January 2013

Working Title: "Economic Policy Uncertainty In China"
Scott Baker, Nick Bloom, Steven J. Davis, and Sophie Wang

A. Overview

To measure economic policy uncertainty for China, we construct a scaled frequency count of articles about policy-related economic uncertainty in the South China Morning Post (SCMP), Hong Kong's leading English-language newspaper. The method follows our news-based indexes of economic policy uncertainty for the United States and other countries.

We proceed as follows. First, we identify SCMP articles about economic uncertainty pertaining to China by flagging all articles that contain at least one term from each of the China EU term sets: {China, Chinese} and {economy, economic} and {uncertain, uncertainty}. Second, we identify the subset of the China EU articles that also discuss policy matters. For this purpose, we require an article to satisfy the following text filter: {policy OR spending OR budget OR political OR "interest rates" OR reform} AND {government OR Beijing OR authorities}} OR tax OR regulation OR regulatory OR "central bank" OR "People's Bank of China" OR PBOC OR deficit OR WTO. We use this compound filter because it outperforms simpler alternatives in our audit study. Third, we apply these requirements in an automated search over every SCMP article published since 1995. This automated search yields a monthly frequency count of SCMP articles about policy-related economic uncertainty. Fourth, we divide the monthly frequency count by the number of all SCMP articles in the same month. We then normalize the resulting series to a mean value of 100 from January 1995 to December 2011 by applying a multiplicative factor.

A few examples clarify how the compound text filter works. If an article includes both "policy" and "government", we regard it as at least partly about government policy. Therefore, if it also contains a word in each of the China EU term sets described above, the article enters our frequency count for economic policy uncertainty. However, the word "policy" is not sufficient for an article to count as policy uncertainty; the article must also contain one of "government", "Beijing" or "authorities". Certain other terms in our text filter – e.g., "tax" or "regulation" – do not involve a compound requirement. We determined when to apply a compound requirement based on our audit study.

Our audit study considers 500 randomly sampled articles drawn from the universe of SCMP articles that satisfy the China EU term sets. The sampling period is January 1995 to February 2012. We subject all 500 sampled articles to human readings to evaluate the accuracy of various text filters. In assessing accuracy, we regard the classifications produced by the human readings as correct.

According to the human readings, 492 of the 500 sampled articles pertain to economic uncertainty for China. The remaining 8 articles were incorrectly flagged by the automated

search method as pertaining to economic uncertainty for China. In other words, the China EU term sets produce a very small false positive error rate for economic uncertainty pertaining to China.

Using automated methods to further classify the articles as about policy-related economic uncertainty, or not, is more challenging. Here as well, however, our preferred text filter (described above) produces good results:

- The policy-related economic uncertainty count produced by automated search methods exhibits a correlation of 0.82 with the true count (human reading) in quarterly time-series data.
- The net error rate produced by automated search methods is nearly uncorrelated (-0.15) with the true count in quarterly time-series data.
- The overall false positive rate produced by the automated method is 0.11. The overall false negative rate is 0.21.

B. Some Details of our China EU and EPU Indexes Based on the SCMP

This section describes our construction of China economic uncertainty and economic policy uncertainty indexes based on the South China Morning Post, a Hong Kong based English-language newspaper. First, we construct China EU and EPU universes automatically through the ProQuest database, then perform auditing and use the audit results to refine our text-string search criteria for identifying articles that discuss policy-related aspects of economic uncertainty in China. We also assess the error rate properties that emerge from our preferred text-string search criteria, and discuss further plan to build weighted China indexes including newspapers from other countries in the East Asia area. We have access to SCMP from three different sources: ProQuest database (1995 - present), Factiva database (2003 - present), and SCMP online archive (recent).

Two major issues emerge along the way: how to identify articles that actually discuss China economic policy uncertainty (i.e. the accuracy concern); and how to reconcile different algorithm rules from various data sources. To address the first issue, we use both automatic algorithm and human review to audit sample articles from SCMP, then refine our EU and EPU filters based on the auditing results. As for the second issue, our construction of SCMP indexes also provides a unique opportunity to test and refine search criteria from different data sources (i.e. between ProQuest and Factiva databases).

Section I describes our preliminary China economic and economic policy uncertainty indexes based on SCMP. Section II explains our methodology for the auditing process. (Sophie: I'm really not sure here. Shall we talk mainly about theoretical or technical methodology for auditing in Section II? And we already wrote appendix for both anyway.) Section III demonstrates statistical audit results of our original China EPU filter, and Section IV proposes and compares various alternative EPU filters based on our audit results. Section V explores the discrepancies between different data sources. Section VI discusses our plan to construct weighted China indexes based on newspapers from a couple of countries in East Asian area.

Section I: Preliminary results of China EU and EPU indexes based on SCMP from ProQuest

We use raw article counts from ProQuest database to construct China economic uncertainty and economic policy uncertainty indexes from Jan 1995 to Mar 2012. The China EU filter is {(China OR Chinese) AND (economic OR economy) AND (uncertain OR uncertainty)}. The China EPU filter includes additional policy-related terms: {policy OR tax OR spending OR regulation OR budget OR deficit OR PBOC OR "People's Bank of China"}. We then read sample articles that feed into our filters, and label major spikes below.

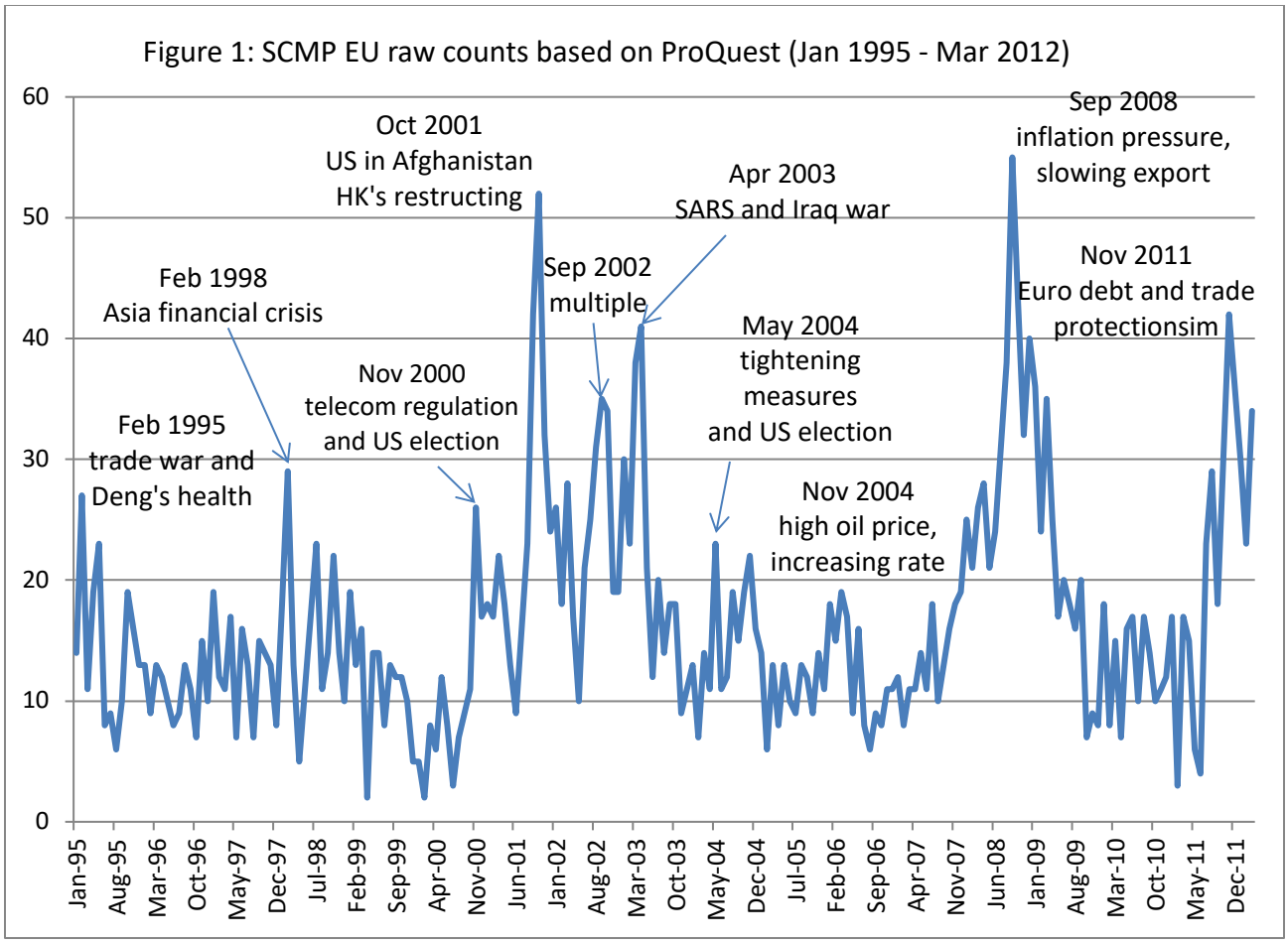


Figure 1: SCMP EU raw counts based on ProQuest (Jan 1995 - Mar 2012)
60
50
40
30
20
10
0
Feb 1995
trade war and
Deng's health
Feb 1998
Asia financial crisis
Nov 2000
telecom regulation
and US election
Oct 2001
US in Afghanistan
HK's restructing
Sep 2002
multiple
Apr 2003
SARS and Iraq war
May 2004
tightening
measures
and US election
Nov 2004
high oil price,
increasing rate
Sep 2008
inflation pressure,
slowing export
Nov 2011
Euro debt and trade
protectionsim
Jan-95
Aug-95
Mar-96
Oct-96
May-97
Dec-97
Jul-98
Feb-99
Sep-99
Apr-00
Nov-00
Jun-01
Jan-02
Aug-02
Mar-03
Oct-03
May-04
Dec-04
Jul-05
Feb-06
Sep-06
Apr-07
Nov-07
Jun-08
Jan-09
Aug-09
Mar-10
Oct-10
May-11
Dec-11

