Filtering with heavy tails

Andrew Harvey (ach34@cam.ac.uk) and Alessandra Luati

Faculty of Economics, University of Cambridge and and Dept. of Statistics, Bologna University

January 2013
http://www.econ.cam.ac.uk/dae/repec/cam/pdf/cwpe1255.pdf

Introduction to dynamic conditional score (DCS) models

A guiding principle is **signal extraction**. When combined with basic ideas of maximum likelihood estimation, the signal extraction approach leads to models which, in contrast to many in the literature, are relatively simple in their form and yield analytic expressions for their principal features.
For estimating location, DCS models are closely related to the unobserved components (UC) models described in Harvey (1989). Such models can be handled using state space methods and they are easily accessible using the STAMP package of Koopman et al (2008).
For estimating scale, the models are close to stochastic volatility (SV) models, where the variance is treated as an unobserved component.

Unobserved component models

A simple Gaussian signal plus noise model is

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim NID\left(0, \sigma_\varepsilon^2\right), \quad t = 1, ..., T$$

$$\mu_{t+1} = \phi\mu_t + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2),$$

where the irregular and level disturbances, ε_t and η_t, are mutually independent. The AR 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) is an ARMA(1,1) process

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

but with restrictions on θ. For example, when $\phi = 1$, $0 \leq \theta \leq 1$. The forecasts from the UC model and RF are the same.

Unobserved component models

The UC model is effectively in state space form (SSF) and, as such, it may be handled by the Kalman filter (KF). The parameters ϕ and q can be estimated by ML, with the likelihood function constructed from the one-step ahead prediction errors.
The KF can be expressed as a single equation. Writing this equation together with an equation for the one-step ahead prediction error, v_t, gives the innovations form (IF) of the KF:

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

The Kalman gain, k_t, depends on ϕ and q.
In the steady-state, k_t is constant. Setting it equal to κ and re-arranging gives the **ARMA(1,1)** model with $\xi_t = v_t$ and $\phi - \kappa = \theta$.

Outliers

Suppose noise is from a heavy tailed distribution, such as Student's t. Outliers.

The RF is still an ARMA(1,1), but allowing the $\tilde{\zeta}_t's$ to have a heavy-tailed distribution does not deal with the problem as a large observation becomes incorporated into the level and takes time to work through the system.

An ARMA models with a heavy-tailed distribution is designed to handle *innovations outliers*, as opposed to *additive outliers*. See the **robustness** literature.

But a *model-based approach* is not only simpler than the usual robust methods, but is also more amenable to diagnostic checking and generalization.

See Lange et al (JASA, 1989) for robustification with the t-distribution.

Unobserved component models for non-Gaussian noise

Simulation methods, such as MCMC, provide the basis for a direct attack on models that are nonlinear and/or non-Gaussian. The aim is to extend the Kalman filtering and smoothing algorithms that have proved so effective in handling linear Gaussian models. Considerable progress has been made in recent years; see Durbin and Koopman (2012).

But simulation-based estimation can be time-consuming and subject to a degree of uncertainty.

Also the statistical properties of the estimators are not easy to establish.

Observation driven model based on the score

The DCS approach begins by writing down the distribution of the $t-th$ observation, conditional on past observations. Time-varying parameters are then updated by a suitably defined filter. Such a model is *observation driven*, as opposed to a UC model which is *parameter driven*. In a *linear Gaussian UC* model, the KF is driven by the one step-ahead prediction error, v_t. The DCS filter replaces v_t in the KF equation by a variable, u_t, that is proportional to the score of the conditional distribution.
The innovations form becomes

$$\begin{aligned} y_t &= \mu_{t|t-1} + v_t, \qquad\qquad t = 1, ..., T \\ \mu_{t+1|t} &= \phi\mu_{t|t-1} + \kappa u_t \end{aligned}$$

where κ is an unknown parameter.

Dynamic location model

$$\begin{aligned} y_t &= \omega + \mu_{t|t-1} + v_t = \omega + \mu_{t|t-1} + \exp(\lambda)\varepsilon_t, \\ \mu_{t+1|t} &= \phi\mu_{t|t-1} + \kappa u_t, \end{aligned}$$

where ε_t is serially independent, standard t-variate and the conditional score is

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

where $v_t = y_t - \mu_{t|t-1}$ is the prediction error and $\varphi = \exp(\lambda)$ is the (time-invariant) scale.

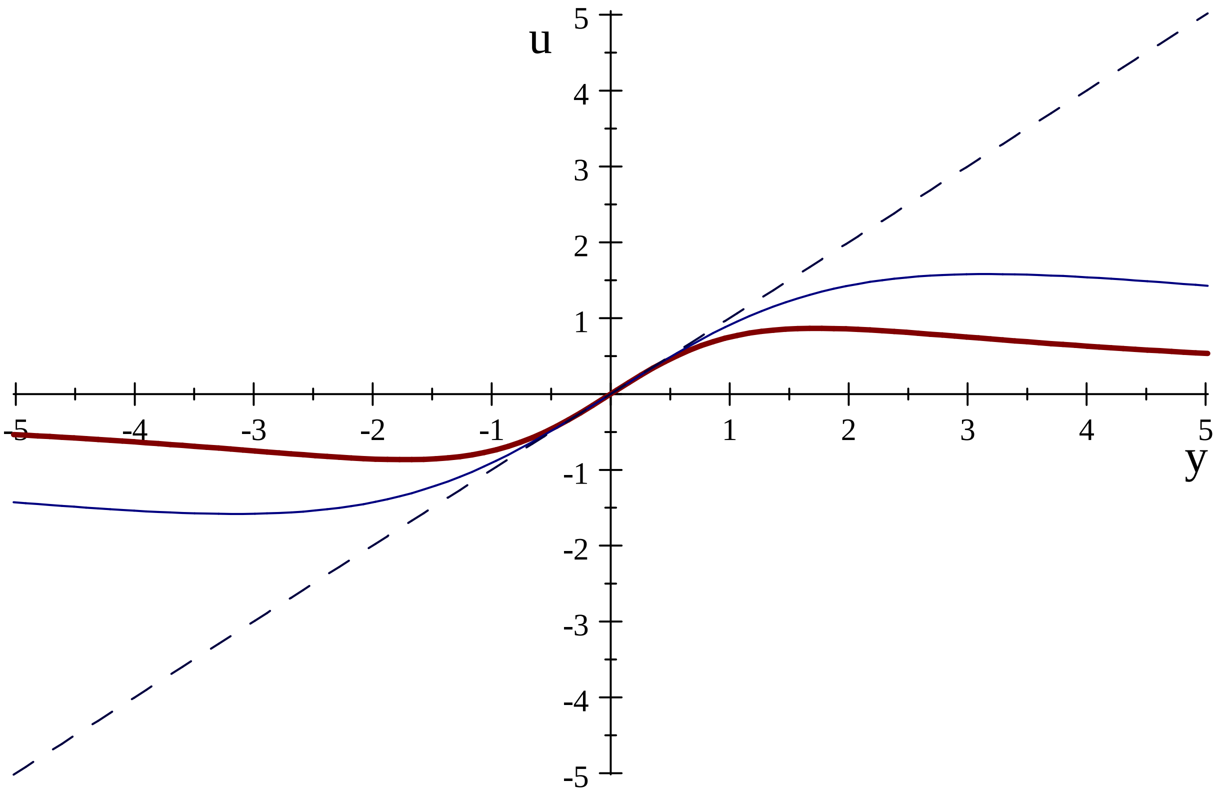
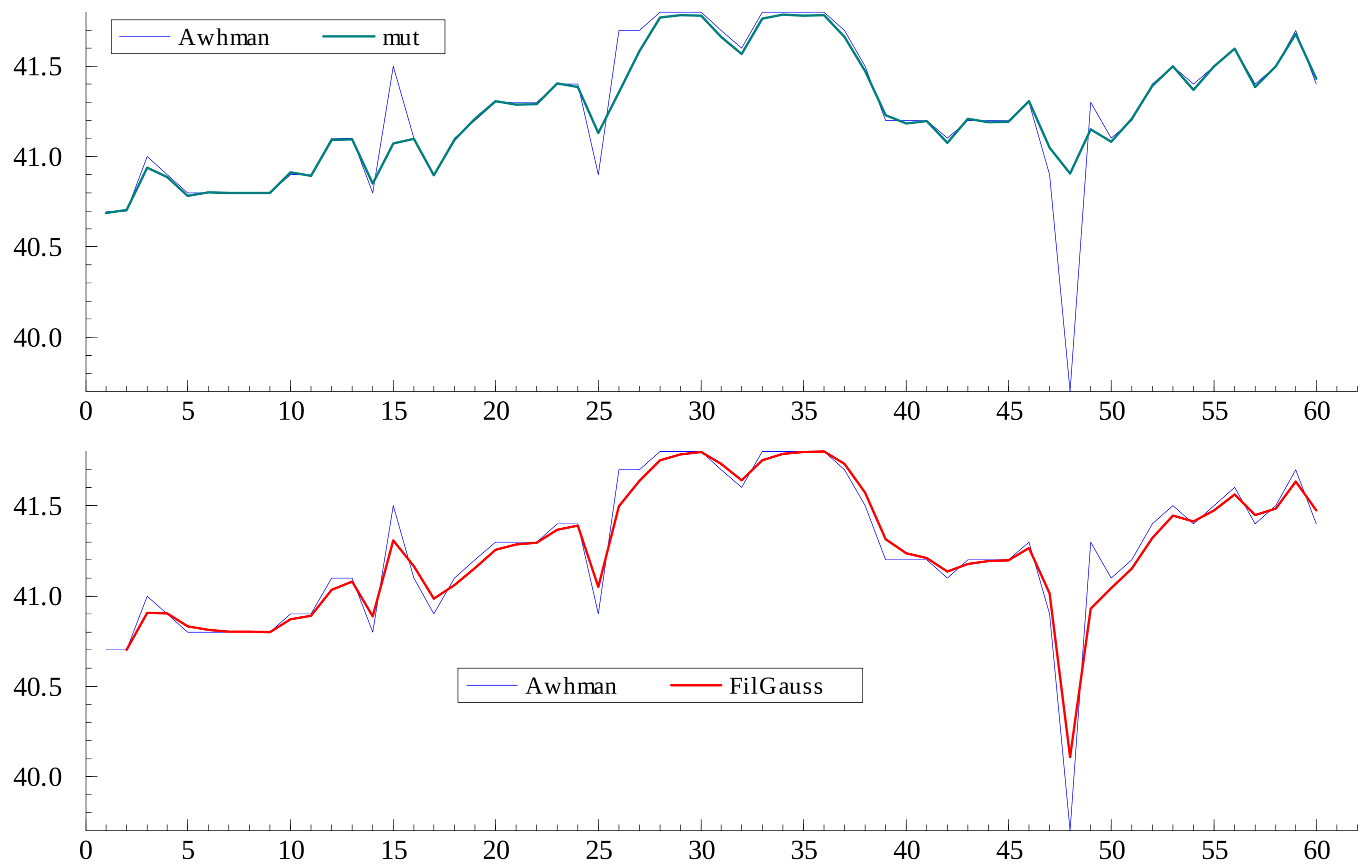


Figure: Impact of u_t for t_ν (with a scale of one) for $\nu = 3$ (thick), $\nu = 10$ (thin) and $\nu = \infty$ (dashed).



