NBER Call for Proposals

Big Data and High-Performance Computing for Financial Economics

Project title: Human Capital and the Lifecycle of Firms

Maxime Bonelli* Jack Liebersohn† Victor Lyonnet†

How much do worker characteristics affect the performance of firms? There is evidence that top executives like CEOs can significantly influence outcomes (Bertrand and Schoar, 2003), but we know less about other employees. There is also little known about whether the correct match of skills and abilities between employees and firms matter more at certain stages in the lifecycle of a firm. For example, can individual employees play a bigger role in the success of startups compared to established firms? We propose to address these questions using a rich employee-employer linked dataset together with machine-learning tools to create new measures of worker skill, ability and human capital and their contribution to firm outcomes. Unexpected employee departures provide exogenous variation with which to identify the causal impact of worker characteristics on firm outcomes.

Our data comes from various administrative sources in France such as *Déclarations Annuelles de Données Sociales* (DADS) and tax filings. Together, this data contains the universe of private firms in France from 1990 to 2017, along with information on their workers, such as worker-level occupation and firms' workforce composition at the establishment level. Worker characteristics also include demographic data such as employee's age, gender, and education as well as employment characteristics such as number of hours worked and employment history for a subset of the labor force. The dataset represents millions of observations each year and high-dimensional characteristics for each employee, containing both structured and unstructured components.

Our project will have two parts. First, we will use machine learning tools to identify patterns between firms' outcomes and their worker characteristics. For instance, we might use variable selection with sparsity and post-selection inference (Belloni et al., 2011; Chernozhukov et al., 2016). Second, we will identify the causal effect of human capital on firms' performance by

*HEC Paris.

†Ohio State University. Contacts: maxime.bonelli@hec.edu, liebersohn.1@osu.edu, lyonnet.1@osu.edu.

studying plausibly exogenous employee departures (following, e.g., Jones and Olken, 2005; Jaravel et al., 2018). Machine learning tools will allow us to explain heterogeneity in the effect of human capital across firm types as well as at different points in their lifecycle (Athey and Imbens, 2016).

Both parts will require us to determine the most relevant measures of human capital and employee characteristics. The high dimensionality of the data, and classification of more than 400 distinct occupations, make this a challenge. An emerging literature in labor economics uses a "task" approach to classify jobs,[1] but occupations might not be readily mapped into a set of skills or tasks.

As a first step, we will map each occupation into a set of abilities and tasks using externally supplied and validated job content measures from the Occupational Information Network (O*NET). Next, we will contrast the results from the task-based measure of human capital with a machine-learning approach for selecting the relevant set and combination of occupation characteristics and associated tasks, employee skills, and employment history. Our results may validate, or possibly challenge, the traditional task-based approach to measuring human capital.

Our project contributes to two distinct research areas: finance research on corporate governance and labor, and research in entrepreneurship. It has the potential to address key unanswered questions in these areas, including: What is the importance of human capital for firms? How do employees characteristics predict the success of startups?

We will improve on the existing literature in several dimensions. First, many existing papers study the role of CEOs and other top executives, showing how the skills and expertise of top workers affect firms' decisions (Güner et al., 2008; Fee et al., 2013; Benmelech and Frydman, 2015; Bandiera et al., 2017) and are tightly linked to corporate performance (Bertrand and Schoar, 2003; Guiso et al., 2015; Bernile et al., 2018). Our paper will extend these findings by studying firms' entire occupational structure. Second, rather than relying on existing measures of human capital (e.g., education or occupational tasks), we will leverage machine learning techniques to determine the most relevant aspects of employees' profile. This can provide the basis of a better measure of human capital.

From a computational perspective, we plan to use Python and R. These programming languages have several packages implementing machine learning tools. Our research team has experience in these programming languages as well as working on high-performance computing clusters.

[1] See, for instance, Autor et al. (2003); Acemoglu and Autor (2011); Autor and Handel (2013).

References

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, volume 4, pages 1043–1171. Elsevier.

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360.

Autor, D. H. and Handel, M. J. (2013). Putting tasks to the test: Human capital, job tasks, and wages. *Journal of Labor Economics*, 31(S1):S59–S96.

Autor, D. H., Levy, F., and Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, 118(4):1279–1333.

Bandiera, O., Hansen, S., Prat, A., and Sadun, R. (2017). Ceo behavior and firm performance. Technical report, National Bureau of Economic Research.

Belloni, A., Chernozhukov, V., and Hansen, C. (2011). Inference for high-dimensional sparse econometric models. *arXiv preprint arXiv:1201.0220*.

Benmelech, E. and Frydman, C. (2015). Military ceos. *Journal of Financial Economics*, 117(1):43–59.

Bernile, G., Bhagwat, V., and Yonker, S. (2018). Board diversity, firm risk, and corporate policies. *Journal of Financial Economics*, 127(3):588–612.

Bertrand, M. and Schoar, A. (2003). Managing with style: The effect of managers on firm policies. *The Quarterly Journal of Economics*, 118(4):1169–1208.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., and Newey, W. K. (2016). Double machine learning for treatment and causal parameters. Technical report, cemmap working paper.

Fee, C. E., Hadlock, C. J., and Pierce, J. R. (2013). Managers with and without style: Evidence using exogenous variation. *The Review of Financial Studies*, 26(3):567–601.

Guiso, L., Sapienza, P., and Zingales, L. (2015). The value of corporate culture. *Journal of Financial Economics*, 117(1):60–76.

Güner, A. B., Malmendier, U., and Tate, G. (2008). Financial expertise of directors. *Journal of financial Economics*, 88(2):323–354.

Jaravel, X., Petkova, N., and Bell, A. (2018). Team-specific capital and innovation. *American Economic Review*, 108(4-5):1034–73.

Jones, B. F. and Olken, B. A. (2005). Do leaders matter? national leadership and growth since world war ii. *The Quarterly Journal of Economics*, 120(3):835–864.