Title: *E*ffects of *M*&A on *E*nergy Use in US Transportation: *R*obust *G*eneralized *E*stimation (*EMERGE*) using Tools of Statistical/Machine Learning

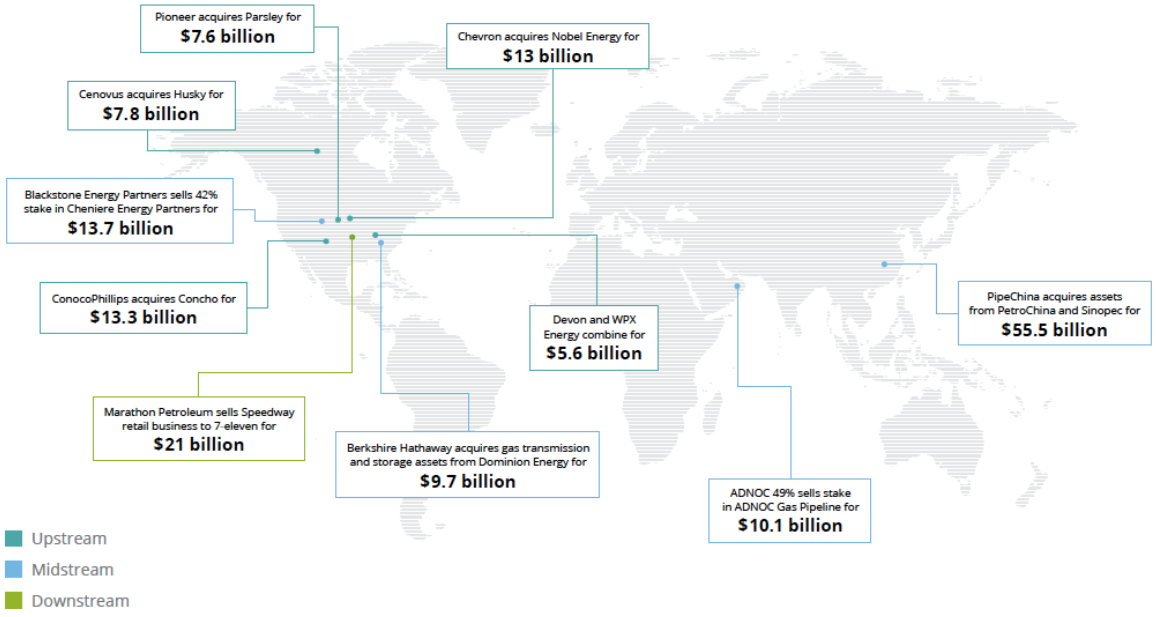
Researchers: Anjishnu Banerjee* and Avik Chakrabarti**

* Associate Professor, Division of Biostatistics, Medical College of Wisconsin (MCW).
** Associate Professor, Department of Economics, University of Wisconsin-Milwaukee (UWM).

Research Question: Is there quantifiable evidence of any *causal* link between mergers and acquisitions (M&A), within and across borders in the oil and natural gas industries, and the level as well as mix of energy consumption in U.S. transportation?

Project Narrative: The U.S. transportation sector accounts for nearly a third of the nation's energy consumption. Petroleum, natural gas, and coal dominate the energy consumption in the U.S. transportation sector, attributable to several modes: Cars, light trucks, and motorcycles (~ 60%) account for the largest share of energy consumed in the transportation sector, followed by other trucks (~ 25%) and aircrafts (~ 10%).[1] The post-pandemic resurgence of M&A activities in the oil and natural gas industries (Figures 1 and 2) would likely have far reaching implications for energy use in transportation.[2]

Figure 1. Top 10 Oil and Gas M&A Transactions (2020)

