobile device”, “user equipment” and “user device”.

Note in the plot in the upper top left corner the black curve, “mobile device”, definitely grows the fastest of the three at the very start of the considered time period. The plot in the upper right corner with the first derivatives at each time point ranks emergence according to the magnitude of the signal. The plot in the lower right corner, depicts the ratio of the first derivative by the signal (net growth). In my opinion this point-wise measure of emergence captures the features we would like to focus on: it distinguishes not only the main features of emergence at the beginning of the series but also doesn’t give weight to the particular magnitude or scale of the signal and tracks the emergence well across periods.



If we consider the entire time period, E_2 ranks “user equipment” as most emergent, followed by “user device” and the least emergent series is “mobile device”.

Ignoring the first two years, when the terms first started to emerge, the ranking stays the same as when considering the whole time period. Note that for the term “user equipment”, \bar{E}_2 remains the same whether we consider the first two years or not. On the other hand E_2 is larger when considering the entire time period than when discarding the first two years. This is an example where E_2 increases as the considered time period expands due to the fact that there are more time periods but doesn’t really indicate that there was more emergence in the wider time span.

If we consider only the time from 2014, then the ranking changes to “user device” as being the most emergent series, followed by “user equipment” and “mobile device”.

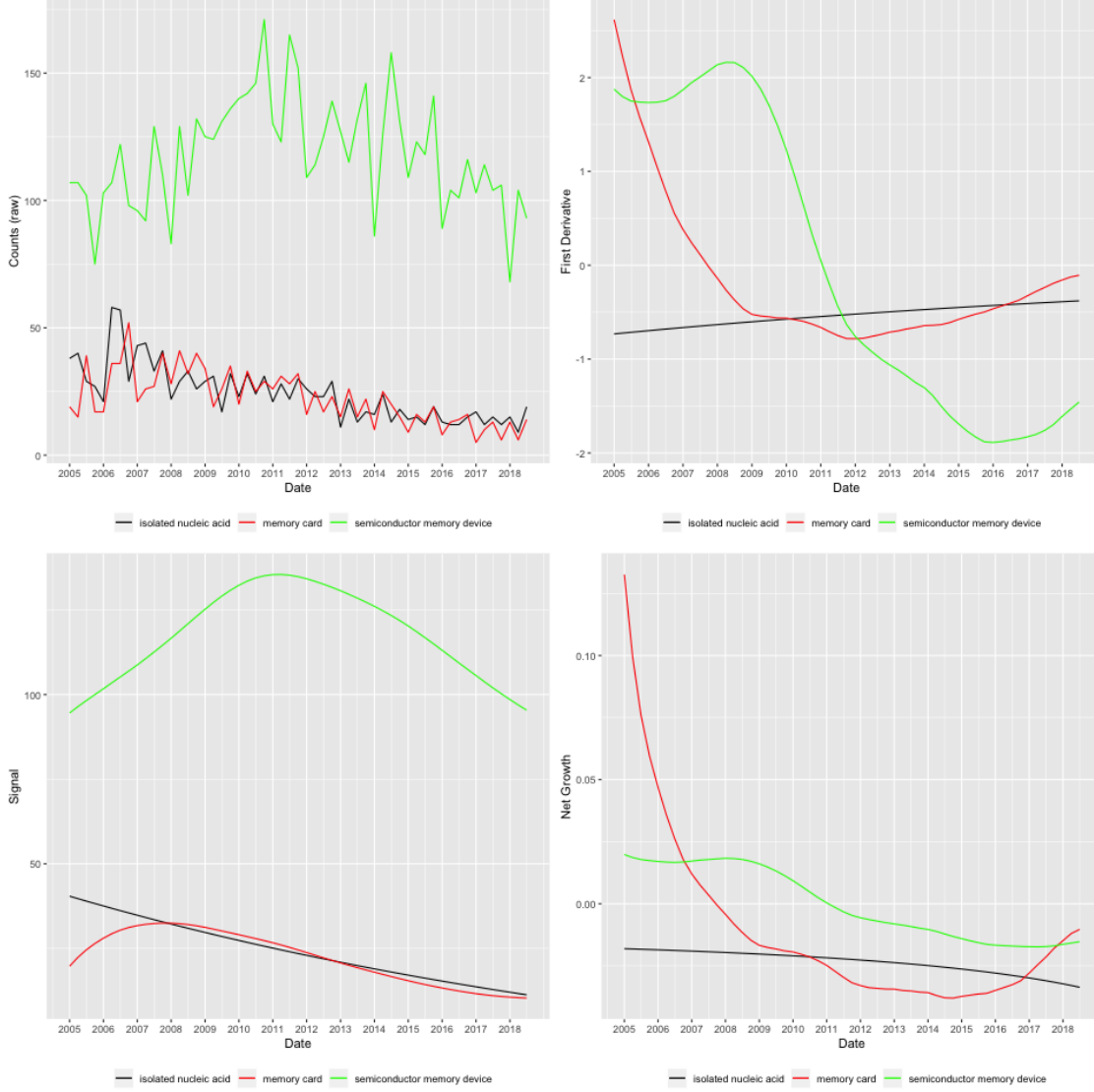
Our assertions above are summarised in the following tables.

