Measuring economic uncertainty using news-media textual data

Peter Eckley

Bank of England

Motivation

- Uncertainty affects economic decision-making
- But is unobservable, and existing proxies have drawbacks
- News-media measures promising (Baker, Bloom, Davis, 2015) but need further development and characterisation

Contribution of this paper

Theory of measurement

Refined methodology

Comparison to stock volatility

Conclusions

Theory of measurement

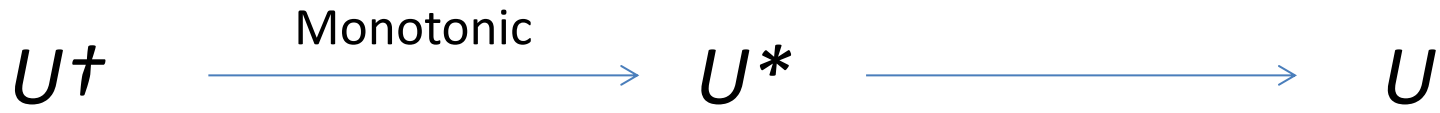
Refined methodology

Comparison to stock volatility

Conclusions

$$U \equiv \frac{\text{\# of articles containing keyphrase(s)}}{\text{total \# of articles in FT}}$$

What is *U* measuring?



Intensity of the **cognitive state** of uncertainty

Propensity to <u>express</u> uncertainty in natural language

Frequency of <u>expressions</u> of uncertainty in natural language

U as an unbiased, consistent estimator of U*
Articles = independent Bernoulli trials
Success = article expresses uncertainty
Pr(Success) = U*
→ *U* ~ scaled Poisson binomial with mean *U**

Noise-to-signal ratio:

Daily	0.87
Weekly	0.21
Monthly	0.05
Annual	0.01

Theory of measurement

Refined methodology

Comparison to stock volatility

Conclusions

Uncertainty keyphrases: baseline

- {“uncertain”, “uncertainty”, “uncertainties”}
- Including “uncertainties” boosts signal by 14% without material noise increase
- Measurement error should be modest *a priori*
  - Fairly compact semantic space
  - Semantics stable over time
  - Negation is rare because already self-negated

Uncertainty keyphrases: extensions?

Added keyphrase	Gain in U	Correlation with baseline (monthly)
unclear*	23%	0.51
not clear*	14%	no data
unpredictabl*	7%	0.59
not sure*	7%	0.47
unsure*	4%	0.45
not certain*	1%	0.38
not predictabl*	0%	0.05
all the above	**54%**	**no data**
risk*	238%	no data

For data availability reasons I stick to “uncertain*”
Advantage of a relatively ‘clean’ measure

Isolating economic uncertainty

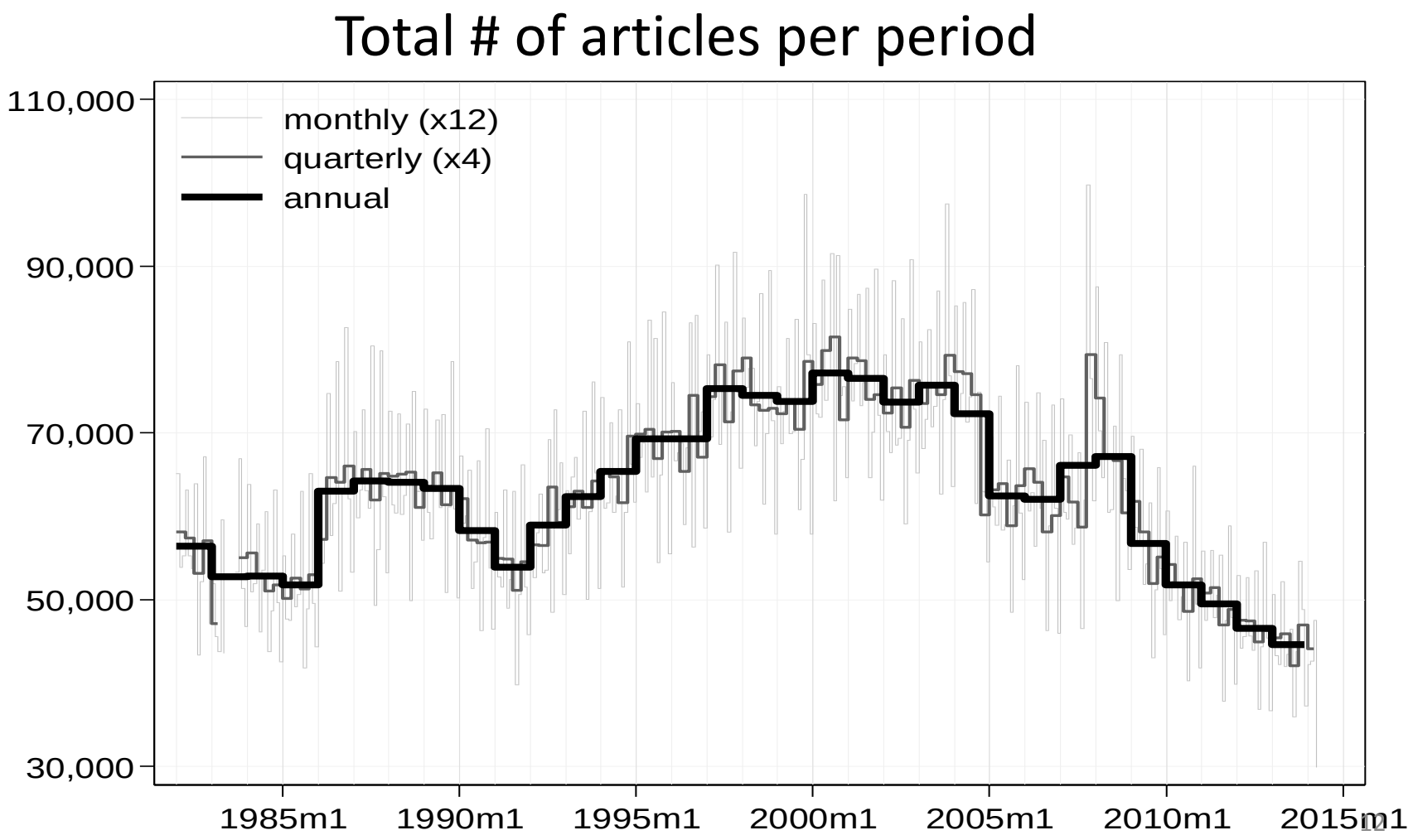
- Literature using generalist newspapers has to filter out non-economic articles
  - Typically only use articles containing “economic” or “economy”
- Unnecessary with FT
  - Almost all articles have an economic angle
  - May even give stronger signal than subset of articles from multiple generalist papers

% of “uncertain*” articles that do NOT contain “economic” or “economy” but DO contain…	
financ*	66%
bank*	34%
debt*	14%
credit*	15%
bond*	11%
equity | equities	16%
money*	18%
business*	38%
profit*	30%
earnings*	15%
revenue*	11%
wage*	2%
government*	36%
politic*	19%
policy*	13%
Eurozone	2%
oil	10%
gas	6%
coal	2%
any of the above	**95%**

