Extreme Value Mixture Modelling: P-Splines+GPD and `evmix`

7th International Conference of the ERCIM WG on Computational and Methodological Statistics (ERCIM 2014)

Alfadino Akbar & Carl Scarrott

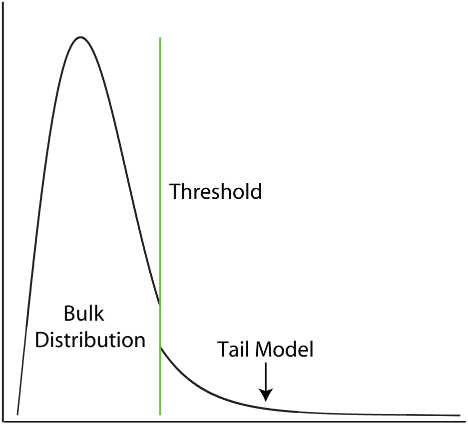
University of Canterbury, New Zealand

December 6, 2014

Talk Outline

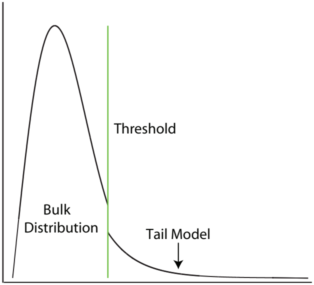
- Intro to Extreme Value Mixture Models
- General Framework for Common Models
- `evmix` package on CRAN
- P-Splines+GPD Model
- Application Results
- Some Advice

Why Use Extreme Value Mixture Models?



- Provide automated and objective “threshold” estimation
- Or avoid threshold choice altogether
- Allow for threshold uncertainty to be taken into account
- **Key issue: sensitivity of tail fit to that of bulk**

Some Terminology



- Tail model typically generalised Pareto distribution (GPD)
- Bulk model has many forms, "loosely" categorised:
  - **parametric**: normal, Weibull, gamma, log-normal, beta
  - **semi-parametric**: mixtures of gamma, normal, log-normal
  - **nonparametric**: mixture of uniforms, kernel density estimation, smoothing polynomials

GPD and Tail Fraction Scaling

- Suppose $X|X > u \sim GPD(\sigma_u, \xi)$ for threshold exceedances:

$$P(X > x|X > u) = \begin{cases} \left[1 + \xi\left(\dfrac{x-u}{\sigma_u}\right)\right]_+^{-1/\xi} & \xi \neq 0, \\ \exp\left[-\left(\dfrac{x-u}{\sigma_u}\right)\right]_+ & \xi = 0. \end{cases}$$

- GPD is a conditional model, to make it unconditional:

$$P(X > x) = \phi_u P(X > x|X > u)$$

- "Tail fraction" above the threshold or "threshold exceedance probability" $\phi_u = P(X > u)$ is an implicit parameter
- Usually estimated using sample proportion, the maximum likelihood estimate
- Classic GPD tail modelling approach

Common Mixture Model Structure

- Common mixture model specification for cumulative distribution function:

$$F(x) = \begin{cases} H(x) & x \leq u, \\ H(u) + (1 - H(u))G(x) & x > u. \end{cases}$$

- $H(x)$ is bulk model cdf and $G(x)$ is the GPD or other conditional tail model for exceedances
- Tail fraction is specified by the bulk model (parameters):
  - $\phi_u = 1 - H(u)$
- Terminology: **bulk model approach**

- Essentially, borrowing information from bulk where you have more data
- Induces sensitivity of tail fit to bulk model performance

General Mixture Model Structure

- Some mixture models use a more general specification:

$$F(x) = \begin{cases} H(x)\dfrac{1-\phi_u}{H(u)} & x \le u, \\ (1-\phi_u) + \phi_u G(x) & x > u. \end{cases}$$

- Extra explicit parameter ϕ_u for tail fraction
- Rescaling of bulk $\frac{1-\phi_u}{H(u)}$ ensures density integrates to unity
- Closer to classical GPD tail modelling approach
- Includes bulk model approach as special case
- Terminology: **parameterised tail fraction approach**

- Extra degree of freedom
- Tail fit robust to bulk model misspecification

Further Niceties

- Mixture models have no requirement of density to be continuous at threshold
  - Note: cdf is continuous, just density is not
  - Usually physically sensible to have continuous density
  - Various parameter constraints to achieve continuity (incl. upto second derivatives)
  - Induces some further dependence between bulk and tail estimates
- Smooth transition functions (Frigessi et al 2003, Holden and Haug 2009) are being developed
  - Weak performance in wide applications, as missing the tail fraction scaling of GPD
  - Promising but more development needed
- **Don't forget that in classical approach that the GPD is a conditional model, so needs appropriate tail fraction scaling!**

