ONLINE APPENDIX TO SECTORAL NEWS FOCUS AND AGGREGATE FLUCTUATIONS

RYAN CHAHROUR, KRISTOFFER P. NIMARK, AND STEFAN PITSCHNER

This appendix complements the analysis in our paper *Sectoral News Focus and Aggregate Fluctuations*. The appendix contains 8 sections and is organized as follows. Section 1 provides additional proofs from Section 3 of the paper. Section 2 contains the sector definitions used for both the macro data and the measures of news focus. Section 3 provides additional details on the macro data. Sections 4 to 7 provide additional details on the news data. Section 8 provides additional empirical evidence related to the mechanism of the model.

1. Additional Proofs

Proof of Proposition 3. The proposition states that for a given $r < n$, the variance of productivity shocks conditional on being reported $var\left(z_i \mid s_i = 1\right)$ is larger than the unconditional variance $var\left(z_i\right)$ and increasing in n.

Proof. We start by proving that $var(z_i \mid s_i = 1) > var(z_i)$. Define the variable $x_i \equiv z_i^2$. Since $E(z_i) = 0$, $E(x_i) = var(z_i)$. Denote the k^{th} order statistic of $\{x_1, x_2, ..., x_n\}$ as $x_{(k)}$ so that

$$x_{(1)} \equiv \min\left\{x_1, x_2, ..., x_n\right\} \tag{1.1}$$

$$x_{(2)} \equiv \min\left\{\left\{x_1, x_2, ..., x_n\right\} \setminus x_{(1)}\right\} \tag{1.2}$$

$$\vdots$$

$$x_{(k)} \equiv \min\left\{\left\{x_1, x_2, ..., x_n\right\} \setminus \left\{x_{(1)}, x_{(2)}, ..., x_{(k-1)}\right\}\right\} \tag{1.3}$$

Note that $s_i = 1$ implies that

$$x_i \in \left\{x_{(n)}, x_{(n-1)}, ..., x_{(n-r+1)}\right\}. \tag{1.4}$$

Since $x_{(k)} \geq x_{(k-j)}$ for any $j > 0$, $x_{(k)}$ first order dominates $x_{(k-j)}$, and hence

$$E\left(x_{(k)}\right) \geq E\left(x_{(k-j)}\right) \tag{1.5}$$

so that

$$var(z_i \mid s_i = 1) \geq var(z_j \mid s_j = 0). \tag{1.6}$$

Combining (1.6) with the fact that

$$var(z_i) = p(s_i = 1)var(z_i \mid s_i = 1) + p(s_i = 0)var(z_j \mid s_j = 0) \tag{1.7}$$

gives the desired result

$$var(z_i) = var(z_i \mid s_i = 1) - p(s_i = 0)\left[var(z_i \mid s_i = 1) - var(z_j \mid s_j = 0)\right] \tag{1.8}$$

$$\leq var(z_i \mid s_i = 1). \tag{1.9}$$

Date: July 8, 2021.

To prove the second part of the proposition, we also need to show that $var(z_i \mid s_i = 1)$ is increasing in n. Using the same notation as above, consider $n = l$, so that the squared values of the reported sectors is the set $\left\{x_{(l)}, x_{(l-1)}, ..., x_{(l-r+1)}\right\}$. Now, consider adding one dimension to the state so that $n = l + 1$. If $x_{l+1} > x_{(l-r+1)}$ the value of one element in the random vector $\left\{x_{(l)}, x_{(l-1)}, ..., x_{(l-r+1)}\right\}$ is replaced by a larger value. The elements in the vector $\left(x_{(l+1)}, x_{(l)}, ..., x_{(l-r+2)}\right)$ are then larger than or equal to the corresponding elements in the vector $\left(x_{(l)}, x_{(l-1)}, ..., x_{(l-r+1)}\right)$. The former vector thus first order dominates the latter vector, implying that $E\left(x_{(l+1)}, x_{(l)}, ..., x_{(l-r+2)}\right) > E\left(x_{(l)}, x_{(l-1)}, ..., x_{(l-r+1)}\right)$. The desired result then follows from the fact that the definition of x_i implies that $E(x_i) = var(z_i)$. □

Proof of Proposition 4. The proposition states that the conditional variance of unreported productivity shocks $var\left(z_j \mid \mathbf{s}, \mathbf{r}, s_j = 0\right)$ is increasing in the minimum value of the reported productivity shocks $\min\left\{|z_i| : s_i = 1\right\}$.

Proof. The news selection function $\mathcal{S}_{|z|}$ implies that

$$p\left(|Z_j| > \min\left\{|Z_i| : s_i = 1\right\} \mid s_j = 0\right) = 0. \tag{1.10}$$

The distribution $p\left(Z_j \mid \mathbf{r}, \mathbf{s}, s_j = 0\right)$ is therefore a truncated normal with density function

$$p\left(Z_j \mid \mathbf{r}, \mathbf{s}, s_j = 0\right) = \begin{cases} 0 \text{ if } Z_j < -\min\left\{|Z_i| : s_i = 1\right\} \\ \\ \frac{\phi(Z_j)}{\Phi(\min\{|Z_i|:s_i=1\}) - \Phi(-\min\{|Z_i|:s_i=1\})} \\ \\ 0 \text{ if } Z_j > \min\left\{|Z_i| : s_i = 1\right\} \end{cases}. \tag{1.11}$$

where ϕ and Φ are the pdf and cdf of the unconditional distribution of Z_j.

A zero mean symmetric two-sided truncated normal distribution with truncation points $-a$ and a is a mean preserving spread of any zero mean symmetric two-sided truncated normal distribution with truncation points $-b$ and b such that $a > b$. The proposition then follows from that a mean preserving spread increases the variance of a distribution. □

Proof of Proposition 5. The proposition states that the mean of reported productivity shocks $E\left(z_i \mid s_i = 1\right)$ is lower than the unconditional mean of productivity shocks and decreasing in the number of sectors n.