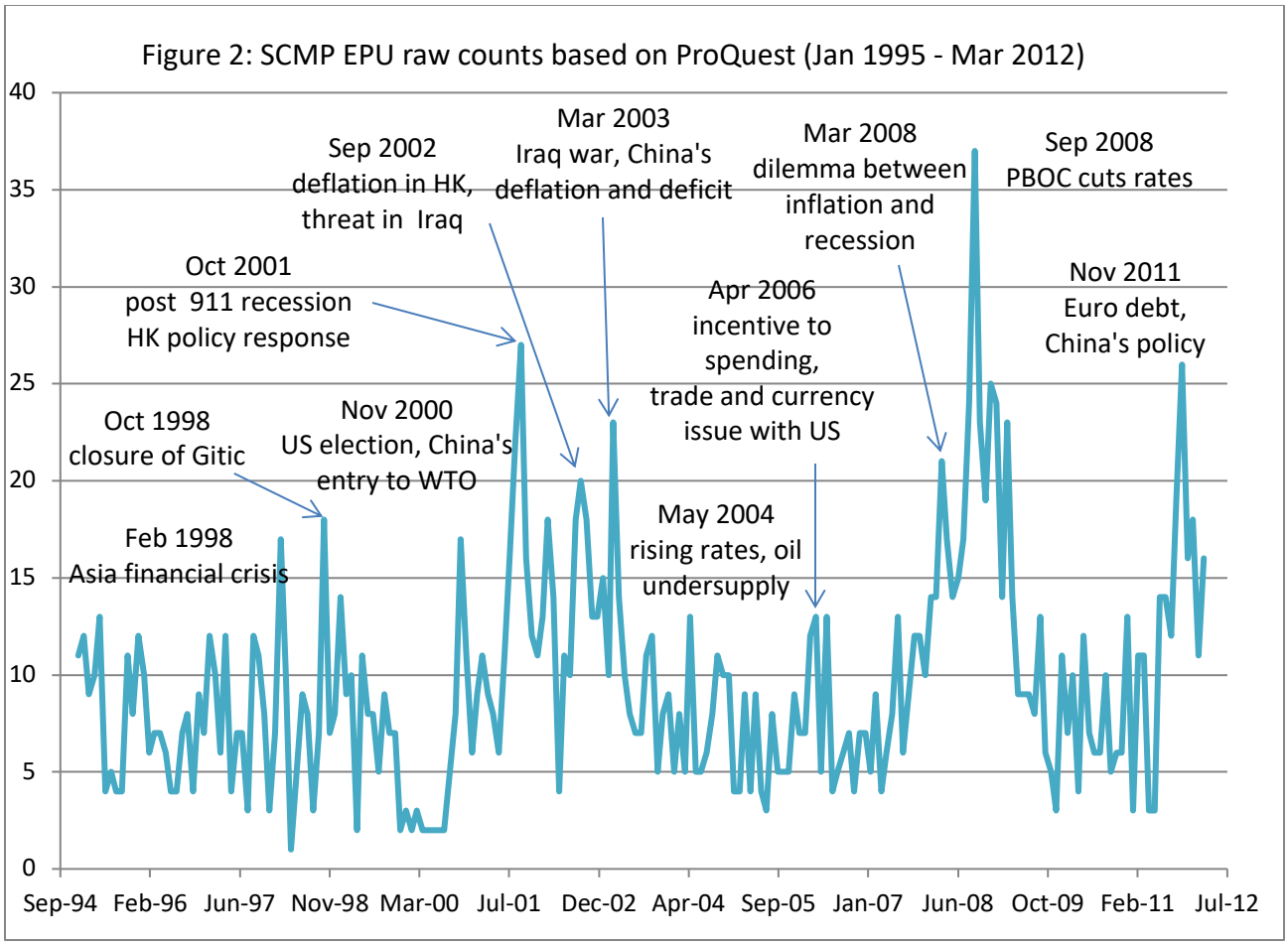
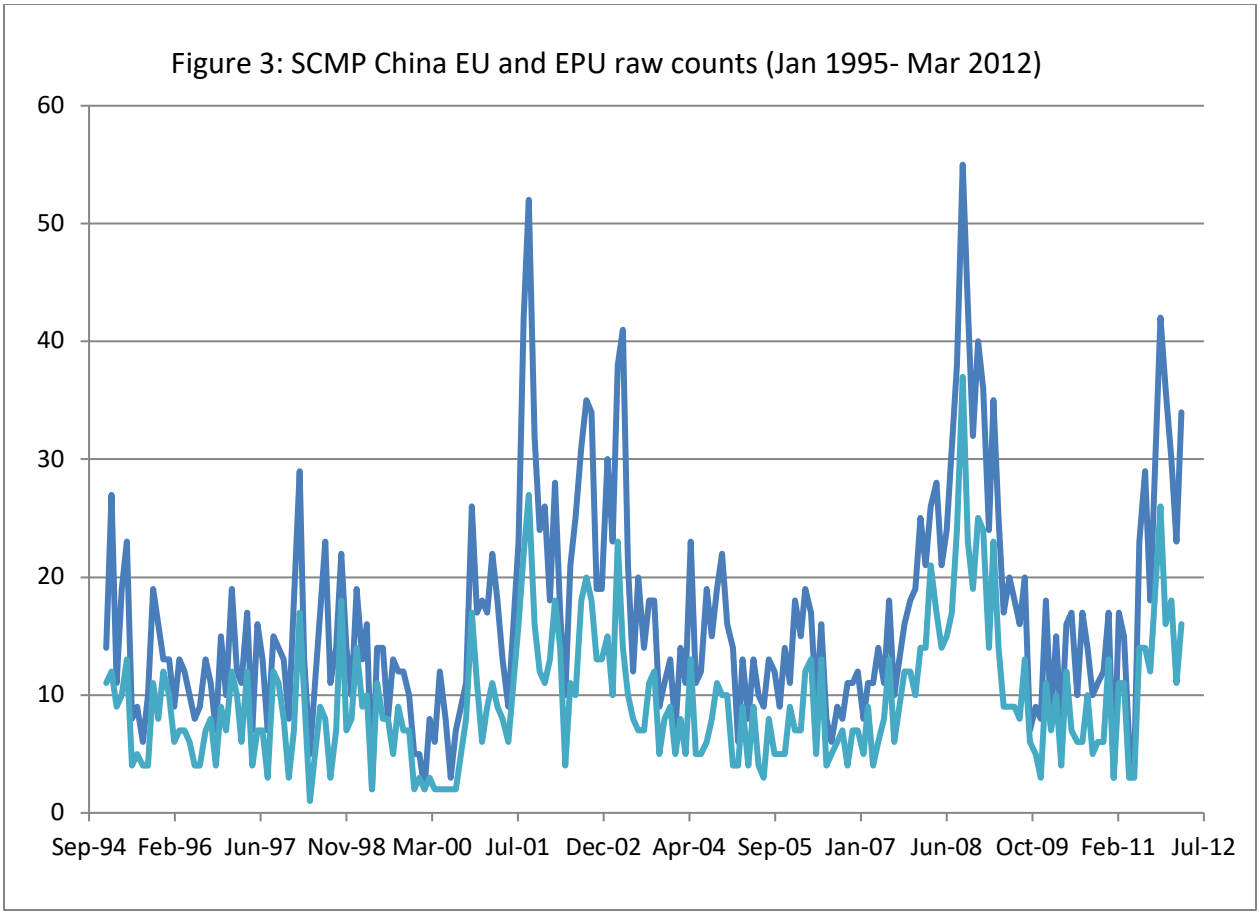