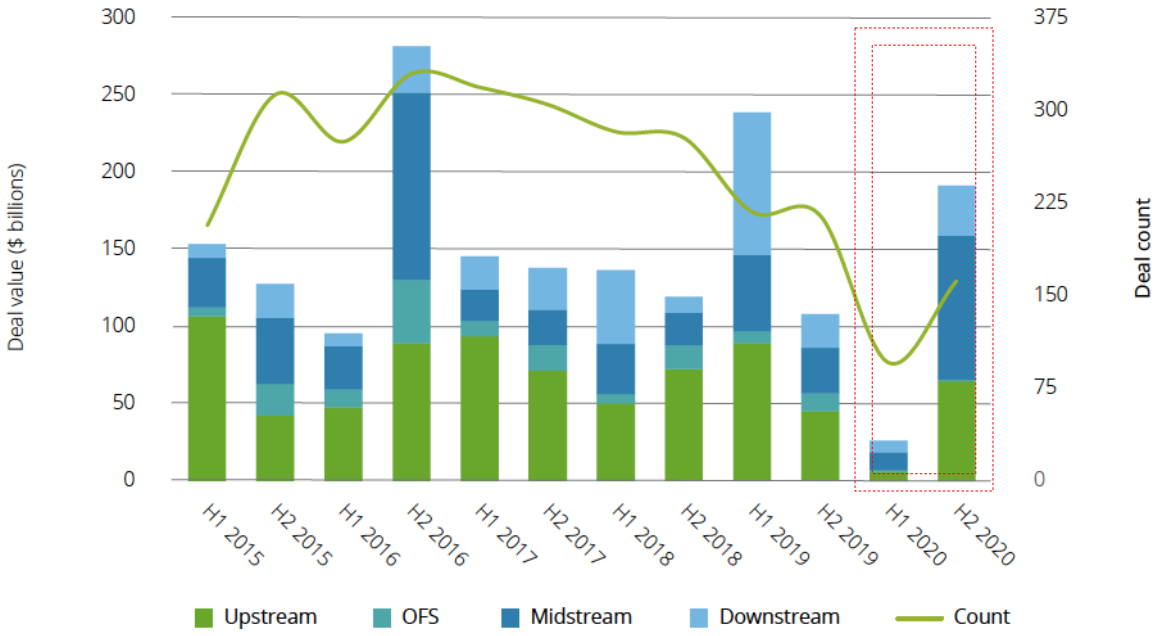


Source: Enverus (2021)

---

[1] See International Energy Outlook (2020).
[2] See Glaeser and Poterba (2020), Glaeser, Gorback, and Redding (2020), Greenstone *et al*. (2017), Knittel *et al*. (2016), Whitefoot and Skerlos (2012), and Greene *et al*. (2005).

Figure 2. M&A Deals in Oil and Natural Gas



Source: Authors' Adaptation of Enverus (2021)

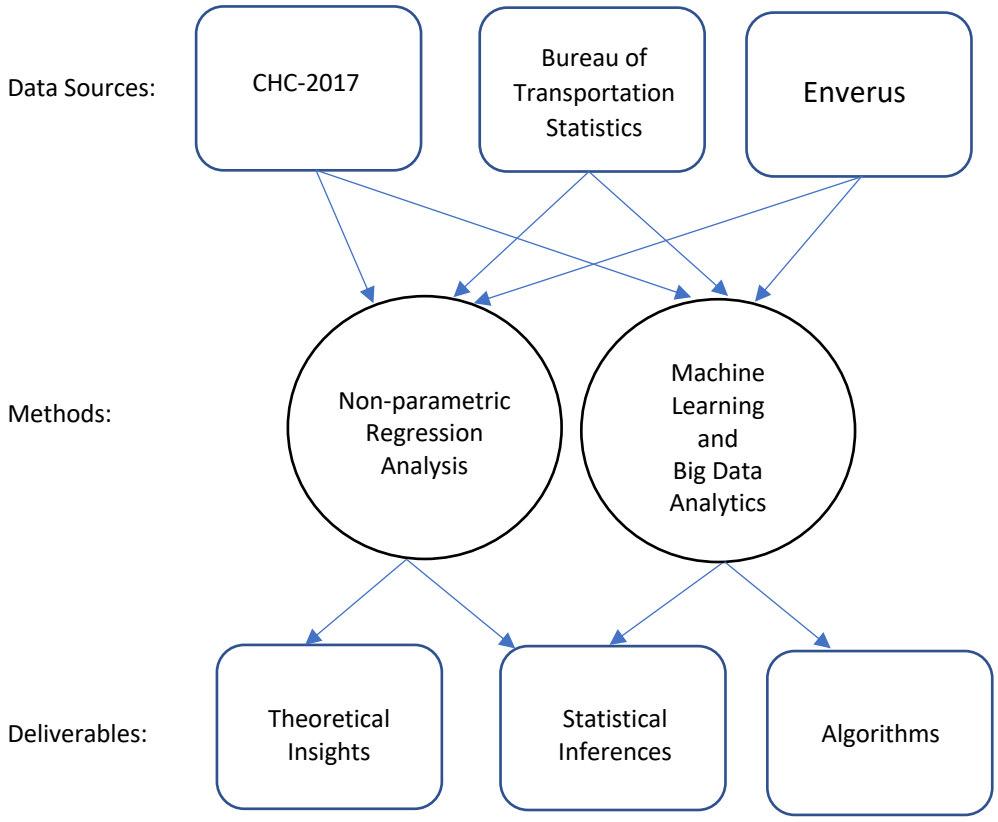
As the energy and transportation sectors undergo digitization and liberalization, such transformations bring to the fore defining characteristics of big data scenarios: sensors, communication, computation, and control capabilities. In liberalized markets, potentials of such big data are realizable when information flows across organizational boundaries are captured through increased digitization and automation of the infrastructure for operational efficiency leading to high-volume, high-velocity, and high-variety data. Chakrabarti, Hsieh, and Chang (2017) have drawn on a rich and growing body of literature[3] on M&A, within and across borders, to construct a unique data-set (CHC-2017, hereinafter), by extracting M&A observations (from Security Data Corporation's (SDC) Mergers and Corporate Transactions database) on individual firms and augmenting this data with detailed information on competition measures spanning 90,614 mergers (22,600 of which are cross-border M&A with a total transaction value of $3.15 trillion) between firms from 86 countries with a total transaction value of $10.49 trillion. After matching CHC-2017 with increasingly accessible data on energy use and transportation, our goal is to determine, by employing state-of-the art techniques of big data analytics and machine learning, how recent M&A activities, within and across borders, in the oil and natural gas industries are likely to affect the level and mix of energy consumption in U.S. transportation.

