Can Social Media Reliably Estimate Unemployment?

Do Lee[1], Manuel Tonneau[1,2,3], Boris Sobol[4],
Nir Grinberg[4], and Samuel Fraiberger[1,3,*]

[1]New York University
[2]University of Oxford
[3]The World Bank
[4]Ben-Gurion University of the Negev

[*]Corresponding author; E-mail: sfraiberger@worldbank.org.

Abstract

Digital trace data holds tremendous potential for measuring policy-relevant outcomes in real-time, yet its reliability is often questioned. Here, we propose a principled yet simple approach: capturing individual disclosures of unemployment using a custom large language model and post-stratification adjustment using inferred user demographics. We show that our methodology consistently outperforms the industry's forecasting average, and can improve the predictions of U.S. unemployment insurance claims, up to two weeks in advance, at the national, state, and city levels at both turbulent and stable times. The results demonstrate the potential of combining large language models with statistical modeling to complement traditional survey methodology, and contribute to better-informed policymaking, especially at turbulent times.

Introduction

Two weeks after COVID-19 was officially declared a pandemic, the number of people filing new claims for unemployment benefits ("UI claims") in the U.S. surged from about 278 thousand to nearly six million. Absent accurate, real-time information about the magnitude of the shock that triggered the worst job crisis since the Great Depression, government agencies across the country quickly became unable to process claims in a timely manner, which had serious economic and psychological ramifications for beneficiaries (*1*).

This episode epitomizes that timely and disaggregated information about the labor market is vital for economic well-being. It improves market efficiency (*2*) and enables the design of evidence-based policies (*3*). However, official statistics are typically available with a considerable lag – especially at high resolution – and are subject to ex-post revisions, which impedes policymakers' ability to alleviate the impact of economic shocks in a timely fashion. For example, U.S. statistics on UI claims are published with a lag of at least four days at the national and state levels as well as for a limited number of cities; other valuable unemployment statistics are only available on a monthly or quarterly basis, with limited coverage at the subnational level. These limitations are even more severe in low- and middle-income countries, where national statistical agencies often lack the resources to consistently collect timely and reliable labor market data (*4, 5*).

In this context, the potential of real-time digital trace data to complement official statistics has been explored extensively over the past decade (*6–8*). Social media data have proven to be a valuable source of information across various domains such as quantifying migration flows (*9*), the impact of natural disasters (*10*), economic mobility and connectedness (*11, 12*), asset markets fluctuations (*13, 14*), economic policy uncertainty (*15*), inflation expectations (*16*), and employment shocks (*17*). Several studies have also identified signals from social media – in-

cluding users' diurnal rhythm (*18*), connectedness and keyword counts (*19, 20*) – that can be correlated with unemployment statistics; however, these approaches fell short of demonstrating sufficient predictive power and robustness to be relied upon in practice.

In this study, we present a principled yet simple methodology for forecasting the number of UI claims in the U.S. at the national, state, and city levels up to two weeks ahead of the official release using unemployment self-disclosures on Twitter. We focus on UI claims as it is the most frequently updated official statistics about the labor market and an important macroeconomic measure for policymakers (*21, 22*), macroeconomic forecasters (*23, 24*), and financial markets (*25*). We identify unemployment disclosures by training a custom Large Language Model (LLM) using Active Learning (*26*), a sampling strategy that maximizes detection performance by letting the model choose the data points from which it learns. Our LLM captures substantially more disclosures of unemployment on Twitter compared to previous approaches without compromising on precision, it identifies disclosures from more people, and yields a more representative sample of unemployed users. Then, using inferred user demographics and census population estimates, we construct a Twitter unemployment index by post-stratifying the proportion of unemployed users to correct for the sample non-representativeness. UI claim predictions are based on an autoregressive model using the Twitter unemployment index, past official statistics, and the industry's consensus forecast (*27*) (see Supplementary Methods). To thoroughly evaluate our methodology, we test model predictions over the course of three years, during both turbulent and "normal" times, evaluate accuracy up to two weeks before the official statistics are released, measure robustness at the national, state, and city levels, and compare performance to the industry's leading consensus forecast. By contrasting our models with previously proposed rule-based approaches (*19*), unweighted variants, and down-sampled versions, we gain insight into the contributing factors for the model's success. Finally, we demonstrate the model's ability to fill in gaps in official statistics.

Methods

Self-disclosures of unemployment status are extremely rare in the sea of social media content. Therefore, we need a large sample of users and a comprehensive approach to detect unemployment self-disclosures. To that end, we query the Twitter API to collect the tweets posted by users with a profile location that uniquely maps to a geographical location in the U.S., and snowball sample additional users mentioned in these tweets (see Supplementary Methods for more details). The dataset analyzed here consists of public tweets posted by a snowball sample of 31.5 million U.S.-based users collected between January 2020 and December 2022. We use two different approaches to identify public self-disclosures of unemployment status in tweets' text. The first is a rule-based model inspired by previous work (*19*), where a tweet is considered disclosing one's unemployment status if it contains any of 75 theoretically-motivated phrases describing job loss such as "I just lost my job" (see Supplementary Fig. S2 for a complete list of phrases and their prevalence). The second approach trains an LLM by following the procedure proposed by Tonneau et al. (*28*). It involves an Active Learning iterative process where at each step, a BERT-based LLM (*29*) is trained on the currently available, manually-labeled tweets, and then model uncertainty is used to select additional tweets to be sent for labeling. The final model, referred to hereafter as JoblessBERT and available publicly[1], is trained on a set of 8,838 tweets (see Supplementary Methods for more details). To construct a daily unemployment index, we calculate the percentage of users who disclosed their employment status (using either the rule-based or JoblessBERT model) out of all active users observed in a sliding window of seven days. We also construct post-stratified versions of the index to adjust for the platform's non-representative user base (*30*) by reweighting users based on inferred age, gender, and location from their profiles to match U.S. general population estimates from the Census Bureau (see Supplementary Methods and Supplementary Fig. S1). Finally, we use an autoregressive

