Opening the "Black Box" of Merger Deals: A Topic Modelling Approach

Sudipto Dasgupta, Fangyuan Ma, Daisy Wang, and Haojun Xie*

Abstract:

Merger deals are often a "black box" for the empirical researcher – apart from standard information provided by established data sources on acquirer and target and the financial terms of the deal, very little is known about the unique features of a deal – why they are done, what are sources of value, what are the key concerns regarding deal success. In this paper, using machine learning techniques, we attempt to provide more granular description of deals. Specifically, we analyze a large set of textual data from M&A conference call transcripts and use topic modeling techniques to uncover the thematic structures. By decomposing call contents into latent topics, we infer the issues of relevance for the deals from the topic distributions, and further examine the value implications of each topic. Our study provides a machine learning solution to decrypting the information exchange between firms and the analysts who participate in these conference calls regarding the M&A deals.

Motivation and Background:

Merger and acquisitions, one of the most important corporate decisions, have significant value consequences for both acquirer and target. M&A deal announcements are very often associated with dramatic market reactions (e.g. Betton, Eckbo, and Thorburn, 2008, and references therein). Which aspects of the deals has the market been react to? Which aspects of the deal does the market like, and which does it not like? Although the finance literature has been engaged in formulating the economic forces and implications of M&As, the issues that genuinely concern the market remain hidden in a black box. In practice, each deal involves multi-dimensional concerns. The value-relevant issues include but are not limited to synergies, management teams and cultural integration, deal structures and financial terms, technology and human resources, market conditions and strategies, etc. The complex nature of merger deals makes it challenging to uncover the set of issues that the market perceives to be relevant from standard financial databases. A thorough study of value implications of various aspects of deals calls for much richer data to expand the matrix observable to the econometrician.

One way to study such market perception is by directly looking into the contents of merger-related communications between the management and market participants. Following a deal announcement, the management often holds a conference call to provide detailed explanations of the deal and to answer questions from analysts and market participants. Unlike other forms of information disclosure, such as SEC filings and website news, the communication in conference calls are interactive. The questions and discussions tend to center around the few issues that are most pivotal to evaluating the specific deal. Moreover, there is no uniform structures for conference calls. The flexible form enables more discussion of forward-looking information that is of most concern to market participants. The M&A conference calls provide an ideal data source to study what issues of merger deals are perceived as value relevant by the financial market.

In this study, we first propose to analyze the contents of M&A conference calls by generating structured data from the call transcripts. Leveraging machine learning techniques, we aim to uncover the topics discussed in each calls and infer the market-perceived pivotal issues from them. We further examine the value implications of each issues by regressing the market reactions to deal announcement on the topic distributions. Finally, the interactive nature of conference calls also allows us to explore the communication dynamics between the management and call participants.

The contents of M&A conference call have not been well studied. To the best of our knowledge, one existing paper (Kimbrough and Louis, 2011) has mainly focused on understanding the determinants and consequences of management's decision of holding an M&A call, while the content of calls remains largely unexplored. The use

* Chinese University of Hong Kong, Business School, 12 Chak Cheung Street, Shatin, N.T., Hong Kong.
Email: s.dasgupta@cuhk.edu.hk; fangyuanma@cuhk.edu.hk; me@daisy.finance; haojunxie@cuhk.edu.hk.

of machine learning techniques enables us to handle large scale of textual data and conduct comprehensive analyses on the linguistic contents.

Research Strategy:

We obtain the M&A conference call transcripts from the Street Event database. Our sample covers all the conference call transcripts associated with M&A deals globally from 2003 to 2017, which includes around 200,000 paragraphs containing 260,000 unique terms. To match the calls from the Street Event database with the M&A deals in SDC, we resort to automated methods to undertake the fuzzy matching process based on names and titles, and further validate the matching by using programs to extract relevant information from transcripts and news. Such automated solutions also enable us to identify the call participants from I/B/E/S database using call transcripts. Our analysis is done in Python and R, the standard languages for machine learning applications.

We face two major challenges. First, we need to dissect the content of the conference call transcripts by attributing texts to various semantic clusters. By far the most common approach to automated content analysis in finance and economics relies on the dictionary methods in which a set of keywords are pre-defined before being searched (e.g. Baker, Bloom, and Davis, 2016). While the dictionary methods are generally efficient for extracting statements on certain focal topics of interest, they are not suitable in this M&A context given that the potential scope of topics could be very large ex-ante. It is practically challenging to impose word representation to all the potential topics, and statistically undesirable to restrict attention to certain set of topics subjectively chosen by researchers.

