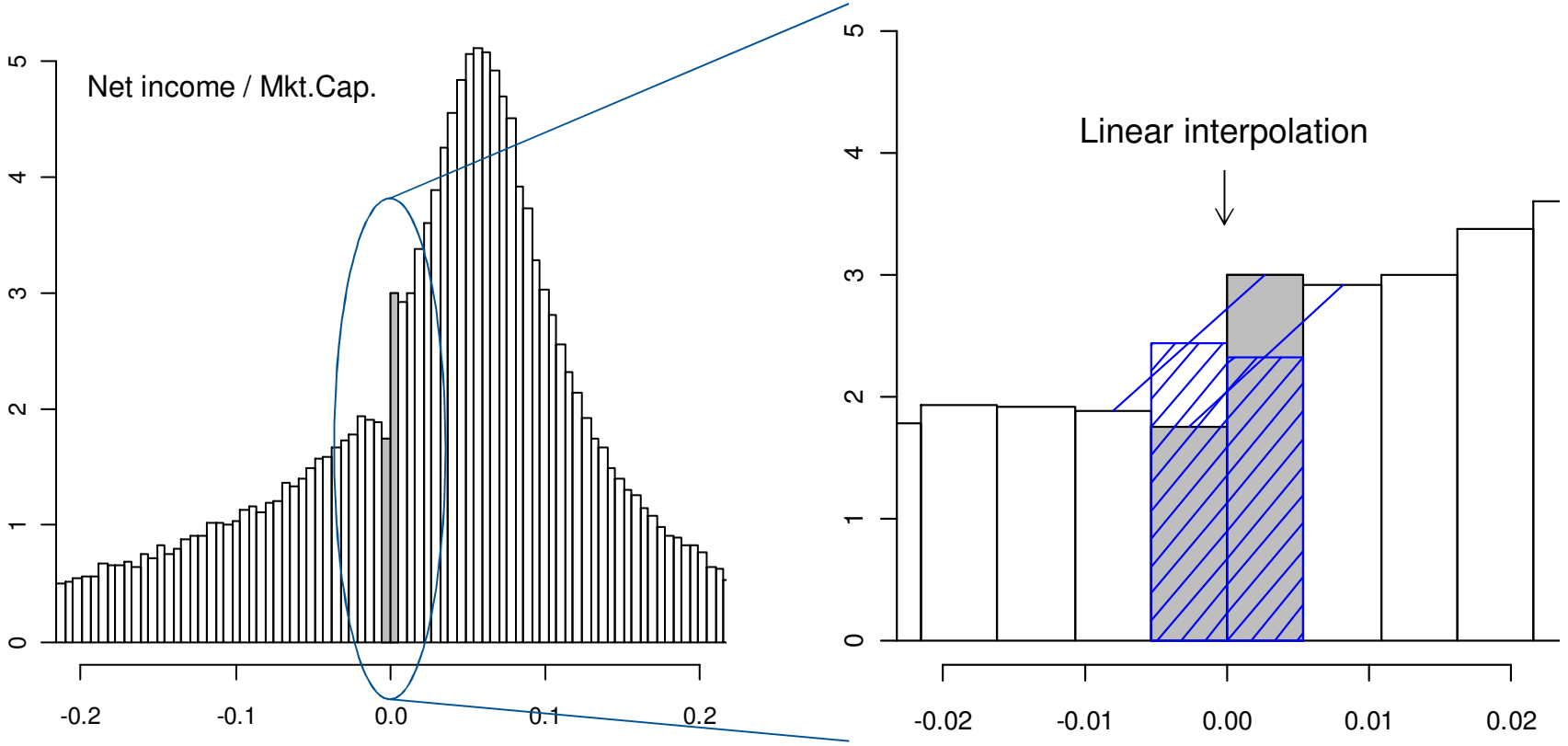
An Improved Test for Earnings Management Using Kernel Density Estimation

