Text as Instruments

James Nesbit

July 17, 2020

Motivation

Body of research uses text to infer latent information

Implicit assumption: unobservables drive both text and data

- Structural shocks drive macro vars and policy documents
- News shocks drive equity returns and media coverage

Link used informally to infer responses to unobservables

- Narrative i.d. of exogenous monetary / fiscal policy
- Relationship between WSJ and stock market activity

This Paper

Model of text formalizing explicit link between text and data

- Word choice depends on exogenous unobservables

Use text model to:

1. Provide framework to think carefully about text generation—endogeneity
2. Guide construction of valid instruments from text
3. Perform inference with 'optimal' text IVs
4. Empirical application
    - Identify monetary policy in SVAR
    - News shocks and equity prices

Motivating Model

Linear IV (general model is GMM):

$$y_t = x_t'\beta_0 + Y_t\theta_0 + e_t, \quad \mathbb{E}[x_t' e_t] = 0, \quad \mathbb{E}[Y_t e_t] \neq 0$$

Assume exists unobservable, exogenous variable z_t^*

- Can used as an instrument for Y_t

Text Model:

T documents; N words; V terms

Context c: event—previous word(s) / term does (not) appear

Utility of term w, in document t, at word n in context c

$$U_{w,c,t,n} = X_t'\gamma + \alpha_{1,w,c} z_t^* + \alpha_{2,w,c} e_t + \epsilon_{w,c,t,n}$$

Why is Word Choice a Function of Unobservables?

$$U_{w,c,t,n} = X_t'\gamma + \alpha_{1,w,c} \underbrace{z_t^*}_{\text{target}} + \alpha_{2,w,c} \underbrace{e_t}_{\text{nuisance}} + \epsilon_{w,c,t,n}$$

e_t in returns—e_t contains past / stale news shocks

e_t in utility—news articles may be about stale news

Assume $\epsilon_{w,c,t,n} \sim$ i.i.d. Gumbel

$$p_t(w|c) = \frac{\exp\{X_t'\gamma + \alpha_{1,w,c}z_t^* + \alpha_{2,w,c}e_t\}}{\sum_{w'=1}^{V} \exp\{X_t'\gamma + \alpha_{1,w',c}z_t^* + \alpha_{2,w',c}e_t\}}$$

Why Context?

Use unconditional frequency of **‘shortfall’** as instrument

High freq of ‘**shortfall**’ when contemporaneous news shock

- Relevance ✓

High freq of ‘**shortfall**’ when stale news shock

- Exogeneity ✗

For example: Record high Apple share prices in 2020:

> *“If you look at our results, our **shortfall** is over 100 percent from iPhone and it’s primarily in greater China,” Apple CEO Tim Cook told CNBC’s Josh Lipton in an interview in **January 2019**.*

Assumption for Instruments

Assumption (Instruments)

Let $\mathcal{J}$ be collection of terms s.t. $|\mathcal{J}^*| > 2$ and

1. $\alpha_{1,w_j,c^*} \neq 0$ for at least one $j \in \mathcal{J}^*$
2. $\alpha_{2,w_j,c^*} = 0$ for all $j \in \mathcal{J}^*$

1. At least one w_j in context c^* is function of z_t^*
2. All w_j in context c^* are **not** functions of e_t

Consider

- w^* = 'shortfall' / 'downgrade' / 'tumble'
- 'yesterday', 'last week',' day before yesterday', 'in a row', three prior weekdays **not** in article

What are Instruments?

For $j \in \mathcal{J}$,

$$p_t(w_j|c^*) = \frac{\exp\{X_t'\gamma + \alpha_{1,w_j,c} z_t^*\}}{\sum_{w'=1}^{V} \exp\{X_t'\gamma + \alpha_{1,w',c} z_t^* + \alpha_{2,w',c} e_t\}}$$

Proposition

For all $j \in \mathcal{J}^*$, $p_t(w_j|c^*)$ are not instruments

Generalized log odds

$$z_t(\mathcal{J}, c, \boldsymbol{\omega}) := \sum_{j \in \mathcal{J}} \omega_j \log(p_t(w_j, c))$$