Proof. Denote the k^{th} order statistic of the vector z as $z_{(k)}$. Then $z_{(k)}$ state-wise dominates $z_{(k-j)}$ for any $j > 0$ so that $E\left(z_{(k)}\right) > E\left(z_{(k-j)}\right)$. To prove the first part of the proposition, note that the elements in $\mathbf{r}$ consists of the first r order statistics of z, so that

$$E\left(z_i \mid s_i = 1\right) < E\left(z_j \mid s_j = 0\right). \tag{1.12}$$

The result then follows from that

$$E\left(z_i\right) = p(s_i = 1)E\left(z_i \mid s_i = 1\right) + p(s_j = 0)E\left(z_j \mid s_j = 0\right) \tag{1.13}$$

$$= 0 \tag{1.14}$$

To prove the second part of the proposition, set the dimension of the state $n = l$. The reported sector shocks in the set $\left\{z_{(1)}, z_{(2)}, ..., z_{(r)}\right\}$ then consists of the first r order statistics of an l dimensional vector. Now, consider adding one dimension to the state so that $n = l+1$.

If $z_{l+1} < z_{(r)}$ the value of one element in the random vector $\left(z_{(1)}, z_{(2)}, ..., z_{(r)}\right)$ is replaced by a smaller value. If $z_{l+1} > z_{(r)}$ the vector is unchanged. The values of the first r order statistics drawn from l sectors thus first order stochastically dominates the first r order statistics drawn from $l+1$ sectors. The conditional mean $E\left(z_i \mid s_i = 1\right)$ is thus decreasing in n. □

Proof of Proposition 6. The proposition states that the expected value of non-reported productivity shocks $E\left(z_j \mid \mathbf{s}, \mathbf{r}, s_j = 0\right)$ is increasing in the maximum value of the reported productivity shocks $\max\left\{z_i : s_i = 1\right\}$.

Proof. The news selection function $\mathcal{S}_{-}$ implies that

$$p\left(Z_j < \max\left\{Z_i : s_i = 1\right\} \mid s_j = 0\right) = 0. \tag{1.15}$$

The conditional distribution of a non-reported sector shocks is thus normal but left-truncated at $\max\left\{Z_i : s_i = 1\right\}$ with expected value given by

$$E\left(Z_j \mid s_j = 0, \max\left\{Z_i : s_i = 1\right\}\right) = \frac{\phi\left(\max\left\{Z_i : s_i = 1\right\}\right)}{1 - \Phi\left(\max\left\{Z_i : s_i = 1\right\}\right)} \tag{1.16}$$

which is increasing in $\max\left\{Z_i : s_i = 1\right\}$. □

2. Sector Definitions

Sectoral data is aggregated from the BEA *summary level* sectors to the sector definitions according to Table A1 below.

Table A1. Sector Definitions

Sector	Sector Name	BEA Codes	NAICS Codes
1	agriculture & forestry	111CA, 113FF	111-115
2	mining	212-213	212-213
3	oil & gas extraction	211	211
4	construction	23	23
5	food & kindred products	311-312	311-312
6	textile mill products	313-314	313-314
7	apparel & leather	315-316	315-316
8	lumber	321	321
9	furniture & fixtures	337	337
10	paper & allied products	322	322
11	printing & publishing	323, 511	323, 511
12	chemicals	325	325
13	petroleum refining	324	324
14	rubber & plastics	326	326
15	non-metallic minerals	327	327
16	primary metals	331	331
17	fabric. metal products	332	332
18	non-electrical machinery	333	333
19	electrical machinery	335	335
20	motor vehicles	3361-3363	3361-3363
21	other transportation equipment	3364-3369	3364-3369
22	instruments	334	334
23	misc. manufacturing	339	339
24	transportation & warehousing	48-49	48-49
25	communications	512-514	512-519
26	electric & gas utilities	22	22
27	wholesale & retail	42, 44-45, 4A0 (42, 44RT)	42, 44-45
28	F.I.R.E.	52-53, OR (52, 531, 532RL)	52-53
29	other services	54-56, 61-62, 71-72, 81	54-56, 61-62, 71-72, 81

Notes: BEA codes from BEA summary level make-use tables. When different, BEA KLEMS sectors used for computing sector-level productivity are listed in parenthesis. NAICS codes used to classify news coverage.

3. Construction of Input/Output Tables and Macro Time Series

The input-output structure for the calibrated economy is derived from the BEA input-output use table *after redefinition* from the 2007 benchmark tables. These data provide nominal input use by each BEA sectors of the output from each BEA sector. "Redefinition" refers to the process in which the BEA reallocates a secondary output from one sector (e.g. restaurants located within hotels) to the industry in which it the primary output (e.g. dining). Input uses are then aggregated across each of our sector definitions. The use tables provide domestic *absorption* of intermediate goods. Because the US is not actually a closed economy, absorption can exceed production (as it does, for example, in the oil and gas extraction industry.) To generate an internally consistent set of input-output tables, we therefore treat all domestic absorption as if it were the produced within the US.

Let M_{ij} be the intermediate use of good j in the production of sector i, and A_i be the total domestic absorption (output less net-exports) of good i, all measured in producer prices. The parameters α_{ij} are then computed as

$$\alpha_{ij} = \frac{M_{ij}}{A_i}.$$

To find the final output shares, β_i, we then compute usage in final output as $F_i = A_i - \sum_j M_{ji}$ and calibrate

$$\beta_i = \frac{F_i}{\sum_i F_i}.$$

Finally, sectoral productivity series are computed from the BLS-BEA multi-factor productivity dataset, using a weighted average of the productivity index in each sub-sector in Table 1 above, with time-varying weights equal to the sub-sector's share in nominal gross output in that sector.

4. News Data: Matched and Unmatched Entity Tags

Sectoral news focus is measured as described in Section 5 and based on Factiva entity tags. The figure below shows the most frequent entity tags for three different groups. The first panel shows the most frequent entity tags that we classify as US companies and that we use to construct our measures of sectoral media focus. The second panel shows the most frequent entity tags that we match to a sector, but that are not classified as US companies and therefore excluded from our measures of news focus. The third panel shows the most frequent entity tags that are not matched to one of our sectors and thus also excluded from our analysis.