Henry Lahr
Centre for Business Research
University of Cambridge

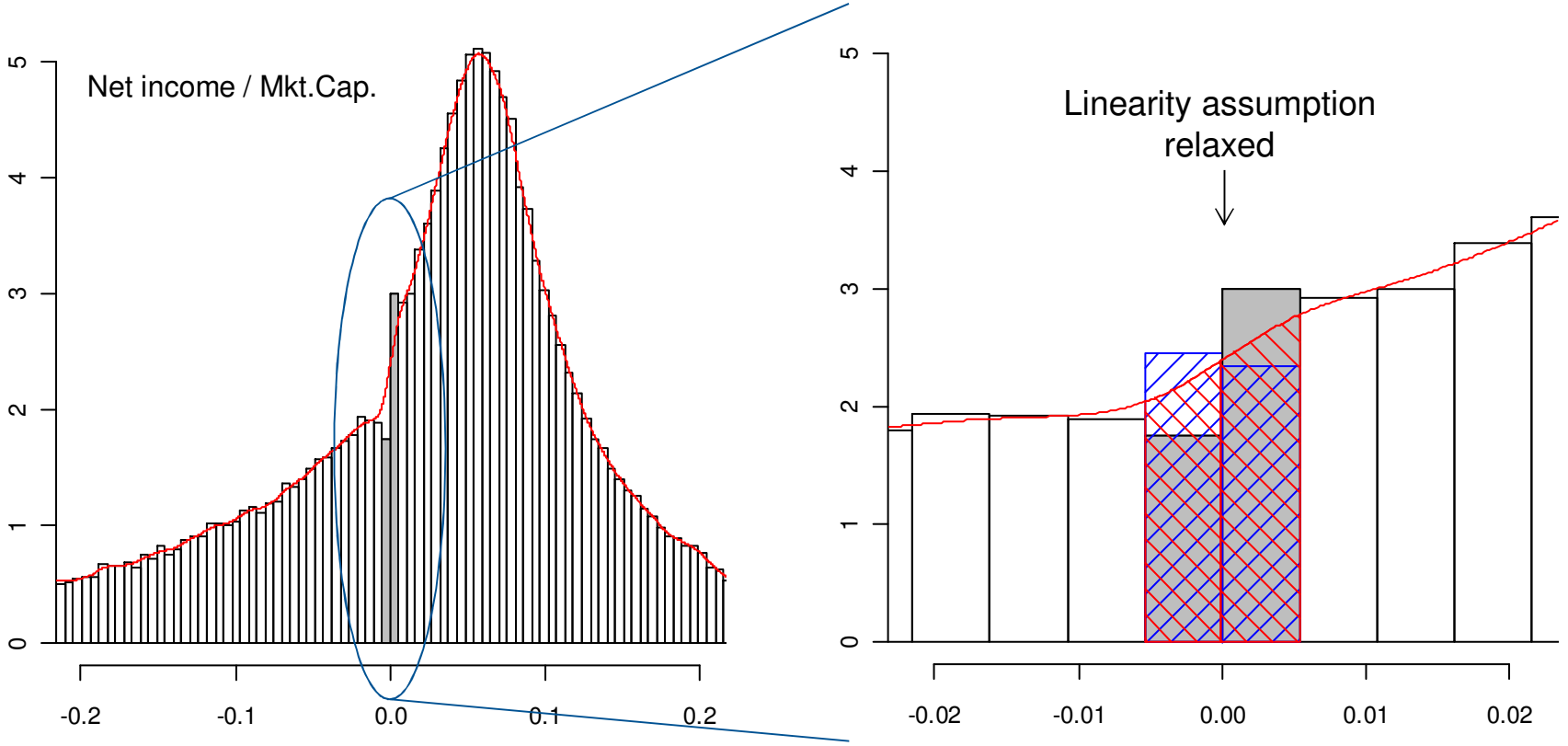
Motivation: Burgstahler / Dichev 1997

- Linear interpolation of histogram bins; t-test against empirical data
- Bin width and location (almost) arbitrary
- Distributional assumption restrictive



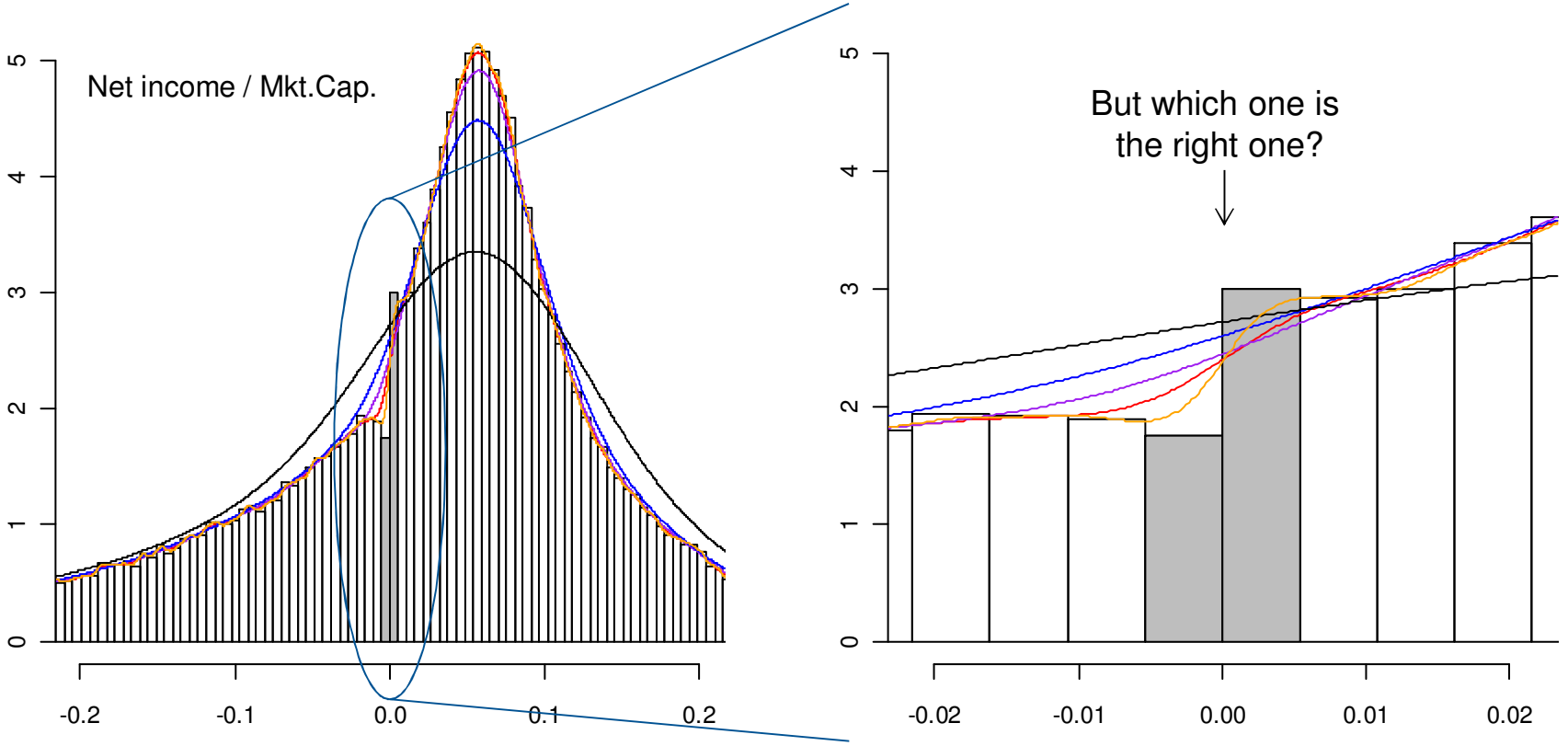
Relaxing the distributional assumption: Kernel density estimation