[1] https://huggingface.co/worldbank/jobless-bert

distributed lag model to predict weekly UI claims, where covariates consist of the Twitter unemployment index, official statistics about past UI claims, and the industry's consensus forecast when it becomes available (full model specifications are in the Supplementary Methods). The autoregressive model is trained on data from a weekly sample of 208 observations spanning 2016-2019 (inclusive) and tested in a weekly sample of 156 observations spanning 2020-2022 (inclusive), which includes the turbulent times of the COVID-19 pandemic. To ensure that the covariates are expressed in comparable units, the UI claims and consensus forecasts are normalized by the size of the labor force during the previous month. Hereafter, we refer to the normalized UI claims as UI claims for brevity.

Results

Detecting disclosures of unemployment status. First, we evaluate the ability of the language models to detect disclosures of a user's unemployment status on Twitter. We find that JoblessBERT considerably improves the classification of unemployment disclosures compared to the rule-based model (Fig. 1A). As one might expect, the rule-based model achieves a high level of precision (93.1%) with a relatively low recall (29.3%). In contrast, our JoblessBERT model retrieves nearly three times more relevant content about unemployment (recall of 76.5%; $P < 0.001$) with the same level of precision. As shown in Fig. 1A, our model maintains high precision (> 0.85) when retrieving more than 90% of the relevant disclosures. A closer examination of the linguistic patterns identified by JoblessBERT reveals that JoblessBERT expands beyond the frequent and intuitive patterns used in previous work (*31*). For example, JoblessBERT picks up expressions that contain spelling mistakes ("neeeeeed a job") and slang ("needa job"), which are prevalent on social media but are unlikely to be pre-conceived. These differences also considerably expand the set of users whose expression is captured: JoblessBERT identifies nearly 13 times more unemployed users than the rule-based model.

JoblessBERT also yields a more representative sample of users than the rule-based model. Examining the proportion of unemployed users in different states (Fig. 1B), we find that the rule-based model under-represents states where unemployment is low and over-represents states where unemployment is high: the slope of a fitted linear model yields a slope of 0.38, which is significantly different from an identity line ($P < 0.001$). In contrast, JoblessBERT's sample is more closely aligned with the actual distribution of unemployment across U.S. states, having a regression slope of 0.86, which is not significantly different from an identity line ($P > 0.15$). Fig. 1C further shows that JoblessBERT's sample is closer to the distribution of unemployment across age brackets in the general population: the proportions of unemployed users in age brackets of 20 years old or older are not statistically different from that of the official data. Notably, JoblessBERT over-represents unemployed youth (below 20), albeit to a lesser extent than the rule-based model, which stems from our sample's skew towards users under 20 years old ($P < 0.001$, see Supplementary Fig. S1). Finally, in terms of gender, Fig. 1D shows that the proportion of women unemployed users in JoblessBERT is closer to that of the official data than the rule-based model, although the proportions in both models are not statistically different from that of the official data ($P > 0.10$). Taken together, these findings highlight that our custom-trained LLM captures a broader variety of linguistic patterns describing unemployment, a substantially larger sample of users disclosing their unemployment status, and a sample that resembles more closely the official statistics of unemployment across states, age brackets, and genders.

Monitoring unemployment in real-time. Next, we investigate whether self-disclosures of unemployment on Twitter can help monitor UI claims. The numbers of UI claims for the current week – which ends on Sunday 12:00 AM – are published on Thursday 8:30 AM the following week. This lag of more than four days has created a space for an industry of professional