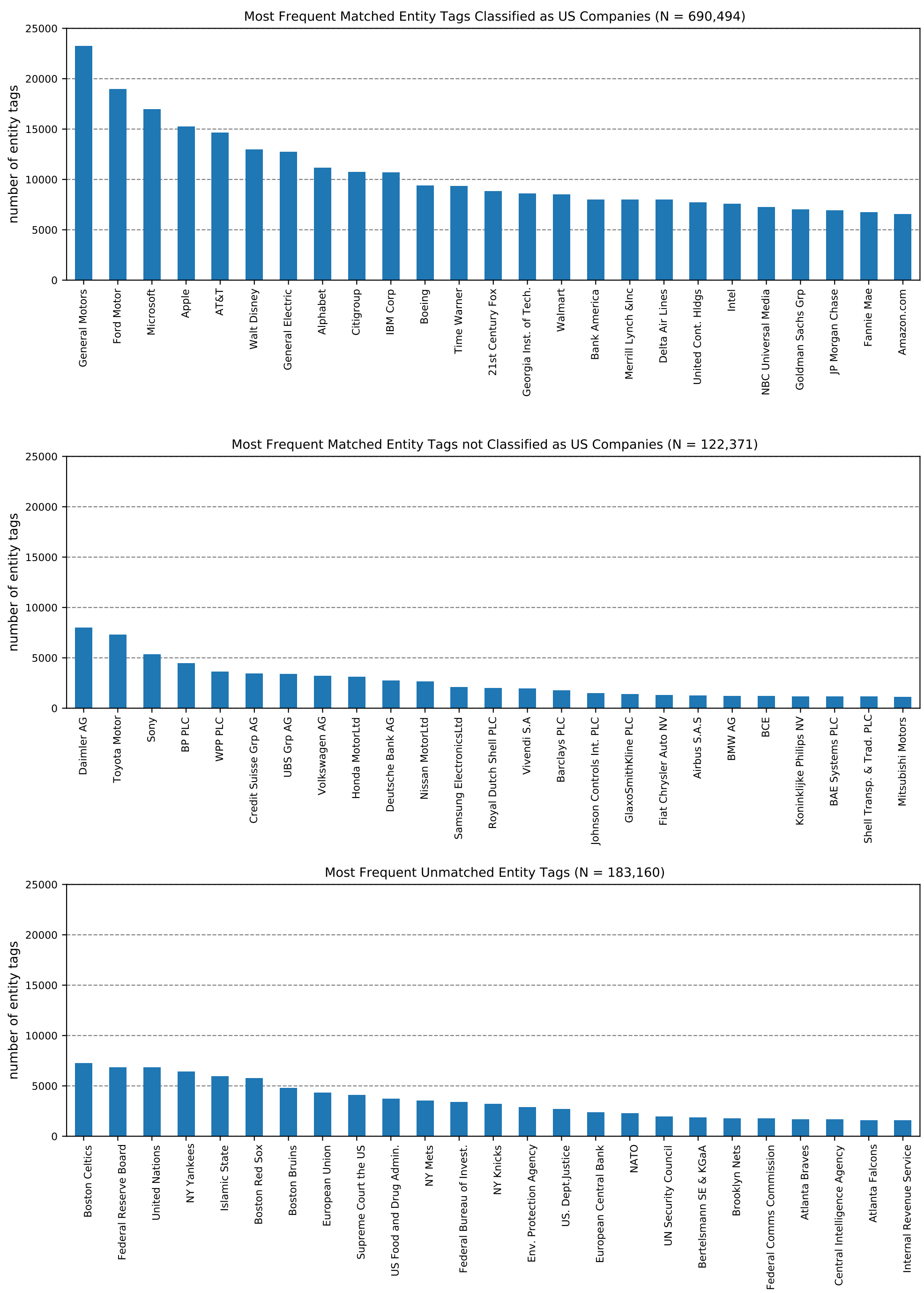

Figure A1. Most Frequent Matched and Unmatched Entity Tags.

5. News Data: Sector Constituents

The figures below show (up to) the 20 most-frequent constituents of each of the 29 sectors. Some company names have been abbreviated.

