

Robust time series models with trend and seasonal components

Michele Caivano¹ · Andrew Harvey² ·
Alessandra Luati³

Received: 13 January 2015 / Accepted: 3 November 2015 / Published online: 10 December 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract We describe observation driven time series models for Student-t and EGB2 conditional distributions in which the signal is a linear function of past values of the score of the conditional distribution. These specifications produce models that are easy to implement and deal with outliers by what amounts to a soft form of trimming in the case of t and a soft form of Winsorizing in the case of EGB2. We show how a model with trend and seasonal components can be used as the basis for a seasonal adjustment procedure. The methods are illustrated with US and Spanish data.

Keywords Fat tails · EGB2 · Score · Robustness · Student's t · Trimming · Winsorizing

JEL Classification C22 · G17

1 Introduction

Time series are often subject to observations that, when judged by the Gaussian yardstick, are outliers. This is a very real issue for many economic time series. Agustin Maravall's seasonal adjustment program, TRAMO-SEATS, which he developed jointly with Victor Gomez, tackles the problem by identifying outliers and, where appropriate, replacing them by dummy variables. Here we take a different approach

✉ Andrew Harvey
ach34@cam.ac.uk

¹ Bank of Italy, Rome, Italy

² University of Cambridge, Cambridge, UK

³ University of Bologna, Bologna, Italy

in which we employ a new class of robust models where the dynamics of the level, or location, are driven by the score of the conditional distribution of the observations. These called dynamic conditional score (DCS) models have recently been developed by [Creal et al. \(2011, 2013\)](#) and [Harvey \(2013\)](#). They are relatively easy to implement and their form facilitates the development of a comprehensive and relatively straightforward theory for the asymptotic distribution of the maximum likelihood estimator.

The changing level of a Gaussian time series is usually obtained from an ARIMA process or explicitly modeled as an unobserved component. The statistical treatment of linear Gaussian unobserved components models is straightforward, with the Kalman filter playing a key role. Additive outliers may be captured by dummy variables. A different way forward is to let the noise have a Student t -distribution, thereby accommodating the outliers. However, the treatment of such a model requires computationally intensive procedures, as described in [Durbin and Koopman \(2012\)](#). The DCS- t model proposed by [Harvey and Luati \(2014\)](#) provides an alternative approach which is observation-driven in that the conditional distribution of the observations is specified. A model of this kind may be compared and contrasted with the methods in the robustness literature; see [Maronna, Martin and Yohai \(2006, ch 8\)](#) and [McDonald and Newey \(1988\)](#), where a parametric approach is called ‘partially adaptive’. Robust procedures for guarding against additive outliers typically respond to large observations in one of two ways: either the response function converges to a positive (negative) constant for observations tending to plus (or minus) infinity or it goes to zero. These two procedures are usually classified as Winsorizing or as trimming. The score for a t -distribution converges to zero and so can be regarded as a parametric form of trimming. Similarly a parametric form of Winsorizing is given by the exponential generalized beta distribution of the second kind (EGB2) distribution. The article by [Caivano and Harvey \(2014\)](#) sets out the theory for the DCS location model with an EGB2 distribution and illustrates its practical value.

The article is organized as follows. Section 2 reviews the idea behind the DCS location model and expands on the reason for using the conditional score. The various classifications of distributions in terms of their tails are then set out and this is followed by a discussion of the score as it relates to location and scale and the link with robust estimation. This material leads on to the contrast between the DCS t and EGB2 models in Sect. 3. Section 4 compares the fit of EGB2 and t -distributions for two US macroeconomic time series. A comparison between the way in which Gaussian and DCS models adapt to structural breaks is made in Sect. 5. This issue is important, because it may be thought that the price paid by DCS models for their robustness is a slow response to structural breaks.

DCS models with trend and seasonal components are described in Sect. 6. These models can be regarded as robust counterparts to the unobserved components ‘basic structural model’ (BSM). Seasonal adjustment can be carried out with the BSM by extracting smoothed components using the standard Kalman filter and smoother. Because DCS models only give filtered components, it is necessary to devise a method for smoothing. In Sect. 7 we apply this method to a monthly series of tourists arriving in Spain and compare the extracted trend and seasonal components with those obtained by fitting a BSM with the outliers handled by dummy variables.

2 Filters, heavy tails and robust estimation

The first sub-section below sets out a simple unobserved components model and shows how the innovations form of the Kalman filter may be adapted to form a DCS model. The second sub-section provides a rationale for the use of the conditional score. The way in which tails of distributions may be classified is reviewed in the third sub-section and in the fourth, tail behaviour is related to the considerations of robustness. The treatment of these topics is more general and integrated than in [Harvey \(2013\)](#).

2.1 Unobserved components and filters

A simple Gaussian signal plus noise model is

$$\begin{aligned}
 y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim NID(0, \sigma_\varepsilon^2), & t &= 1, \dots, T, \\
 \mu_{t+1} &= \phi\mu_t + \eta_t, & \eta_t &\sim NID(0, \sigma_\eta^2), & &
 \end{aligned}
 \tag{1}$$

where the irregular and level disturbances, ε_t and η_t respectively, are mutually independent and the notation $NID(0, \sigma^2)$ denotes normally and independently distributed with mean zero and variance σ^2 . The autoregressive parameter is ϕ , while the signal-noise ratio, $q = \sigma_\eta^2/\sigma_\varepsilon^2$, plays the key role in determining how observations should be weighted for prediction and signal extraction. The reduced form (RF) of (1) is an ARMA(1,1) process

$$y_t = \phi y_{t-1} + \xi_t - \theta \xi_{t-1}, \quad \xi_t \sim NID(0, \sigma^2), \quad t = 1, \dots, T \tag{2}$$

but with restrictions on θ . For example, when $\phi = 1, 0 \leq \theta \leq 1$.

The UC model in (1) is effectively in state space form and, as such, it may be handled by the Kalman filter (KF); see [Harvey \(1989\)](#). The parameters ϕ and q may be estimated by maximum likelihood (ML), with the likelihood function constructed from the one-step ahead prediction errors. The KF can be expressed as a single equation which combines $\mu_{t|t-1}$, the optimal estimator of μ_t based on information at time $t - 1$, with y_t in order to produce the best estimator of μ_{t+1} . Writing this equation together with an equation that defines the one-step ahead prediction error, v_t , gives the innovations form of the KF:

$$\begin{aligned}
 y_t &= \mu_{t|t-1} + v_t, \\
 \mu_{t+1|t} &= \phi\mu_{t|t-1} + k_t v_t.
 \end{aligned}
 \tag{3}$$

The Kalman gain, k_t , depends on ϕ and q . In the steady-state, k_t is constant. Setting it equal to κ in (3) and re-arranging gives the ARMA model (2) with $\xi_t = v_t$ and $\phi - \kappa = \theta$.

When the noise in (1) comes from a heavy-tailed distribution such as Student’s t it can give rise to observations which, when judged against the yardstick of a Gaussian

distribution, are additive outliers. As a result fitting a Gaussian model is inefficient and may even yield estimators which are inconsistent. Simulation methods, such as Markov chain Monte Carlo (MCMC) and particle filtering, provide the basis for a direct attack on such non-Gaussian models; see [Durbin and Koopman \(2012\)](#). However, simulation-based estimation can be time-consuming and subject to a degree of uncertainty. In addition the statistical properties of the estimators are not easy to establish.

The DCS approach begins by writing down $f_t(y_t)$, the distribution of the t -th observation conditional on past observations. The time-varying parameter is then updated by a suitably defined filter. Such a model is said to be observation driven. In a linear Gaussian UC model, the KF depends on the one step-ahead prediction error. The main ingredient in the DCS filter for non-Gaussian distributions is the replacement of v_t in the KF equation by a variable, u_t , that is proportional to the score of the conditional distribution; compare Maronna, Martin and Yohai ([2006](#), p. 272–4) and the references therein. Thus the second equation in (3) becomes

$$\mu_{t+1|t} = \phi \mu_{t|t-1} + \kappa u_t, \quad (4)$$

where $u_t = \partial \ln f_t(y_t) / \partial \mu_{t|t-1}$ and κ is treated as an unknown parameter. This filter could be regarded as an approximation to the computer intensive solution for the parameter driven unobserved components model. The attraction of regarding it as a model in its own right is that it becomes possible to derive the asymptotic distribution of the maximum likelihood estimator and generalize in various directions.