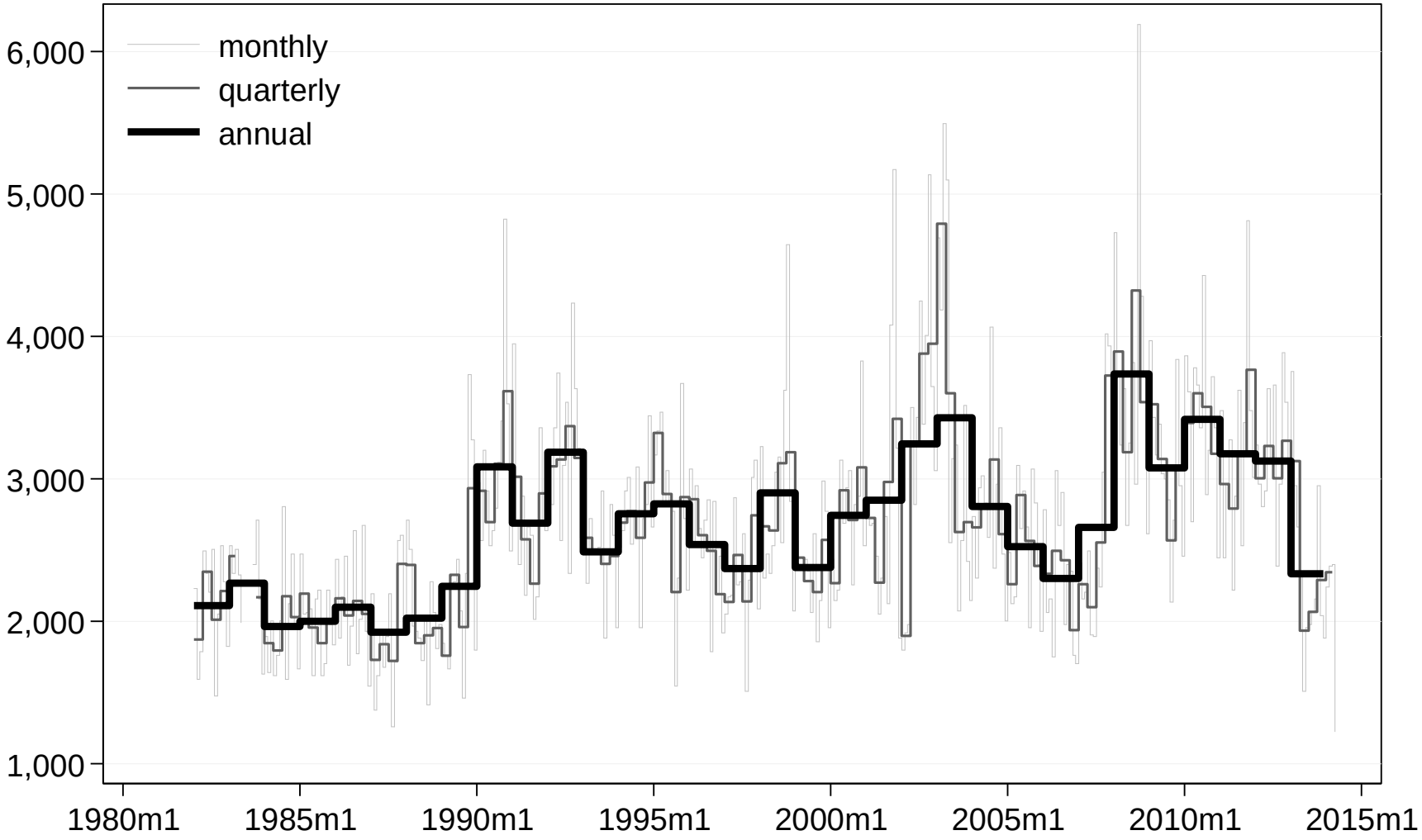
De-duplication

Ratio of daily FT record counts in Factiva vs. Nexis UK

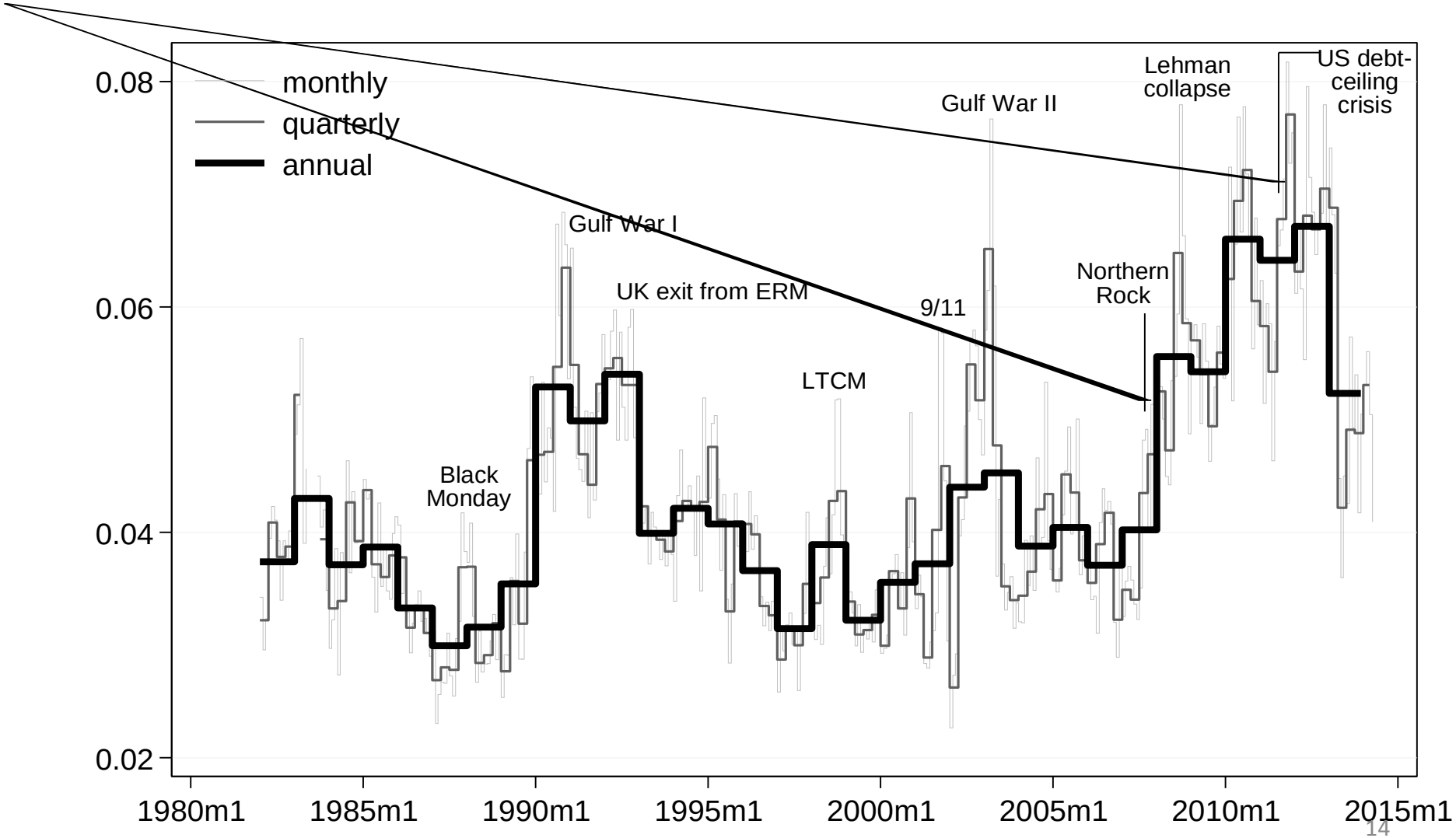
Normalising by news volume



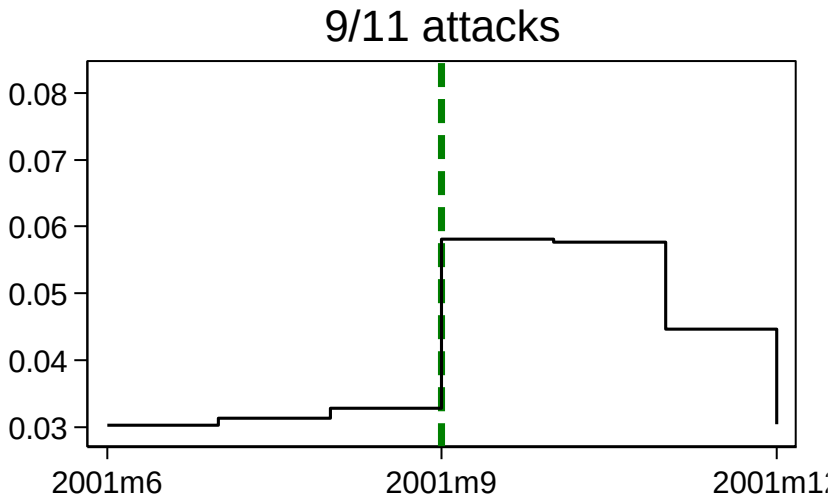
Without normalisation (numerator of *U*)