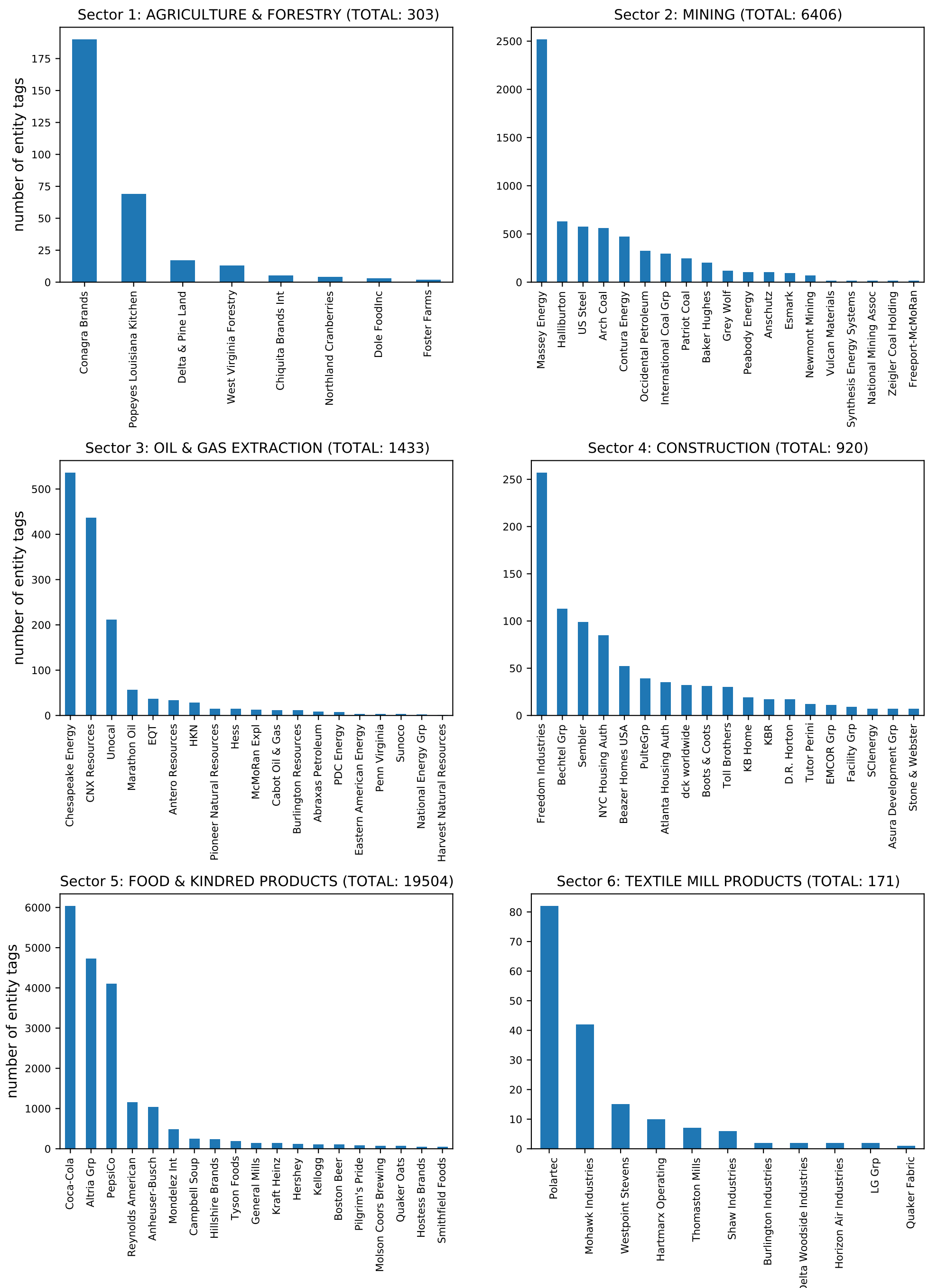

Figure A2. Most Frequent Sector Constituents: Sectors 1-6.

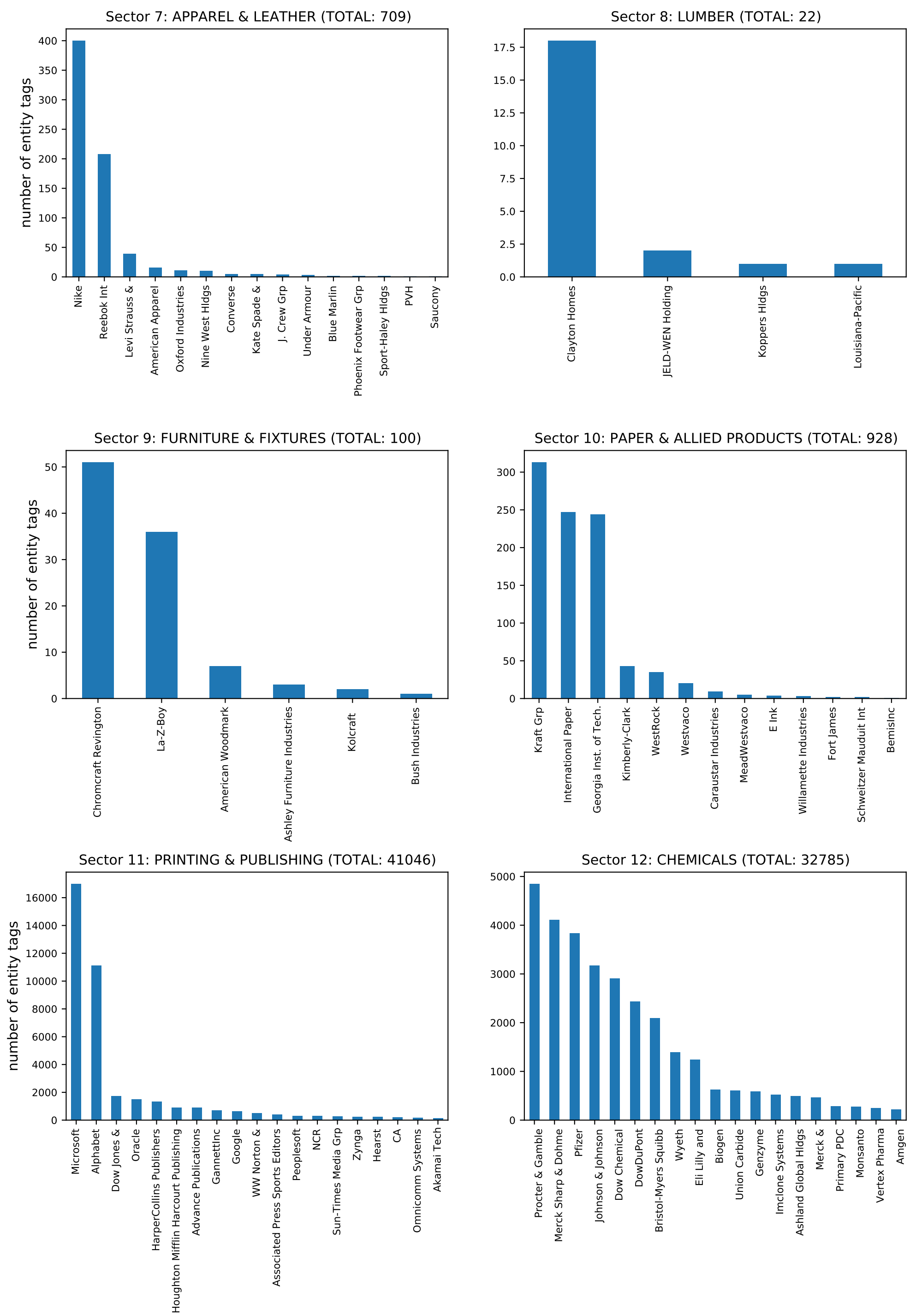

FIGURE A3. Most Frequent Sector Constituents: Sectors 7-12.

