Trend Estimation

Proietti, Tommaso

SEFEMEQ, University of Rome "Tor Vergata"

24 March 2010

Trend Estimation

Tommaso Proietti,* SEFEMEQ, University of Rome "Tor Vergata"

Trend estimation deals with the characterization of the underlying, or long–run, evolution of a time series. Despite being a very pervasive theme in time series analysis since its inception, it still raises a lot of controversies. The difficulties, or better, the challenges, lie in the identification of the sources of the trend dynamics, and in the definition of the time horizon which defines the long run. The prevalent view in the literature considers the trend as a genuinely latent component, i.e. as the component of the evolution of a series that is persistent and cannot be ascribed to observable factors. As a matter of fact, the univariate approaches reviewed here assume that the trend is either a deterministic or random function of time.

A variety of approaches is available, which can be classified as nonparametric (kernel methods, local polynomial regression, band-pass filters, and wavelet multiresolution analysis), semiparametric (splines and Gaussian random fields) and parametric, when the trend is modeled as a stochastic process. They will be discussed with respect to the additive decomposition of a time series $y(t)$:

$$y(t) = \mu(t) + \epsilon(t), \quad t = 1, \dots, n, \tag{1}$$

where $\mu(t)$ is the trend component, and $\epsilon(t)$ is the noise, or irregular, component. We assume throughout that $\epsilon(t) = 0$ is a zero mean stationary process, whereas $\mu(t)$ can be a random or deterministic function of time. The above decomposition bears different meanings in different fields. In experimental sciences $\epsilon(t)$ is usually interpreted as a pure measurement error, so that a signal is observed with superimposed random noise. However, in behavioral sciences such has economics, quite often $\epsilon(t)$ is interpreted as a stationary stochastic cycle or as the transitory component of $y(t)$. The underlying idea is that trends and cycles can be ascribed to different economic mechanisms. Moreover, according to some approaches $\mu(t)$ is an underlying deterministic function of time, whereas for other it is a random function (e.g. a random walk, or a Gaussian process), although this distinction becomes more blurred in the case of splines. For some methods, like band pass filtering, the underlying true value $\mu(t)$ is defined by the analyst via the choice of a cutoff frequency which determines the time horizon for the trend.

The simplest and historically oldest approach to trend estimation adopted a global polynomial model for μ_t: $\mu(t) = \sum_{j=0}^{p} \beta_j t^j$. The statistical treatment, based on least squares, is provided in Anderson (1971). It turns out that global polynomials are amenable to mathematical treatment, but are not very flexible: they can provide bad local approximations and behave rather weirdly at the beginning and at the end of the sample period, which is inconvenient for forecasting purposes. More up to date methodologies make the representation more flexible either assuming that certain features, like the coefficients or the derivatives, evolve over time, or that a low order polynomial representation is adequate only as a local approximation.

Local polynomial regression (LPR) is a nonparametric approach that assumes that $\mu(t)$ is a smooth but unknown deterministic function of time, which can be approximated in a neighborhood of time t by a polynomial of degree p of the time distance with time t. The polynomial is fitted by locally weighted least squares, and the weighting function is known as the kernel. LPR generates linear signal extraction filters (also known as moving average filters) whose properties depend on three key ingredients: the order of the approximating polynomial, the size of the neighborhood, also known as the bandwidth, and the choice of the kernel function. The simplest example is the arithmetic moving average $m_t = \frac{1}{2h+1} \sum_{j=-h}^{h} y_{t+j}$,

*To appear in the INTERNATIONAL ENCYCLOPEDIA OF STATISTICAL SCIENCES, Springer

which is the LPR estimator of a local linear trend ($p = 1$) in discrete time using a bandwidth of $2h + 1$ consecutive observations and the uniform kernel.

Trend filters that arise from fitting a locally weighted polynomial to a time series have a well established tradition in time series analysis and signal extraction; see Kendall, Stuart and Ord (1983) and Loader (1999). For instance, the Maculay's moving average filters and the Henderson (1916) filters are intergral part of the X-12 seasonal adjustment procedure adopted by the US Census Bureau. The methodology further encompasses the Nayadara-Watson kernel smoother.

An important class of nonparametric filters arises from the frequency domain notion of a band-pass filter, that is popular in engineering. An ideal low-pass filter retains only the low frequency fluctuations in the series and reduces the amplitude of fluctuations with frequencies higher than a cutoff frequency ω_c. Such a filter is available analytically, but unfeasible, since it requires a doubly infinite sequence of observations; however, it can be approximated using various strategies (see Percival and Walden, 1993). Wavelet multiresolution analysis provides a systematic way of performing band-pass filtering.