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



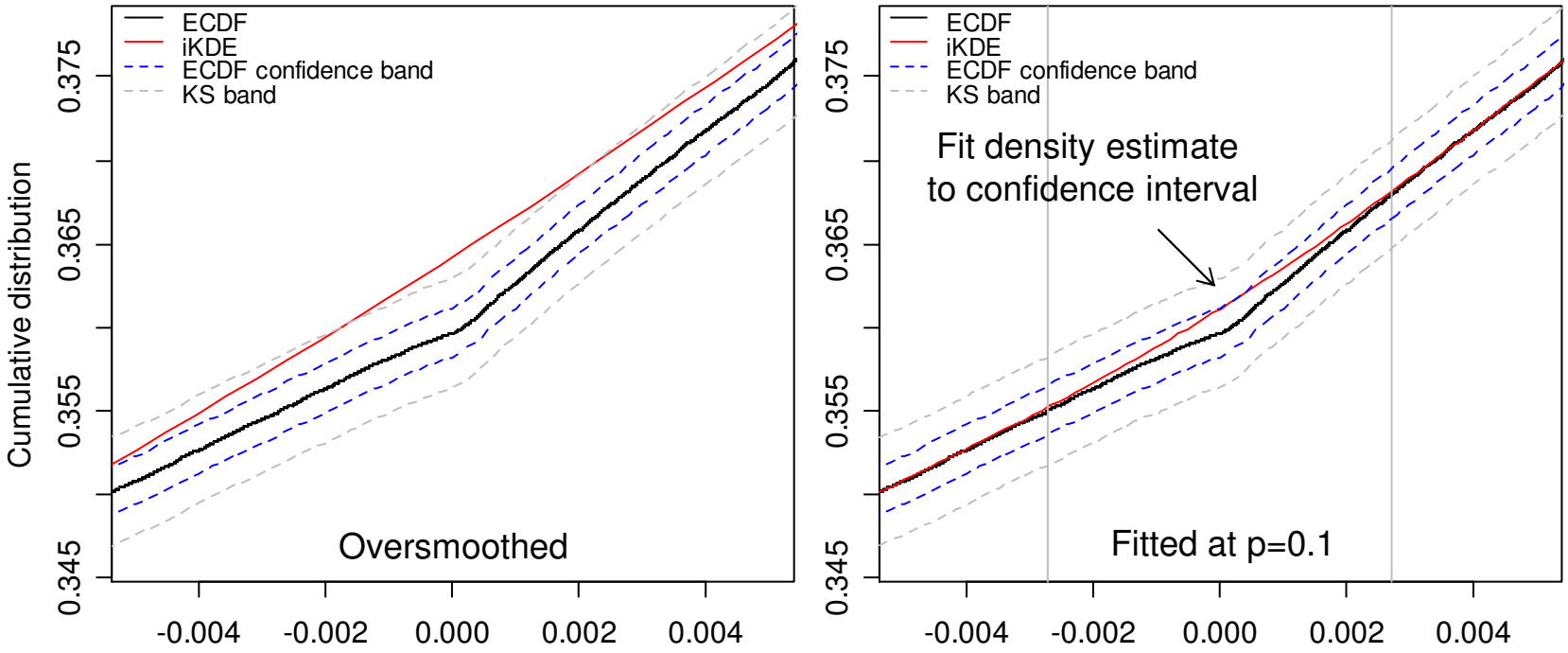
Testing against a reference distribution

