Measuring economic uncertainty using news-media textual data

Peter, Eckley

University of Oxford, Bank of England

30 January 2015

Measuring economic uncertainty using news-media textual data

An improved methodology and a detailed comparison to stock returns volatility

Draft – comments welcome

1 March 2016

Peter Eckley, Department of Economics, University of Oxford

(Supervised by Professor Christopher Bowdler, Department of Economics, University of Oxford)

This work was supported by the Economic and Social Research Council grant number ES/I903887/1.

I thank seminar participants at the 9th International Conference on Computational and Financial Econometrics (CFE 2015), the Bank of England, University of Oxford for comments.

Correspondence address: peter.eckley@gmail.com

Abstract: We develop a news-media textual measure of aggregate economic uncertainty – the fraction of Financial Times articles containing uncertainty-related keyphrases – for 1982–2014 at daily to annual frequencies. We contribute to the literature in three areas.

First, we provide a measurement framework that links observed expressions of uncertainty in newspaper articles to a latent propensity to express uncertainty, which we argue is an ordinal proxy for the uncertainty that matters for economic decision-making, namely the intensity of the cognitive state of uncertainty. We use this framework to estimate how the noise-to-signal ratios varies with sample size (or frequency) and show that noise variance is modest at monthly and lower frequencies, and approaching signal variance at daily frequency.

Second, we study key choices in the empirical implementation of such measures more deeply than has been done previously, focusing on uncertainty keyphrase selection, isolating economic uncertainty, de-duplication of articles, and appropriate scaling of the uncertainty measure, with a critique of scaling methods commonly used in the literature. Our findings provide empirical foundations for the extant literature, and evidence-based recommendations for methodological improvements.

Third, we conduct the first detailed comparative analysis of a news-media uncertainty measure with another uncertainty proxy, stock returns volatility. Our narrative analysis establishes the plausibility of our news-media measure. Our quantitative analysis reveals a strong relationship to stock volatility on average. But this relationship breaks down periodically, with timing that suggests that the semantics of the word "uncertainty" may be biased towards *downside* uncertainty or risk. Finally, we establish the absence of Granger causation between the measures down to daily frequency, except for a one-day lead of stock volatility over news-media uncertainty, which is to be expected given that the FT is published before the market opens.

Contents

1 Introduction

1.1 Motivation

Uncertainty is widely believed to affect a wide range of economic and financial decisions and outcomes, especially those involving a long or indefinite horizon, such as company capital investment and hiring, durables consumption, precautionary savings, creation of credit with a long maturity, and long-term financial investment by households.

The uncertainty that is causal for economic decisions, and thereby for economic outcomes, is uncertainty as perceived by humans about factors entering their economic decisions. We argue that this uncertainty can be conceptualised as a cognitive state variable.

Unfortunately, this cognitive state of uncertainty is not directly observed. A variety of uncertainty proxies have been proposed[1]. However, there has been limited exploration of how these relate to the cognitive state in theory, and how they relate to one another empirically.

A new class of uncertainty proxies has recently been proposed, based on automated analysis of news-media textual data (Alexopoulos & Cohen, 2009, 2015; Baker, Bloom, & Davis, 2013, 2015). In essence, these proxies measure the fraction of newspaper articles on a given topic (e.g. 'the economy' or 'economic policy') that express uncertainty, as judged by the appearance of specified uncertainty-related keyphrases in the article.

A fundamental *advantage of natural language measures* is that they provide a relatively direct expression of cognitive state. However, there has been little work to date on: the *measurement framework* linking these proxies to cognitive state, *key choices in empirical implementation*, and *comparison to other uncertainty proxies*. This paper begins to address these gaps, as we summarise briefly below. The corpus used in our empirical work comprises all FT articles published between 1 January 1982 and 30 April 2014.

1.2 Research summary

1.2.1 Advantages of natural language measures

Natural language expressions of uncertainty, such as those recorded in news-media textual data, provide a relatively direct window on the latent intensity of the cognitive state of uncertainty. Only surveys that directly elicit uncertainty perceptions capture similarly direct expressions of uncertainty. These are relatively rare[2], and typically cast the question in terms of a forecast range or quantiles or moments of a probability distribution, thus excluding uncertainty related to 'unknown unknowns' and inchoate uncertainty (see below). Other uncertainty measures tend to be based on observed outcomes several steps removed from cognitive state, and thus more vulnerable to conflation with non-uncertainty factors[3].

---

[1] These include volatility (realised or options-implied) of financial returns or of aggregate economic indicators, measures of scale of distributional/density forecasts, revision volatility of forecasts, size of forecast errors, cross-sectional dispersion of (i.e. disagreement between) forecasts, and frequency of particular internet searches (see Bloom (2014) for an overview).

[2] The Bank of Italy survey of firms used in Bontempi, Golinelli, & Parigi (2010) is one example.

[3] Realised volatilities of financial asset returns are the result of many individual trading decisions, which may conflate idiosyncratic and time-varying risk-aversion, liquidity premia, counterparty risk, and technical factors.

Two additional advantages of natural language approaches serve to motivate a larger research program of which this paper is a building block, even though they are not realised in this paper. First, the subject of uncertainty is often stated explicitly in the text, so uncertainty can be fairly unambiguously decomposed by topic. Second, the author and/or reader of the text may identifiable, so interpersonal heterogeneity in uncertainty perceptions can be examined. The diversity of financial markets instruments may compete with the topic coverage of the financial news-media (though in wider corpora there is surely more diverse topic coverage, especially for longer histories), and survey data can identify the respondent, but neither does both at once. While there can be significant barriers to exploiting this textual data (access, digitisation, cleaning, structuring, semantic analysis, etc.) the reward is a wealth of granular information.

1.2.2 Measurement framework

We propose a model linking the cognitive state of uncertainty to the frequency of observed expressions of uncertainty in news-media articles. We assume that the cognitive state of uncertainty has an *intensity*, and that the propensity of a journalist to express uncertainty in a news article, is monotonically increasing in that intensity. (Journalists' cognitive states of uncertainty and their propensities to express uncertainty are of interest because they plausibly reflect – and may also affect – a central tendency in perceived uncertainty across a wider population.) We then aim to estimate the propensity to express uncertainty, and interpret it as an ordinal measure of the cognitive state of uncertainty.

Each article can be considered as a Bernoulli trial in which success is defined as the article expressing uncertainty. Our model assumes that the probability of expressing uncertainty equals the latent propensity to express uncertainty, and that these Bernoulli trials are approximately independent, so that their sum is binomial distributed with mean equal to the propensity to express uncertainty. Under these conditions the observed fraction of articles expressing uncertainty is an unbiased and consistent estimator of the propensity to express uncertainty. We also explore relaxing these conditions. In conclusion then, the fraction of articles expressing uncertainty provides an ordinal measure of the cognitive state of uncertainty.

We derive a rough indication of the theoretical noise-to-signal ratio in our uncertainty measure, where the noise arises from sampling variance in the observation fraction of articles expressing uncertainty, when the number of articles is finite. We show that the noise-to-signal ratio is negligible at annual frequency, and that noise is of comparable magnitude to signal at daily frequency, so that the measure is still useable in statistical modelling at high frequency, but the error-in-variables issue should be recognised.

1.2.3 Key choices in empirical implementation

Textual uncertainty measures are based on a fraction of articles in a given period that express uncertainty. Following the literature, we classify an article as expressing uncertainty if it contains one or more instances uncertainty keyphrases, namely "uncertain", "uncertainty", or "uncertainties"[4]. However, we go further and show that adding further sensible keyphrases (e.g. "unclear*", "unsure*", using wildcard notation) would provide only a modest boost to signal

---

Implied volatilities based on options or variance swap prices conflate assumptions embedded in the asset-pricing models used. Forecast disagreement could reflect heterogeneous beliefs that are nonetheless held with great certainty. Forecast revision volatility could reflect updating of beliefs that were previously held with great certainty.

[4] The plural, excluded in some of the seminal work without good reason, boosts the signal by 14%.

strength, at the expense of muddying the semantics, and introducing poorly understood and potentially material noise. The exception is "risk*". This appears almost three times as often as "uncertain*" but has complex and sufficiently different semantics, that it would need a separate study.

To isolate *economic* uncertainty, we focus on articles about economic topics. We show that this includes the large majority of FT articles. We can therefore include all FT articles in our counts – rather than only articles containing a small number of 'economic' keyphrase – without inducing large measurement error from the inclusion of articles on non-economic topics. The four-fold increase in the number of articles reduces noise-to-signal ratio by on the order of 30%. We analyse the relative effects of different economic keyphrase sets found in the literature, to help inform comparisons between papers using these. We also caution that the emerging practice of requiring uncertainty and economic keyphrases to occur in close proximity (not just in the same article) – which has the laudable aim of more precisely isolating *economic* uncertainty – can result in very low article counts and thus large sampling variance (noise).

Duplication of articles is a common problem in computational analysis of textual data, yet to the best of our knowledge no systematic analysis of duplication in the commonly used data sources has been undertaken. We show that duplication is common in the database we use, and develop a simple de- duplication algorithm based on patterns in the duplication.

Finally, we argue that the various approaches in the literature to scaling economic uncertainty article counts – with the intention of controlling for time-varying news volume – are not theoretically well grounded. Instead, uncertainty article on a given topic should be scaled by the number of articles *on that topic*[5]. Other approaches will induce high-frequency noise into the measure, as well as low frequency trends that are driven by factors other than uncertainty. Empirically, this concern appears to be most important when comparing *levels* of the uncertainty measure between widely-spaced periods. For *movements* at timescales where secular trends in article counts are less important, the choice of scaling method appears to be less important as measured by the correlation between uncertainty measures using the different scaling methods, but our preferred method does yield substantially lower noise variance.

1.2.4 Comparison to other uncertainty proxies

With our measure in hand, we compare variation over time to that in the widely-cited economic policy uncertainty (EPU) measure of Baker et al. (2015). The two co-move strongly as expected. However, since they are targeting uncertainty about different topics (general economic matters vs. only economic policy) this is a far as we take the comparison.

The rest of our analysis comprises the first in-depth comparison of a news-media uncertainty measure with a non-text-based measure. We consider stock returns volatility, which has widely been used as proxy for economic uncertainty. Our narrative analysis shows that our news-media uncertainty measure moves in line with intuition about uncertainty around key narrative events. News-media uncertainty remained high for several years after the onset of the global financial crisis, even as the *level* stock volatility was artificially suppressed by central bank intervention. Nevertheless, the *movements* in the two measures continued to be strongly correlated. Indeed, the

[5] Since we use the FT the total article counts and economic article counts coincide, but this is not the case in the rest of the literature.

strong relationship holds even up to daily frequency, and we find that the only Granger causation between them – a one-day lead of stock volatility on news-media uncertainty – can be explained by the fact that the FT is published before markets open, so can only incorporate information that emerges during the trading day, in the next day's edition. This is consistent with both measures incorporating a common uncertainty signal, and points to news-media measures being useable at high frequency, even based on the relatively limited news volume from a single newspaper.

Interestingly though, the rolling correlations reveal that the long-term average correlation obscures switching behaviour between periods of high vs. low correlation. Notably, the timing of the periods of low correlation coincides with economic booms, in which perceptions of downside risk may be have been muted. Volatility weights upside and downside risk equally. But if the news-media measure attaches greater weight to downside risk, for example due to the semantics of the word "uncertain", this might explain the dynamics. Testing this hypothesis is a key direction for future research.

1.3 Structure of the paper

The rest of this paper is structured as follows. Section 2 briefly reviews the related literature. Section 3 lays out our measurement framework. Section 4 introduces our data. Section 5 provides our empirical implementation. Section 6 presents our comparative analysis. Section 7 draws conclusions and Section 8 suggests directions for future research.

2 Literature review

The literature on news-media measures of uncertainty is small as yet. The seminal work is Alexopoulos & Cohen (2009), which measured the frequency of articles in the New York Times containing the keywords "uncertainty" (or "uncertain") and "economic" (or "economy") for 1929-2008 (thus missing most of the recent crisis period). The authors contrast the effect of news-media uncertainty with that of stock volatility in separate low dimensional VARs and together in a trivariate VAR with various US output variables. This implies an *indirect* comparison of the measures, but the authors do not go beyond a cursory verbal analysis in comparing the two uncertainty proxies *directly*. The authors extend this work in a published version (Alexopoulos & Cohen, 2015), using longer keyphrase lists and conducting more comparative analysis between their measures, VXO and the EPU measure of Baker et al. (2013).

The next most closely related work is Baker et al. (2013). This focuses on economic *policy* uncertainty and so counts only articles that contain particular policy-related keyphrases in addition to "uncertain" or "uncertainty". The authors' comparison of this measure to the US implied stock returns volatility index known as the VIX is limited to a graphical comparison of the movements of this measure around a small subset of large stock market index jumps and reporting the full sample Pearson's correlation of 0.578[6]. Their analysis is extended in a more recent version (Baker et al., 2015) of their working paper.

A couple of recent papers have used news-media uncertainty measures as one component in composite uncertainty indices, deployed in macroeconomic VARs for the UK. However, these papers contain very limited analysis of the composite measure itself or of the relationship between the

[6] They also report a correlation of 0.733 using a variant of their index focussed on articles containing terms that they *a priori* relate to the equity markets, rather than to economic policy *per se*.

news-media component and the other components (Dendy, Mumtaz, & Silver, 2013; Haddow, Hare, Hooley, & Shakir, 2013).

To the best of our knowledge there has been no detailed study of key aspects of the construction methodology including de-duplication and scaling by news volume. Nor has there been any detailed comparison of a news-media uncertainty measure with common alternative uncertainty proxies, such as stock volatility.

A larger parallel literature seeks to extract measures of general tone or sentiment (rather than uncertainty specifically) from news-media textual data using larger dictionaries of keywords. The focus is typically on predicting the level of stock returns, though a couple of papers have examined the link between sentiment measures (not uncertainty) and volatility (e.g. Kothari, Li, & Short, 2009; Tetlock, 2007). The only sentiment analysis work specifically on uncertainty that we know of is Loughran & McDonald (2011), which builds a dictionary of 285 words "denoting uncertainty, with emphasis on the general notion of imprecision rather than exclusively focussing on risk". The authors examine the relationship between frequency of these words in the text of US company 10-K filings and changes in company stock returns, volume, and returns volatility around the filing publication date. They find a significant positive correlation between post-publication return volatility and occurrence of uncertainty words[7].

A wider literature develops measures of economic uncertainty using *non-textual* data. Recent contributions include Jurado et al. (2015), Rossi & Sekhposyan (2015), and Segal, Shaliastovich, & Yaron (2015) which decomposes uncertainty into 'good' (upside) and 'bad' (downside) components.

Finally, our comparative analysis also connects to the literature on the correlates and determinants of stock returns volatility. In particular, if news-media uncertainty is interpreted as reflecting fundamentals then its correlation with volatility would support the hypothesis that the volatility is partly determined by fundamentals. The framework for our quantitative analysis bears similarities to Campbell, Lettau, Malkiel, & Xu (2001) though they compare stock returns volatility with disaggregated component of itself, rather than with uncertainty.

3 Measurement framework

This Section lays out our measurement framework.

Section 3.1 argues that the propensity to express uncertainty in news-media articles, U^*, can be interpreted as an ordinal measure of the intensity of the cognitive state of uncertainty, which is ultimately what we want to measure.

Section 3.2 explores the conditions under which our economic uncertainty measure, namely the fraction of articles expressing uncertainty, U, is an unbiased and consistent estimator of the propensity to express uncertainty, U^*.

[7] Our tentative understanding of their results (they do not provide an extended discussion or sufficient detail to construct a definitive interpretation for oneself) is that a one standard deviation increase in the frequency of uncertainty-related keywords in a 10-K is associated with a 10 percentage point increase in post-publication return volatility, which would be quite substantial compared to a typical annualised volatility of 10 to 20 percentage points.

Section 3.3 provides indicative estimates of the noise-to-signal ratio (NSR) in U to be expected at different frequencies, and demonstrates that the NSR is small enough to make U an empirically useful proxy for U^* even down to daily frequency.

3.1 Cognitive state and the propensity to express uncertainty

We aim to measure the intensity of the FT journalist's cognitive state of uncertainty.

Models of cognition in the context of real-world economic decisions are under-developed. The empirical work needed to guide model development (e.g. neuro-economics) is also in its infancy. Most extant models of economic decision-making restrict uncertainty to mean-preserving spread in the distribution of a stochastic variable, i.e. the decision-maker behaves as if they are able to enumerate all possible outcomes and assign a probability to each. This excludes less quantifiable or Knightian (1921) uncertainty. More recent models of decision-making under ambiguity are promising[8], but still require enumerable outcomes. This excludes uncertainty arising from the belief that there may be 'unknown unknowns', and inchoate uncertainty closer to Keynes' (1936) 'animal spirits'[9].

Yet it seems plausible that the cognitive state of uncertainty ('cognitive state' for brevity) – however it is modelled – can be considered to have an *intensity*. While we do not observe this intensity directly, we do observe the shadow it casts in natural language expressions of uncertainty. Natural language (notably the word "uncertainty") is free to encompass whatever composite of uncertainty concepts is most congruent with the cognitive state. It further seems plausible that the propensity to express uncertainty in natural language statements ('the propensity'), which we denote U^*, is monotonically increasing in this intensity[10].

An estimate of the propensity to express uncertainty is then an ordinal measure of the intensity of cognitive state of uncertainty[11]. The propensity can be estimated from the relative frequency of realised expressions of uncertainty in a natural language corpus.

3.2 Empirical estimator

Our empirical estimator of the latent propensity to express uncertainty, U^*, is the observed fraction of articles that express uncertainty, which we denote U. In other words, $U \equiv m/n$, where n is the total number of articles published in the given period (time subscripts left implicit), and m is the number of those articles that express uncertainty.

U is discretised for finite on to the set $\left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\right\}$ but the discreteness becomes small relative to the scale of variation in U^* as n becomes large. For example, at daily frequency n is typically on the order of 200, so that the spacing of discretisation points is 0.005, which is much smaller than the empirically estimated standard deviation of U^* of 0.029 in Table 1.

We seek inference about U^* based on n and m (which are combined in U). For simplicity and to avoid having to make assumptions about the process that generates n, or about the latent distribution from which n is drawn, we condition inference on observed n. We will thus be using m

---

[8] See Etner, Jeleva, & Tallon (2012) for a review. Empirical applications are few as yet.

[9] For a discussion of the different types of uncertainty see e.g. Marakova (2014).

[10] This monotonicity is preserved under any deviations from truthful revelation that can be modelled as a monotonically increasing transformation to the benchmark propensity under truthful revelation (e.g. understatement of uncertainty due to conservative house style in the FT).

[11]

or equivalently (conditional on n), U, for inference. The resulting efficiency loss is likely small, since any variation in n that might have conveyed information about U^* is constrained by the relatively stable size of a daily edition (the expanded editions in the weeks following the collapse of Lehman Brothers are rare exceptions).

To derive the properties of $U|U^*, n$, we model m as a sum of n i.i.d. Bernoulli trials with success probability equal to U^*. Each trial represents an article and ‘success’ in a trial corresponds to that article expressing uncertainty[12]. To preview Section 5.1, we classify an article as expressing uncertainty if it contains at least one instance of the keywords “uncertain”, “uncertainty”, or “uncertainties”. The expected probability of an article expressing uncertainty is thus U^*, and the variance is $U^*(1-U^*)$.

m is thus conditionally binomially distributed, i.e. $m|U^*, n \sim \mathrm{Bin}(n, U^*)$, and $U|U^*, n$ is distributed as a scaled binomial, with expectation and variance as follows:

$$E_{U|U^*,n}[U|U^*,n] = E_{m/n|U^*,n}\left[\frac{m}{n}|U^*,n\right] = \frac{E_{m|U^*,n}[m|U^*,n]}{n} = U^* \tag{1}$$
$$Var_{U|U^*,n}\,[U|U^*,n] = Var_{m/n|U^*,n}\left[\frac{m}{n}|U^*,n\right] = \frac{Var_{m|U^*,n}[m|U^*,n]}{n^2} = \frac{U^*(1-U^*)}{n} \xrightarrow[n\to\infty]{} 0$$

where the first equality in each equation follows by the definition of U, the second equality from the properties of expectation and variance operators, and the third from standard results for the moments of the binomial distribution applied to m, and cancellation of common terms in numerator and denominator.

Thus U is an unbiased and consistent[13] (in the large n sense) estimator of U^*. Furthermore, by correspondence with standard results for estimators of the binomial distribution, U is the maximum likelihood estimator (MLE) of U^* under our modelling assumptions, and thus asymptotically efficient within the class of estimators that condition on n.

The estimation error $U - U^*$ is distributed as a re-centred binomial with zero mean. By the de-Moivre–Laplace theorem, for large n, this is well approximated by a normal distribution, so that U is reasonably approximated as a variable centred on U^* and subject to classical measurement error whose size is decreasing in n.

If m encompasses multiple topics (but with the mix of topics assumed to be fixed over time – see discussion below) and/or publication days then the Bernoulli success probabilities (the latent uncertainties) may vary between articles. However, so long as the draws remain independent, the their sum is, by definition, distributed as Poisson binomial. This is a generalisation of the binomial distribution. The mean of m will equal the means of the means (sic) of the individual Bernoulli trials (i.e. the mean of the latent uncertainties). Similarly, the variance of m will equal the mean of the variances of the individual Bernoulli trials, and the classical error approximation for large n will still hold. Hence U will be an unbiased and consistent estimator of for the mean of the latent uncertainties across the articles within the aggregation set. In particular, if aggregation is over time

[12] The restriction to a binary classification of articles is a simplifying device. More generally each article could be assigned an uncertainty score based on a more sophisticated analysis of the text.
[13] By Chebyshev's inequality, unbiasedness and variance tending to zero are sufficient for consistency.

(the aggregation set is over multiple days) then U estimates the mean of the daily U^* values across that period.

Note that when uncertainty varies between articles in the aggregation set, the variance of our measure is smaller than it would be in if the uncertainty were equal across those articles, which is the case which uniquely maximise the variance[14]. Thus we can think of the simplified binomial case as giving an upper bound on measurement error.

Our simple model neglects that the topics of the articles published in a given period are sampled from a larger topic space. To the extent that the latent uncertainties are heterogeneous across topics, such topic sampling will induce an additional variance (noise) component in $U|U^*, n$. The variance will be smaller with

- aggregation over longer periods, which allows for richer sampling of the topic space;
- more evenly distributed topic weights, which makes it less likely that the mean uncertainty of a finite sample of topics will deviate substantially from the mean over the population of topics;
- less disperse distribution of uncertainties between topics (in the limiting case with zero dispersion of uncertainties between topics, all topics have the same uncertainty, and $U|U^*, n$ collapses to U^* in the limit of large n)..

Quantitative modelling of this variance component, to determine its relative importance, would be a useful direction for future research.

The assumption of independence of Bernoulli trials greatly simplifies modelling, but is not innocuous. While the independence assumption is only required between articles published within the same measurement period, it is not difficult to imagine mechanisms that would induce serial dependence, especially at short lags. However, we lack a clear prior on even the sign of any serial correlation. Some mechanisms could induce negative serial correlation: if uncertainty on a given topic has been expressed in a recent article it may now be 'old news' and thus less likely to stimulate the production of, or be expressed in, a new article; or uncertainty may have been so frequently mentioned that it comes to be taken as given and thus less likely to be mentioned explicitly. Other mechanisms could induce positive serial correlation, beyond that arising from persistence in the underlying cognitive uncertainty: expressions of uncertainty in the recent articles may lead to uncertainty becoming part of the boilerplate narrative in articles on certain subjects. We thus leave further exploration of serial dependence between articles to future work.

3.3 Noise-to-signal ratio

Using the simple model above, we can obtain a rough indication of the relative scale of noise in U versus signal (latent U^*) as a function of n. The noise is the sampling variance in the stochastic generating process for U. In short, we use the above model of how a given latent U^* gives rise to U, to partition the observed variance of U into noise and signal components, whose relative magnitude depends on n.

[14] Too see this note that the sum is of the form $\sum_{\tau \in \mathrm{T}} U_\tau^*(1 - U_\tau^*)$ where T is the aggregation set of articles and U_τ^* is the latent uncertainty for the article indexed by τ. The summand is concave, so the sum is maximised when the U_τ^* are all equal.

We summarise the scale of the noise by the conditional variance $Var_{U-U^*|n}[U - U^*|n]$ and of the signal by $Var_{U^*|n}[U^*|n]$, and consider their ratio:

$$NSR(n) \equiv \frac{Var_{U-U^*|n}[U - U^*|n]}{Var_{U^*|n}[U^*|n]} \tag{2}$$

Using the results from Section 3.2 – that U is an unbiased estimator of U^*, and that the case with equal propensity to express uncertainty across all articles in the period gives an upper bound on $Var_{U|U^*,n}[U|U^*, n]$ – we can obtain an expression for $NSR(n)$ in terms of observables only. Replacing the population statistics in that expression with their sample analogues gives the following estimator:

$$\widehat{NSR(n)} = \frac{(n-1)\widehat{\sigma^2_{U,n}}}{n\widehat{\sigma^2_{U,n}} - \widehat{\overline{U_n}}\left(1 - \widehat{\overline{U_n}}\right)} - 1 \tag{3}$$

where $\widehat{\overline{U_n}}$ and $\widehat{\sigma^2_{U,n}}$ are respectively the conventional sample estimators of the expectation and variance of U based on observations at the periodicity corresponding to n. See Appendix B for the derivation. As expected, $\widehat{NSR(n)}$ tends to zero in the limit of large n.

We also obtain an estimate of the true latent variance of U^*:

$$\widehat{\sigma^2_{U^*,n}} = \widehat{\sigma^2_{U,n}}/\left(1 + \widehat{NSR(n)}\right) \tag{4}$$

Numerical estimates of $\widehat{NSR(n)}$ and $\widehat{\sigma^2_{U^*,n}}$ are shown in Table 1 below. These are based on simplifying assumptions, but should suffice to indicate the orders of magnitude involved. Noise is less than 0.6% the size of the signal at annual frequency and probably still small enough to neglect at the monthly frequency at which the main parts of our comparative analysis are conducted. However, at daily frequencies, the noise is comparably sized to (and quite possibly larger than) the signal, and this will cause a non-trivial errors-in-variables problem in Granger causation analysis in Section 6.3.4.

Table 1: Noise-to-signal ratio (NSR) – estimates for daily to annual frequencies

Frequency	Daily	Weekly	Monthly	Quarterly	Annual
Typical n	200	1200	5200	15600	64200
$\widehat{\overline{U_n}}$	0.0436	0.0432	0.0432	0.0432	0.0431
$\sqrt{\widehat{\sigma^2_{U,n}}}$	0.0211	0.0142	0.0124	0.0117	0.0105
$\boldsymbol{\widehat{NSR(n)} \leq}$	**0.871**	**0.205**	**0.054**	**0.020**	**0.006**
$\sqrt{\widehat{\sigma^2_{U^*,n}}}$	0.0289	0.0156	0.0127	0.0118	0.0105

Notes: $\widehat{\overline{U_n}}$ and $\widehat{\sigma^2_{U,n}}$ taken from Table 7 on pg. 53.

The signal variance decreases as the frequency decreases, consistent with reversion-to-the-mean behaviour that is also observed in other measures of uncertainty such as stock returns volatility.

4 Data

4.1 News-media textual data

m and n, from our measurement framework in Section 3.2, are operationalised as U (not to be confused with the italicised U, our uncertainty measure, which is a ratio) and T, derived from the full text of the London print editions of the Financial Times (FT). U is calculated as the ratio of U and T. Lower frequency observations are generated by separately summing U and T, then calculating U from the sums (as opposed to averaging daily U).

The consistent publication schedule, audience, topic coverage, and format associated with a long-established newspaper like the FT makes it easier to construct an uncertainty measure that is comparable from one period to the next, than is the case with other less structured textual data such as tweets or blogs.

Using a single publication minimises the variation in structure and format which might induce unobserved variation in the correspondence between U and U^*. The FT is the leading daily business news publication in the UK, with an audience and content focus particularly suitable for measuring the uncertainty of major decision makers about business, economic and financial matters (see Appendix A.1).

That said, in order to analyse the effects of some empirical implementation choices on a generalist paper, we also obtain article counts for The Times of London, for parts of the overall sample period.

Our primary source of article counts is the Dow Jones Factiva database. Other similar databases, including Nexis UK, are used to help clean and de-duplicate the data (see Appendix A.2). The main sample we use for quantitative analysis spans 1984-2012 (see Appendix A.3).

4.2 Stock returns volatility

Aggregate volatility, σ, is estimated from realised holding returns on the Datastream Global Equity Indices Total UK Market index[15]. These are highly correlated with returns on the more commonly used FTSE100 and FTSE All Share indices (0.994), but span our full sample period unlike the latter indices (which start in 3 January 1984 and 31 December 1985 respectively). Options-implied (ex-ante) volatility would be preferable in principle, but is only available from the late 1990s for the UK, and is practically indistinguishable by eye from the realised measure.

Non-parametric estimates of σ are straightforwardly obtained from daily returns. At weekly and lower frequency we estimate σ as the sample standard deviation of daily returns multiplied by the conventional $\sqrt{252}$ annualising factor (Merton, 1980), in line with references to stock volatility in our literature. However, at daily frequency the estimate is based on a single returns observation, and the Merton method is sensitive to outliers. We therefore use we use the daily absolute returns (DAR) method of Schwert (1989a)[16] which is more resistant to outliers.

For the regression-based Granger causality tests at daily frequency in Section 6.3.4, the substantial noise in the DAR estimates of σ would raises concerns about errors-in-variables and consequent

[15] Datastream series TOTMKUK.

[16] See his footnote 4: multiply absolute daily returns by $\sqrt{\pi/2}$ (corresponding to the ratio of the variance to the expectation of the absolute value of the deviation from the mean, assuming a normal distribution of returns) and then by $\sqrt{252}$ to annualise.

parameter bias that could obscure interpretation of the test results. For this purpose then, we use realized volatility (RV) estimates of Heber, Lunde, Shephard, & Sheppard (2009) based on intra-day tick data. However, these estimates are not available for periods before 2000 and they refer to the narrower FTSE100 index[17], hence why we use the less precise estimates above in the rest of our work, where the measurement error is less problematic. The Pearson's correlation coefficient between the RV estimates and the daily absolute return (DAR) over 2000–2012 was 0.451, largely reflecting the noise in the DAR estimates which is probably also responsible for the relatively low daily correlations reported in Table 10.

5 Empirical implementation

This Section examines – more closely than in the extant literature – how some of the key choices in implementing a news-media uncertainty measure can affect the results, and recommends methodological improvements. Section 5.1 discusses uncertainty keyphrase selection. Section 5.2 shows how we isolate 'economic' uncertainty. Section 5.3 documents duplication in Factiva's FT records, and provides a method to remove it. Section 5.4 explains the importance of scaling the measure appropriately, and critiques the methods used in the extant literature.

5.1 Uncertainty keyphrase selection

We operationalise m from Section 3.2, by classifying an article as expressing uncertainty if its full text (including headline) contains any of the keywords "uncertain", "uncertainty", or "uncertainties", and denoting the count by U. We collectively refer to these keywords as "uncertain*", using wildcard notation[18]. "Uncertainties" is neglected in most of the extant literature[19]. Including it increases U by 14.3% without materially changing the semantic range, thus improving the signal-to-noise ratio.

This set of keywords is largely immune to problems often associated with keyphrase counting. Their semantics are *a priori* highly congruent with our latent uncertainty concept (it is difficult to imagine a sense of "uncertain*" that is both far from this concept and in common use), and relatively stable over our sample period (cf. internet-related words such as "surf" and "web"). Undetected negation is not a material problem because "uncertain*" is already self-negated and further negation, whether outright[20] or qualified[21] is rare, as expected for professionally edited news copy.

---

[17] This covers 80% by market capitalisation of the Datastream index mentioned above.

[18] Conventionally the wildcard notation would include any word starting with the string "uncertain", including the remaining part of speech, "uncertainly", and typos (e.g. "uncertainity", "uncertainy" and instances where the space before the next word has been dropped, such as "uncertaintyin"). Our data collection did not cover these, but they occur very infrequently: in our full sample period only 382 articles contain "uncertainly", of which 61 contain the included keyphrases; 130 articles would be added if including typos.

[19] Haddow et al. (2013) count only "uncertainty". Alexopoulos & Cohen (2009) and Baker et al. (2013) also count only "uncertainty" and "uncertain". (Alexopoulos & Cohen refer to another paper of theirs that counted "risk" instead, but this is neither published nor available online.) Most recently however, (Alexopoulos & Cohen (2015) do include "uncertainties".

[20] In the nearly two million articles published during our sample period, "not uncertain*" appears in only twelve articles, "no longer uncertain*" and "nor uncertain*" in two, and "never uncertain*" in one. Other negations, including phrasal forms, are presumably similarly rare.

[21] In a human audit of 4,300 articles mentioning "uncertain" or "uncertainty" and "economic" or "economy", Baker et al. (2013) found that "only 1.8% of articles about economic policy uncertainty discuss low or declining policy uncertainty".

Further keyphrases may be used to express our latent uncertainty concept. We constructed the candidate list in Table 2 by a traversal[22] of WordNet[23] (Cognitive Science Laboratory of Princeton University, 2010). We further added "risk*" which did not emerge by this method but is arguably related to our latent uncertainty concept (24% of Factiva FT records containing "uncertain*" also contain "risk*").

Adding such keyphrases to our measure may add both signal and noise (due to use of senses not congruent with our latent uncertainty concept). An upper bound on the incremental signal gain (corresponding to no non-congruent uses of the keyphrases) is the ratio of the number of articles containing the keyphrases to be added *and not* "uncertain*", to the number of articles containing "uncertain*". In practice, noise will subtract from this upper bound.

Table 2: Additional keyphrases – basic relationship to "uncertain*", 1984–2012

Keyphrase	Ratio of U using only this keyphrase to U using only "uncertain*"		Correlation of a variant of U, calculated using only this keyphrase, to baseline U, calculated using only "uncertain*"			
	standalone	incremental	weekly	monthly	quarterly	yearly
unclear*	0.26	0.23	0.42	0.51	0.56	0.62
not clear*	n/d†	0.14	n/d†	n/d†	n/d†	n/d†
unpredictabl*	0.08	0.07	0.35	0.59	0.71	0.85
not sure*	0.08	0.07	0.43	0.47	0.66	0.72
unsure*	0.05	0.04	0.24	0.45	0.63	0.81
not certain*	0.15	0.01	0.29	0.38	0.61	0.77
not predictabl*	0.00	0.00	0.02	0.05	0.09	0.16
any the above	1.55	0.54	n/d	n/d	n/d	n/d
risk*	2.96	2.38	n/d	n/d	0.65	0.71

Notes: 'standalone' refers to the ratio of the number of articles containing the given keyphrase to the number containing "uncertain*"; 'incremental' instead excludes articles that contain "uncertain*" from the numerator. It represents the proportional increase in the number of articles that would be classified as uncertain, if adding this one keyphrase to our baseline keyphrase set {"uncertain*"}. n/d = not calculable from our dataset. † Some figures were not available for "not clear*" due to a data collection error (omitting the wildcard in the database search) that could not be rectified by the time it was discovered because I no longer had access to the data source.