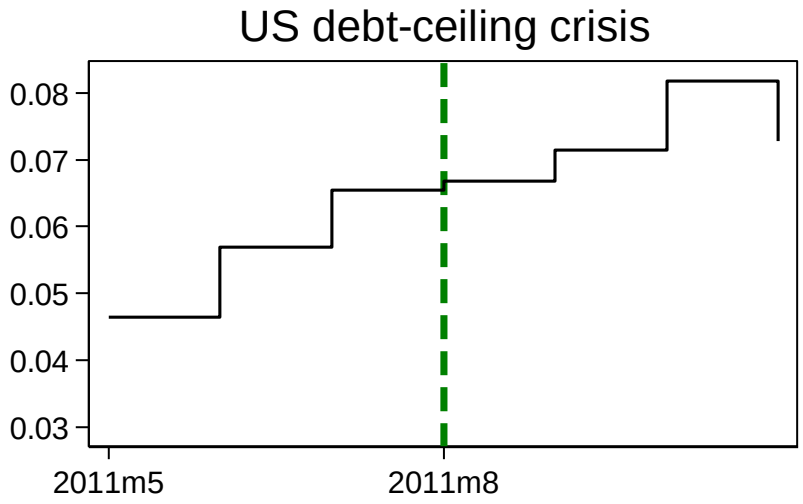


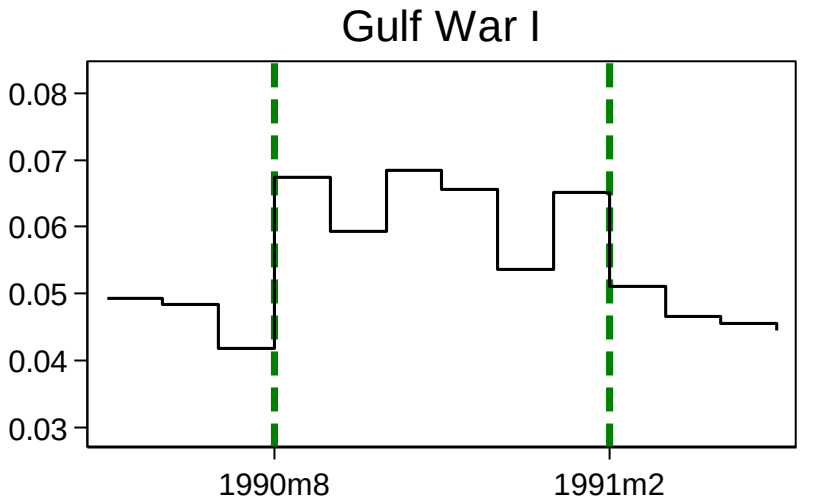
With normalisation (news-media uncertainty, *U*)

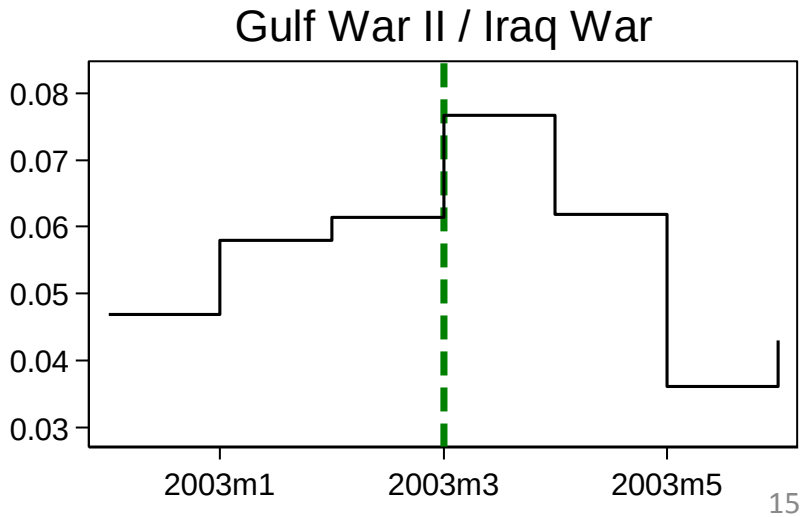


Moves around major narrative events









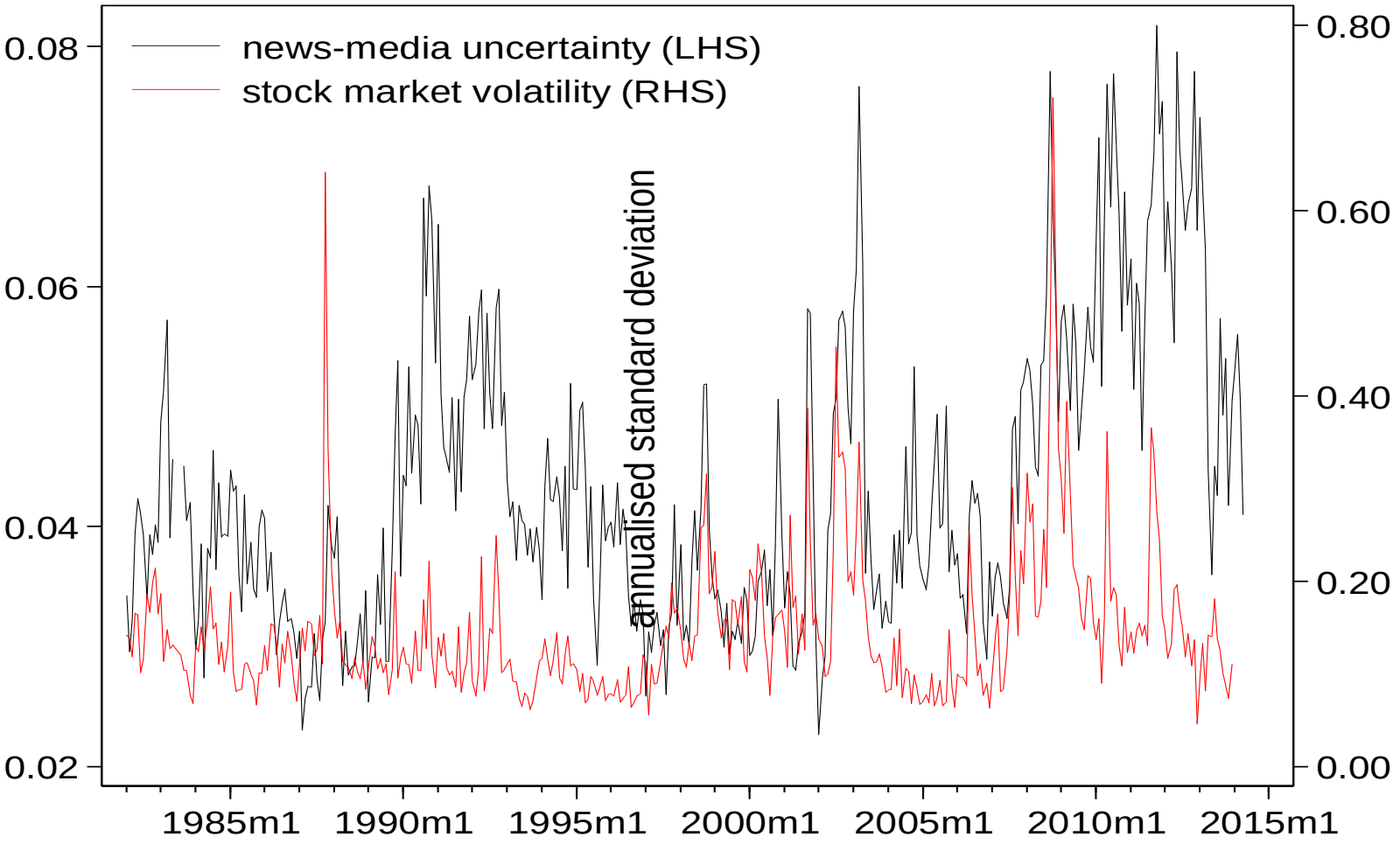
Theory of measurement

Refined methodology

Comparison to stock volatility

Conclusions

U vs. σ


news-media uncertainty (LHS)
stock market volatility (RHS)
annualised standard deviation
0.08
0.06
0.04
0.02
0.80
0.60
0.40
0.20
0.00
1985m1
1990m1
1995m1
2000m1
2005m1
2010m1
2015m1