2.2 Why the Score?

Suppose that at time $t - 1$ we have $\hat{\theta}_{t-1}$ the ML estimate of a parameter, θ . Then the score is zero, that is

$$\frac{\partial \ln L_{t-1}(\theta)}{\partial \theta} = \sum_{j=1}^{t-1} \frac{\partial \ln \ell_j(\theta)}{\partial \theta} = 0 \quad \text{at} \quad \theta = \hat{\theta}_{t-1}, \quad (5)$$

where $\ell_j(\theta; y_j) = f(y_j; \theta)$. When a new observation becomes available, a single iteration of the method of scoring gives

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{I_t(\hat{\theta}_{t-1})} \frac{\partial \ln L_t(\theta)}{\partial \theta} = \hat{\theta}_{t-1} + \frac{1}{I_t(\hat{\theta}_{t-1})} \frac{\partial \ln \ell_t(\theta)}{\partial \theta},$$

where $I_t(\hat{\theta}_{t-1}) = t \cdot I(\hat{\theta}_{t-1})$ is the information matrix for t observations and the last expression follows because of (5). For a Gaussian distribution a single update goes straight to the ML estimate at time t (recursive least squares).

Remark 1 We may sometimes be able to choose a link function so that the information quantity does not depend on θ .

As $t \rightarrow \infty$, $I_t(\widehat{\theta}_{t-1}) \rightarrow \infty$ so the recursion becomes closed to new information. If it is thought that θ changes over time, the filter needs to be opened up. This may be done by replacing $1/t$ by a constant, which may be denoted as κ . Thus

$$\widehat{\theta}_t = \widehat{\theta}_{t-1} + \kappa \frac{1}{I(\widehat{\theta}_{t-1})} \frac{\partial \ln \ell_t(\theta)}{\partial \theta}.$$

With no information about how θ might evolve, the above equation might be converted to the predictive form by letting $\widehat{\theta}_{t+1|t} = \widehat{\theta}_t$. Thus

$$\widehat{\theta}_{t+1|t} = \widehat{\theta}_{t|t-1} + \kappa \frac{1}{I(\widehat{\theta}_{t|t-1})} \frac{\partial \ln \ell_t(\theta)}{\partial \theta}. \tag{6}$$

For a Gaussian distribution in which θ is the mean and the variance is known to be σ^2 ,

$$\frac{1}{I(\widehat{\theta}_{t|t-1})} \frac{\partial \ln \ell_t(\theta)}{\partial \theta} = y_t - \widehat{\theta}_{t|t-1}$$

and (6) is an exponentially weighted moving average (EWMA).

If there is reason to think that the parameter tends to revert to an underlying level, ω , the updating scheme might become

$$\widehat{\theta}_{t+1|t} = \omega(1 - \phi) + \phi\widehat{\theta}_{t|t-1} + \kappa \frac{1}{I(\widehat{\theta}_{t|t-1})} \frac{\partial \ln \ell_t(\theta)}{\partial \theta}$$

where $|\phi| < 1$. This scheme corresponds to a first-order autoregressive process. More generally we might introduce lags so as to smooth out the changes or allow for periodic effects.

2.3 Heavy tails

The Gaussian distribution has kurtosis of three and a distribution is said to exhibit *excess kurtosis*, or to be *leptokurtic*, if its kurtosis is greater than three. Although some researchers take excess kurtosis as defining heavy tails, it is not, in itself, an ideal measure, particularly for asymmetric distributions. Most classifications in the insurance and finance literature begin with the behaviour of the upper tail for a non-negative variable, or one that is only defined above a minimum value; see [Embrechts et al. \(1997\)](#). The two which are relevant here are as follows.

A distribution is said to be *heavy-tailed* if

$$\lim_{y \rightarrow \infty} \exp(y/\alpha) \overline{F}(y) = \infty \quad \text{for all } \alpha > 0, \tag{7}$$

where $\bar{F}(y) = \Pr(Y > y) = 1 - F(y)$ is the survival function. When y has an exponential distribution, $\bar{F}(y) = \exp(-y/\alpha)$, so $\exp(y/\alpha)\bar{F}(y) = 1$ for all y . Thus the exponential distribution is not heavy-tailed.

A distribution is said to be *fat-tailed* if, for a fixed positive value of η ,

$$\bar{F}(y) = cL(y)y^{-\eta}, \quad \eta > 0, \tag{8}$$

where c is a non-negative constant and $L(y)$ is slowly varying,¹ that is

$$\lim_{y \rightarrow \infty} \frac{L(ky)}{L(y)} = 1.$$

The parameter η is the tail index. The implied PDF is a *power law* PDF

$$f(y) \sim cL(y)\eta y^{-\eta-1}, \quad y \rightarrow \infty, \quad \eta > 0, \tag{9}$$

where \sim is defined such that $a(x) \sim b(x)$ as $x \rightarrow x_0$ if $\lim_{x \rightarrow x_0} (a/b) \rightarrow 1$. The m -th moment exists if $m < \eta$. The Pareto distribution is a simple case in which $\bar{F}(y) = y^{-\eta}$ for $y > 1$. If a distribution is fat-tailed then it must be heavy-tailed, but the converse is not true; see Embrechts, Kluppelberg and Mikosch (1997, p. 41–2).

The above criteria are related to the behavior of the conditional score and whether or not it discounts large observations. This, in turn, connects to robustness, as shown in the sub-section following. More specifically, consider a power law PDF, (9), with y divided by a scale parameter, φ , so that $\bar{F}(y/\varphi) = cL(y/\varphi)(y/\varphi)^{-\eta}$ and $f(y) \sim cL(y)\varphi^{-1}\eta(y/\varphi)^{-\eta-1}$. Then

$$\partial \ln f / \partial \varphi \sim \eta / \varphi \quad \text{as } y \rightarrow \infty \tag{10}$$

and so the score is bounded. With the exponential link function, $\varphi = \exp(\lambda)$, $\partial \ln f / \partial \lambda \sim \eta$ as $y \rightarrow \infty$. Similarly as $y \rightarrow 0$, $\partial \ln f / \partial \lambda \sim \bar{\eta}$.

The logarithm of a variable with a fat-tailed distribution has exponential tails. Let x denote a variable with a fat-tailed distribution in which the scale is written as $\varphi = \exp(\mu)$ and let $y = \ln x$. Then for large y

$$f(y) \sim cL(e^y)\eta e^{-\eta(y-\mu)}, \quad \eta > 0, \quad \text{as } y \rightarrow \infty,$$

whereas as $y \rightarrow -\infty$, $f(y) \sim cL(e^y)\bar{\eta}e^{\bar{\eta}(y-\mu)}$, $\bar{\eta} > 0$. Thus y is not heavy-tailed, but it may exhibit excess kurtosis. The score with respect to location, μ , is the same as the original score with respect to the logarithm of scale and so tends to η as $y \rightarrow \infty$.

¹ More generally regularly varying is $\lim_{y \rightarrow \infty} (L(ky)/L(y)) = k^\beta$; see Embrechts, Kluppelberg and Mikosch (1997, p. 37, 564). Fat-tailed distributions are regularly varying with $\eta = -\beta > 0$.