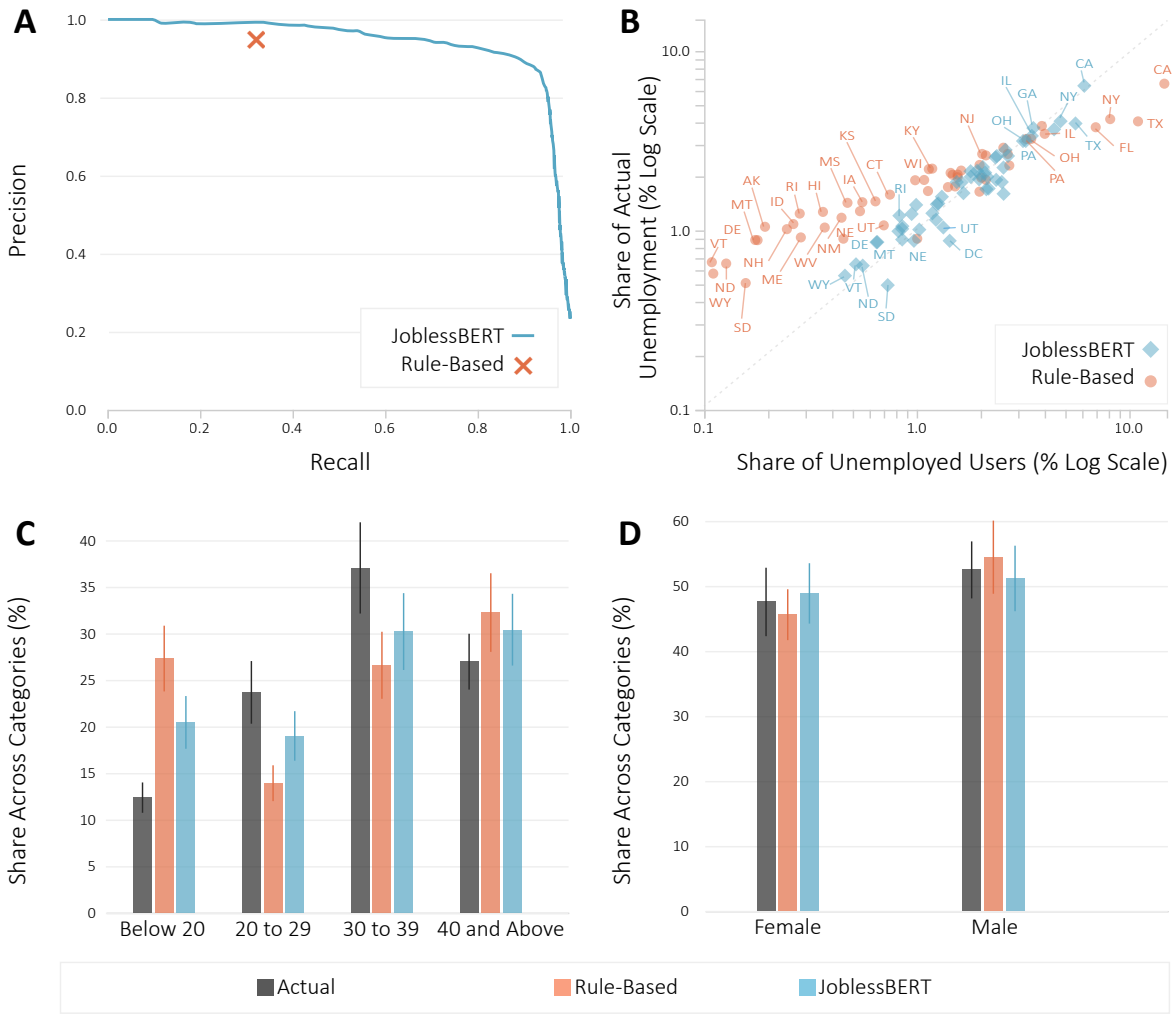

Fig. 1: Detecting disclosures of unemployment status.

(A) Precision-recall curve. **(B-D)** Distributions of unemployed users and actual unemployment by state, age bracket, and gender. Precision-recall curve has been computed using an evaluation sample of 3,546 tweets produced in prior work (*28*). Inferences on gender and age are available for the 23 million users in our sample with a valid profile picture. *JoblessBERT* outperforms the rule-based model, both in terms of precision and recall of detecting unemployment self-disclosures, and in producing a sample that is more representative of the general U.S. population.

forecasters, who publish their estimates almost two days before the work week ends, on Friday morning (Fig. 2B). In contrast, disclosures of unemployment by Twitter users, signaling potential eligibility for unemployment benefits, are available continuously throughout the week. Therefore, we construct daily estimates of weekly UI claims as the proportion of Twitter users who disclosed being unemployed in a seven-day sliding window out of all active users during the time window. We distinguish four types of such unemployment indices: unweighted indices, which are based on the raw numbers produced by the two language models (rule-based and JoblessBERT), and post-stratified indices, which are reweighted based on U.S. census population estimates from the previous month and using inferred age, gender, and state information of users (see Supplementary Methods for details).

Figure 2A shows the actual UI claims and the four indices constructed based on Twitter disclosures of unemployment on a logarithmic scale (10-based). Each series is normalized by its average value in the first month of January 2020, indicating, for instance, that UI claims rose by 20 orders of a magnitude two weeks after COVID-19 was declared a pandemic (on March 28, 2020) relative to actual claims recorded in January 2020. As shown in the figure, the unweighted indices underestimate changes in UI claims, particularly during volatile months of the pandemic, while the post-stratified indices are significantly closer to the actual UI claims ($P < 0.001$). The RMSE of JoblessBERT indices relative to UI claims are significantly lower than those of the rule-based indices ($P < 0.001$), with the post-stratified JoblessBERT index significantly outperforming all other indices ($P < 0.001$). While the gains of the post-stratified JoblessBERT index over the post-stratified rule-based index may seem small on a logarithmic scale, they are meaningful in absolute values. At the height of the pandemic (March to June 2020), the post-stratified rule-based index under-estimates UI claims, on average, by 54.5% more than the post-stratified JoblessBERT index (RMSE of 7.74 and 5.01 standard deviations of UI claims, respectively), which translates to underestimation of 872,978 claims during this

period ($P < 0.001$). On an average week during the more stable times after June 2020, the differences between the post-stratified rule-based and post-stratified JoblessBERT indices are smaller (RMSE of 1.14 and 0.77 standard deviations of UI claims, respectively), which translates to an average underestimation of 287,036 claims during this period ($P < 0.01$).