By far the largest potential signal gain comes from "risk*" (238%, see Table 2). The Pearson's correlation of a variant of U using "risk*" with our baseline version using "uncertain*" is 0.71 at annual frequency, which is substantial, consistent with our prior of a close semantic relationship, but substantially less than unity, consistent with the presence of noise.

The maximum potential gain from other keyphrases is second-order: at most 23% from a single keyphrase ("unclear*") and 54% from all of them simultaneously. The correlations with our baseline measure are again large but substantially below unity (e.g. 0.62 for "unclear*"). The potential for noise is non-trivial. For example, "unclear*" may be commonly used to express inexactness of knowledge about past or present facts, rather than the forward-looking uncertainty more typically expressed by "uncertain*".

---

[22] We started from synsets (essentially, sets of synonyms) containing the word "uncertain*" and followed semantic links to related words. If those possessed frequently used senses that were congruent with our latent uncertainty concept then we added them to the list and continued the traversal, else we stopped the traversal in that direction. We added direct negations of antonyms in the resulting list (e.g. "not clear" cf. "unclear").

[23] A widely used semantic lexicon designed to support automatic text analysis.

The rest of this paper analyses only our baseline version of U using "uncertain*". This is because, unfortunately, we obtained data for the additional keyphrases only after the rest of this paper was complete, and the data we do now have for "risk*" is only down to monthly frequency. Clearly a priority in future research should be to examine measures including "risk*". That said, the semantics of "risk*" and its relationship to "uncertain*" are not straightforward[24] and our focus on "uncertain*" has the advantage of producing a relatively 'clean' measure which is comparable to the extant literature.

5.2 Isolating 'economic' uncertainty

The extant literature has used generalist newspapers which cover topics largely unrelated to the economy. To filter out uncertainty about irrelevant topics, authors count only articles that contain at least one economy-related keyphrase. However, the keyphrase lists employed are usually quite basic[25]. Most extant literature uses {"economic", "economy"} and we denote the corresponding article count by E[26]. This discards many economy-related articles, and leaves a question mark over the representative of this subset of articles for the larger set of economy-related articles.

By using only FT articles, we can count over *all* articles, whose number we denote T, and thus largely sidestep problems associated with economic keyphrase filters. Less than 8% of FT articles are *not* about economic topics, as shown in Table 3. The list there comprises some obvious economy-related terms and could be extended, so 8% is an upper bound. There is a trade-off here between two sources of noise: first, from including some articles on non-economic topics; and second increasing the fraction of economic articles included in our sample and thus reducing the sampling variance of our estimates of U^*.

On the first source of noise, the correlation between the measure estimated on all articles, versus only economic articles, is approximately 0.94 (in our study sample, see Table 5), which suggests that these two definitions are measuring substantially the same thing, and noise from non-economic articles is not a major concern.

On the second source of noise, counting across all economic articles increases the number of economic articles sampled by a factor of more than four, since only 19% of articles contain "economic" OR "economy". This translates into an estimated 30% reduction in noise-to-signal ratio at monthly frequency (see Appendix B.2). Note also that attempting to bulk up the number of articles containing "economic" or "economy" by adding The Times of London to the corpus, alongside the FT, as is done in the UK EPU measure of Baker et al. (2015), would boost the sample size by a factor of only around one half, so that the number of articles sampled would still fall far short of that obtained from all FT articles[27]. On balance therefore, we expect our approach to yield a substantial net reduction in noise relative to the standard approach in the literature.

Table 3: Fraction of FT articles containing economy-related keyphrases, 1984–2012

[24] Five of the six senses of "risk" recorded in WordNet refer to possibility or likelihood of loss or negative outcome, rather than either uncertainty *per se* or stochastic risk.

[25] Alexopoulos & Cohen (2015) is a recent exception, using 11,725 economic-finance terms and partial phases. The present work was at an advanced stage when that paper was published, and collecting data to analyse the extended keyphrase list would have been prohibitively time-consuming.

[26] We required exact word matches, which excludes plurals, following Baker et al. (2013), whose method was clarified in personal communications during January 2016. See also Alexopoulos & Cohen (2009).

[27] In Table 4 and Table 5, compare the sum of $\widehat{\mathrm{E}}$ estimates for The Times and The FT, with $\widehat{\mathrm{T}}$ for the FT.

financ*	59%
econ*	21%
market*	39%
bank*	24%
("building society" OR "building societies")	1%
debt*	9%
credit*	9%
bond*	6%
loan*	5%
(equity OR equities)	9%
money*	12%
(currency OR currencies)	5%
"exchange rate*"	1%
derivative*	1%
(company OR companies)	47%
business*	27%
profit*	20%
earning*	11%
revenue*	8%
wage*	2%
government*	28%
tax*	16%
politic*	14%
policy*	11%
Eurozone	1%
oil	7%
gas	4%
coal	1%
commodit*	3%
any of the above	**92%**
none of the above	8%

Notes: "*" denotes a wildcard. Percentages are calculated from un-de-duplicated search result counts.

The comparability of E in different time periods is assured by the relatively time-invariant semantics of the words "economic" and "economy". The comparability of T depends on relatively time-invariant topic coverage of the FT, which we expect holds to a reasonable approximation given the FT's long pedigree.[28]

The semantic range (or topic coverage) of 'economic' corresponding to T, counted within the FT, is potentially broader than that corresponding to E, counted within the FT or The Times. For example, in articles on the macro-economy and public economic policy, journalists might have a higher propensity to use the words "economic" or "economy" when in articles on particular companies or industries.

We can test for a difference in the semantic range of T and E in the FT. First, if there were *no* difference then "uncertain*" should appear with the same frequency in both T as in E. In other words, a pair of indicator variables for the occurrence in a given article of "uncertain*" and of

[28] Of course on a given day not all topics are drawn from the distribution over all topics. This induces additional noise (sampling variance, larger at higher frequency) in our measure, the modelling of which we leave to future research.

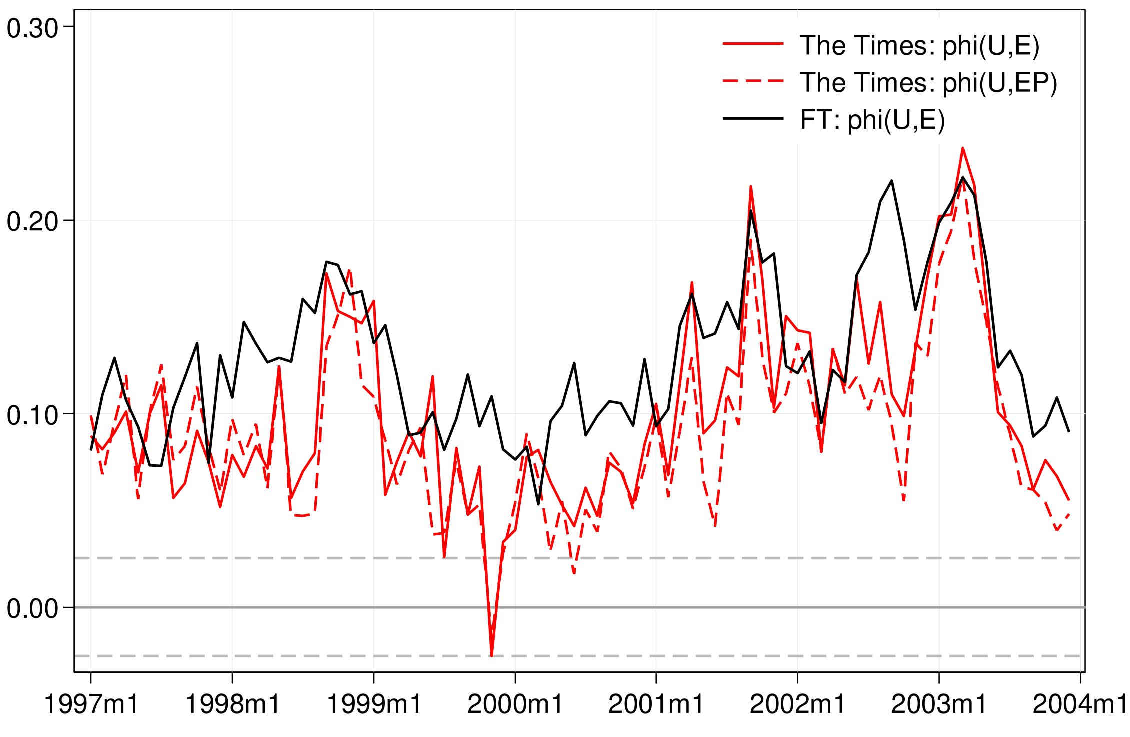
"economic" OR "economy", would be uncorrelated across many articles. Figure 1 shows that this correlation is significant, but modest and fairly stable over time at around 0.1 to 0.2 – suggestive of a fairly moderate and stable semantic difference.

Second, the time-variation in U/T and EU/E is strongly correlated in the FT, as seen in the upper panel of Figure 11, and in the correlation coefficient of 0.938 (see Table 5). This is consistent with a similar pattern of time-variation in uncertainty within T and E, which is in turn consistent with T and E having a similar semantic range (though correlation between the uncertainties over two different semantic ranges would also produce this result).

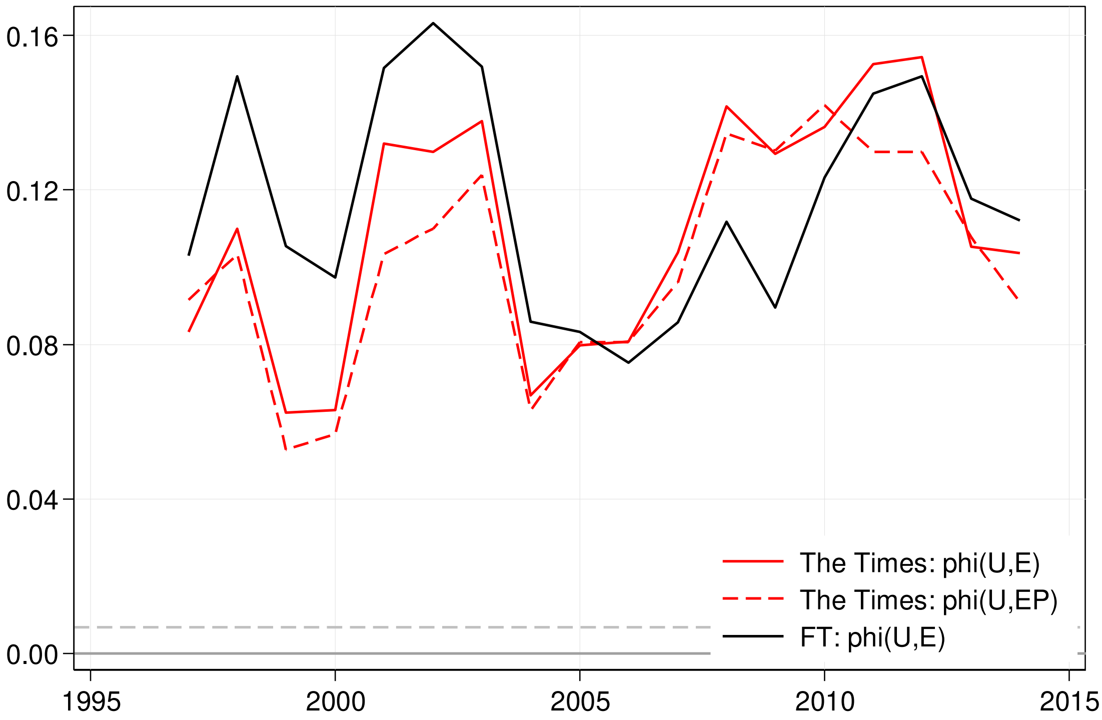
Surprisingly, similar results hold for The Times. The association between occurrence of "uncertain*" and "economic" OR "economy" within a given article, is similar to that in the FT; and the time-series correlation between U/T and EU/E (0.878) is only slightly lower than in the FT. This might be in part due to the fact that uncertainty references are disproportionately weighted towards economic uncertainty (EU/E is around one third, cf. only around one tenth for U/T). But these results could also arise if uncertainties about economic and non-economic topics (e.g. politics and social policy) are positively correlated. In any case, these results suggest that filtering on economic keyphrases constitutes a second-order refinement in The Times, raising the question of whether this is also true for other generalist newspapers. Future research should probe this with a proper textual / semantic analysis.

Figure 1: Correlation of article-level indicator variables for the occurrence of "uncertain*" and of ("economic" OR "economy")

Monthly



Annual



Notes: Based on undeduplicated article counts. Grey dashed lines are indicative thresholds for significant difference from zero at the 95% level, derived from the fact that the correlation coefficient is approximately distributed as $\chi^2(1)/\mathrm{T}$, and substituting for T its sample mean value for The Times. (Strictly, the threshold would vary with T as it varies over time and between The Times and FT. But the resulting variation in the threshold is small compared to the scale of variation in the correlation coefficient.)

Recent work has extended the economic keyphrase list used for selecting articles on economic topics in generalist newspapers. Dendy et al. (2013) include "econ*" (using the wildcard notation introduced above). We denote the corresponding article count by E^*. This picks up around 20% more articles in The Times (see $\widehat{\mathrm{E}^*/\mathrm{E}}$ in Table 4) than just "economic" OR "economy", and around 14% more in the FT (Table 5). Most of the gain is due to the plurals "economics", "economies". It is therefore unsurprising that the percentage gain in news volume is relatively stable, as can be seen in the upper panels of Figure 2: at monthly frequency E^*/E has a standard deviation equal to around 2% of its mean in both newspapers (see $\widehat{\sigma_{\mathrm{E}^*/\mathrm{E}}}$ in Table 4 and Table 5), shows negligible trend or seasonality, and does not exhibit strong autocorrelation.[29]

The same authors only classify an article as expressing economic uncertainty if the uncertainty keyphrase occurs within five words of the economic keyphrase. Alexopoulos & Cohen (2015) allow a twenty-word window. The laudable aim is to increase the likelihood that the expression of uncertainty actually refers to the economy, which might not be the case if multiple topics are discussed in a single article, only one of which is the economy. However, even with the wider window, the resulting article counts, denoted $\mathrm{EU20}$, are often so small at monthly frequency (median of seven articles and several months with zero articles in our study sample – see notes to Table 4) that the resulting quantisation induces considerable sampling variance. This can be seen in

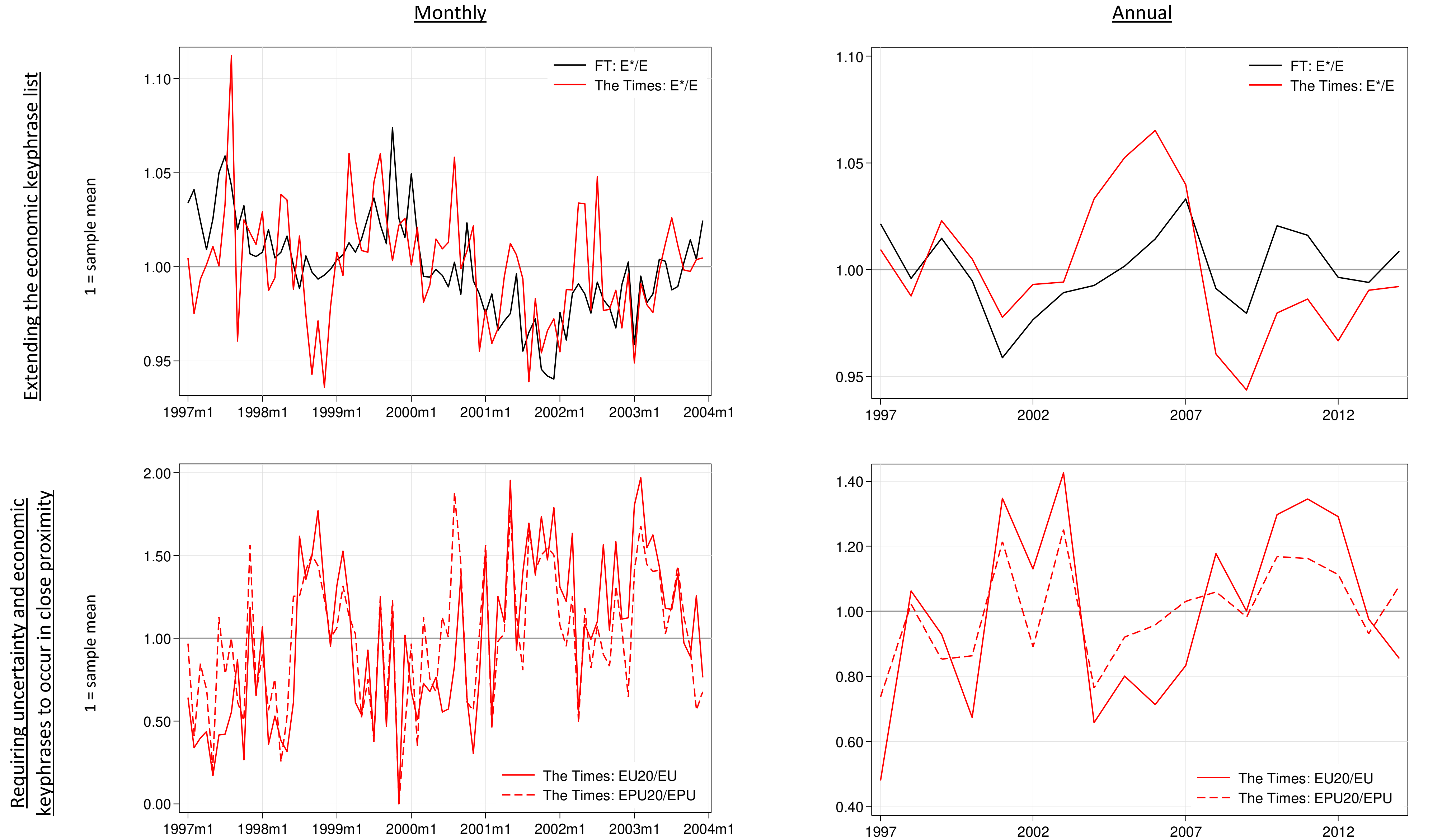
[29] Interestingly, the time-profile of E^*/E is similar The Times and the FT, which suggests this factor is not just due to sampling variance. For example, it could be due to shifts in the relative rate of news arrival (then picked up in both newspapers) between news about the economy ("economic") vs. the economics discipline ("economics"), or the domestic "economy" vs. global "economies". But the variation in E^*/E is of second order compared to other factors examined here, we leave further investigation of this to future research.

the lower panels of Figure 2. The standard deviation of the ratio $\mathrm{EU20}/\mathrm{EU}$ is the same order of magnitude to that in the benchmark uncertainty measure, at monthly frequency (see $\sigma_{\widehat{\mathrm{E^*/E}}} / \sigma_{\widehat{\mathrm{EU/E}}}$ in Table 4). The problem is worse at higher frequency – around 60% of days in our study sample have $\mathrm{EU20}$ equal to zero – and for narrower topics. For example, we show the effect on economic policy uncertainty article counts ($\mathrm{EPU20}/\mathrm{EPU}$) using the keyphrase lists in Baker et al. (2015)[30]. Counting over multiple newspapers will mitigate this to some extent. However, Alexopoulos & Cohen (2015) uses a single newspaper, and Dendy et al. (2013) only three, so these concerns may not be immaterial. Any trend in these factors is limited, and autocorrelation relatively weak, but the substantial monthly seasonality bears further investigation in future research, given that it is not obviously an artefact from a structural break, and is not sensitive to extending the lag order of the regression.

In conclusion, we recommend using "econ*" rather than {"economic", "economy"} to obtain modest but non-trivial gains in sample size, and consequent reduction in noise arising from sampling variance, without a large change in semantics; but advise caution when considering filtering rules that substantially lower article counts.

[30] We count an article in $\mathrm{EPU20}$ if the uncertainty keyphrase appears within twenty words of *either* the economic or policy keyphrases.

Figure 2: The effect of alternative approaches to isolating economic uncertainty, on article counts



Notes: Based on undeduplicated article counts. Ratios are normalised by their own mean over the sample plotted. Due to the time-intensive nature of data collection, in the lower panels we show results only for The Times, for which E and EPU are smaller than in the FT, and therefore the problems discussed in the main text are more starkly illustrated.

5.3 De-duplication of articles

Duplication is a common problem in computational analysis of textual data, yet despite Factiva FT records being used in a growing literature, to the best of our knowledge no systematic analysis of duplication in this data source has been published to date. This Section begins to address that gap.

In short, there is substantial duplication, with a variety of causes and thus no single simple pattern over time or across articles. We estimate that on average 6.4% of Factiva FT records are duplicates, but these are clustered in particular periods, with the daily duplication rate varying between nil and more than 100% (so more than one duplicate per article).

If duplication affected all articles on a given day in the same way, then it would cancel out in the ratio that we take when scaling uncertainty article counts. Unfortunately, manual investigation showed the duplication rate varies across articles even on the same day, especially before 1993, and did not show a regular pattern over time. Hence deduplication is necessary.

Records that are identical to one another in all fields except the unique identifier (accessionNo) assigned by Factiva, are unambiguously duplicates of the same underlying FT article. However, most records that one would want to identify as duplicates – i.e. that correspond to the same item in the facsimile copy of the day's FT edition in Gale database – exhibit some variation in the article text. Where to draw the line between near duplicates and distinct articles is ultimately a matter of judgement. Near duplicates may exhibit:

- cosmetic variation: white space, punctuation, coding of non-alphanumeric characters such as currency symbols, and case (especially in the headline);
- headline formatting: for example, with the section name (e.g. "COMPANIES AND MARKETS") pre-pended in one version and not in another;
- substantive textual variations that a human reader would identify as relatively minor differences, which might arise from corrections, editorial changes in later editions, and localisation in regional editions[31] (e.g. additional explanation in the US edition in an article mentioning the UK Royal family).

We designed a parsimonious algorithm to identify likely duplicates that avoids the complexity associated with fuzzy text matching, while removing most of the duplicates. The key was to understand the nature of the variation that renders the corresponding database records non-identical. We targeted our diagnostic efforts on periods when the ratio of the total article count in Factiva to that in Nexis UK, shown in Figure 3 before any de-duplication, deviated substantially and persistently from unity[32]. (We believe that Nexis UK contains relatively few duplicates for two reasons. First, its daily article count changes slowly over the last 30 years, consistent with our prior expectation, and larger discrete jumps were usually consistent with the facsimile record. Second, manual examination of all FT records in Nexis UK, for randomly selected days, revealed relatively few duplicates, as well as contents closely matching ProQuest and the facsimile copies in Gale. Thus

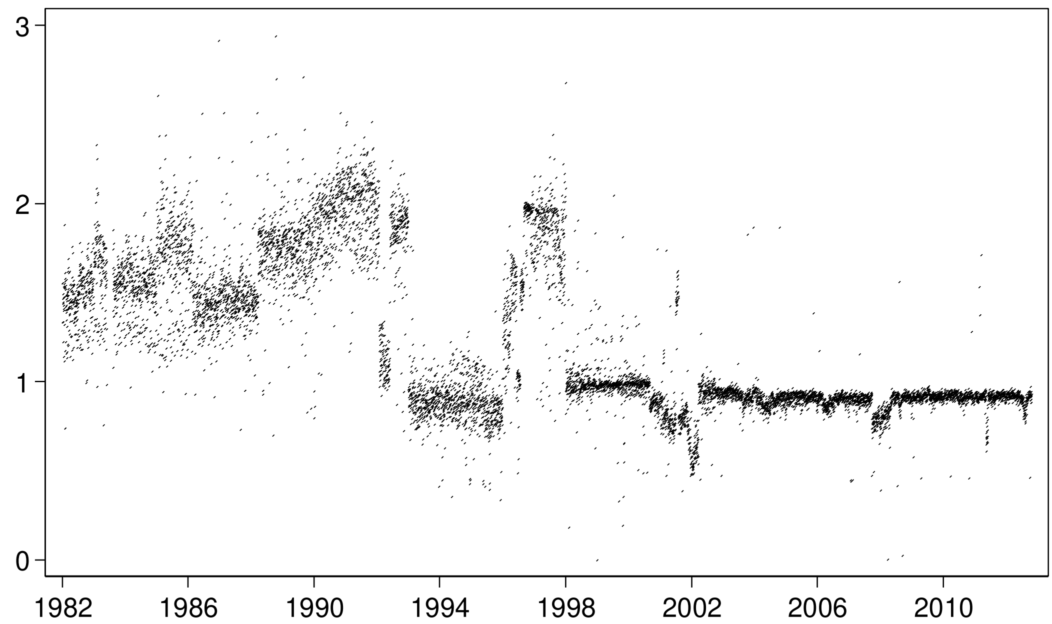
[31] In theory these should not appear in the dataset for the London print edition, but in practice the archiving does not appear to have always been so clean.

[32] A small fraction of the deviations from a unit ratio in Figure 1 are due to variations between and within databases (primarily comparing 1980s and later periods) in how a daily FT edition is split into 'articles' for database storage (see footnote 33 for an example).

Nexis UK seems unlikely to suffer from serious omissions. In retrospect, though Factiva is more widely used in this nascent literature, we would have used Nexis UK as our source, had the switching costs not become prohibitive by the time we established this.)

From Figure 3 it is clear that there are few duplicates since 26 June 2008, consistent with Factiva's claim, in its documentation, to have de-duplicated content added since that date. Our algorithm identifies only two duplicates (manually verified) beyond that date. Furthermore, duplication rates are mostly low beyond 2003. However, duplication rates are particularly erratic prior to 1993 and persistently high (around 2) during 1996-7.

Figure 3: Ratio of daily FT record counts in Factiva vs. Nexis UK, 1 January 1982–30 April 2014



Date axis tick marks are for 1 January in years labelled.
Ratio is missing for a single observation, 23 January 1999, due to a database error in Nexis UK.
Outliers truncated, numbering 31, of which 7 are for days with no articles in Nexis UK but non-zero article count in Factiva, implying an infinite ratio.

Each iteration of algorithm development focused on a cluster of days with a ratio far from unity, manually checking for false positives in candidate duplicates and false negatives (residual duplicates) in the records retained by the last iteration of the algorithm, and parsimoniously modifying the algorithm in an attempt to avoid these.

Duplicates also arise from records with identical text being tagged with different company codes by Factiva. Patterns in the Factiva unique record identifier suggest that this may be the result of Factiva's tagging algorithm having being run at different times, without cleaning up the legacy copies.

Our final algorithm counts two Factiva records as duplicates if all the following attributes are equal:

- publication date
- uncertainty keyphrase counts in each part of the article (headline, lead paragraph, tail paragraphs)

- headline *after* removing all non-alphanumeric characters (including punctuation and whitespace) and converting to uppercase
- (for publication dates up to 31 Dec 1988 only[33]) company codes applied by Factiva.

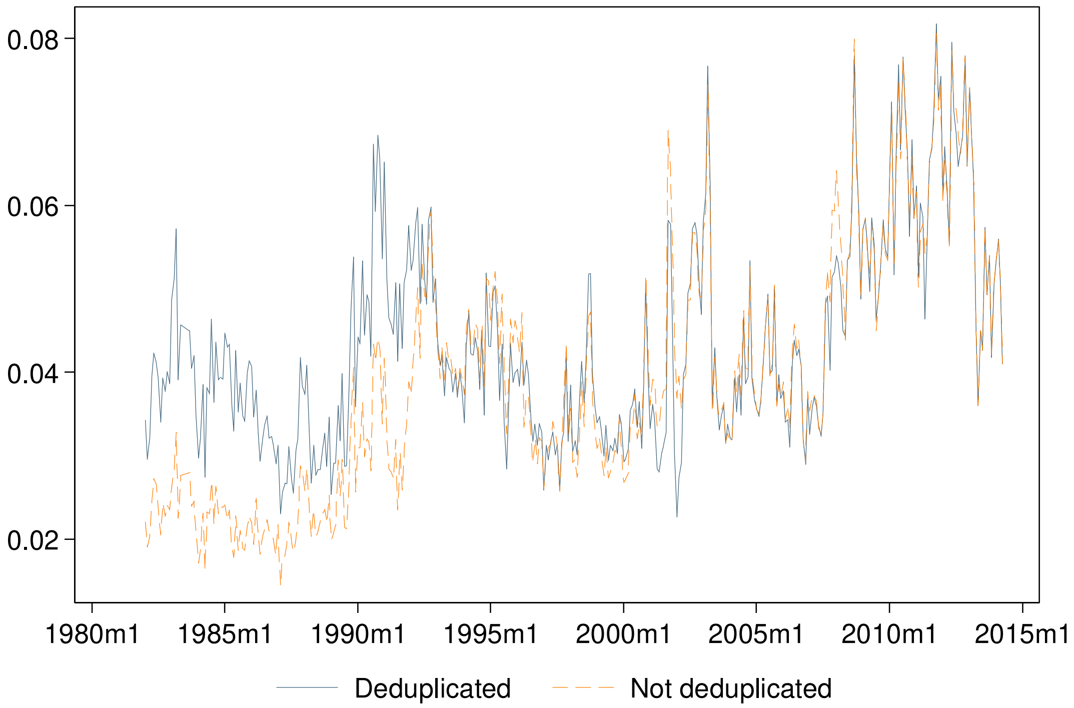
We retain a single record from each duplicate set, taking the union of company codes sets where these differ, and retaining the record with the larger number of text characters where this differs.

We believe this delivers a reasonable approximation to a deduplicated set of articles, though it does not handle substantive textual variations (see above) and could be improved in future work.

For technical reasons we were only able to apply this bottom-up de-duplication algorithm to articles containing our "uncertainty*" keyphrases. For total article counts, and other article counts used in auxiliary analyses in Section 5, we use a top-down method, starting from total article counts in Nexis UK and applying the adjustments listed in Appendix A.5.

The effect of duplication on U can be seen in Figure 4, which compares U calculated from Factiva data before and after deduplication (full sample correlation 0.851). Differences are concentrated before 1993.

Figure 4: U calculated on data before and after de-duplication, 1982m1–2014m4



[33] Summary articles about share stakes, appointments, and annual/interim report releases are split into separate items with identical headlines (e.g. "Share stakes") prior to this date, and requiring company codes to match prevents these being incorrectly identified as duplicates. Meanwhile, duplicates with different company codes appear to be more prevalent in the 1990s, so that removing this field from the duplicates criterion prevents those duplicates being missed.

After de-duplication we have 1,998,165 unique articles in our full sample, of which 1,858,424 fall during our main 1984-2012 sample period.

5.4 Scaling the uncertainty measure

5.4.1 Rationale for our method

The uncertainty article counts must be appropriately scaled to remove the effect of substantial secular and high frequency variation, documented below, that is probably driven – at least in part – by factors other than uncertainty.

Under our model of 'propensity to express uncertainty' in Section 3, each article on a given topic constitutes an opportunity for uncertainty to be expressed. Increasing the number of articles published on any topic will, all else equal, increase the number of articles expressing uncertainty by the same proportion. It is the fraction of articles rather than the total count that gives us a direct estimate of the propensity.

To isolate the propensity to express uncertainty from changes in news volume we must scale the uncertainty article count by the total article count *on the topic* whose uncertainty we are trying to measure. So if the numerator counts only articles containing economic keyphrases and uncertainty keyphrases (denoted EU), as is common in the literature because the corpus includes generalist newspapers, then the denominator should count only articles containing economic phrases (denoted E). Since we count uncertainty in all FT articles (U; see Section 5.2) we must scale by the count of all FT articles (denoted T).

Contrary to this, most of the literature scales EU by T rather than by E (Alexopoulos & Cohen, 2015; Baker et al., 2013; Dendy et al., 2013). This is inappropriate because T may vary for reasons unrelated to the economy, for example the launch of a new newspaper section or supplement, or drift in the share of the newspaper devoted to non-economic topics. Including this variation in the denominator will induce artefacts in the uncertainty measure.

Another practice, common in the early literature, is not to scale at all (Alexopoulos & Cohen, 2009; Baker et al., 2013; Haddow et al., 2013). This would only be appropriate if the number of economic articles per period was constant over time, which we show below is not the case. Furthermore, if the article data contain duplicates, and the duplication rate varies over time (as is the case in pre-2003 Factiva data, see Figure 3), then failing to scale the measure is particularly problematic, because the duplication rate will be conflated into the uncertainty measure.

The effect of these alternative scaling practices, relative to our recommended practice, can be analysed by decomposing U as follows:

$$\text{If } U \equiv \frac{\mathrm{EU}}{\mathrm{E}} \Rightarrow \frac{\mathrm{EU}}{\mathrm{T}} = \frac{\mathrm{E}}{\mathrm{T}} \cdot U \text{ and } \mathrm{EU} = \mathrm{E} \cdot U \tag{5}$$

Scaling by T induces an additional (potentially time-varying) factor of E/T, relative to our target measure U. Not scaling induces a factor of E. We now analyse these factors empirically, along with other factors related to various aspects of the scaling methods observed in the literature.

5.4.2 Empirical comparison to alternative methods

5.4.2.1 Methodology

For the FT we have de-duplicated counts for U and T spanning our full sample period, as used in the construction of our baseline uncertainty measure U. These are plotted in Figure 5 through Figure 7. For The Times and for other article counts we collected data for shorter 'study samples', to keep the data collection effort manageable. The study samples all start in 1997, coinciding with the start date of the UK measure of Baker et al. (2015), and are described in the notes to Table 4.

We denote the generic factor by x. Multiple factors may be multiplied together to get to convert a particular uncertainty measure to the benchmark uncertainty measure that we argued above is appropriate. Since we are primarily interested in the relative variation in our uncertainty measure, and less so the absolute scale of that variation, we study x after scaling by its sample[34] mean $\hat{\bar{x}}$ at the given frequency, i.e. $\tilde{x} \equiv x/\hat{\bar{x}}$. This also facilitates comparison between different factors.

Various properties of the scaled factors are summarised in Table 4 below, which complements the time series plots in Figure 7 through Figure 11.

The scale of variability in $\tilde{x}$ is summarised by an estimate of its standard deviation, $\widehat{\sigma_{\tilde{x}}}$, and compared to the standard deviation of the corresponding benchmark uncertainty measure – $\widehat{\sigma_{\widetilde{\mathrm{EU/E}}}}$ for The Times, and $\widehat{\sigma_{\widetilde{\mathrm{U/T}}}}$ for the FT (see Section 5.2). We also reported the sample correlation $\hat{\rho}$ between $\tilde{x}$ and the scaled benchmark measure, though do not rely too heavily on these we, especially where the series may be non-stationary, leading to a risk of 'spurious' correlations.