Proposition

Any $\boldsymbol{\omega}$ s.t. $\sum_{j \in \mathcal{J}} \omega_j = 0, \quad z_t(\mathcal{J}^*, \boldsymbol{\omega}, c^*)$ is an instrument for Y_t

Estimation

Don't observe z_t, only observe estimate $\hat{z}_t$. Two asym regimes:

1. *Relative* documents $T/N \to 1$—news articles / tweets
2. *Large* documents $T\log(T)/N \to 0$—policy documents

$$g_T(\theta, \hat{z}_t) = g_T(\theta, z_t) + \frac{1}{T}\sum_{t=1}^{T} G_{z_t}(\hat{z}_t - z_t) \qquad (\bigstar)$$

Consistency under both regimes

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^{T} G_{z_t}(\hat{z}_t - z_t)\right\| \\
&\quad \leq \sup_{t}\|\hat{z}_t - z_t\| \sup_{\theta\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^{T} G_{z_t}\right\| = O_p\left(\sqrt{\frac{\log(T)}{N}}\right) O_p(1)
\end{aligned}$$

Asymptotic Normality

$$g_T(\theta, \hat{z}_t) = g_T(\theta, z_t) + \frac{1}{T}\sum_{t=1}^{T} G_{z_t}(\hat{z}_t - z_t) \qquad (\bigstar)$$

Relative documents, $T/N \to 1$

$$avar(\hat{\theta}) = H\mathbb{E}\left[(g_T(\theta_0, z_t) + G_z(\hat{z}_t - z_t))(g_T(\theta_0, z_t) + G_z(\hat{z}_t - z_t))'\right] H'$$

- Generated regressors problem

Large documents, $T\log(T)/N \to 0$

- Same as observed z_t

$$\sqrt{T}\sup_t \|\hat{z}_t - z_t\| \left\| \frac{1}{T}\sum_{t=1}^{T} G_{z_t} \right\| = O_p\left(\sqrt{\frac{T\log(T)}{N}}\right) O_p(1)$$

Optimal Instruments in Class of Generalized Log Odds

Under relative documents asymptotics, $T/N \to 1$

- Consistently estimate avar for each $\boldsymbol{\omega} \implies$ optimize

In linear IV, can get closed form solutions

$$g_T(\theta_0, z_t) + G_z(\hat{z}_t - z_t) = z_t e_t + e_t(\hat{z}_t - z_t)) = \hat{z}_t e_t$$
$$avar(\hat{\theta}) = H\mathbb{E}[\hat{z}_t \hat{z}_t' e_t^2]H'$$

Weight on a term w_i depends:

- Negatively on term contribution to variance—$1/p_t(w_i)$
- Negatively on term contribution to mean—$\log p_t(w_i)$
- Positively on 'relevance' of term—α_{w_i}

Empirical Application - News Shocks

News articles from Reuters 2006–2013

- 86 companies
- $25,852$ articles

Sentiment = log of sum of positive and negative terms

50 positive terms

- ‘repurchase’, ‘surpass’, ‘upgrade’, ‘undervalue’, ‘surge’, ‘customary’, ‘jump’, ‘declare’, ‘rally’, ‘discretion’, ‘beat’

50 negative terms

- ‘shortfall’, ‘downgrade’, ‘disappointing’, ‘tumble’, ‘blame’, ‘hurt’, ‘auditor’, ‘plunge’, ‘slowdown’

Instruments

Terms

- 50 positive sentiment words
- 50 negative sentiment words

Context

- 'yesterday', 'last week',' day before yesterday', 'in a row', three prior weekdays not mentioned
- 63% articles satisfy

Market return is included as covariate

Results

	OLS	Unitary	Optimal
Sentiment	0.0020	0.0042	0.0034
		(0.0007)	(0.0005)

Unitary weights put 1 on positive terms, -1 on negative terms

1% increase in sentiment causes 0.0034 higher return

Conclusion

Building fully structural model of text data is difficult

Motivated by recent NLP research, work with parametric assumption on word choice given context

Give sufficient conditions to use generalized log odds as optimal instrument

Perform inference with under 2 document size regimes

Empirical application of news shocks on returns