Using the post-stratified Twitter indices, we next examine the ability of a dynamic model to predict UI claims up to two weeks in advance of the official data release. To identify any predictive gains from our Twitter-based indices, we consider three specifications of an autoregressive distributed lag model: (i) a baseline "consensus model" that only uses past UI claims releases and professional consensus forecasts, (ii) a "rule-based model" that adds to the baseline model the post-stratified rule-based Twitter index, and (iii) a "JoblessBERT model" that adds to the baseline model the post-stratified JoblessBERT Twitter index. (see Supplementary Methods for full model specifications). Figure 2C shows the RMSE of the three models as a function of time relative to the end of the measurement week. For example, the consensus model's RMSE starts at 0.67 UI claims standard deviations on day -10 (two weeks before data release), drops to 0.46 standard deviations on day -3 when the official release about the previous week becomes available, and reaches 0.43 standard deviations after the consensus forecast is published on day -2. Across all three models, a lower RMSE is obtained as the release date draws closer, but there are important differences between models. First, there is a clear rank ordering between models, where on an average two-week period before data release of actual UI claims (the forecast target), the rule-based index reduces the RMSE of the baseline model by 28.5%, and the JoblessBERT index reduces it by 54.3%. Moreover, the figure shows that the starting point for the JoblessBERT model two weeks ahead of the data release is on par with the performance of the baseline model at the end of the measurement week ($d = 0$), when the baseline model has access to much more recent information.

It is also important to examine the model response to economic shocks. A pivotal example

of such a shock occurred during the first week after COVID-19 was declared a pandemic (March 14 to March 21, 2020), when UI claims jumped from about 252 thousand claims at the beginning of the week to 2.9 million claims at the end of it – an astounding increase of 4.1 standard deviations. The consensus model failed to anticipate the spike: using the industry's estimate two days before the week ended, the model predicted only 327.2 thousand claims. On the same day, the rule-based model "sensed" the sudden change and predicted UI claims to reach 2.32 million, underestimating the true value by 20.5% and an improvement over the 88.8% underestimation of the consensus model. Finally, the JoblessBERT model forecasted 2.66 million UI claims two days before the week ended and 2.8 million claims on the day before the official release of 2.9 million. These results suggest that JoblessBERT could play a key role in an early warning system that senses changes in the labor market. A subsample of users is sufficient to retain much of the predictive accuracy with a substantially smaller sample size (see Supplementary information, section S3 for details).

Monitoring unemployment subnationally. Focusing on national trends may obfuscate large variability in unemployment across local labor markets (*32*). Tracking sub-national dynamics is critical for understanding spatial heterogeneities as they occur, especially during a crisis, and for designing place-based policies (*33*). Therefore, we evaluate the predictive performance of our models at the sub-national level by estimating a separate model for each U.S. state and city. Since the consensus forecast is only available at the national level, we do not include it in our subnational models (see Supplementary Methods).

In line with the national-level results, we find that JoblessBERT robustly outperforms other models across all U.S. states (Fig. 3A). On an average two-week period before data release of actual UI claims, JoblessBERT's predictions are 36.2% more accurate than the baseline ($P < 0.001$) and 20.6% more accurate than the rule-based model ($P < 0.001$). As shown