Induced factors are more problematic if they induce spurious trends, serial correlation, or seasonality in the measure. We test for these problems via parameter estimates in the following regressions:

$$\text{Annual: } \tilde{x}_t = \alpha_0 + \delta t + \alpha_1 \tilde{x}_{t-1} + \alpha_2 \tilde{x}_{t-2} + \varepsilon_t \tag{6}$$

$$\text{Monthly: } \tilde{x}_t = \alpha_0 + \delta t + \alpha_1 \tilde{x}_{t-1} + \alpha_2 \tilde{x}_{t-2} + \sum_{m=2}^{12} \gamma_m + \varepsilon_t \tag{7}$$

$$\text{Daily: } \tilde{x}_t = \alpha_0 + \delta t + \alpha_1 \tilde{x}_{t-1} + \alpha_2 \tilde{x}_{t-2} + \sum_{dow} \gamma_{dow} + \varepsilon_t \tag{8}$$

where $dow \in \{Tue, Wed, Thu, Fri, Sat\}$ and Monday is the base level. The aim here is to identify deviations from White noise, rather than a full characterisation of the dynamics, so we stop short of a more sophisticated data-driven approach to lag choice, specification of the deterministic trend, and attempts to distinguish structural breaks from large persistent shocks – which would be at least as problematic as autoregression and seasonality if present.

[34] Note the sample span is shorter at higher frequency, in order to make the data collection effort feasible. Thus results for different frequencies are not strictly comparable with one another, though the relative orders of magnitude should provide reasonable indicators of the relative scale of effects.

5.4.2.2 Results

Not scaling at all

Consider the factor of T that would be induced in our baseline FT-based uncertainty measure U/T, if failing to scale U at all. This exhibits strong secular trends and considerable high frequency volatility, as seen in Figure 5.

Some of the secular variation coincides with movements in uncertainty – for example, the spike in 2008Q4-2009Q1 obviously *coincides* with elevated uncertainty around the onset of the Global Financial Crisis – and may in part be driven by uncertainty (though probably also a temporary surge in the supply of and demand for economic news). But much of the variation in T is not obviously linked to uncertainty, and is more likely driven by structural change and other unexplained factors.

Pronounced secular trends – almost a factor of two between the annual sample minimum and maximum – coincide with structural shifts at the FT: an increase through the 1990s with the phased launch of new international and online editions, and a secular decline since the mid-2000s with the shift towards the online publication, including items not appearing in the print edition. A decline in the share of very short articles through the 1980s – which appears to result from mix of changes in actual FT article length, and in the method for dividing the FT into 'articles' for storage in news databases[35] – tends to reduce the number of articles (though there are apparently countervailing factors since the net effect is muted). The steps up in 1986Q1/2 and down in 2004Q3 do not have an obvious cause.

Only a small part of the substantial monthly volatility is explained by variation in the number of publication days per calendar month. Other causes might include irregular publication of supplements that – if approximately randomly distributed over time – would be smoothed out at lower frequency consistent with what we observe.

Daily news volumes also display substantial day-of-week seasonality, seen in Figure 6 and Table 5, that is unlikely to be driven by uncertainty.

The effect of failing to scale U by T can be seen by comparing U and U respectively in Figure 7. At quarterly frequency uncertainty would be judged to be higher around the Iraq War in 2004Q1 than following the collapse of Lehman Brothers in 2008Q4, and at all frequencies the uncertainty ordering of the collapse of Lehman Brothers and the Eurozone crisis would be reversed.

Turning to our shorter study samples, the counts for which are not deduplicated in contrast with the full sample results discussed above, we can see the particularly pernicious of effect of failing to scale the uncertainty article counts when there is substantial duplication, as during 1997 in the in the upper panels of Figure 8. Failing to scale here would lead one to think that uncertainty was twice as high in 1997 and then abruptly dropped by half on 1 January 1998!

---

[35] For example, many of the short articles in Factiva that are associated with the left tail of the distribution that is seen in the early 1980s, refer to news snippets or corporate appointments announcements that are grouped under a larger heading in the facsimile copy and under a single article in Nexis, and come to be grouped similarly in Factiva in later years.

The step change in article counts in The Times between 2008 and 2009, extends to T so is not just a pivot towards a greater economic news *share*, but also an increase in the total newspaper size.

Even in periods of less dramatic variation, eyeballing Figure 8 suggests the scale of variation in T, E and EP is on the order of 20% of the scale of variation in the benchmark uncertainty measure.

Figure 5: Total FT article counts, T, full sample 1982m1-2014m4 (after de-duplication)

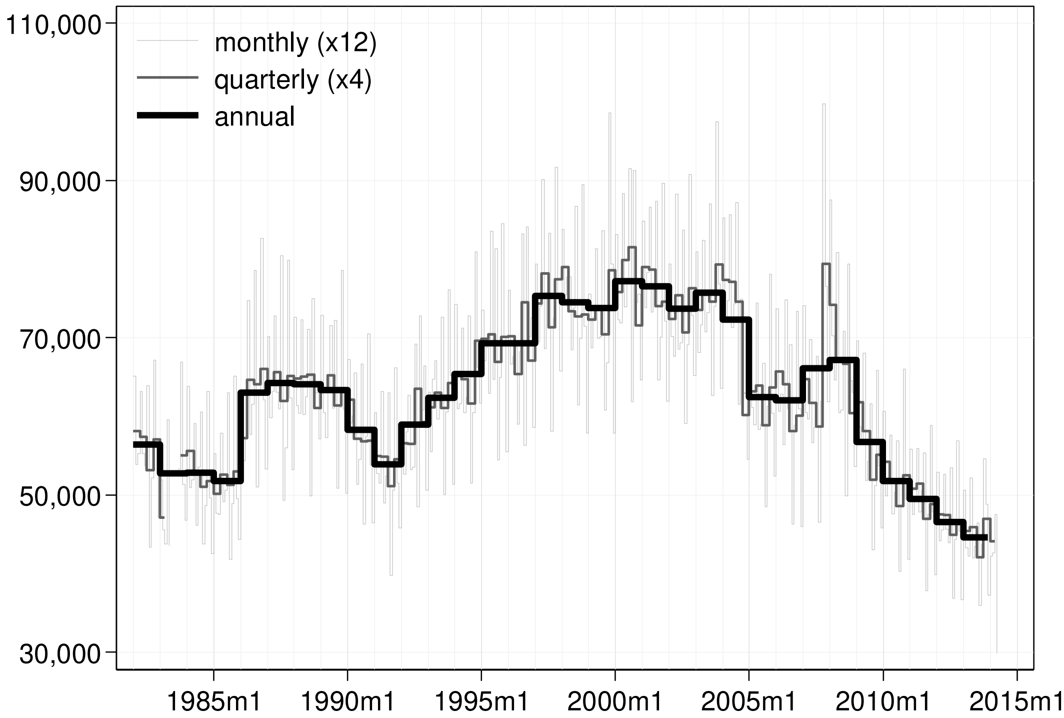
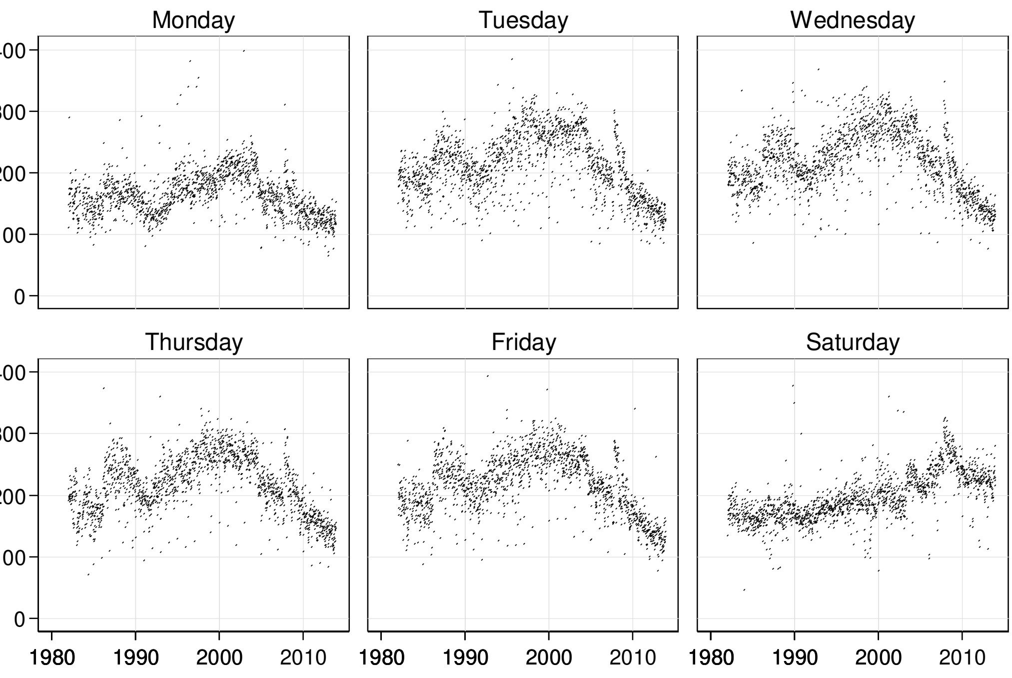


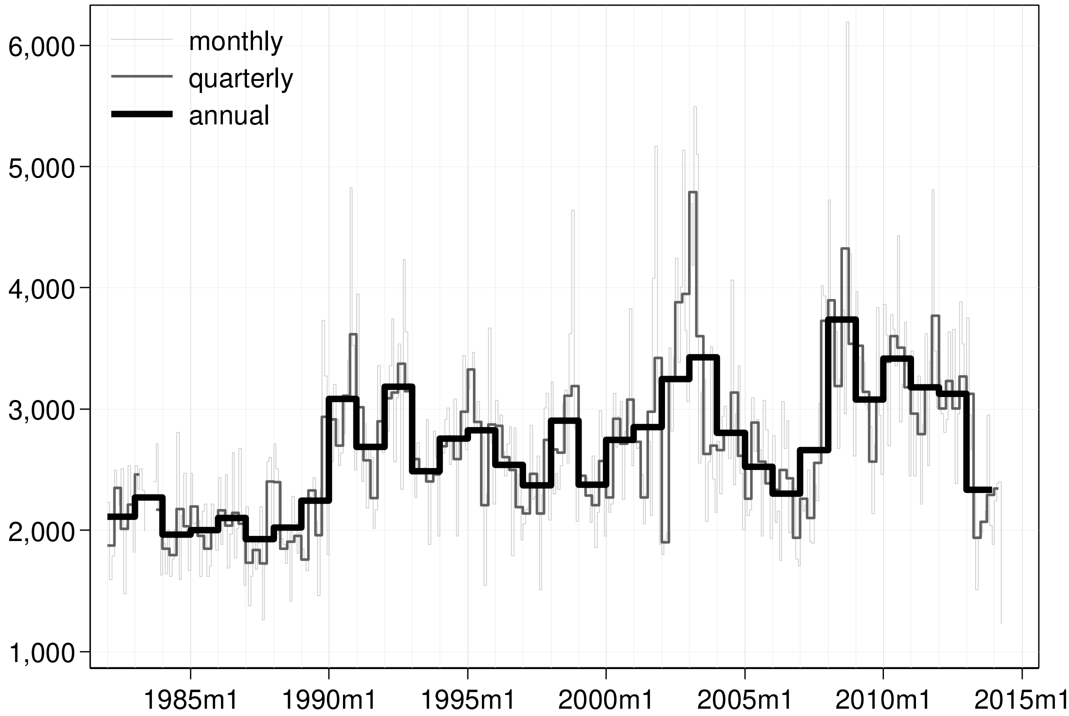
Figure 6: Total FT news volume, T, by day of week (after de-duplication)



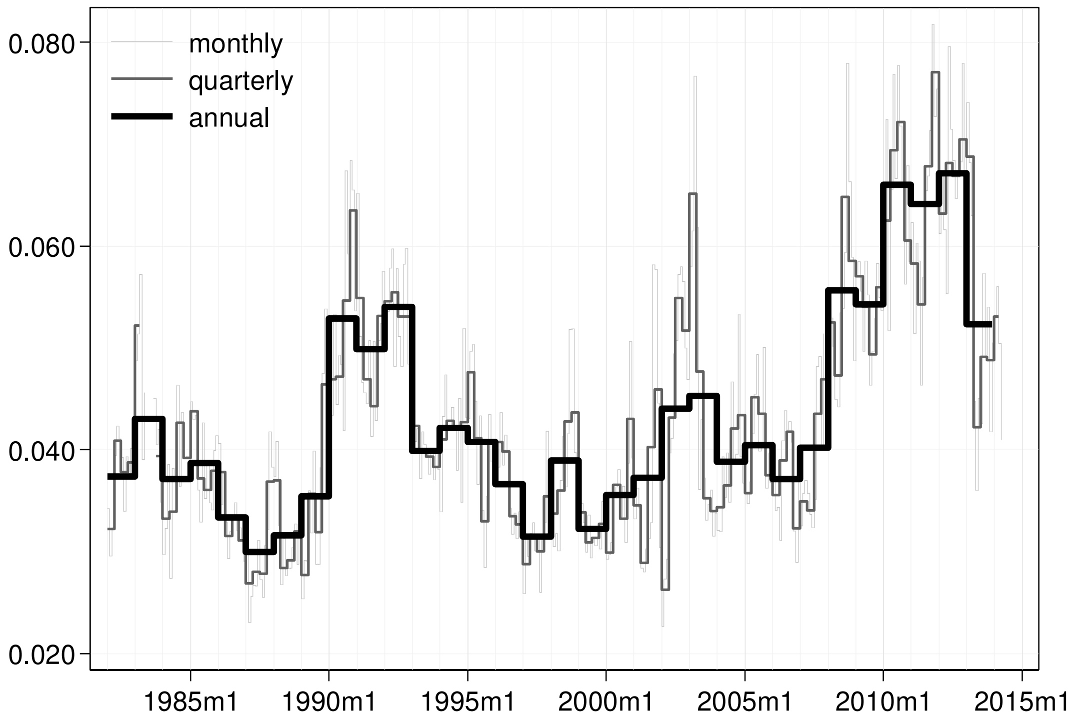
Note: observations with more than 400 articles per days were truncated from the plot.

Figure 7: The effect of failing to scale by economic news volume: U compared to U

U



U



Scaling by total article counts rather than economic article counts

Scaling the measure by T rather than E or EP (for economic *policy*), induces a noise factor, $\mathrm{E/T}$ or $\mathrm{EP/T}$ respectively, whose standard deviation around 20-50% as large as that of signal, depending on whether the frequency and choice of newspaper.

The most notable feature of the time variation in this noise factor, shown in the lower panels of Figure 8, is the sharp step up coincident with the onset of the recent financial crisis – roughly a 50% increase in the factor. As discussed above, while this may in part be driven by the associated uncertainty, it is probably also driven in part simply by the greater supply of and demand for news about topics associated with references to "economic" or "economy", and we have no way of separate out these two parts. Therefore, scaling by T will induce a (partly) spurious jump in the uncertainty measure at this point.

Using proxies for total article counts

Even when aiming to scaling by T, some researchers have been unable to obtain total article counts, citing technical barriers, so instead use the number of articles containing "the" (Dendy et al., 2013), or "today" (Baker et al., 2015) arguing that these are neutral words.

Given the above mentioned problems with scaling by T, we would not recommend either approach. However, in order to illuminate the impact on comparability of results within the literature, and the choice between available methods, we plot $\mathrm{T/T_{today}}$ and $\mathrm{T/T_{the}}$ in Figure 9 and report statistical properties in Table 4 and Table 5.

$\mathrm{T_{the}}$ is clearly preferable *a priori*. Researchers probably resort to $\mathrm{T_{today}}$ where search engines do not allow common keywords like "the" as search terms. Empirically, the standard deviation of the corresponding factor of $\mathrm{T/T_{the}}$ that would be induced in the uncertainty measure is only 2% as large as the variation in U. The analogous result for $\mathrm{T_{today}}$ is 13%.[36]

Note that even the factor $\mathrm{T/T_{the}}$ is not innocuous in the longer run: after a several years of relatively stability, there is an approximate 10% jump up coincident with the onset of the recent financial crisis. The cause is unclear, but the effect will be to drive spurious movement in the uncertainty measure.

Comparison of uncertainty measures with alternative scaling methods

Finally, we compare uncertainty measures obtained under different scaling methods, applied to the same data from our study samples. The results are plotted in Figure 10 and Figure 11. In the main text we report correlations at each of the three frequencies of our study samples, namely annual (monthly) [daily]. When comparing correlations at different frequencies, bear in mind that the sample periods are longer at lower frequencies.

In the FT, $\mathrm{U/T}$ and $\mathrm{EU/E}$ – both of which are scaled in line with our recommended method – are strongly correlated (0.77 (0.94) [0.60]). This is consistent with our claim that the noise caused by including some non-economic articles in U and T, is modest for the FT. Meanwhile the noise from sampling variance is around 30% lower in $\mathrm{U/T}$ than in $\mathrm{EU/E}$ at monthly frequency, and still lower at

[36] Given time constraints on data collection, we only obtained these time-series results for The Times. However, we confirmed on aggregate article counts for the annual study period (1997-2014) that the mean of these ratios are similar for the FT: 7.62 for $\mathrm{T/T_{today}}$ and 1.02 for $\mathrm{T/T_{the}}$.

daily frequency. The similarly high correlation between $\mathrm{U/T}$ and $\mathrm{EU/E}$ in The Times (0.85 (0.88) [0.37]) is less expected, given that $\mathrm{U/T}$ includes more non-economic articles than the FT.

Given our theory arguments that $\mathrm{EU/T}$ is not appropriately scaled, the very high correlation between $\mathrm{EU/T}$ and $\mathrm{EU/E}$ (0.98 (0.97) [0.97] for the FT and 0.90 (0.97) [0.88] for The Times) is perhaps surprising. This result suggests that the empirical importance of scaling by T vs. by E is not so important for *movements* at timescales where secular trends in article counts are less important, even though it matters when comparing uncertainty *levels* between more widely-spaced periods as discussed above. Since the correlations between $\mathrm{U/T}$ and $\mathrm{EU/E}$ are somewhat lower, it would appear that the method for isolating 'economic' articles, and the differences in the semantics/topic definition implied by different choices, are more important, at least in our study samples. That said, our preferred method does yield around 20% lower noise variance.

Comparing within the same measure between the FT and The Times, we find strong correlation in $\mathrm{EU/E}$ in our annual and monthly study samples, consistent with both capturing a similar economic uncertainty signal; and negative correlation in the daily study sample, the reason for which is unclear (0.87 (0.80) [-0.29]). $\mathrm{EU/T}$ exhibits similar levels of correlation of 0.86 (0.82) [-0.26] because $\mathrm{E/T}$ happens to co-move strongly in the two newspapers, which might not have been the case with other publications.

Since the numerator of $\mathrm{U/T}$ captures non-economic as well as economic uncertainty in The Times, we expect this measure to have lower correlation between The Times and the FT. The high correlation in the monthly study sample is thus surprising. It may not be representative of the wider sample, since the correlation is considerably lower in the longer annual study sample, due to diverging trends in the denominator T between the two papers (0.45 (0.79) [-0.13]).

Table 4: Properties of ratios and scale factors induced by alternative scaling methods, and of uncertainty measures – The Times of London

Factor, x	Unconditional distribution							Time-series structure in $\tilde{x}$								
	LR mean	Variability in $\tilde{x}$ vs. benchmark uncertainty measure						LR trend	Autocorrelation				Seasonality			
	Annual	Monthly			Daily			Annual	Monthly		Daily		Monthly		Daily	
	$\hat{\hat{x}}$	$\widehat{\sigma_{\tilde{x}}}$	$\widehat{\sigma_{\tilde{x}}}/\sigma_{\widehat{\mathrm{EU/E}}}$	$\hat{\rho}(\tilde{x},\widehat{\mathrm{EU/E}})$	$\widehat{\sigma_{\tilde{x}}}$	$\widehat{\sigma_{\tilde{x}}}/\sigma_{\widehat{\mathrm{EU/E}}}$	$\hat{\rho}(\tilde{x},\widehat{\mathrm{EU/E}})$	$\hat{\delta}$	$\widehat{\alpha_1}$	$\widehat{\alpha_2}$	$\widehat{\alpha_1}$	$\widehat{\alpha_2}$	P-val.	Size	P-val.	Size
Ratios of article counts from alternative methods for isolating economic uncertainty																
$\mathrm{E^*/E}$	1.209†††	0.023***	0.07	-0.413***	n/d			-0.002*	0.184	0.246**	n/d		0.039	0.013	n/d	
	(0.009)	(0.002)						(0.001)	(0.125)	(0.101)				(0.009)		
$\mathrm{EU20/EU}$	0.343†††	0.407***	1.18	0.544***	n/d			0.016	0.183*	0.395***	n/d		0.043	0.331**	n/d	
	(0.022)	(0.027)						(0.012)	(0.104)	(0.107)				(0.148)		
$\mathrm{EPU20/EPU}$	0.473†††	0.325***	0.94	0.402***	n/d			0.012*	0.240**	0.085	n/d		0.069	0.238**	n/d	
	(0.017)	(0.025)						(0.006)	(0.102)	(0.117)				(0.100)		
Factors induced by not scaling at all																
T	88834†††	0.080***	0.23	0.009	0.180***	0.18	0.121	0.038**	0.635***	-0.145	0.586***	0.140**	0.000	0.068**	0.000	0.596***
	(4887)	(0.008)			(0.022)			(0.004)	(0.112)	(0.135)	(0.133)	(0.056)		(0.033)		(0.050)
E	7048†††	0.127***	0.37	0.457***	0.239***	0.24	-0.045	0.061***	0.535***	-0.059	0.290*	0.101	0.000	0.095***	0.061	0.100**
	(657)	(0.011)			(0.025)			(0.009)	(0.124)	(0.121)	(0.158)	(0.117)		(0.034)		(0.051)
$\mathrm{E^*}$	8456†††	0.121***	0.35	0.404***	n/d			0.059***	0.527***	-0.111	n/d		0.000	0.085***	n/d	
	(743)	(0.010)						(0.008)	(0.124)	(0.123)				(0.027)		
EP	3591†††	0.160***	0.46	0.356***	0.285***	0.28	-0.071	0.063***	0.497***	0.0s48	0.256	0.081	0.000	0.088***	0.674	0.110
	(357)	(0.013)			(0.032)			(0.010)	(0.125)	(0.113)	(0.164)	(0.138)		(0.019)		(0.115)
Factors induced by scaling by total article counts rather than economic article counts																
$\mathrm{E/T}$	0.077†††	0.100***	0.29	0.581***	0.211***	0.21	-0.094	0.023***	0.579***	0.131	-0.031	0.004	0.048	0.048*	0.025	0.494***
	(0.004)	(0.009)			(0.034)			(0.005)	(0.123)	(0.128)	(0.136)	(0.143)		(0.026)		(0.164)
$\mathrm{EP/T}$	0.039†††	0.134***	0.39	0.425***	0.297***	0.29	-0.105	0.025***	0.441***	0.215*	0.045	0.039	0.034	0.064	0.006	0.533***
	(0.002)	(0.010)			(0.036)			(0.007)	(0.114)	(0.120)	(0.140)	(0.124)		(0.047)		(0.194)
Factors induced by using proxies for total article counts																
$\mathrm{T/T}_{today}$	8.372†††	0.046***	0.13	-0.265***	0.189***	0.19	0.079	0.002	0.262***	0.043	0.024	-0.115	0.018	0.024*	0.000	0.421***
	(0.134)	(0.004)			(0.021)			(0.003)	(0.102)	(0.114)	(0.123)	(0.117)		(0.013)		(0.056)
$\mathrm{T/T}_{the}$	1.053†††	0.007***	0.02	0.341***	n/d			0.008***	0.422***	0.342***	n/d		0.009	0.008**	n/d	
	(0.013)	(0.001)						(0.001)	(0.141)	(0.089)				(0.003)		
Uncertainty measures																
$\mathrm{EU/T}$	0.005***	0.420***	1.21	0.974***	1.001***	0.99	0.881***	0.032**	0.754***	-0.022	-0.207	-0.370***	0.277	0.235*	0.059	1.096***
	(0.000)	(0.037)			(0.087)			(0.014)	(0.095)	(0.101)	(0.125)	(0.125)		(0.127)		(0.361)
$\mathrm{EU/E}$	0.063†††	0.346***	1	1	1.010***	1	1	0.010	0.608***	0.097	-0.143	-0.257	0.435	0.162*	0.018	0.620**
	(0.003)	(0.031)			(0.103)			(0.010)	(0.111)	(0.109)	(0.124)	(0.205)		(0.097)		(0.255)
$\mathrm{U/T}$	0.016†††	0.169***	0.49	0.878***	0.437***	0.43	0.369***	0.001	0.645*	-0.073	-0.168	0.112	0.007	0.073**	0.905	0.112
	(0.001)	(0.017)			(0.045)			(0.006)	(0.139)	(0.114)	(0.162)	(0.165)		(0.035)		(0.103)

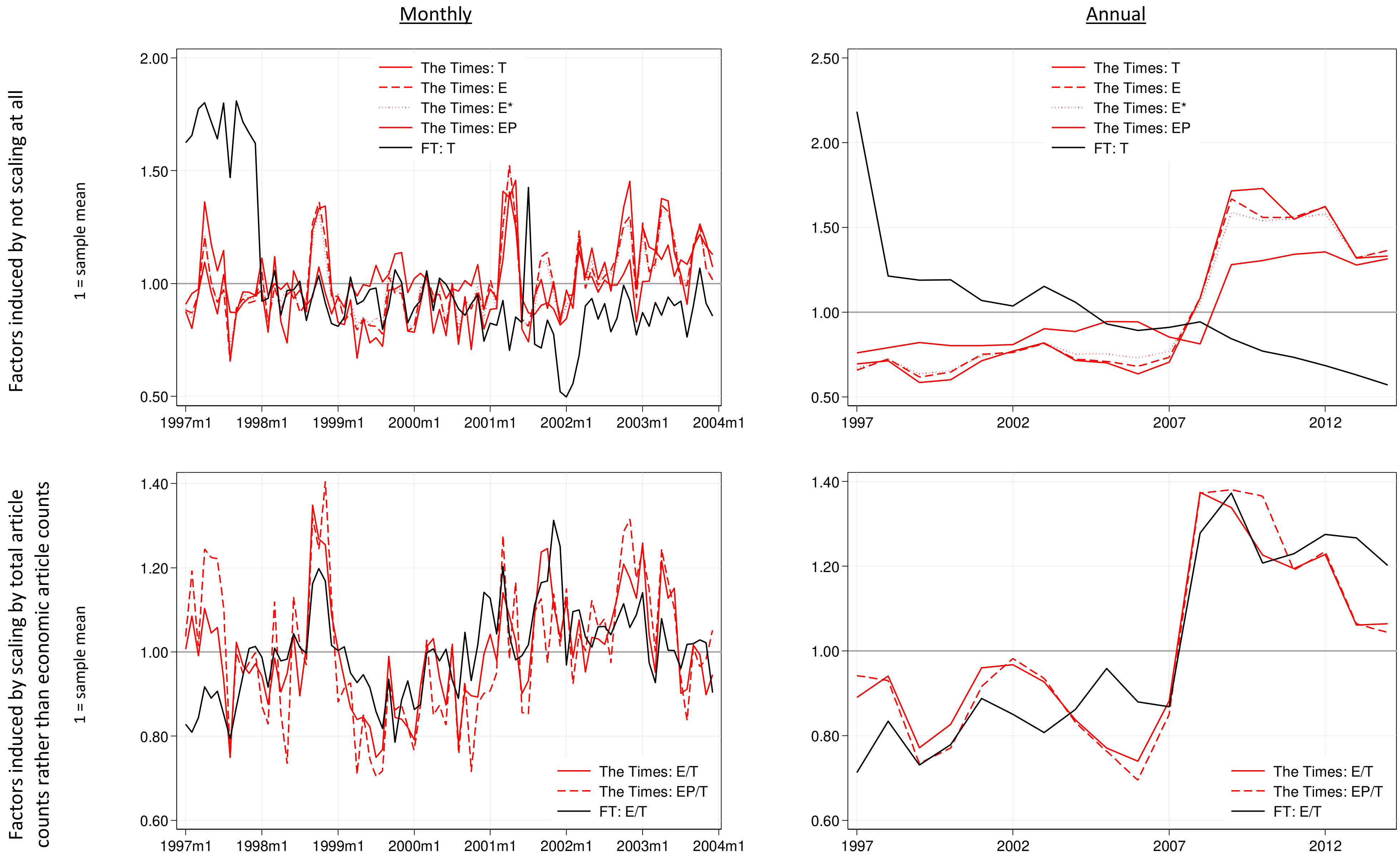
Notes: *Sample.* All rows based on the same samples, which are shorter at higher frequency, and with some variables not collected at daily frequency (resulting in n/d = no data), to make data collection feasible: 18 annual observations for 1997-2014; 84 monthly observations for 1997m1-2003m12; 48 daily observations for 1997m1-1997m2, after excluding days with zero articles. x denotes the variable in the left-hand column of each row. See main text for definitions. U, which denotes the count of articles containing uncertainty keyphrases, should not be confused with U, which denotes the ratio that is our uncertainty measure. *Unconditional distribution.* The estimate of the long-run (LR) mean of x, $\hat{\hat{x}}$, is the parameter estimate on the constant in a regression of x on a constant in our annual sample. $\tilde{x} \equiv x/\hat{\hat{x}}$. The estimate of the standard deviation of $\tilde{x}$, $\widehat{\sigma_{\tilde{x}}}$, is the parameter estimate on the residuals from a regression of residuals on a constant, where the residuals come from a regression of $\tilde{x}$ on a constant. $\widehat{\sigma_{\tilde{x}}}$ is compared to a benchmark measure ($\widehat{\mathrm{EU/E}}$ for The Times, to isolate articles on economic topics, as discussed in Section 5.2). ρ denotes Pearson's product-moment correlation. *Time-series structure.* The long-run (LR) trend parameter $\hat{\delta}$ is estimated from regression (6) in the main text, in the annual sample. The remaining parameter estimates in this section come from the regressions (7) and (8). The p-value for seasonality is from the Wald test of the joint exclusion restriction on all seasonal dummies. The 'size' of the seasonality is estimated as the mean absolute value of the parameter estimates on each of the seasonal dummies. *Inference.* HAC-robust standard errors reported in parentheses. *, **, *** indicate significant difference from zero at 10%, 5% and 1% levels respectively; †, ††, ††† similarly for difference from unity.

Table 5: Properties of ratios and scale factors induced by alternative scaling methods, and of uncertainty measures – Financial Times

Factor, x	Unconditional distribution							Time-series structure in $\tilde{x}$								
	LR mean	Variability in $\tilde{x}$ vs. benchmark uncertainty measure						LR trend	Autocorrelation				Seasonality			
	Annual	Monthly			Daily			Annual	Monthly		Daily		Monthly		Daily	
	$\hat{\tilde{x}}$	$\widehat{\sigma_{\tilde{x}}}$	$\widehat{\sigma_{\tilde{x}}}/\widehat{\sigma_{\widetilde{\mathrm{U/T}}}}$	$\hat{\rho}(\tilde{x}, \widetilde{\mathrm{U/T}})$	$\widehat{\sigma_{\tilde{x}}}$	$\widehat{\sigma_{\tilde{x}}}/\widehat{\sigma_{\widetilde{\mathrm{U/T}}}}$	$\hat{\rho}(\tilde{x}, \widetilde{\mathrm{U/T}})$	$\hat{\delta}$	$\widehat{\alpha_1}$	$\widehat{\alpha_2}$	$\widehat{\alpha_1}$	$\widehat{\alpha_2}$	P-val.	Size	P-val.	Size
Ratios of article counts from alternative methods for isolating economic uncertainty																
$\mathrm{E^*/E}$	1.143*** (0.005)	0.019*** (0.002)	0.10	-0.575***	n/d			0.000 (0.001)	0.503*** (0.111)	0.254** (0.125)	n/d		0.341	0.005 (0.003)	n/d	
Factors induced by not scaling at all																
T	67554*** (5664)	0.217*** (0.024)	1.10	-0.265**	0.183*** (0.014)	0.50	-0.195	-0.055*** (0.014)	0.570*** (0.180)	0.276* (0.163)	0.048 (0.171)	0.123 (0.125)	0.000	0.150** (0.073)	0.000	0.371*** (0.049)
E	15200*** (741)	0.190*** (0.019)	0.96	-0.077	0.218*** (0.045)	0.59	-0.176	-0.017 (0.011)	0.570*** (0.139)	0.165 (0.117)	0.134 (0.118)	0.056 (0.056)	0.000	0.172* (0.094)	0.000	0.318*** (0.074)
$\mathrm{E^*}$	17373*** (858)	0.196*** (0.020)	1.00	-0.128	n/d			-0.017 (0.011)	0.595*** (0.142)	0.175 (0.121)	n/d		0.000	0.172* (0.094)	n/d	
Factors induced by scaling by total article counts rather than economic article counts																
E/T	0.237*** (0.013)	0.080*** (0.007)	0.41	0.588***	0.183*** (0.055)	0.50	-0.101	0.037*** (0.004)	0.627*** (0.129)	0.089 (0.149)	0.109* (0.055)	0.015 (0.043)	0.014	0.042 (0.041)	0.044	0.196** (0.085)
Uncertainty measures																
EU/T	0.022*** (0.002)	0.302*** (0.027)	1.54	0.944***	0.493*** (0.060)	1.35	0.618***	0.053*** (0.010)	0.888*** (0.099)	-0.073 (0.115)	0.090 (0.154)	-0.027 (0.163)	0.394	0.101 (0.065)	0.160	0.541*** (0.195)
EU/E	0.091*** (0.004)	0.239*** (0.024)	1.21	0.938***	0.513*** (0.064)	1.40	0.601***	0.017** (0.007)	0.910*** (0.124)	-0.123 (0.117)	0.132 (0.142)	-0.064 (0.175)	0.269	0.069* (0.037)	0.033	0.653*** (0.189)
U/T	0.047*** (0.003)	0.197*** (0.019)	1	1	0.366*** (0.046)	1	1	0.040*** (0.006)	0.882*** (0.104)	-0.180 (0.122)	0.562*** (-0.174)	-0.174 (0.147)	0.008	0.058*** (0.020)	0.064	0.407** (0.188)

Notes: See notes to Table 4. The benchmark measure is $\widetilde{\mathrm{U/T}}$ for the FT (see Section 5.2). Due to the time-intensive nature of data collection, for the FT we collected only a subset of the variables that we collected for The Times.