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



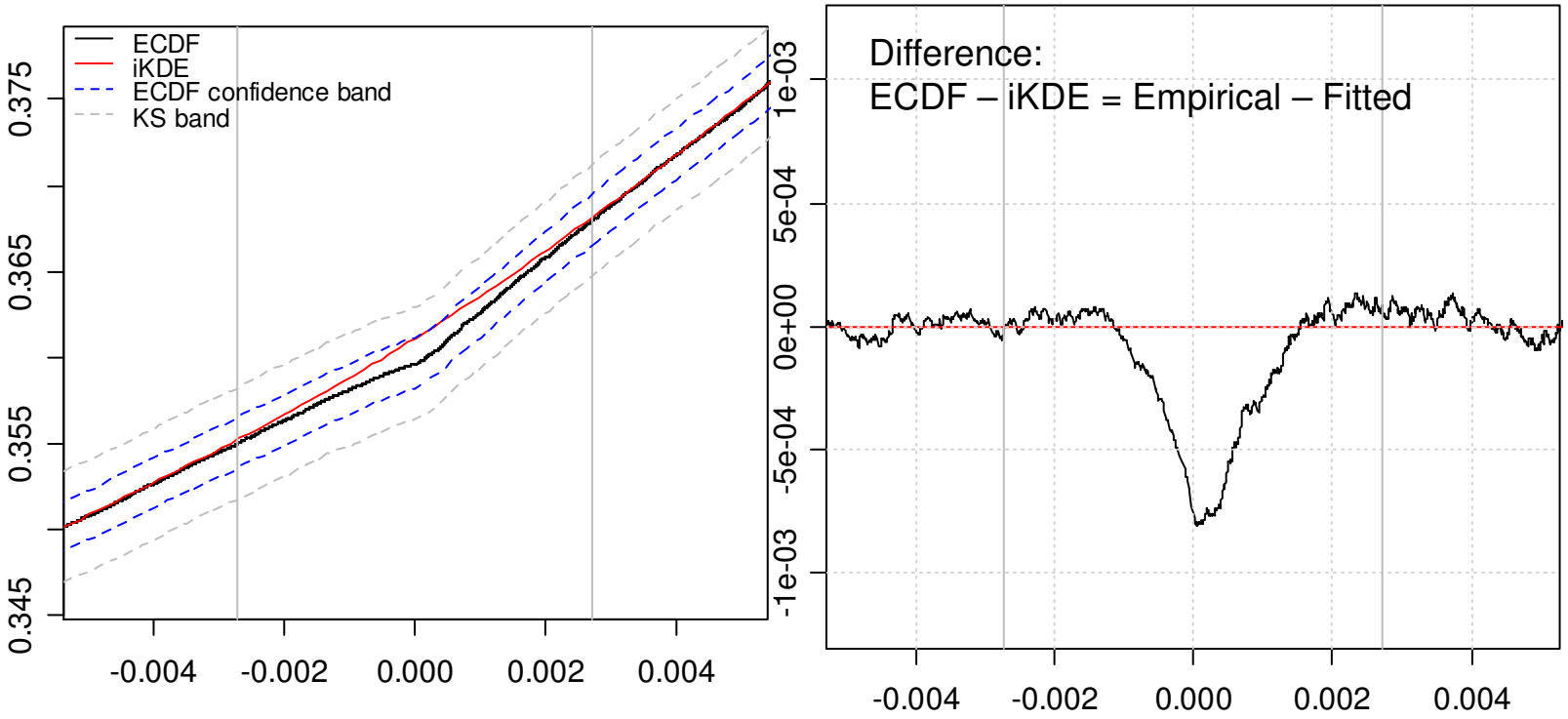
Constructing the reference distribution

- 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



Finding discontinuities

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



Data

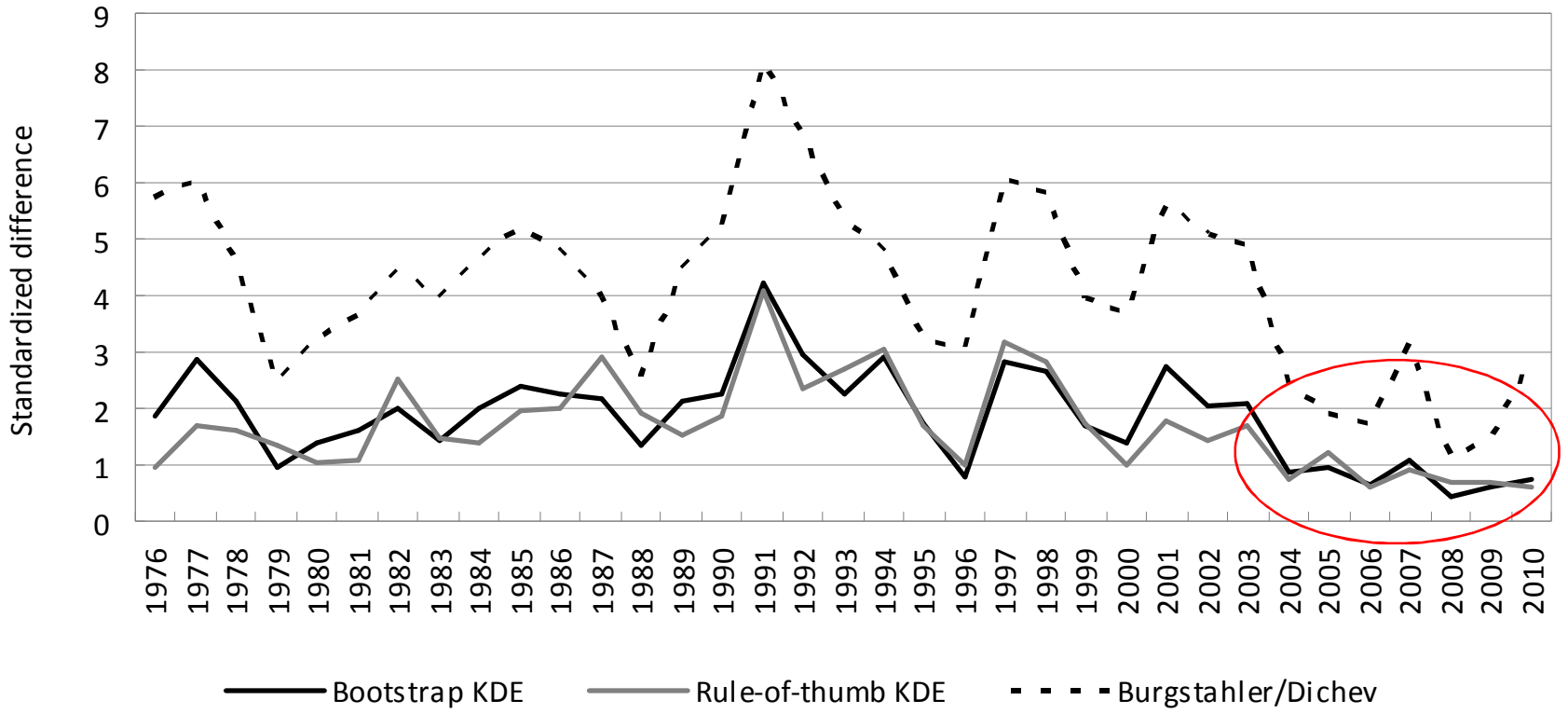
- US company accounts from Compustat
- Financial institutions excluded
- 1976-2010
- Net income, lagged market capitalisation
- N=174009 (standardised earnings), N=163664 (earnings changes)