`evmix` Package in `R`

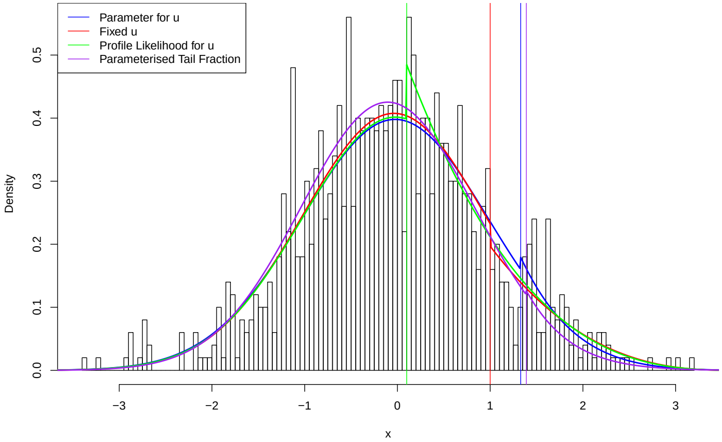
- General goal:
  - Suite of tools for extreme value threshold estimation and uncertainty quantification
- Named after `evd` package as similar syntax for basic GPD and threshold diagnostic plots
- Current release has:
  - most extreme value mixture models in the current literature
  - model fit diagnostic plots for all of them
  - Maximum likelihood estimation with either:
    - fixed threshold;
    - profile likelihood for threshold; or
    - combine threshold with other model parameters
  - Variants of all models with constraint of continuity of density at threshold
  - threshold diagnostic plots (MRL, threshold stability, Hill/AltHill/smooHill plots)
- Available on CRAN for download
- Any feedback and bug reports welcome!

Example Usage 1

- Example of fitting variants of the normal bulk with GPD tail
- Different inference approaches for threshold

```
set.seed(1234)
x = rnorm(1000)

# Assume bulk model tail fraction by default and
# threshold as parameter so maximised wrt as with other parameters
fit = fnormgpd(x)

# Can apply fixed threshold approach (if threshold pre-chosen)
fit.u = fnormgpd(x, useq = 1, fixedu = TRUE)

# Profile likelihood search over sequence of thresholds, then fixed
fit.profu = fnormgpd(x, useq = seq(0, 2, 0.01), fixedu = TRUE)

# Change to parameterised tail fraction
fit.profu.phiu = fnormgpd(x, useq = seq(0, 2, 0.01), fixedu = TRUE, phiu = FALSE)
```

Example Usage 2



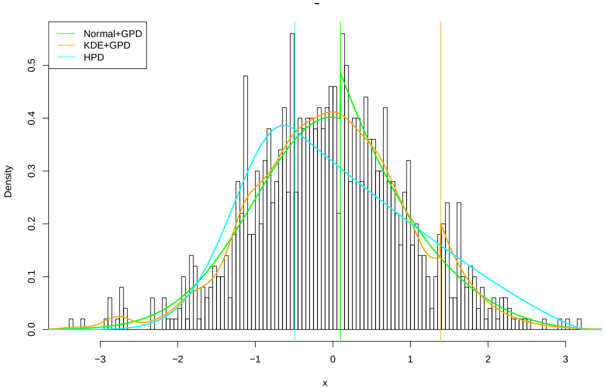
Example Usage 3

- Nonparametric KDE's use cross-validation likelihood so much slower:

```
# Nonparametric KDE for bulk model
fit.kde = fkdengpd(x, useq = seq(0, 2, 0.01), fixedu = TRUE)
```

- Hybrid Pareto (no tail fraction scaling at all)

```
# Hybrid Pareto
fit.hpd = fhpd(x)
```