Basic properties

$$u_t = (1 - b_t)(y_t - \mu_{t|t-1}), \qquad (1)$$

where

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

is distributed as $beta(1/2, \nu/2)$. The $u_t's$ are $IID(0, \sigma_u^2)$ and symmetrically distributed.
The fact that (2) has a beta distribution follows from the property of the t-distribution

Basic properties

The filter may be generalized to:

$$\mu_{t+1|t} = \phi_1 \mu_{t|t-1} + ... + \phi_p \mu_{t-p+1|t-p} + \kappa_0 u_t + \kappa_1 u_{t-1} + ... + \kappa_r u_{t-r}.$$

Such a filter is denoted as $QARMA(p, r)$. The full model will be called $DCS - t - QARMA(p, r)$. It corresponds to an unobserved component signal plus noise model in which the signal is $ARMA(p, r)$.

Basic properties

In the Gaussian case $u_t = v_t$. If q is defined as $max(p, r+1)$, we may write

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + v_t - (\phi_1 - \kappa_0)v_{t-1} - ... - (\phi_q - \kappa_q)v_{t-q},$$

which is an $ARMA(p, q)$ with MA coefficients $\theta_i = \phi_i - \kappa_{i-1}, i = 1, .., q$. The invertibility conditions apply to $\theta_i = \phi_i - \kappa_{i-1}, i = 1, .., q$ rather than to $\kappa_i, i = 0, ...q$. But more generally, for a $t_\nu-$distribution with $\nu < \infty$,

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \kappa_0 u_{t-1} + ... + \kappa_q u_{t-q} + v_t - \phi_1 v_{t-1} - ... - \phi_p v_{t-p}$$

and the MA disturbances are not identically distributed as each is a different combination of variables, u_t and v_t, which have different (non-normal) distributions. In fact they do not all have the same variances. The process is still $ARMA(p, q)$, but the MA coefficients are not $\phi_i - \kappa_{i-1}, i = 1, .., q$.

Basic properties: moments

When (**??**) is stationary, the location can be written as an infinite moving average,

$$\mu_{t|t-1} = \omega + \sum_{j=1}^{\infty} \psi_j u_{t-j}, \qquad \sum_{j=1}^{\infty} \psi_j^2 < \infty,$$

where $\omega = \delta/(1 - \phi_1 - ... - \phi_p)$, so

$$y_t = \omega + \sum_{j=1}^{\infty} \psi_j u_{t-j} + v_t.$$

The existence of moments of y_t is not affected by the dynamics.

Basic properties: autocovariances*

In the first -order model the form of the ACF is that of an ARMA(1,1) since

$$\rho_\nu(1) = \left[\kappa + \frac{\nu}{3+\nu}\frac{\kappa^2\phi}{1-\phi^2}\right] / \left[\frac{\nu+1}{\nu-2} + \frac{\nu}{3+\nu}\frac{\kappa^2}{1-\phi^2}\right]$$

depends on κ and ϕ, but thereafter $\rho_\nu(\tau) = \phi\rho_\nu(\tau-1), \quad \tau = 2, 3, ...$

Maximum likelihood estimation

The log-likelihood function for the DCS-t model is

$$\begin{aligned} \ln L(\boldsymbol{\psi}, \nu) &= T \ln \Gamma\left((\nu+1)/2\right) - \frac{T}{2}\ln\pi - T\ln\Gamma\left(\nu/2\right) \\ &\quad -\frac{T}{2}\ln\nu - T\ln\varphi - \frac{(\nu+1)}{2}\sum_{t=1}^{T}\ln\left(1+\frac{(y_t-\mu_{t|t-1})^2}{\nu\varphi^2}\right). \end{aligned}$$

Maximization of the log-likelihood function with respect to the unknown dynamic parameters in the vector $\boldsymbol{\psi}$ and the scale and shape parameters, λ and ν, can be carried out by numerical optimization.

Maximum likelihood estimation: information matrix

Let $y_t \mid Y_{t-1}$ have a t_ν-distribution with $\mu_{t|t-1}$ generated by the first-order model. Then, assuming that $|\phi| < 1$ and $b < 1$,

$$\mathbf{I}\begin{pmatrix}\boldsymbol{\psi}\\ \lambda\\ \nu\end{pmatrix} = \begin{bmatrix} \frac{\nu+1}{\nu+3}\exp(-2\lambda)\mathbf{D}(\boldsymbol{\psi}) & 0 & 0 \\ 0 & \frac{2\nu}{\nu+3} & \frac{1}{(\nu+3)(\nu+1)} \\ 0 & \frac{1}{(\nu+3)(\nu+1)} & h(\nu)/2 \end{bmatrix},$$

where $h(\nu)$ is a function of ν (involving trigamma functions) and

$$\mathbf{D}\begin{pmatrix}\kappa\\ \phi\\ \omega\end{pmatrix} = \frac{1}{1-b}\begin{bmatrix} \sigma_u^2 & \frac{a\kappa\sigma_u^2}{1-a\phi} & 0 \\ \frac{a\kappa\sigma_u^2}{1-a\phi} & \frac{\kappa^2\sigma_u^2(1+a\phi)}{(1-\phi^2)(1-a\phi)} & 0 \\ 0 & 0 & \frac{(1-\phi)^2(1+a)}{1-a} \end{bmatrix}$$

Maximum likelihood estimation: information matrix

$$\begin{aligned} a &= \phi - \kappa\frac{\nu}{\nu+3}, \\ b &= \phi^2 - 2\phi\kappa\frac{\nu}{\nu+3} + \kappa^2\frac{\nu\left(\nu^3+10\nu^2+35\nu+38\right)}{(\nu+1)(\nu+3)(\nu+5)(\nu+7)}, \end{aligned}$$

Figure shows a plot of b against κ for $\phi = 0.9$ and $\nu = 6$. The admissible range is slightly bigger than in the Gaussian case where it is $-0.1 < \kappa < 1.9$.

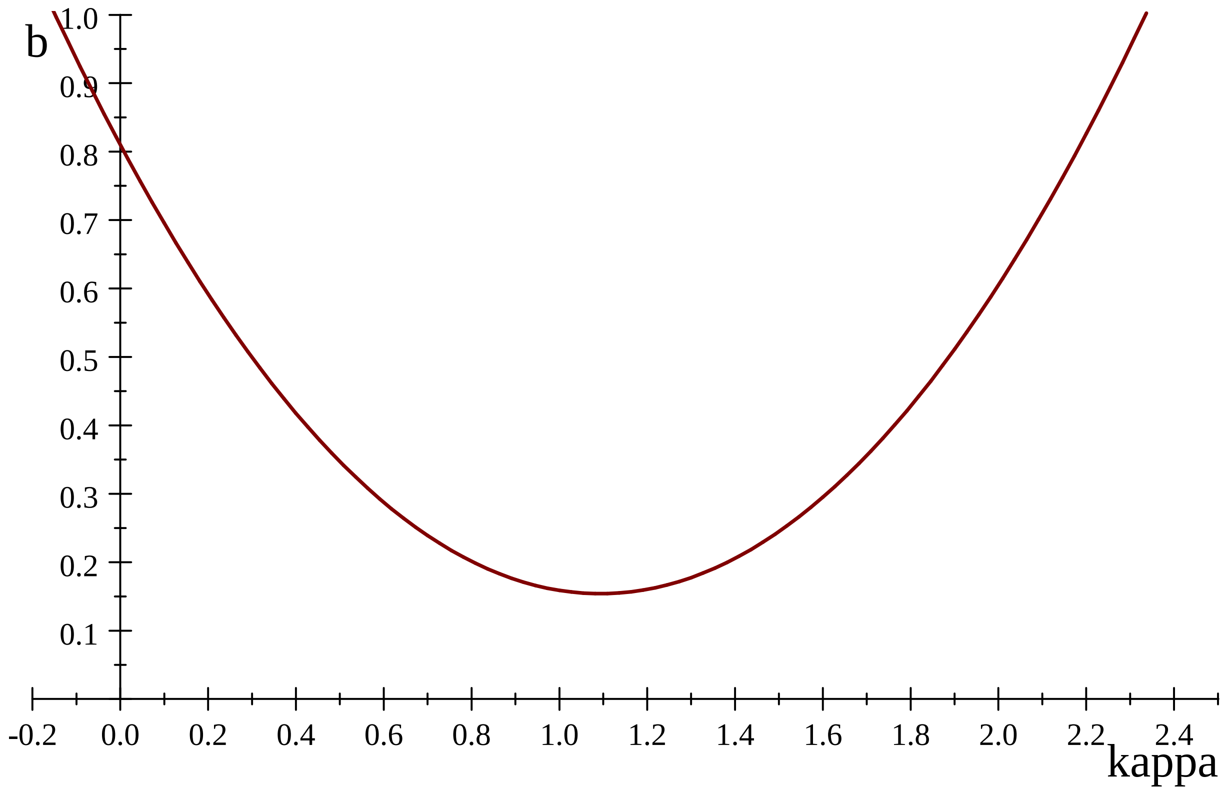


Figure: Plot of b against κ for $\phi = 0.9$ and $\nu = 6$

Maximum likelihood estimation: Gaussian model

For a Gaussian model, $b < 1$ provided that $\phi - 1 < \kappa < \phi + 1$.
The reduced form is the $ARMA(1,1)$ process

$$y_t = \phi y_{t-1} + v_t - \theta v_{t-1}.$$

The condition for strict invertibility in the ARMA(1,1) model is $|\theta| < 1$ and since $\theta = \phi - \kappa$, invertibility ensures that $b < 1$. The condition $\theta \neq \phi$ is needed for identifiability and this condition is equivalent to $\kappa \neq 0$.
When ϕ is known,

$$Var(\widetilde{\kappa}) = 1 - b = 1 - (\phi - \kappa)^2,$$

which is consistent with the standard $MA(1)$ result, $Var(\widetilde{\theta}) = 1 - \theta^2$.