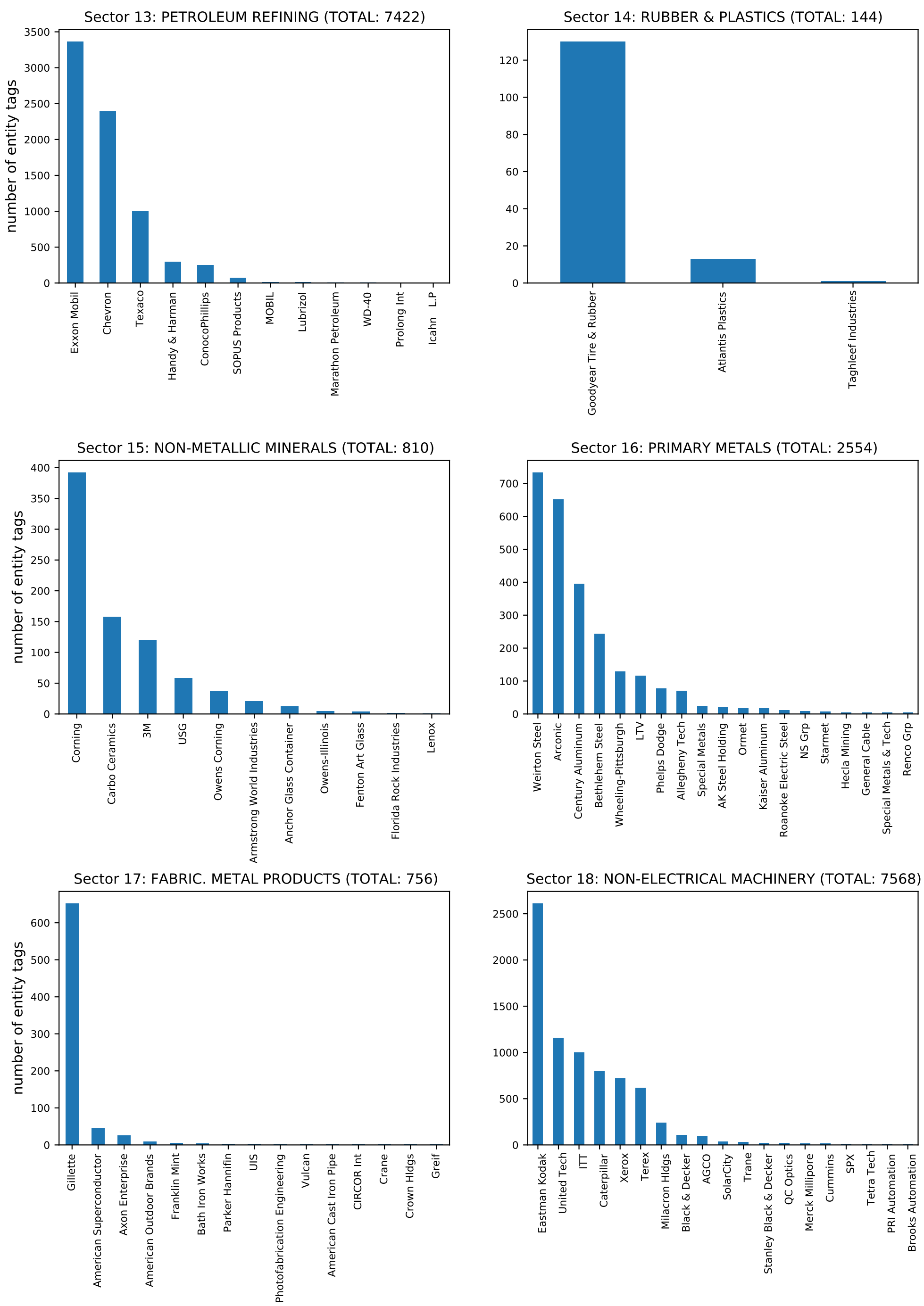

Figure A4. Most Frequent Sector Constituents: Sectors 13-18.