Figure 2: SCMP EPU raw counts based on ProQuest (Jan 1995 - Mar 2012)
40
35
30
25
20
15
10
5
0
Sep-94
Feb-96
Jun-97
Nov-98
Mar-00
Jul-01
Dec-02
Apr-04
Sep-05
Jan-07
Jun-08
Oct-09
Feb-11
Jul-12
Oct 1998
closure of Gitic
Feb 1998
Asia financial crisis
Nov 2000
US election, China's
entry to WTO
Oct 2001
post 911 recession
HK policy response
Sep 2002
deflation in HK,
threat in Iraq
Mar 2003
Iraq war, China's
deflation and deficit
Apr 2006
incentive to
spending,
trade and currency
issue with US
May 2004
rising rates, oil
undersupply
Mar 2008
dilemma between
inflation and
recession
Sep 2008
PBOC cuts rates
Nov 2011
Euro debt,
China's policy



Section II: Audit Methodology for China SCMP-Based Index

Two distinct measurement concerns arise in connection with our news-based indexes of uncertainty. One concern is whether specific text-string search criteria accurately identify the set of articles that discuss a certain type of uncertainty. A second concern is whether an accurate count for news articles about a particular type of uncertainty provides a good indicator for that type of uncertainty. We refer to these as the "accuracy" concern and the "suitability" concern. The primary objective of our audit study is to address accuracy concerns related to our China EPU index. We then use the audit results to refine our text-string search criteria for identifying articles that discuss policy-related aspects of economic uncertainty in China.

For a given set of news sources, we seek to measure and characterize the (scaled) number of articles that discuss policy-related and overall economic uncertainty. Let X^{PU} and X^{EU} denote

the true number of articles about policy-related economic uncertainty and overall economic uncertainty, respectively. We do not observe X^{PU} or X^{EU}. Instead, we derive measured counts of articles about policy-related and overall economic uncertainty by conducting automated searches according to our text-string search criteria.

Consider a time-series model for the relationship between the measured and true count of articles about policy-related economic uncertainty:

$$Y_t^{PU} = X_t^{PU} + FP_t^{PU} - FN_t^{PU}, \qquad (1)$$

where Y^{PU} is the measured count of articles about policy uncertainty, and FP^{PU} and FN^{PU} are the numbers of false positives and false negatives in the measured count. A false positive occurs when the search criteria identifies an article as about policy uncertainty when it is, in fact, about something else. By contrast, a false negative occurs when our search criteria fail to identify an article about policy-related aspects of economic uncertainty.

We conduct an audit study below on the set of articles identified by: {economic, economy} AND {uncertain, uncertainty} AND {China, Chinese}. We draw a random sample of 500 articles from this set for human review and coding and for comparison to the coding produced by automated methods.

<u>Constructing the Audit Universe and Count of Articles about "Economic Uncertainty"</u>

1. Determine the audit universe AU for the source and sample period:
    a. "Crawl" the source and sample period to identify all articles in *AU.*
    b. Identify and store hyperlinks to each article in AU in the Audit Universe file.
    c. Create a unique ID for each article and label the locally stored version of the article. Count the number of articles in AU for each newspaper-month.

<u>Identifying the Audit Sample and Storing</u>

2. Randomly sample from AU with sampling probability, *p*:

a. Load the list of articles in *AU* into STATA, and run "gen u_value = runiform()", Store the resulting u_value in the *AU* file.
b. Apply the sampling probability, *p,* by sampling all articles in *AU* with *p* < u_value.
c. Download each sampled article in PDF or other standard format, and store it in a dedicated directory for the source.

3. Create a single Audit Sample file that reports the following monthly data for each source:
    a. Source
    b. Year and month and day
    c. Number of articles in the AU for the source-month
    d. Sampling probability, *p,* used for the source-month
    e. Number of sampled articles for the source-month

<u>Performing the Audit</u>

4. Read each article. Set *EU* = 1 if article discusses economic uncertainty; set *EU* = 0, otherwise. We require an article to contain remarks about one or more aspects of economic uncertainty in order to code it as $EU = 1$.
5. Set $Y = 1$ if the article discusses policy-related aspects of economic uncertainty; set $Y = 0$ otherwise. We require an article to contain remarks about policy-related aspects of economic uncertainty in order to code it as $Y = 1$.
6. Set $M^k = 1$ if the article is in PU^k; that is, if policy filter f^k classifies the article as about policy-related aspects of economic uncertainty. Set $M^k = 0$, otherwise.
7. Identify and annotate false positives and false negatives:
    a. If $Y = 0$ and $M^k = 1$ (False Positive), identify and list the terms in f^k that appear in the article. Identify and store all terms in f^k for each article that satisfies $Y = 0$ and $M^k = 1$.
    b. If $Y = 1$ and $M^k = 0$ (False Negative), identify and list terms in the article used to discuss policy-related aspects of economic uncertainty, and that are potentially useful in refining the policy filter.

8. Provide additional classifications of articles that discuss policy-related aspects of economic uncertainty. That is, for articles in PU^H ($Y = 1$):
    a. (Direction) If passages in the article about policy-related aspects of economic uncertainty mainly discuss actual or prospective declines in policy-related economic uncertainty, set $Down = 1$. Otherwise, set $Down = 0$.
    b. (Mention US) If passages about policy-related aspects of economic uncertainty mention developments, concerns or consequences related to US policy makers or their actions, set $MentionUS = 1$. Otherwise, set $MentionUS = 0$.
    c. (Mainly US) If passages about policy-related aspects of economic uncertainty mainly discuss developments, concerns or consequences related to US policy makers or their actions, set $MainlyUS = 1$. Otherwise, set $MainlyUS = 0$. Note that $MainlyUS$ involves a stricter test than $MentionUS$. If $MainlyUS = 1$ for a given article, then $MentionUS = 1$ as well.
    d. (Mention Other Foreign) If passages about policy-related aspects of economic uncertainty mention developments, concerns or consequences related to other foreign policy makers or their actions, set $MentionForeign = 1$. Otherwise, set $MentionForeign = 0$.
    e. (Mainly Other Foreign) If passages about policy-related aspects of economic uncertainty mainly discuss developments, concerns or consequences related to foreign policy makers or their actions, set $MainlyForeign = 1$. Otherwise, set $MainlyForeign = 0$.
    f. (About Mainland) Set $AboutMainland = 1$ if the article contains any discussion of policy-related concerns and uncertainties in mainland China, including concerns and uncertainties related to U.S. and other foreign policy developments, actions and consequences, and including developments related to Hong Kong or Taiwan; otherwise, set $AboutMainland = 0$.
    g. (About Hong Kong) Set $AboutHK = 1$ if the article contains any discussion of policy-related concerns and uncertainties in Hong Kong; otherwise, set $AboutHK = 0$.

h. (About Taiwan) Set $AboutTaiwan = 1$ if the article contains any discussion of policy-related concerns and uncertainties in Taiwan; otherwise, set $AboutTaiwan = 0$.

i. (Type) Identify the type of policy uncertainty:

   i. (Who) If the article discusses uncertainty about the outcome of elections or other contests for political power (who will win), set $Who = 1$. Otherwise, set $Who = 0$.

   ii. (Actions) If the article discusses uncertainty about what actions will be taken by political leaders or government officials (what they will do, or when), set $Actions = 1$. Otherwise, set $Actions = 0$.

   iii. (Effects) If the article discusses uncertainty about the effects of past, present or future policy decisions or government actions, set $Effects = 1$. Otherwise, set $Effects = 0$.

Section III: Audit Results of the Original China EPU Filter Based on SCMP (Jan 95 - Feb 12)

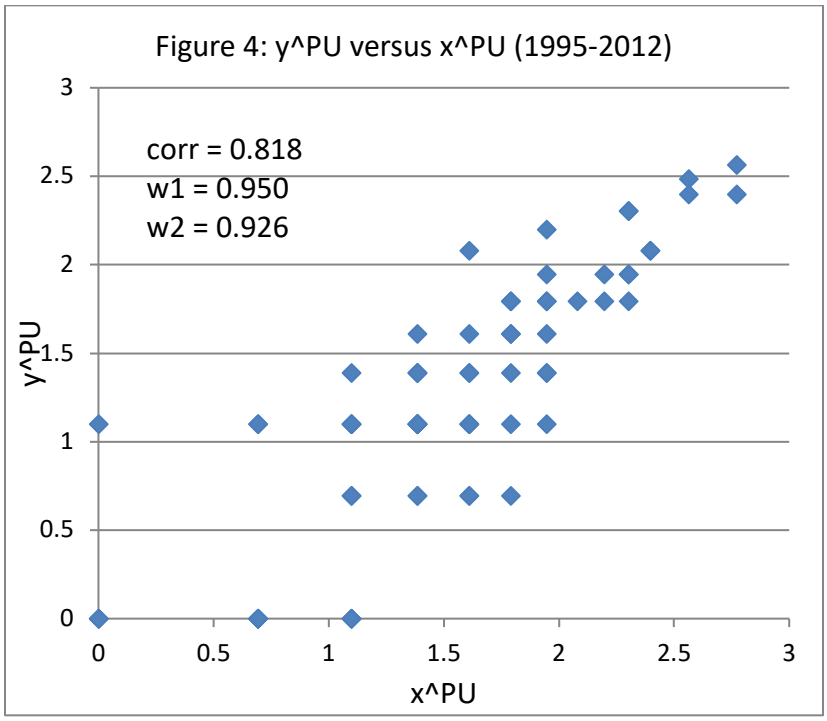
We conduct an audit study on the set of articles identified by: {economic, economy} AND {uncertain, uncertainty} AND {China, Chinese}. We draw a random sample of articles from this set for human review and coding and for comparison to the coding produced by automated methods. Our current policy filter for the ProQuest database is f = {policy, tax, spending, regulation, PBOC, central bank, People's Bank of China, budget, deficit}.