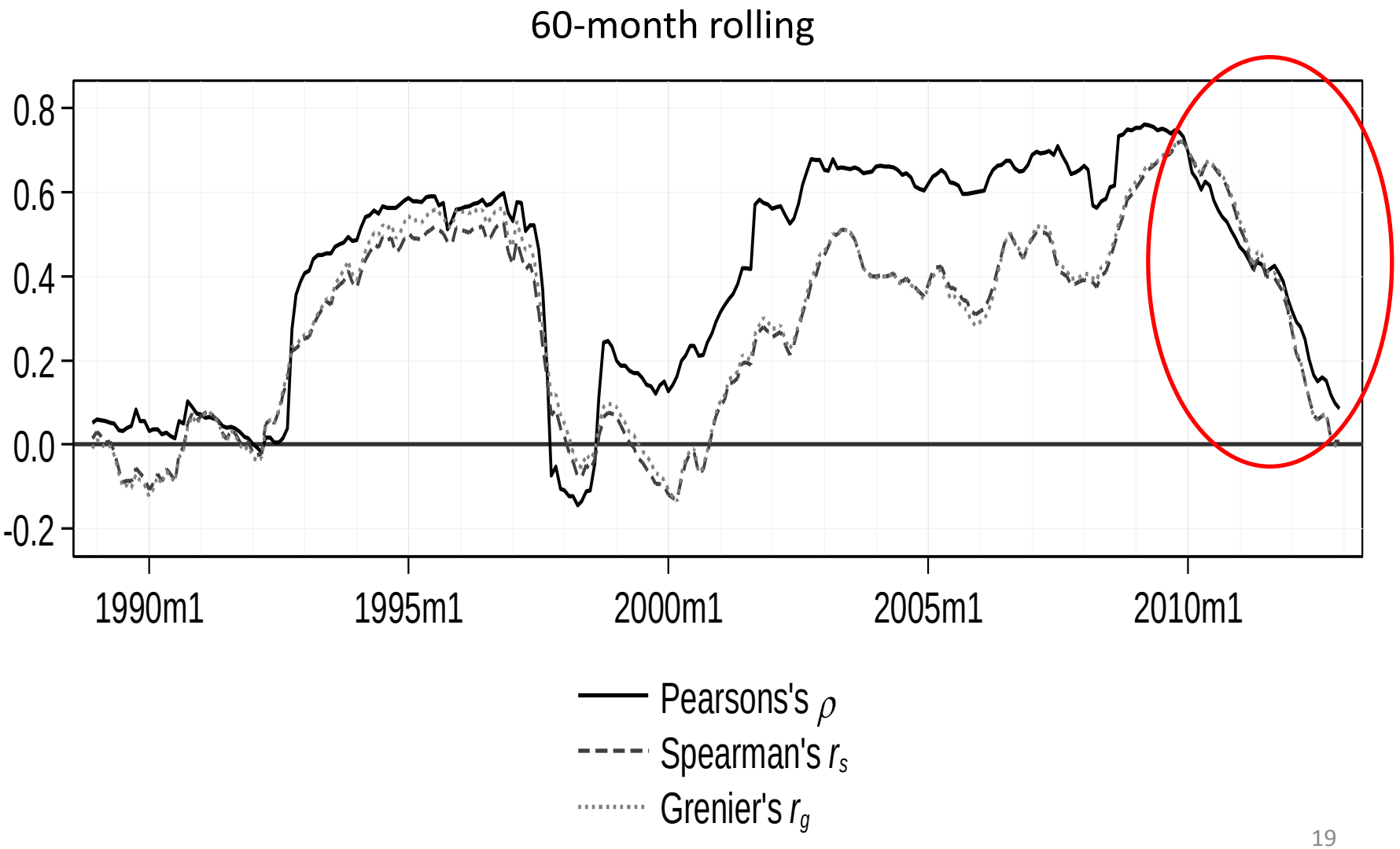
Granger causality tests

- Bivariate daily VAR with RV estimates* of σ (2000-2012)
- $U \nrightarrow \sigma$ [p=0.487 on joint exclusion of all lags]
- $\sigma \rightarrow U$ [p=0.010 on joint exclusion of all lags]
- But markets lead by only one day, consistent with FT being published before markets open

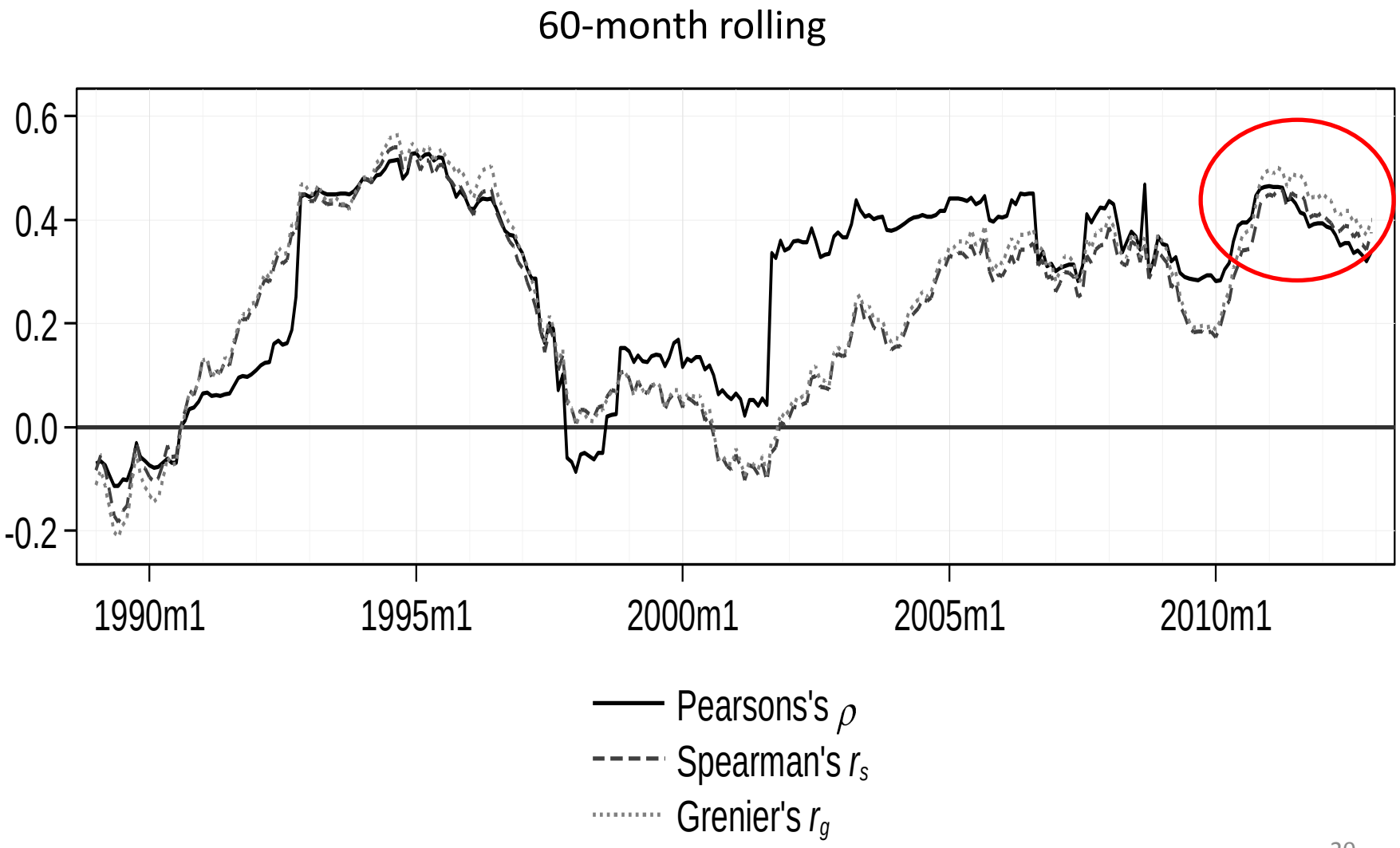
  [p=0.202 for joint exclusion restriction on σ lags 2+]

* Realized volatility (RV) estimates of Heber, Lunde, Shephard, & Sheppard (2009) based on intra-day tick data