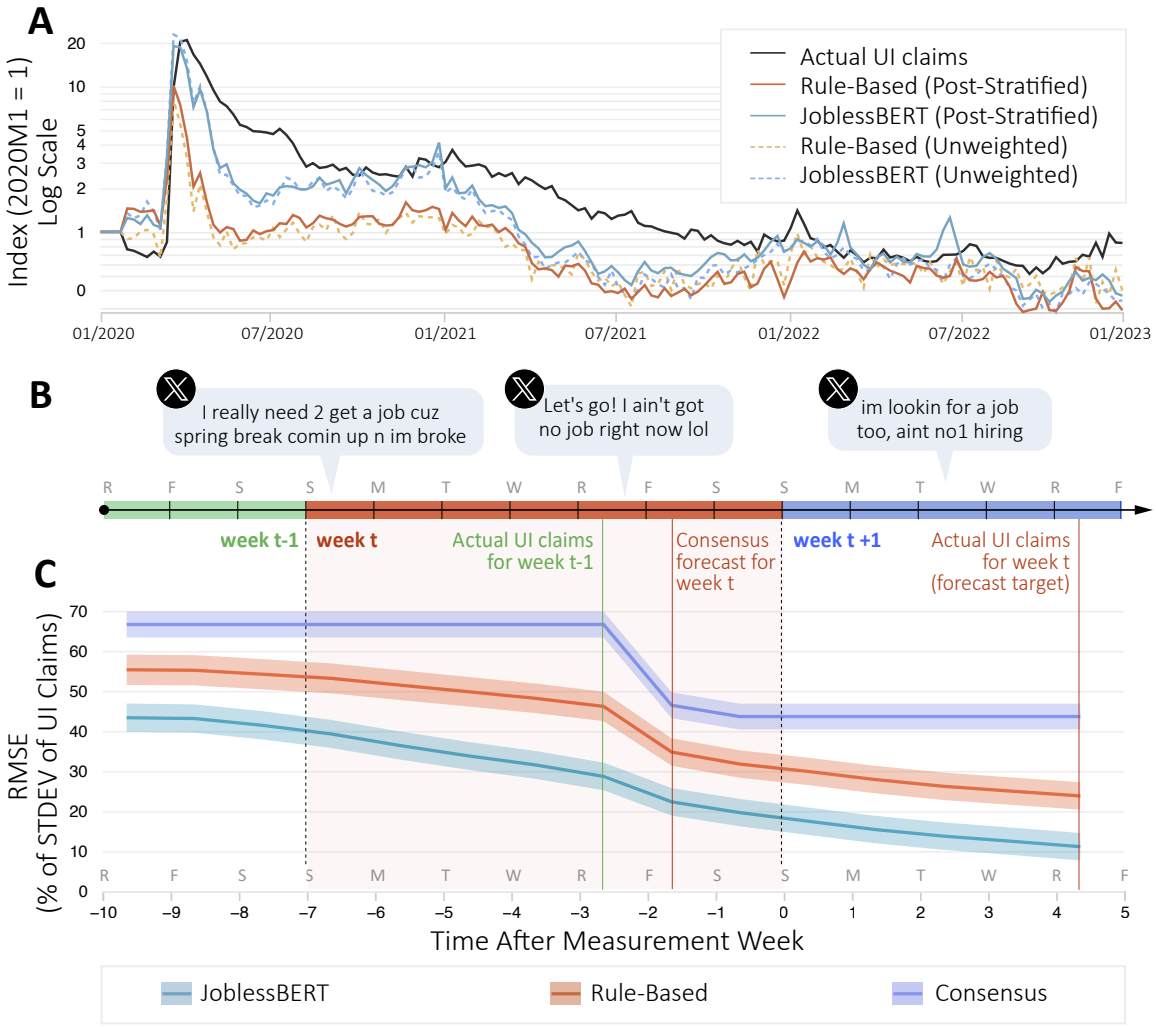

Fig. 2: Predicting the U.S. weekly initial claims for unemployment insurance (UI).
(A) Time series of the unemployment indices. **(B)** Timeline of the real-time data flow. **(C)** National-level predictions of unemployment insurance claims in the U.S. as a function of the forecasting horizon.

in Fig. 3B, the JoblessBERT model yields substantial error reduction two weeks ahead of the state data release compared to the baseline's prediction using all available information the day before the official release. It is also important to note that the accuracy of both the rule-based and JoblessBERT models steadily improve over time as more social media disclosures become available.

To push our approach to the limit, we evaluate the performance of our models at the city level using data from 45 cities across the U.S. (See Supplementary Methods). As shown in Fig. 3C, JoblessBERT's model continues to outperform all other models at the city level, with RMSE reductions of 23.0% over the baseline and 8.4% over the rule-based model, respectively ($P < 0.001$). Comparing the best-performing models at the city and state levels (i.e. JoblessBERT the day prior to data release), we find that state-level predictions are significantly more accurate than city-level predictions, leading to RMSEs of 0.45 and 0.73 standard deviations, respectively ($P < 0.001$). This suggests that our approach begins to reach a performance limit at the city level. To further investigate model behavior at the limit, we examine predictive performance as a function of variability in claims and the adoption of Twitter at the city level. Panels D and E of Fig. 3 show city-level RMSE as a function of the average change in UI claims (panel D), and as a function of the estimated Twitter penetration rate (panel E). Clearly, performance varies across cities, but an overall trend is observed in panels D and E: predictions are more accurate when UI claims are changing and when Twitter's adoption is higher.

Finally, to test the ability of our approach to compensate for gaps in official statistics, we evaluate the performance of the JoblessBERT model in ten "holdout cities," where official UI claim numbers are rarely or irregularly updated (see Supplementary Methods M8). As lagged variables are often missing in holdout cities, we substitute the city autoregressive terms in our models with state-level UI claims (see Supplementary Methods). Panels D and E in Fig. 3 (as well as Supplementary information, Fig. S3) show that forecast errors for holdout cities

(hollow points) are comparable to those of cities with regularly updated UI claims (full points). These results indicate that the JoblessBERT predictions are valuable even when actual UI claim numbers are unavailable during training, suggesting that signals extracted from social media can fill gaps in traditional measures of unemployment at the city level.

Discussion

This study demonstrates that an LLM-based index of self-disclosures of unemployment status extracted from social media can serve as a leading indicator of unemployment in the U.S. It shows that coupling an LLM with Active Learning can yield substantially more relevant content about unemployment than existing rule-based approaches without compromising retrieval quality. The results also suggest that the additional content identified by the classifier is not merely a replication of the same linguistic patterns but rather a diverse set of expressions by a more representative sample of the target population. Post-stratification using inferred demographics of users considerably improves the alignment of our Twitter-based unemployment index with UI claims. Incorporating this index in a predictive model significantly enhances the accuracy of UI claim forecasts at national, state, and city levels up to two weeks before official data releases, outperforming professional forecasters, particularly during significant changes in unemployment.