Maximum likelihood estimation: Monte Carlo experiments

Parameter			ML estimates for $T = 1000$				
ϕ	κ		ϕ	κ	λ	ω	ν
0.8	0.5	RMSE	0.037	0.053	0.035	0.093	1.161
		ASE	0.037	0.043	0.029	0.094	0.844
0.8	1.0	RMSE	0.250	0.067	0.031	0.144	0.920
		ASE	0.240	0.045	0.029	0.147	0.844
0.95	0.5	RMSE	0.015	0.048	0.035	0.244	1.100
		ASE	0.012	0.038	0.029	0.269	0.844
0.95	1.0	RMSE	0.012	0.064	0.031	0.387	0.882
		ASE	0.010	0.043	0.029	0.484	0.844

Application to US GDP

A Gaussian AR(1) plus noise model with a constant, was fitted to the growth rate of US Real GDP, defined as the first difference of the logarithm, using the STAMP 8 package. The data were quarterly, from 1947(2) to 2012(1), and the parameter estimates were as follows:

$$\widetilde{\phi} = 0.501, \quad \widetilde{\sigma}_{\eta}^{2} = 7.62 \times 10^{-5}, \quad \widetilde{\sigma}_{\varepsilon}^{2} = 2.30 \times 10^{-5}, \quad \widetilde{\omega} = 0.0078.$$

There was little indication of residual serial correlation, but the Bowman-Shenton statistic is 30.04, which is clearly significant as 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.

Application to US GDP

DCS-location-t model. The estimated degrees of freedom of 6.3 means that the DCS filter is less responsive to more extreme observations, such as the fall of 2009(1).

Parameter	κ	ϕ	λ	ω	ν
Estimate	0.520	0.497	-4.878	0.0079	6.303
Num SE	0.098	0.102	0.073	0.0009	2.310
ASE	0.090	0.140	0.057	0.0009	1.807

Higher-order models and the state space form

The observation in the state space form is related to an $m \times 1$ state vector, $\boldsymbol{\alpha}_t$, through a measurement equation,

$$y_t = \mathbf{z}'\boldsymbol{\alpha}_t + \varepsilon_t, \quad t = 1, ..., T,$$

where $\mathbf{z}$ is an $m \times 1$ vector and ε_t is a serially uncorrelated disturbance with $E(\varepsilon_t) = \mathbf{0}$ and $Var(\varepsilon_t) = \sigma_\varepsilon^2$. The transition equation is

$$\boldsymbol{\alpha}_{t+1} = \mathbf{T}\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t, \quad t = 1, ..., T.$$

The Kalman filter can be written as a single set of recursions going directly from $\boldsymbol{\alpha}_{t|t-1}$ to $\boldsymbol{\alpha}_{t+1|t}$, that is

$$\boldsymbol{\alpha}_{t+1|t} = \mathbf{T}\boldsymbol{\alpha}_{t|t-1} + \mathbf{k}_t v_t, \quad t = 1, ..., T,$$

where $v_t = y_t - \mathbf{z}_t'\boldsymbol{\alpha}_{t|t-1}$ is the innovation and $f_t = \mathbf{z}'\mathbf{P}_{t|t-1}\mathbf{z} + \sigma_\varepsilon^2$ is its variance. The gain vector, $\mathbf{k}_t$, is

$$\mathbf{k}_t = (1/f_t)\mathbf{T}\mathbf{P}_{t|t-1}\mathbf{z}, \quad t = 1, ..., T. \tag{3}$$

Higher-order models and the state space form

Re-arranging the KF equations gives the innovations form

$$\begin{aligned} y_t &= \mathbf{z}'\boldsymbol{\alpha}_{t|t-1} + v_t, \quad t = 1, ..., T, \\ \boldsymbol{\alpha}_{t+1|t} &= \mathbf{T}\boldsymbol{\alpha}_{t|t-1} + \mathbf{k}_t v_t. \end{aligned} \tag{4}$$

A general location DCS model may be set up in the same way as the innovations form of a Gaussian state space model. The model corresponding to the steady-state of (4) is

$$\begin{aligned} y_t &= \omega + \mathbf{z}'\boldsymbol{\alpha}_{t|t-1} + v_t, \quad t = 1, ..., T, \\ \boldsymbol{\alpha}_{t+1|t} &= \boldsymbol{\delta} + \mathbf{T}\boldsymbol{\alpha}_{t|t-1} + \boldsymbol{\kappa} u_t. \end{aligned} \tag{5}$$

Higher-order models: asymptotic theory

The quantities a and b become

$$\mathbf{A}(\nu) = \mathbf{T} - \{\nu/(\nu+3)\}\boldsymbol{\kappa}\mathbf{z}',$$

and

$$\begin{aligned} \mathbf{B}(\nu) &= \mathbf{T}\otimes\mathbf{T} + \frac{\nu}{\nu+3}(\boldsymbol{\kappa}\mathbf{z}'\otimes\mathbf{T} + \mathbf{T}\otimes\boldsymbol{\kappa}\mathbf{z}') \\ &+ \frac{\nu\left(\nu^3 + 10\nu^2 + 35\nu + 38\right)}{(\nu+1)(\nu+3)(\nu+5)(\nu+7)}\boldsymbol{\kappa}\mathbf{z}'\otimes\boldsymbol{\kappa}\mathbf{z}' \end{aligned}$$

The asymptotic theory requires that the roots of the $m^2 \times m^2$ matrix **B**(ν) **have modulus less than one**.
For a Gaussian model this will be the case if the roots of **A** have modulus less than one because $\mathbf{B} = \mathbf{A}\otimes\mathbf{A}$.

Trend and seasonality

Stochastic trend and seasonal components may be introduced into UC models for location. These models, called structural time series models, are implemented in the STAMP package of Koopman *et al* (2009).
The Gaussian random walk plus noise or *local level* model is

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim NID(0, \sigma^2_\varepsilon), \\ \mu_t &= \mu_{t-1} + \eta_t, & \eta_t &\sim NID(0, \sigma^2_\eta), \end{aligned}$$

where $E(\varepsilon_t \eta_s) = 0$ for all t and s. The signal noise ratio is $q = \sigma^2_\eta / \sigma^2_\varepsilon$.
The KF is an EWMA

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

Local level

For the DCS-t filter

$$\begin{aligned} y_t &= \mu_{t|t-1} + v_t, \\ \mu_{t+1|t} &= \mu_{t|t-1} + \kappa u_t. \end{aligned}$$

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

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

A sufficient condition for the weights on current and past observations to be non-negative is that $\kappa(1-b_t) < 1$ and, because $0 \leq b_t \leq 1$, this is guaranteed by $0 < \kappa \leq 1$.
The restriction that $\kappa \leq 1$ is much stricter than is either necessary or desirable. Indeed the argument based on matching autocorrelations suggests an admissible range of $0 \leq \kappa \leq (\nu+1)/(\nu-2)$.

Local level

As regards asymptotic properties,

$$Var(\widetilde{\kappa}) = \left(2\kappa\frac{\nu}{\nu+3} - \kappa^2\frac{\nu\left(\nu^3+10\nu^2+35\nu+38\right)}{\left(\nu+1\right)\left(\nu+3\right)\left(\nu+5\right)\left(\nu+7\right)}\right)\left(\frac{\nu+3}{\nu}\right)^2.$$

In contrast to the case when $|\phi| < 1$, it is necessary that $\kappa > 0$. For finite degrees of freedom, the upper bound will be greater than the value of 2 for a Gaussian model.

Local level

Fitting a local level DCS model (initialized with $\mu_{2|1} = y_1$) to seasonally adjusted monthly data on U.S. Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing (AWHMAN) from February 1992 to May 2010 (220 observations) gave

$$\widetilde{\kappa} = 1.246 \qquad \widetilde{\lambda} = -3.625 \qquad \widetilde{\nu} = 6.35$$

with numerical (asymptotic) standard errors

$$SE(\widetilde{\kappa}) = 0.161(0.090) \qquad SE(\widetilde{\lambda}) = 0.120(0.062) \qquad SE(\widetilde{\nu}) = 1.630(1.991)$$

A drift term was initially included but it was statistically insignificant. The value of b is 0.151. Although $\widetilde{\kappa}$ is greater than one, the resulting filter is perfectly consistent with the properties of the series. Figure shows (part of) the series together with the contemporaneous filter, which for the random walk is $\mu_{t|t} = \mu_{t+1|t}$. Unusually large prediction errors result in a small value of $\kappa(1 - b_t)$ and most of weight in the filter is assigned to $\mu_{t|t-1}$.

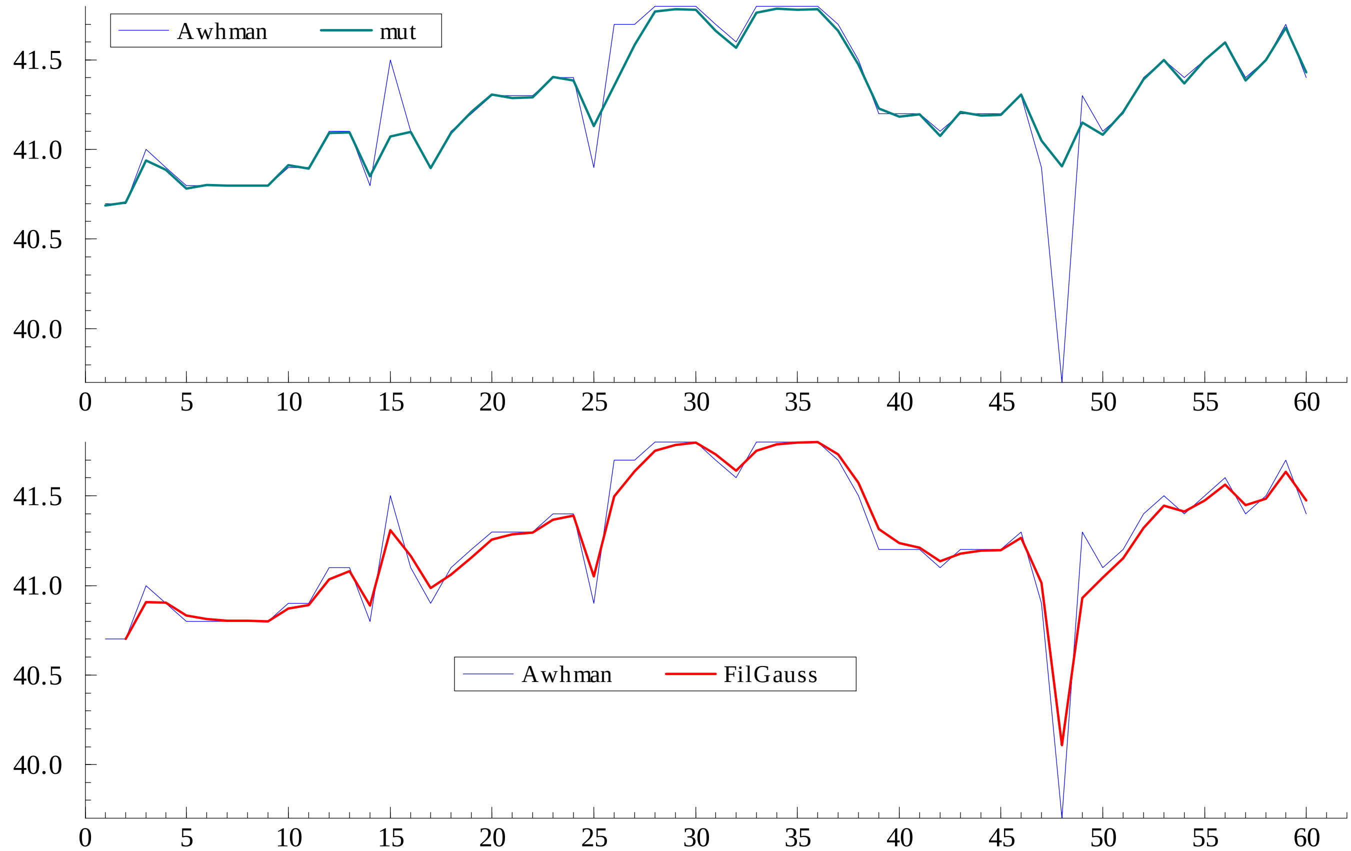


Figure: DCS and Gaussian (bottom panel) local level models fitted to US average weekly hours of production