Data and Methods: The energy sector, on the one hand, generates vast amounts of big data through meters, sensor networks, customer payments, credit history, satellite imagery, etc. It is not surprising that energy companies are turning to the idea of leveraging big data analytics for performance optimization and improved service delivery. A combination of technologies, and a more complex world demanding greater agility and new competences impact all aspects of the energy sector and manifest themselves in changing patterns of consumption, new ways of asset

[3] See Beladi and Chakrabarti (2019) and Beladi, Chakrabarti, and Marjit (2013).

optimization, and cross-industry partnerships. On the other hand, increasing access to big data in real time and near real time on an unprecedented scale has inspired fast and comprehensive data collection capabilities to support both current ways of doing business and future innovations in the transportation industry. With this backdrop, figure 3 provides a schematic presentation of the data sources and methods, to be used in our *EMERGE* study, with specific reference to deliverables.

Figure 3. Data Sources, Methods, and Deliverables



While harnessing the rich and growing pool of information, without entertaining "Big data hubris" (i.e. any presumption that big data can substitute rather than complement conventional data collection and analysis), we recognize that dimensionality of data does not address foundational issues such as measurement, validity, and dependencies. For instance, and most importantly, we directly address the issue of potential misspecifications in our analyses. More specifically, in matching CHC-2017 with the vast, heterogeneous, and moving pool of data on energy use and transportation, we recognize that such a join can inaccurately specify matches due to incorrectly tagged keywords or missing potential matches, possibly leading to both false positives and false negatives. We specifically aim at addressing this important issue of potential mis-specification by employing a Bayesian Random Forest specification for robust quantification of non-linear patterns in our regression analyses and test if our model is robust and, if so, to what degree of mis-specification of covariates. From an analytic perspective, the complexities involved in studying the associations between M&A and energy use, can be grouped into two categories: a) Nested observations, within industries/entities of interest and b) imperfectly observed variables of interest. The overall mathematical model, to assess the effect of M&A (within and across borders) in the

oil and natural gas industries on the level and mix of energy consumption in U.S. transportation, can be represented as,

$$y_{it} = f(I_{it}, \mathbf{X}_{it}) + \epsilon_{it}$$

where, for any unit (i) of analysis (e.g. trucking companies), y_{it} is the value of the dependent variable of interest (e.g. energy consumption in transportation) at time t; I_{it} is a categorical variable indicating whether a trucking company i consumes energy supplied by any firm involved in M&A at time t; $\mathbf{X}_{it}$ is a matrix of observable controls; and ϵ_{it} is the unexplained model noise term. To the extent that I_{it} may be imperfectly observed, we propose the following hierarchical formulation. Let T_{it} be the true unobserved value of the categorical variable indicating whether a trucking company i, for instance, consumes energy supplied by any firm involved in M&A at time t and I_{it} be its deduced value from a probabilistic matching algorithm, possibly incorrectly specified. Consider the vector of conditional probabilities $\boldsymbol{\pi_{it}} = \{\boldsymbol{\pi_{it}}^j\}$ where $\boldsymbol{\pi_{it}}^j = \Pr(T_{it} \in j \mid I_{it})$. A multivariate model can be constructed for estimation of these conditional probabilities, such as

$$\boldsymbol{\pi_{it}}^j = g(U_i, V_j, I_{it})$$

where U_i and V_j represent observed characteristics. Since the estimated value of y_{it} itself depends on the covariate I_{it}, we use a Bayesian formulation, where U_i and V_j are encoded into prior parameters. Once these conditional probabilities are estimated, they replace their counterparts reducing our model to

$$y_{it} = f(\boldsymbol{\pi_{it}}, \mathbf{X}_{it}) + \epsilon_{it}$$

As a by-product of the above formulation, we will incorporate dependence structures, such as nested industries or other groupings of interest, through the dependence imposed on the noise term (ϵ_{it}). Typical machine learning algorithms assume independent observations, which are not realistic in the present context. A more flexible dependent structure, meshed within the estimation algorithm itself, will be imposed through the Bayesian formulation.[4]