We use the *probabilistic topic modeling* approach, a set of algorithms that aim to extract thematic information from large scale of documents (Blei, 2012). An important advantage of *topic modeling* over dictionary methods is that it obviates the need for labeling, as the algorithm directly learns topics from the original documents. The variation in all the vocabulary terms is used to determine which words are important for discriminating topics. In a typical *probabilistic topic model*, we regard the observed documents as the outcome of a data generating process that includes hidden thematic structures -- the topics, per-document topic distributions, and the per-document per-word topic assignments. The central computational problem for *topic modeling* is to infer such hidden structures from the term-document matrix. It thus achieves dimensionality reduction by representing each document with a low-dimensional topic distribution. In our study, we exploit the recent extensions of *latent Dirichlet allocation* (LDA, as in Blei, Ng, and Jordan, 2003), such as *correlated topic model* (Blei and Lafferty, 2007) to accommodate the co-occurrence of M&A relevant issues in conference calls, and *Dirichlet-multinomial regression models* (Mimno and McCallum, 2008) to incorporate the additional deal characteristics into topic models.

The second challenge of this research is to control for the selection issue that managers endogenously choose whether to hold a call, and the potential topic distribution of deals without calls could be systematically different from those observed, should a conference call be held. To draw unbiased inferences, we adopt an *endogenous switching model* in which the managers are assumed to decide whether or not to hold an M&A call taking into account the ex-ante expected market reactions under two choices. The market reactions conditional on not having a call is modeled as a function of the firm and the deal characteristics, while function of the reactions for having a call is further augmented with the topic distribution of the conference call. The estimated parameters on topic weights should reflect the average market perception about the value associated with each topics of M&A.

Preliminary Results:

We have performed a pilot test using the *structural topic model* (Roberts, Stewart, and Airoldi, 2016) allowing the topic contents and prevalence to vary across the acquirers' industry. The algorithm performs well as the topic clustering appears to have natural economic meanings (see details in the appendix). Preliminary regressions shows that more discussions on *Cost & Synergy* in the Q&A session is associated with more positive abnormal returns of the acquirer around deal announcement, while such discussions in the management presentation is not significant in explaining market reactions. We also find more discussions on *Team & Culture* and *Technology* are generally associated with negative market reactions, suggesting that market concerns about the high uncertainty of deals that could only be evaluated based on such soft information.

Reference:

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593-1636.

Betton, S., Eckbo, B. E., & Thorburn, K. S. (2008). Corporate takeovers. In *Handbook of Empirical Corporate Finance* (pp. 291-429).

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993-1022.

Blei, D. M. (2012). Surveying a suite of algorithms that offer a solution to managing large document archives. *Communication of the ACM*, *55*(4), 77-84.

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, *1*(1), 17-35.

Kimbrough, M. D., & Louis, H. (2011). Voluntary disclosure to influence investor reactions to merger announcements: An examination of conference calls. *The Accounting Review*, *86*(2), 637-667.

Mimno, D., & McCallum, A. (2008). Topic models conditioned on arbitrary features with Dirichlet-multinomial regression. *In Uncertainty in Artificial Intelligence*, (pp.411-418).

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, *111*(515), 988-1003.

Appendix: Preliminary Results

We estimate the *structural topic model* with 15 topics and acquirer's (one-digit SIC) industry as the covariates and report the outcomes below.

The table below shows the top keywords for each topic. We name each topic based on these keywords.

	Topics	**Top Keywords**							
1	Location & Government	nordic	sweden	netherlands	kong	belgium	spain	european	brazilian
2	Business & Production	fuel	gas	grain	ontario	plant	producer	oil	renovation
3	Market condition	cyclicality	margin	seasonality	recession	pricing	lumpy	seasonal	slowdown
4	Deal negotiation	speculate	premature	diligence	auction	lawyer	wait	anything	bidding
5	Financial projection	expensed	goodwill	depreciation	amortized	accretion	taxable	formas	intangible
6	Team & Culture	team	cultural	stakeholder	incredible	impressed	breed	thrive	culture
7	Cost & Synergy	overhead	synergy	headcount	saving	purchasing	duplicate	phasing	redundancy
8	Small words	luck	bye	alright	recorded	instruction	congrats	mute	session
9	Technology	connecting	electronically	wireline	fulfillment	installation	manufacturer	mobility	digital
10	Disclaimer	www	cautionary	undue	uncertainty	undertakes	statement	cautioned	webcast
11	Strategic prospects	illustrates	headquartered	diversifies	comprised	approximately	disciplined	unparalleled	demonstrates
12	Small words	awful	thing	stuff	bunch	digest	chasing	really	think
13	Deal structures	bond	debt	covenant	repurchase	tranche	dividend	payout	libor
14	Product & Distribution	promotional	catalog	wal	mart	grocery	promotion	channel	distributing
15	Small words	yes	actually	sorry	say	basically	answer	question	repeat