An alternative way of overcoming the limitations of the global polynomial model is to add polynomial pieces at given points, called knots, so that the polynomial sections are joined together ensuring that certain continuity properties are fulfilled. Given the set of points $t_1 < \ldots < t_i < \ldots t_k$, a polynomial spline function of degree p with k knots $t_1, \ldots, t_k$ is a polynomial of degree p in each of the $k + 1$ intervals $[t_i, t_{i+1})$, with $p - 2$ continuous derivatives, whereas the $p - 1$-st derivative has jumps at the knots. It can be represented as follows:

$$\mu(t) = \beta_0 + \beta_1(t - t_1) + \cdots + \beta_p(t - t_1)^p + \sum_{i=1}^{k} \eta_i (t - t_i)_+^p, \tag{2}$$

where the set of functions

$$(t - t_i)_+^p = \begin{cases} (t - t_i)^p, & t - t_i \geq 0, \\ 0, & t - t_i < 0 \end{cases}$$

defines what is usually called the truncated power basis of degree p.

According to (2) the spline is a linear combination of polynomial pieces; at each knot a new polynomial piece, starting off at zero, is added so that the derivatives at that point are continuous up to the order $p - 2$. The most popular special case arises for $p = 3$ (cubic spline); the additional *natural boundary conditions*, which constrain the spline to be linear outside the boundary knots, is imposed. See Green and Silverman (1994) and Ruppert, Wand and Carroll (2003).

An important class of semiparametric and parametric time series models are encompassed by (2). The piecewise nature of the spline "reflects the occurrence of structural change" (Poirier, 1973). The knot t_i is the timing of a structural break. The change is "smooth", since certain continuity conditions are ensured. The coefficients η_i, which regulate the size of the break, may be considered as fixed or random. In the latter case $\mu(t)$ is a stochastic process, η_i is interpreted as a *random shock* that drives the evolution of $\mu(t)$, whereas the truncated power function $(t - t_i)_+^p$ describes its *impulse response function*, that is the impact on the future values of the trend.

If the η_i's are considered as random, the spline model can be formulated as a linear mixed model, which is a traditional regression model extended so as to incorporate random effects. Denoting $\mathbf{y} = [y(t_1), \ldots, y(t_n)]'$, $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_n]'$, $\boldsymbol{\epsilon} = [\epsilon(t_1), \ldots, \epsilon(t_n)]'$, $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\eta}$,

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\epsilon} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{3}$$

where the t-th row of $\mathbf{X}$ is $[1, (t - 1), \ldots, (t - 1)^p]$, and $\mathbf{Z}$ is a known matrix whose i-th column contains the impulse response signature of the shock η_i, $(t - t_i)_+^p$.

The trend is usually fitted by penalized least squares (PLS), which chooses $\boldsymbol{\mu}$ so as to minimize

$$(\mathbf{y} - \boldsymbol{\mu})'(\mathbf{y} - \boldsymbol{\mu}) + \lambda \int \left[\frac{d^{p-1}\mu(t)}{dt^{p-1}} \right]^2 dt, \tag{4}$$

where $\lambda \geq 0$ is the smoothness parameter.

PLS is among the most popular criteria for designing filters that has a long and well established tradition in actuarial sciences and economics (see Whittaker, 1923, Leser, 1963, and, more recently, Hodrick and Prescott, 1997). Under Gaussian independent measurement noise minimizing the PLS criterion amounts to finding the conditional mode of μ given $\mathbf{y}$, This is a solution to the smoothing problem. If $\mu(t)$ is random, the minimum mean square estimator of the signal is $\mathrm{E}(\mu(t)|\mathbf{y})$. If the model (1) is Gaussian, these inferences are linear in the observations. The computations are carried out efficiently by the Kalman filter and the associated smoother (see Wecker and Ansley, 1983).

The linear mixed model representation (3) encompasses other approaches, according to which the component $\mathbf{Z}\boldsymbol{\eta}$ is a Gaussian random process (Rasmussen and Williams, 2006), or a (possibly nonstationary) time series process with a Markovian representation, such as in the structural time series approach see Harvey (1989), and in the canonical decomposition of time series (see Hillmer and Tiao, 1986). The Markovian nature of the opens the way to the statistical treatment by the state space methodology and signal extraction is carried out efficiently by the Kalman filter and smoother. Popular predictors, such as exponential smoothing and Holt and Winters, arise as special cases (see Harvey, 1989). The representation theory for the estimator of the trend component, Wiener-Kolmogorov filter, is established in Whittle (1983).