Out of the 500 audited articles that pass the China economic uncertainty filter, there are 492 articles that actually report China economic uncertainty (i.e. X^EU). 305 articles pass our China economic policy uncertainty filter (i.e. Y^PU), while human review identifies 389 articles about economic policy uncertainty (i.e. X^PU). The automated searching result and human review are very similar, with a 0.82 correlation during the full sample period (see Figure 4).

Table 1: summary of SCMP auditing results for full sample and two subsamples

Time	Y	X	FP	FN	y	x	p	n

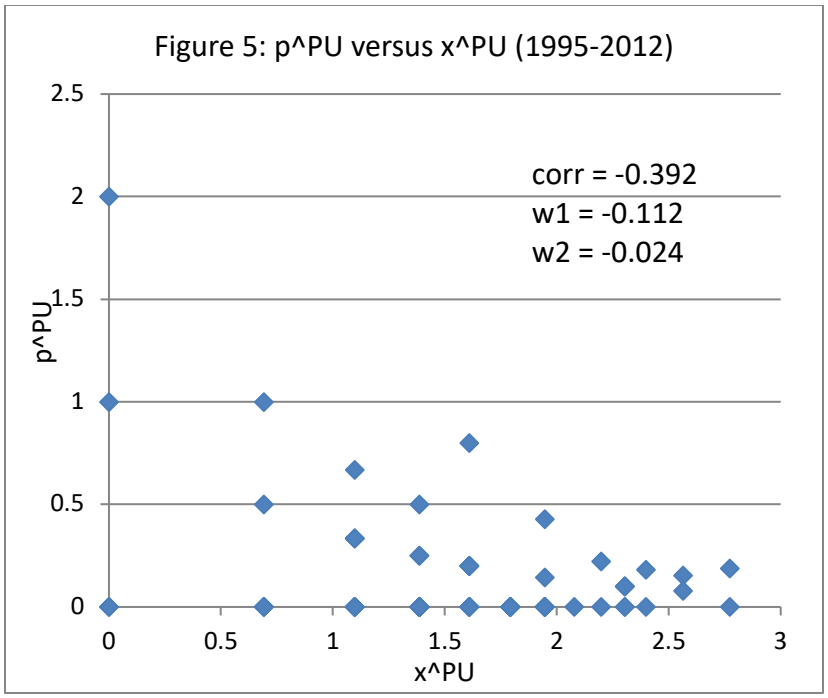
Full sample	305	389	40	122	5.7203	5.9636	0.1028	0.3136
1995-2003	156	202	21	65	5.0499	5.3083	0.1040	0.3218
2004-2012	149	187	19	57	5.0039	5.2311	0.1016	0.3048



*note 1: We pool to the quarterly level here and in the other charts, because the monthly count of audited articles is often very small.

*note 2: In Figure 4, w1 stands for correlation weighted by Y^PU, and w2 stands for correlation weighted by y^PU. While in Figures 5 and 6, w1 stands for correlation weighted by X^PU, and w2 stands for correlation weighted by x^PU.

Figure 5 and Figure 6 demonstrates that the two measurement errors (i.e. false positive rate and false negative rate) are both mildly negative correlated with the "true" EPU level. This indicates that we are not systematically exaggerating the level of policy uncertainty in the periods we identify as EPU spikes.



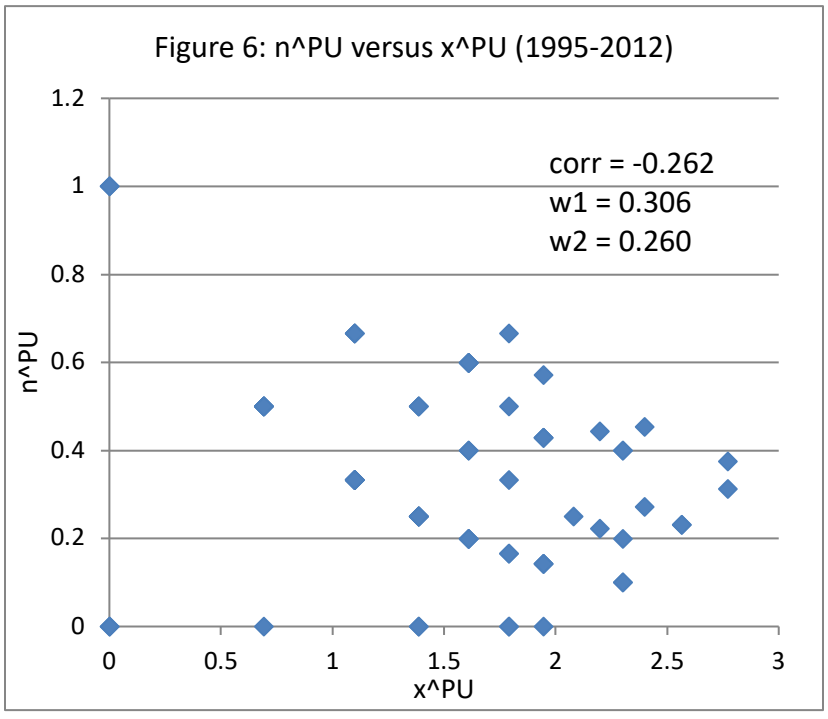


Table 2 and Table 3 below decompose our audit results into different category in more detail, such as various policy categories, related-countries, policy action versus effect, and policy uncertainty going down ratio.

Table 2: Intensity and Composition of Policy-Related "true" EU by Time Period.

time period	1995-2003	2004-2012	Full sample
Economic Uncertainty	125.0	128.1	126.5
Economic Policy Uncertainty	99.0	101.0	100
1. Monetary policy	23.5	43.2	32.9
2. Taxes, spending & fiscal	24.5	33.5	28.8
2a. Fiscal Policy	11.3	16.8	13.9
2b. Taxes	10.3	16.8	13.4
2c. Government spending	9.8	10.8	10.3
3. Entitlement programs	9.8	7.6	8.7
4. Health care	3.4	2.2	2.8
5. Financial regulation	23.0	22.2	22.6
6. Labor regulation	7.8	4.9	6.4
7. Energy & environmental	9.3	13.0	11.1
8. National security	17.1	8.6	13.1
9. Sovereign debt & currency	23.0	24.9	23.9
10. Trade policy	22.1	10.8	16.7
11. Competition policy	11.8	4.3	8.2
12. Legal policy	12.7	10.3	11.6
13. Political conflict & leader	44.1	26.5	35.7
14. Foreign investment	8.8	9.2	9.0
Sum of Rows 1 to 14	241.2	221.0	231.6
Policy Uncertain / Econ Uncertain	0.792	0.789	0.791

Table 3: Additional Classification of Policy-Related "true" Economics Uncertainty.

time period	1995-2003	2004-2012	Full sample
Economic Policy Uncertainty	99.0	101.0	100
1. Direction down	8.8	5.9	7.5
2. Mention US	38.2	35.1	36.8
3. Mainly US	11.8	8.1	10.0
4. Mention other foreign	27.0	21.6	24.4
5. Mainly other foreign	13.7	9.2	11.6
6. About mainland	80.9	85.4	83.0
7. About Hong Kong	58.3	39.4	49.4
8. About Taiwan	11.8	7.6	9.8
9. Who political election	20.1	9.2	14.9
10. Policy actions	77.0	72.9	75.1
11. Policy Effects	62.3	75.1	68.4

*note 3: Table 2 and Table 3 are based on "true" (i.e. human review) economic and economic policy uncertainty. All the entries are based on human review of the 500 out of 3396 SCMP articles from Jan 1995 to Feb 2012 in the ProQuest database.

*note 4: The “true” economic and economic policy uncertainty article counts are 492 and 389, respectively, in our audit sample, which contains 500 randomly selected articles that pass the economic uncertainty filter. 255 EU and 202 PU articles in the subsample 1995-2003; and 237 EU and 187 PU articles in the subsample 2004-2012.