2.4 Robust estimation

The location-dispersion model is

$$y_t = \mu + \varphi \varepsilon_t, \quad t = 1, \dots, T, \tag{11}$$

where ε_t is a standardized variable with PDF denoted $\exp \rho(\varepsilon_t)$ and the scale, φ , is called the dispersion for y_t ; see Maronna, Martin and Yohai (2006, p. 37–8). The density for y_t is

$$f(y_t; \mu, \varphi, \xi) = \varphi^{-1} \exp \rho((y_t - \mu)/\varphi),$$

where ξ denotes one or more shape parameters, and the scores for μ and φ are given by differentiating $\ln f(y_t) = \rho((y_t - \mu)/\varphi) - \ln \varphi$. The score for location is

$$\frac{\partial \ln f_t}{\partial \mu} = \frac{\partial \rho(z_t)}{\partial \mu} = \psi_L(z_t),$$

where $z_t = (y_t - \mu)/\varphi$, whereas the score for scale is

$$\frac{\partial \ln f_t}{\partial \varphi} = \frac{\partial \rho(z_t)}{\partial \varphi} - \frac{1}{\varphi} = \psi_S(z_t) - \frac{1}{\varphi}.$$

Note that $\psi_S(z_t) = \varphi^{-1} z_t \psi_L(z_t)$. When the scale is parameterized with an exponential link function, $\varphi = \exp \lambda$, the score is

$$\frac{\partial \ln f}{\partial \lambda} = \frac{\partial \rho(z_t)}{\partial \lambda} - 1 = z_t \psi_L(z_t) - 1. \tag{12}$$

If $y_t = \ln x_t$, where $x_t = \varphi \varepsilon_t$ as in (11) and then the logarithm of the scale parameter for x_t , that is $\ln \varphi$, becomes the location for y_t . Hence $\psi_L(y_t) = \psi_S(x_t)$.

The ML estimators are asymptotically efficient, assuming certain regularity conditions hold. More generally $\rho(\cdot)$ may be any function deemed to yield estimators with good statistical properties. In particular, the estimators should be robust to observations which would be considered to be outliers for a normal distribution. When normality is assumed, the ML estimators of the mean and variance are just the corresponding sample moments, but these can be subject to considerable distortion when outliers are present. Robust estimators, on the other hand, are resistant to outliers while retaining relatively high efficiency when the data are from a normal distribution.

The M-estimator, which features prominently in the robustness literature, has a Gaussian response until a certain threshold, K , whereupon it is constant; see Maronna, Martin and Yohai (2006, p. 25–31). This is known as Winsorizing as opposed to trimming, where observations greater than K in absolute value are given a weight of zero.²

² In both cases a (robust) estimate of scale needs to be pre-computed and the process of computing M-estimates is then often iterated to convergence.

3 DCS location models

The stationary first-order DCS model corresponds to the Gaussian innovations form, (3), and is

$$\begin{aligned} y_t &= \mu_{t|t-1} + v_t = \mu_{t|t-1} + \exp(\lambda)\varepsilon_t, & t = 1, \dots, T, \\ \mu_{t+1|t} &= \delta + \phi\mu_{t|t-1} + \kappa u_t, & |\phi| < 1, \end{aligned} \quad (13)$$

where $\omega = \delta/(1 - \phi)$ is the unconditional mean of $\mu_{t|t-1}$, ε_t is a serially independent variate with unit scale³ and u_t is proportional to the conditional score, that is $u_t = k \cdot \partial \ln f(y_t | y_{t-1}, y_{t-2}, \dots) / \partial \mu_{t|t-1}$, where k is a constant.

More generally, an ARMA-type model of order (p, r) is

$$\mu_{t+1|t} = \delta + \phi_1 \mu_{t|t-1} + \dots + \phi_p \mu_{t-p+1|t-p} + \kappa_0 u_t + \kappa_1 u_{t-1} + \dots + \kappa_r u_{t-r}. \quad (14)$$

In the Gaussian case $u_t = y_t - \mu_{t|t-1}$ and if q is defined as $\max(p, r + 1)$, then y_t is an $ARMA(p, q)$ process with MA coefficients $\theta_i = \phi_i - \kappa_{i-1}$, $i = 1, \dots, q$. Nonstationary ARIMA-type models may also be constructed as may structural times series models with trend and seasonal components. Explanatory variables can be introduced into DCS models, as described in [Harvey and Luati \(2014\)](#).

Maronna, Martin and Yohai (2006, Sect 8.6 and 8.8) give a robust algorithm for AR and ARMA models with additive outliers. For a first-order model their filter is essentially the same as (13) except that their dynamic equation is driven by a robust ψ -function and they regard the model as an approximation to a UC model.⁴

3.1 Student t model

The Student t_ν distribution has fat tails for finite degrees of freedom, ν , with the tail index given by ν . Moments exist only up to and including $\nu - 1$. The excess kurtosis, that is the amount by which the normal distribution's kurtosis of three is exceeded, is $6/(\nu - 4)$, provided that $\nu > 4$.

When the location changes over time, it may be captured by a model in which, conditional on past observations, y_t has a t_ν -distribution with $\mu_{t|t-1}$ generated by a linear function of

$$u_t = \left(1 + \nu^{-1} e^{-2\lambda} (y_t - \mu_{t|t-1})^2\right)^{-1} v_t, \quad t = 1, \dots, T, \quad (15)$$

³ The standard deviation is $\sqrt{\nu/(\nu - 2)}$ times the scale.

⁴ Muler, Peña and Yohai (2009, p. 817) note two shortcomings of the estimates obtained in this way. They write: 'First, these estimates are asymptotically biased. Second, there is not an asymptotic theory for these estimators, and therefore inference procedures like tests or confidence regions are not available.' They then suggest a different approach and show that it allows an asymptotic theory to be developed.

where $v_t = y_t - \mu_{t|t-1}$ is the prediction error. Differentiating the log-density shows that u_t is proportional to the conditional score, $\partial \ln f_t / \partial \mu_{t|t-1} = (v + 1)v^{-1} \exp(-2\lambda)u_t$. For low degrees of freedom, the score function is such that observations that would be seen as outliers for a Gaussian distribution are far less influential. As $|y| \rightarrow \infty$, the response tends to zero. Redescending M-estimators, which feature in the robustness literature, have the same property. For example, Tukey’s biweight function is $\psi(z) = [1 - (x/4.685)^2]_+^2 z$, where $[\]_+$ denotes the positive part of the term in $[\]$. This function implements ‘soft trimming’, as opposed to metric trimming, where $\psi(z) = z$ for $|z| \leq K$ and is zero thereafter. The t score function is even softer.

The variable u_t can be written $u_t = (1 - b_t)(y_t - \mu_{t|t-1})$, where

$$b_t = \frac{(y_t - \mu_{t|t-1})^2 / v \exp(2\lambda)}{1 + (y_t - \mu_{t|t-1})^2 / v \exp(2\lambda)}, \quad 0 \leq b_t \leq 1, \quad 0 < v < \infty, \quad (16)$$

is distributed as $beta(1/2, v/2)$; see Harvey (2013, Chap. 3). Because the u_t ’s are $IID(0, \sigma_u^2)$, $\mu_{t|t-1}$ is weakly and strictly stationary so long as $|\phi| < 1$. Although determining the statistical properties of $\mu_{t|t-1}$ requires assuming that it started in the infinite past, the filter needs to be initialized in practice and this may be done by setting $\mu_{1|0} = \omega$.

All moments of u_t exist and the existence of moments of y_t is not affected by the dynamics. The autocorrelations can be found from the infinite MA representation; the patterns are as they would be for a Gaussian model.

Maximum likelihood estimation is straightforward and for a first-order dynamic equation, as in (13), an analytic expression for the information matrix is available.

There are a number of ways in which skewness may be introduced into a t-distribution. One possibility is the method proposed by Fernandez and Steel (1998). There is a minor technical issue in that the score is not differentiable at the mode but as, Zhu and Galbraith (2011) show, the asymptotic theory for the ML estimator still goes through in the usual way. The asymptotic theory for the DCS skew-t location model also goes through; see Harvey (2013, Sect 3.11).