The analysis of economic time series has contributed to trend estimation in several ways. The first contribution is the attempt to relate the trend to a particular economic mechanism. The issue at stake is whether $\mu(t)$ is better characterized as a deterministic or stochastic trends. This problem was addressed in a very influential paper by Nelson and Plosser (1982), who adopted the (augmented) Dickey Fuller test for testing the hypothesis that the series is integrated of order 1, I(1), implying that $y(t)-y(t-1)$ is a stationary process versus the alternative that it is trend-stationary, e.g. $m(t) = \beta_0 + \beta_1 t$. Using a set of annual U.S. macroeconomic time series they are unable to reject the null for most series and discuss the implications for economic interpretation. The trend in economic aggregate is the cumulative effect of supply shocks, i.e. shocks to technology that occur randomly and propagate through the economic system via a persistent transmission mechanism.

A fundamental contribution is the notion of cointegration (Engle and Granger, 1987), according to which two or more series are cointegrated if they are themselves nonstationary (e.g. integrated of order 1), but a linear combination of them is stationary. Cointegration results from the presence of a long run equilibrium relationship among the series, so that the same random trends drive the nonstationary dynamics of the series; also, part of the short run dynamics are also due to the adjustment to the equilibrium.

A third contribution, related to trend estimation, is the notion of spurious cycles that may result from inappropriate detrending of a nonstationary time series. This effect is known as the Slutzky-Yule effect, and concerned with the fact that an ad hoc filter to a purely random series can introduce artificial cycles.

Finally, large dimensional dynamic factor models have become increasingly popular in empirical macroeconomics. The essential idea is that the precision by which the common components are estimated can be increased by bringing in more information from related series: suppose for simplicity that $y_i(t) = \theta_i \mu(t) + \epsilon_i(t)$, where the i-th series, $i = 1, \dots, N$, depends on the same stationary common factor, which is responsible for the observed comovements of economic time series, plus an idiosyncratic component, which includes measurement error and local shocks. Generally, multivariate methods provide more reliable measurements provided that a set of related series can be viewed as repeated measures of the same underlying latent variable. Stock and Watson (2002) and Forni *et al.* (2000) discuss the conditions on μ_t and ϵ_{it} under which dynamic or static principal components yield consistent estimates of the underlying factor μ_t as both N and the number of time series observations tend to infinity.

References

Anderson, T.W. (1971), *The Statistical Analysis of Time Series*, Wiley, New York.

Engle, R.F. and Granger, C.W.J. (1987). Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55, 251–276.

Forni, M., Hallin, M., Lippi, F., Reichlin L. (2000). The Generalized Dynamic Factor Model: Identification and Estimation. *Review of Economics and Statistics* 82, 540-554.

Green P.J. and Silverman, B.V. (1994) *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach*. Chapman & Hall, London.

Harvey, A.C. (1989). *Forecasting, Structural Time Series and the Kalman Filter*, Cambridge University Press, Cambridge, UK.

Henderson, R. (1916). Note on Graduation by Adjusted Average, *Transaction of the Actuarial Society of America*, 17, 43-48.

Hillmer, S.C. and Tiao G.C. (1982) An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association*, 77, 63-70.

Hodrick, R., and Prescott, E.C. (1997). Postwar U.S. Business Cycle: an Empirical Investigation, *Journal of Money, Credit and Banking*, 29, 1, 1-16.

Kendall M., Stuart, A., and Ord, J.K. (1983). *The Advanced Theory of Statistics*, Vol 3. C. Griffin.

Leser, C.E.V. (1961). A Simple Method of Trend Construction, *Journal of the Royal Statistical Society B*, 23, 91-107.

Loader, C. (1999). *Local regression and likelihood*. Springer-Verlag, New York

Nelson, C.R., and Plosser, C.I. (1982). Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics*, 10, 139-62.

Percival D., Walden A. (1993). *Spectral Analysis for Physical Applications*. Cambridge University Press.

Poirier, D.J. (1973). Piecewise regression using cubic splines, *Journal of American Statistical Association*, 68, 515-524.

Rasmussen, C.E. and Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*. The MIT Press.

Ruppert D., Wand, M.J. and Carroll R.J. (2003). *Semiparametric regression*, Cambridge University Press.

Stock, J. H. and Watson, M.W. (2002b). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97 1167-1179.

Watson, G.S. (1964). Smooth regression analysis, *Shankya* Series A, 26, 359-372.

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

Whittaker, E. (1923). On new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41, 63–75.

Whittle P. (1983) *Prediction and Regulation by Linear Least Squares Methods*, Second edition. Basil Blackwell, Oxford.