Term	σ_ϵ	$E_1 = \sum_i d_i$	$E_2 = \sum_i d_i / \mu_i$	$E_2 = \sum_{i \geq 9} d_i / \mu_i$	$E_2 = \sum_{i \geq 37} d_i / \mu_i$
mobile device	38.838	643.448	3.894	2.858	0.436
user equipment	26.169	524.989	4.371	3.702	0.739
user device	16.673	317.314	3.984	3.542	0.957

Term	$\bar{E}_1 = \frac{1}{55} \sum_i d_i$	$\bar{E}_2 = \frac{1}{55} \sum_i d_i / \mu_i$	$\bar{E}_2 = \frac{1}{47} \sum_{i \geq 9} d_i / \mu_i$	$\bar{E}_2 = \frac{1}{19} \sum_{i \geq 37} d_i / \mu_i$
mobile device	11.699	0.071	0.061	0.023
user equipment	9.545	0.079	0.079	0.039
user device	5.769	0.072	0.075	0.050

EXAMPLE. *Emerging technologies data (cont.). Comparing declining series*

We consider the series of counts for “isolated nucleic acid”, “memory card” and “semiconductor memory device”. The terms are mostly declining in counts. For a few years at the start of the considered time period (2005 - 3rd quarter 2018) “memory card” and “semiconductor memory device” had increasing counts. “semiconductor memory device” has the largest counts throughout the period and also the largest E_2 and \bar{E}_2 values. Without considering the first two years, “memory card” and “isolated nucleic acid” have the same \bar{E}_2 , not E_2 , scores. The same holds for the time period from 2014.