Local linear trend

The DCS local linear trend filter is

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

The initialization $\beta_{3|2} = y_2 - y_1$ and $\mu_{3|2} = y_2$ can be used, but, as in the local level model, initializing in this way is vulnerable to outliers at the beginning. *Estimating the fixed starting values, $\mu_{1|0}$ and $\beta_{1|0}$, is a better option.*

An IRW trend in the UC local linear trend model implies the contraint $\kappa_2 = \kappa_1^2/(2-\kappa_1), \quad 0 < \kappa_1 < 1$, 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.

Stochastic seasonal

A fixed seasonal pattern may be modeled as $\gamma_t = \sum_{j=1}^{s} \gamma_j z_{jt}$, where s is the number of seasons and the dummy variable z_{jt} is one in season j and zero otherwise. In order not to confound trend with seasonality, the coefficients, γ_j, $j = 1, ..., s$, are constrained to sum to zero.
The seasonal pattern may be allowed to change over time by letting the coefficients evolve as random walks. If γ_{jt} denotes the effect of season j at time t and $\boldsymbol{\gamma}_t = (\gamma_{1t}, ..., \gamma_{st})'$, then

$$\boldsymbol{\gamma}_t = \boldsymbol{\gamma}_{t-1} + \boldsymbol{\omega}_t, \qquad t = 1, ..., T,$$

where $\boldsymbol{\omega}_t$ is a normally distributed, zero-mean vector of disturbances.

Stochastic seasonal

Although all s seasonal components are continually evolving, only one affects the observations at any particular point in time, that is $\gamma_t = \gamma_{jt}$ when season j is prevailing at time t. The requirement that $\sum_{j=1}^{s} \gamma_{jt} = 0$, is enforced by the restriction that the disturbances sum to zero at each point in time.
This restriction is implemented by the correlation structure in

$$Var(\boldsymbol{\omega}_t) = \sigma_{\omega}^2 \left(\mathbf{I} - s^{-1}\mathbf{ii}' \right)$$

where $\boldsymbol{\omega}_t = (\omega_{1t}, ..., \omega_{st})'$ and $\mathbf{i}$ is a vector of ones, coupled with initial conditions requiring that the seasonals sum to zero at $t = 0$.

Stochastic seasonal

In the state space form, the transition matrix is just the identity matrix, but the $\mathbf{z}$ vector must change over time to accommodate the current season. Apart from replacing $\mathbf{z}$ by $\mathbf{z}_t$, the form of the KF remains unchanged. Adapting the innovations form to the DCS observation driven framework gives

$$y_t = \mathbf{z}_t' \boldsymbol{\alpha}_{t|t-1} + v_t, \qquad \boldsymbol{\alpha}_{t+1|t} = \boldsymbol{\alpha}_{t|t-1} + \boldsymbol{\kappa}_t u_t,$$

where $\mathbf{z}_t$ picks out the current season, $\gamma_{t|t-1}$, that is $\gamma_{t|t-1} = \mathbf{z}_t' \boldsymbol{\alpha}_{t|t-1}$. The only question is how to parameterize $\boldsymbol{\kappa}_t$.

Stochastic seasonal

The seasonal dummies in the UC model are constrained to sum to zero and the same is true of their filtered estimates. Thus $\mathbf{i}'\boldsymbol{\kappa}_t = 0$ in the Kalman filter and this property should carry across to the DCS filter. If κ_{jt}, $j = 1, .., s$, denotes the $j-th$ element of $\boldsymbol{\kappa}_t$, then in season j we set $\kappa_{jt} = \kappa_s$, where κ_s is a non-negative unknown parameter, while

$$\kappa_{it} = -\kappa_s/(s-1) \qquad i \neq j.$$

The amounts by which the seasonal effects change therefore sum to zero.

Basic structural model

The seasonal recursions can be combined with the trend filtering equations to give a structure similar in form to that of the Kalman filter for the stochastic trend plus seasonal plus noise UC model, sometimes known as the ‘basic structural model’. Thus

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

The initial conditions at time $t = 0$ are estimated by treating them as parameters; there are $s - 1$ seasonal parameters because the remaining initial seasonal state is minus the sum of the others.

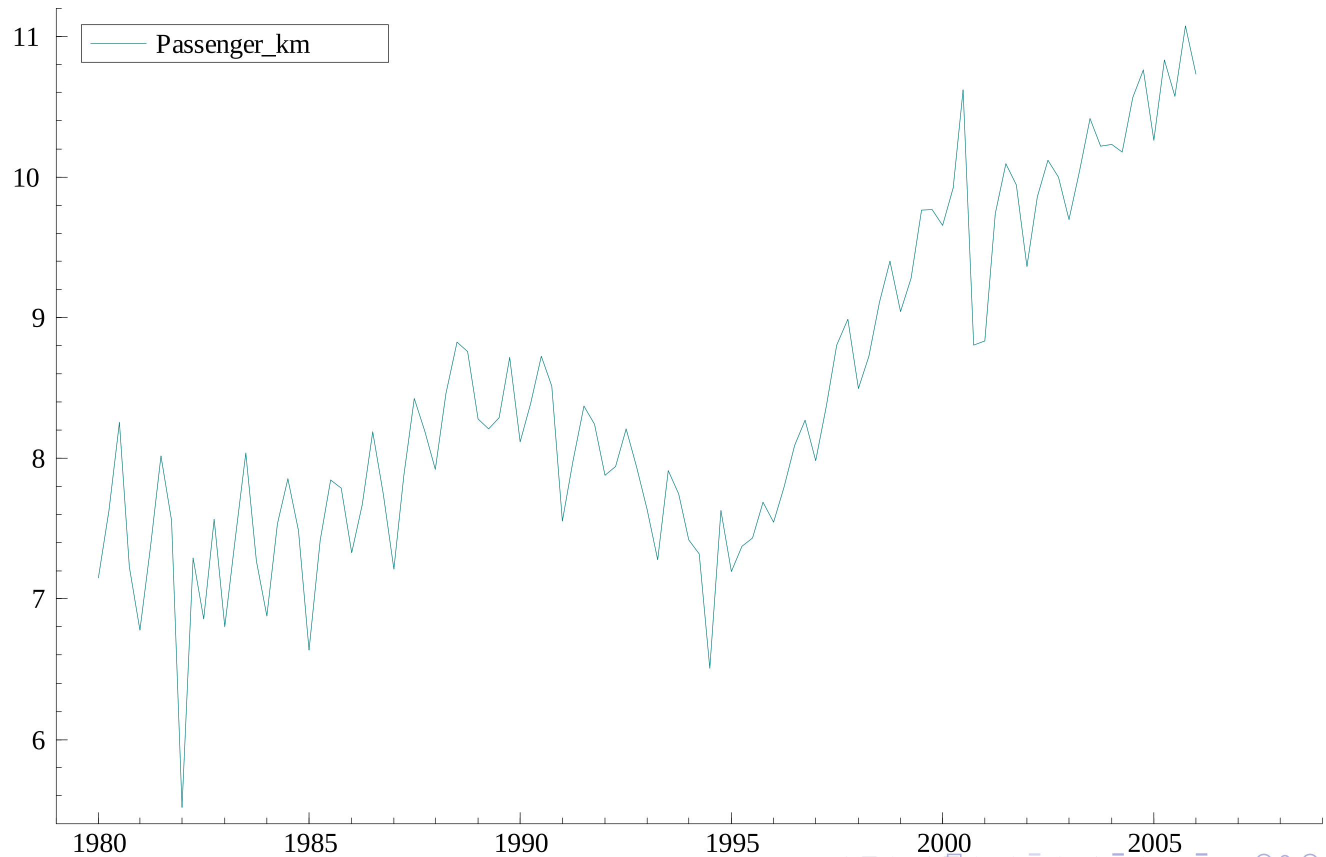


Figure: Logarithm of National Rail Travel in the UK (number of kilometres)

Application to rail travel

An unobserved components model was fitted to the rail series using the STAMP 8 package of Koopman et al (2009). Trend, seasonal and irregular components were included but the model was augmented with intervention variables to take out the effects of observations that are known to be unrepresentative.
The intervention dummies were:
(i) the train drivers strikes in 1982(1,3);
(ii) the Hatfield crash and its aftermath, 2000(4) and 2001(1); and
(iii) the signallers strike in 1994(3).

Application to rail travel

Fitting a DCS model with trend and seasonal avoids the need to deal explicitly with the outliers. The ML estimates for the parameters in a model with a random walk plus drift trend are

$$\widetilde{\kappa} = 1.421(0.161) \qquad \widetilde{\kappa}_s = 0.539\ (0.070) \qquad \widetilde{\lambda} = -3.787\ (0.053)$$
$$\widetilde{\nu} = 2.564\ (0.319) \qquad \widetilde{\beta} = 0.003\ (0.001)$$

with initial values $\widetilde{\mu} = 2.066(0.009)$, $\widetilde{\gamma}_1 = -0.094(0.007)$, $\widetilde{\gamma}_2 = -0.010(0.006)$ and $\widetilde{\gamma}_3 = 0.086(0.006)$.
The figures in parentheses are numerical standard errors. The last seasonal is $\widetilde{\gamma}_4 = 0.018$; it has no SE as it was constructed from the others.