The index can have several important applications. It can inform policymakers of changes in the labor market sooner, up to two weeks in advance, which can be critical for a timely response to economic volatility. The granular geographic resolution of the index can aid decision-makers in devising policies tailored to the places they aim to impact, which can lead to more effective interventions. The index can also uncover measurement errors in official statistics: for example, during one week in May 2020, official UI claims in Connecticut reportedly increased sharply from 36,148 to 298,680, only to be corrected the following week to reflect a slight decrease to

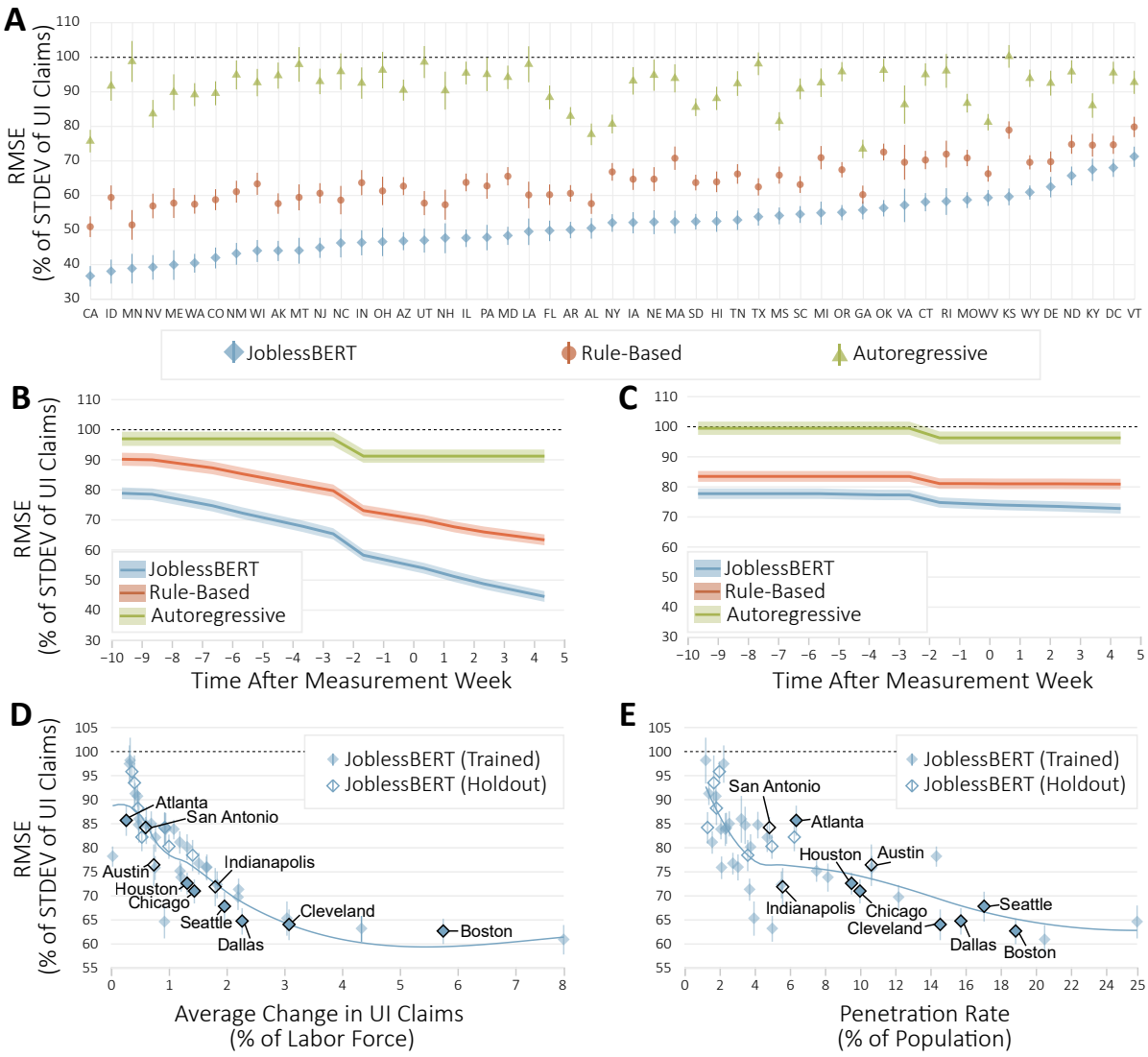

Fig. 3: Sub-national predictions.

(A) RMSE of state-level predictions by state. **(B)** RMSE of state-level predictions as a function of the forecasting horizon. **(C)** RMSE of city-level predictions (trained cities) as a function of the forecasting horizon. **(D)** RMSE by city as a function of the average change in UI claims. **(E)** RMSE by city as a function of the Twitter penetration rate.

the level of 30,046 claims (*8*). The absence of such a spike in the JoblessBERT index could have assured government officials, and perhaps even the market, that the official measurement was off. Moreover, sensing changes in particular geographical locations using social media data could help surveyors decide to shift their samples to areas where changes are happening, leading to more accurate estimates and tighter error bounds. Finally, social protection agencies could use these social media indicators of unemployment to advertise training programs to relevant audiences and help connect job seekers with relevant opportunities.