3.2 Exponential generalized beta distribution model

The exponential generalized beta distribution of the second kind (EGB2) is obtained by taking the logarithm of a variable with a generalized beta distribution of the second kind (GB2) distribution. What was the logarithm of scale in GB2 now becomes location in EGB. A reparameterization in terms of the standard deviation, σ , gives the PDF as

$$f(y; \mu, \sigma, \xi, \varsigma) = \frac{h \exp\{\xi h(y - \mu)/\sigma\}}{\sigma B(\xi, \varsigma) (1 + \exp\{h(y - \mu)/\sigma\})^{\xi + \varsigma}},$$

where ξ and ς are shape parameters which determine skewness and kurtosis and $h = \sqrt{\psi'(\xi) + \psi'(\varsigma)}$. The EGB2 distribution has exponential tails and all moments exist. When $\xi = \varsigma$, the distribution is symmetric; for $\xi = \varsigma = 1$ it is a logistic

distribution and when $\xi = \varsigma \rightarrow \infty$ it tends to a normal distribution. When $\xi = \varsigma = 0$ the distribution is double exponential or Laplace. The distribution is positively (negatively) skewed when $\xi > \varsigma$ ($\xi < \varsigma$) and its kurtosis decreases as ξ and ς increase. Skewness ranges between -2 and 2 and kurtosis lies between 3 and 9 . There is excess kurtosis for finite ξ and/or ς . The EGB2 distribution is more peaked than a t with the same kurtosis, but as kurtosis increases, the differences between the peaks become more marked.

Multiplying the score function with respect to location by the variance gives

$$u_t = \sigma^2 \frac{\partial \ln f_t}{\partial \mu_{t|t-1}} = \sigma h [(\xi + \varsigma)b_t(\xi, \varsigma) - \xi], \tag{17}$$

where

$$b_t(\xi, \varsigma) = \frac{e^{(y_t - \mu_{t|t-1})h/\sigma}}{e^{(y_t - \mu_{t|t-1})h/\sigma} + 1}.$$

Because $0 \leq b_t(\xi, \varsigma) \leq 1$, it follows that as $y \rightarrow \infty$, the score approaches an upper bound of $\sigma\sqrt{2}$, whereas $y \rightarrow -\infty$ gives a lower bound of $-\sigma\sqrt{2}$. As $\varsigma = \xi \rightarrow \infty$, the distribution becomes normal with $u_t \simeq y_t - \mu_{t|t-1}$.

Figure 1 shows the score functions for EGB2 and t distributions with a standard deviation of one and an excess kurtosis of two. The shape parameters for the two distributions are $\xi = 0.5$ and $\nu = 7$. As can be seen, the score for the t -distribution is re-descending, reflecting the fact that it has fat tails, whereas the EGB2 score is bounded.

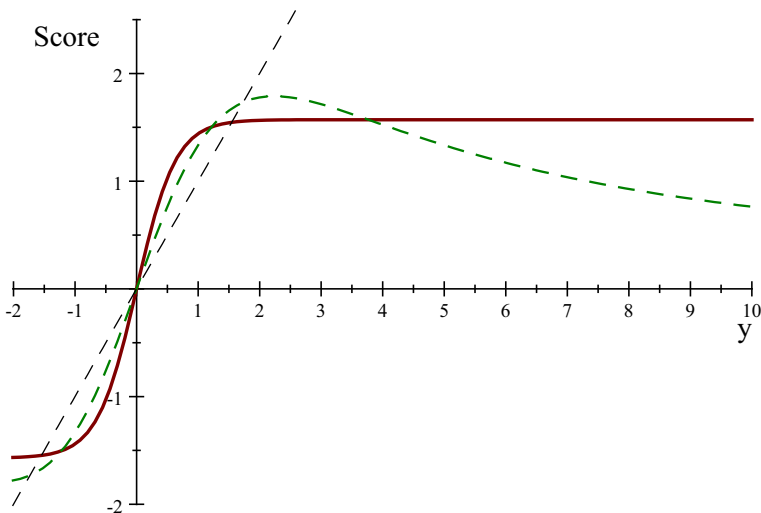


Fig. 1 Score functions for t and EGB2 (solid line) with excess kurtosis of two, together with the (linear) score for a normal distribution

4 Macroeconomic time series

Dynamic location models were fitted to the growth rate of US GDP and industrial production using EGB2, Student’s *t* and normal distributions. GDP is quarterly, ranging from 1947q1 to 2012q4. Industrial production data are monthly and range from January 1960 to February 2013. All data are seasonally adjusted and taken from the Federal Reserve Economic Data (FRED) database of the Federal Reserve of St. Louis.

When first-order Gaussian models are fitted, there is little indication of residual serial correlation. There is excess kurtosis in all cases, but no evidence of asymmetry. For example, with GDP the Bowman-Shenton statistic is 30.04, which is clearly significant because the distribution under the null hypothesis of Gaussianity is χ^2_2 . The non-normality clearly comes from excess kurtosis, which is 1.9, rather than from skewness, which is only 0.18 (with a *p* value of 0.24). Comparing the residuals with a fitted normal shows them to have a higher peak at the mean, as well as heavier tails; see Fig. 2.

Tables 1 and 2 report the estimation results and Table 3 compares goodness of fit. The Student-*t* model and the EGB2 outperform the Gaussian model with the shape parameter, ν or ξ , confirming the excess kurtosis. The λ parameter is the logarithm of scale, ν , but the estimates of σ are shown because these are comparable across different distributions. For GDP the EGB2 models gives a slightly better fit, whereas the *t*-distribution is better for industrial production. However, the differences between the two are small compared with the Gaussian model.

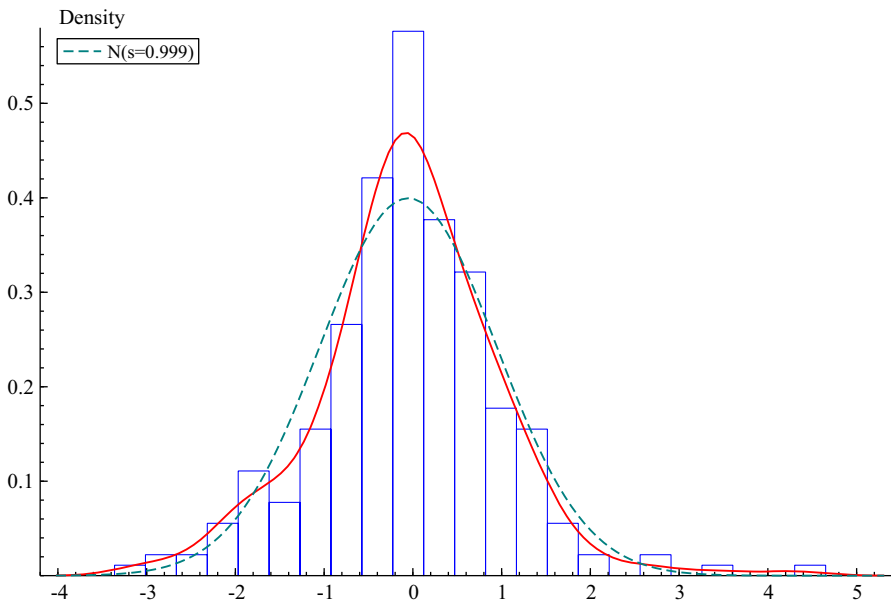


Fig. 2 Residuals from fitting a first-order Gaussian model to GDP

Table 1 US GDP (quarterly)

	κ	ϕ	ω	λ	ξ (or ν)	σ
EGB2	0.30	0.50	0.008	-5.40	0.88	0.0091
Num SE	(0.063)	(0.103)	(0.001)	(0.324)	(0.394)	
Asy SE	(0.054)	(0.143)	(0.001)	(0.415)	(0.502)	
t	0.50	0.50	0.008	-4.88	6.49	0.0091
Num SE	(0.094)	(0.103)	(0.001)	(0.071)	(2.364)	
Asy SE	(0.089)	(0.141)	(0.001)	(0.056)	(1.887)	
Gaussian	0.35	0.49	0.008	-4.70	—	0.0091
Num SE	(0.058)	(0.112)	(0.001)	(0.044)		
Asy SE	(0.061)	(0.141)	(0.001)	(0.044)		