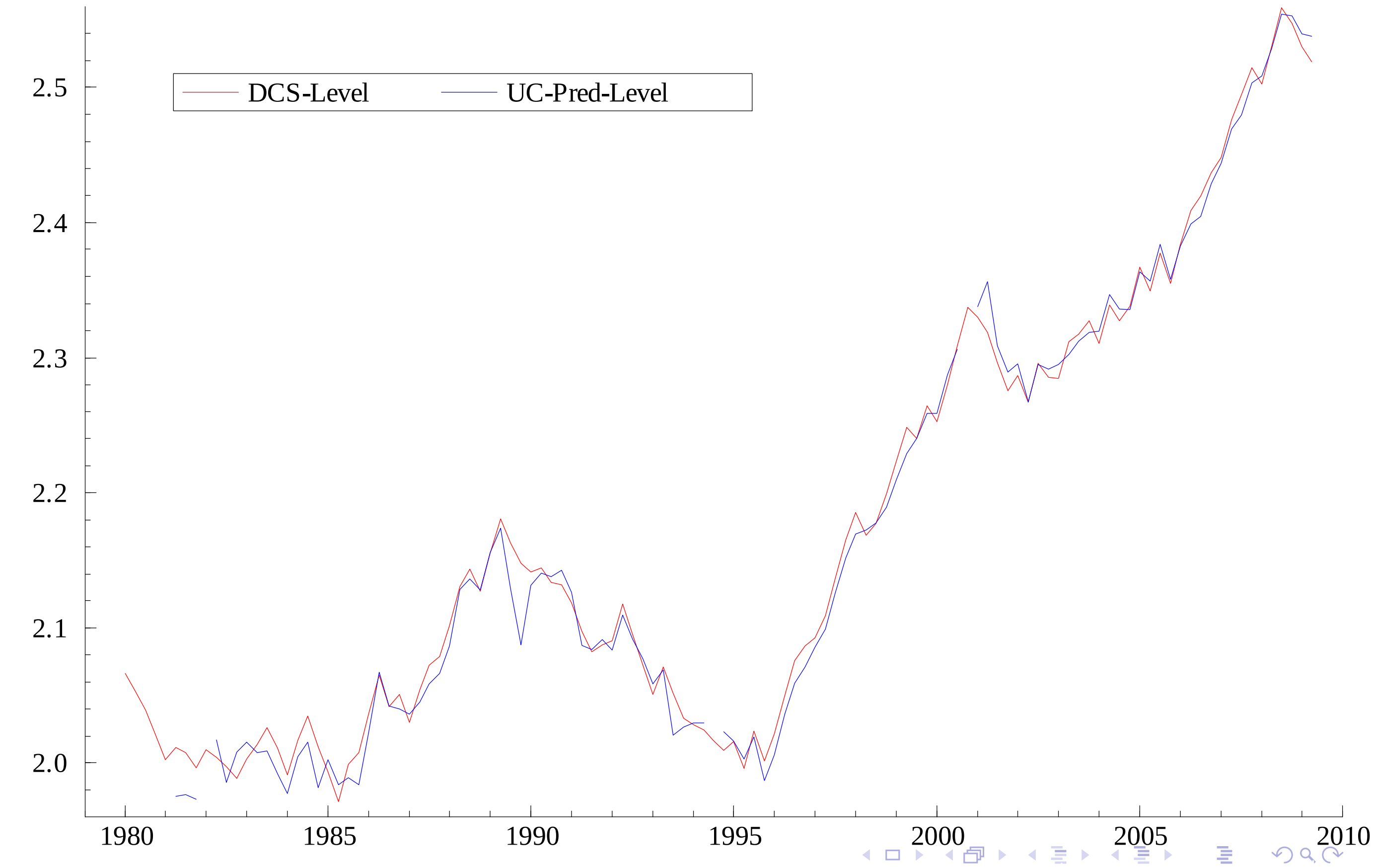


Figure: Trends in National Rail Travel from UC and DCS models

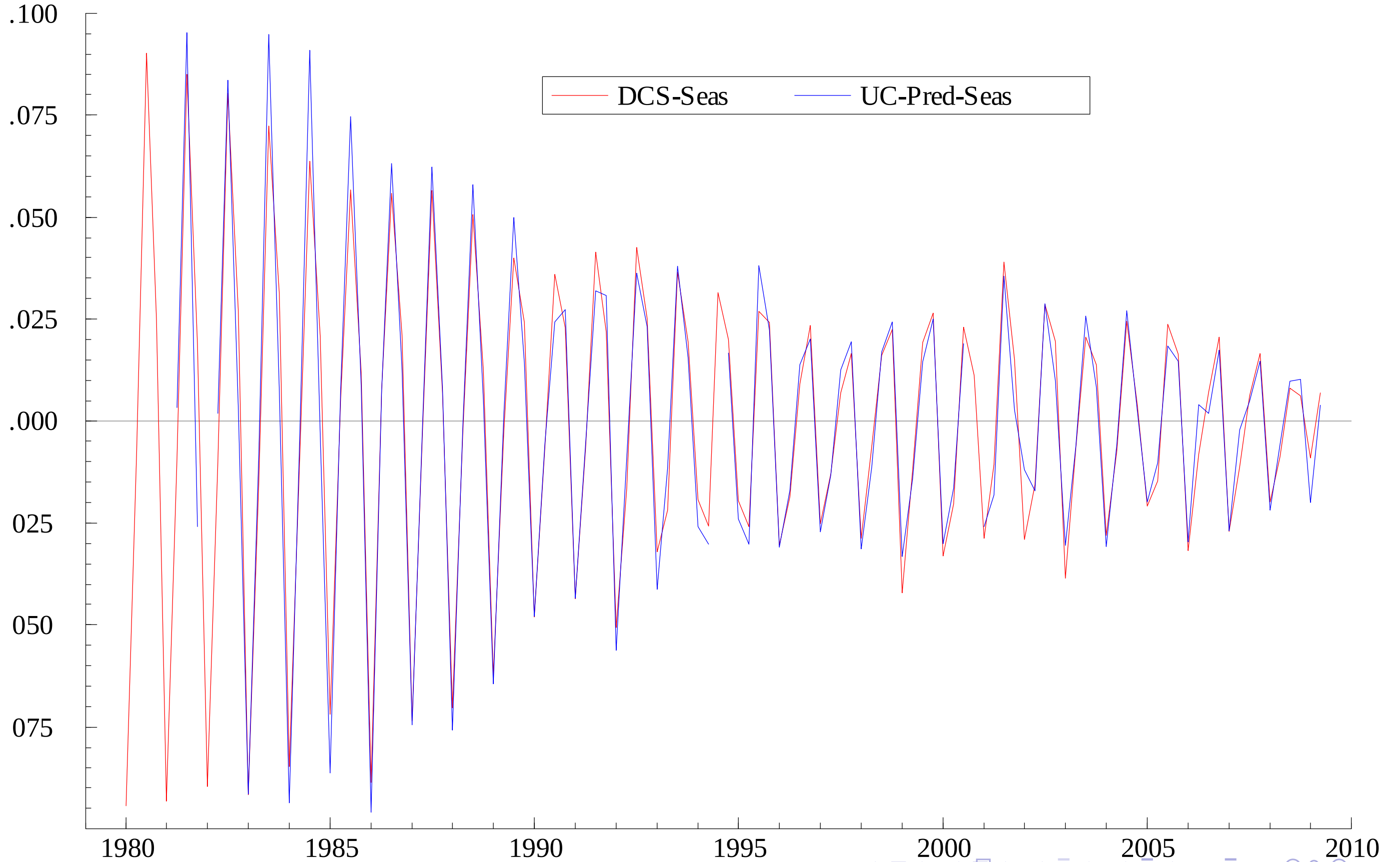


Figure: Seasonals in National Rail Travel from UC and DCS models

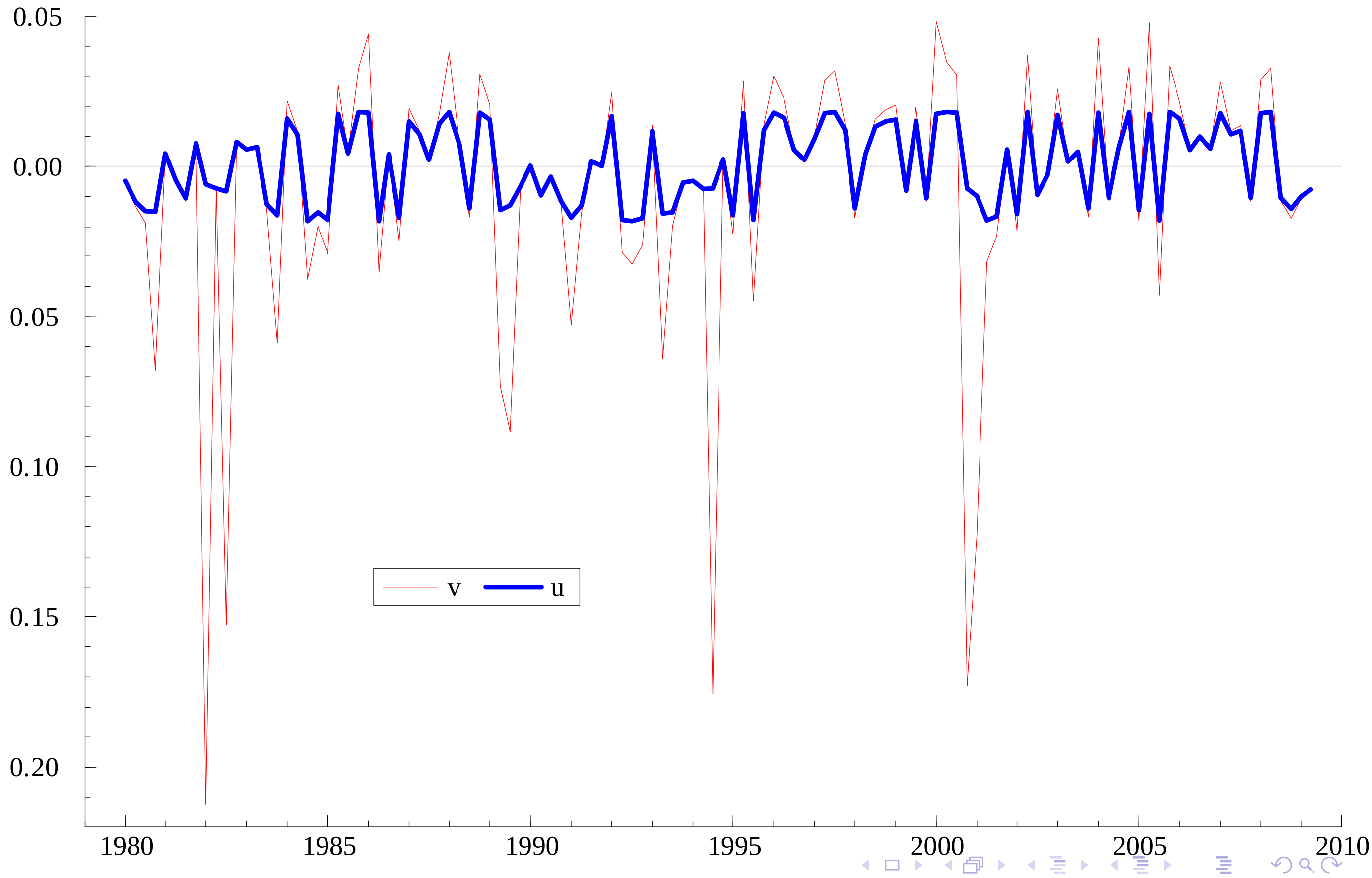


Figure: Residuals and scores from DCS-t model

Rail travel

Figure shows the residuals, that is the one-step ahead prediction errors, and score for the DCS model. The outliers, which were removed by dummies in the UC model, show up clearly in the residuals.
In the score series the outliers are downweighted and the autocorrelations for the score are slightly bigger than those of the residuals presumably because they are not weakened by aberrant values. The Box-Ljung $Q(12)$ statistic is 19.78 for the score and 12.40 for the residuals.
If it can be assumed that only the number of fitted dynamic parameters affects the distribution of the Box-Ljung statistic, its distribution under the null hypothesis of correct model specification is χ^2_{10}, which had a 5% critical value of 18.3. Thus the scores reject the null hypothesis, albeit only marginally, while the residuals do not. Having said that, the score autocorrelations do not exhibit any clear pattern and it is not clear how the dynamic specification might be improved.

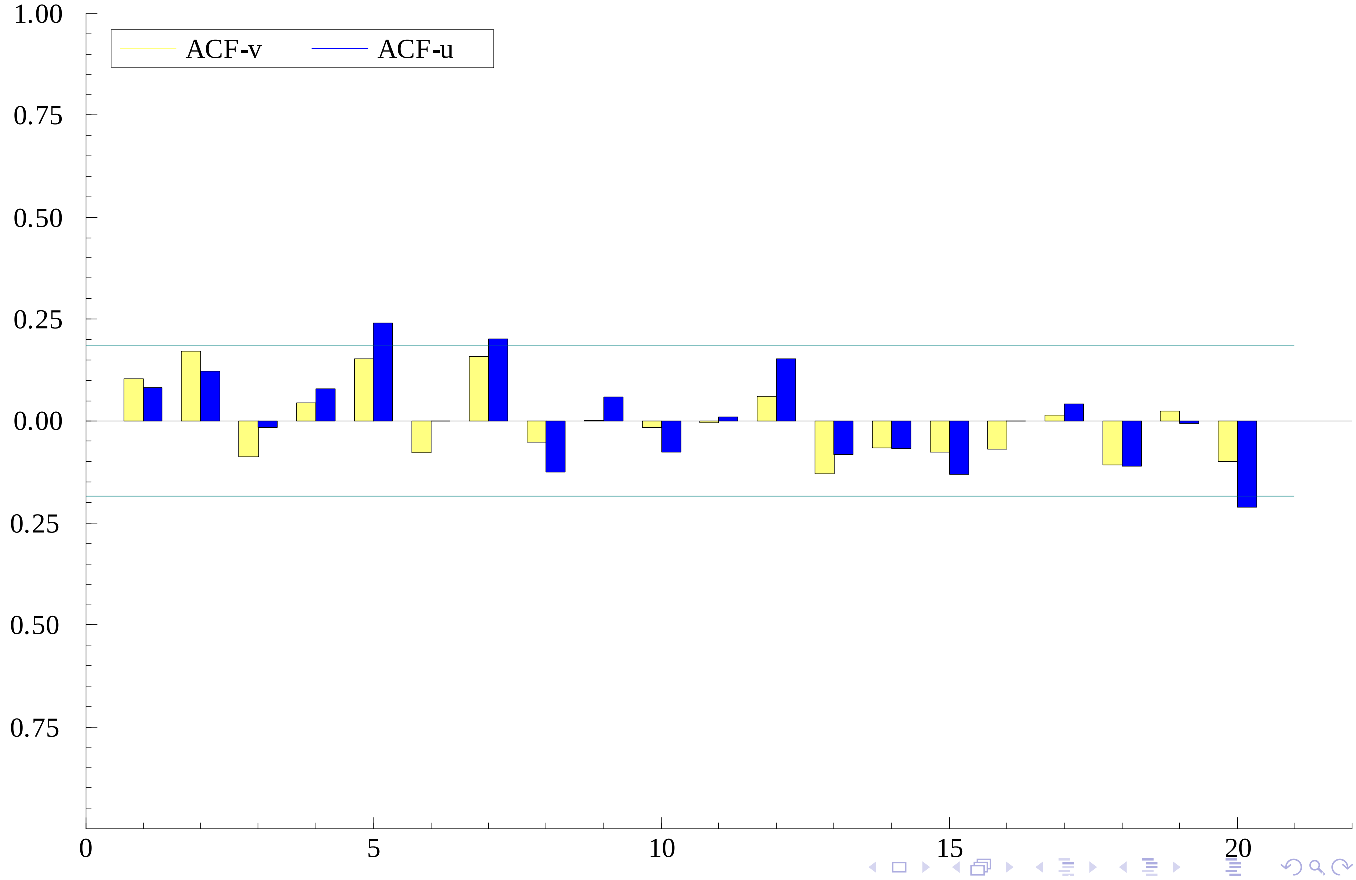


Figure: Residual correlograms for irregular and score residuals from DCS-t model fitted to National Rail Travel (lines are $\pm 1/\sqrt{T}$)

Explanatory variables

The location parameter may depend on a set of observable explanatory variables, denoted by the $k \times 1$ vector $\mathbf{w}_t$, as well as on its own past values and the score. The model can be set up as

$$y_t = \mu^{\dagger}_{t|t-1} + \mathbf{w}'_t \gamma + \varepsilon_t \exp(\lambda), \qquad t = 1, ..., T,$$

where $\mu^{\dagger}_{t|t-1}$ could be a stationary process or a stochastic trend.
The model may be augmented by a seasonal component.
If it is possible to make a sensible guess of initial values of the explanatory variable coefficients, the degrees of freedom parameter, ν, and the dynamic parameters, ϕ and κ for a stationary first-order model or β and κ for a random walk with drift, can be estimated by fitting a univariate model to the residuals, $y_t - \mathbf{w}'_t \widehat{\gamma}$, $t = 1, .., T$. These values are then used to start off numerical optimization with respect to all the parameters in the model.

Asymptotic theory