Motivation for P-splines+GPD

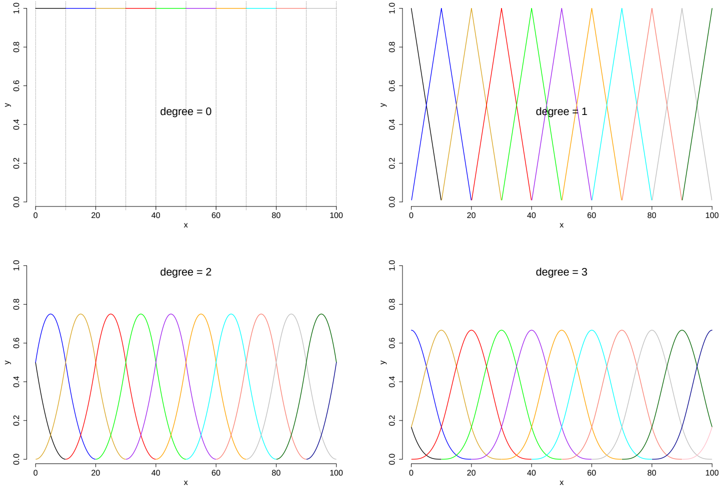
- MacDonald et al. (2011) developed extreme value mixture models with nonparametric kernel density estimator (KDE) for the bulk model
- KDE does not perform well for bounded support, e.g. for pole at boundary
- Leads to leakage past boundary and bias near boundary when the density is non-zero
- MacDonald et al (2013) used boundary corrected KDE, for a wide range of boundary correction approaches
- Identified some challenges:
  - quick and dirty approaches for boundary correction don't improve bias much;
  - more sophisticated approaches for correcting the boundary bias usually have high computational overhead (e.g. renormalisation of KDE to make it proper); and
  - extensions of boundary corrected KDE's to non-stationary (multidimensional) problems not well developed

P-Splines Based Density Estimation

- Proposed by Eilers and Marx (1996) combines B-splines with flexible penalty constraints
- Their approach:
  - Histogram binning on fine mesh to get counts
  - Poisson regression on counts to estimate spline coefficients - mixed model representation with penalty
  - Penalty magnitude estimated using statistics which aim to prevent overfitting (e.g. AIC/BIC, cross-validation RMSE)
- Very heuristic justification to their approach, but it is flexible and is seeing wide application

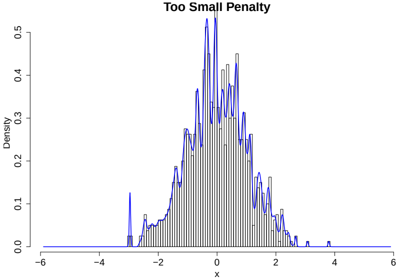
- B-splines naturally have bounded supported due to knots and multi-dimensional smoothing easy using tensor products
- Note for extremists - ignore rules of thumb for specifying bins, knots, etc. they often don't work well for heavy tails

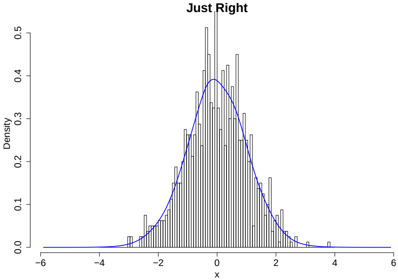
B-Splines

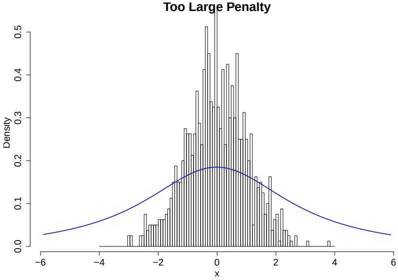


- local basis functions, piecewise polynomials of fixed degree
- need to define knots and degree
- Not natural B-splines (adapted behaviour at boundary)

P-Splines







- histogram binning not critical, provided the histogram DE is coarse so is not smoothing
- knots aren't critical, provided you have plenty!
- penalty aims to prevent overfitting

Flexible Penalties

- Penalties are usually specified using difference in coefficients $\alpha_0, \alpha_1, ..., \alpha_k$
- Use delta notation $\Delta\alpha_i = \alpha_i - \alpha_{i-1}$
- Simple form of penalty:

$$\lambda \sum_i (\Delta\alpha_i)^2$$

- λ controls strength of penalty
- Conceptual idea: equal $\alpha_i = \alpha$ then get uniform density, larger differences in neighbouring α_i's means more roughness
- Higher order penalties recommended, e.g. second order $\Delta^2\alpha_i = (\alpha_i - \alpha_{i-1}) - (\alpha_{i-1} - \alpha_{i-2})$
- Efficient computation compared to traditional spline penalties
- Local basis and penalty difference matrix both sparse

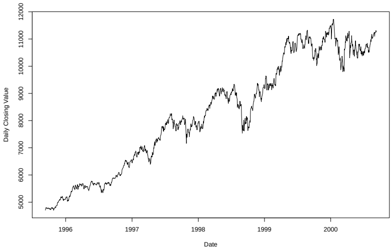
P-Splines + GPD Extreme Value Mixture Model

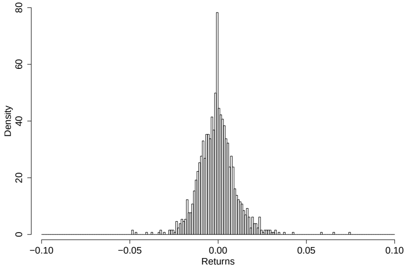
- The bulk model cdf $H(x)$ is the P-splines DE:

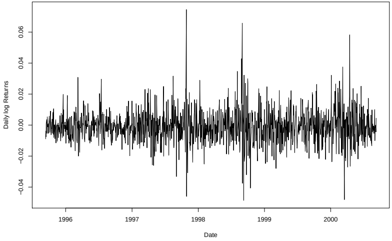
$$F(x) = \begin{cases} H(x)\dfrac{1-\phi_u}{H(u)} & x \leq u, \\ (1-\phi_u) + \phi_u G(x) & x > u. \end{cases}$$

- Two stage MLE inference following Cabras and Castellanos (2009):
  - MLE for P-spline density, with penalty chosen by AIC/BIC/CVRMSE;
  - Assume P-splines are fixed when fitting mixture model (threshold and GPD parameters);
- Profile likelihood estimation for threshold (advised approach)
- Investigating combined penalized likelihood approaches (and avoid binning step)

Application: Dow Jones returns

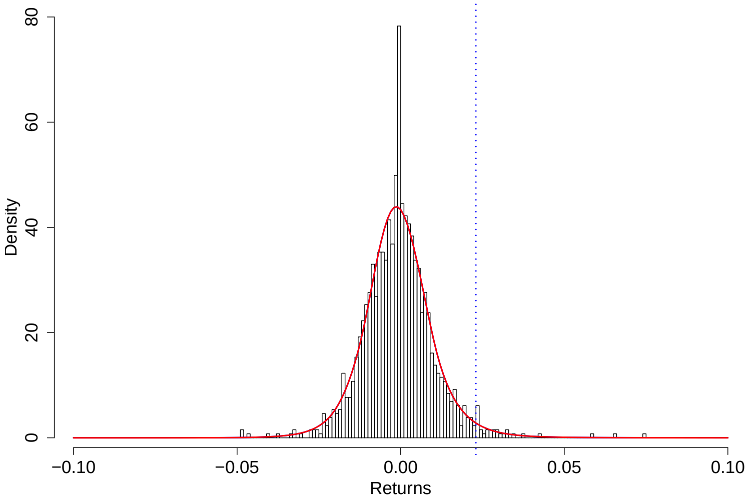




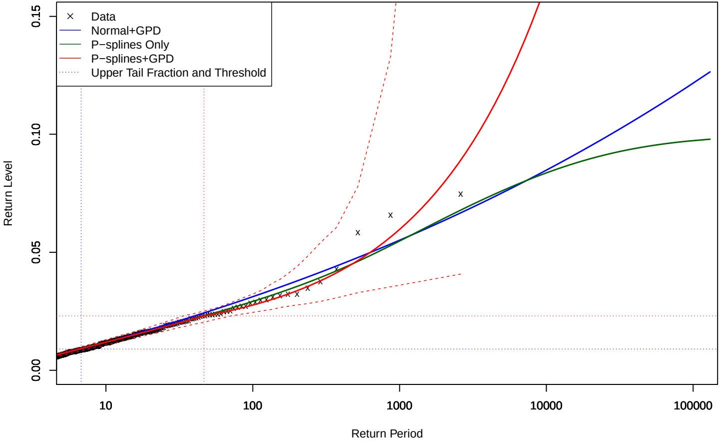


- `data(dowjones)` from `ismev` package
- Daily closing price 1996-2000
- Log returns

Application: Dow Jones returns



Application: Dow Jones returns



- P-splines DE (green) has bounded support, so short tailed behaviour
- P-splines+GPD and Normal+GPD differing thresholds and upper tail behaviour, but appear to be within sample variation

The Good and the Bad!

- The good:
  - conceptually simple and reasonably computational efficient compared to many of the more usual smoothing splines
  - naturally accounts for bounded support (still boundary bias?)
  - easy to build in continuity constraints on PDF (lose degrees of freedom from P-spline fit, rather than GPD parameters)
  - straightforward extensions to nonstationary problems using tensor products of B-splines
- The bad:
  - log-link in Poisson regression leads to no closed form for CDF, so computationally inefficient!
  - needs many knots for heavy tails, non-regular knots are possible but specification of sensibly behaved penalties messier

References and Website

Review paper:

Scarrott and MacDonald (2012). A review of extreme value threshold estimation and uncertainty quantification. REVSTAT Statistical Journal 10(1), 33-60.
(all references in here)

Package: `evmix` available on CRAN (all feedback appreciated)

Website:
`http://www.math.canterbury.ac.nz/~c.scarrott/evmix`

Thanks for your attention...