More generally, the approach used in this work bears promise for other forecasting or estimation tasks that might benefit from the aggregation of public opinion. The general approach of iterative training of an LLM to capture a more diverse set of linguistic forms can be applied to identify other rare forms of expression such as symptoms of a relatively rare disease or hateful and violent speech. The relatively simple adjustment and prediction methods used in this study provide an interpretable solution that facilitates the inspection of model outcomes and any anomalous predictions it may generate. As such, this modeling approach can potentially supplement additional estimation tasks in other fields including economics (e.g., inflation expectations, perception of policy-related outcomes), public health (e.g., rise in particular symptoms), politics (e.g., candidate support), environmental protection (e.g., climate change awareness), and more.

Limitations and future directions. The current study uses simple aggregation methods to construct indicators and parsimonious linear models to construct predictions. Richer time series models with additional predictors and non-linearities can further improve the predictions. As social media becomes progressively more prevalent, network effects could increase users' incentives to signal their unemployment status (*34, 35*), and in turn decrease forecast errors even further. This study focuses on English posts in the U.S., but the same statistical approach can

be applied to other languages, particularly local languages spoken in developing countries. The added value of our approach is potentially high in countries whose statistical agencies lack the resources to regularly collect reliable labor market data (*4, 36*). This approach could be replicated on other social media platforms with better coverage in these developing countries, such as Facebook. Of course, the social benefits of the approach laid out in this work depend on the availability of large-scale social media data, and it remains an open question whether social media platforms will continue to provide such access in the future.

Data availability

All data necessary to evaluate the final conclusions in the paper are publicly available and stored at: `https://github.com/dqlee2/twitter_unemployment`. For restricted-access data, please email D.L. (dql204@nyu.edu) or M.T. (manuel.tonneau@oii.ox.ac.uk) or S.P.F. (sfraiberger@worldbank.org).

Code availability

All data and code necessary to evaluate the final conclusions in the paper are publicly available and stored at: `https://github.com/dqlee2/twitter_unemployment`

References

1. M. P. Bitler, H. W. Hoynes, D. W. Schanzenbach, The Social Safety Net in the Wake of COVID-19. *Brookings Papers on Economic Activity* pp. 119–145 (2020).

2. R. Barnichon, A. Figura, Labor Market Heterogeneity and the Aggregate Matching Function. *American Economic Journal: Macroeconomics* **7**, 222 (2015).

3. D. Card, J. Kluve, A. Weber, Active Labour Market Policy Evaluations: A Meta-Analysis. *The Economic Journal* **120**, F452 (2010).

4. S. Devarajan, Africa's Statistical Tragedy. *Review of Income and Wealth* **59**, S9 (2013).

5. W. Bank, *World Development Report 2021: Data for Better Lives*, World Development Report (World Bank Publications, 2021).

6. R. Chetty, J. N. Friedman, M. Stepner, O. I. Team, The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data. *The Quarterly Journal of Economics* **139**, 829 (2023).

7. N. Woloszko, Nowcasting with panels and alternative data: The OECD weekly tracker. *International Journal of Forecasting* **40**, 1302 (2024).

8. W. D. Larson, T. M. Sinclair, Nowcasting Unemployment Insurance Claims in the Time of COVID-19. *International Journal of Forecasting* **38**, 635 (2022).

9. J. Palotti, *et al.*, Monitoring of the Venezuelan Exodus through Facebook's Advertising Platform. *PLOS ONE* **15**, 1 (2020).

10. Y. Kryvasheyeu, *et al.*, Rapid Assessment of Disaster Damage using Social Media Activity. *Science advances* **2**, e1500779 (2016).

11. R. Chetty, *et al.*, Social Capital I: Measurement and Associations with Economic Mobility. *Nature* **608**, 108 (2022).

12. R. Chetty, *et al.*, Social Capital II: Determinants of Economic Connectedness. *Nature* **608**, 122 (2022).

13. J. Bollen, H. Mao, X. Zeng, Twitter Mood Predicts the Stock Market. *Journal of computational science* **2**, 1 (2011).

14. M. Zamani, H. A. Schwartz, Using Twitter Language to Predict the Real Estate Market. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics* **2**, 28 (2017).

15. J. S. Becerra, A. Sagner, Twitter-Based Economic Policy Uncertainty Index for Chile. *Revista de Análisis Económico* **38**, 41 (2023).

16. C. Angelico, J. Marcucci, M. Miccoli, F. Quarta, Can We Measure Inflation Expectations using Twitter? *Journal of Econometrics* **228**, 259 (2022).

17. J. L. Toole, *et al.*, Tracking employment shocks using mobile phone data. *Journal of The Royal Society Interface* **12**, 20150185 (2015).

18. A. Llorente, M. Garcia-Herranz, M. Cebrian, E. Moro, Social Media Fingerprints of Unemployment. *PLOS ONE* **10**, 1 (2015).

19. D. Antenucci, M. Cafarella, M. Levenstein, C. Ré, M. D. Shapiro, Using Social Media to Measure Labor Market Flows., *Working Paper 20010*, National Bureau of Economic Research (2014).