Table 2 US Industrial production (monthly)

	κ	ϕ	ω	λ	ξ (or ν)	σ
EGB2	0.20	0.85	0.003	-6.05	0.55	0.0069
Num SE	(0.033)	(0.036)	(0.001)	(0.214)	(0.147)	
Asy SE	(0.027)	(0.040)	(0.001)	(0.287)	(0.196)	
t	0.40	0.85	0.002	-5.25	4.49	0.0071
Num SE	(0.060)	(0.036)	(0.001)	(0.046)	(0.743)	
Asy SE	(0.055)	(0.040)	(0.001)	(0.038)	(0.634)	
Gaussian	0.25	0.83	0.002	-4.95	—	0.0071
Num SE	(0.032)	(0.041)	(0.001)	(0.028)		
Asy SE	(0.035)	(0.046)	(0.001)	(0.028)		

Table 3 US macroeconomic series—Model comparison

	Log-Likelihood	AIC	BIC
GDP			
EGB2	868.376	-6.566	-6.498
t	868.242	-6.565	-6.497
Gaussian	862.212	-6.526	-6.472
Industrial production			
EGB2	2291.66	-7.168	-7.133
t	2293.56	-7.174	-7.139
Gaussian	2255.21	-7.057	-7.029

5 Structural breaks

It might be thought that the EGB2 and t filters will be less responsive to a permanent change in the level than the linear Gaussian filter. However, for moderate size shifts,

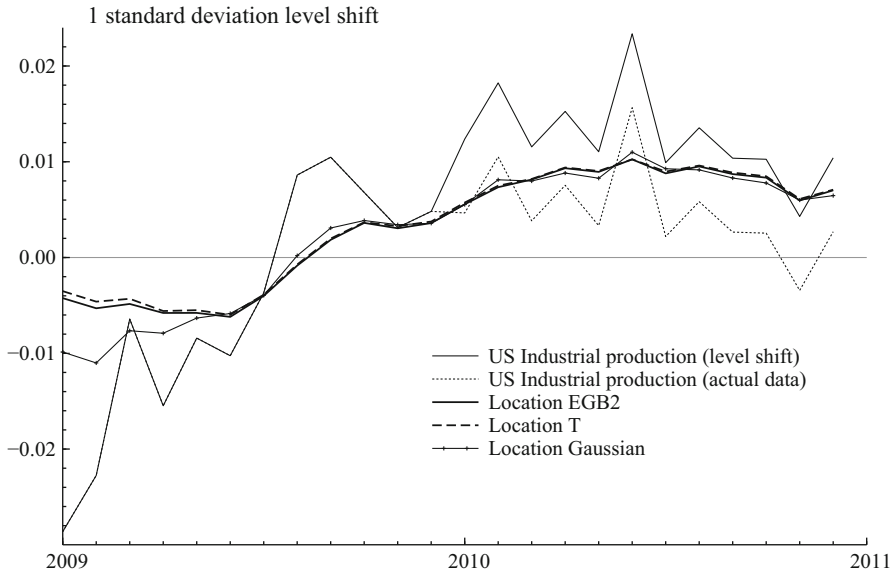


Fig. 3 Response of Gaussian, EGB2 and t-filters to a one SD shift in level in January 2010

the score functions in Fig. 1 suggest that this might not be the case, because only for large observations is the Gaussian response bigger than the response of the robust filters. For example, for the logistic (EGB2 with unit shape parameters), the score is only smaller than the observation (and hence the linear filter) when it is more than (approximately) 1.6 standard deviations from the mean. The behaviour of the t-filter is similar.

In order to investigate the issue of adapting to a permanent change in level, an upward shift was added to the US industrial production data at the beginning of 2010. The size of the shift was calibrated so as to be proportional to the sample standard deviation of the series. The results are shown in the Figs. 3 and 4. For a one standard deviation shift the paths of all three filters after the break are similar. For two standard deviations the Gaussian filter adapts more quickly, but it is still well below the new level initially and after 9 or 10 months it is virtually indistinguishable from the other two filters.

6 Trend and seasonality

The simplest nonstationary unobserved components model is the Gaussian local level model, that is a random walk plus noise. In this case the Kalman filter is an exponentially weighted moving average (EWMA) of past observations. For the DCS-t filter

$$y_t = \mu_{t|t-1} + v_t,$$

$$\mu_{t+1|t} = \mu_{t|t-1} + \kappa u_t.$$

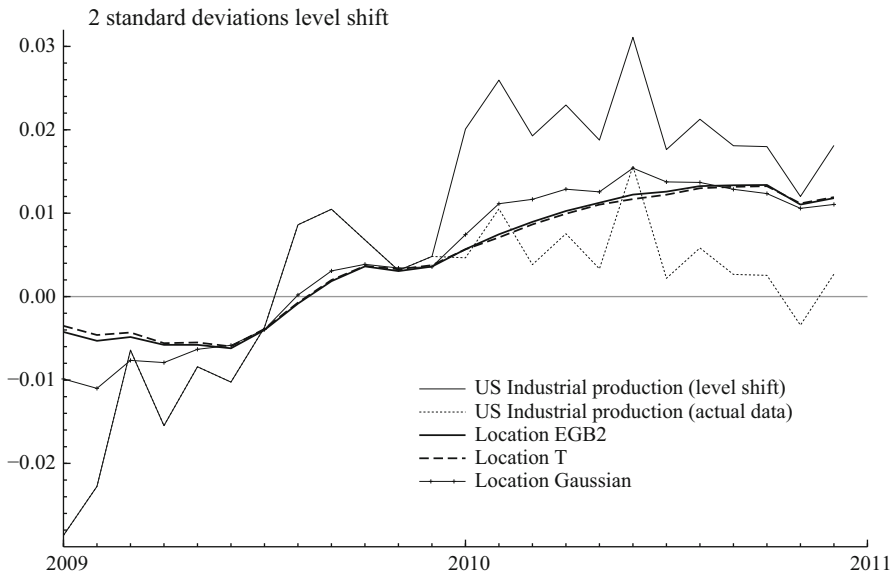


Fig. 4 Response of Gaussian, EGB2 and t-filters to a two SD shift in level in January 2010

and the initial value, $\mu_{1|0}$, is treated as an unknown parameter that needs to be estimated along with κ and ν . Since $u_t = (1 - b_t)(y_t - \mu_{t|t-1})$, re-arranging the dynamic equation gives

$$\mu_{t+1|t} = (1 - \kappa(1 - b_t))\mu_{t|t-1} + \kappa(1 - b_t)y_t,$$

which can be regarded as an EWMA modified to deal with outliers: when y_t^2 is large, b_t is close to one.

Generalizing the above model to include a slope and seasonals provides the basis for a robust treatment of seasonal adjustment.

6.1 Basic structural model

Stochastic trend and seasonal components may be introduced into Gaussian unobserved components models for location. These models, called structural time series models, are implemented in the STAMP package of [Koopman et al. \(2009\)](#). The BSM is made up of mutually independent trend, seasonal and irregular components. Thus

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \quad t = 1, \dots, T,$$

where μ_t is a local linear trend with a stochastic level and slope and γ_t is a stochastic seasonal. The trend is

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2), \\ \beta_t &= \beta_{t-1} + \zeta_t, \quad \zeta_t \sim NID(0, \sigma_\zeta^2), \end{aligned}$$

where ζ_t and η_t are independent of each other and of ε_t . The random walk plus noise or local level model is a special case. As regards the seasonal component, let γ_{jt} denote the effect of season j at time t and define $\gamma_t = (\gamma_{1t}, \dots, \gamma_{st})'$. The full set of seasonals evolves as a multivariate random walk

$$\gamma_t = \gamma_{t-1} + \omega_t, \quad t = 1, \dots, T,$$

where $\omega_t = (\omega_{1t}, \dots, \omega_{st})'$ is a zero mean disturbance with $Var(\omega_t) = \sigma_\omega^2 (\mathbf{I} - s^{-1}\mathbf{1}\mathbf{1}')$, with $\sigma_\omega^2 > 0$. Although all s seasonal components are continually changing, only one affects the observations at any particular time, that is $\gamma_t = \gamma_{jt}$ when season j is prevailing at time t . The requirement that the seasonal components always sum to zero is enforced by the restriction that the disturbances sum to zero at each t . This restriction is implemented by the correlation structure of ω_t , where $Var(\mathbf{1}'\omega_t) = 0$, coupled with initial conditions constraining the seasonals to sum to zero at $t = 0$.