*note 5: Raw article counts of policy-related news are scaled by the number of months in each period to get the frequencies, then expressed as a percentage of the “true” economic policy uncertainty index value for the full sample period, shown in the Table 2 and Table 3.

Section IV: Comparison and Refinement of Alternative China EPU Filters Based on SCMP

First, we present relevant key words in our 40 FP and 122 FN out of 500 auditing articles.

Table 4: relevant key words in FP articles from SCMP auditing results

search term	policy	spending	budget	tax	deficit	regulation
article counts	16	12	8	7	4	2

Table 5: relevant key words in FN articles from SCMP auditing results

Text	Beijing, authorities	political	Interest rates	reform	WTO	regulatory
Articles	23	21	14	13	8	7
Text	Tightening measures	Stimulus measure	Trade agreement	Five-Year-Plan	Austerity measures	
Articles	5	3	2	2	2	

Based on Table 4 and Table 5 above, we propose a list of modification to our text search criteria that will help to reduce the false positive and false negative error rates.

Alternative 1: “Beijing authorities”

{China OR Chinese} AND {economy OR economics} AND {uncertain OR uncertainty} AND {{{policy OR spending OR budget OR political OR "interest rates" OR "reform"} AND {government OR "Beijing authorities"}} OR tax OR regulation OR regulatory OR "central bank" OR "People's Bank of China" OR PBOC OR deficit}

Alternative 2: “Beijing authorities” + “WTO”

{China OR Chinese} AND {economy OR economics} AND {uncertain OR uncertainty} AND {{{policy OR spending OR budget OR political OR "interest rates" OR "reform"} AND {government OR "Beijing authorities"}} OR tax OR regulation OR regulatory OR "central bank" OR "People's Bank of China" OR PBOC OR deficit OR WTO}

Alternative 3: “Beijing OR authorities”

{China OR Chinese} AND {economy OR economics} AND {uncertain OR uncertainty} AND {{{policy OR spending OR budget OR political OR "interest rates" OR "reform"} AND {government OR Beijing OR authorities}} OR tax OR regulation OR regulatory OR "central bank" OR "People's Bank of China" OR PBOC OR deficit}

Alternative 4: “Beijing OR authorities” + “WTO”

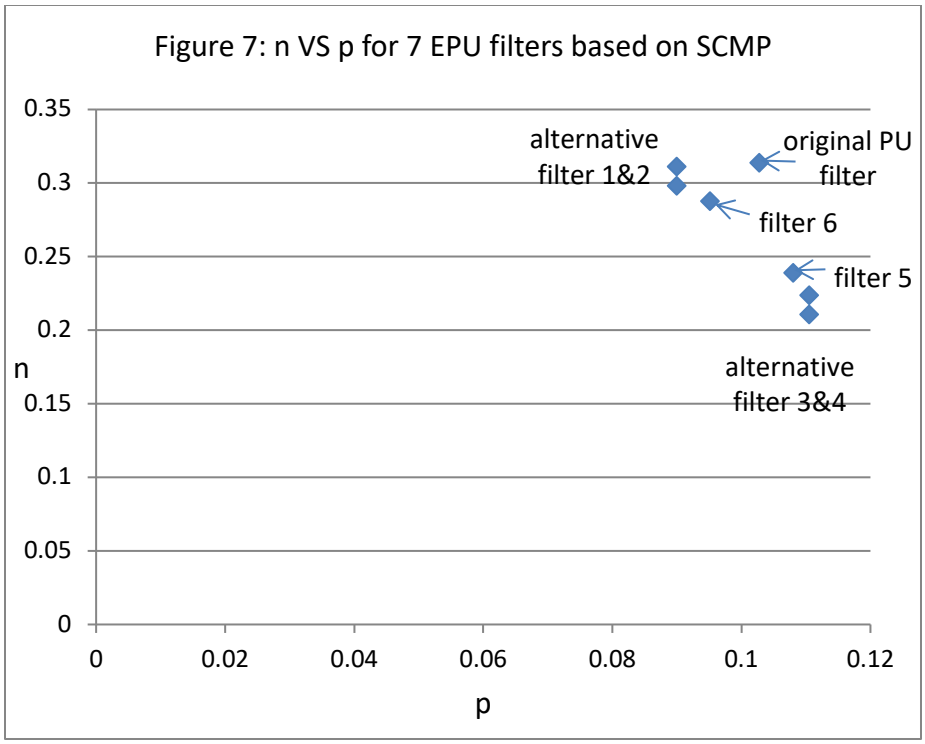
{China OR Chinese} AND {economy OR economics} AND {uncertain OR uncertainty} AND {{{policy OR spending OR budget OR political OR "interest rates" OR "reform"} AND {government OR Beijing OR authorities}} OR tax OR regulation OR regulatory OR "central bank" OR "People's Bank of China" OR PBOC OR deficit OR WTO}

Alternative 5: “Beijing”

{China OR Chinese} AND {economy OR economics} AND {uncertain OR uncertainty} AND {{{policy OR spending OR budget OR political OR "interest rates" OR "reform"} AND {government OR Beijing}} OR tax OR regulation OR regulatory OR "central bank" OR "People's Bank of China" OR PBOC OR deficit}

Alternative 6: “Beijing authorities” + “WTO” + Additional Policy-Related Terms

{China OR Chinese} AND {economy OR economics} AND {uncertain OR uncertainty} AND {{{policy OR spending OR budget OR political OR "interest rates" OR "reform" OR “tightening measures” OR “stimulus” OR “trade agreement” OR “Five-year Plan”} AND {government OR "Beijing authorities"}} OR tax OR regulation OR regulatory OR "central bank" OR "People's Bank of China" OR PBOC OR deficit OR WTO}



As shown from Figure 7, the alternative EPU filter 2, 4 and 6 emerge as winners (we demonstrate more detailed statistical results for these three filters and original EPU filter in Table 4 below).

Table 6: statistical summary of the original EPU filter and 6 alternative filters based on SCMP

	Filter0	Filter1	Filter2	Filter3	Filter4	Filter5	Filter6
corr(y,x)	0.82	0.74	0.75	0.82	0.82	0.81	0.75
corr(p,x)	-0.39	-0.48	-0.50	-0.58	-0.54	-0.59	-0.54
corr(n,x)	-0.26	0.01	-0.13	-0.00	-0.15	-0.01	-0.13
corr(p,n)	-0.04	-0.03	-0.17	-0.02	-0.04	-0.04	-0.15
corr(x,p-n)	-0.17	-0.37	-0.31	-0.50	-0.37	-0.50	-0.35
P	0.103	0.090	0.090	0.111	0.111	0.108	0.095
N	0.314	0.311	0.298	0.224	0.211	0.239	0.288
p + n	0.417	0.401	0.388	0.335	0.322	0.347	0.383
p - n	-0.211	-0.221	-0.208	-0.113	-0.100	-0.131	-0.193

From Table 6, we notice that both false positive rate and false negative rate are negatively correlated with true economic policy uncertainty. This implies that we are not systematically exaggerating the level of policy uncertainty in the periods we identify as EPU spikes.

Table 7a: SCMP auditing results for the original EPU filter

Time	Y	X	FP	FN	y	x	p	n
Full sample	305	389	40	122	2.4843	2.5900	0.1028	0.3136
1995-2003	156	202	21	65	2.1931	2.3054	0.1040	0.3218
2004-2012	149	187	19	57	2.1732	2.2718	0.1016	0.3048

Table 7b: SCMP auditing results for Filter 2 ("Beijing authorities" + "WTO")

Time	Y	X	FP	FN	y	x	p	n
Full sample	308	389	35	116	5.7301	5.9636	0.0900	0.2982
1995-2003	158	202	17	62	5.0626	5.3083	0.0842	0.3069
2004-2012	150	187	18	54	5.0106	5.2311	0.0963	0.2888