Next, we look into the keywords of each covariates for each topic as a validation of the model.

For the topics that we expect cross-industry variations in vocabularies, such as Topic 2, 7, 9, 11 and 14, we find clear industrial heterogeneity in the covariate keywords. The following table reports the covariates' keywords for Topic 7 (Cost & Synergy) as an example.

Acquirer's Industry	**Top Keywords of Covariates for Topic 7 (Cost & Synergy)**							
Agriculture	harvest	switch	opex	soft	saving	cost	overhead	synergy
Mining	acreage	drilled	acre	drill	foot	texas	play	basin
Construction	equipment	panel	labor	saving	efficiency	cost	capex	overhead
Manufacturing	capacity	saving	synergy	manufacturing	ton	production	procurement	plant
Transportation	capex	synergy	saving	cost	integration	cingular	efficiency	opex
Wholesale	branch	sourcing	saving	cost	synergy	purchasing	integration	technical
Retail	eckerd	store	foot	capex	pipeline	saving	overlap	department
Financial	save	synergy	cost	saving	revenue	land	royalty	efficiency
Service	synergy	saving	license	cost	amp	integration	capex	park
Public administration	maintenance	saving	cost	synergy	productivity	footprint	land	capex

For the other topics that we do not expect cross-industry variations in vocabularies, we indeed find no clear industrial heterogeneity in the covariate keywords. The following table reports the covariates' keywords for Topic 5 (Financial projections) as an example.

Acquirer's Industry	**Top Keywords of Covariates for Topic 5 (Financial projections)**							
Agriculture	accretion	guidance	tax	million	salary	ebitda	quarter	refinancing
Mining	capex	percent	fee	guidance	million	break	roughly	quarter
Construction	tax	guidance	million	ebitda	fair	quarter	accounting	amortization
Manufacturing	tax	guidance	accretion	million	expense	amortization	quarter	intangible
Transportation	tax	guidance	million	ebitda	income	accretion	multiple	quarter
Wholesale	accretion	tax	guidance	amortization	million	eps	ebitda	fiscal
Retail	tax	guidance	million	ebitda	accretion	fiscal	quarter	cad
Financial	tax	accretion	adjustment	assuming	estimate	intangible	million	correct
Service	tax	ebitda	million	quarter	guidance	eps	amortization	deferred
Public administration	tax	accretion	amortization	guidance	disclosed	goodwill	million	accounting

In the table below, we show the results of regressing the acquirer's abnormal returns (controlling for Fama-French three factors) around the deal announcement window on the logarithm of length of discussions (number of words) on each topic in the Q&A sessions of M&A call transcripts, controlling for the endogenous selection issue of call decisions with the *endogenous switching model*. The acquirer's past history of making M&A deals and holding M&A conference calls are used as the instrument for their first stage decision of having a call or not.

Regression results from *Endogenous Switching Model*

CAR [-1,+5] Acquirer	
Locations & Government	0.0013
	(0.28)
Business & Production	0.0044
	(1.31)
Market condition	0.0038
	(0.69)
deal negotiation	0.0015
	(0.17)
financial projection	-0.0092
	(-1.55)
Team & Culture	**-0.0064***
	(-1.89)
cost & synergy	**0.0085***
	(1.92)
Small Words	**0.019****
	(2.46)
Technology	**-0.0064***
	(-1.96)
disclaimer	-0.0080
	(-1.34)
Strategic prospects	-0.0013
	(-0.35)
Small Words	0.0041
	(0.44)
deal structures	0.0049
	(1.07)
product & distribution	-0.000092
	(-0.02)
Small Words	-0.0086
	(-0.60)
Instrument	Acquirer past Deal/ Call
Deal and Acquirer Characteristics	Yes
Acquirer Industry Dummy	Yes
Target Industry Dummy	Yes
Year Dummy	Yes