The multiplicative seasonal ARIMA model known as the ‘airline model’ is

$$\Delta \Delta_s y_t = (1 + \theta L)(1 + \Theta L^s) \xi_t, \quad \xi_t \sim NID(0, \sigma_\xi^2)$$

Box and Jenkins (1976, pp. 305–6) gave a rationale for this model in terms of EWMA’s at monthly and yearly intervals. The reduced form of the BSM is $\Delta \Delta_s y_t \sim MA(s + 1)$. Maravall (1985), compared the autocorrelation functions of $\Delta \Delta_s y_t$ for the BSM and airline model for some typical values of the parameters and finds them to be quite similar, particularly when the seasonal MA parameter, Θ , is close to minus one. In the limiting case when Θ is equal to minus one, the airline model is equivalent to a BSM with $\sigma_\zeta^2 = \sigma_\omega^2 = 0$.

6.2 Stochastic trend and seasonal in the DCS model

The DCS model for trends and seasonals,

$$y_t = \mu_{t|t-1} + \gamma_{t|t-1} + v_t, \quad t = 1, \dots, T,$$

has a structure which is similar to that of the innovations form of the Kalman filter for the BSM.

The filter for the trend is

$$\begin{aligned} \mu_{t+1|t} &= \mu_{t|t-1} + \beta_{t|t-1} + \kappa_1 u_t \\ \beta_{t+1|t} &= \beta_{t|t-1} + \kappa_2 u_t. \end{aligned}$$

An integrated random walk (IRW) trend in the UC local linear trend model implies the constraint $\kappa_2 = \kappa_1^2 / (2 - \kappa_1)$, $0 < \kappa_1 < 1$, which may be found from Harvey

(1989, p. 177). The restriction can be imposed on the DCS- t model by treating $\kappa_1 = \kappa$ as the unknown parameter, but without unity imposed as an upper bound.

The filter for the seasonal is

$$\gamma_{t|t-1} = \mathbf{z}'_t \gamma_{t|t-1}, \quad \gamma_{t+1|t} = \gamma_{t|t-1} + \kappa_t u_t,$$

where the $s \times 1$ vector \mathbf{z}_t picks out the current season from the vector $\gamma_{t|t-1}$. If κ_{jt} , $j = 1, \dots, s$, denotes the j th element of κ_t , then in season j we set $\kappa_{jt} = \kappa_s$, where κ_s is a non-negative unknown parameter, whereas $\kappa_{it} = -\kappa_s/(s-1)$, $i \neq j$, $i = 1, \dots, s$. The amounts by which the seasonal effects change therefore sum to zero. The initial conditions at time $t = 0$ are estimated by treating them as parameters.

The above filter may be regarded as a robust version of the well-known Holt-Winters filter; see Harvey (1989, p. 31). However, it differs from Holt-Winters in the Gaussian case by enforcing the restriction that the seasonals sum to zero. This is an important advantage.

6.3 Seasonal adjustment

In contrast to the Gaussian BSM, the DCS model has no exact solution for smoothing. Some possibilities are suggested in Harvey (2013, Sect 3.7), but these are difficult to generalize beyond the local level model. The best way to employ the DCS model for seasonal adjustment is to use it to mitigate the effects of outliers by modifying them rather than eliminating them by dummy variables. A dummy variable effectively means that the corresponding observation is treated as though it were missing; in other words it corresponds to hard trimming.

We first fit the DCS model and use it to construct pseudo-observations from the signal, that is $\mu^s_{t|t-1} = \mu_{t|t-1} + \gamma_{t|t-1}$. Thus

$$y_t \left(\mu^s_{t|t-1} \right) = \mu^s_{t|t-1} + u_t, \quad t = 1, \dots, T. \quad (18)$$

For the DCS- t model we can write

$$y_t \left(\mu^s_{t|t-1} \right) = (1 - b_t) y_t + b_t \mu^s_{t|t-1}, \quad t = 1, \dots, T,$$

where b_t is as in (16) with $\mu_{t|t-1}$ replaced by $\mu^s_{t|t-1}$. In the EGB2, we work directly with (18).

The pseudo-observations are now used to estimate the parameters in a BSM and the signal is estimated by smoothing. This new signal, denoted $\mu^s_{t|T}$, is then used to construct new pseudo-observations as

$$y_t \left(\mu^s_{t|T} \right) = \mu^s_{t|T} + u_t, \quad t = 1, \dots, T, \quad (19)$$

with b_t replaced by

$$b_{t|T} = b_{t|T}(\mu_{t|T}^s) = \frac{(y_t - \mu_{t|T}^s)^2 / v \exp(2\lambda)}{1 + (y_t - \mu_{t|T}^s)^2 / v \exp(2\lambda)}$$

for the t-distribution and by

$$b_{t|T} = \frac{\exp\left\{\left(y_t - \mu_{t|T}^s\right) / h / \sigma\right\}}{1 + \exp\left\{\left(y_t - \mu_{t|T}^s\right) h / \sigma\right\}}$$

for the EGB2. The BSM may be re-estimated and the whole process iterated until convergence. If desired, λ and σ can be updated at each step using the sample variance of the u_t^s s. The relationship between variance of the u_t^s s and λ for the t-distribution is given in Harvey (2013, p. 62). For the EGB2 (with $\xi = \varsigma$) it follows from Caivano and Harvey (2014) that $\sigma_u^2 = \sigma^2 h^2 \xi^2 / (2\xi + 1)$.

The seasonally adjusted observations are constructed as

$$y_t^s = y_t - \gamma_{t|T} = \mu_{t|T} + v_{t|T}, \quad t = 1, \dots, T, \tag{20}$$

where $v_{t|T} = y_t - \mu_{t|T}^s$. If it is felt that the outliers should be modified to make them less extreme, we could let $y_t^s = \mu_{t|T} + u_{t|T}$.

The above procedure could be implemented with TRAMO-SEATS rather than the unobserved components BSM.

7 Tourists in Spain

The logarithm of the number of tourists entering Spain from January 2000 to April 2014 (source: Frontur) is plotted in Fig. 5. Comparing a Gaussian unobserved components model with dummy variables with a DCS-t model provides a contrast between hard and soft trimming.

Fitting a BSM gives a Bowman-Shenton statistic of 19.29 for the residuals so normality is clearly rejected. (The 1% critical value for a χ_2^2 is 9.21.) The automatic outlier detection option in STAMP finds five outliers, three of which are in 20012, along with a small structural break which, although not sudden, is attributed to April 2008. The corresponding (edited) output is shown below. Fitting a BSM with dummy variables included for the outliers and break gives the smoothed components shown in Fig. 5.