Figure 8: Factors induced in the uncertainty measure (cf. U) by alternative scaling methods



Notes: Based on undeduplicated article counts. Ratios are normalised by their own mean over the sample plotted.

Figure 9: Factors induced in the uncertainty measure (cf. U) by using proxies for total article counts

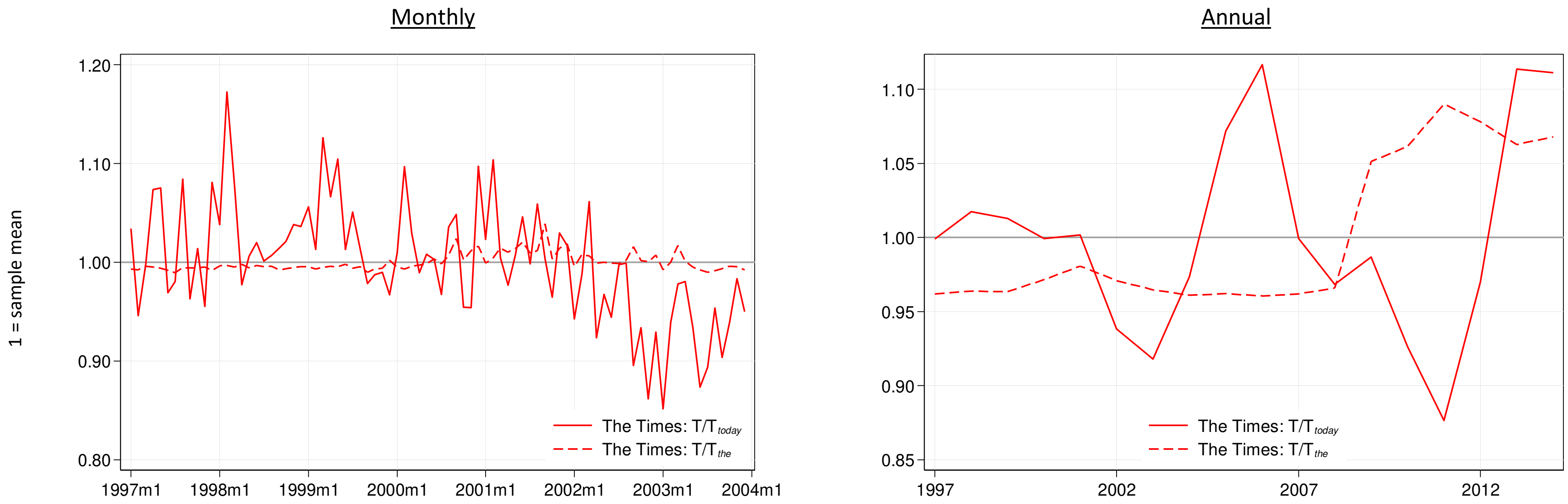


Figure 10: Comparing between the same uncertainty measures in the FT vs. The Times

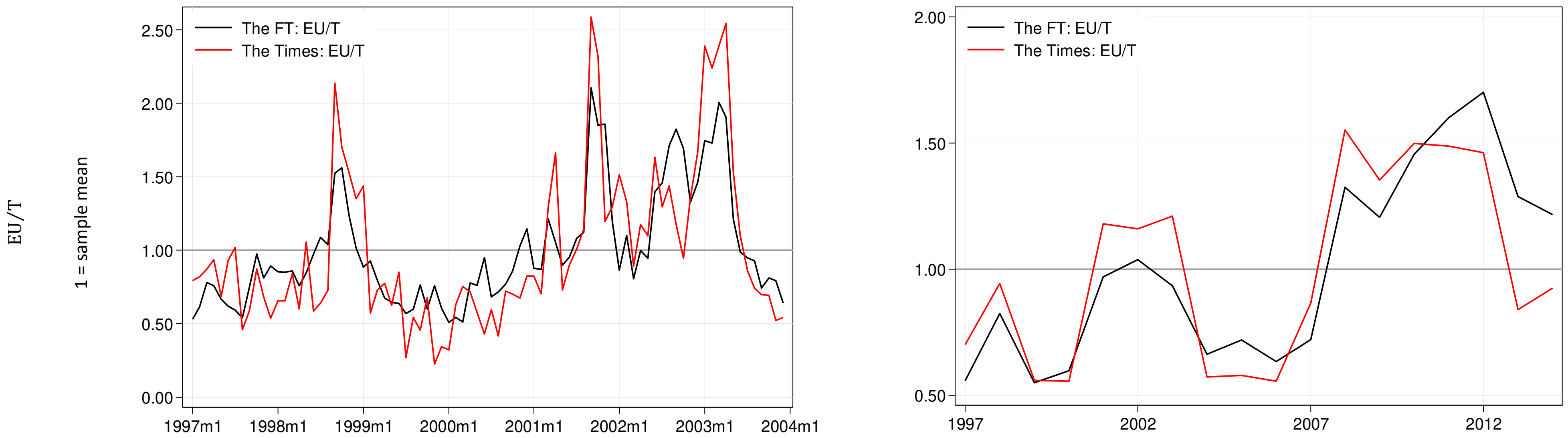
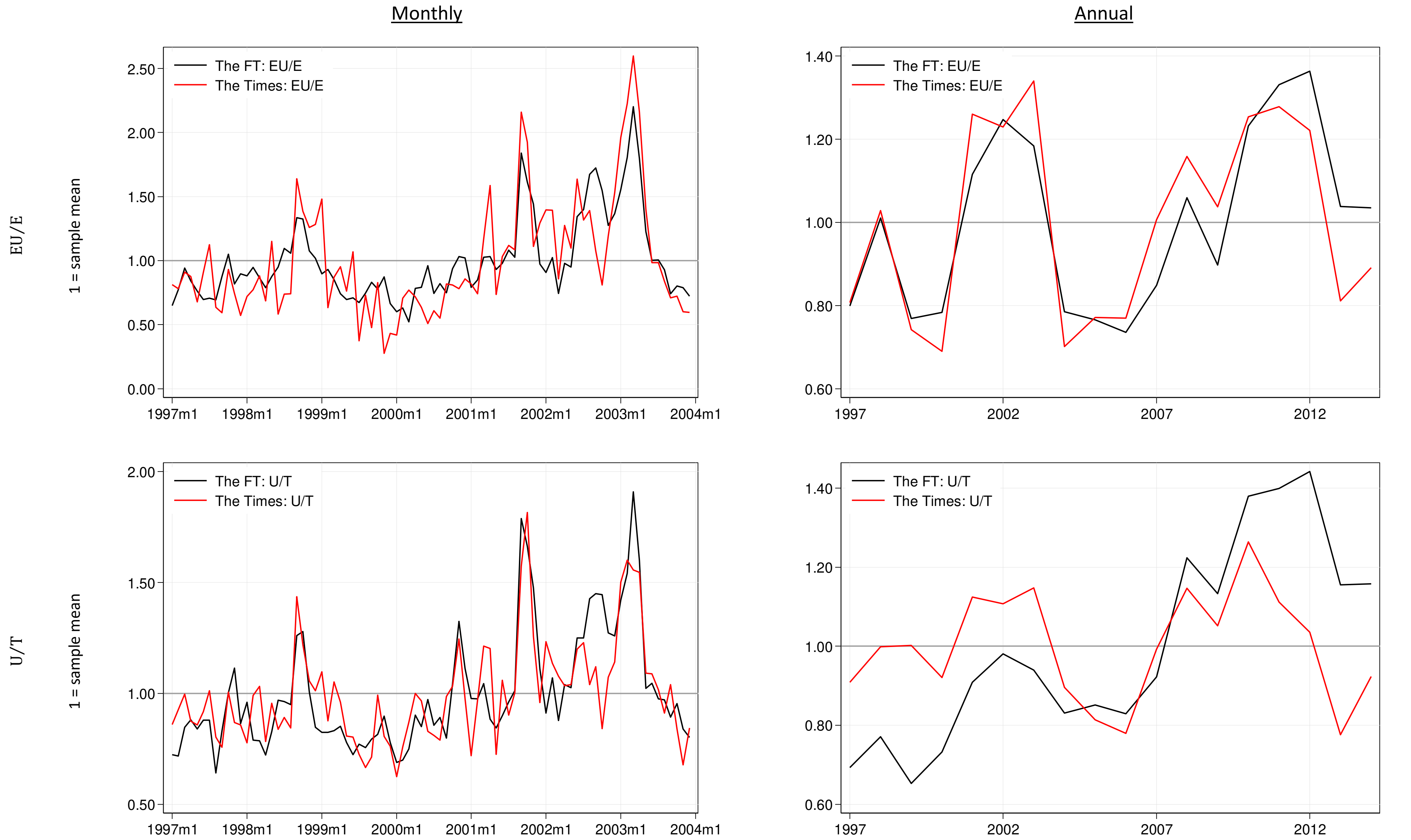
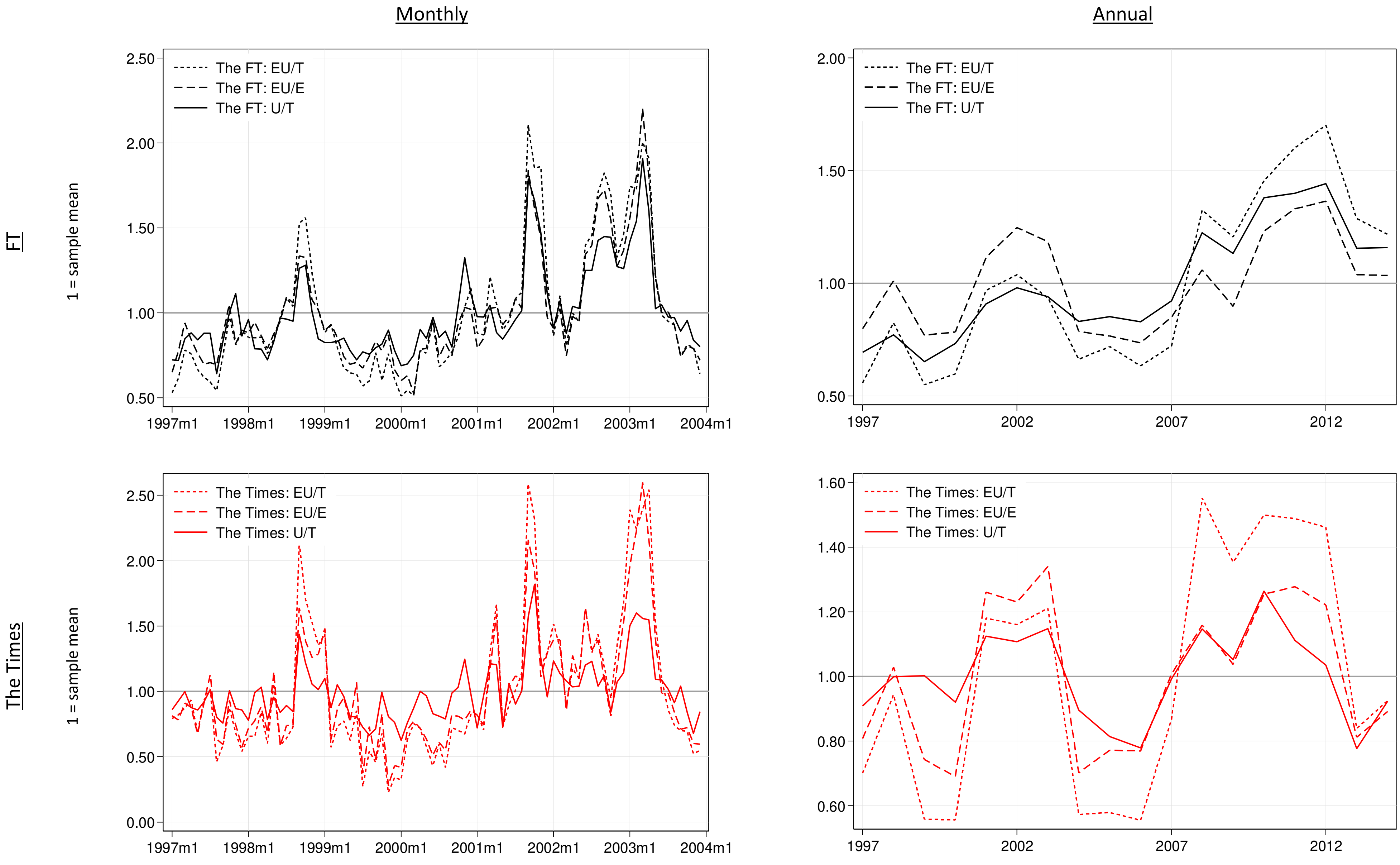


Figure 10: cont'd



Notes: Based on undeduplicated article counts, to better illustrate the effects of approaches used in the literature, where deduplication is not considered. Values for U/T shown for the FT may therefore differ from those shown in Section 6. Ratios are normalised by their own mean over the sample plotted.

Figure 11: Comparing between alternative uncertainty measures on the same newspaper



Notes: See notes to Figure 10.

6 Comparison to other uncertainty proxies

This Section compares U to other popular uncertainty measures. The first is the widely-cited economic *policy* uncertainty (EPU) measure of Baker et al. (2015) (Section 6.1). The second is stock returns volatility, σ, for which compare in considerable detail through narrative analysis (Section 6.2) and quantitative analysis (Section 6.3), focussing on patterns time-variation in contemporaneous association between the measures (Section 6.3.3) and Granger causality tests in the high-frequency bivariate dynamics (Section 6.3.4).

6.1 Comparison to the EPU measure of Baker et al. (2015)

As a first step in our comparative analysis,

Figure 12: Comparison of U and EPU measure of Baker et al. (2015) at monthly frequency
plots U against the widely-cited economic policy uncertainty (EPU) measure of Baker et al. (2015)[37]. We show the UK variant of EPU, which in the version as at October 2015 is based on counts of news articles from the FT and The Times of London, scaled by total articles.

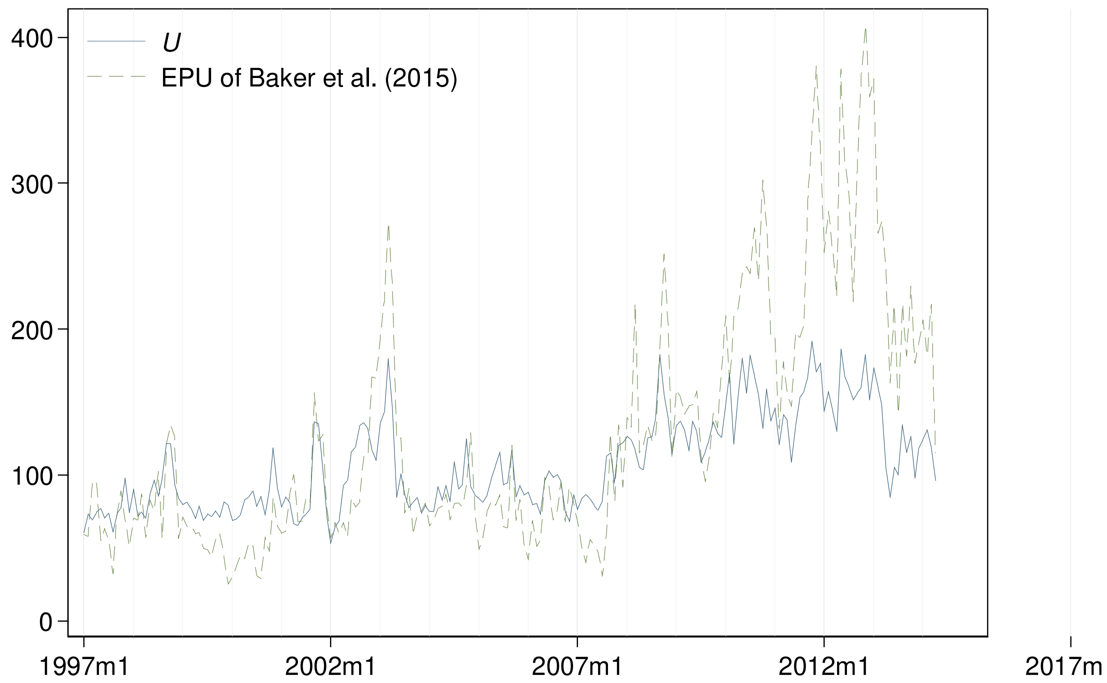
While the two measures track uncertainty about different sets of economic topics – all economic topics vs. economic *policy* only – they are strongly correlated (Pearson's rho = 0.86, Spearman's rho = 0.84) and exhibit similar time profiles.

The most obvious difference if that EPU peaks substantially higher around 2011-13. This coincides with an intense period of economic policy uncertainty about the Eurozone, with substantial spillovers to UK economic policy. Note that the measures are normalised to their own means for 1997-2010 (the period used by Baker et al. (2015)) so the fact that EPU appears higher than U on the chart does not mean that a larger fraction of articles referenced economic policy uncertainty. Indeed, economic policy articles are a subset of all economic articles. Rather, it indicates that EPU became increasingly intense as the crisis wore on, consistent with an understanding of the global financial crisis as having started in the markets but then increasingly over time moved onto public sector balance sheets through bailouts and fiscal stimulus, so that the uncertainty was progressively translated into the policy sphere. According the EPU then, the uncertainty surrounding the later phases of the crisis was approaching double that in the immediate aftermath of the Lehman collapse.

The EPU measure is also conflated with time-variation in the economic policy share of total news volume, due to the choice of scaling as discussed in Section 5.4. To the extent that economic policy topics take a higher share of news volume in the later stages of the crisis, over and beyond any part of that increase that is driven by uncertainty, the measure will be distorted higher. It would be impossible to know for sure how much of the elevation in EPU is due to this spurious effect rather than the latent economic policy uncertainty that EPU is attempting to measure, without establishing the factors driving the economic policy share of news volume.

[37] Data downloaded from http://www.policyuncertainty.com/europe_monthly.html on 31 January 2016.

Figure 12: Comparison of U and EPU measure of Baker et al. (2015) at monthly frequency



6.2 Narrative analysis

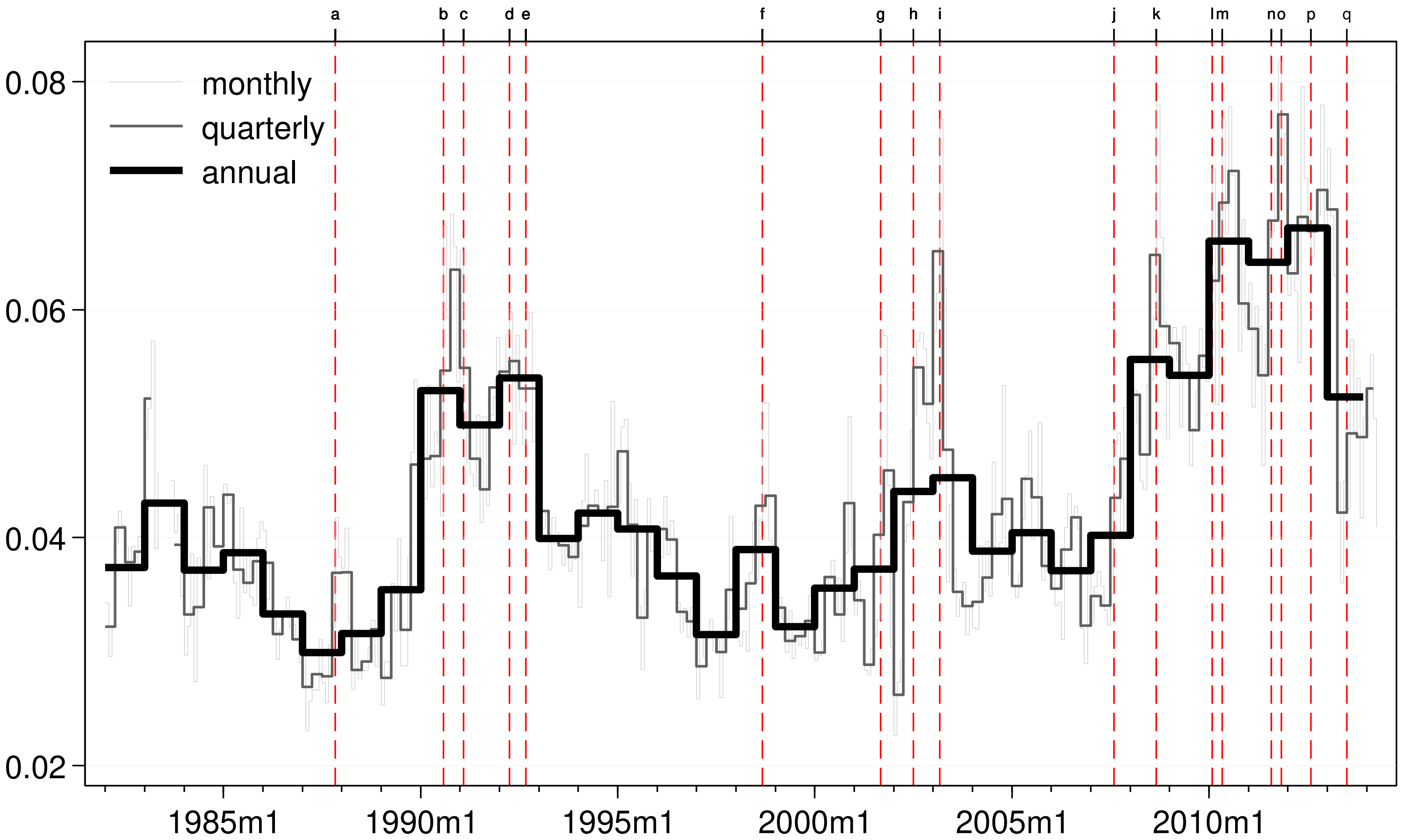
U and σ are plotted separately in Figure 13 and Figure 14 respectively with monthly, quarterly and annual frequencies overlaid, and a similar set of events labelled as in the related literature[38]. Figure 15 through Figure 17 show the same data with U and σ (or $\ln\sigma$)[39] overlaid for ease of comparison, in levels and first differences. Additional zoomed-in plots of selected events are interleaved with the discussion below.

We focus on the monthly series, in line with related literature (e.g. Campbell et al., 2001; Schwert, 1989b). Dynamics related to events of interest can be obscured at much lower frequency, and the amount of detail can become unmanageable at much higher frequency, although we use weekly views for some examples.

[38] E.g. Haddow et al. (2013) on UK data and Figure 1 of Bloom (2009) on US data. As is natural for a UK series, we include more UK/EU focussed events and fewer US-specific events than Bloom (2009).

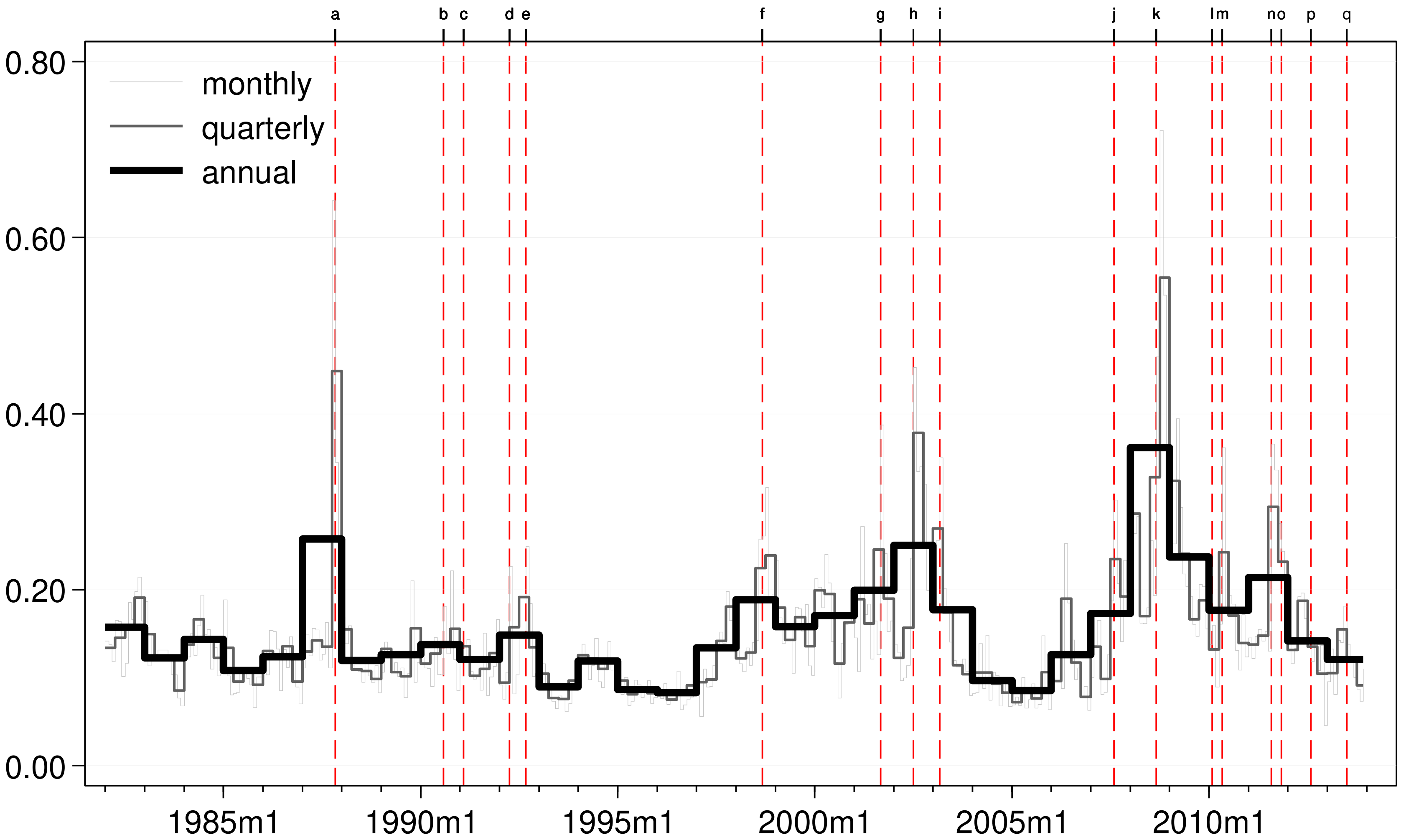
[39] Absent theory to pin down the functional form of the expected relationship between U and σ, plotting them against each other in levels, rather than subject to some other monotonic transformation, is arbitrary, albeit in line with the few precedents in the literature (e.g. Baker et al., 2013; Haddow et al., 2013). We show a plot with $\ln\sigma$ (Figure 16) because naturally brings out more of the temporal structure in volatility during periods when it is relatively low (e.g. the early 1990s). However, this does not yield major new insights, nor did experimentation with other monotonic transformations and quantile plots.

Figure 13: Aggregate news-media uncertainty, U



Notes: See Table 6 on pg.41 for a description of the events associated with vertical red dashed lines. The annual measure is pro-rated for 1983 due to database gap for 2 June to 8 August; the corresponding months and quarters are set to missing.

Figure 14: Aggregate stock returns volatility, σ



Notes: See Table 6 on pg.41 for a description of the events associated with vertical red dashed lines.

Table 6: List of events lines in Figure 13 and Figure 14

Label	Date	Description
a	19 Oct 1987	Black Monday stock market crash
b	2 Aug 1990	Gulf War: Operation Desert Shield starts
c	28 Feb 1991	Gulf War: Operation Desert Storm ends
d	9 Apr 1992	UK election
e	16 Sep 1992	UK exits European Exchange Rate Mechanism (ERM)
f	23 Sep 1998	Collapse of Long-Term Capital Management (LTCM) [dated to the recapitalisation]
g	11 Sep 2001	9/11 terrorist attacks in the US
h	Jul 2002	WorldCom accounting scandal
i	20 Mar 2003	Iraq War starts
j	9 Aug 2007	Subprime crisis onset: BNP Paribas freezes funds exposed to subprime mortgages
k	15 Sep 2008	Lehman Brothers bankruptcy
l	15 Mar 2010	Greek government debt crisis
m	12 May 2010	UK hung parliament [6-12 May]
n	2 Aug 2011	US government debt ceiling deadline
o	27 Oct 2011	EU bailout fund expanded to EUR1 trillion
p	26 Jul 2012	ECB governor Draghi promises to do ‘whatever it takes to save the Euro’
q	8 Jul 2013	French President Francois Hollande claims ‘crisis in the Eurozone is over’

Figure 15: U and σ overlaid, by frequency

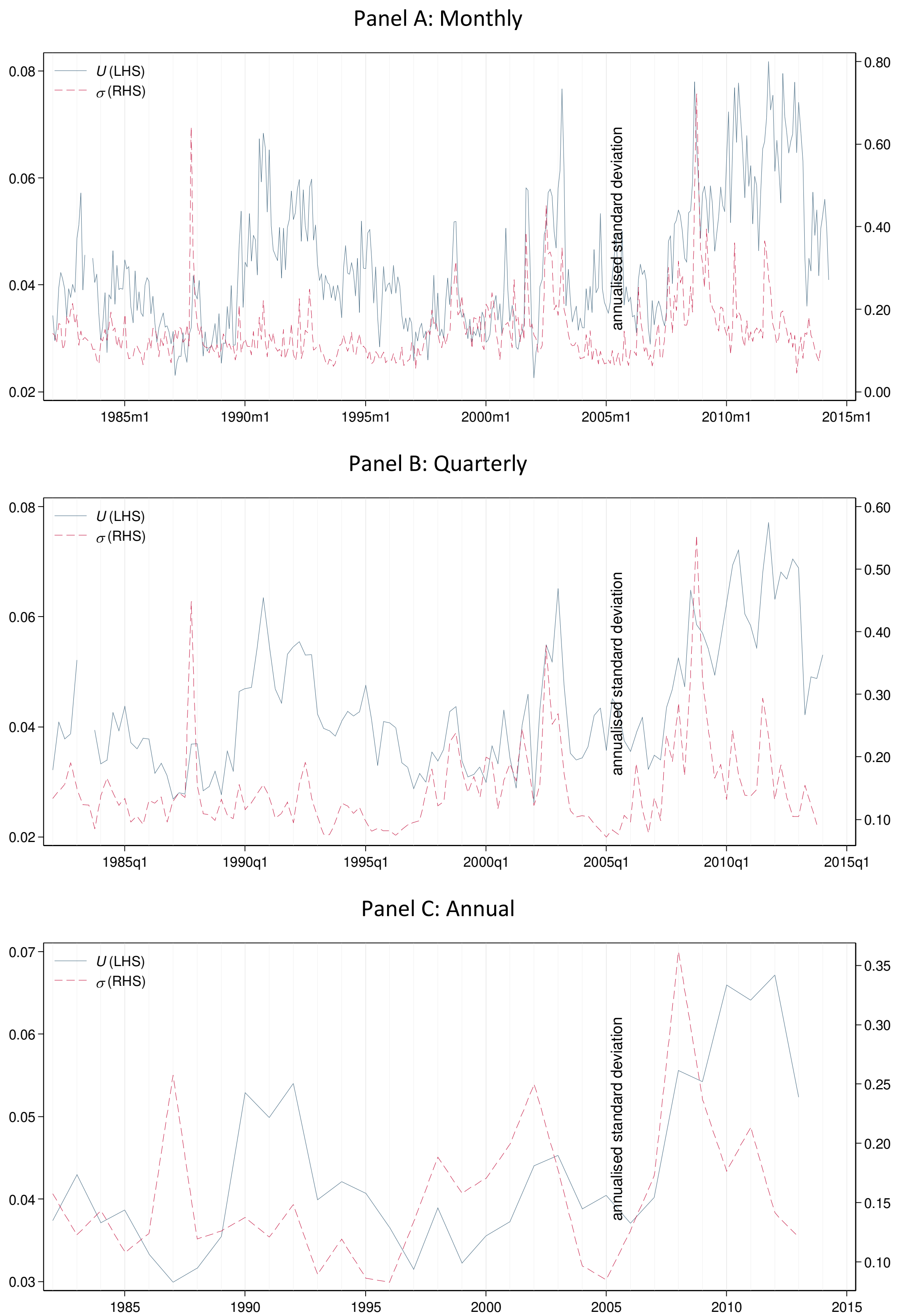
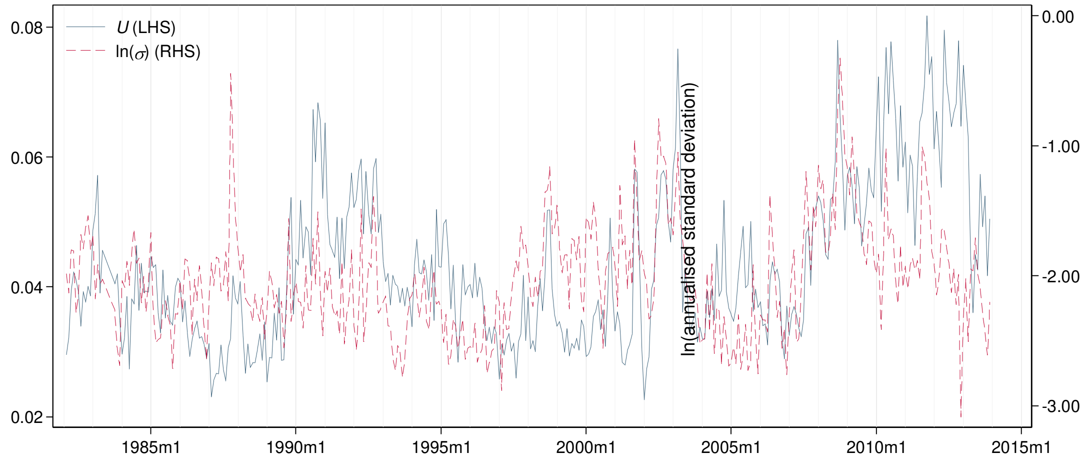
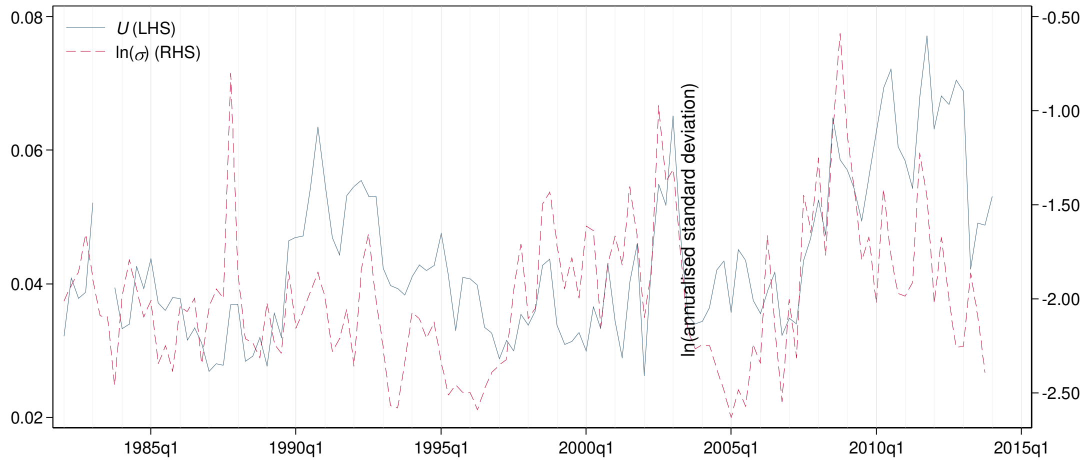