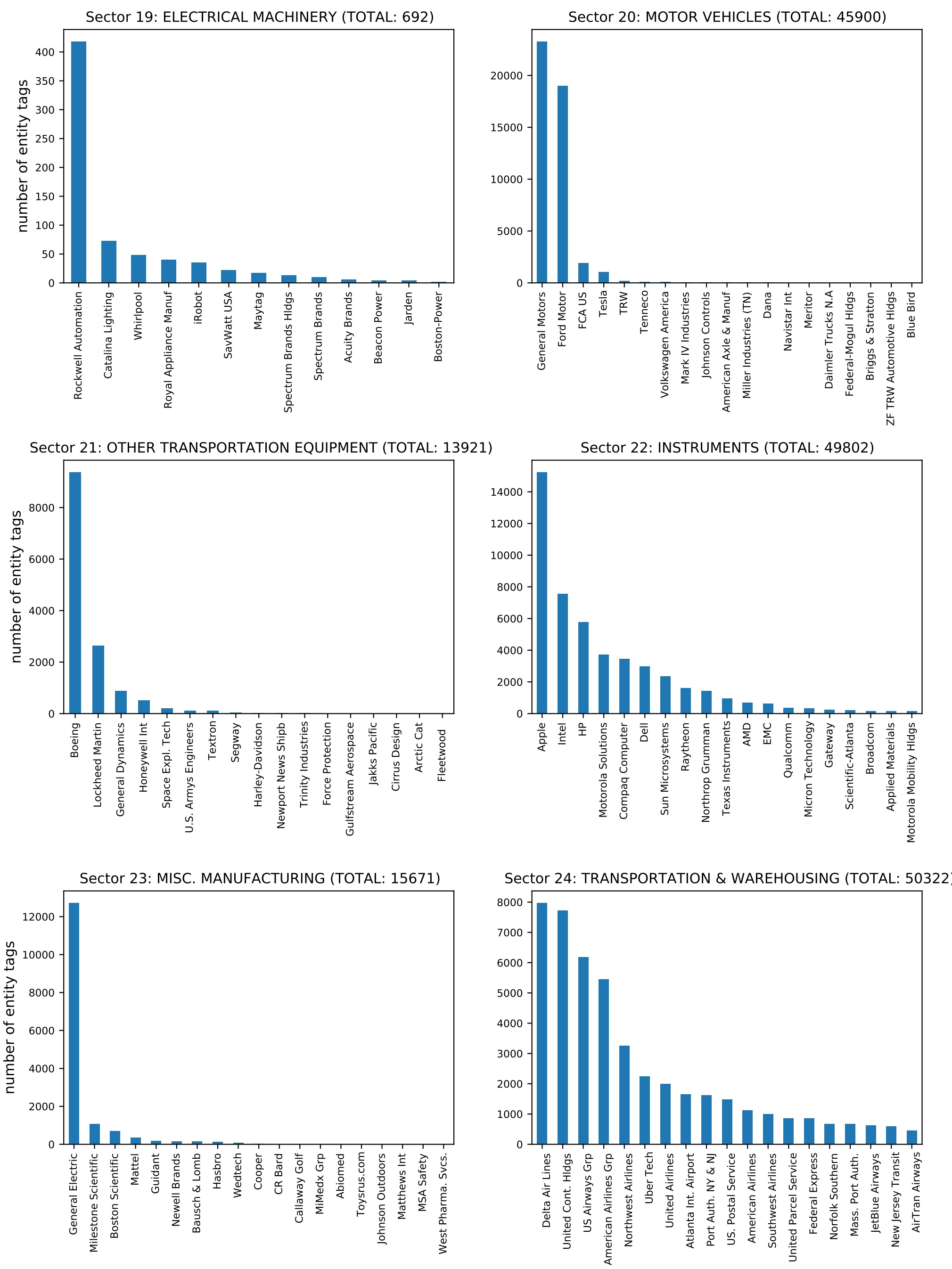

Figure A5. Most Frequent Sector Constituents: Sectors 19-24.

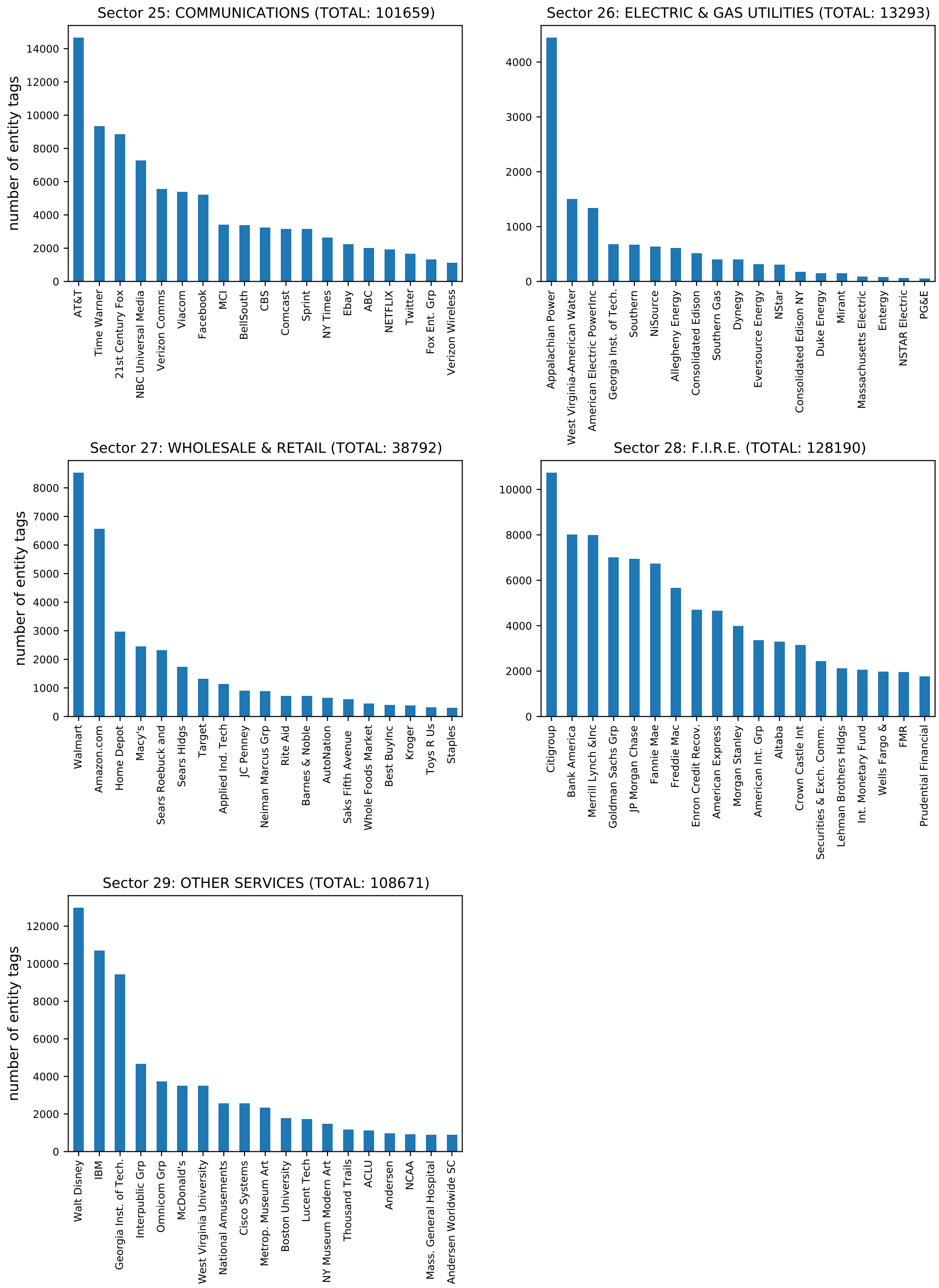

Figure A6. Most Frequent Sector Constituents: Sectors 25-29.