Correlation of *U* and *σ*

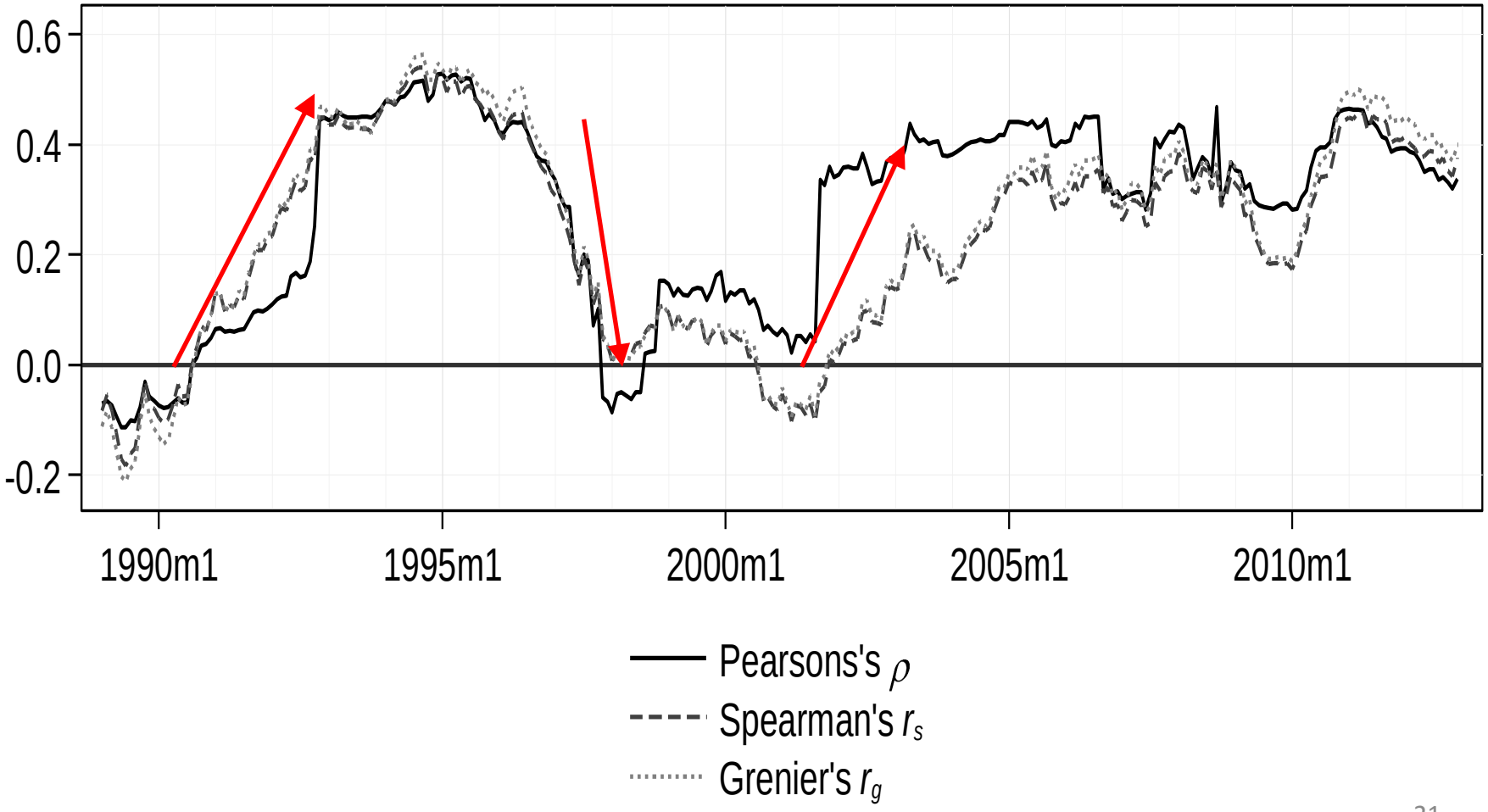


Correlation of ΔU and $\Delta\sigma$



Switching between high and low correlation

60-month rolling



Possible explanation: asymmetry in *U*

- Upside risk not naturally expressed as “uncertainty”
- Whereas σ treats upside and downside symmetrically
- Circumstantial evidence consistent with hypothesis
  - Correlation lower when perceived risk more skewed to the upside, e.g. in booms
  - Correlation higher when downside risk is higher, which under our hypothesis will be when U (or $|\Delta U|$) is larger.

    Consistent with this, correlation is 0.331 for U above its median vs. 0.076 below. Similar result with $|\Delta U|$, and not for placebo with σ above vs. below its median.

Theory of measurement

Refined methodology

Comparison to stock volatility

Conclusions

Conclusions

- Theory of measurement clarifies what we are measuring, and how much noise to expect
- Some simple methodology refinements give material reduction in measurement error
- Others are shown to have a modest impact, which gives greater confidence in the apparently simplistic approach of counting small numbers of keywords
- Relationship between U and σ is time-varying . We proposed an explanation for further testing.

Questions?

Key references

- **Alexopolous & Cohen (2009).** "Uncertain times, uncertain measures." *University of Toronto Department of Economics*
- **Baker & Bloom (2013).** "Measuring Economic Policy Uncertainty." *www.policyuncertainty.com*
- **Bekaert et al (2009).** "Risk, uncertainty and asset prices." *Journal of Financial Economics*
- **Bloom (2014).** "Fluctuations in Uncertainty." *Journal of Economic Perspectives*

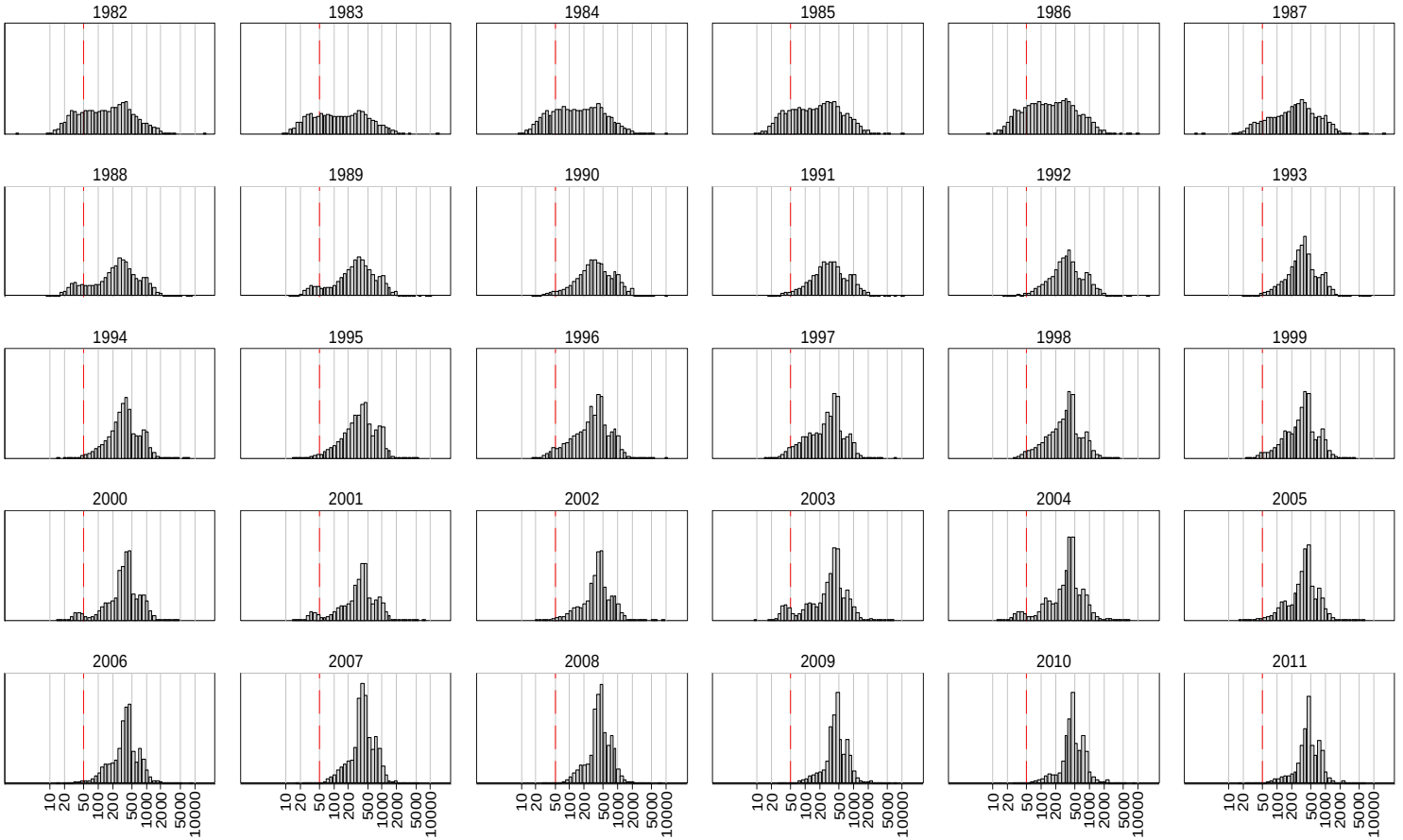
APPENDIX

U as an estimator of *U**: sampling noise in finite samples

- Discretisation grid c. 0.005 at daily frequency (c. 200 articles)
- cf. mean[U] ~ 0.04, sample var[U] ~ 0.01-0.02
- Estimated noise-to-signal ratio (using model-based partition of *U*)
  - Daily: 0.871
  - Weekly: 0.205
  - Monthly: 0.054
  - Annual: 0.006