Figure 16: U and $\ln(\sigma)$ overlaid, by frequency

Panel A: Monthly



Panel B: Quarterly



Panel C: Annual

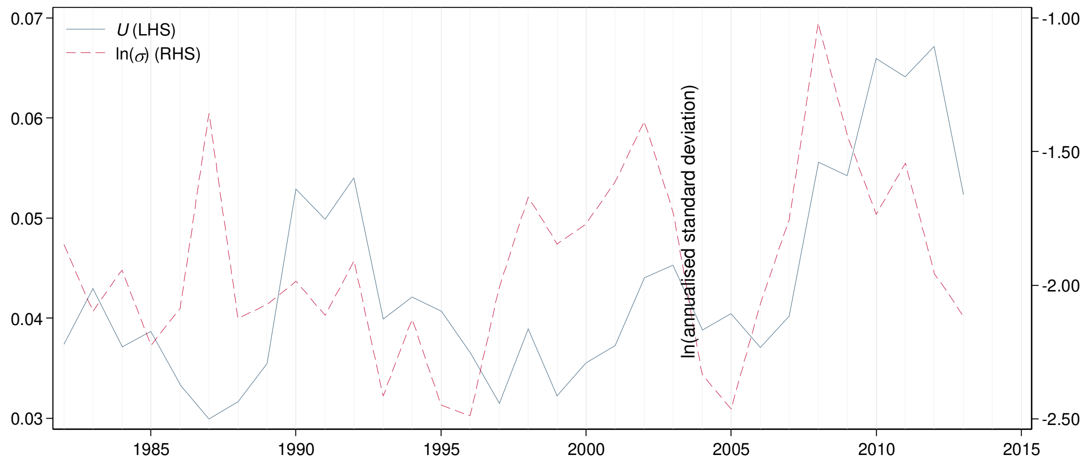
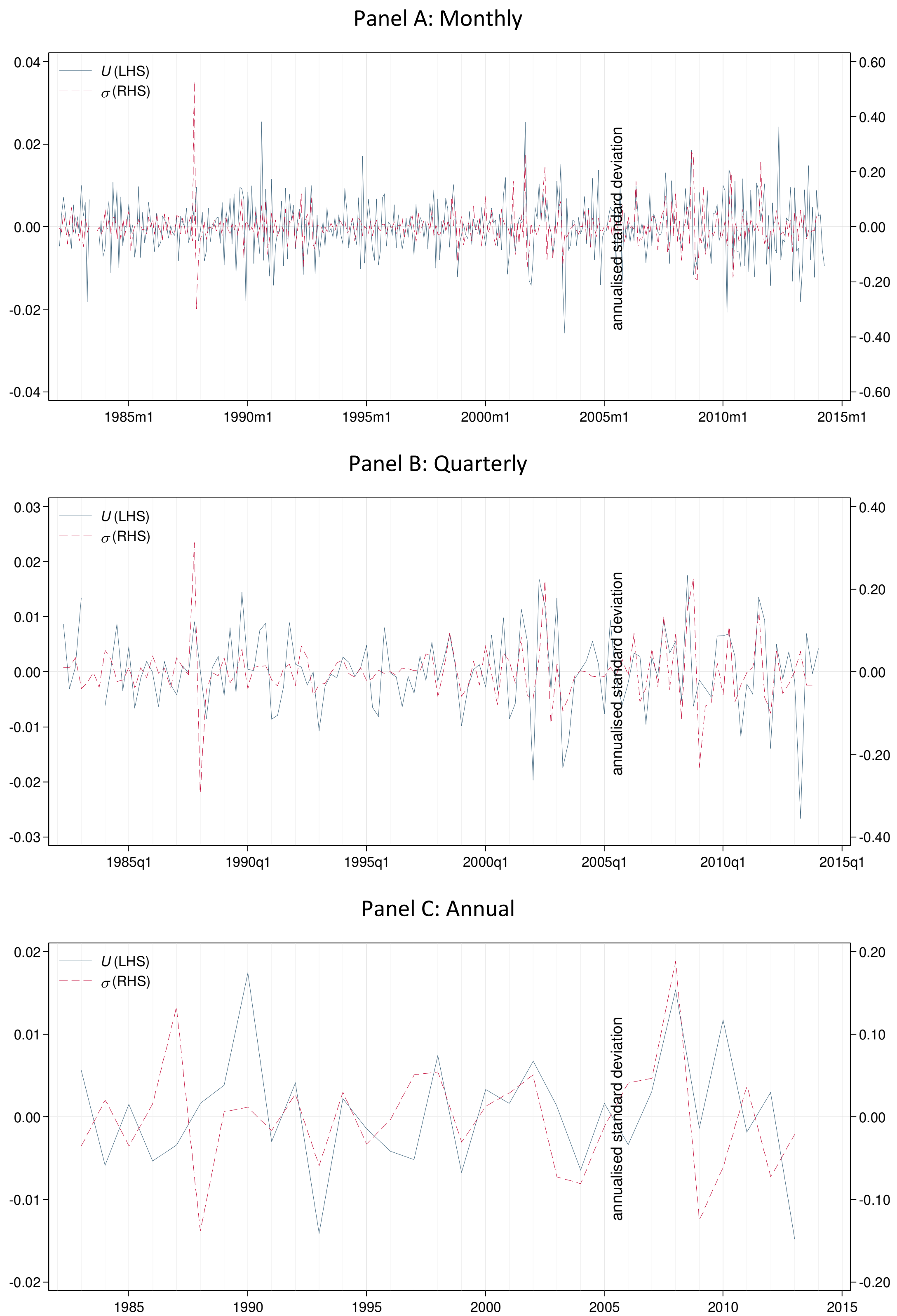


Figure 17: First differences – aggregate news-media uncertainty, ΔU vs. stock returns volatility, $\Delta\sigma$

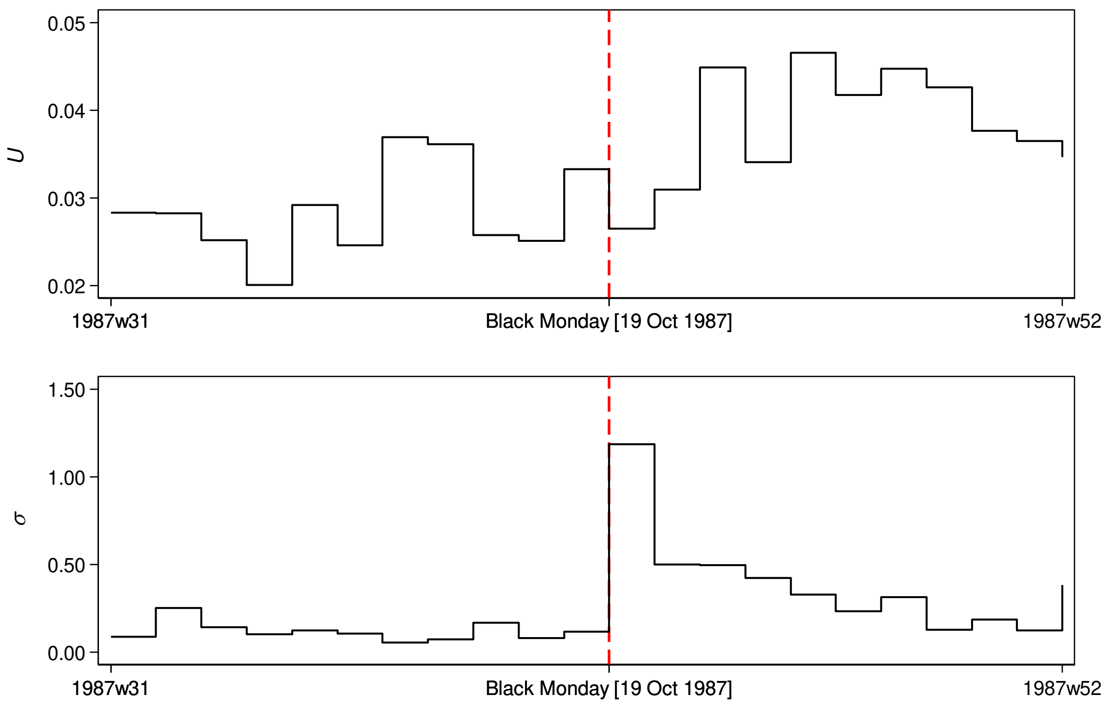


Let us start by noting two differences between U and σ that are apparent over the full sample.

First, the time profile of σ is dominated by large movements over short periods around major financial dislocations: the stock market crash of Black Monday (19 October 1987), and the collapse of Lehman Brothers (15 September 2008). U is elevated around these events too, but also reaches comparable levels at certain times in the first half of the 1990s, and in the years of the Global Financial Crisis (GFC) and Eurozone (EZ) crisis. We look at these periods in more detail below, but overall it appears that U may offer a broader-based view of uncertainty than σ. Naturally, this conclusion is tempered if we apply a skew-reducing transform to σ, such as $\ln(\sigma)$ (Figure 16), or if we consider only ordinal information.

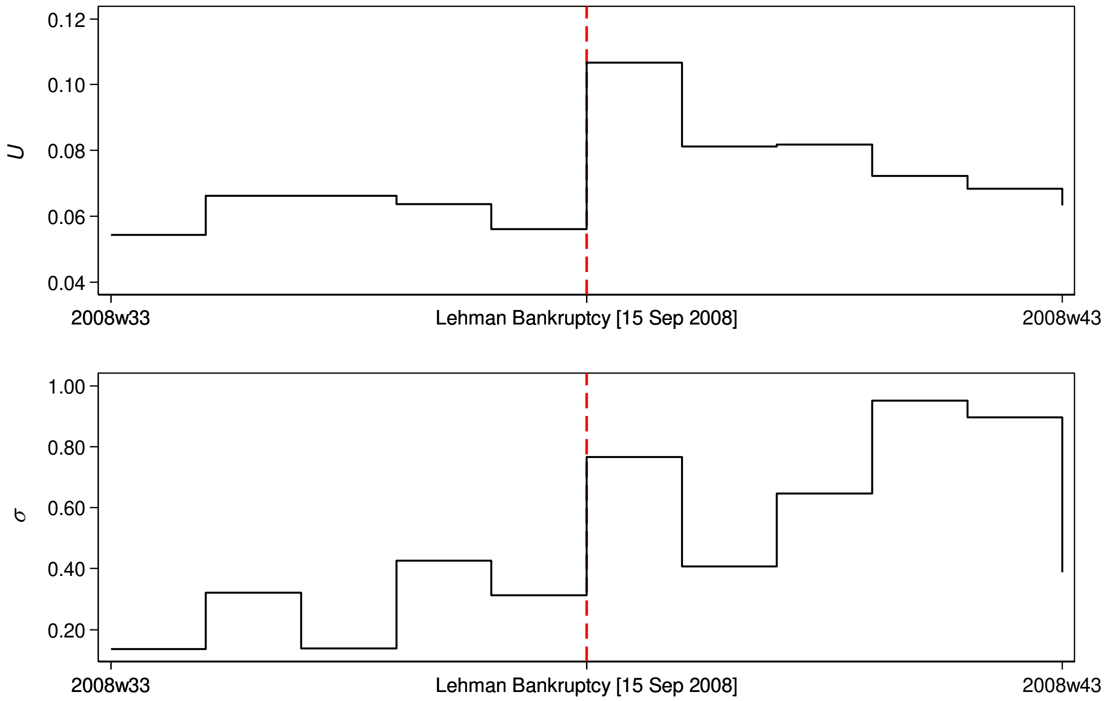
Second, the measures respond differently to major financial dislocations. Black Monday is associated with the largest positive monthly and quarterly increase in σ, to reach its second highest level in our sample. By comparison, the response of U is modest, similar to the first principal component of the battery of uncertainty proxies in Haddow et al. (2013). Here, U may better reflect the level of uncertainty in the real economy, since transmission of the stock market dislocation to the real economy was relatively muted – indeed the 1980s ended in the so-called Lawson boom. The two-week lag in the rise in U is atypical. Usually U responds more promptly to major narrative events, as we will see below.

'Black Monday' – 19 October 1987
Weekly



The collapse of Lehman Brothers coincides with large spikes in both U and σ to near their sample maxima. At monthly frequency the local peak (also the sample maximum) of σ is delayed until following month (October 2008). This is at least in part because, with the emergency market closures, fewer than half of the trading days in September fell after the Lehman collapse. By contrast, U is able to reflect the spike in uncertainty more promptly because the newspapers kept publishing during this period. However, σ then drops back sharply within a few months and continues to trend down through the end of the sample, whereas U remains elevated until 2013q1, consistent with the conventional view that economic uncertainty remained elevated for years after the onset of the crisis.

Lehman Bankruptcy – 15 September 2008
Weekly

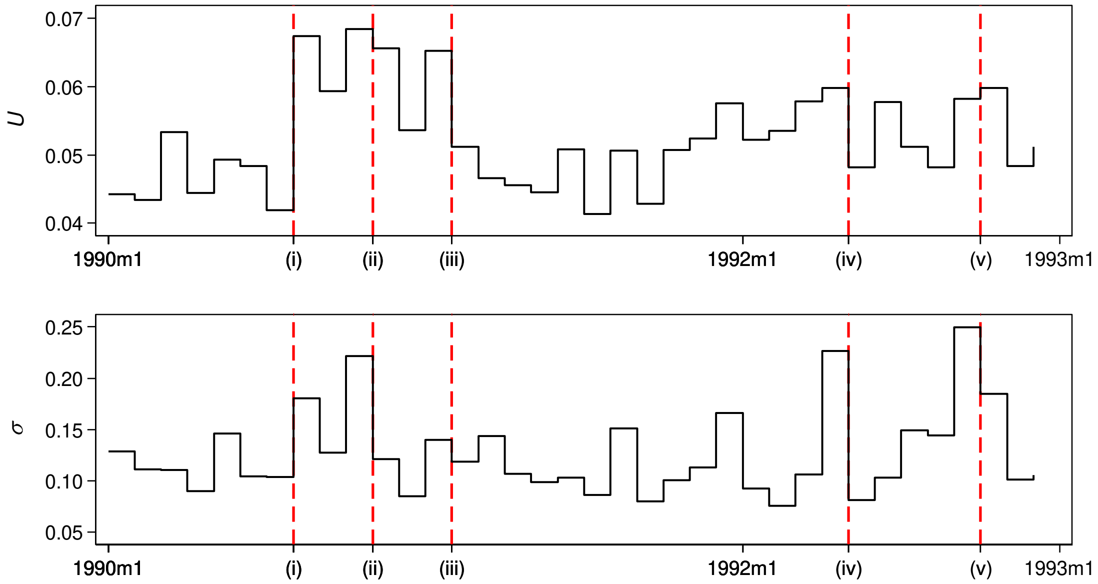


Arguably then, U is a more convincing indicator of the level of latent uncertainty during and following financial crises, even if the reason for the lag in response after Black Monday is unclear.

Let us now zoom in on a few periods that exhibit particularly substantial movement in U and/or σ.

In the early 1990s we see a sustained rise in U during the Gulf War (2 August 1990 to 28 Feburary 1991). σ rose at the start of the war but this higher level was not sustained. The rise in November 1990 coincides with the leadership battle in the ruling Conservative Party, leading to substantial political uncertainty and ultimately the ouster of Prime Minister Thatcher of 28 November 1990. Both measures show a drop off in uncertainty after the national election of 9 April 1992, and a rise coincident with the crisis around the UK's membership of the European Exchange Rate Mechanism (ERM) which was resolved by a forced exit from the mechanism in 16 September 1992.

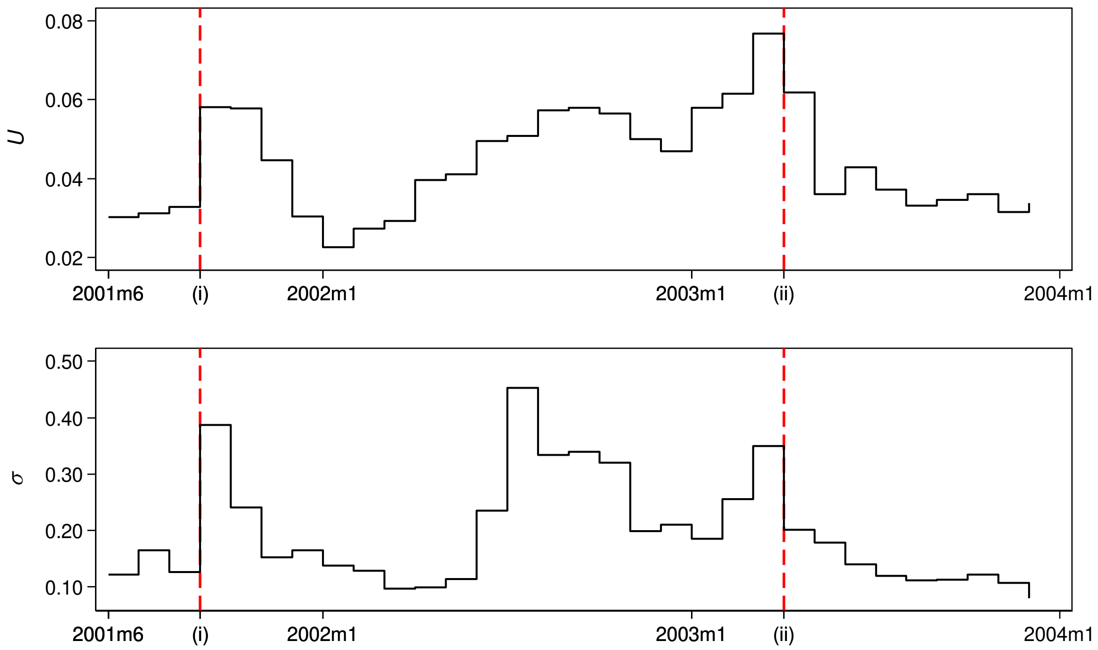
Early 1990s
Monthly



(i) Gulf War begins [2 Aug 1990]
(ii) Ouster of Prime Minister Thatcher [28 Nov 1990]
(iii) Gulf War ends [28 Feb 1991]
(iv) UK election [9 Apr 1992]
(v) UK exits ERM [16 Sep 1992]

The terrorist attacks in the US on Tuesday, 9 September 2001 ("9/11") and the subsequent war in Iraq are associated with clear peaks in uncertainty.

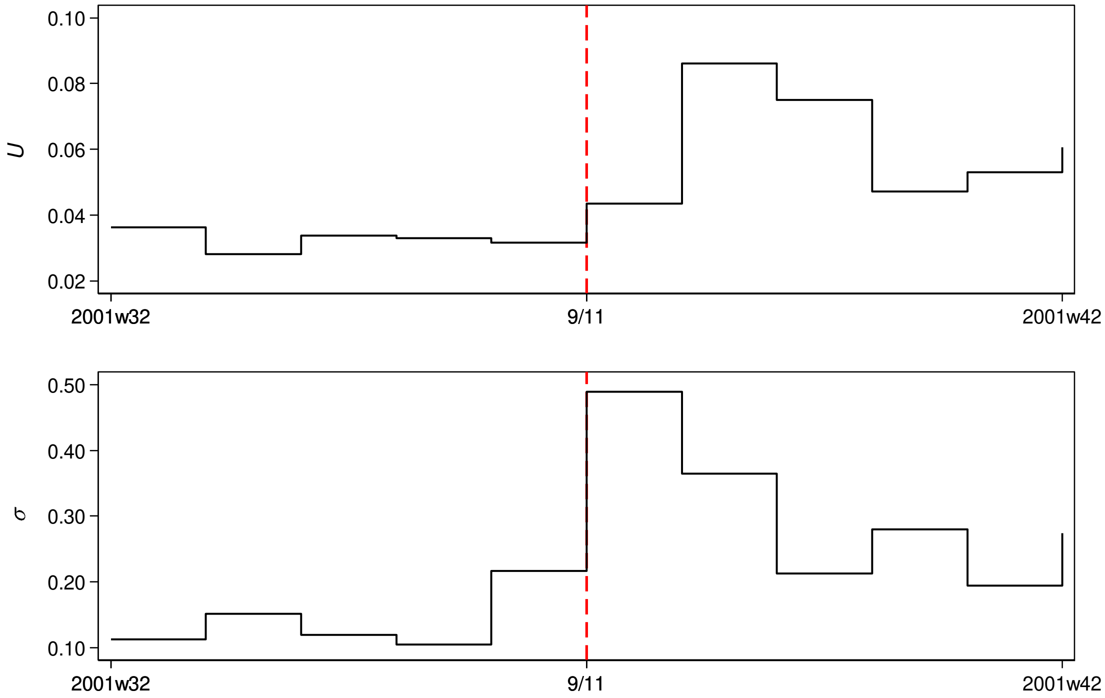
2001-3
Monthly



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

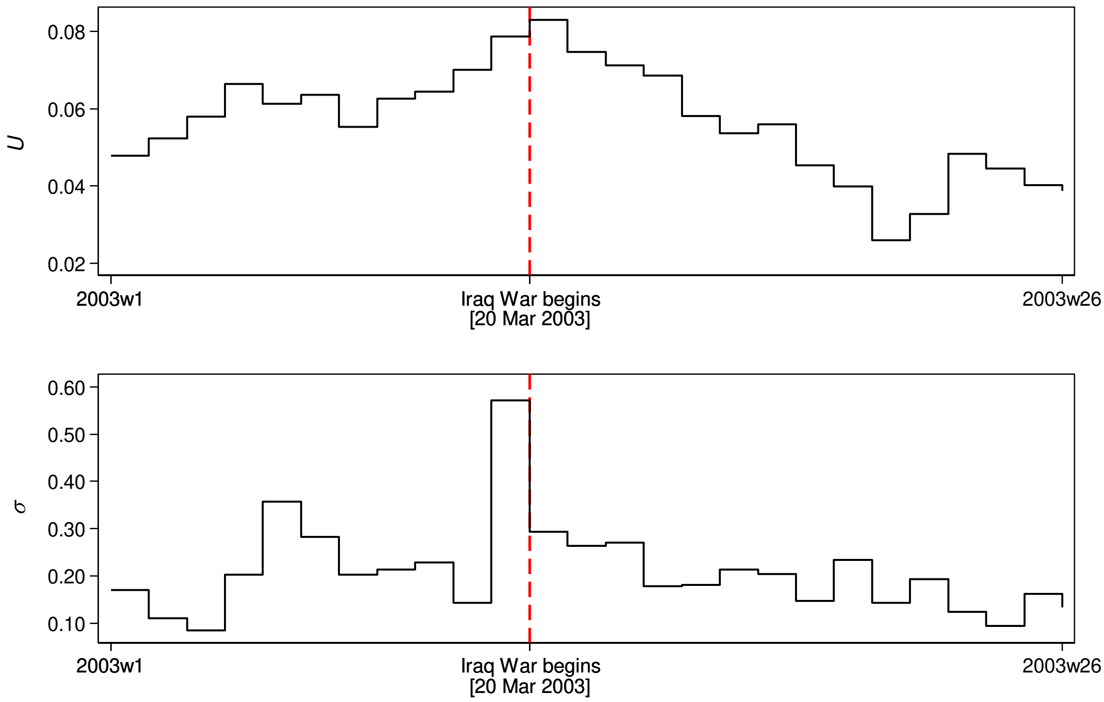
9/11 is the quintessential shock event. It is associated with a correspondingly sharp jump in both U and σ, although at weekly frequency the full jump in U is delayed by a week, at least in part because two of the week's six print editions pre-dated the attacks (of course the same is true for the daily returns from which σ is estimated, but the spike in volatility in the second part of the week sufficed to create a peak for the week overall).

9/11



The Iraq War, by comparison, came after a multi-month build-up during which there was great uncertainty as to whether and when the war would commence, and which nations would join the US. This can be seen in both U and σ. The ramp up through 2002 may also be associated with the ongoing war in Afghanistan. When the US and its allies did finally attack on 20 Mar 2003, they toppled the incumbent regime in days, and much of the pent up uncertainty dissipated, though this occurs more gradually in U than in σ.

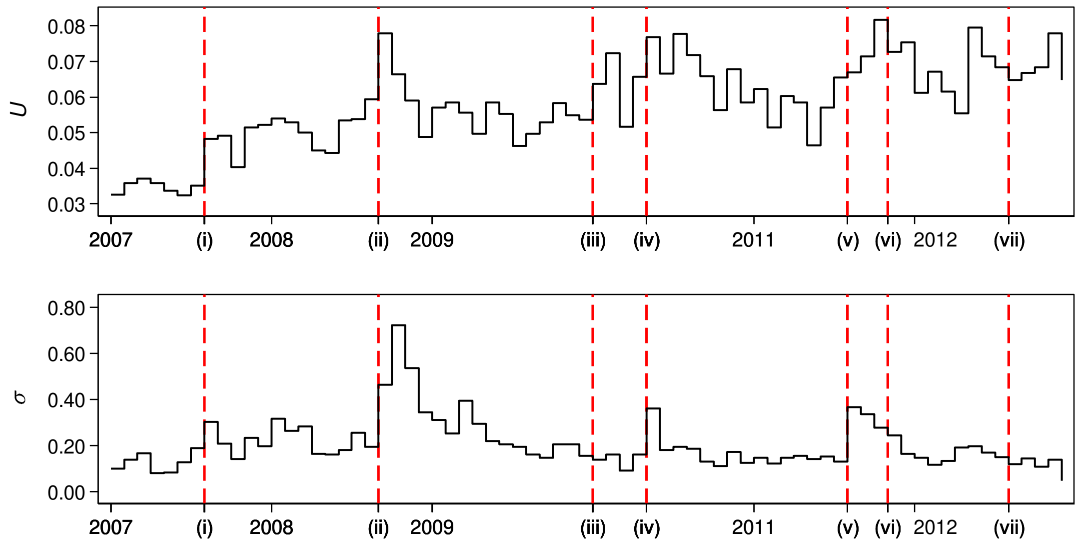
Iraq War
Weekly



Moving on to the several years spanned by the Global Financial Crisis (GFC) and Eurozone (EZ) Crisis, we can see structure in U and σ consistent with conventional narrative accounts[40]. Both measures exhibit an uptick coincident with the event conventionally used to date the start of the GFC: BNP Paribas freezing three of its funds, which contained CDOs with exposure to subprime mortgages, on 9 August 2007. As discussed above, the collapse of Lehman Brothers stands out clearly in both measures.

Global Financial Crisis & Eurozone Crisis

Monthly



(i) BNP Paribas freezes funds exposed to subprime mortgages [9 Aug 2007]
(ii) Lehman Brothers collapses [15 Sep 2008]
(iii) Onset of of Greek debt crisis [early 2010]
(iv) UK hung parliament [6-12 May 2010]
(v) US government debt ceiling deadline [2 Aug 2011]
(vi) EU bailout fund expanded to 1 trillion Euro [27 Oct 2011]
(vii) Draghi's "whatever it takes to save the Euro" [26 Jul 2012]

However, there are also clearly differences between U and σ in the post-Lehman period, beyond simply the suppression in the level of σ noted above.

Much of the subsequent variation in U can be linked circumstantially to events in the EZ crisis, or other key events. In conventional accounts the EZ crisis began with the Greek debt crisis in early 2010, at which time U rises but σ remains relatively muted. The sharp fall in U in 2013Q2 (see Figure 13) appears to be associated with a period of relative calm such that the French President felt able to claim in a speech on 8 June 2013 that "the crisis in the Eurozone is over"[41], even if this may have been premature in retrospect.

Significant uncertainty-related events in the post-Lehman period are so numerous, drawn out, and overlapping in time that persuasively identifying the relationship between all substantial movements in U and particular events is a demanding task beyond the present scope. However, we offer a couple of examples by way of illustration.

---

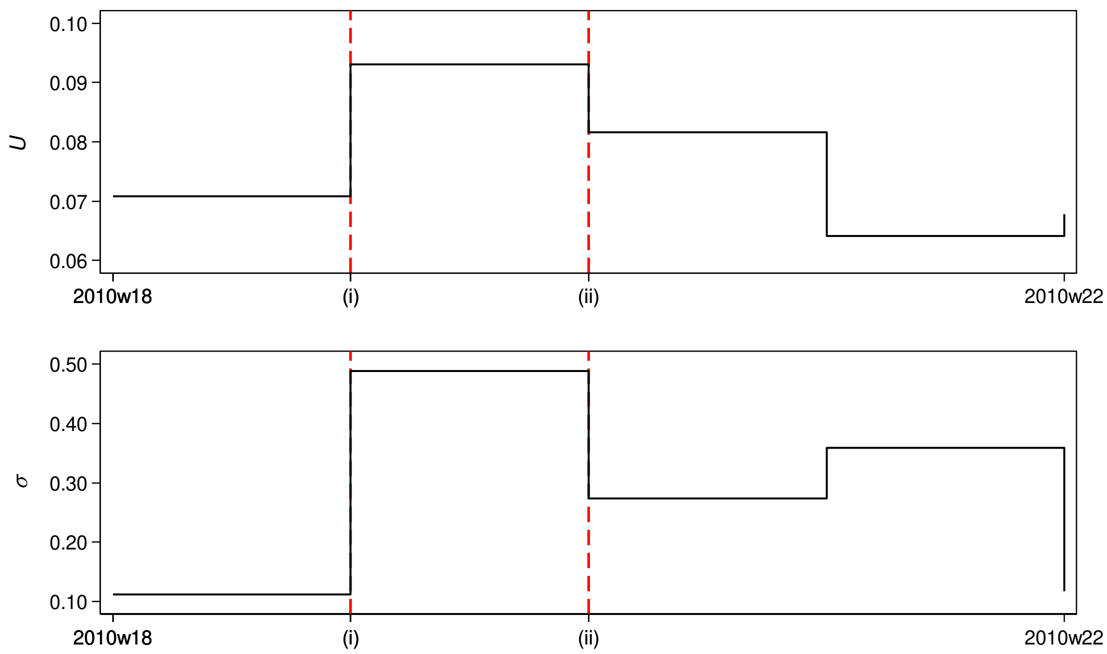
[40] The Guardian's (2014) Eurozone crisis timeline was particularly helpful in this analysis.

[41] Source: http://www.theguardian.com/world/2013/jun/09/francois-holland-eurozone-crisis-over

First, the national elections in the UK on 6 May 2010 led to a hung parliament for the first time since 1974, creating a great deal of uncertainty as to who would end up governing the country and setting policy. In the week between the election, and the formation of a coalition government on 12 May which dispelled this uncertainty, both U and σ rose sharply to around their 99th percentiles (see Figure below). The most obvious EZ crisis event that might be conflated in the dynamics here is the agreement on 10 May of EZ governments to a EUR500trillion rescue plan. This was obviously a significant event, but one that was neither sudden nor unexpected by the time it arrived. The initial turbulence of the Greek debt crisis had passed by this point, with Greece having accepted a bailout on 15 April 2010, and the next bailout crisis (in Hungary) was not to occur until July.

UK Hung Parliament – May 2010

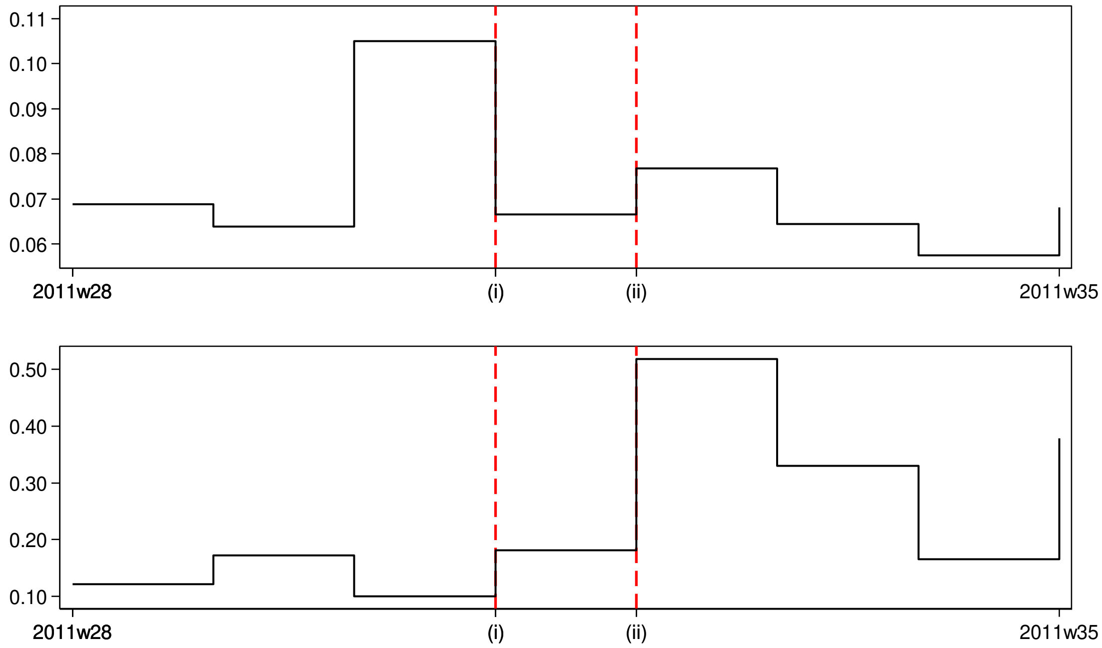
Weekly



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

The second sustained ramp up in U during the EZ crisis, from June to October 2011, coincides with events on both sides of the Atlantic. In the US, Congress was threatening to not extend the US government debt ceiling by the deadline of 2 August 2011, and thus force the US into default, generating uncertainty about the knock on effects for the global financial system. In the EZ, the government debt crisis that had already prompted bailouts for Greece, Ireland and Portugal was threatening to engulf Spain and Italy too. Thus on 8 August 2011 the ECB re-activated its Securities Markets Program to buy Spanish and Italian government bonds in large quantities. The intra-month timing in U suggests that the FT focussed more on uncertainty associated with the US debt ceiling deadline. σ does not indicate substantial uncertainty in advance of either event. Furthermore, σ declines in subsequent months, though remaining elevated.

US debt ceiling crisis and ECB action to stem contagion, August 2011



(i) US government debt ceiling deadline (2 August)
(ii) ECB starts buying Spanish and Italian bonds (8 August)

Finally, the jump up in U in May 2012, to its second highest value in the sample, coincides with a resurgence of the EZ crisis. On 6 May the Greek elections resulted in a majority for parties opposing the international bailout. However, they failed to form a coalition government, so new polls were announced for 17 June, leaving substantial uncertainty about the Greek bailout and therein the future of the Eurozone. On top of this, Spain's 4th largest bank asked for a bailout on 25 May.