Date	Estimate	SE	't-stat'
Outlier 2002(3)	0.14295	0.02818	5.07283
Outlier 2002(4)	-0.14438	0.02824	-5.11205
Outlier 2002(8)	0.10473	0.02775	3.77429

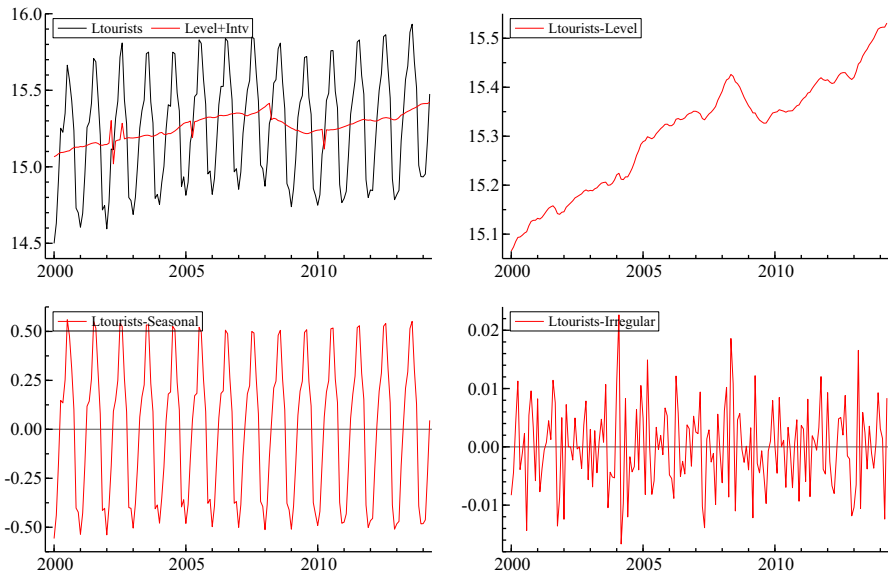


Fig. 5 Smoothed estimates of components from BSM with outliers treated by dummy variables

```

Outlier 2005 (4)      -0.10592  0.02766  -3.82885
Outlier 2010 (4)     -0.12736  0.02767  -4.60211

Level break 2008 (4) -0.10926  0.02429  -4.49895
    
```

The ML estimates for the parameters of a DCS-t random walk model (the drift was not significant) with a seasonal component are as follows:

$$\begin{aligned}
 \tilde{\kappa} &= 0.4906(0.0666) & \tilde{\kappa}_s &= 1.0068(0.1172) \\
 \tilde{\nu} &= 6.006(1.5678) & \tilde{\lambda} &= -3.2573(0.0612)
 \end{aligned}$$

with initial values $\tilde{\mu}_0 = 15.1018(0.0174)$ and $\tilde{\gamma}_0 = [-0.5795(0.0270), -0.4761(0.0269), -0.1780(0.0264), 0.1733(0.0244), 0.1475(0.0291), 0.2648(0.0261), 0.5791(0.0259), 0.4566(0.0209), 0.3157(0.0327), 0.1167(0.0304), -0.3862(0.0289), -0.4339]$, for the seasonal factors. The figures in parentheses are numerical standard errors and the initial value for the last seasonal factor is $\tilde{\gamma}_{0,12} = 0.018$ which is constructed by from the others and thus has no standard error.

The filtered DCS-t trend shown in Fig. 6 appears not to be affected by the outliers, which are downweighted, and the trend, shown in the two top panels, adjusts to the fall in 2008 as quickly as the Gaussian filter from the outlier-adjusted BSM which is fitted with a dummy for a break in 2008(4). The figure also shows the seasonal component and contrasts the prediction errors (irregular component) and scores; it is evident that extreme observations are downweighted by the score. The autocorrelation functions

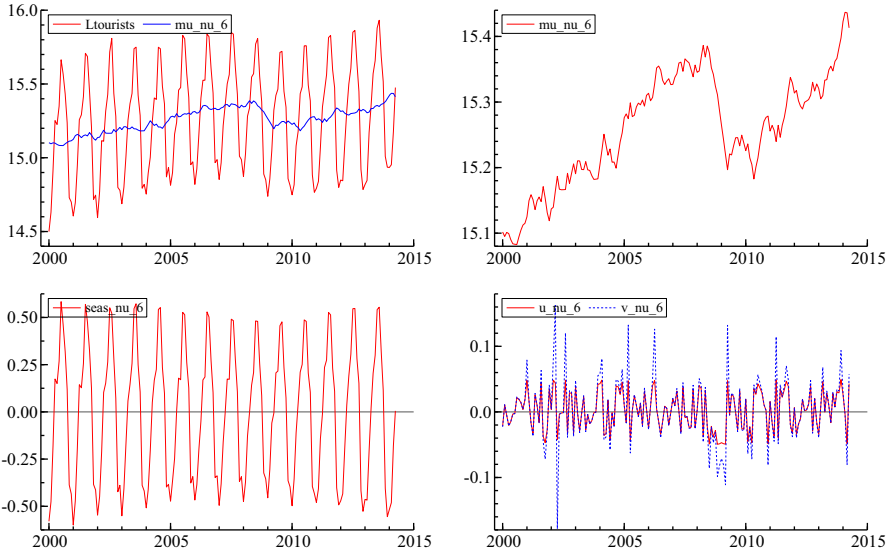


Fig. 6 Filtered estimates from DCS-t filter

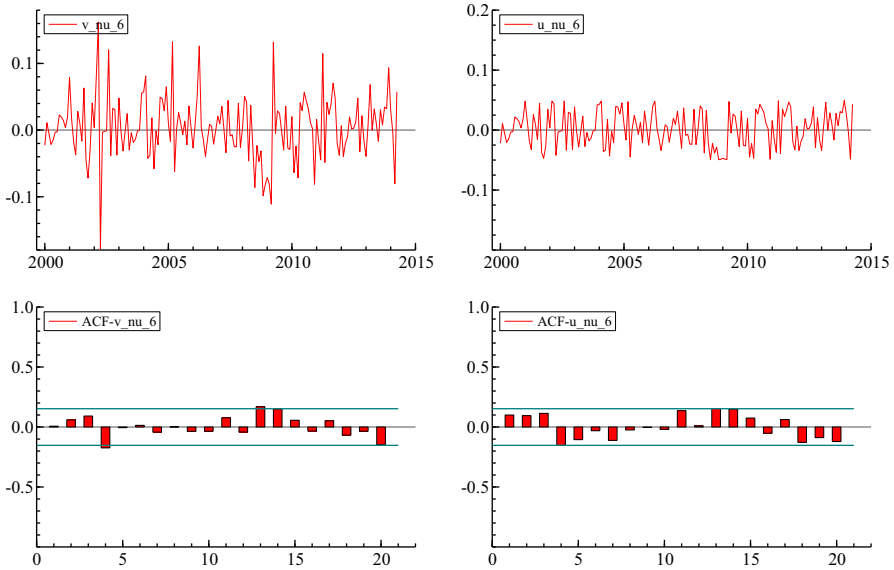


Fig. 7 Residuals and scores, together with ACFs

of irregular and score are shown in Fig. 7. The autocorrelations of the scores tend to be slightly bigger in absolute value, reflecting the fact that they are not weakened by outliers in the same way as the raw predictions errors.

Figure 8 shows the filtered trend together with the smoothed trends for the first three iterations, obtained as described in Sect. 6.3. The trend at the second iteration is almost indistinguishable from the one obtained at the third.