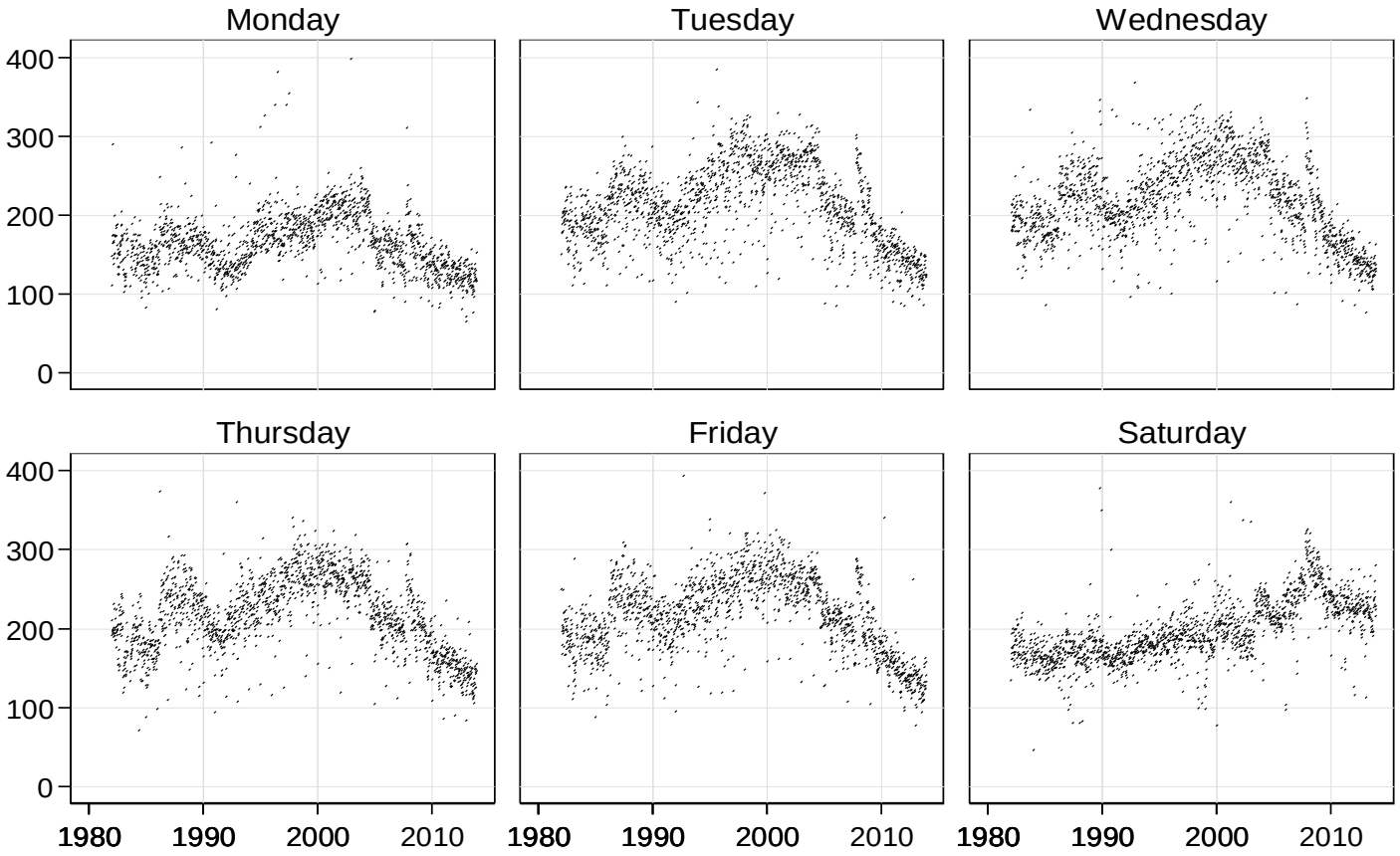
$$\frac{Var_{U-U^*|n}[U-U^*|n]}{Var_{U^*|n}[U^*|n]}$$

Article length: secular variation in the distribution



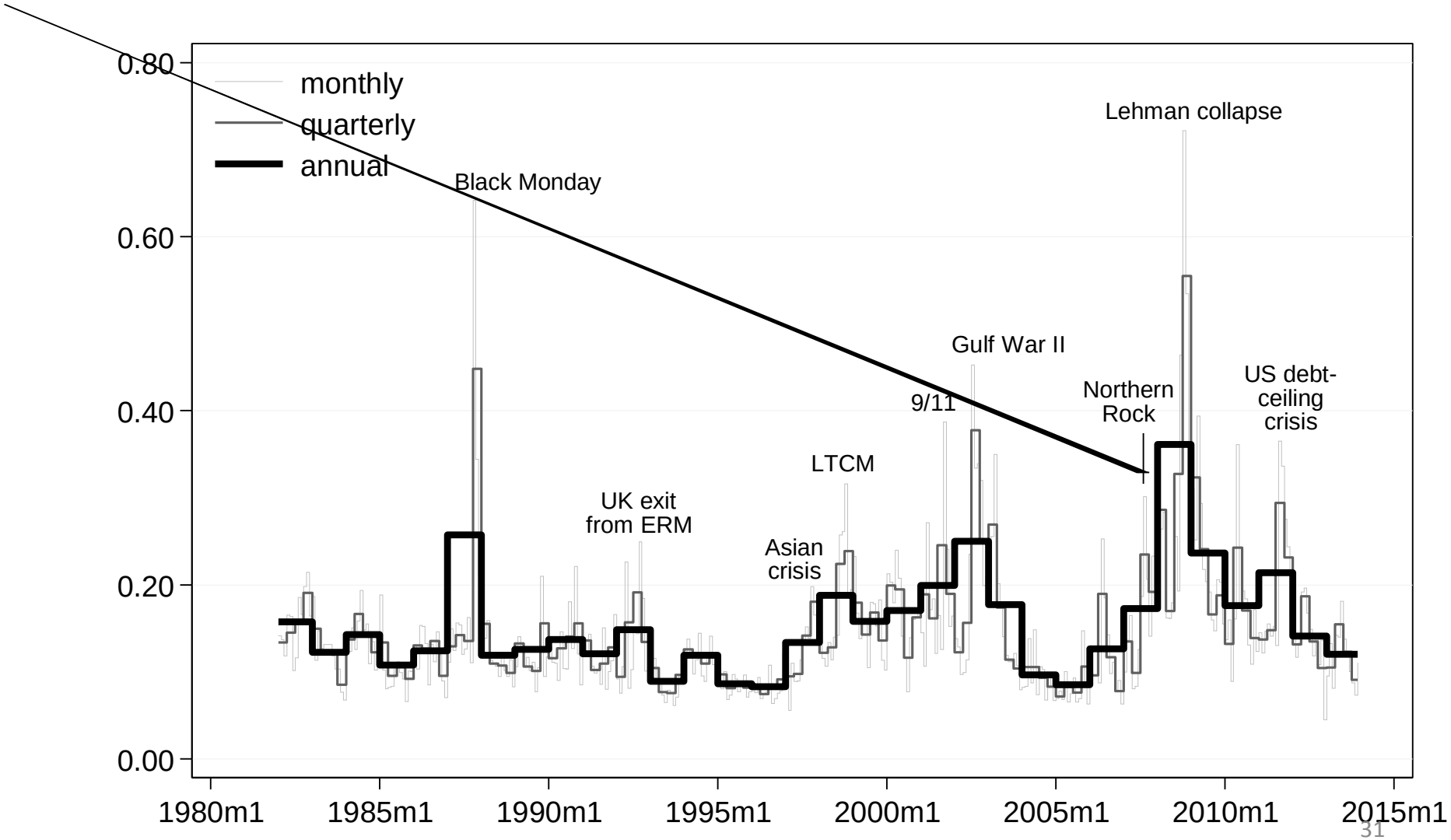
article word count (log scale)