6.3 Quantitative analysis

6.3.1 Sample summary statistics

Table 7 reports sample summary statistics for U and σ in levels and first differences. The mean of U is around 4.3%, and the interquartile range is 3.4% to 5.1% at monthly frequency. The mean of σ is around 15%, and the interquartile range is 9.9% to 16.5%.

Both variables are approximately log-normally distributed (though $\sqrt{U}$ gives a better approximation to normality for the daily data used in Section 6.3.4). They exhibit lower variance at lower frequency. consistent with mean reversion, and higher excess kurtosis at higher frequency. First differences are closer to normally distributed, with zero mean consistent with stationarity (the skewness in $\Delta\sigma$ is primarily due to a few large positive outliers). Similarly, Table 8 shows that the distribution of positive vs. negative movements is equally balanced.

Table 7: Sample summary statistics, 1984–2012

Variable / frequency		Mean	Std. dev.	Percentiles 25th	50th	75th	Skew-ness	Kurtosis	# of obs.
U	daily (ex.)	0.0443	0.0216	0.0289	0.0407	0.0559	0.972	4.500	7326
	daily	0.0436	0.0211	0.0286	0.0404	0.0550	0.963	4.558	8909
	weekly	0.0432	0.0142	0.0329	0.0402	0.0508	0.951	3.908	1508
	monthly	0.0432	0.0124	0.0336	0.0399	0.0510	0.895	3.165	348
	quarterly	0.0432	0.0117	0.0340	0.0410	0.0506	0.849	2.975	116
	annual	0.0431	0.0105	0.0366	0.0399	0.0499	0.972	2.927	29
σ	daily	0.148	0.151	0.050	0.109	0.195	3.361	26.079	7326
	weekly	0.139	0.096	0.081	0.115	0.165	3.346	23.738	1508
	monthly	0.147	0.082	0.099	0.124	0.165	2.834	15.218	348
	quarterly	0.151	0.075	0.104	0.132	0.170	2.461	11.283	116
	annual	0.157	0.062	0.120	0.142	0.177	1.372	5.146	29
ΔU	daily (ex.)	0.0000	0.0235	-0.0145	-0.0001	0.0146	0.068	3.782	7325
	daily	0.0000	0.0235	-0.0147	-0.0001	0.0146	0.071	3.902	8908
	weekly	0.0000	0.0107	-0.0064	0.0001	0.0067	0.159	4.623	1507
	monthly	0.0001	0.0070	-0.0042	0.0003	0.0044	0.128	4.067	347
	quarterly	0.0003	0.0068	-0.0034	0.0001	0.0049	-0.042	3.390	115
	annual	0.0011	0.0067	0.0034	0.0016	0.0036	0.462	3.642	28
$\Delta\sigma$	daily	0.000	0.184	-0.088	-0.000	0.088	0.038	11.485	7325
	weekly	-0.000	0.088	-0.041	0.001	0.039	0.974	23.603	1507
	monthly	-0.000	0.068	-0.029	0.002	0.023	1.736	16.419	347
	quarterly	-0.000	0.074	-0.034	-0.003	0.026	0.474	8.362	115
	annual	-0.000	0.070	-0.047	0.009	0.039	0.378	3.756	28

Notes: daily (ex.) excludes non-trading days as per the baseline daily VAR analysis in Section 6.3.4.

Table 8: Sample sign distribution of first differences, 1984–2012

Frequency	ΔU % of obs. by sign <0	=0	>0	Sign symmetry. p-value	$\Delta\sigma$ % of obs. by sign <0	=0	>0	Sign symmetry. p-value	# of obs.
daily (ex.)	50.3	0.2	49.5	0.240	50.0	0	50.0	0.157	7325
daily (full)	50.1	0.2	49.7	0.171	n/a			n/a	8908
Weekly	49.8	0	50.2	0.797	49.7	0	50.3	0.701	1507
Monthly	48.7	0	51.3	0.560	52.5	0	47.6	0.326	347
Quarterly	48.7	0	51.3	0.781	53.9	0	46.1	0.395	115
Annual	43.9	0	57.1	0.451	46.4	0	53.6	0.706	28

Notes: see also notes to Table 7. Percentages may not sum to 100% across rows due to rounding. Column entitled 'Sign symmetry p-value' reports two-sided p-values for the null hypothesis that 50% of signs are strictly positive. P-values are obtained by normal bootstrap with 999 resamples over non-overlapping blocks spanning 2 calendar months (rounded up at non-monthly frequencies) (see Section 0). We also verified that 50% sign share was encompassed by 90% confidence intervals based on the bias-corrected and accelerated bootstrap of Efron (1987) (not reported for the sake of space).

6.3.2 Time-series properties

Both U and σ are strongly and significantly autocorrelated at a lag of one month, both in the full sample (see Table 9) and in most rolling windows (see Figure 18), though autocorrelation of U is more stable over time. U is more persistent than σ.

Table 9: Monthly autocorrelations, 1984–2012

	u		σ	
ρ_1	0.83	***	0.66	***
ρ_2	0.76	***	0.48	***
ρ_3	0.70	***	0.38	***
ρ_4	0.64	***	0.31	***
ρ_6	0.59	***	0.30	***
ρ_9	0.53	***	0.23	**
ρ_{12}	0.46	***	0.16	

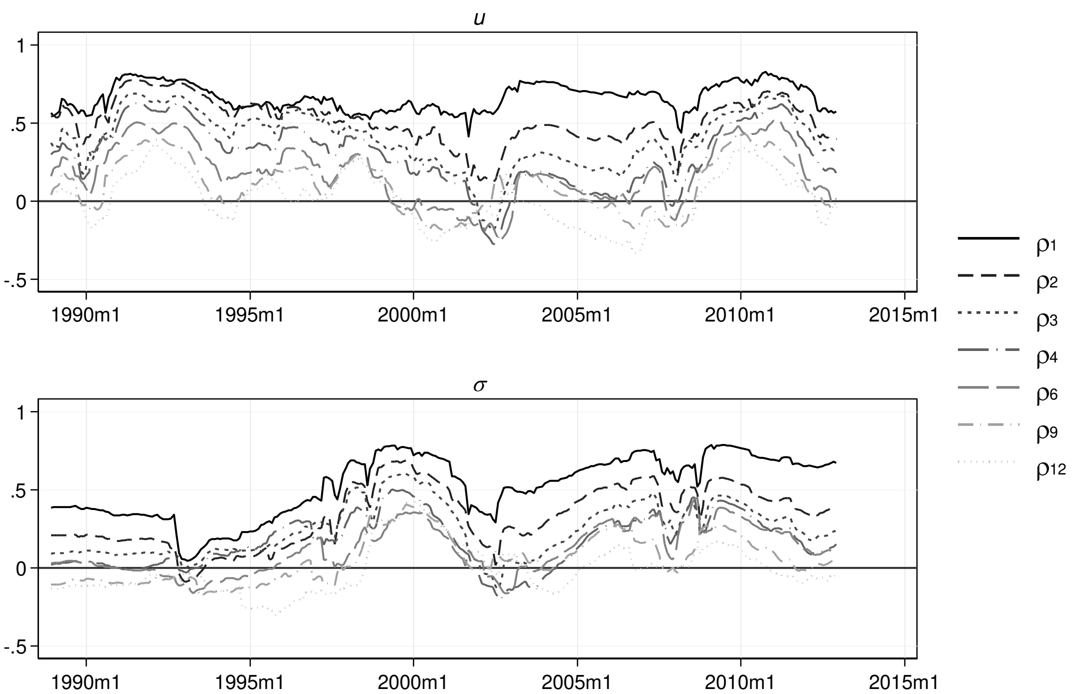
Notes: ρ_l is the sample autocorrelation coefficient at lag l. *, **, *** indicate significance at 10%, 5% and 1% levels respectively. Critical values are approximate, based on the assumption that the time series are Gaussian. Under the null that autocorrelation ρ_l equals zero (in which case the bootstrap is not required to account for serial dependence) ρ_l is normally distributed around zero with standard error calculated using Bartlett's formula for MA(l) processes.

How should we understand the greater persistence of U compared to σ? Suppose we can interpret the *impulses* in both U and σ as reflecting *impulses* in latent uncertainty. Then at least one of U and σ must have different persistence than latent uncertainty, and this would blur that measures accuracy as a proxy for short-run movements.

One can imagine σ settling *before* latent uncertainty if there is persistent uncertainty but the flow of new information has limited impact on the forecast distribution of returns, so that classically rational investors have no reason to modify their positions after adjustments to the initial uncertainty impulse have been made. The news media may still report the information along with uncertainty references, so that u does not settle prematurely. Alternatively, U might settle *after* latent uncertainty if news articles make retrospective references to past uncertainty. Distinguishing between these possibilities in future research might be achieved by, for example, classifying the textual uncertainty references into current vs. retrospective.

Higher order correlation is varies more over time. The half-life of U (the lag at which autocorrelation drops below 0.5) ranges from around one to six months. The half-life of σ is often less than a month though goes up to around four months in the late 1990s.

Figure 18: Autocorrelations of U and σ within a rolling 60-month window



Note: the sharp drop in σ autocorrelation in October/November 1992 is associated with the October 1987 stock market crash passing out of the rolling window.

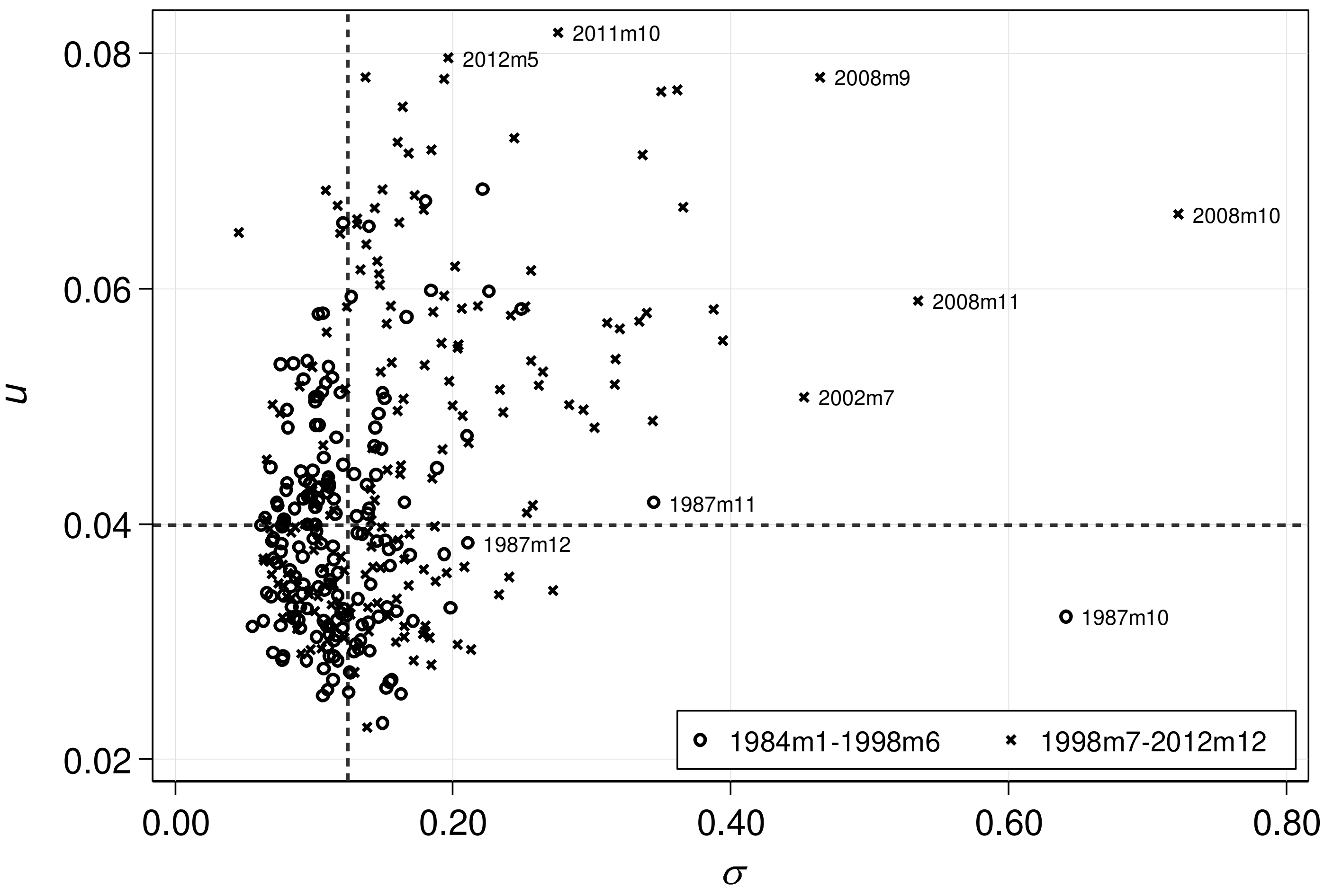
No unit root behaviour is evident from eyeball examination of Figure 15 above. U eventually retreats from the sustained highs of the early 1990s and the recent financial crisis. The recent literature has generally modelled σ as fractionally co-integrated but covariance stationary. DF-GLS tests of Elliott et al. (1996) rejected the null of a unit root on daily data for 1984m1-2012m12 (see Appendix, pg. 81, Table 12). This rejection was robust to subsample analysis either side of the time points where the data was most suggestive of a potential break (e.g. at the onset of the recent financial crisis) and to change of frequency to monthly.

6.3.3 Contemporaneous association

Cross-plots of U vs. σ and ΔU vs. $\Delta\sigma$ are shown in Figure 19 and Figure 20 below.

We do not show best-fit lines between U and σ for two reasons. First, standard regression methods assume that the independent variable is fixed and that the dependent variable is responsible for all variance around the best fit line, whereas no such asymmetry suggests itself in our context. Indeed, the gradient of best fit lines (including robust and non-parametric variants such as the Theil-Sen median slope, and the median regression line) varied widely depending of which of U and σ was cast as the dependent variable. Second, the magnitudes of regression slope parameters on these variables have no direct economic interpretation given the arbitrary nature of the scale on U.

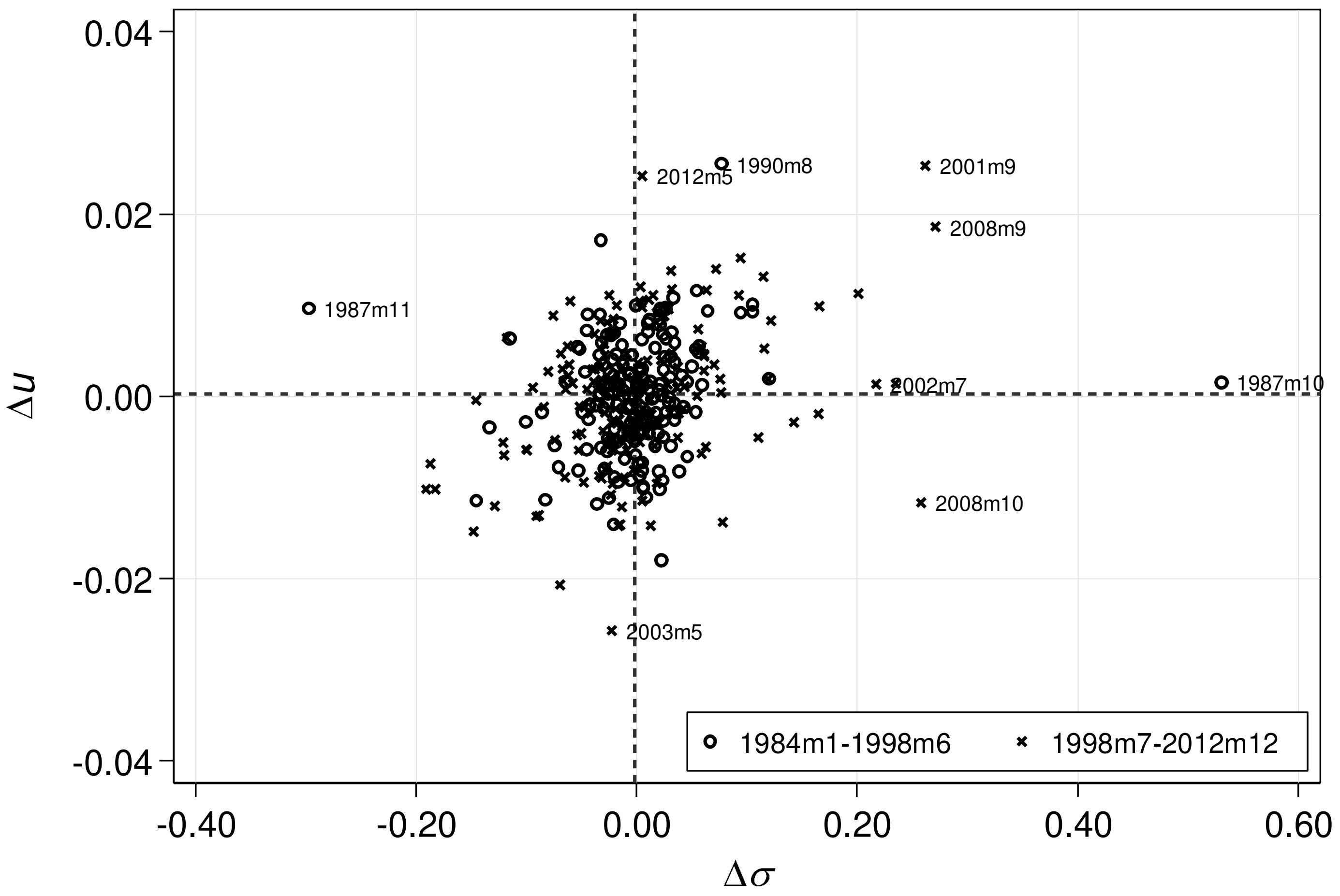
Figure 19: Cross-plot of levels of U and σ, 1984-2012



Note: dashed lines show univariate unconditional medians.

Notes: Observations from the first and second halves of the sample are given different marker symbols, to provide one rough visualisation of temporal change in the bivariate distribution. More of the largest observations fall in the second half of the sample period, and mostly relate to the financial crisis period at the end of the sample, and to the October 1987 crash.

Figure 20: Cross-plot of first differences ΔU and $\Delta\sigma$, 1984-2012



Note: dashed lines show univariate unconditional medians.

Notes: see notes to Figure 19.

6.3.3.1 Methodology

Measures of correlation

We use four measures of correlation/association – one cardinal, three ordinal – which make different trade-offs between efficiency and robustness.

Pearson's product-moment correlation ρ measures the degree of linear association between the variables. However, we only consider U to be an ordinal proxy for U^*, and we lack a clear model of the relationship between σ and U^*, so a priori we only have reason to expect the relationship to be monotonic, not necessarily linear. Also, ρ is sensitive to outliers, as seen, for example, in the abrupt jumps in rolling values of ρ in Figure 21.

Rank correlations are more outlier-resistant and capture the strength of monotonic association. Spearman's rank correlation r_s is a straightforward application of Pearson's ρ on ranks. Kendall's τ_a is calculated by permuting over all pairs of observations. It is the difference between the number of pairs whose ordering is the same according to both variables (concordance) and the number of pairs whose ordering is different (discordance), as a fraction of all pairs. r_s is perhaps more familiar via analogy to ρ, but more reliable methods for obtaining confidence intervals are available for τ_a than for r_s[42], and τ_a is easier to interpret precisely than r_s: τ_a equals the difference in probability of the two covariates 'agreeing' versus 'disagreeing', in the sense that for a randomly selected pair of observations the ordering of the observations (according which is larger) is the same according to both covariates.

Greiner's $r_g \equiv \sin\left(\frac{\pi}{2}\tau_a\right)$ transforms τ_a so as to render it comparable with ρ[43] while retaining the outlier-resistance, and invariance to monotonic transforms of U and σ, enjoyed by τ_a. If there exists a pair of monotonic transformations under which U and σ become bivariate normal (and we know that the log transform is already a reasonable approximation), then r_g is equal to the ρ that would be obtained on the transformed covariates, but without us having to know what those transformations are.

Sign concordance measures are still more robust to outliers, and to possible disturbances from non-uncertainty components of the measures, albeit at an efficiency cost. In the present context these are only useful for differences. For simplicity of interpretation we focus on the fraction of observations, $S \in [0,1]$, that have $sign(\Delta U) = sign(\Delta\sigma)$ where $sign(\cdot) \equiv \begin{cases} 1\ if & \cdot > 0 \\ 0\ if & \cdot \leq 0 \end{cases}$.

Bootstrapped p-values and confidence intervals

HAC-robust variance estimators are not generally available for correlation statistics, and the 'classical' approximations for p-values and confidence intervals (CIs) assume independent observations and, in some cases, normally distributed data, which is clearly not congruent with the

[42] This will matter more when estimating confidence intervals/p-values for the differences between degree of association in different subsets of the data, as is done below.

[43] Originally the one-to-one mapping between r_g and τ_a was derived under the assumption that the correlates are bivariate normal. However, as Newson (2002) highlights, it is not affected by odd-numbered moments (such as skewness) and is expected to hold approximately for a wide range of continuous bivariate distributions.

substantial and significant autocorrelation (see Table 9 and Figure 18) and non-normality (see Table 7) in our time-series data.

We therefore use the block bootstrap of Kunsch (1989), which resamples over non-overlapping blocks thus allowing for serial dependence within blocks. Bootstrap p-values are derived from bootstrap standard errors under the approximation that the sampling distribution is normal. Bootstrap CIs are the bias-corrected and accelerated CIs of Efron (1987) which do not assume a normal sampling distribution. All bootstrap results are based on 999 replications[44] and implemented using the -bootstrap- command in Stata.

Block length is important. For analysis in levels, where rolling autocorrelations are mostly damped to insignificance by 6 months, but sometimes persist longer, we use blocks spanning 6 months and 12 months (as a sensitivity check). This allows the majority of the serial dependence patterns to be captured within blocks (in light of the autocorrelation u and σ reported below) while also providing for a reasonable number of resampling units (key to an effective bootstrap). For first differences where autocorrelation is weaker, we use blocks spanning 2 month and 6 months.

Since the bootstrap is only asymptotically correct, we also show ‘classical’ p-values and CIs where they are available and space permits.

6.3.3.2 Full sample correlations

Correlation measures for the full sample are reported in Table 10, and sign concordances are reported in Table 11. In short, U and σ are strongly positively correlated.

Correlation in levels is positive, substantial (ρ on the order of 0.3-0.4), comparably sized at weekly through annual frequencies (daily results are discussed separately below), and statistically significant at the 99% level except for annual data where the confidence interval is wide due to the small number of observations. This is qualitatively similar to Baker et al.’s (2013) finding of a Pearson correlation of 0.578 between their economic policy uncertainty measure and the VIX for 1990-2012. Rank correlation coefficients are only slightly smaller than Pearson’s ρ, indicating that ρ is not just driven by outliers. The CI of [0.087,0.304] on Kendall’s τ_a for monthly data tells us that, with 95% confidence, for a randomly selected pair of months, U and σ are between 8.7% and 30.4% more likely to agree than to disagree in their ordering of the months.

Correlation in first differences is similarly positive, substantial, significant, robust, and comparably sized at monthly through annual frequencies (weekly results are discussed separately below). For a randomly selected pair of months, U and σ are between 9.7% and 24.6% more likely to agree than to disagree in their ordering. The sign concordances of first differences show a similar pattern across frequencies (first column of Table 11). U and σ move in the same direction in 57.1% of months, and this is significantly different, at the 99% level, to the ‘coin-toss’ benchmark of 50%. The probability of sign concordance does not vary strongly or significantly with the direction of movement (last four columns of Table 11).

[44] This choice, rather than a round number such as 1000, avoids the need for interpolation in estimating results for conventional confidence levels.

The much weaker correlation at daily frequency, and in first differences at weekly frequency, is at least partly due to noisier estimates of σ at higher frequency, but *a priori* could also be partly due to dynamics causing interdependence to operate at lags and leads at higher frequency and thus not be reflected into contemporaneous correlations.

We focus on monthly data for the deeper analysis of contemporaneous correlations in the remainder of this Section, and consider the higher frequency dynamics in Section 6.3.4.

Table 10: Contemporaneous cross-correlations in levels and first differences, 1984–2012

		N	Pearson's ρ		Spearman's r_s		Greiner's r_g		Kendall's τ_a	
Levels	daily	7326	0.126*** †††	*[0.103,0.148]* [0.062,0.204] ^{6} [0.064,0.194]^{12}	0.089*** †††	*[0.066,0.112]* [0.041,0.151] ^{6} [0.040,0.156]^{12}	0.093*** †††	*[0.069,0.117]* [0.042,0.158] ^{6} [0.042,0.163]^{12}	0.059*** †††	*[0.044,0.075]* [0.027,0.101] ^{6} [0.027,0.104]^{12}
	weekly	1508	0.305*** †††	*[0.258,0.350]* [0.167,0.419] ^{6} [0.163,0.442]^{12}	0.237*** †††	*[0.188,0.283]* [0.123,0.344] ^{6} [0.116,0.372]^{12}	0.245*** †††	*[0.194,0.294]* [0.129,0.356] ^{6} [0.120,0.382]^{12}	0.157*** †††	*[0.124,0.190]* [0.082,0.232] ^{6} [0.076,0.250]^{12}
	monthly	348	0.393*** †††	*[0.298,0.477]* [0.202,0.517] ^{6} [0.188,0.544]^{12}	0.314*** †††	*[0.215,0.404]* [0.135,0.458] ^{6} [0.105,0.499]^{12}	0.313*** †††	*[0.216,0.405]* [0.136,0.459] ^{6} [0.103,0.497]^{12}	0.203*** †††	*[0.139,0.266]* [0.087,0.304] ^{6} [0.066,0.331]^{12}
	quarterly	116	0.396*** †††	*[0.223,0.534]* [0.192,0.544] ^{6} [0.161,0.565]^{12}	0.364*** †††	*[0.189,0.508]* [0.153,0.527] ^{6} [0.135,0.556]^{12}	0.365*** †††	*[0.208,0.504]* [0.167,0.526] ^{6} [0.136,0.555]^{12}	0.238*** †††	*[0.134,0.337]* [0.107,0.353] ^{6} [0.087,0.374]^{12}
	annual	29	0.302 †	*[-0.094,0.588]* [-0.094,0.568]^{12}	0.249	*[-0.146.0.552]* [-0.159,0.552]^{12}	0.253	*[-0.128,0.568]* [-0.166,0.561]^{12}	0.163	*[-0.080,0.387]* [-0.106,0.377]^{12}
First differences	daily	7325	-0.025**	*[-0.047,-0.002]* [-0.056,0.008]^{2} [-0.055,0.014]^{6}	-0.020*	*[-0.043,0.003]* [-0.046,0.010]^{2} [-0.048,0.011]^{6}	-0.021*	*[-0.046,0.003]* [-0.049,0.010]^{2} [-0.050,0.012]^{6}	-0.014*	*[-0.029,0.002]* [-0.031,0.006]^{2} [-0.032,0.008]^{6}
	weekly	1507	0.049*	*[-0.001,0.099]* [-0.018,0.131]^{2} [-0.023,0.126]^{6}	0.028	*[-0.022,0.078]* [-0.044,0.087]^{2} [-0.044,0.089]^{6}	0.030	*[-0.026,0.086]* [-0.047,0.093]^{2} [-0.047,0.096]^{6}	0.019	*[-0.017,0.055]* [-0.030,0.059]^{2} [-0.030,0.061]^{6}
	monthly	347	0.272*** †††	*[0.171,0.366]* [0.159,0.394]^{2} [0.119,0.408]^{6}	0.243*** †††	*[0.140,0.338]* [0.144,0.356]^{2} [0.113,0.378]^{6}	0.256*** †††	*[0.142,0.365]* [0.151,0.377]^{2} [0.118,0.401]^{6}	0.165*** †††	*[0.091,0.238]* [0.097,0.246]^{2} [0.076,0.263]^{6}
	quarterly	115	0.414*** †††	*[0.243,0.549]* [0.239,0.555]^{3} [0.297,0.518]^{6}	0.418*** †††	*[0.248,0.552]* [0.237,0.570]^{3} [0.258,0.558]^{6}	0.447*** †††	*[0.254,0.605]* [0.257,0.603]^{3} [0.277,0.584]^{6}	0.295*** †††	*[0.164,0.415]* [0.164,0.412]^{3} [0.178,0.397]^{6}
	annual	28	0.283	*[-0.113,0.575]* [-0.082,0.680]^{12}	0.220	*[-0.175.0.531]* [-0.150,0.585]^{12}	0.245	*[-0.198,0.605]* [-0.143,0.638]^{12}	0.158	*[-0.126,0.417]* [-0.091,0.441]^{12}

Notes: N is the number of observations. *, **, *** indicate significance at 10%, 5% and 1% levels respectively, based on approximate two-sided 'classical' p-values (see below). Similarly †, ††, ††† indicate significance but based on normal block bootstrap two-sided p-values (with block size the smaller of those used in constructing CIs) using the methods described in Section 0. Brackets [] contain 95% confidence intervals (CIs). 'Classical' CIs are *italicised*. Block bootstrap CIs are non-italicised and the block length in months is noted in a superscript suffix. Shorter blocks are used for first differences than for levels because autocorrelation is shorter-lived in first differences. 'Classical' p-values and CIs for ρ and r_s are obtained using Fisher's approximation, which assumes that observations are independent and bivariate normally distributed and the sample size is large. Under these assumptions, the sampling distributions of the hyperbolic arctangents of ρ and r_s (Fisher's z-transform) are asymptotically normal with mean $\rho + 2\rho/(N-1)$ or $r_s + 2r_s/(N-1)$ respectively (we subtract the second term in each expression as a bias adjustment) and variance approximately equal to $1/(N-3)$. 'Classical' p-values and CIs for τ_a and $r_g \equiv \sin(\pi\tau_a/2)$ are obtained by the method of Newson (2005), using the jackknife and Taylor polynomial approximations, after applying Fisher's z transform to stabilise variances as recommended in Edwardes (1995).

Table 11: Contemporaneous sign concordance in first differences, 1984–2012

	N	All		$\Delta U \leq 0$		$\Delta U > 0$		$\Delta\sigma \leq 0$		$\Delta\sigma > 0$	
daily	7325	0.498	[0.485,0.512]^{2} [0.485,0.511]^{6}	0.498	[0.485,0.512]^{2} [0.485,0.511]^{6}	0.498	[0.483,0.514]^{2} [0.483,0.513]^{6}	0.503	[0.486,0.517]^{2} [0.489,0.517]^{6}	0.493	[0.479,0.508]^{2} [0.479,0.507]^{6}
weekly	1507	0.518	[0.489,0.546]^{2} [0.486,0.544]^{6}	0.515	[0.481,0.546]^{2} [0.476,0.542]^{6}	0.520	[0.488,0.552]^{2} [0.491,0.557]^{6}	0.515	[0.480,0.547]^{2} [0.481,0.546]^{6}	0.520	[0.485,0.553]^{2} [0.487,0.553]^{6}
monthly	347	0.571†††	[0.523,0.624]^{2} [0.513,0.631]^{6}	0.598 †††	[0.520,0.671]^{2} [0.524,0.667]^{6}	0.545	[0.476,0.612]^{2} [0.472,0.616]^{6}	0.555 †	[0.483,0.621]^{2} [0.481,0.619]^{6}	0.58† 8†	[0.512,0.653]^{2} [0.519,0.656]^{6}
quarterly	115	0.635†††	[0.530,0.722]^{3} [0.526,0.713]^{6}	0.679 †††	[0.553,0.789]^{3} [0.536,0.796]^{6}	0.593	[0.462,0.719]^{3} [0.474,0.704]^{6}	0.613 †	[0.434,0.726]^{3} [0.471,0.721]^{6}	0.66† 0†	[0.509,0.776]^{3} [0.526,0.780]^{6}
annual	28	0.607	[0.393,0.750]^{12}	0.583	[0.308,0.875]^{12}	0.625	[0.375,0.857]^{12}	0.538	[0.231,0.778]^{12}	0.667	[0.353,0.857]^{12}

Notes: N is the number of observations at the given frequency. In each of the remaining columns the left-most number is the fraction S of observations for which $sign(\Delta U) = sign(\Delta\sigma)$, with $sign(\cdot)$ as defined in the main text. †, ††, ††† indicate significant difference from 0.5 at 10%, 5% and 1% levels respectively, based on block bootstrap p-values (described in Section 0) for the shorter of the block lengths used for the CIs. Brackets [] contain block bootstrap 95% confidence intervals (CIs) and the block length in months is noted in a superscript suffix. CIs are shown for different block lengths to allow assessment of their sensitivity (or lack thereof) to block length.

6.3.3.3 Temporal variation in correlations

The full sample correlations conceal substantial temporal variation, as clearly illustrated in Figure 21 using rolling 60-month windows[45]. Occasional large jumps in Pearson's ρ betray its sensitivity to outliers, but the more robust rank correlation measures exhibit similar temporal profiles to ρ (albeit with less abrupt changes), as do correlations in first differences and sign concordance of first differences.

The correlation between U and σ drops off sharply from early 2010, falling to nearly zero by the end of the sample at end 2012. By contrast, the correlation and sign concordance between ΔU and $\Delta\sigma$ rose through 2010 and remained high to the end of the sample. The timing coincides with the start of the Eurozone crisis, conventionally dated to the emergence of the Greek government debt crisis in late 2009/early 2010.

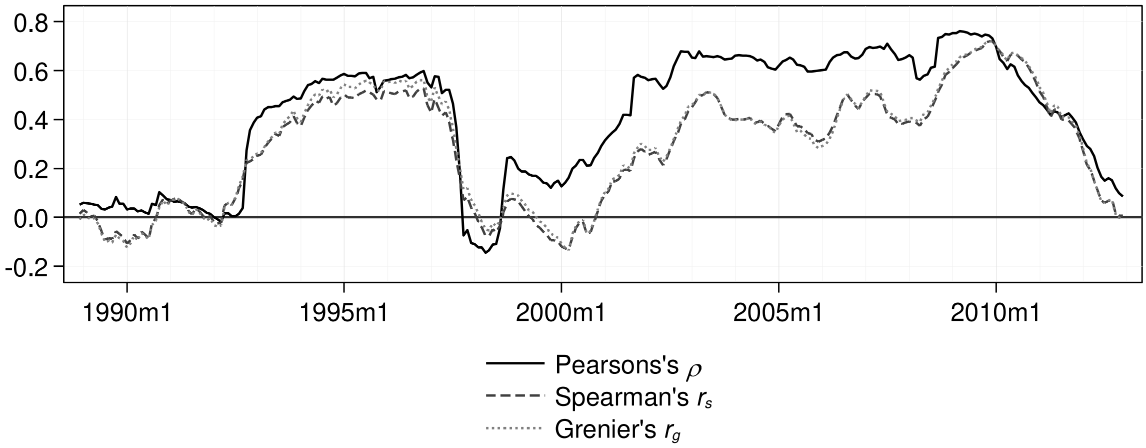
One potential explanation might be that the extraordinary interventions by fiscal and monetary authorities during the Eurozone crisis artificially supressed the level of σ (even while movements around the supressed base level continued to reflect movements in uncertainty), whereas the news-media continued to express the strongly elevated level of uncertainty, causing a disconnect in levels between U and σ. Of course this is speculative, and does not account for the fact that the decline in σ began earlier in 2009, nor why earlier interventions such as TARP in the US, and the £200billion of asset purchases undertaken by the UK authorities between March 2009 and January 2010, were not associated with a similar decline in correlation.