Summary: In sum, contemporary market, technology, and policy drivers of globalization have prompted widespread collection of information leading to the accumulation of big data on energy use by stakeholders across industries, resulting in nested observations, which are addressed by the proposed model above. At the same time, there is mounting evidence of transformational changes in the transportation industry that have important implications for energy use. Still, modeling big data on energy use in transportation remains at its infancy. Any meaningful role of the Department of Transportation (DOT), particularly in future proofing infrastructure as well as enabling adaptability and resilience, must rely on quantifiable evidence of any *causal* link between domestic and cross-border M&A in the oil and natural gas industries and the level as well as mix of energy consumption in U.S. transportation. As such our finding(s) will have far reaching implications for research and policy priorities of the DOT.

Team Composition: Dr. Banerjee will use his expertise in network based statistical learning, including Bayesian and classical multilevel models. Dr. Chakrabarti will use his expertise in econometric modeling to detect any latent pattern in the data and meaningfully interpret any causal link, discovered through the process of this empirical exercise, that can place the new finding(s)

[4] See Banerjee et al. (2013a, 2013b, 2013c).

in the context of the extant literature. A research assistant will help in data cleaning, developing machine learning models. She/he will contribute to manuscript writing as well.

Conflict of Interest Statement: The research team has no financial or other interest in or any tie to the transportation industry that might bear on the proposed work.

References

Banerjee, Anjishnu, Jared Murray, and David Dunson, (2013a). "Bayesian learning of joint distributions of objects." *Artificial Intelligence and Statistics*. PMLR.

Banerjee, Anjishnu, David B. Dunson, and Surya T. Tokdar (2013b). "Efficient Gaussian process regression for large datasets." *Biometrika* 100.1: 75-89.

Banerjee, A., Vogelstein, J. and Dunson, D., (2013c). "Parallel inversion of huge covariance matrices." arXiv preprint arXiv:1312.1869.

Beladi, H. and Chakrabarti, A., (2019). Multidivisional firms, internal competition, and comparative advantage," *Journal of International Economics* 116, 50-57.

Beladi, H., Chakrabarti, A., and Marjit, S. (2013). "Cross-border mergers in vertically related industries," *European Economic Review* 59, 97-108.

Chakrabarti, A., Hsieh, Y., and Chang, Y. (2017). "Cross-border mergers and market concentration in a vertically related industry: Theory and evidence," *Journal of International Trade and Economic Development*, 26(1): 111-130, 2017.

Energy Information Administration, 2020. International Energy Outlook 2020.

Energy Sector Management Assistance Program. (2017). Energy analytics for development: Big data for energy access, energy efficiency, and renewable energy. ESMAP Knowledge Series;No. 027/17. World Bank, Washington, DC.

Enverus, 2021. https://www.enverus.com/solutions/energy-analytics/ep/prism/core/m-and-a-data/

Glaeser, E.L. and Poterba, J.M. (2020). "Economic analysis and infrastructure investment" National Bureau of Economic Research No. w28215.

Glaeser, E.L., Gorback, C. and Redding, S.J. (2020). "JUE Insight: How much does COVID-19 increase with mobility? Evidence from New York and four other US cities." *Journal of Urban Economics*, https://doi.org/10.1016/j.jue.2020.103292.

Graham C. (2017). "The future is not what it used to be: oil and gas strategies for a carbon-conscious world," *The APPEA Journal*, 57(2): 459-61.

Greene, D.L., Patterson, P.D., Singh, M., Li, J., (2005). "Feebates, rebates and gas guzzler taxes: a study of incentives for increased fuel economy," *Energy Policy*, 33 (6) 757–775.

Greenstone, M., Sunstein, C.R. and Ori, S. (2017). "The next generation of transportation policy," *Policy Proposal* 2017-02, Energy Policy Institute, University of Chicago.

Knittel, C., Metaxoglou, K. and Trindade, A. (2016). "Are we fracked? The impact of falling gas prices and the implications for coal-to-gas switching and carbon emissions," *Oxford Review of Economic Policy*, 32(2): 241-259.

Whitefoot, K.S. and Skerlos, S.J. (2012). "Design incentives to increase vehicle size created from the US footprint-based fuel economy standards," *Energy Policy*, 41: 402-411.