6. News Data: Sectoral News Focus Over Time For All 29 Sectors

The figures below show the measures of sectoral news focus for each of the 29 sectors. The time series are defined as in Section 3.

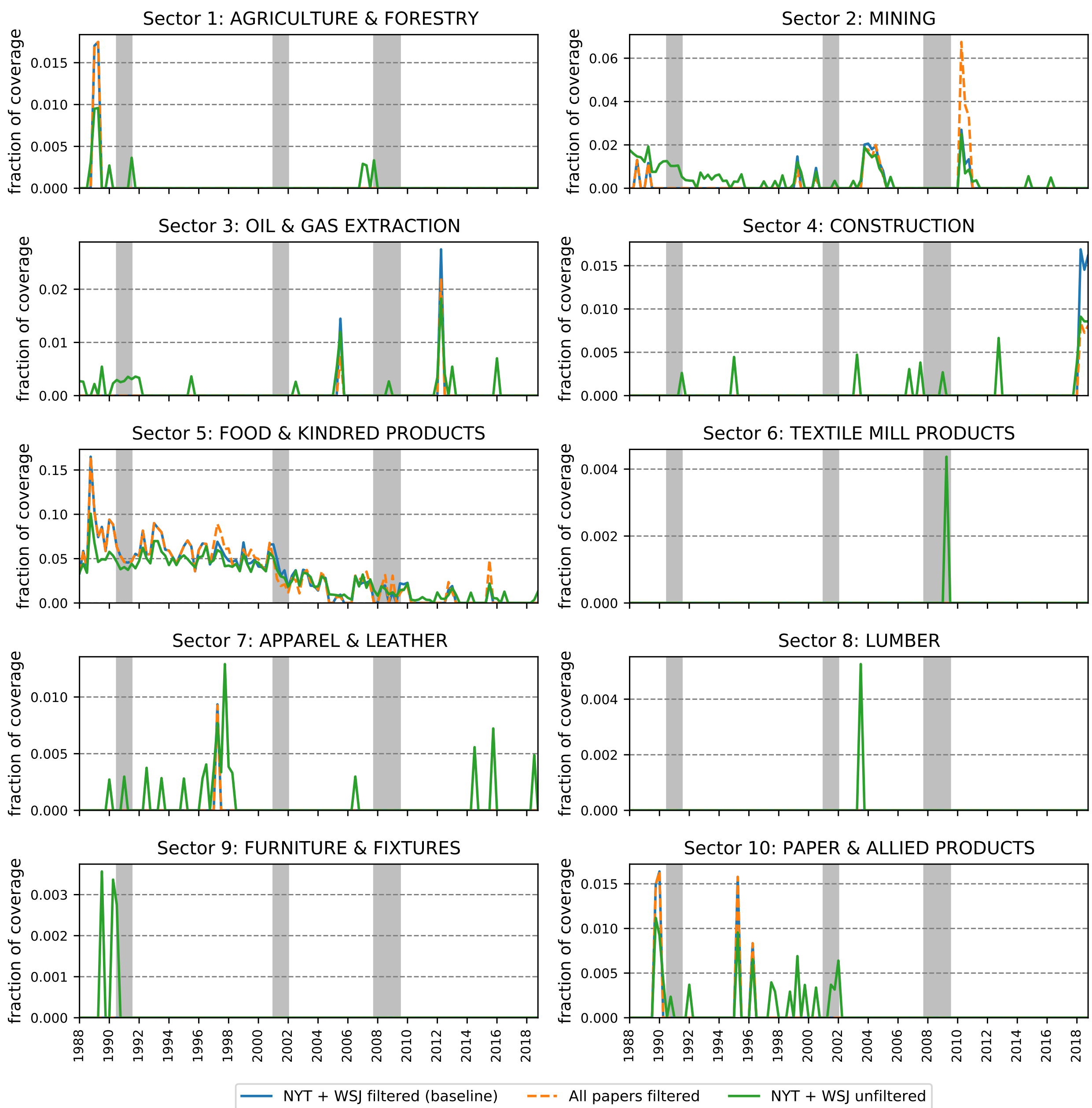

Figure A7. Sectoral News Focus: Sectors 1-10.

FIGURE A8. Sectoral News Focus: Sectors 11-20.