The correlation between U and σ also exhibits switching behaviour between sustained periods of very low correlation and sustained periods of very high correlation. The consistency of this pattern across rank and sign concordance measures demonstrates that this is not an artefact of outliers moving in and out of the rolling window. The sustained nature of the movements, even allowing for the smoothing inherent in a rolling analysis, suggests some underlying structure.

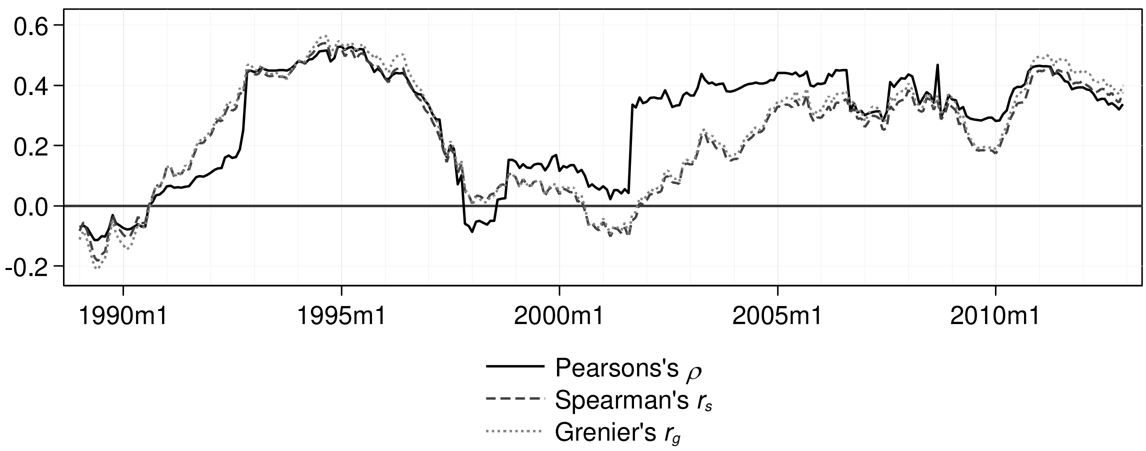
[45] This is the same window lengths used by Campbell et al. (2001) in their investigation of the relationship between disaggregated components of volatility. Five-year rolling correlations on weekly and quarterly data show similar time profiles, but the rise in correlation of levels around the onset of the financial crisis and the subsequent drop is less pronounced (relative to the early-mid 1990s hump) at lower frequency.

Figure 21: Contemporaneous cross-correlations within rolling 60-month windows

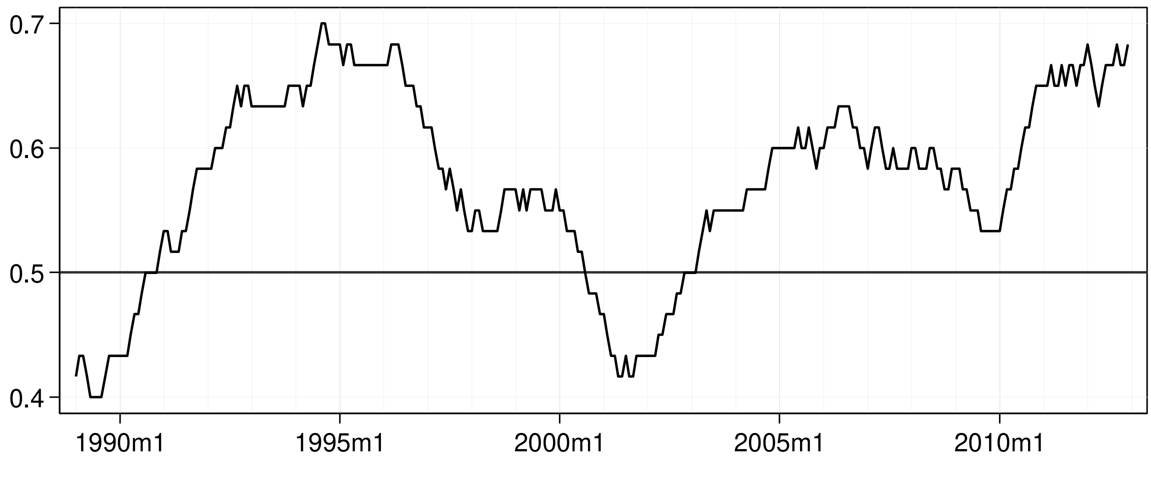
Levels, U and σ



First differences, Δu and $\Delta\sigma$



Fraction of observations, S, with sign agreement between ΔU and $\Delta\sigma$



Notes:

1. date axes indicate end date of the five-year rolling window.
2. τ_a and r_g are related to one another by a time-invariant monotonic transform, so there is no incremental information from displaying both. We display r_g because it is more intuitively comparable with ρ and r_s.

What might cause this switching behaviour? We offer two hypotheses, which are not mutually exclusive.

First, journalists may have a lower propensity to use the word “uncertainty” to express upside risk than to express equally sized downside risk. U would then tend to underweight upside risk whereas σ would weight upside and downside risk equally. Correlation between U and σ may then be weaker in periods when upside risk is higher. Since perceived upside risk is probably higher during booms, we would expect lower correlation during booms. This is broadly consistent with what we observe: the periods of lower correlation broadly correspond respectively to the late 1980s macroeconomic upswing known as the ‘Lawson boom’ (after the then UK Chancellor) and to the dot-com boom of the late 1990s.

A few recent papers that attempt decompose uncertainty into upside and downside components, provide some corroborating evidence for our hypothesis, albeit with different empirical uncertainty measures than used here. Rossi & Sekhposyan (2015) identify uncertainty with a measure of the size of realised error[46] on GDP forecasts, which the authors interpret as a measure of unpredictability, which they further assert is associated with uncertainty. Two of the three extended periods of upside uncertainty shown in their results for our sample period coincide broadly with the periods of low correlation identified above. The third, around 1992-1993 does not. (See their Figure 2, second and fourth panels.) That said, they use US rather than UK data. Segal, Shaliastovich, & Yaron (2015) identify ‘good’ (‘bad’) or upside (downside) uncertainty with positive (negative) realised semi-volatilities of the US industrial production growth rate, using the estimator of Barndorff-Nielsen, Kinnebrock, & Shephard (2008). They also show upside uncertainty being elevated in the mid to late 1990s, at to a lesser extent in the late 1980s, relative to the rest of our sample period (see their Figure 2). Feunou, Jahan-Par, & Tedongap (2010) show similar timings for elevated S&P500 upside volatility, estimated from a binormal-GARCH model (see their Figure 2, Panels E and F).

Future research could test our hypothesis more rigorously by applying the present correlation framework to the semi-volatilities of aggregate stock index returns, and through manual semantic analysis of occurrences of “uncertain*” in FT articles published in the late 1990s about dot-com stocks[47].

Second, risk aversion (distinct from the perceived degree of risk itself) has been estimated to account for perhaps one quarter of stock returns volatility (Bekaert, Engstrom, & Xing, 2009). Temporal variation in risk aversion might therefore contribute to temporal variation in the correlation between U and σ. Future research could use the framework of Bekaert, Engstrom, & Xing (2009) to decompose σ into risk aversion and risk components, and examine how these relate to U.

[46] Specifically, the difference between the 0.5 (representing the median) and the quantile of the realised forecast error with respect to its historic distribution

[47] An outline of such a research program might be: i) manually tag statements that express or imply uncertainty about the future; ii) manually classify these by direction of risk (upside/downside/symmetric) and index the words or phrases used; iii) examine correlations between direction of risk and particular words or phrases.

6.3.3.4 Further structure in the correlation

Finally, we document how the correlations between the movements of U and σ, and their signs and magnitudes, vary with the level of U and σ, and with the magnitude and direction of their movements. The aim is to establish basic empirical facts against which future theories of the relationship between U and σ can be developed and tested.

We segment observations into a series of two-by-two grids[48] and estimate correlation and its significance within each cell and in the corresponding margins, along with the p-value for the difference between each pair of cells or margins. Figure 22 and Figure 24 segment the sample by whether the lagged levels $L.U$ and $L.\sigma$ are above (or at) or below their medians (referred to as 'high' and 'low' for ease of exposition). We use lags because contemporaneous levels are by construction correlated with the first differences, the variation in whose correlation we are trying to measure. Figure 23 segments by whether the magnitude of first differences $|\Delta U|$ and $|\Delta\sigma|$ are above (or at) or below their medians. Figure 25 segments the sample by the sign of first differences.

Correlation in movements and concordance in their direction are substantial and significant when $L.U$ is high (bottom rows of Figure 22 and Figure 24) but weak and insignificant when the level $L.U$ is low (top rows). This true whether $L.\sigma$ is high or low, and thus also in the margin[49]. By contrast, these measures of co-movement do *not* vary strongly or significantly when depending on whether $L.\sigma$ is high versus low.

The pattern is similar for concordance in the direction of movements when segmenting by the magnitude of movements $|\Delta U|$ and $|\Delta\sigma|$. There is significantly stronger concordance when there is a large movement U compared to essentially no concordance when the movement in U is small. By contrast, the magnitude of movement in σ does not make a significant difference.

The lack of concordance when *both* U and σ are low might be partly due to a lower signal-to-noise ratio, but the asymmetry in dependence of correlation/concordance on U vs. σ demands a more structural explanation.

The magnitude of movements in U and σ is fairly well correlated when they are moving in the same direction (diagonal cells in Figure 25) – more strongly so when they are both falling rather than rising, though the difference is not significant. Unsurprisingly, correlation in magnitude of movements is weaker when the measures move in opposite directions.

[48] Experiments with more granular segmentation left too few observations per cell to conduct useful inference.
[49] In Figure 24, the value of $\rho = 0.278$ when $L.U$ is low and $L.\sigma$ is high (and thus also the marginal $\rho = 0.144$ for $L.U$ low) is driven by outliers. The more robust rank correlation measures show much smaller values here.

Figure 22: Segmentation of sign concordance S, by lagged levels of U and σ

L.u \ ***L.σ***	< median		≥ median	all
< median	102 0.539 0.220	← 0.906	71 0.437 0.861	173 0.497 0.533
	↑ 0.203	↖ †† 0.024	↑ ††† 0.000	↑ ††† 0.003
≥ median	71 0.606 †† 0.035	← 0.199	103 0.670 ††† 0.000	174 0.644 ††† 0.000
all	173 0.566 † 0.062	← 0.442	174 0.575 †† 0.024	347 0.571 ††† 0.008

Quick key (details in notes)

N_1 S_1 Sig. level P-value	Test of $S_1 \leq 0.5$ (one-sided)
↑ Sig. level P-value	Test of $S_2 < S_1$ (one-sided)
N_2 S_2 Sig. level P-value	Test of $S_2 \leq 0.5$ (one-sided)

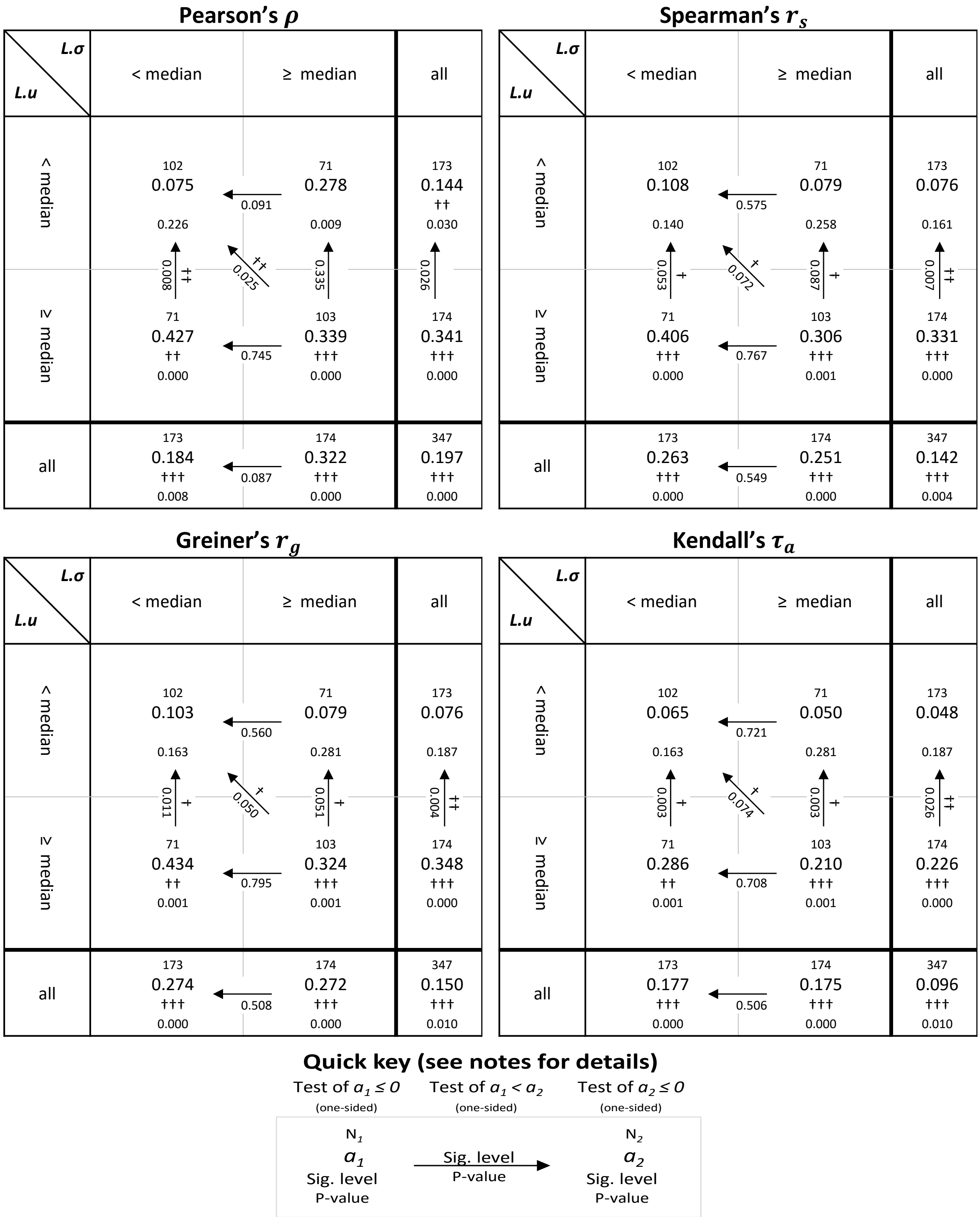
Figure 23: Segmentation of sign concordance S, by size of movements $|\Delta U|$ and $|\Delta\sigma|$

|Δu| \ ***|Δσ|***	< median		≥ median	all
< median	103 0.495 0.543	← 0.470	70 0.500 0.500	173 0.497 0.533
	↑ † 0.101	↖ ††† 0.005	↑ ††† 0.007	↑ ††† 0.005
≥ median	71 0.600 † 0.063	← 0.174	103 0.681 ††† 0.000	174 0.644 ††† 0.001
all	173 0.543 0.154	← 0.185	174 0.598 ††† 0.008	347 0.571 ††† 0.008

Notes: <u>Within cells and margins:</u> Reported p-values are one-sided for the null that the fraction of observations, S, exhibiting sign agreement is less than 0.5. These p-values are derived from the normal block bootstrap (see Section 0) using 6-month blocks. †, ††, ††† indicate significant rejection of the same null at 10%, 5% and 1% levels respectively, but based on whether 0.5 lies beyond the bias-corrected and accelerated block bootstrap confidence limit.

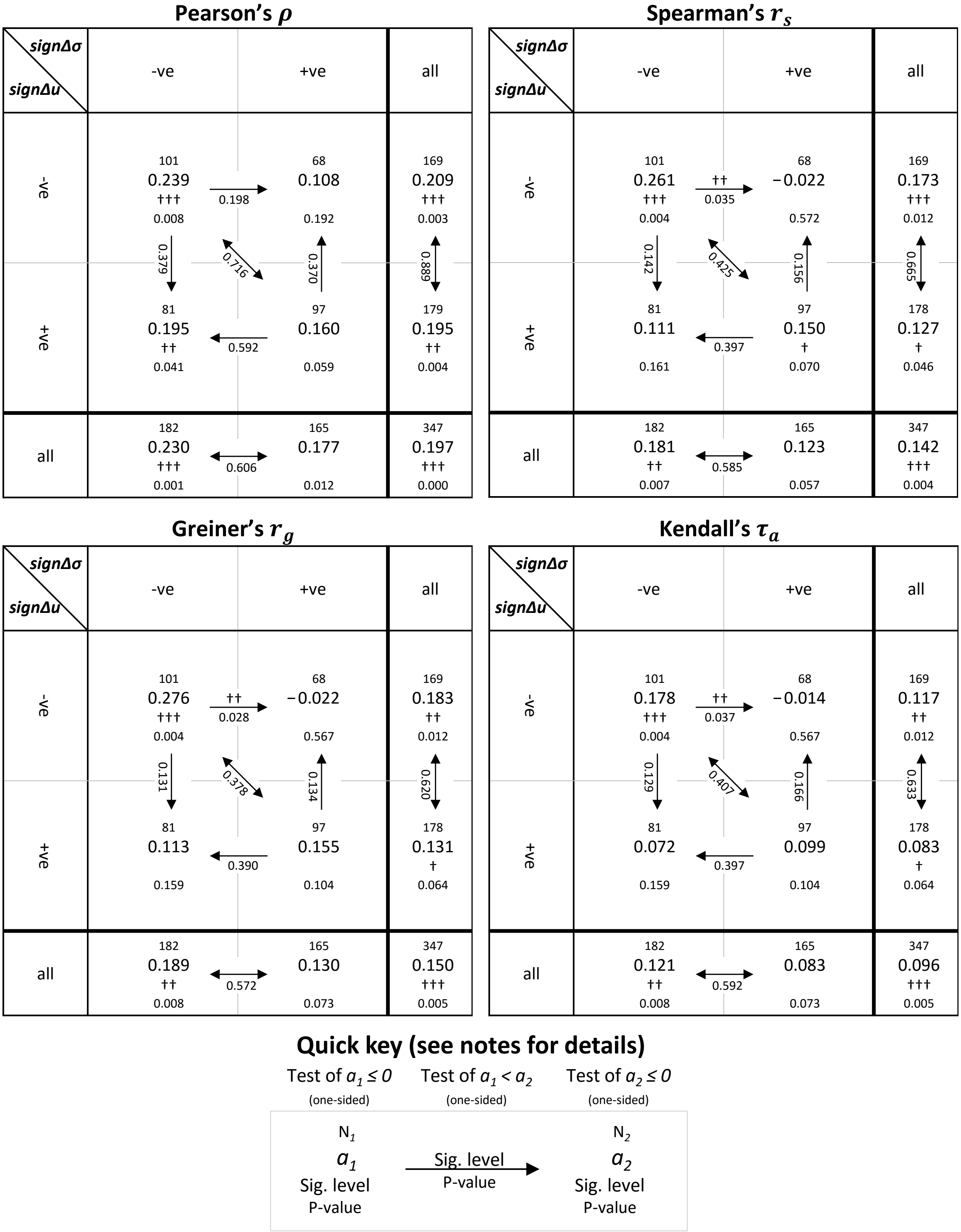
<u>Comparisons between cells and between margins (indicated by arrows):</u> Reported p-values are one-sided for the null that S, in the cell (or margin) at the arrow's tail, is less than in the cell (or margin) at the arrow's head. The choice of one-sided p-values reflects our prior expectation that S will be higher when the segmenting variables are larger. †, ††, ††† indicate significant rejection of the same null, at 10%, 5% and 1% levels respectively, but based on whether zero is contained within one-sided block bootstrap bias-corrected and accelerated confidence limits.

Figure 24: Correlation of ΔU and $\Delta\sigma$, segmented by level of U and σ



Notes: Within cells and margins: Reported p-values are one-sided for the null that the association measure is less than zero. These p-values are derived 'classical' method s described in notes to Table 10, assuming independence of observations. †, ††, ††† indicate significant rejection of the same null at 10%, 5% and 1% levels respectively, but based on whether zero is lies beyond the bias-corrected and accelerated block bootstrap confidence limit, which allows for serial dependence within 6-month blocks (see Section 0).
Comparisons between cells and between margins (indicated by arrows): Reported p-values are one-sided for the null that the measure of association, in the cell (or margin) at the arrow's tail, is greater than in the cell (or margin) at the arrow's head. The choice of one-sided p-values reflects our prior expectation that association will be higher when the segmenting variables are larger. †, ††, ††† indicate significant rejection of the same null at 10%, 5% and 1% levels respectively, but based on whether zero is contained within one-sided block bootstrap bias-corrected and accelerated confidence limits.
Outliers are behind the $\rho = 0.278$ value for $L.U <$ median and $L.\sigma \ge$ median, but do not so strongly affect the other more outlier-resistant correlation measures.

Figure 25: Correlation of $|\Delta U|$and $|\Delta\sigma|$, segmented by sign of movements $sign(\Delta U)$ and $sign(\Delta\sigma)$



Notes: see notes to Figure 24, with the exception that p-values and CIs for comparisons on the diagonal and in the margins are two-sided (indicated by a bidirectional arrow) because we lack a clear prior on the sign of the differences.

6.3.4 Granger causality tests

The correlated components of U and σ may be plausibly interpreted as components of latent uncertainty that are reflected in both measures. In this Section we ask whether this information is incorporated more quickly into one measure than into the other, or equivalently whether one or both of the measures can help to forecast the other.

To investigate this we test for Granger causality in a bivariate vector autoregression (VAR), in a similar spirit to Campbell et al. (2001) (albeit they compare aggregate stock volatility with industry- and firm-level volatility, rather than with news-media uncertainty). *A priori* we expect any Granger causation to be visible only at high frequency, given the strong incentives for rapid incorporation of information in both the news media and the stock market, so we conduct analysis at daily frequency.

Using realized volatility estimates of σ based on intra-day tick data should keep noise in σ manageable. In Section 3.3 we estimated that the noise in U is approaching the same scale as the signal U^* at daily frequency, so errors-in-variables is a non-trivial concern. However, we draw some comfort from the fact that from our large number of observations we are able to identify a simple pattern of Granger-causation that can be given a straightforward explanation, which suggests that that the regressions are not overwhelmed by attenuation bias for example.

6.3.4.1 Baseline specification

To ensure a sample without gaps, and to simplify analysis, the calendar of 'days' is constructed to include only the 3,247 days that are both London Stock Exchange trading days and FT publication days, so that the 'day' after Friday is Monday (or Tuesday or Wednesday on Bank Holiday weekends) and we discard Saturday editions of the FT. Note that the distribution of U is little affected by whether Saturdays are included or not (see Table 7 on pg. 53) and the information loss is likely modest because if latent uncertainty is very high on a non-trading day then the strong autocorrelation documented above means it is also more likely to be high on the next trading day and thus expressed in U.

We apply skew-reducing transforms $\sqrt{U}$ and $\ln(\sigma)$ to render the variables approximately normal before entering them in the VAR. This mitigates the misspecification that would otherwise arise from the right skew in $U$and σ feeding through to the VAR residuals (which account for around 60% of the variance of U and 20% for σ).

We augmented the VAR with a full set of day-of-week dummies to control for weekly seasonality in U which, though an order of magnitude smaller than the interquartile range of the measure, was not trivial[50]. To these we added a small number of individual day dummies[51] for the largest and visually obvious outliers to mitigate possible distortion of inference. Estimates were very similar, and our conclusions unchanged, if these dummies were omitted.

Maximum lag length was set to 11 based on the Hannah-Quinn information criterion.

No major misspecifications are suggested by standard residual diagnostic plots (see Figure 26, Appendix D). No serious parameter instability is exhibited by the rolling parameter estimates in Figure 27. As anticipated by Granger (1998), some of our formal misspecification tests reject the null

[50] This is seasonality that remains even after scaling by economic news volume as discussed in Section 5.4.
[51] 17/04/2000, 17/07/2000, 11/09/2000, 22/05/2001, 11/09/2001, 26/06/2002, 20/09/2002, 07/07/2005, 07/11/2007, 08/09/2008, 24/08/2011, 17/01/2012, 28/12/2012

because we have orders of magnitude more observations and thus greater power than in the typical quarterly macroeconomic VAR. However, the *size* of misspecifications is modest. Residual serial correlation coefficients at the first lag are −0.008 and −0.006 for the U and σ equations respectively, and are similarly small at longer lags (see residual diagnostic plots). Residual skewness is −0.077 and 0.321, comparable to levels deemed acceptable in the leading reference text on VARs by Juselius (2006). Residual kurtosis is 3.24 and 3.91, not too far from the normal value of 3.

6.3.4.2 Results and interpretation

We find no evidence of Granger causation from $U \rightarrow \sigma$: the p-value for a Wald test of the joint exclusion restriction on all lags of U in the σ equation is 0.487.

There is however statistically significant evidence of Granger causation from $\sigma \rightarrow U$: the p-value for a Wald test of the joint exclusion restriction on all lags of σ in the U equation is 0.010. This is primarily attributable to the first daily lag of σ in the U equation, for which the parameter is around twice as large as subsequent lags and has a p-value of 0.006. All subsequent lags are individually and jointly insignificant at the 95% level (joint p-value 0.202).

This tidy lag pattern is consistent with the intra-day timing of events: the FT is published in the morning before the markets open (recall our dataset corresponds to the print edition), thus uncertainty events occurring later in the day cannot be reflected in U until the following day, but such events occurring before market close may be reflected in σ that same day.

The joint dynamics are heavily dominated by autoregression. The statistically significant cross-equation effect identified above is very small in magnitude: the estimated response of U to a one standard deviation impulse in σ never exceeds 0.006 standard deviations of U. That said, we cannot rule out the possibility of attenuation in parameter estimates due to measurement error in U.

It is conceivable that a richer econometric specification might capture some incremental forecasting power of U for σ, which would be of interest in risk management and quantitative trading strategies. However, the lack of significant Granger causation $U \rightarrow \sigma$ in our linear VAR suggests that any such gains are likely to be modest.

Combined with the earlier evidence of strong correlation between the measures, these Granger causation results are consistent with the hypothesis that U and σ rapidly and completely incorporate the information from some common latent variable, which we suggest has natural interpretation as a composite of latent uncertainty components.

6.3.4.3 Robustness

These results, including the pattern of parameter estimates and conclusions from the Granger causality tests, were robust to alternative maximum lag lengths (including 5 and 22 as selected by Schwarz Information Criterion and the Akaike Information Criterion), stopping the sample before 2007m7 (an early conventional date for the start of the financial crisis) and indeed subsampling in many other ways as seen in the rolling analysis (Figure 27, Appendix D), omission of dummies, and including Saturday's articles in the calculation of U for the following Monday instead of dropping them.

7 Conclusions

We have provided a measurement framework in which the fraction of news-media articles about a given topic that express uncertainty, U can be interpreted as an estimator of the latent propensity to express uncertainty, U^*, which in turn we claim is an ordinal measure of the latent intensity of the cognitive state of uncertainty. With loss of generality, whether an article expresses uncertainty or not, can be modelled as a Bernoulli distribution. If the distributions for all articles within the same measurement period are mutually independent, then the fraction of articles that express uncertainty is an unbiased and consistent estimator of the propensity to express uncertainty, with approximately normal measurement error for large numbers of articles. Combining this model with the empirical within mean and variance of the distribution over this fraction we show that the noise-to-signal ratio is negligible at annual frequency, and only on the order of 5% at monthly frequency. At daily frequency the noise is of the same order of magnitude as the signal, so is still useable (and produces sensible results in our correlational analysis) but prompt consideration of errors-in-variables in the VAR regressions we use for Granger causality testing.

We draw several conclusions from our analysis of key choices in empirical implementation of news-media textual measures of uncertainty:

- Uncertainty keyphrase lists should include "uncertainties" (often neglected in the literature) in addition "uncertain" and "uncertainty", since this boosts signal by 14% without adding noise.
- No other candidate keyphrases would produce a large signal boost without risking the introduction of substantial noise. Adding all seven candidates identified by our semantic analysis, leaving aside "risk*" would increase signal by 54%. "Risk*" occurs in three times as many articles as "uncertain*", but its semantic relationship to the latter is complex, and we lack the data needed to probe this, so we leave it for future work.
- The method used in the literature to isolate *economic* uncertainty – counting only articles that "economic" or "economy" – sacrifices around three quarters of articles about economic topics in a bid to remove articles about other topics. Restricting the corpus to the FT, rather than a wider range of generalist newspapers, avoids the need for such filtering: at least 92% of articles without those keyphrases contain other obviously economy-related keyphrases, and the percentage would grow if we extended our fairly modest keyphrase list. The c.30% reduction in noise-to-signal ratio arising from lower sampling variance probably outweighs the noise from introducing some non-economic articles into the counts.
- Where filtering articles using economic keyphrases is necessary – i.e. when including generalist newspapers in the corpus, as done in most of the literature – using "econ*" rather than just "economic" OR "economy" boosts the signal by c.20%, while retaining similar semantics, and without inducing substantial trend, autocorrelation, or seasonality.
- Where requiring uncertainty and economic keyphrases to occur in close proximity (not just in the same article) in order to count as an expression of economic uncertainty, attention should be paid to how low this drives the article counts. Very small article counts can be a cause of large sampling variance (i.e. noise).
- We analyse the relative effects of different economic keyphrase sets found in the literature, to help inform comparisons between papers using these. We also caution that the emerging practice of requiring uncertainty and economic keyphrases to occur in close proximity (not just in the same article) – which has the laudable aim of more precisely isolating *economic* uncertainty – can result in very low article counts and thus large sampling variance (noise).

- There is substantial and irregular in Factiva's FT records, particularly in earlier years of the sample. This has not been considered in the previous literature that uses Factiva or similar databases. It causes non-trivial distortion to the uncertainty measure. We provide a simple de-duplication algorithm.
- Due attention should be given to the method of scaling uncertainty article counts. Inappropriate scaling, or the use of proxies in place of actual article counts, can induce noise factors with a standard deviation up to half as large as U, as well as discrete jumps in some cases. Inappropriate scaling can also endanger the comparability of the measure between widely spaced time points. For example, the unscaled measure says that uncertainty was higher during the Iraq War than in the aftermath of the Lehman Brother collapse, and lower during the Eurozone crisis than after Lehman. The scaled measure reverses these orderings.

Overall, our empirical analysis of U suggests that it is plausible proxy for aggregate uncertainty. It moves in ways that one would expect of latent uncertainty both in the broad sweep of the last thirty years, including around major narrative events that are conventionally associated with elevated uncertainty. U is also strongly, significantly and robustly correlated with another popular uncertainty proxy, stock returns volatility. The significant Granger causation between them is attributable to intra-day timing (publication of the FT before the markets open), which is consistent with the hypothesis that U and σ efficiently incorporate information from a common underlying uncertainty factor. That said, the implausible decline in the level of σ following the on-set of the recent financial crisis how σ may be susceptible to artificial suppression due to invention by the official sector, in which case U may provide a more reliable guide to the level of uncertainty at times of major financial dislocation.

The switching behaviour between sustained periods of high correlation versus very low correlation, begs explanation. We have hypothesised two potential causes – a bias in the semantics of "uncertain*" towards downside uncertainty, and time-varying risk-aversion – and outlined how these could be tested in future research. We have also documented further structure in the relationship between U and σ that can inform future model building and testing.

These results provide empirical foundations for the emerging literature that uses similar news-media uncertainty measures, and illuminate the path to further developing such measures. From the perspective of the literature on stock volatility, these results also support the thesis that volatility is connected to uncertainty about fundamentals (which are here captured in news-media references).

8 Directions for future research

The literature on news-media textual uncertainty measures is young and advances could be sought in several directions.

The measurement methodology based on simple keyphrase counts has yielded interesting results without the complexity associated with deeper parsing of the article text. We therefore suggest two priorities in this direction. First, probe the reasons for the breakdown in correlation between news-media uncertainty and stock returns volatility during the late 1980s and late 1990s booms. Conduct a deeper semantic analysis of "uncertain*", and its relationship to "risk*" and to semi-volatilities, beginning with dot-com stocks in the late 1990s where perceived risk was probably skewed to the upside; and test if news-media uncertainty remains correlated to downside semi-volatility during these periods. Second, refine the keyphrase counting method to account for the number of keyphrase occurrences *within* the article, their prominence (e.g. headline vs. tail paragraphs), and perhaps the more easily identifiable qualifiers of degree (e.g. “very” appearing immediately before “uncertain”). In the longer term, there is much work to be done drilling down on uncertainty about sub-topics.

The comparative analysis should be extended in two directions. First, more robust characterisation of the switching behaviour apparent in the correlations between U and σ (e.g. using a Markov switching model), and the dependence of the correlation on the level of U (e.g. estimating tail dependence using a Gumbel copula, or in a dynamic setting using a smooth transition VAR). Second, adding other extant uncertainty proxies, such as those in Haddow et al. (2013), and other news sources, with priority on those which are likely to have incremental information on economic uncertainty relative to the FT, such as newswires.

References

Alexopoulos, M., & Cohen, J. (2009). *Uncertain times, uncertain measures* (No. 352). *University of Toronto, Department of Economics Working Papers*.

Alexopoulos, M., & Cohen, J. (2015). The power of print: Uncertainty shocks, markets, and the economy. *International Review of Economics and Finance*, *40*, 8–28. http://doi.org/10.1016/j.iref.2015.02.002

Baker, S. R., Bloom, N., & Davis, S. J. (2013). *Measuring Economic Policy Uncertainty*. Retrieved from www.policyuncertainty.com

Baker, S. R., Bloom, N., & Davis, S. J. (2015). Measuring Economic Policy Uncertainty. *Working Paper*, (October). Retrieved from http://www.nber.org/papers/w21633

Barndorff-Nielsen, O. E., Kinnebrock, S., & Shephard, N. (2008). *Measuring downside risk — realised semivariance*.