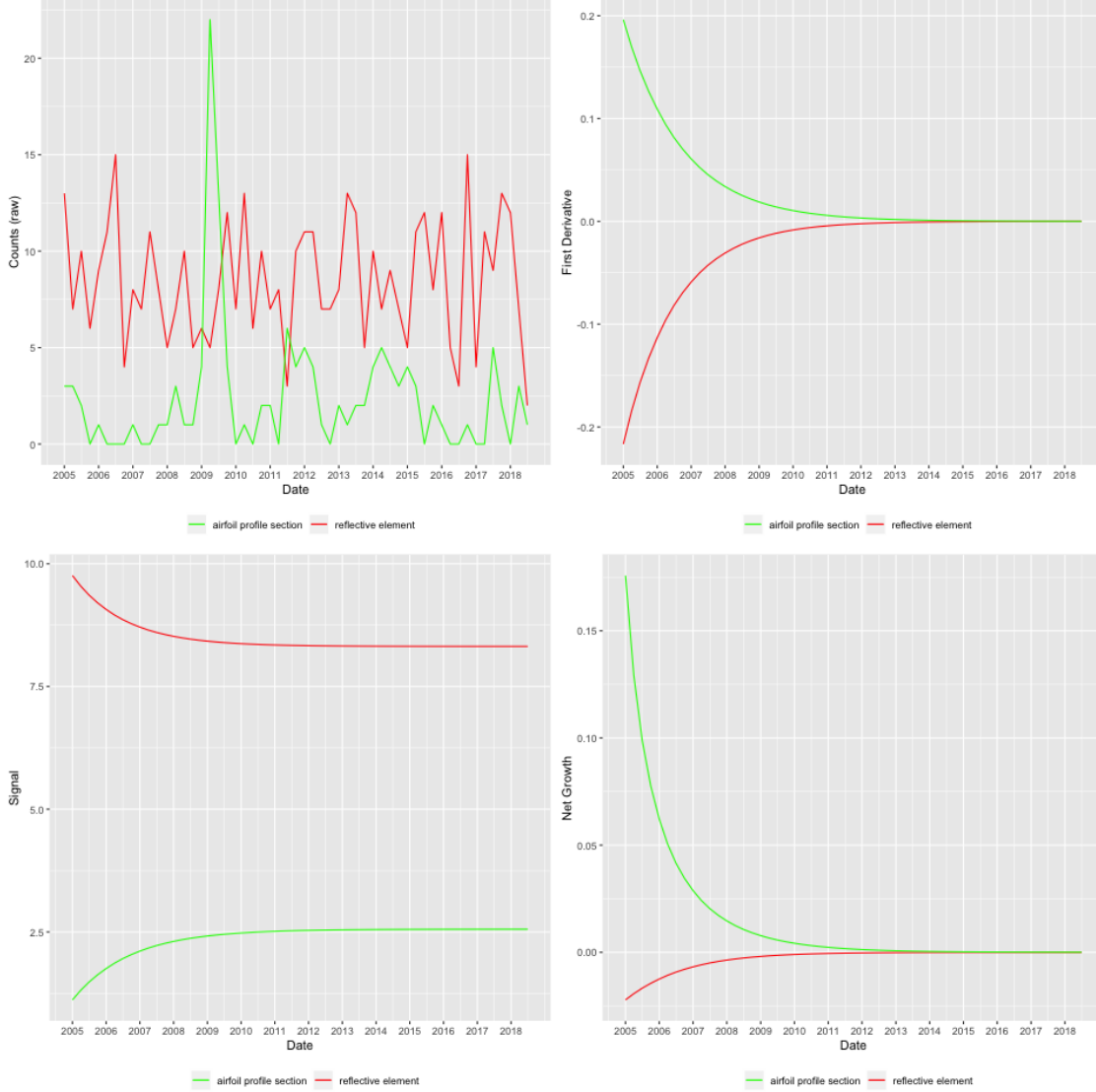


Term	σ_ϵ	$E_1 = \sum_i d_i$	$E_2 = \sum_i d_i / \mu_i$	$E_2 = \sum_{i \geq 9} d_i / \mu_i$	$E_2 = \sum_{i \geq 37} d_i / \mu_i$
isolated nucleic acid	6.801	-29.518	-1.294	-1.146	-0.545
memory card	6.760	-8.838	-0.578	-1.111	-0.546
semiconductor memory device	16.522	-0.634	-0.001	-0.142	-0.292

Term	$\bar{E}_1 = \frac{1}{55} \sum_i d_i$	$\bar{E}_2 = \frac{1}{55} \sum_i d_i / \mu_i$	$\bar{E}_2 = \frac{1}{47} \sum_{i \geq 9} d_i / \mu_i$	$\bar{E}_2 = \frac{1}{19} \sum_{i \geq 37} d_i / \mu_i$
isolated nucleic acid	-0.537	-0.024	-0.024	-0.029
memory card	-0.161	-0.011	-0.024	-0.029
semiconductor memory device	-0.012	0.000	-0.003	-0.015

EXAMPLE. *Emerging technologies data (cont.). Static series*

We consider the series of counts for “reflective element” and “airfoil profile section”. We see that the index \bar{E}_2 is very near zero, regardless of the considered time period. Note the robustness of the index: the model does not follow a single peak in counts for “airfoil profile section”.



Term	σ_ϵ	$E_1 = \sum_i d_i$	$E_2 = \sum_i d_i / \mu_i$	$E_2 = \sum_{i \geq 9} d_i / \mu_i$	$E_2 = \sum_{i \geq 37} d_i / \mu_i$
refl.element	3.104	-1.444	-0.159	-0.046	-0.0005
airf.prof.sect.	3.466	1.439	0.863	0.193	0.0027

Term	$\bar{E}_1 = \frac{1}{55} \sum_i d_i$	$\bar{E}_2 = \frac{1}{55} \sum_i d_i / \mu_i$	$\bar{E}_2 = \frac{1}{47} \sum_{i \geq 9} d_i / \mu_i$	$\bar{E}_2 = \frac{1}{19} \sum_{i \geq 37} d_i / \mu_i$
refl.element	-0.026	-0.003	-0.001	-2e-05
airf.prof.sect.	0.026	0.016	0.004	1e-04

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