An Improved Test for Earnings Management Using Kernel Density Estimation

Henry Lahr
Centre for Business Research
University of Cambridge
Motivation: Burgstahler / Dichev 1997

- Linear interpolation of histogram bins; t-test against empirical data
- Bin width and location (almost) arbitrary
- Distributional assumption restrictive
Relaxing the distributional assumption: Kernel density estimation

- Bollen / Pool, 2009 (JoF)
- Estimating a reference distribution by kernel density estimation
- Problem: What is a ‘good’ reference distribution?

Net income / Mkt.Cap.

Linearity assumption relaxed
Testing against a reference distribution

- Multiple reasonable kernel density estimates possible
- Spurious significance if density is oversmoothed (=bandwidth $h$ large)
- Bin origin and boundaries still arbitrary

Net income / Mkt.Cap.

But which one is the right one?
Constructing the reference distribution

- Fit depends on choice of bandwidth and kernel function
- **Idea:** Bootstrap empirical cumulative distribution, construct confidence levels, select bandwidth such that reference distribution is indistinguishable from empirical distribution

![Graph showing ECDF, iKDE, confidence band, and KS band with annotations for Oversmoothed and Fitted at p=0.1]
Finding discontinuities

- Remaining difference can identify candidates for discontinuities
- No need for histogram bins
- Maximum difference at 0.0000642 → Candidate for discontinuity
Data

- US company accounts from Compustat
- Financial institutions excluded
- 1976-2010
- Net income, lagged market capitalisation
- N=174009 (standardised earnings), N=163664 (earnings changes)

- EPS estimates and actual values from I/B/E/S
- 1986-2010
- Forecast errors based on I/B/E/S-adjusted and manually adjusted estimates and actuals
- Forecast errors scaled by lagged market capitalisation
# Standardised earnings (Net income\(_t\) / Market cap\(_{t-1}\))

<table>
<thead>
<tr>
<th>Period/N Interval</th>
<th>Epanechnikov kernel</th>
<th>Gaussian kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bootstrap KDE</td>
<td>Rule-of-thumb KDE</td>
</tr>
<tr>
<td>1976--1994 ([-h,0]]</td>
<td>h=0.005739</td>
<td>-6.542</td>
</tr>
<tr>
<td>N=78923 ([0,h]]</td>
<td>h=0.010622</td>
<td>4.593</td>
</tr>
<tr>
<td></td>
<td>h=0.011515</td>
<td></td>
</tr>
<tr>
<td>1995--2010 ([-h,0]]</td>
<td>h=0.011515</td>
<td>-3.179</td>
</tr>
<tr>
<td>N=96426 ([0,h]]</td>
<td>h=0.005385</td>
<td>2.250</td>
</tr>
<tr>
<td></td>
<td>h=0.005385</td>
<td></td>
</tr>
<tr>
<td>1976--2010 ([-h,0]]</td>
<td>h=0.002670</td>
<td>-5.942</td>
</tr>
<tr>
<td>N=175349 ([0,h]]</td>
<td>h=0.010622</td>
<td>5.038</td>
</tr>
</tbody>
</table>
Standardised earnings over time (Epanechnikov kernel)

- Bootstrap KDE
- Rule-of-thumb KDE
- Burgstahler/Dichev
Earnings changes over time (Epanechnikov kernel)
Earnings changes over time (Gaussian kernel)
EPS forecast errors (scaled by market cap.)

a) Median forecast 1 month before earnings announcement

b) Median forecast 3 months before earnings announcement

→ Not possible to construct reference distributions including zeroes

Note: Forecast errors based on “unadjusted” actuals, scaled by market capitalisation in t-1 year
EPS forecast errors (“adjusted” actuals)

\[
\text{Error}_{t^*t} = (\text{AdjActual}_{t^*} \times \text{AdjFac}_{t^*} - \text{MedEst}_{t^*}) \times \frac{\text{Shares}_{t^*}}{\text{MktCap}_{t-1}}
\]

- Bootstrap KDE does **not** detect any discontinuity
- Rule-of-thumb KDE significant for **all** forecast horizons
- Burgstahler/Dichev test significant for **all** forecast horizons

→ Note the many small rounding errors(?): For example, an actual EPS value of 1.16 that is shrunk to 0.0967 due to a 12:1 stock split would lead to a forecast error of 0.0967 × 12 − 1.16 = 0.0004 per share.
**EPS forecast errors ("unadjusted" actuals)**

\[
\text{Error}_{t^*t} = (\text{UnadjActual}_{t^*} \times \text{AdjFac}_{t^*} / \text{AdjFac}_t - \text{MedEst}_{t^*}) \times \text{Shares}_{t^*} / \text{MktCap}_{t-1}
\]

- Bootstrap KDE significant up to 3 months, but sign indicates problem
- Rule-of-thumb KDE significant for all forecast horizons
- Burgstahler/Dichev test significant for all forecast horizons
EPS forecast errors (+ random [-0.005,0.005] error)

\[ \text{Error}_{t^*/t} = (\text{UnadjActual}_{t^*} \times \text{AdjFac}_{t^*}/\text{AdjFac}_t - \text{MedEst}_{t^*} + e) \times \text{Shares}_{t^*}/\text{MktCap}_{t-1} \]

- Bootstrap KDE significant at 6 and 9 months horizon
- Rule-of-thumb KDE significant at all forecast horizons
- Burgstahler/Dichev test significant at all except 6-month horizon

(Real) earnings management at long horizons vs. no (accruals-based) earnings management at short horizons?
Conclusion

- Use bootstrap to find reference distribution
- Generalization of prior approaches to identifying discontinuities
- No prior knowledge about the true distribution necessary
- Kernel function is the only parameter to choose
  (Epanechnikov-kernel works fine due to cut-off property; using
  Gaussian kernels can go badly wrong)

- Fewer discontinuities in earnings in recent years
- Discontinuities continuously detected in earnings changes
- Little evidence for discontinuity in forecast errors
- Rounding of EPS numbers strongly interferes with distributional
  approach
Contact

Centre for Business Research
University of Cambridge
http://www.cbr.cam.ac.uk

Henry Lahr
Email: h.lahr@cbr.cam.ac.uk