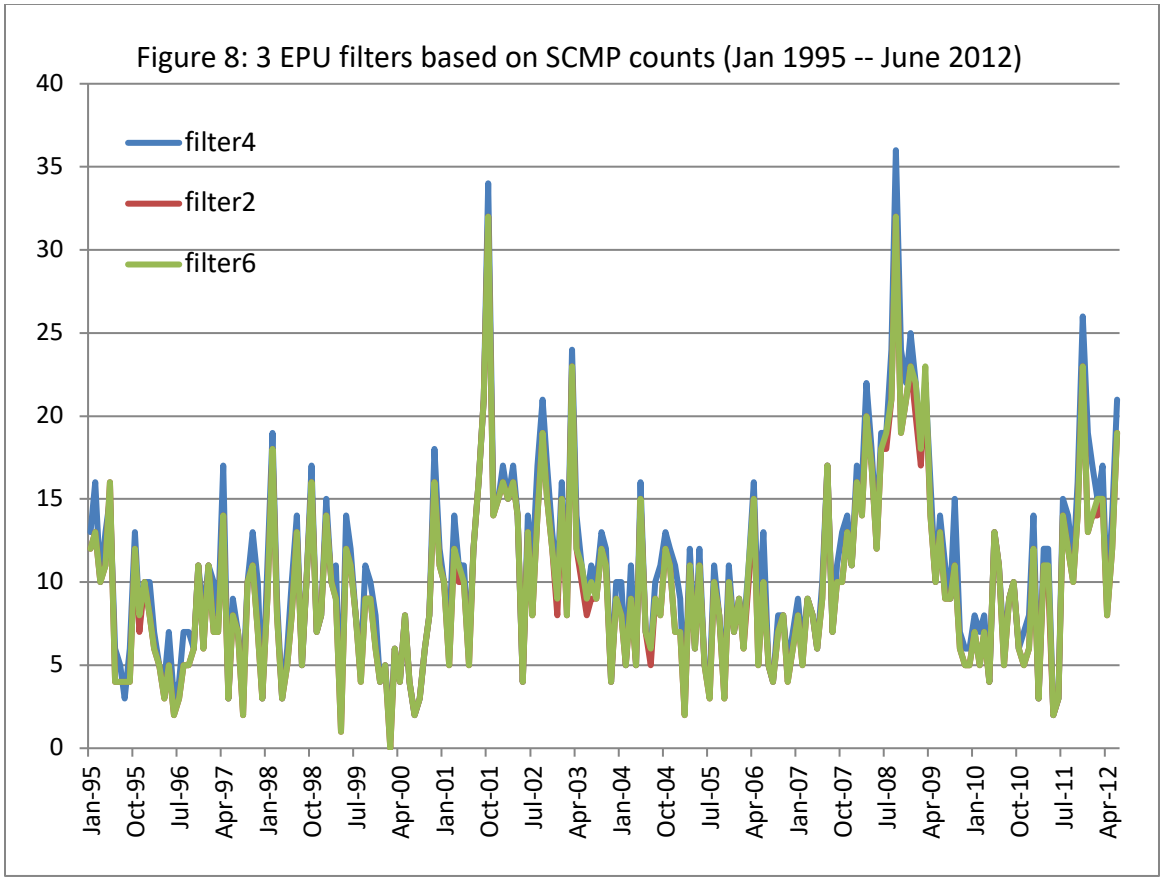
Table 7c: SCMP auditing results for Filter 4 ("Beijing OR authorities" + "WTO")

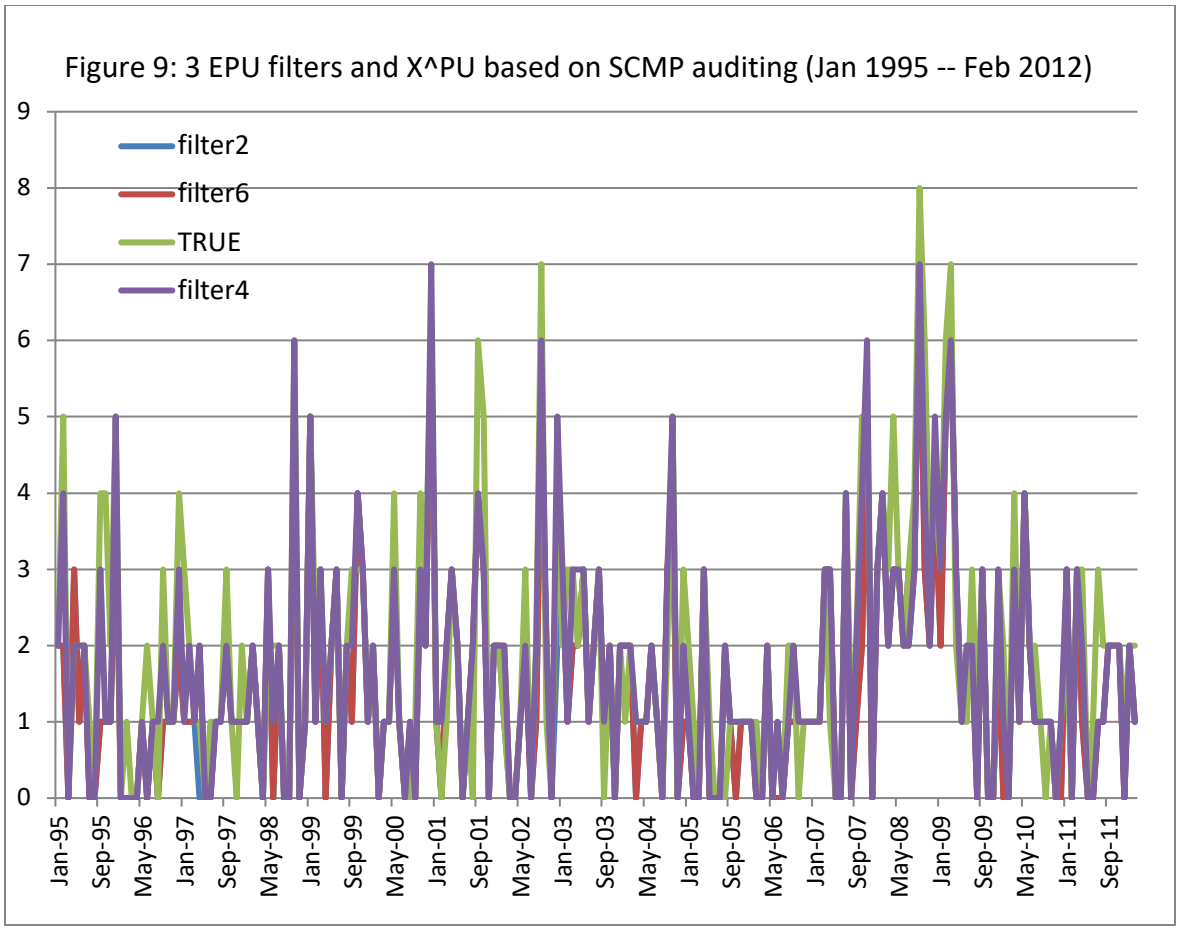
Time	Y	X	FP	FN	y	x	p	n
Full sample	350	389	43	82	5.8579	5.9636	0.1105	0.2108
1995-2003	183	202	24	43	5.2095	5.3083	0.1188	0.2129
2004-2012	167	187	19	39	5.1180	5.2311	0.1016	0.2086

Table 7d: SCMP auditing for Filter 6 ("Beijing authorities" + "WTO" + Additional Terms)

Time	Y	X	FP	FN	y	x	p	n
Full sample	314	389	37	112	5.7494	5.9636	0.0951	0.2879
1995-2003	163	202	20	59	5.0938	5.3083	0.0990	0.2921
2004-2012	151	187	17	53	5.0173	5.2311	0.0909	0.2834

In Figure 8, we generate EPU counts on SCMP articles from ProQuest database based on alternative filter 2, 4, and 6. Filter 4 seems to capture more spikes compared with other filters. Figure 9 demonstrates our auditing results of three alternative filters and true EPU, based on 500 articles of South China Morning Post.





Section V: Explore the Discrepancies of Different Data Sources Based on SCMP

Since both ProQuest and Factiva database have access to SCMP from Jan 2003, we compare the article counts from these two databases to explore the discrepancies of difference data sources. Four causes are identified for different counts. While our major problem can be solved through adjusting the filters, the remaining 3 causes (which are statistically less significant) seem to be harder to deal with.

First, the search algorithm works differently on the two sources. While Factiva searches terms strictly identical to the entries, ProQuest is more flexible. For example, “tax” also picks up closed related words like “taxes” and “taxation”. Second, Factiva database contains some articles that never get published on the SCMP. To reach to this conclusion, I first make a list of the articles shown in Factiva but not in ProQuest, then search for those articles in the SCMP website. None

of these articles are found in the SCMP website, so I suspect these articles were never published. Thirdly, ProQuest also covers abstract summary and reader's letter, etc. So the counts of ProQuest could be larger than those of Factiva sometimes. Finally, for the most recent months, ProQuest tends to neglect a few articles and add them into the database later. On the contrast, Factiva does not have this issue according to my investigation.

Thus, to minimize the discrepancies and to capture all the relevant articles, I propose to adjust for search terms in Factiva as the following:
For China economic uncertainty index, use the search term {(china OR Chinese) AND (economic OR economy) AND (uncertainty OR uncertain OR uncertainties)}.
For China economic policy uncertainty index, use the additional search term (policy OR policies OR spending OR regulation OR budget OR budgets OR deficit OR 'central bank' OR PBOC OR 'People's Bank of China' OR tax OR taxes OR taxation).

Table 8 below demonstrates the discrepancies of article counts after we adjust for search terms in the Factiva. Figure 7 also looks promising using refined search terms for Factiva.