- 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

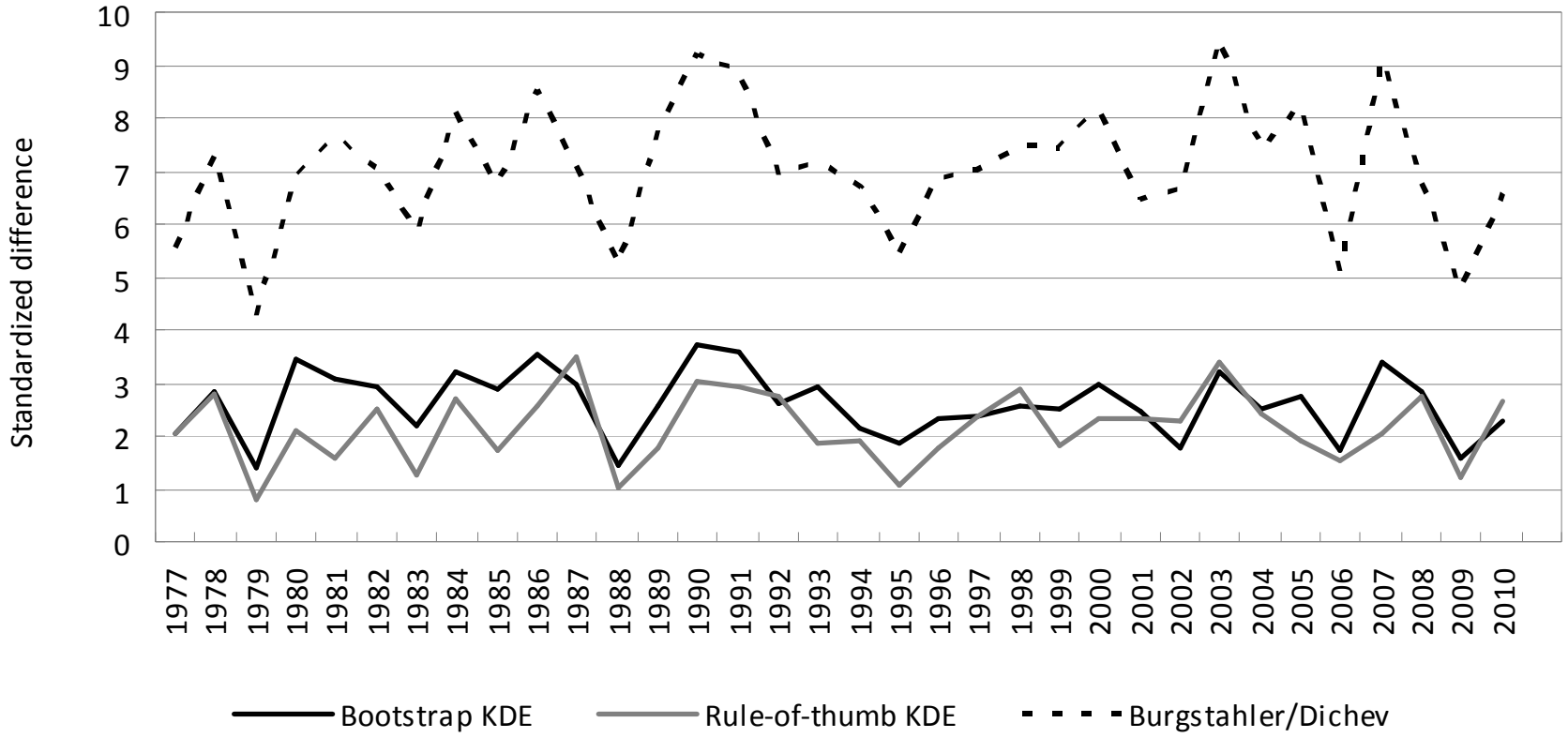
Standardised earnings (Net income$_t$ / Market cap$_{t-1}$)