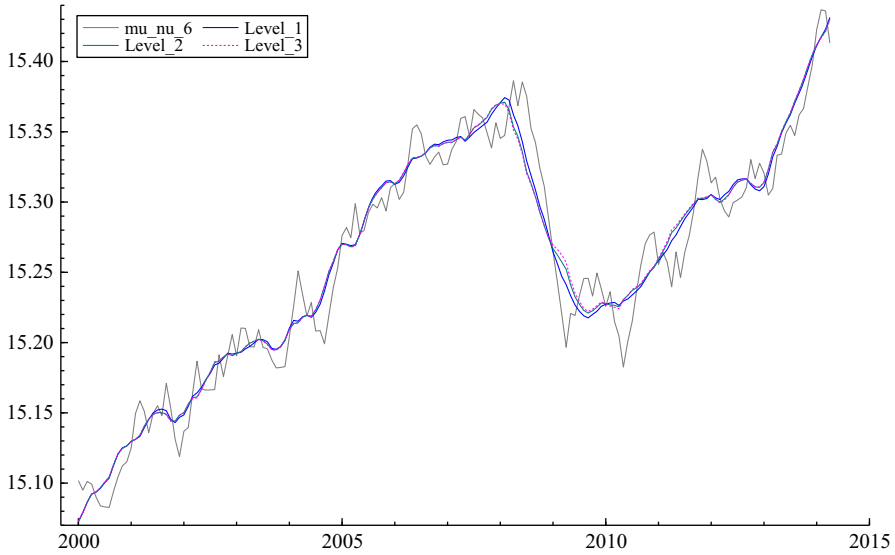


Fig. 8 Smoothed trends from DCS-t model

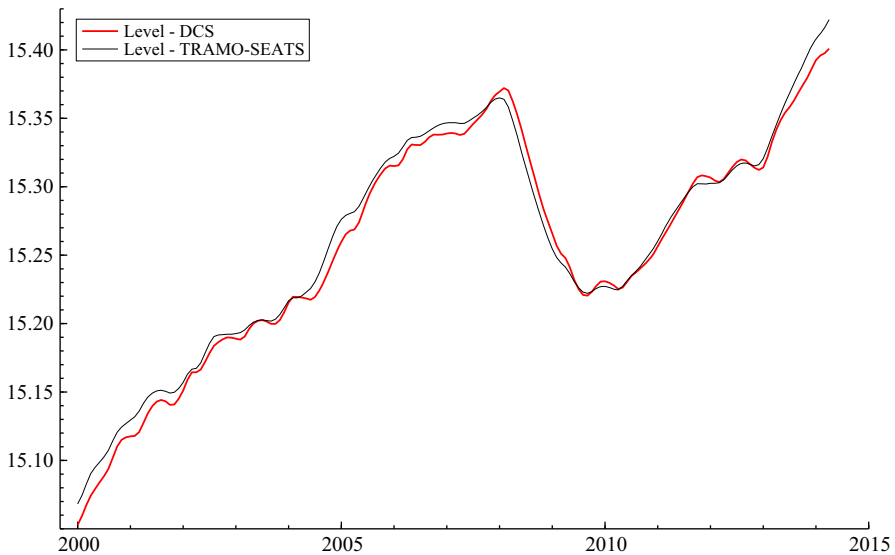


Fig. 9 TRAMO-SEATS trend and TRAMO-SEATS-DCS trend

The above procedure was repeated, but using TRAMO-SEATS to obtain a smoothed estimate of the trend from the DCS pseudo-observations. The result is shown in Fig. 9, together with the smoothed trend obtained from the full (automatic) TRAMO-SEATS procedure. As can be seen, the two trends are very close. Figure 10 shows the corresponding smoothed irregular components.

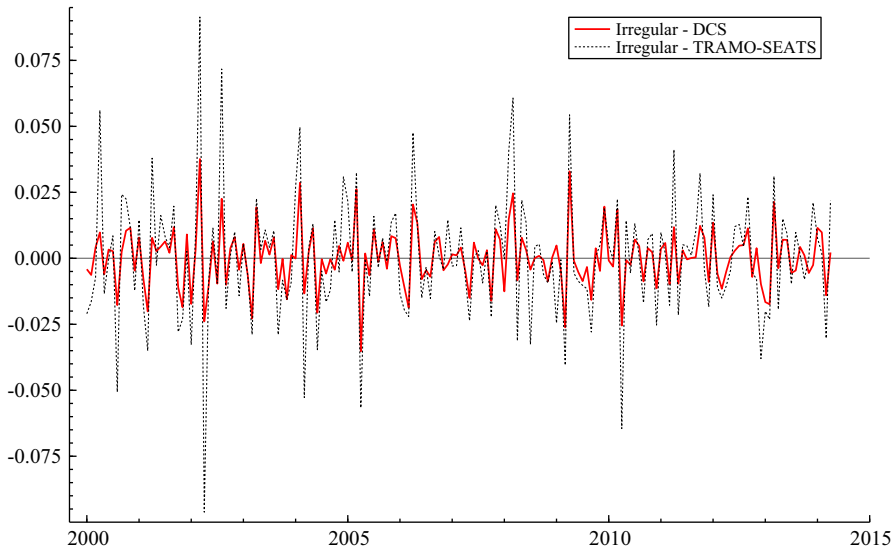


Fig. 10 Irregular components for TRAMO-SEATS and TRAMO-SEATS-DCS

We also fitted an EGB2 model, but the fit was not as good as for the DCS-t model. Generally the results were similar except that the seasonal pattern for EGB2 was found to be deterministic. The ML estimates for the asymmetric model were:

$$\begin{aligned} \tilde{\kappa} &= 0.32(0.07) & \tilde{\kappa}_s &= 0 \\ \tilde{\xi} &= 0.74(0.55) & \tilde{\zeta} &= 0.66(0.55) & \tilde{\lambda} &= -3.03(0.07) \end{aligned}$$

The two shape parameters are very close and there is virtually no loss in imposing symmetry, that is $\xi = \zeta$.

8 Conclusions

This article has shown how DCS models with changing location and/or scale can be successfully extended to cover EGB2 conditional distributions. Most of the theoretical results on the properties of DCS-t models, including the asymptotic distribution of ML estimators, carry over to EGB2 models. However, whereas the t-distribution has fat-tails, and hence subjects extreme observations to a form of soft trimming, the EGB2 distribution has light tails (but excess kurtosis) and hence gives a gentle form of Winsorizing. The examples show that the EGB2 distribution can give a better fit to some macroeconomic series.

The way in which DCS models respond to breaks was examined and it was shown that, contrary to what might be expected, they adjust almost as rapidly as Gaussian models.

A seasonal adjustment procedure may be carried out with DCS models that include both trend and seasonal components. Having fitted the DCS model, the scores are used

to adjust the data before smoothing using a standard Gaussian model. Two or three iterations seem to be sufficient. The method was illustrated with data on tourists in Spain. There is a case for a structural break in April, 2008, but the DCS model quickly adjusts to it and indeed it could be reasonably argued that it is better to let the change take place over several months rather than assigning it to just one. In summary, our new DCS procedure, like TRAMO-SEATS, provides a practical approach to seasonal adjustment in the presence of outliers.

Acknowledgments We are grateful to the Gabriele Fiorentini and a referee for helpful comments on the first draft.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Caivano M, Harvey AC (2014) Time series models with an EGB2 conditional distribution. *J Time Ser Anal* 34:558–571
- Creal D, Koopman SJ, Lucas A (2013) Generalized autoregressive score models with applications. *J Appl Econ* 28:777–795
- Creal D, Koopman SJ, Lucas A (2011) A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *J Bus Econ Stat* 29:552–563
- Durbin J, Koopman SJ (2012) Time series analysis by state space methods, 2nd edn. Oxford Statistical Science Series, Oxford
- Embrechts P, Kluppelberg C, Mikosch T (1997) Modelling extremal events. Springer, Berlin
- Fernandez C, Steel MFJ (1998) On Bayesian modeling of fat tails and skewness. *J Am Stat Assoc* 99:359–371
- Harvey AC (1989) Forecasting, structural time series models and the kalman filter. Cambridge University Press, Cambridge
- Harvey AC (2013) Dynamic models for volatility and heavy tails., Econometric Society Monograph Cambridge University Press, Cambridge, New York
- Harvey AC, Luati A (2014) Filtering with heavy tails. *J Am Stat Assoc* 109:1112–1122
- Koopman SJ, Harvey AC, Doornik JA, Shephard N (2009) STAMP 8.2 structural time series analysis modeller and predictor. Timberlake Consultants Ltd, London
- Maravall A (1985) On structural time series models and the characterization of components. *J Bus Econ Stat* 3:350–355
- Maronna R, Martin D, Yohai V (2006) Robust statistics: theory and methods. Wiley, Chichester
- McDonald JB, Newey WK (1988) Partially adaptive estimation of regression models via the generalized t distribution. *Econ Theory* 4:428–457
- Muler N, Pena D, Yohai VJ (2009) Robust estimation for ARMA models. *Ann Stat* 37:816–840
- Zhu D, Galbraith JW (2011) Modelling and forecasting expected shortfall with the generalised asymmetric student-t and asymmetric exponential power distributions. *Journal of Empir Financ* 18:765–778