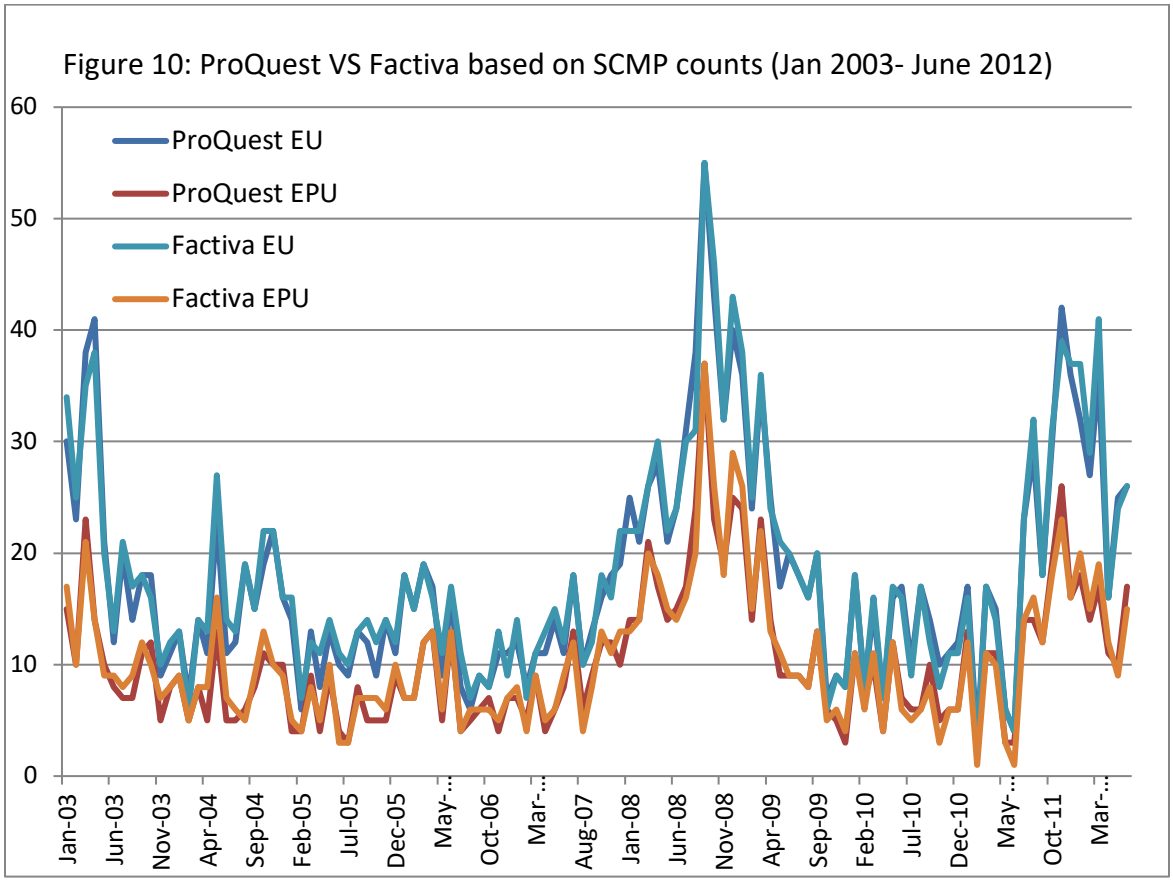
Table 8a: Discrepancies between Factiva and ProQuest based on SCMP (Jan 03–Dec 11)

	ProQuest	Factiva (old filter)	Factiva (new filter)
China EU	1891	1479	1933
China EPU	1097	746	1110

Table 8b: Discrepancies between Factiva and ProQuest based on SCMP (Jan 03–June 12)

	ProQuest	Factiva (old filter)	Factiva (new filter)
China EU	2053	1603	2106
China EPU	1184	797	1200

*Note 6: In Table 8, 'Factiva (old filter)' stands for the same search terms as those in the ProQuest, and 'Factiva (new filter)' stands for my proposed adjusting search terms.



Section VI: Weighted China Indexes Based on Newspapers from Countries in the East Asia

Moving forward, we plan to expand our news data sources and construct weighted China economic and economic policy uncertainty indexes based on newspapers from several countries in the East Asia area (including Japan, Taiwan and South Korea). We will discuss specific issues in the follow-up documents, such as using both English and Japanese news articles to deal with translation problem, and adding additional search terms to purify China-related articles in the Taiwan newspapers.

For example, in aggregating over the 3 newspapers in Taiwan and Hong Kong, we apply two principles: firstly, give the same weight to HK as to Taiwan; secondly, adjust for differences across papers in the average number of articles about economic uncertainty.

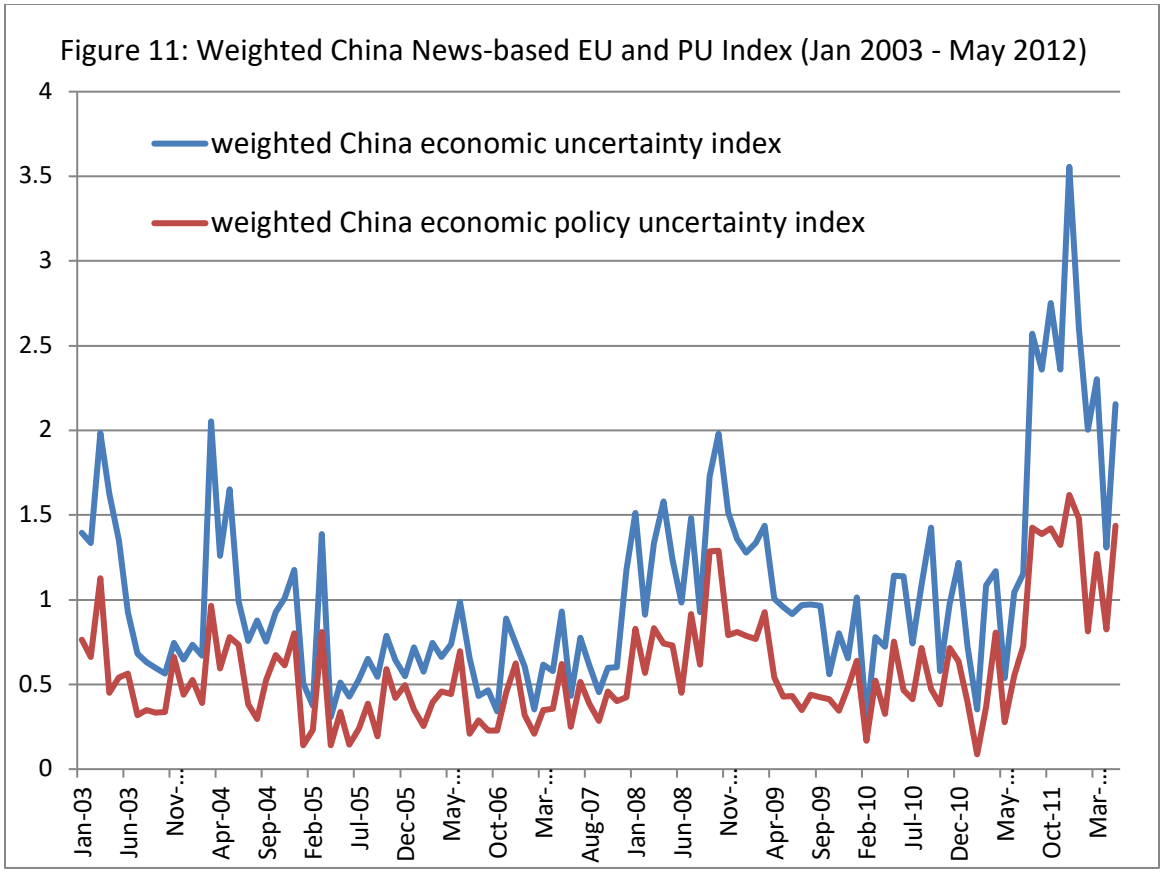
The formulas for weighted indexes are shown below:

(i) Compute the China EU index as

$$ChinaEU_t = (1/2)\frac{a_t^{EU}(HK)}{\overline{a}^{EU}(HK)} + (1/4)\sum_{j \in Taiwan}\frac{a_t^{EU}(j)}{\overline{a}^{EU}(j)}$$

(ii) Compute the China PU index as

$$ChinaPU_t = (1/2)\frac{a_t^{PU}(HK)}{\overline{a}^{EU}(HK)} + (1/4)\sum_{j \in Taiwan}\frac{a_t^{PU}(j)}{\overline{a}^{EU}(j)}$$