20. D. Proserpio, S. Counts, A. Jain, The Psychology of Job Loss: Using Social Media Data to Characterize and Predict Unemployment. *Proceedings of the 8th ACM Conference on Web Science* pp. 223–232 (2016).

21. R. Gordon, Green Shoot or Dead Twig: Can Unemployment Claims Predict the End of the American Recession?, *VoxEU Column*, Center for Economic and Policy Research (2009).

22. D. Lewis, K. Mertens, J. H. Stock, U.S. Economic Activity During the Early Weeks of the SARS-Cov-2 Outbreak., *Working Paper 26954*, National Bureau of Economic Research (2020).

23. D. Borup, E. C. M. Schütte, In Search of a Job: Forecasting Employment Growth Using Google Trends. *Journal of Business & Economic Statistics* **40**, 186 (2022).

24. P. C. Zheng, Predicting Unemployment from Unemployment Insurance Claims. *Indiana Business Review* **95**, 1 (2020).

25. D. Aaronson, *et al.*, Forecasting Unemployment Insurance Claims in Realtime with Google Trends. *International Journal of Forecasting* **38**, 567 (2022).

26. B. Settles, Active learning literature survey., *Tech. Rep. 1648*, University of Wisconsin-Madison Department of Computer Sciences (2009).

27. Bloomberg L.P., Economic Forecasts. (2022). Accessed through the Bloomberg Terminal on 7/5/2022.

28. M. Tonneau, D. Adjodah, J. Palotti, N. Grinberg, S. Fraiberger, Multilingual Detection of Personal Employment Status on Twitter. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics* **1**, 6564 (2022).

29. M. Burtsev, *et al.*, DeepPavlov: Open-Source Library for Dialogue Systems. *Proceedings of ACL 2018, System Demonstrations* pp. 122–127 (2018).

30. A. Mislove, S. Lehmann, Y.-Y. Ahn, J.-P. Onnela, J. N. Rosenquist, Understanding the Demographics of Twitter Users. *Proceedings of the International AAAI Conference on Web and Social Media* (2011).

31. A. Aizenman, C. M. Brennan, T. Cajner, C. L. Doniger, J. Williams, Measuring Job Loss during the Pandemic Recession in Real Time with Twitter Data., *Finance and Economics Discussion Series 2023-035*, Board of Governors of the Federal Reserve System (U.S.) (2023).

32. A. Bilal, The Geography of Unemployment. *The Quarterly Journal of Economics* **138**, 1507 (2023).

33. P. Kline, E. Moretti, Place Based Policies with Unemployment. *American Economic Review* **103**, 238 (2013).

34. Y. M. Ioannides, L. Datcher Loury, Job Information Networks, Neighborhood Effects, and Inequality. *Journal of Economic Literature* **42**, 1056 (2004).

35. M. Bhole, A. Fradkin, J. Horton, Information About Vacancy Competition Redirects Job Search., *SocArXiv p82fk*, Center for Open Science (2021).

36. M. Jerven, *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It* (Cornell University Press, 2013).

37. M. Graham, S. A. Hale, D. Gaffney, Where in the world are you? geolocation and language identification in twitter. *The Professional Geographer* **66**, 568 (2014).

38. M. Tonneau, *et al.*, *Proceedings of the 8th Workshop on Online Abuse and Harms (WOAH 2024)*, Y.-L. Chung, *et al.*, eds. (Association for Computational Linguistics, Mexico City, Mexico, 2024), pp. 283–311.

39. Z. Wang, *et al.*, Demographic Inference and Representative Population Estimates from Multilingual Social Media Data. *The World Wide Web Conference* p. 2056–2067 (2019).

40. D. Lazer, R. Kennedy, G. King, A. Vespignani, The Parable of Google Flu: Traps in Big Data Analysis. *Science* **343**, 1203 (2014).

41. L. Liu, H. R. Moon, F. Schorfheide, Panel Forecasts of Country-Level Covid-19 Infections. *Journal of Econometrics* **220**, 2 (2021). Pandemic Econometrics.

42. F. X. Diebold, R. S. Mariano, Comparing Predictive Accuracy. *Journal of Business & Economic Statistics* **13**, 253 (1995).

43. H. Ledford, Researchers Scramble as Twitter Plans to End Free Data Access. *Nature* **614**, 602 (2023).

44. W. Wang, D. Rothschild, S. Goel, A. Gelman, Forecasting Elections with Non-Representative Polls. *International Journal of Forecasting* **31**, 980 (2015).

Acknowledgments

The study's protocol was reviewed by NYU's IRB and was determined non-human participant research. We thank A. Grishchenko for outstanding visual design work; M. Bailey and D. Lazer for useful discussions and feedback. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development / World Bank and its affiliated organizations or those of the Executive Directors of the World Bank or the governments they represent.

Author contributions

Conceptualization: N.G. and S.P.F. Methodology: D.L., M.T., N.G., and S.P.F. Investigation: D.L., M.T., N.G., and S.P.F. Visualization: D.L., M.T., N.G., and S.P.F. Project administration:

D.L., M.T., B.S., N.G., and S.P.F. Supervision: N.G. and S.P.F. Writing—original draft: D.L., M.T., B.S., N.G., and S.P.F. Writing—review and editing: D.L., M.T., B.S., N.G., and S.P.F.

Competing interests

The authors declare no competing interests.