		Bootstrap KDE			Rule-of-thumb KDE		Burgstahler/Dichev	
Period/N	Interval	Std.Diff.	P(normal)	P(bin.)	Std.Diff.	P(normal)	Std.Diff.	P(normal)
Epanechnikov kernel								
			h=0.005739		h=0.010622		h=0.005739	
1976--1994	]-h,0]	-6.542	0.000	0.000	-8.311	0.000	-14.292	0.000
N=78923	]0,h]	4.593	0.000	0.000	6.277	0.000	10.023	0.000
			h=0.011515		h=0.011474		h=0.011515	
1995--2010	]-h,0]	-3.179	0.001	0.001	-3.215	0.001	-6.449	0.000
N=96426	]0,h]	2.250	0.024	0.025	2.200	0.028	3.090	0.002
			h=0.005385		h=0.009622		h=0.005385	
1976--2010	]-h,0]	-5.942	0.000	0.000	-7.221	0.000	-12.415	0.000
N=175349	]0,h]	5.038	0.000	0.000	5.897	0.000	10.006	0.000
Gaussian kernel								
			h=0.002670		h=0.010622		h=0.002670	
1976--1994	]-h,0]	-7.098	0.000	0.000	-12.345	0.000	-10.240	0.000
N=78923	]0,h]	6.261	0.000	0.000	8.597	0.000	7.571	0.000
			h=0.004986		h=0.011474		h=0.004986	
1995--2010	]-h,0]	-3.641	0.000	0.000	-5.470	0.000	-4.715	0.000
N=96426	]0,h]	3.165	0.002	0.002	1.692	0.091	3.891	0.000
			h=0.002474		h=0.009622		h=0.002474	
1976--2010	]-h,0]	-6.693	0.000	0.000	-11.243	0.000	-9.433	0.000
N=175349	]0,h]	6.276	0.000	0.000	7.547	0.000	7.981	0.000

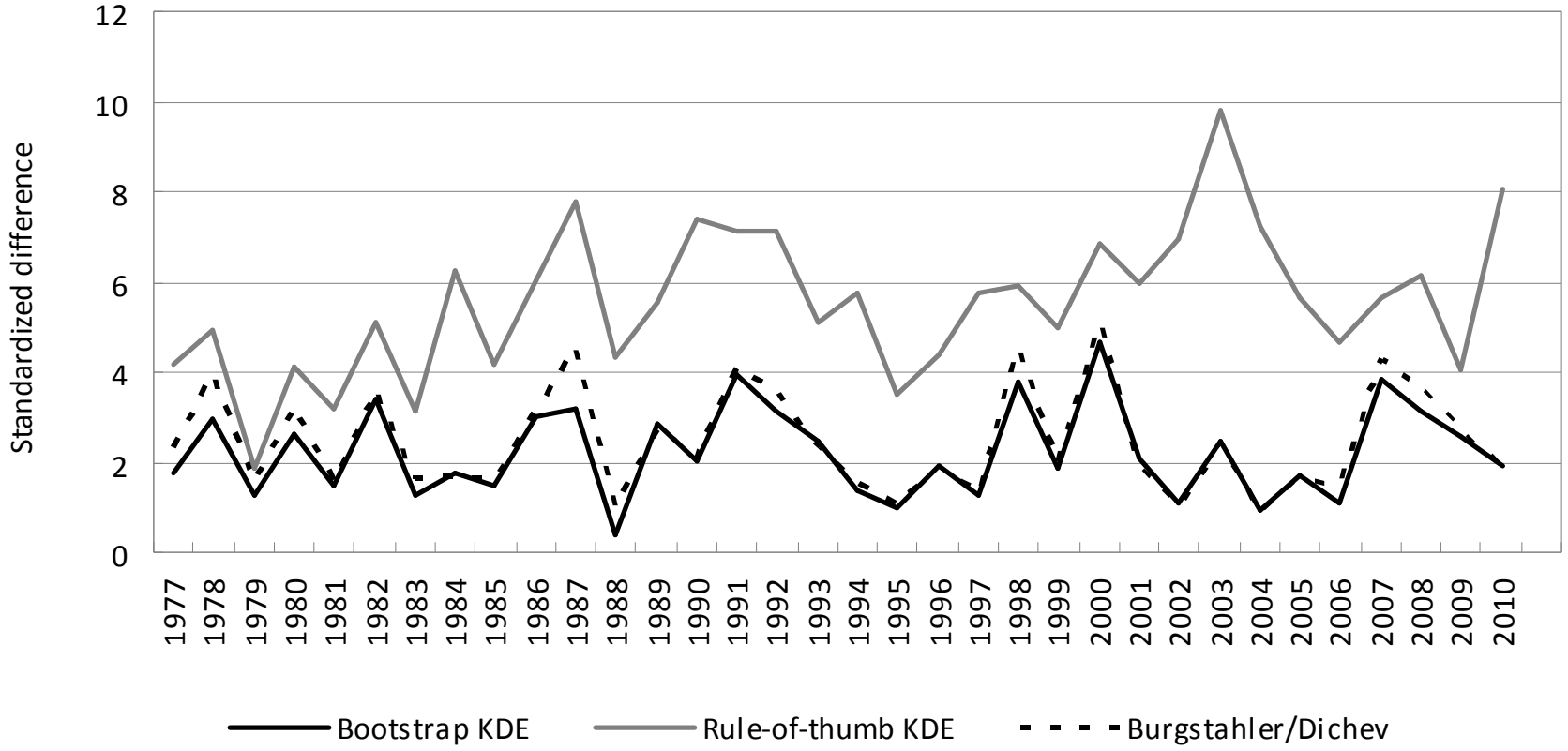
Standardised earnings over time (Epanechnikov kernel)