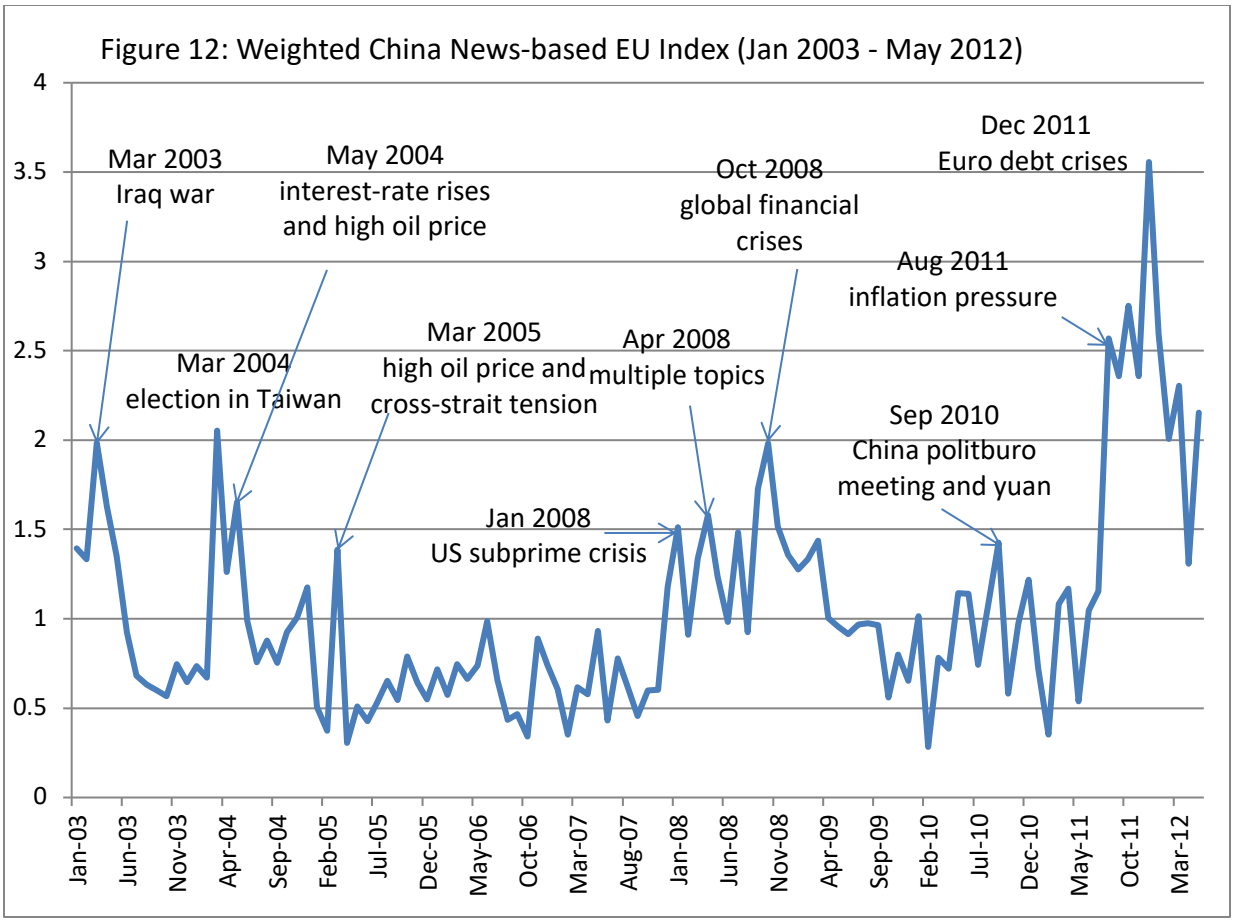


Figure 12: Weighted China News-based EU Index (Jan 2003 - May 2012)
4
3.5
3
2.5
2
1.5
1
0.5
0
Mar 2003
Iraq war
May 2004
interest-rate rises
and high oil price
Oct 2008
global financial
crises
Dec 2011
Euro debt crises
Aug 2011
inflation pressure
Mar 2004
election in Taiwan
Mar 2005
high oil price and
cross-strait tension
Apr 2008
multiple topics
Sep 2010
China politburo
meeting and yuan
Jan 2008
US subprime crisis
Jan-03
Jun-03
Nov-03
Apr-04
Sep-04
Feb-05
Jul-05
Dec-05
May-06
Oct-06
Mar-07
Aug-07
Jan-08
Jun-08
Nov-08
Apr-09
Sep-09
Feb-10
Jul-10
Dec-10
May-11
Oct-11
Mar-12

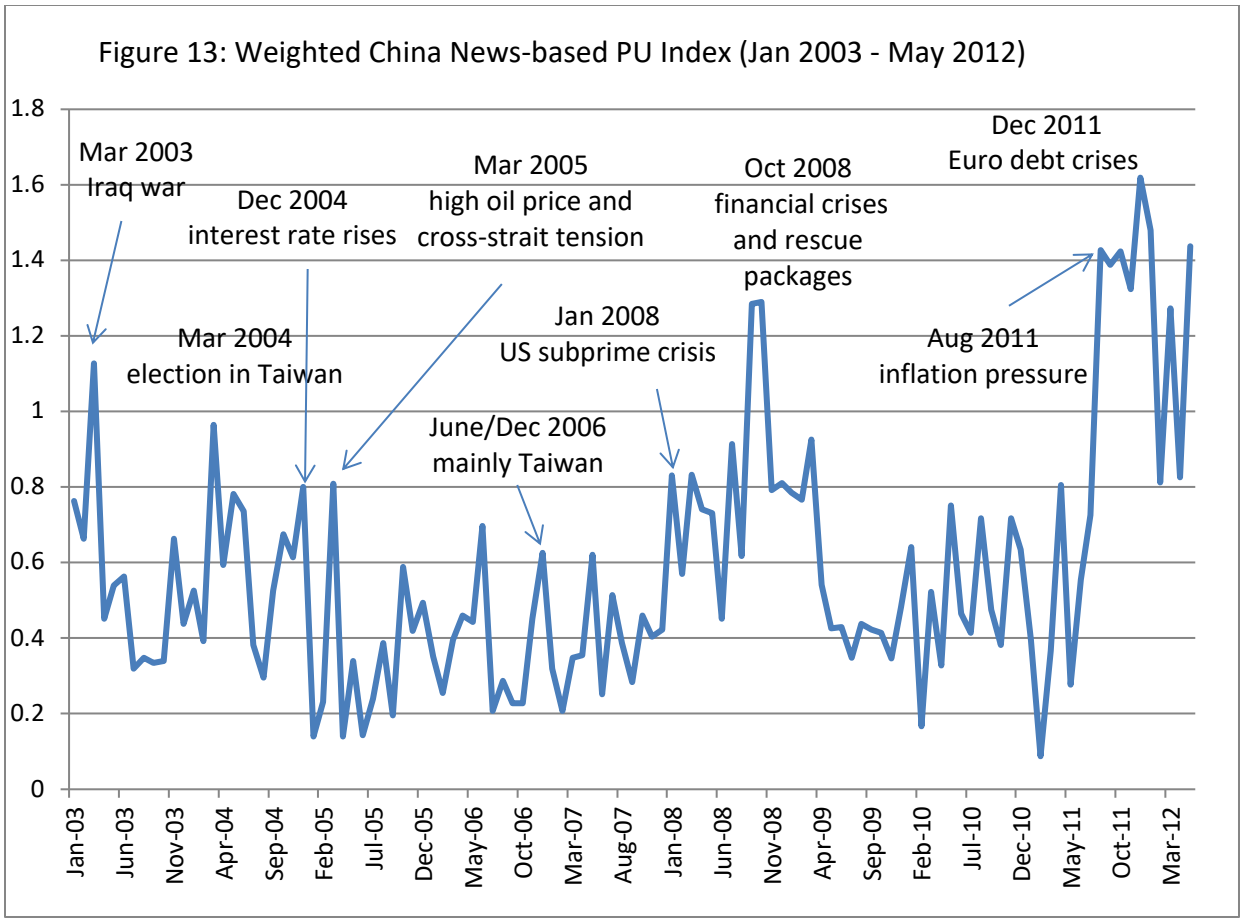


Figure 13: Weighted China News-based PU Index (Jan 2003 - May 2012)
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Mar 2003
Iraq war
Dec 2004
interest rate rises
Mar 2005
high oil price and
cross-strait tension
Oct 2008
financial crises
and rescue
packages
Dec 2011
Euro debt crises
Mar 2004
election in Taiwan
Jan 2008
US subprime crisis
Aug 2011
inflation pressure
June/Dec 2006
mainly Taiwan
Jan-03
Jun-03
Nov-03
Apr-04
Sep-04
Feb-05
Jul-05
Dec-05
May-06
Oct-06
Mar-07
Aug-07
Jan-08
Jun-08
Nov-08
Apr-09
Sep-09
Feb-10
Jul-10
Dec-10
May-11
Oct-11
Mar-12

References

Baker, Scott, Nicholas Bloom and Steven J. Davis, 2013. "Measuring Economic Policy Uncertainty," working paper.