Consider a model with a stationary first-order component. Assume that the explanatory variables are weakly stationary with mean $\boldsymbol{\mu}_w$ and second moment Λ_w and are strictly exogenous in the sense that they are independent of the ε_t's and therefore of the u_t's. Assuming that $b < 1$ and $\kappa \neq 0$, the limiting distribution of $\sqrt{T}(\widetilde{\kappa} - \kappa, \widetilde{\phi} - \phi, \widetilde{\gamma}' - \gamma', \widetilde{\lambda} - \lambda, \widetilde{\nu} - \nu)'$ is multivariate normal with mean vector zero and covariance matrix given by the inverse of

$$\mathbf{I}\begin{pmatrix} \boldsymbol{\psi} \\ \lambda \\ \nu \end{pmatrix} = \begin{bmatrix} \frac{\nu+1}{\nu+3}\exp(-2\lambda)\mathbf{D}(\boldsymbol{\psi}) & 0 & 0 \\ 0 & \frac{2\nu}{\nu+3} & \frac{1}{(\nu+3)(\nu+1)} \\ 0 & \frac{1}{(\nu+3)(\nu+1)} & h(\nu)/2 \end{bmatrix},$$

but with $\boldsymbol{\psi}$ replaced by $(\kappa, \phi)'$ and $\mathbf{D}(\boldsymbol{\psi})$ replaced by

Asymptotic theory

$$\mathbf{D}\begin{pmatrix} \kappa \\ \phi \\ \gamma \end{pmatrix} = \frac{1}{1-b}\begin{bmatrix} A & D & \mathbf{0}' \\ D & B & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_w \end{bmatrix},$$

where

$$\mathbf{C}_w = (1+\phi^2)\Lambda_w - 2\phi\Lambda_w(1) + 2a(1-a)^{-1}(1-\phi)^2\boldsymbol{\mu}_w\boldsymbol{\mu}_w',$$

with $\Lambda_w(1) = E(\mathbf{w}_t\mathbf{w}_{t-1}') = E(\mathbf{w}_{t-1}\mathbf{w}_t')$.

Asymptotic theory

Corollary. When $\mu^{\dagger}_{t|t-1}$ is known to be a random walk with drift, β, and $\mu^{\dagger}_{1|0}$ is fixed and known,

$$\mathbf{D}\begin{pmatrix}\kappa\\ \gamma\\ \beta\end{pmatrix}=\frac{1}{1-b}\begin{bmatrix}\sigma_u^2 & \mathbf{0}' & \mathbf{0}'\\ \mathbf{0} & \mathbf{C}_{\Delta w} & \boldsymbol{\mu}_{\Delta w}\\ \mathbf{0} & \boldsymbol{\mu}'_{\Delta w} & 1\end{bmatrix},$$

where $\boldsymbol{\mu}_{\Delta w}=E(\Delta\mathbf{w}_t)$ and $\mathbf{C}_{\Delta w}=E(\Delta\mathbf{w}_t\Delta\mathbf{w}'_t)$.
Assume $b<1$ and $\mathbf{C}_{\Delta w}$ is positive definite.

Asymptotic theory

The first differences of the explanatory variables must be weakly stationary but their levels may be nonstationary. Then the covariance matrix of the limiting distribution of $\sqrt{T}\widetilde{\gamma}$ is

$$Var(\widetilde{\gamma})=\left(\frac{2\kappa\nu}{\nu+1}-\kappa^2\frac{\nu\left(\nu^3+10\nu^2+35\nu+38\right)}{(\nu+1)^2\,(\nu+5)\,(\nu+7)}\right)e^{2\lambda}(\mathbf{C}_{\Delta w}-\boldsymbol{\mu}_{\Delta w}\boldsymbol{\mu}'_{\Delta w})^{-1}$$

In principle, the above Corollary may be extended to models where seasonals are included.