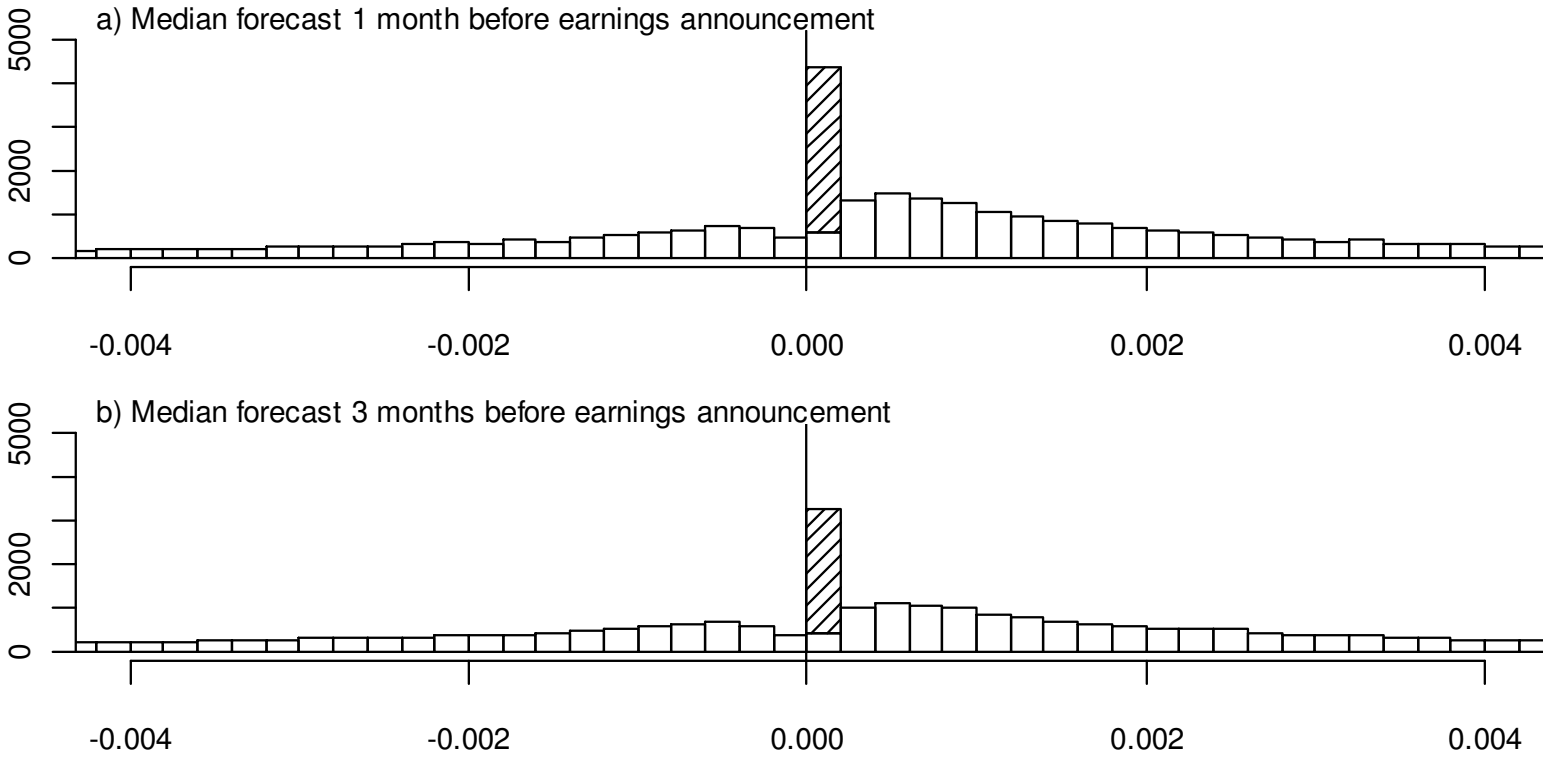
Earnings changes over time (Epanechnikov kernel)



Earnings changes over time (Gaussian kernel)



EPS forecast errors (scaled by market cap.)

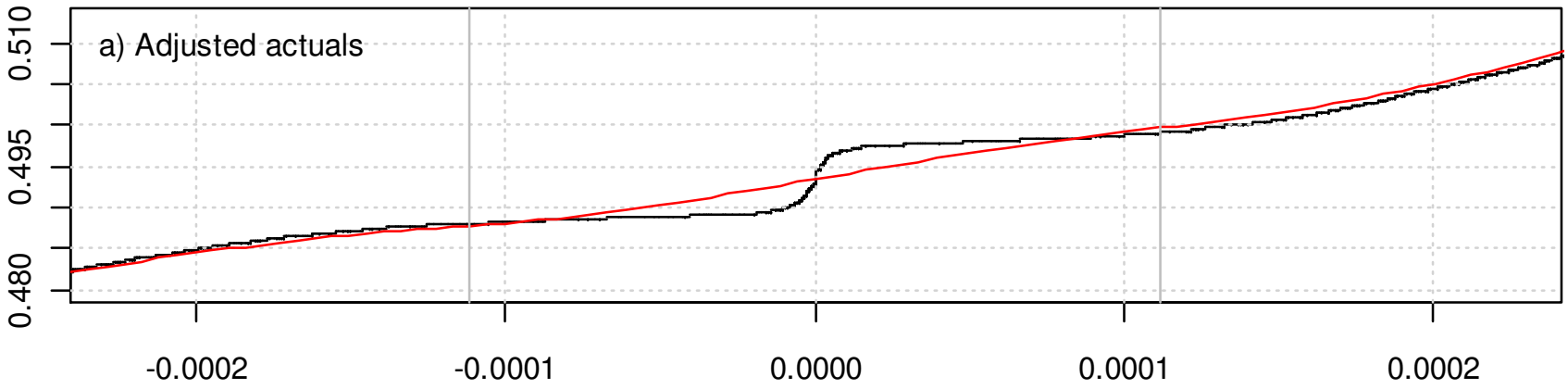


→ Not possible to construct reference distributions including zeroes

Note: Forecast errors based on "unadjusted" actuals, scaled by market capitalisation in t-1 year