News volumes: day-of-week seasonality

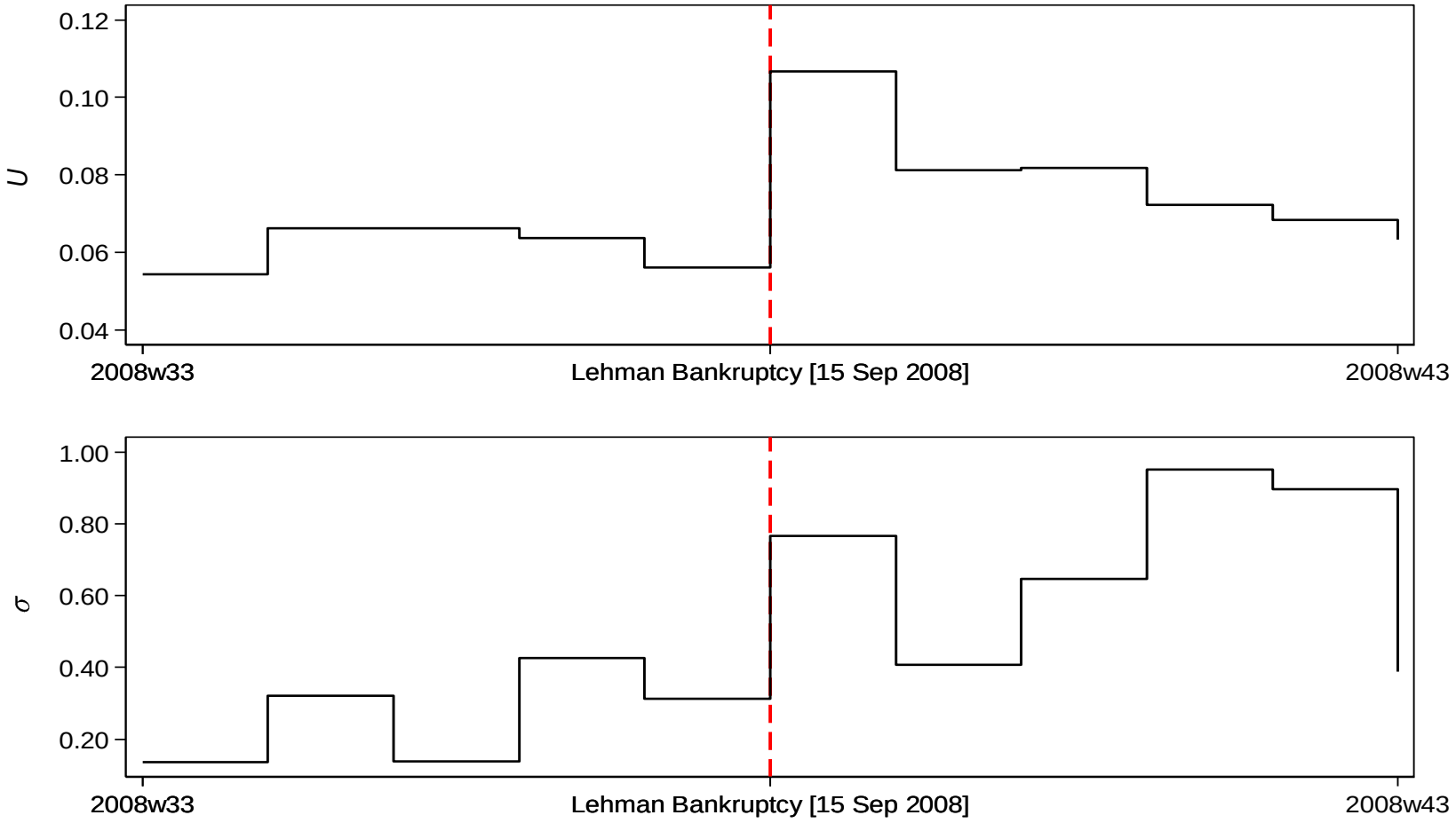


Note: observations with more than 400 articles per days were truncated from the plot.