Rail travel

Potential explanatory variables for the rail travel series of Sub-section 6.5 are: (i) Real GDP (in £2003 prices), (ii) Real Fares, obtained by dividing total revenue by the number of kilometres travelled and the retail price index (RPI), and (iii) Petrol and Oil index (POI), divided by RPI. The fares series was smoothed by fitting a univariate UC model.

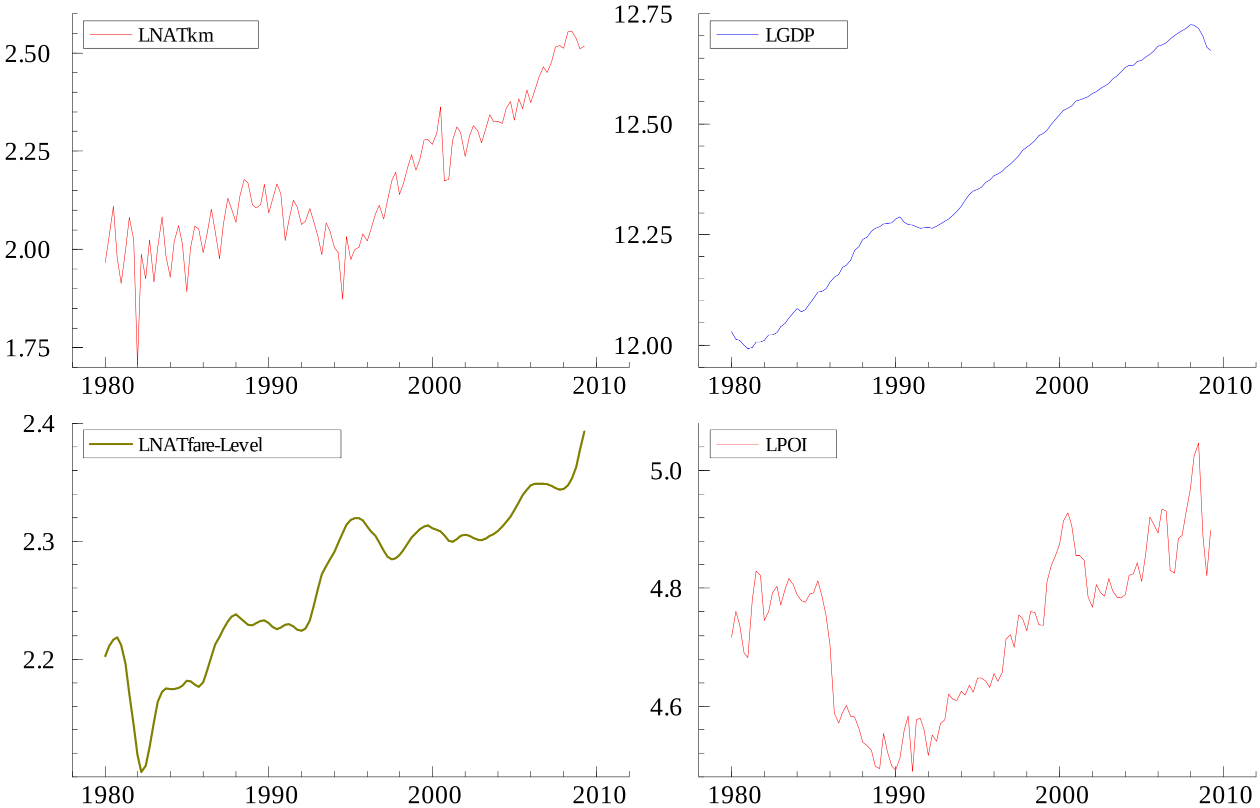


Figure: Rail travel in the UK and explanatory variables

Andrew Harvey (ach34@cam.ac.uk) and Aless Filtering with heavy tails / 63

Application to rail travel: unobserved components model

STAMP gave the following estimates for the coefficients of the logarithms of the explanatory variables:
GDP was 0.716 (0.267), fares was -0.416 (0.245) and POI was 0.050 (0.065).
All the estimates are all plausible. The coefficient of the petrol index is not statistically significant at any conventional level, but at least it has the right sign.
Failure to deal with outliers in a time series regression can lead to serious distortions and this is illustrated when the intervention variables are not included -
the fare estimate is *plus* 0.28 !

Rail travel: DCS-t without seasonal

When rail travel was seasonally adjusted by removing the seasonal component obtained from the univariate DCS-t model and LPOI was also seasonally adjusted, estimating the DCS-t model without a seasonal gave

$$\begin{aligned} \widetilde{\kappa} &= 1.346(0.151) \qquad \widetilde{\lambda} = -3.879\ (0.102) \\ \widetilde{\nu} &= 2.436\ (0.534) \qquad \widetilde{\beta} = 0.001\ (0.002), \end{aligned}$$

where the figures in parentheses are ASEs.
The ASEs calculated for the coefficients of LGDP, Lfare (level) and LPOI (seasonally adjusted) using $Var(\widetilde{\gamma})$ were 0.251, 0.246 and 0.050 respectively.
These figures are close to the standard errors (numerical ?) for the UC model (with seasonal component). (The estimated SEs obtained from a UC model fitted to seasonal adjusted data were similar).

Rail travel: DCS-t with seasonal

Fitting a DCS-t model with seasonal gave

$$\begin{array}{lllll} \widetilde{\kappa} & = & 2.212 & \widetilde{\kappa}_s = 0.771 & \widetilde{\lambda} = -4.059 \\ \widetilde{\nu} & = & 2.070 & \widetilde{\beta} = 0.0004 & \end{array}$$

with initial values $\widetilde{\mu} = -6.162$, $\widetilde{\gamma}_1 = -0.084$, $\widetilde{\gamma}_2 = -0.007$ and $\widetilde{\gamma}_3 = 0.070$.
The coefficients of the explanatory variables were:

$$LGDP = 0.734 \qquad Lfare = -0.427 \quad LPOI = 0.056$$

The Box-Ljung $Q(12)$ statistic is 5.30 for the score and 16.12 for the residuals. This result is surprising because in the univariate model the $Q - statistic$ for the score was bigger than that of the residuals.

Application to rail travel

A good deal, but by no means all, of the growth in rail travel from the mid-nineties is due to the increase in GDP. The continued fall after the economy had moved out of the recession of the early nineties is partly explained by the fact that fares increased sharply in 1993 in anticipation of rail privatisation and continued to increase till 1995.
Nevertheless, as is apparent from the Figure, there remain long-term movements in rail travel that cannot be accounted for by the exogenous variables.

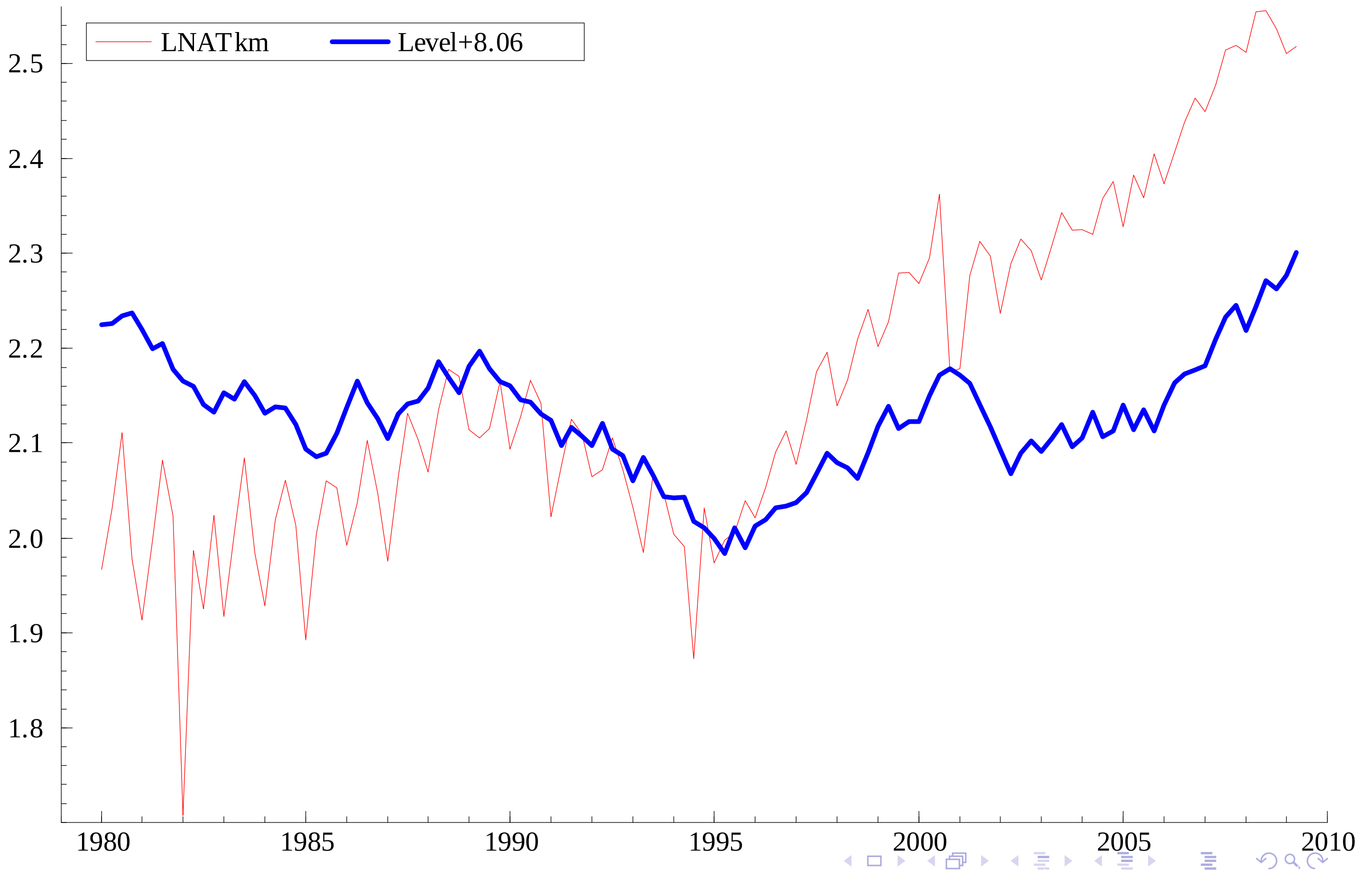


Figure: Trend in rail travel after explanatory variabes have been taken into account. (A constant has been added to the trend so that it is at a level

Stochastic location and stochastic scale

The Student t model for time-varying location may be combined with Beta-t-EGARCH. In other words $y_t \mid Y_{t-1}$ has a t_ν distribution with mean $\mu_{t|t-1}$ and scale $\exp(\lambda_{t|t-1})$, that is

$$y_t = \mu_{t|t-1} + \varepsilon_t \exp(\lambda_{t|t-1}), \qquad t = 1, ..., T.$$

The structure of the information matrix in the static model is such that the form of the dynamic equations for $\mu_{t|t-1}$ and $\lambda_{t|t-1}$ is unchanged. The Beta-t-EGARCH score is

$$u_t = \frac{(\nu+1)(y_t - \mu_{t|t-1})^2}{\nu \exp(2\lambda_{t|t-1}) + (y_t - \mu_{t|t-1})^2} - 1$$

Estimation by ML is straightforward. Unfortunately the presence of $\lambda_{t|t-1}$ in the part of the information matrix associated with $\mu_{t|t-1}$ means that the basic Theorem cannot be applied. Some other route is needed to establish consistency and asymptotically normality of the ML estimators.