EPS forecast errors (“adjusted” actuals)

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

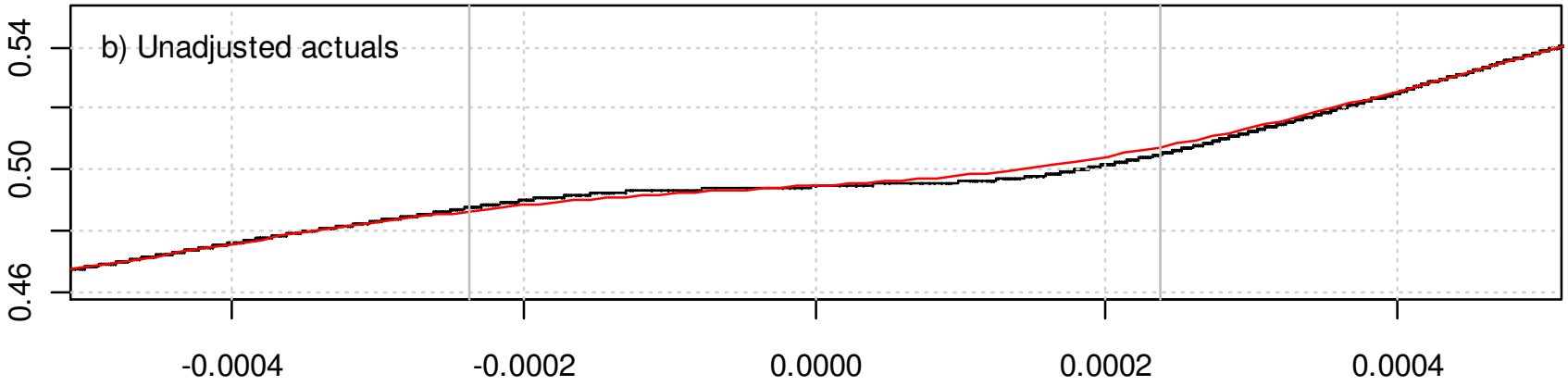


- Bootstrap KDE does **not** detect any discontinuity
- Rule-of-thumb KDE significant for **all** forecast horizons
- Burgstahler/Dichev test significant for **all** forecast horizons

→ 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)

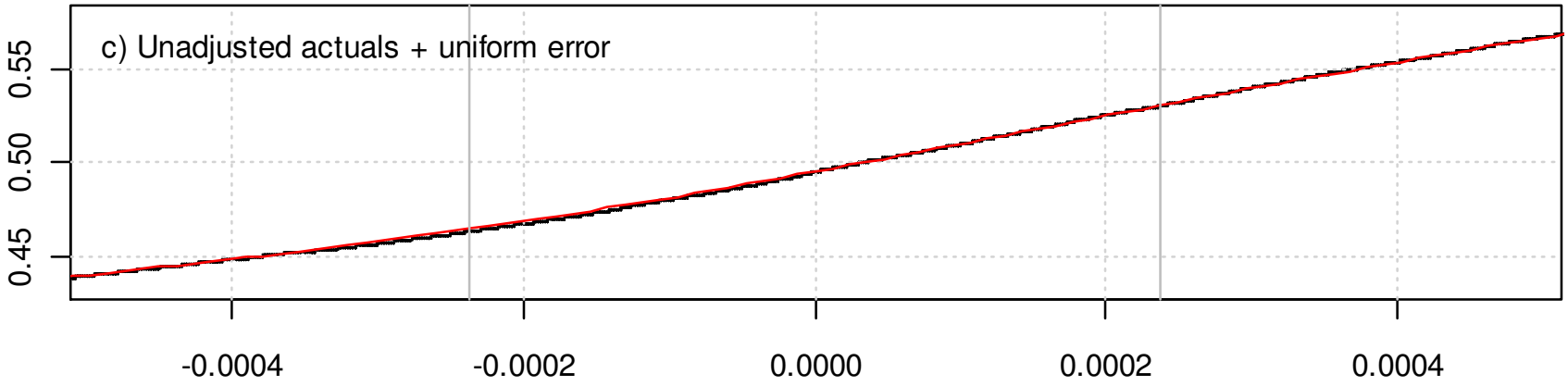
$$\text{Error}_{t^*/t} = (\text{UnadjActual}_t \times \text{AdjFac}_{t^*}/\text{AdjFac}_t - \text{MedEst}_{t^*}) \times \text{Shares}_{t^*}/\text{MktCap}_{t-1}$$



- 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)

$$Error_{t^*/t} = (UnadjActual_t \times AdjFac_{t^*}/AdjFac_t - MedEst_{t^*} + e) \times Shares_{t^*}/MktCap_{t-1}$$



- 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