Bekaert, G., Engstrom, E., & Xing, Y. (2009). Risk, uncertainty, and asset prices. *Journal of Financial Economics*, *91*(1), 59–82. http://doi.org/10.1016/j.jfineco.2008.01.005

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, *77*(3), 623–685.

Bloom, N. (2014). Fluctuations in Uncertainty. *Journal of Economic Perspectives*, *28*(2), 153–176. http://doi.org/10.1257/jep.28.2.153

Bontempi, M. E., Golinelli, R., & Parigi, G. (2010). Why demand uncertainty curbs investment: Evidence from a panel of Italian manufacturing firms. *Journal of Macroeconomics*, *32*(1), 218–238.

Campbell, J. Y., Lettau, M., Malkiel, B. G., & Xu, Y. (2001). Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk. *The Journal of Finance*, *56*(1), 1–43. Retrieved from http://links.jstor.org/sici?sici=0022-1082(200102)56:1<1:HISBMV>2.0.CO;2-7

Cognitive Science Laboratory of Princeton University. (2010). About WordNet. Retrieved from http://wordnet.princeton.edu

Dendy, C., Mumtaz, H., & Silver, L. (2013). *An uncertainty index for the UK 1986-2012*.

Edwardes, M. D. (1995). A confidence interval for Pr(X < Y) − Pr(X > Y) estimated from simple cluster samples. *Biometrics*, *51*(2), 571–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7662846

Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association*, *82*(397), 171–185.

Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). The Econometric Society. *Econometrica*, *64*(4), 813–836.

Etner, J., Jeleva, M., & Tallon, J. M. (2012). Decision theory under ambiguity. *Journal of Economic Surveys*, *26*(2), 234–270. http://doi.org/10.1111/j.1467-6419.2010.00641.x

Feunou, B., Jahan-Par, M. R., & Tedongap, R. (2010). *Modeling Market Downside Volatility*.

Granger, C. W. J. (1998). Extracting information from mega-panels and high-frequency data. *Statistica Neerlandica*, *52*(3), 258–272. http://doi.org/10.1111/1467-9574.00084

Haddow, A., Hare, C., Hooley, J., & Shakir, T. (2013). *Macroeconomic uncertainty: what is it, how can we measure it and why does it matter? Bank of England Quarterly Bulletin*.

Heber, G., Lunde, A., Shephard, N., & Sheppard, K. (2009). Oxford-Man Institute's realized library, version 0.2. Retrieved January 20, 2015, from http://realized.oxford-man.ox.ac.uk/data

Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. *American Economic Review*, *105*(3), 1177–1216. http://doi.org/10.1257/aer.20131193

Juselius, K. (2006). *The Cointegrated VAR model: methodology and applications*. *Advanced Texts in*

Econometrics. Oxford: Oxford University Press.

Keynes, J. M. (1936). *General theory of employment, interest and money*. Atlantic Publishers & Dist.

Knight, F. H. (1921). *Risk, Uncertainty, and Profit*. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Co. Retrieved from http://www.econlib.org/library/Knight/knRUP.html

Kothari, S. P., Li, X., & Short, J. E. (2009). The Effect of Disclosures by Management, Analysts, and Business Press on Cost of Capital, Return Volatility, and Analyst Forecasts: A Study Using Content Analysis. *The Accounting Review*, *84*(5), 1639–1670.

Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, *17*(3), 1217–1241.

Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, *LXVI*(1), 35–65.

Marakova, S. (2014). *Risk and Uncertainty: Macroeconomic Perspective* (Economics and Business Working Papers No. 129).

Mead, N., & Blight, G. (2014). Eurozone crisis timeline. Retrieved January 23, 2015, from http://www.theguardian.com/business/interactive/2012/oct/17/eurozone-crisis-interactive-timeline-three-years

Merton, R. C. (1980). On estimating the expected return on the market. *Journal of Financial Economics*, *8*(4), 323–361.

Newson, R. (2002). Parameters behind “nonparametric” statistics: Kendall’s tau, Somers' D and median differences. *The Stata Journal*, *2*(1), 45–64.

Newson, R. (2005). Confidence intervals for rank statistics: Somers’ D and extensions. *The Stata Journal*, *6*, 497–520.

Ng, S., & Perron, P. (1995). Unit root tests in ARMA models with data-dependent models for the selection of the truncation lag. *Journal of the American Statistical Association*, *90*(429), 268–281.

Ng, S., & Perron, P. (2001). Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power. *Econometrica*, *69*(6), 1519–1554. Retrieved from http://www.jstor.org/stable/2692266

Rossi, B., & Sekhposyan, T. (2015). Macroeconomic Uncertainty Indices Based on Nowcast and Forecast Error Distributions. *American Economic Review: Papers & Proceedings*.

Schwert, G. W. (1989a). Business cycles, financial crises, and stock volatility. *Carnegie-Rochester Conference Series on Public Policy*, *31*, 83–126.

Schwert, G. W. (1989b). Why Does Stock Market Volatility Change Over Time? *The Journal of Finance*, *XLIV*(5), 1115–1153.

Segal, G., Shaliastovich, I., & Yaron, A. (2015). Good and bad uncertainty: Macroeconomic and financial market implications. *Journal of Financial Economics*, *117*(2), 369–397. http://doi.org/http://dx.doi.org/10.1016/j.jfineco.2015.05.004

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, *LXII*(3), 1139–1168.

Appendices

A. News-media data

A.1. Readership of the Financial Times

While it is now a global newspaper, with more copies sold abroad than in the UK since September 1998, the FT still has a wider coverage and more detailed analysis of news on UK-listed companies than other national daily newspaper, and it was the unrivalled UK business daily throughout our sample period. It continues to have a large UK readership: the UK print edition had an estimated daily average audience of 319,000, and the FT claimed to have 457,938 unique UK readers across all platforms (including web, mobile etc.) in 2012[52].

Evidence from various surveys suggests that among business decision-makers the FT daily reaches:

- 20% of 1.8 million business 'Purchase Decision Makers'[53] in the UK, and 31% among the sub-segment who 'worked on international business strategies in the past 12 months', which seems likely to correlate with being a key influencer in major investment decisions
- 24% of 435,000 senior business decision leaders across Europe[54]
- 28% of 3,900 senior finance staff of large organisations across the world (both non-financial and banks) responsible for raising finance from capital markets[55].

Among institutional investors the FT daily reaches:

- 36% of senior decision-makers in buy-side financial institutions globally[56]
- 79% of institutional buy-side investors with more than USD100million under management and based in UK/ROI[57].

Finally, compared to other publications, the FT was considered the most credible media owner in the reporting of financial and economic issues, by those who personally managed assets worth USD5billion or more (a universe of 2,522 individuals).

A.2. Auxiliary databases

To help in cleaning and de-duplicating the data, quality checking the results, and constructing a canonical calendar of FT publication days (see Appendix A.2) for use in daily frequency analyses, we also referred to records of the FT in Nexis UK, Proquest ABI/Inform[58], and Gale FT Historical Archive (for facsimile copies).

---

[52] Source: http://www.fttoolkit.co.uk/admediakit/pdfs/adga/Adga_Certificate_April_12_to_March_13.pdf, retrieved on 6 Nov 2013.
[53] Source: British Business Survey (BBS) 2011.
[54] Defined as "C-suites, Head of Department, other senior management and directors/VPs […] who sit in industrial and commercial companies with 250 or more employees, and if company turnover is greater than £40m the threshold is reduced to 150+ employees". Source: Business Elite Europe (BE:EUROPE) 2013 survey. We did not have access to segmentation for the UK
[55] Source: Global Capital Markets Survey (GCMS) 2011.
[56] Source: Global Capital Markets Survey (GCMS) 2011.
[57] Source: Worldwide Professional Investment Community (PIC) Study 2010, Erdos & Morgan Inc.
[58] Factiva coverage of the FT nominally begins 1 January 1981 but coverage for 1981 is unreliable. Nexis UK coverage starts on 1 January 1982; Proquest on 31 May 1996.

A.3. Maximal sample period

Our sample spans 1 January 1982 (earliest available machine-readable electronic copies) to 30 April 2014, excluding a gap in all available electronic databases from 2 June to 8 August 1983 inclusive (see Appendix A.2) which we reflect in missing values for 1983m6-1983m8 and 1983q2-1983q3, and pro-rata scaling for calendar year 1983. In quantitative analyses we use the longest span of complete calendar years (to ensure a common sample across all frequencies) for which we have continuous coverage (to avoid the complications of gaps in the time series) at the article level (to enable de-duplication; for 2013m11-2014m4 we only have top down counts from search results), i.e. 1984-2012.

A.4. Canonical set of FT publication days

We constructed a canonical set of FT publication days by cross-referencing daily article counts from Factiva, Nexis UK; monthly article counts from Proquest; selected daily facsimile copies of the print newspaper from Gale; and miscellaneous other sources used to establish reasons for non-publication on certain days (e.g. print stoppages).

The FT was usually published daily on Monday through Saturday throughout the sample period. We counted non-Sunday days with 50 or more articles after de-duplication as publication days. Typically an FT issue contains over 100 articles, but occasionally slimmer issues are published (e.g. on 27 December in some years). However, on a random subsample of the 76 days with between zero and 49 records in Factiva, all records were misclassified by date, appearing on a consecutive day in the Gale facsimile copy. We discarded all articles on such days rather than attempt to manually re-classify them, given the effort that would be required and the tiny scale of the resulting error in comparison to other potential sources of noise.

Most canonical non-publication days were either Sundays or Bank Holidays, but also included an FT coverage gap in all available electronic databases from 2 June to 8 August 1983 inclusive, and a handful of disparate days with no articles in any of the available databases, for assorted reasons[59].

A.5. Canonical daily total article counts

Due the greater volatility of duplication rates in Factiva, and the impossibility of reliably de-duplicating this total without access to the full text for every FT article (which we did not have, unlike with the subset of articles containing uncertainty keyphrases or tagged with particular companies) we fell back on the Nexis UK daily FT record count as our canonical daily total FT article count, subject to the following adjustments:

- For January 1990 to January 1992, June 1992 to December 1992, and October to November 2007 inclusive, we reduce the Nexis UK count by 20% as an approximate adjustment for the residual duplication observed during these periods.
- On seven publication days when Nexis UK has zero records, we substitute the Factiva record count (always greater than 50 in these cases) adjusted by a multiplying factor equal the linearly interpolated mean of the ratio of Nexis to Factiva records. We do likewise for 10 publication days when Nexis UK has more than zero but fewer than 50 records (Factiva has more than 50 records in all these cases).

[59] 11 August 1982, 22 September 1982, 27 June 1984 (printers stoppages). Friday 31 December 1999 (perhaps due to 'millennium bug' concerns and preparations); Saturday 11 April 1998 (Easter); Saturday 9 October 1982, Saturday 26 November 1983, Friday 6 July 1984 (no apparent reason).

- On five days when the canonical count is still between 50 and 70 records, we interpolate in the same way but using item counts and ratios from the Gale and Nexis databases rather than Factiva and Nexis. This is to avoid low (and erroneous) outliers.
- On Saturday 2 January 1999, which is a publication day but has no records in Nexis UK, due to a database error, we use the Factiva record count (which appears to be stable and reliable around that time).

B. Noise-to-signal ratio

B.1. Derivation of the estimator

We summarise the scale of the noise by the conditional variance $Var_{U-U^*|n}[U - U^*|n]$ and of the signal by $Var_{U^*|n}[U^*|n]$. Their ratio is:

$$NSR(n) \equiv \frac{Var_{U-U^*|n}[U - U^*|n]}{Var_{U^*|n}[U^*|n]} = \frac{Var_{U|n}[U|n] + Var_{U^*|n}[U^*|n] - 2Cov[U, U^*|n]}{Var_{U^*|n}[U^*|n]} \tag{9}$$
$$= \frac{Var_{U|n}[U|n]}{Var_{U^*|n}[U^*|n]} - 1$$

The first equality (after the identity) follows from the standard identity expanding the variance of a sum of random variables. The second equality follows from applying the law of total covariance, and the result that U is an unbiased estimator of U^*, to the covariance term[60], then cancelling terms and rearranging.

We can estimate $Var_U[U|n]$ directly from our sample. Under our model of the data generating process we can estimate $Var_{U^*}[U^*|n]$ as follows. Start from the law of total variance, use the unbiasedness property mentioned above (the second equality):

$$Var_{U|n}[U|n] = E_{U^*|n}\left[Var_{U|U^*,n}[U|U^*,n]|n\right] + Var_{U^*|n}\left[Var_{U|U^*,n}[U|U^*,n]|n\right] \tag{10}$$
$$= E_{U^*|n}\left[Var_{U|U^*,n}[U|U^*,n]|n\right] + Var_{U^*|n}[U^*|n]$$

$$\Rightarrow Var_{U^*|n}[U^*|n] = Var_{U|n}[U|n] - E_{U^*|n}\left[Var_{U|U^*,n}[U|U^*,n]|n\right] \tag{11}$$

For the variance in the second term on the RHS of (12), recall from Section 3.1 that the binomial case (with equal propensity to express uncertainty across all articles) gives an upper bound, so that where the n articles are obtained by aggregating over multi-day periods, with potentially varying latent uncertainties, this variance will be less than or equal to the expression in (1) on pg. 7. Hence:

$$E_{U^*|n}\left[Var_{U|U^*}[U|U^*,n]|n\right] \leq E_{U^*|n}\left[\frac{U^*(1-U^*)}{n}|n\right] \tag{12}$$
$$= \frac{1}{n}\left(E_{U^*|n}[U^*|n] - E_{U^*|n}[U^{*2}|n]\right)$$
$$= \frac{1}{n}\left(E_{U^*|n}[U^*|n]\left(1 - E_{U^*|n}[U^*|n]\right) - Var_{U^*|n}[U^*|n]\right)$$

[60] $Cov[U, U^*|n] = E_{U^*|n}[Cov[U, U^*|U^*, n]|n] + Cov\left[E_{U|n}[U|U^*, n], E_{U^*|n}[U^*|U^*, n]|n\right]$ $= 0 + Cov[U^*, U^*|n] = Var_{U^*|n}[U^*|n]$ where the first equality is the law of total covariance, and the second equality follows from the fact that U is an unbiased estimator of U^*, i.e. $E_{U|U^*,n}[U|U^*, n] = U^*$.

where the second equality follows from the definition of variance in terms of expectations ($Var_{U^*|n}[U^*|n] = E_{U^*}\left[U^{*2}|n\right] - E_{U^*|n}[U^*|n]^2$).

Substituting (12) in (10) and rearranging gives:

$$Var_{U^*|n}[U^*|n] \geq \frac{1}{n-1}\Big(n \cdot Var_{U|n}[U|n] - E_{U^*|n}[U^*|n]\big(1 - E_{U^*|n}[U^*|n]\big)\Big) \tag{13}$$

Next note that

$$E_{U|n}[U|n] = E_{U^*|n}\left[E_{U|U^*,n}[U|U^*,n]\right] = E_{U^*|n}[U^*|n] \tag{14}$$

where the first equality is the law of total expectation, and the second equality follows from the unbiasedness of our estimator, as expressed in (1) on pg. 7. Substituting (14) in (13), we arrive at:

$$Var_{U^*|n}[U^*|n] \geq \frac{1}{n-1}\Big(n \cdot Var_{U|n}[U|n] - E_{U|n}[U|n]\big(1 - E_{U|n}[U|n]\big)\Big) \tag{15}$$

Finally, substituting (15) in (9) and replacing population moments with their sample analogues (indicated by hats), we arrive at an estimated upper bound for NSR in terms of observables:

$$\widehat{NSR(n)} = \frac{(n-1)\widehat{\sigma^2_{U,n}}}{n\widehat{\sigma^2_{U,n}} - \widehat{\overline{U_n}}\big(1 - \widehat{\overline{U_n}}\big)} - 1 \tag{16}$$

where $\widehat{\overline{U_n}}$ and $\widehat{\sigma^2_{U,n}}$ are respectively the standard sample estimators of the expectation and variance of U based on observations at the periodicity corresponding to n.

Finally, substituting (16) in (9), replacing population moments with their sample analogues, and rearranging, we have an estimate of the true latent variance of U^*, $\widehat{\sigma^2_{U^*,n}} = \widehat{\sigma^2_{U,n}}/\big(1 + \widehat{NSR(n)}\big)$.

B.2. Using all FT articles vs. only articles containing economic keyphrases

The sampling variance associated with measuring uncertainty across *all* FT articles ($\mathrm{U/T}$, in the notation of Section 5.4), cf. only those containing "economic" or "economy" ($\mathrm{EU/E}$), can be estimated in each case using the NSR estimator above:

$$NSR\widehat{(\mathrm{EU/E})} = \frac{\left(\widehat{\overline{\mathrm{E}}} - 1\right)\widehat{\sigma_{\mathrm{EU/E}}}^2}{\widehat{\overline{\mathrm{E}}}\widehat{\sigma_{\mathrm{EU/E}}}^2 - \widehat{\overline{\mathrm{EU/E}}}\left(1 - \widehat{\overline{\mathrm{EU/E}}}\right)} - 1 \approx 0.075 \tag{17}$$

$$NSR\widehat{(\mathrm{U/T})} = \frac{\left(\widehat{\overline{\mathrm{T}}} - 1\right)\widehat{\sigma_{\mathrm{U/T}}}^2}{\widehat{\overline{\mathrm{T}}}\widehat{\sigma_{\mathrm{U/T}}}^2 - \widehat{\overline{\mathrm{U/T}}}\left(1 - \widehat{\overline{\mathrm{U/T}}}\right)} - 1 \approx 0.053 \tag{18}$$

where the quantities on the right-hand sides are obtained by the method, and on the monthly sample, described in the notes to Table 4 on pg. 30 (though not all the quantities used here are reported in that Table). The ratio of $NSR\widehat{(\mathrm{U/T})}$ to $NSR\widehat{(\mathrm{EU/E})}$ is 0.71, corresponding to an approximately 30% reduction in NSR due to using *all* FT articles.

C. Unit root tests

Table 12: Unit-root tests on news-media uncertainty, U, 1984m1–2012m12

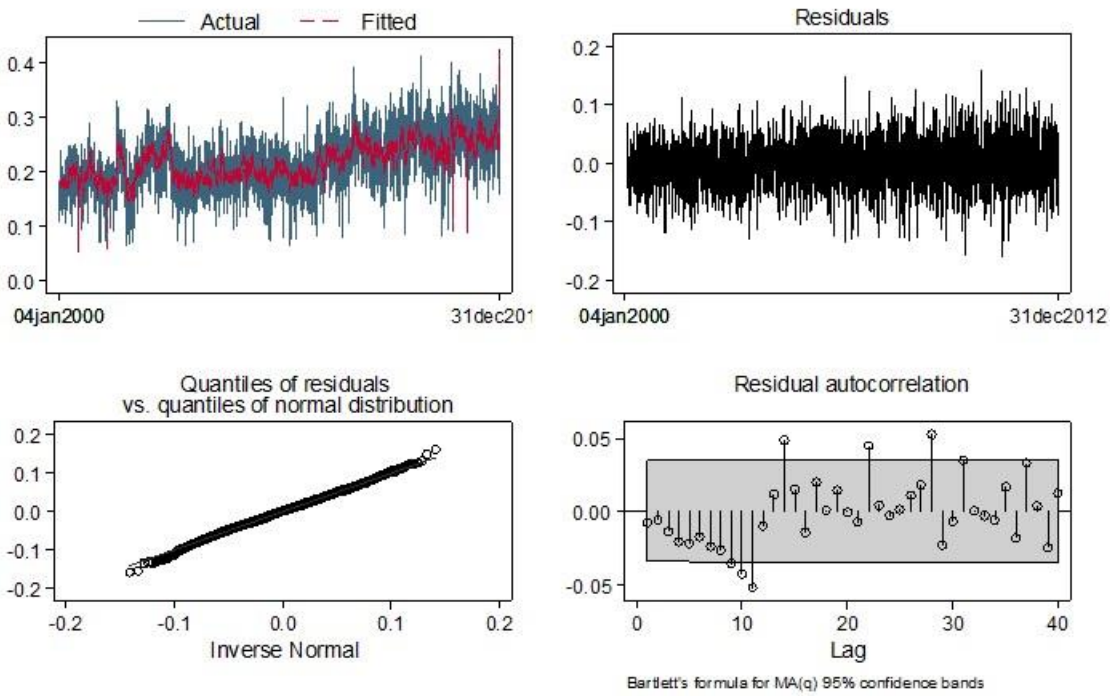
Deterministic trend?	Lag criterion	Lags	α		1%	5%	10%
Linear	Ng-Perron	34	-2.893	**	-3.480	-2.835	-2.547
	Min SC	16	-4.241	***	-3.480	-2.838	-2.550
	Min MAIC	34	-2.893	**	-3.480	-2.835	-2.547
None	Ng-Perron	34	-2.368	**	-2.580	-1.947	-1.623
	Min SC	16	-3.402	***	-2.580	-1.949	-1.625
	Min MAIC	34	-2.368	**	-2.580	-1.947	-1.623

Notes: α is the parameter on the lagged dependent variable in the DF-GLS test of Elliott et al. (1996). The null is that U has a unit root, and the alternative is stationarity or linear trend-stationarity (corresponding to "None" or "Linear" in the "Deterministic trend" column). Lag selection is according to one of three alternative standard criterion: sequential-t algorithm of Ng & Perron (1995), minimum Schwarz information criterion (min SC), and minimum modified information criterion of Ng & Perron (2001). Critical values shown in the last three columns are interpolated as per Stata command −dfgls−. All tests use 7290 observations. *, ** and *** indicate rejection of the null at 10%, 5% and 1% levels respectively.

D. VAR results

Figure 26: Residual diagnostics for daily VAR, 2000–2012

Dependent variable: $\sqrt{U_t}$



Dependent variable: $\ln(\sigma_t)$

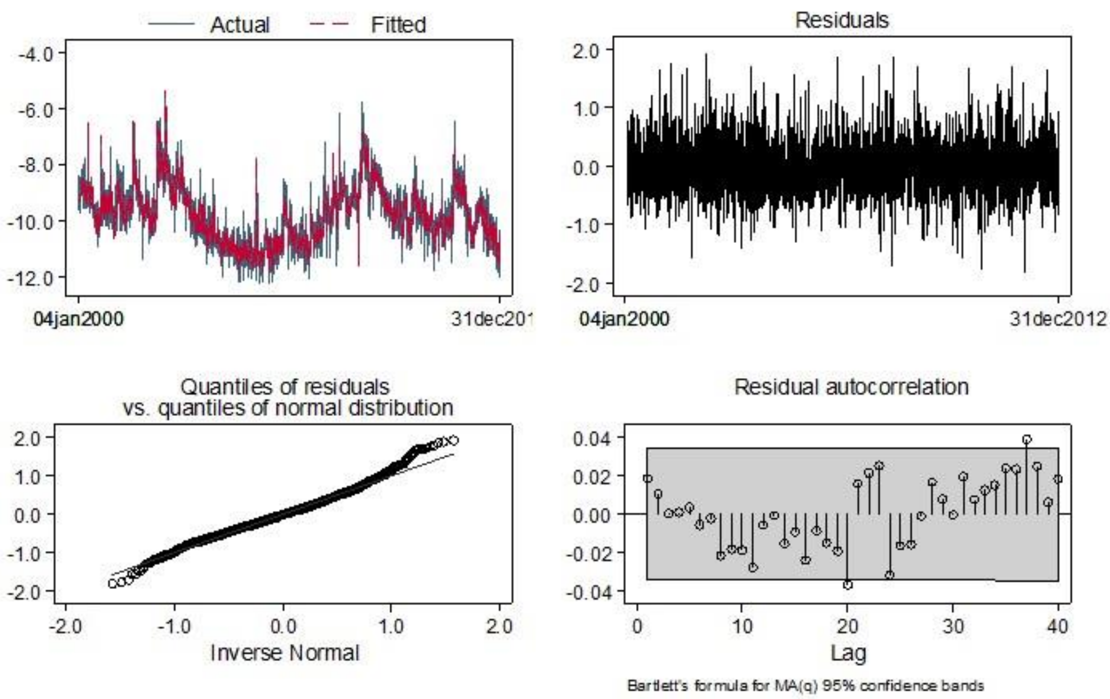
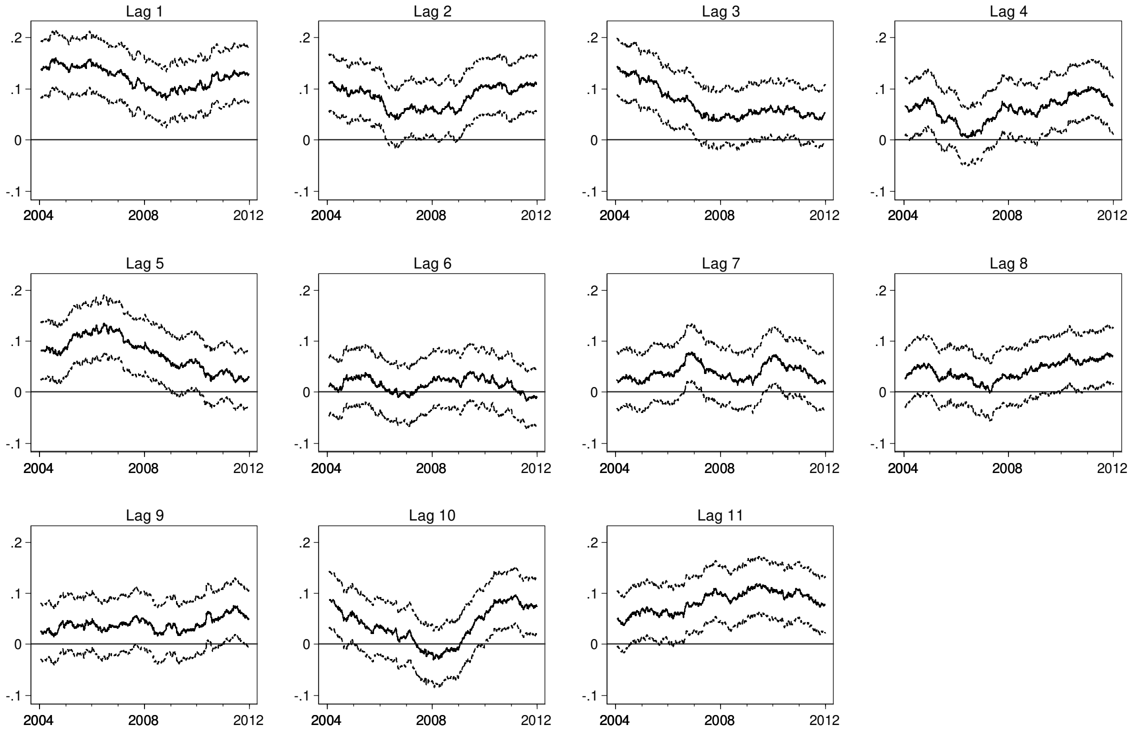


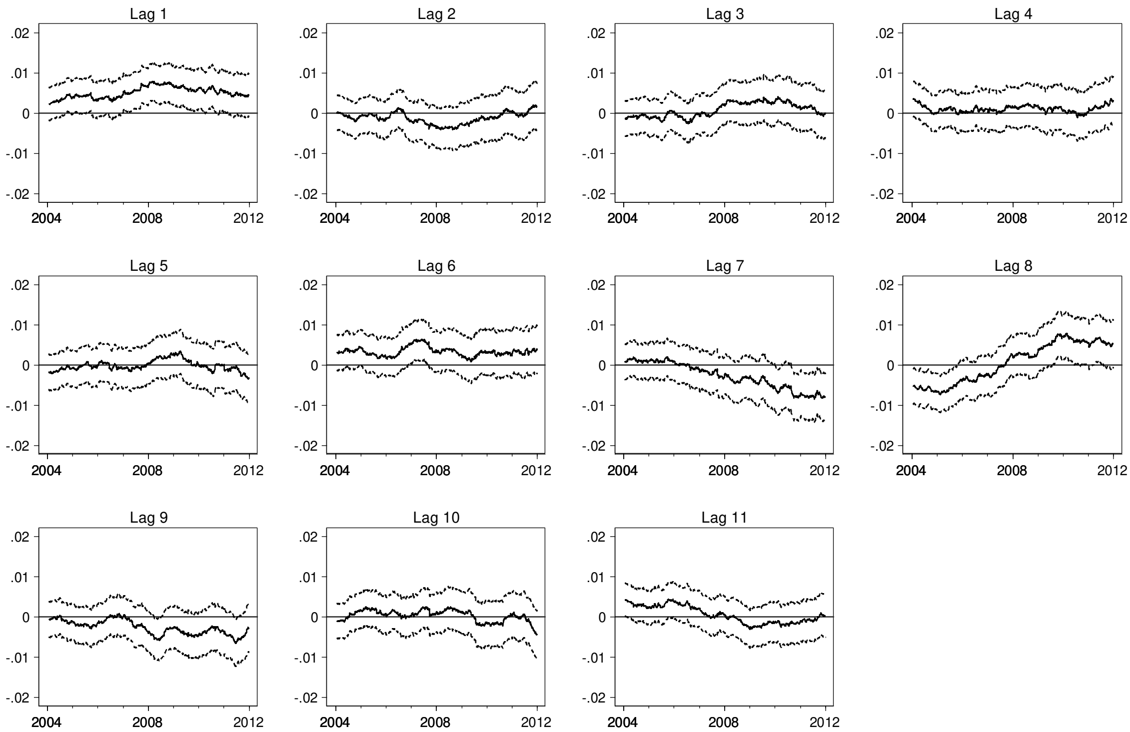
Figure 27: Rolling parameter estimates, daily VAR model, 2000–2012

Dependent variable: U

Explanatory variables: lags of U



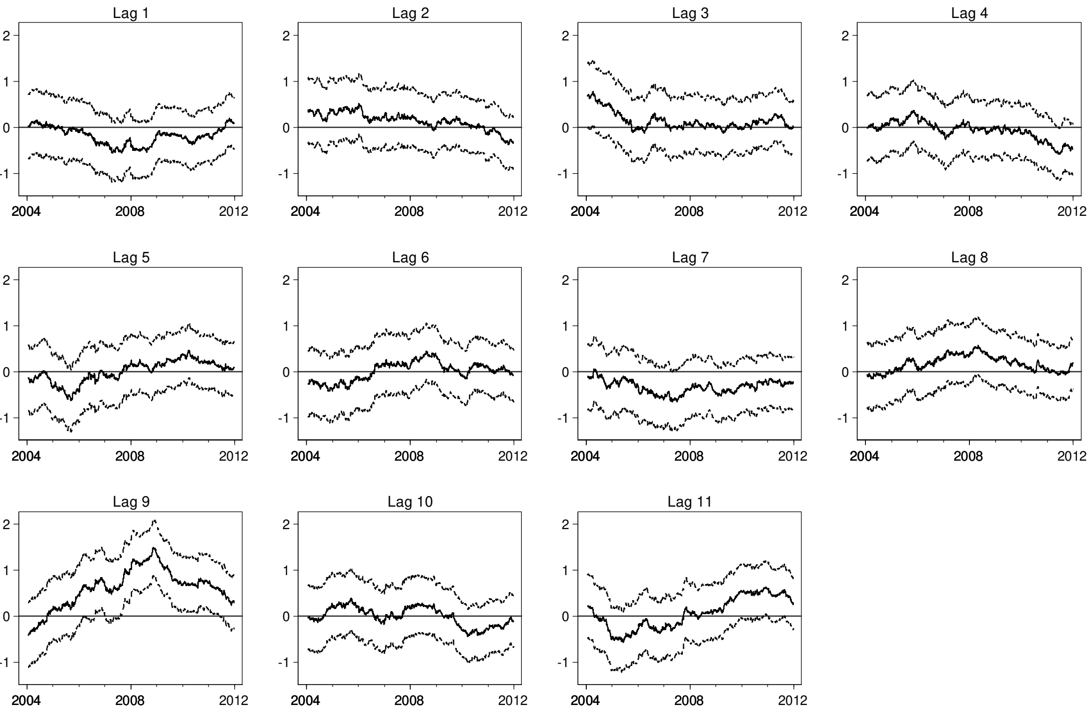
Explanatory variables: lags of σ



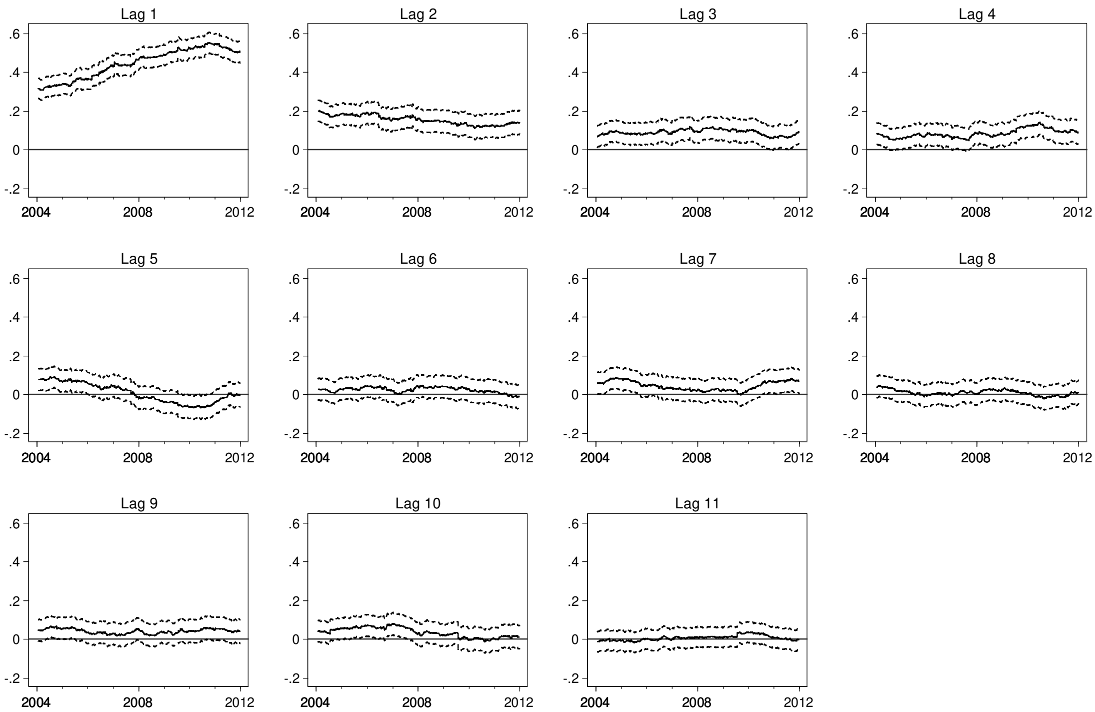
Note: horizontal axis indicates end of five year rolling window.

<u>Dependent variable:</u> σ

<u>Explanatory variables:</u> lags of U



<u>Explanatory variables:</u> lags ofσ



Note: horizontal axis indicates end of five year rolling window.