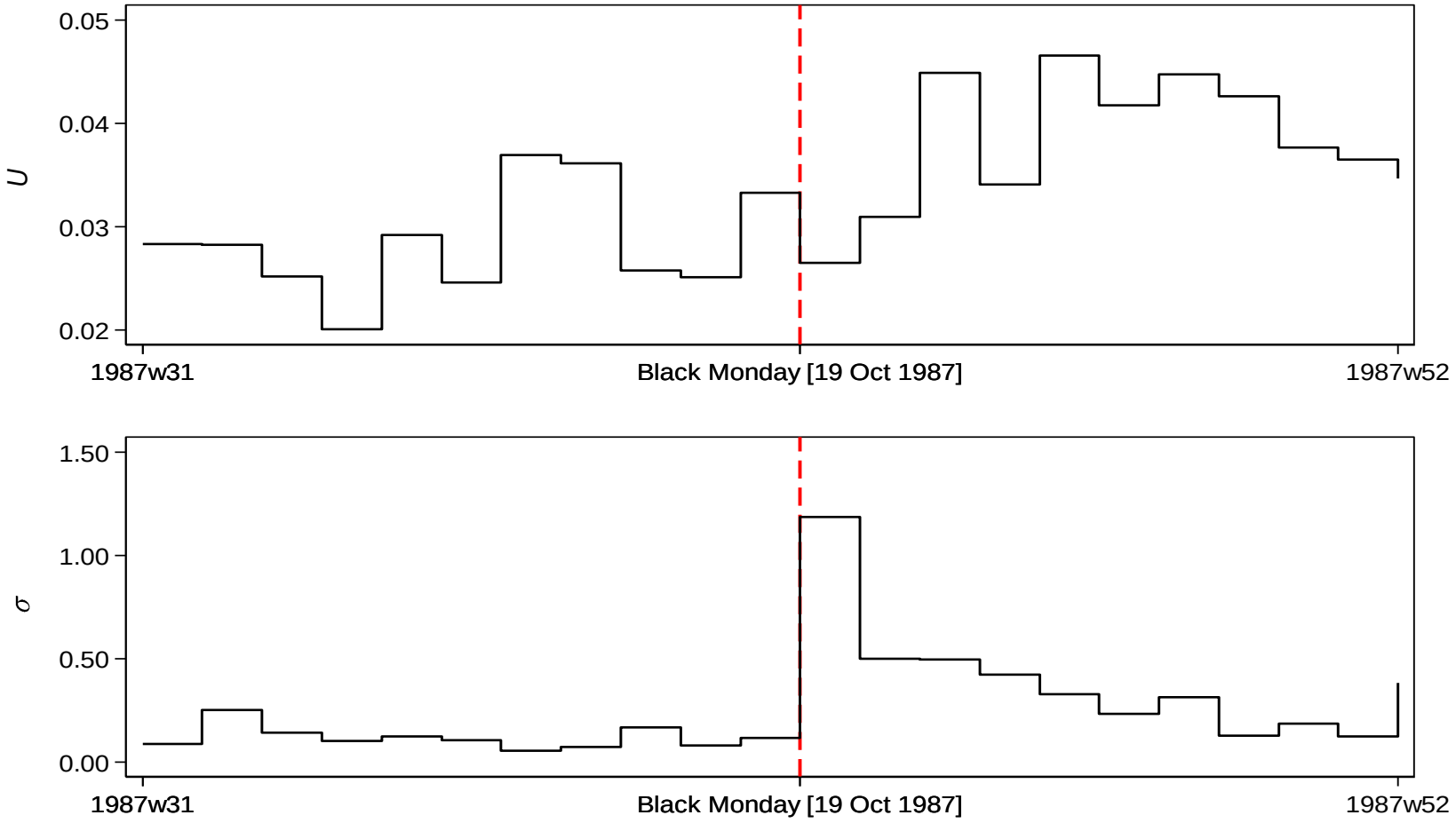
Stock returns volatility, σ



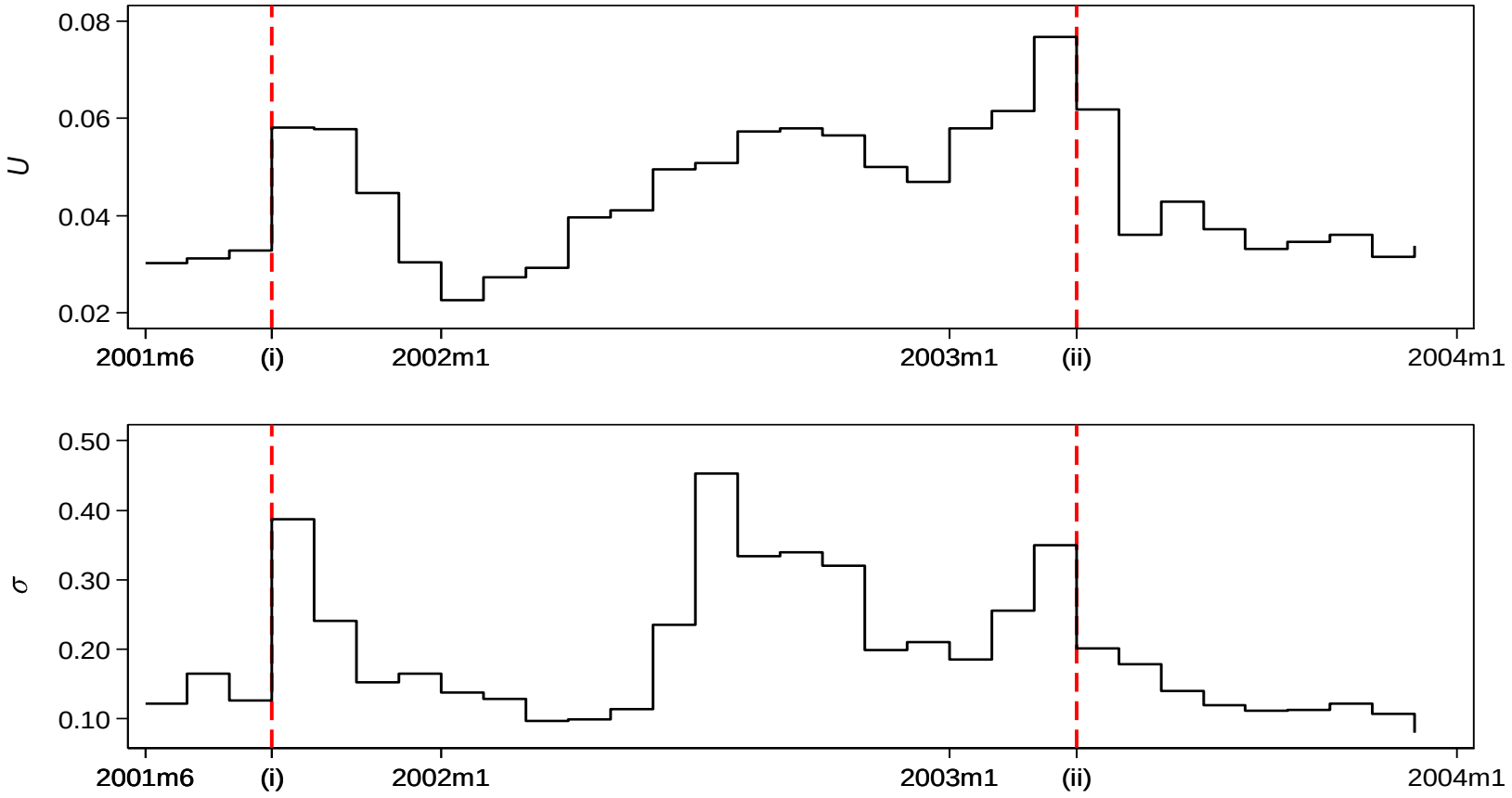
Lehman collapse 2008



Black Monday 1987

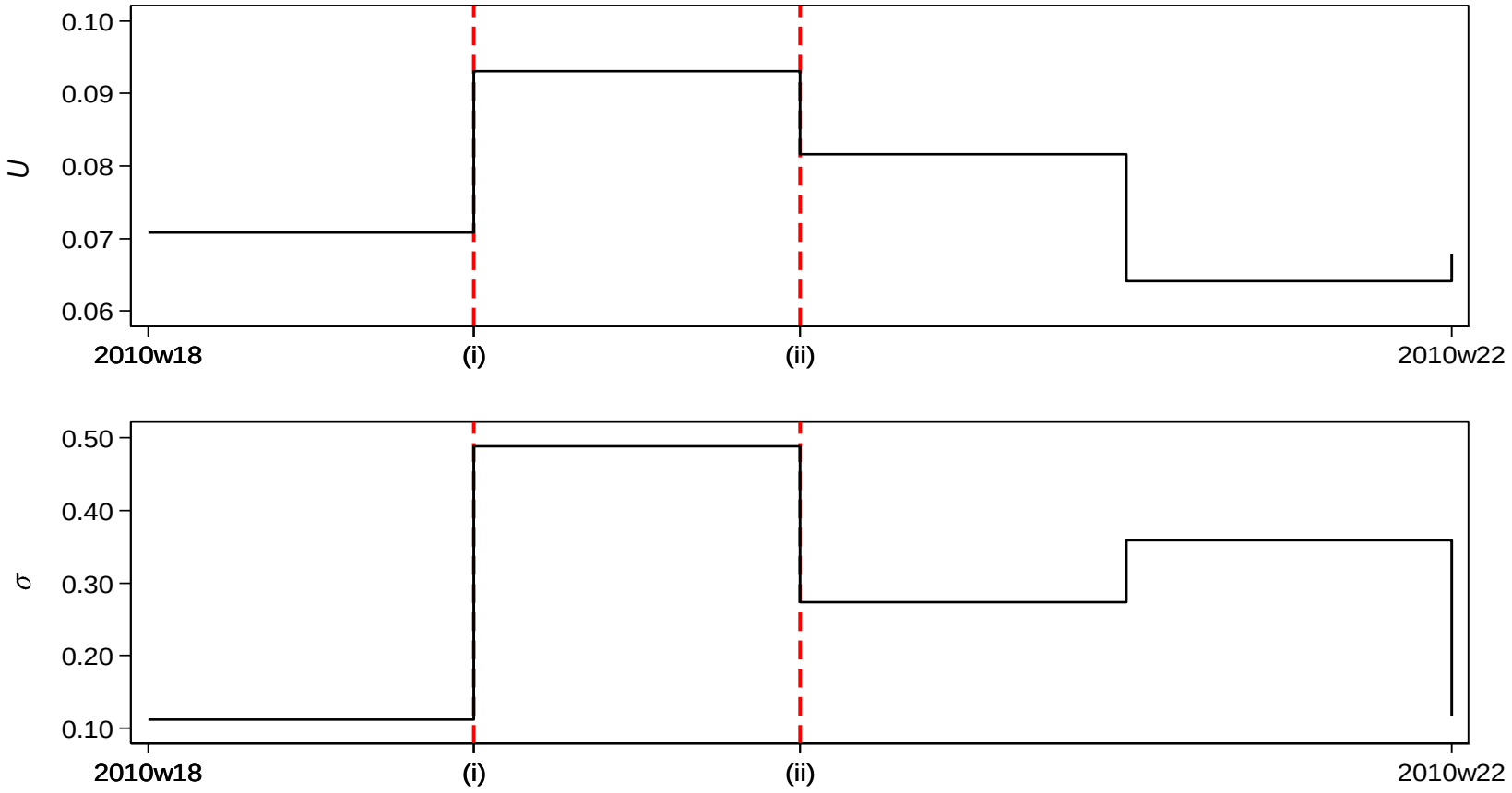


9/11 and Iraq War



(i) Terrorist attacks in US [9 Sep 2001]
(ii) Iraq War begins (20 Mar 2003)

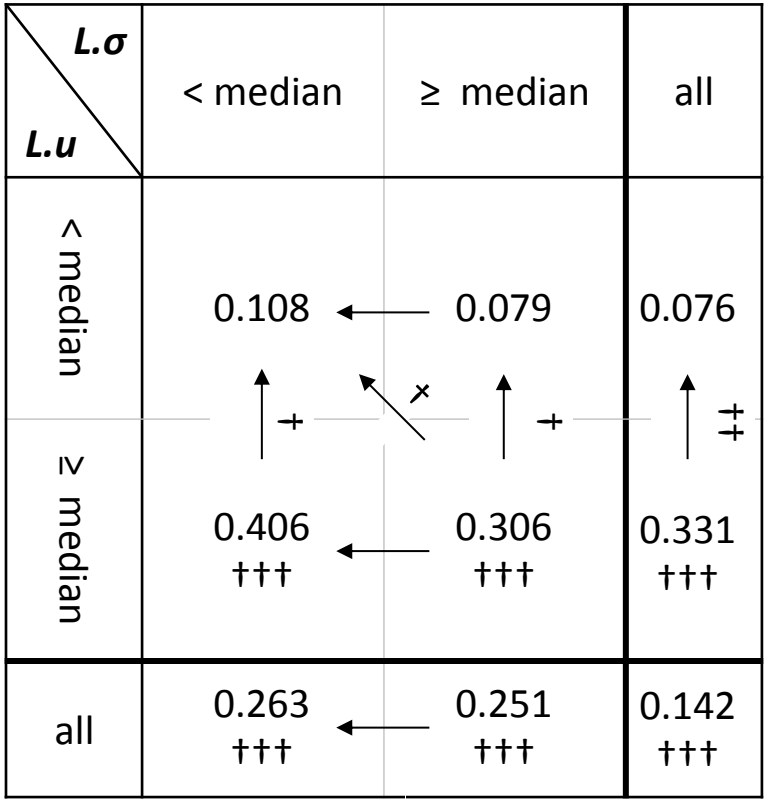
UK hung parliament May 2010



(i) UK national elections result in hung parliament for first time since 1974 [6 May 2010]
(ii) Coalition government formed [12 May 2010]

Rank correlation segmented by levels

monthly differences, 1984-2012



Note: significance levels estimated using block bootstrap to account for autocorrelation

Sign concordance segmented

monthly differences, 1984-2012

L.σ / *L.u*	< median	≥ median	all
< median	0.539	0.437 xx	0.497 +++
≥ median	0.606 ††	0.670 †††	0.644 †††
all	0.566 †	0.575 ††	0.571 †††

|*Δσ*| / |*Δu*|	< median	≥ median	all
< median	0.495 +	0.500 xxx +++	0.497 +++
≥ median	0.600 †	0.681 †††	0.644 †††
all	0.543	0.598 †††	0.571 †††

Note: significance levels estimated using block bootstrap to account for autocorrelation