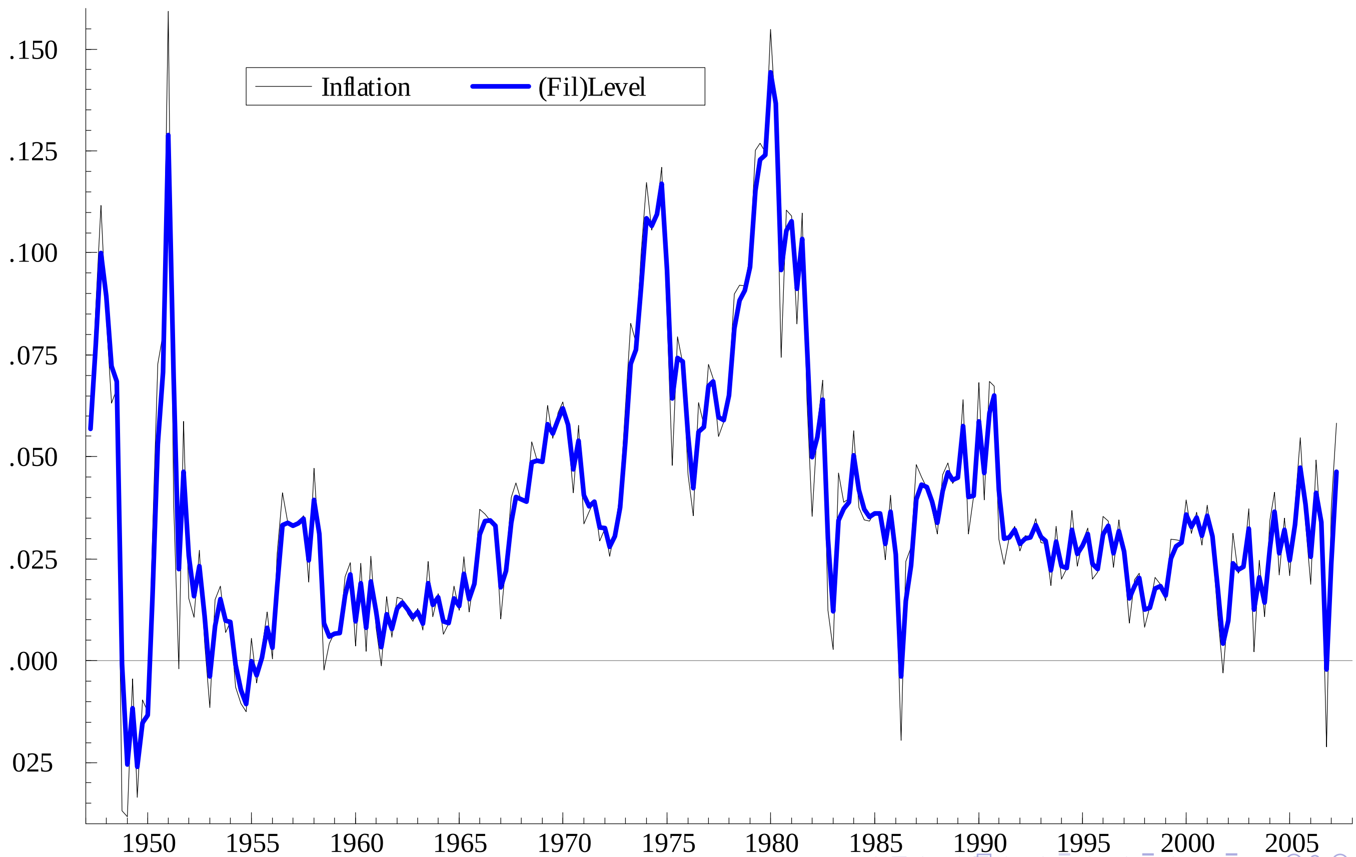
Example - seasonally adjusted rate of inflation in the US. The rate of inflation is often a random walk plus noise. Thus for the DCS-t model

$$\mu_{t+1|t} = \mu_{t|t-1} + \kappa^{\dagger} u_t,$$

where u_t is the level score and the dagger serves to differentiate $\kappa^{\dagger}$ from the κ parameter in the dynamic scale equation.

For the Gaussian unobserved components model, u_t is the prediction error and $\kappa^{\dagger}$ is the Kalman gain.

Fitting such a model using the STAMP package gave an estimate of 0.579 for $\kappa^{\dagger}$. The plot of the filtered level, $\mu_{t+1|t}$, shows it to be sensitive to extreme values, while the ACF of the absolute values of the residuals provides strong evidence of serial correlation in variance.



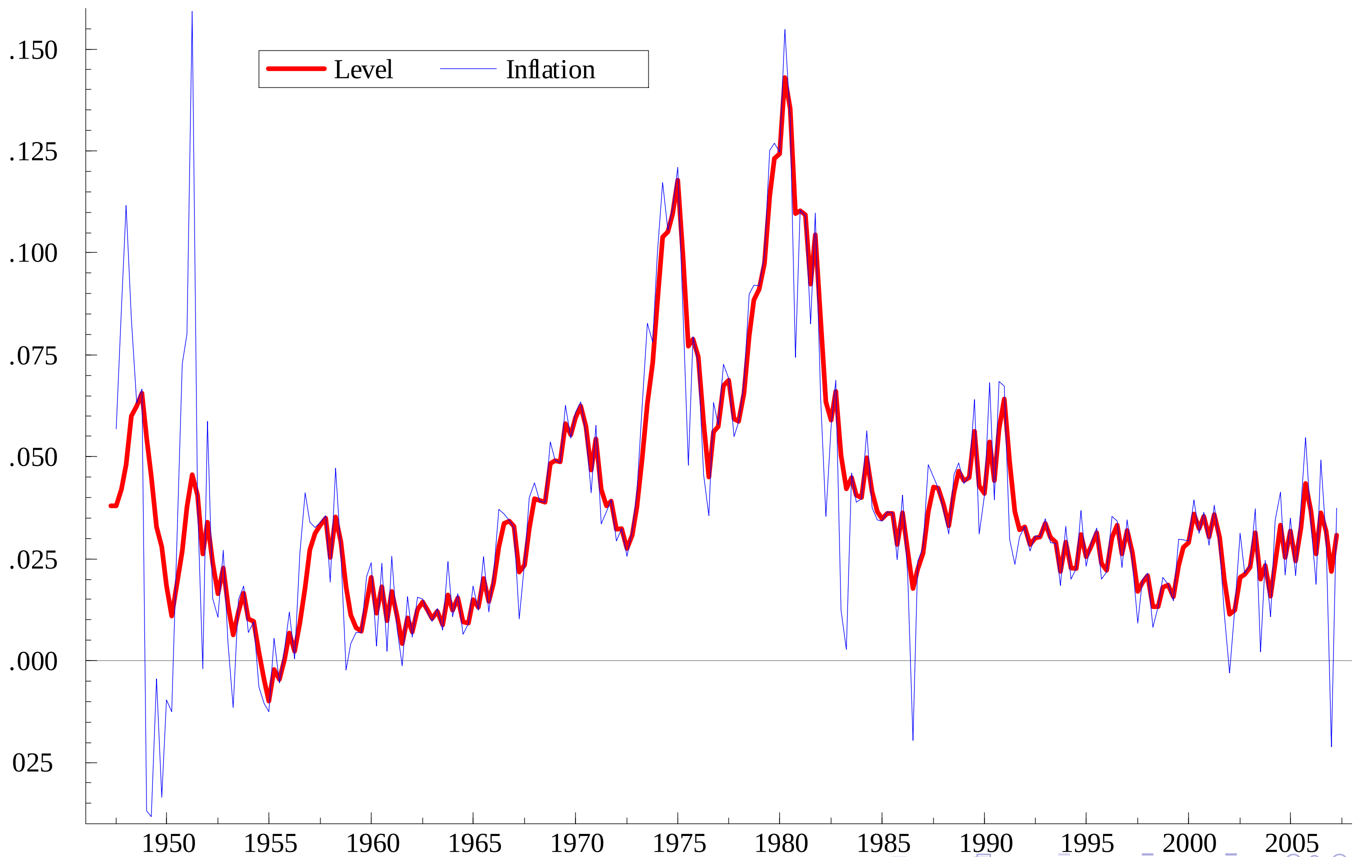
Andrew Harvey (ach34@cam.ac.uk) and Aless Filtering with heavy tails / 63

Figure: Filtered estimates of level from a Gaussian random walk plus noise fitted

Estimating a model in which filtered location is a random walk and scale evolves as a first-order Beta-t-EGARCH process gives the following ML estimates:
for location, $\widetilde{\kappa}^{\dagger} = 0.699(0.097)$,

and for scale, $\widetilde{\delta} = -0.370(0.214)$, $\widetilde{\phi} = 0.912(0.051)$ and $\widetilde{\kappa} = 0.$
$118(0.041)$, with $\widetilde{\nu} = 11.71(4.58)$.

The filtered DCS estimates respond less to extreme values than those from the homoscedastic Gaussian model.



Andrew Harvey (ach34@cam.ac.uk) and Aless Filtering with heavy tails / 63

Figure: Estimated level of US inflation from a random walk plus noise model with

Conclusions

DCS filter enables a time series model to be handled robustly.

Model-based approach, based on a t-distribution, is relatively simple, both conceptually and computationally, and is amenable to diagnostic checking and generalization.

Consider stationary models and then move on to include trend and seasonal components.

The same techniques could be applied to robustify ARIMA and seasonal ARIMA models.

Optimal forecasts can be computed recursively, either as in an ARMA model or by using the SSF, and multi-step conditional distributions can be easily constructed by simulation.

Explanatory variables.

Other generalizations are possible. eg a skewed-t model may be adopted using the method used by Harvey and Sucarrat (2012) for a volatility model.