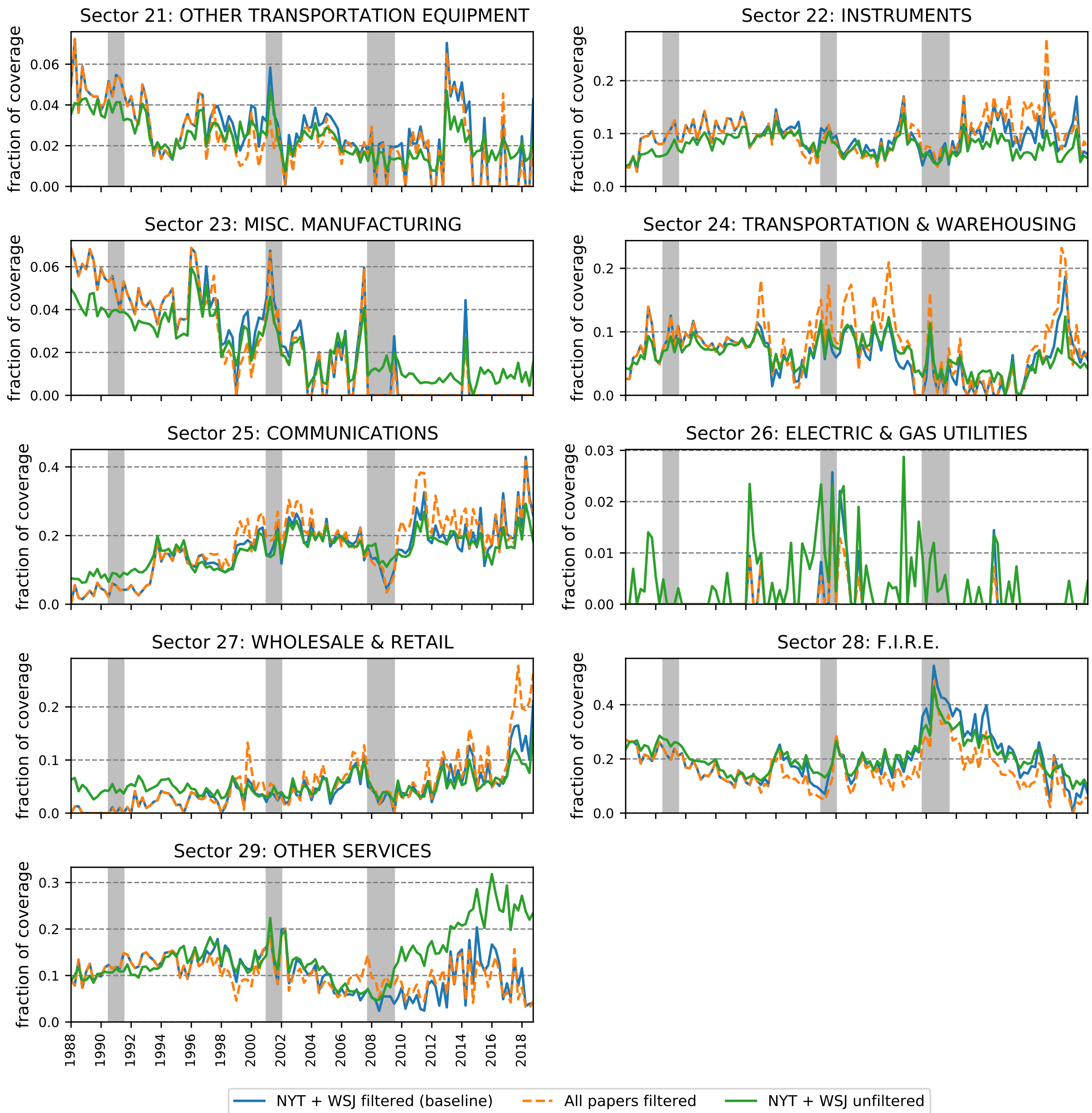

FIGURE A9. Sectoral News Focus: Sectors 20-29.

7. News Data: Search terms used for ngram-based industry news measures

Section 5 contains alternative news coverage measures based on n-grams that occur in news articles and refer directly to specific industries or sectors. The search terms used for these news measures are shown here. Each search term is also used with the word "industry" instead of "sector".

Motor vehicles: auto sector, auto industry, automotive sector, automotive industry, car sector, car industry

F.I.R.E.: financial sector, real estate sector, audit sector, subprime sector, securities sector, mutual fund sector, insurance industry, banking sector, credit card sector, housing sector, mortgage sector, finance sector, accounting sector

Food and kindred products: cattle sector, tobacco sector, poultry sector, citrus sector, tequila sector, organic food sector, ecigarette sector, meat sector, beer sector, fishing sector, liquor sector, seafood sector, food sector, pistachio sector, agriculture sector, grocery sector, beef sector

8. Additional empirical evidence

This appendix describes in more detail the additional empirical evidence referred to at the end of Section 7.

News coverage and sectoral correlations

If the mechanism in the model is relevant in reality, we would expect that output in sectors that are over-represented in the news are more strongly correlated with aggregate output relative to less-reported sectors of the same size. We therefore first run a regression of average sectoral news coverage on Domar weights

$$\frac{1}{T}\sum_{t=1}^{T} f_{i,t} = \gamma_f + \beta_f \lambda_i + \varepsilon_i^f \tag{8.1}$$

where $f_{i,t}$ denotes the fraction of news coverage received by sector i in period t. A positive residual ε_i^f implies that the sector is over-represented in the news relative to its economic size. (The same information is contained in Figure 6.) We then run the regression

$$\rho_{i,y} = \gamma_\rho + \beta_\rho \lambda_i + \varepsilon_i^\rho \tag{8.2}$$

where $\rho_{i,y}$ denotes the correlation between gross output growth in sector i and aggregate output growth. A positive residual ε_i^ρ indicates that sector i is more strongly correlated with aggregate output than what would be implied by its economic size alone.

The correlation of the residuals from the two regressions is 0.21, suggesting that sectors that are over-represented in the news are indeed also more strongly correlated with aggregate output, as predicted by our model.

News coverage-weighted productivity and aggregate output

Our model predicts that productivity in a given sector has a bigger impact on aggregate output when the sector in question is in the news. We therefore compute two news-weighted aggregate productivity series as follows

$$z_t^f \equiv \sum_{i=1}^{n} f_{i,t} z_{i,t}, \quad z_t^{f,\lambda} \equiv \sum_{i=1}^{n} \left(f_{i,t} - \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i} \right) z_{i,t}. \tag{8.3}$$

The measure z_t^f simply weighs sectoral productivity in period t by the fraction of news coverage a sector received in that period. The correlation between z_t^f and aggregate output growth is 0.46. Given the strong correlation between news coverage and the size of a sector, this positive correlation is unsurprising. The second measure, $z_t^{f,\lambda}$ therefore weighs sectoral productivity by the fraction of news coverage in period t that is not simply a reflection of the sector's size. The correlation between the second measure $z_t^{f,\lambda}$ and aggregate output growth is 0.26, again consistent with the idea that a sector has a bigger impact on aggregate output when it is widely reported on by news media.

In the analysis above, we argued that time-varying media focus could generate output fluctuations that a linear filter would attribute to common non-productivity shocks. We therefore also compute the correlation between the weighted productivity measure $z_t^{f,\lambda}$ and the time series of common shocks from the Atalay (2017) filter. The correlation between $z_t^{f,\lambda}$ and the common shocks produced by Atalay's high elasticity calibration is 0.32. The high-elasticity calibration attributes more of the variance to common shocks than Atalay's low elasticity calibration. The correlation of $z_t^{f,\lambda}$ with the common shocks extracted using the latter calibration is also positive, but substantially lower at 0.08.

References

[1] Atalay, E., 2017. How important are sectoral shocks?. American Economic Journal: Macroeconomics, 9(4), pp.254-80.