CASCADE ANALYTICS

Microbusiness Density

Anthony Pérez Eisenbarth and Brian Pérez Eisenbarth

Abstract

We study the density of microbusiness activity in local labor markets across the United States. Making use of data provided by GoDaddy's Venture Forward Project, we forecast the density of microbusiness activity across 3,150 counties, 949 core based statistical areas (CBSAs), and 50 states. Our methods combine the approaches of forecast combination and hierarchical forecasting to reach a Symmetric mean absolute percentage error (SMAPE) of 2.8%

January 9, 2023

aperez@cascade-analytics.com, bperez@cascade-analytics.com

Table of Contents

1 Introduction

In developed countries, microbusinesses (those employing fewer than 10 people) and home-based businesses have been systematically overlooked in urban economic development thinking.

2 Data

In 2018, GoDaddy began a project to analyze the economic impact of the microbusinesses our GoDaddy customers created across the U.S. GoDaddy establishes microbusiness density by measuring the number of GoDaddy-registered domains with active websites in a geographic region; these regions are aggregated results the Core-Based Statistical Area (CBSA), County and State level. The distinguishing characteristic of a microbusiness is a website. Some small businesses with websites are included in the count of microbusinesses. But the dataset also includes many businesses that operate online only, that are side gigs for their owners, or that have never registered with a government entity, and thus are not included in official measures of small businesses.

3 Methods

3.1 Forecast Models

In this section we briefly review the forecasting models to be combined. Our focus is on forecasts of the microbusiness activity as measured by microbusiness density and the number of active microbusiness firms at forecast horizons between 1 and 12 months. The following model series were considered for the model pool utilized in the production of forecasts:

- Naive models
- AR models
- Random walk models
- Exponential Smoothing (ETS), which selects the best model underlying one of the fifteen exponential smoothing methods based on the minimization of a pre-specified information criterion [13].
- ARIMA models with parameter space search implemented by the `Fable` and `forecast` packages in `R` [12].
- Theta models following the method introduced by [1] for the M3 Competition.
- Vector Autoregression (VAR) models
- Linear time series models, linear regression fits of a series on various types of a trend (constant, linear, logarithmic) [stats::lm]

3.2 Loss Function

Throughout the analysis, we let Y be the random variable that generates the value to be forecast. To begin with, we restrict attention to point forecasts, f, which are functions of the available data at the time the forecast is made. Hence, if we collect all relevant information at the time of the forecast into the outcome z of the random variable Z, then the forecast is $f(Z)$. Exactly how z maps into the forecast $f(z)$ depends on a set of unknown parameters, Θ, that typically have to be estimated from data, in which this dependency can be emphasized by writing $f(z, \Theta)$.

The loss function is a function $\mathcal{L}(f, Y, Z)$ that maps the data, Z, outcome, Y, and forecast, f, to the real number line: for any set of values for these random variables the loss function returns a single number. The loss function describes in relative terms how bad any forecast might be given the outcome and possibly other observed data that accounts for any state dependence in the loss.

Forecasts are generally viewed as "poor" if they are far from the observed realization of the outcome variable. However, point forecasts aim to estimate not the realization of the outcome but rather a function of its distribution [9, 2]. For example, in the case of mean squared loss, this is the variance of Y given Z plus the squared difference between the forecast and the mean of Y given Z. Both the conditional mean and variance of Y are functions of Θ and z.

The error measure used in the GoDaddy Kaggle Competition for the case of point forecasts is the Symmetric Mean Absolute Percentage Error (sMAPE). This is defined

CASCADE
ANALYTICS

as,

$$\text{SMAPE} = \frac{2}{n} \sum_{t=1}^{n} \frac{|A_t - F_t|}{|A_t| + |F_t|} \tag{3.1}$$

sMAPE essentially measures the normalized absolute differences between the predicted and actual values of the series. As a measure it was introduced the overcome the drawbacks of the Mean Absolute Percentage Error (MAPE).[1] sMAPE uses percentage errors that are scale independent, intuitive to understand and part of an everyday vocabulary.

3.3 Combination Forecasts

Rather than rely upon a single model to produce forecasts, we utilize the forecast combination approach to combine a variety of forecast models. Combining forecasts from multiple models often outperforms forecasts from a single model [6, 8]. Forecast combinations have flourished remarkably in the forecasting community and, in recent years, have become part of the mainstream of forecasting research and activities. Combining multiple forecasts produced from a single (target) series is now widely used to improve accuracy through the integration of information gleaned from different sources, thereby mitigating the risk of identifying a single "best" forecast. Combination schemes have evolved from simple combination methods without estimation, to sophisticated methods involving time-varying weights, nonlinear combinations, correlations among components, and cross-learning. They include combining point forecasts and combining probabilistic forecasts.[2]

The initial idea of forecast combination methods dates to the late 1960s, perhaps most notably to [3]. An extensive literature has built upon the topic, which is comprehensively reviewed and summarized in [8] and by [28]. For any variable of interest there are (at least) three different ways of forecasting using combinations: (a) combining forecasts from different univariate models, (b) decomposition of the series into components which are modeled and recombined, and (c) modeling with different sets of predictors and combining the forecasts.

The empirical work supporting the validity of combining forecasts and its superior performance to single selection has been consistently proven throughout the past decades [16, 17]. [5] provided an extensive bibliographical review of the early work on the combination of forecasts, and then addressed the issue that the arithmetic means often dominate more refined forecast combinations. [18] concluded empirically that a larger number of individual methods included in the simple average scheme would help improve the accuracy of combined forecasts and reduce the variability associated with the selection of methods. While [29] tested five early procedures for estimating weights and found that results pointed towards a simple unweighted average as the superior choice. [26] examined the performance of several univariate models for the prediction of US economic time series, finding that combinations of simple univariate

---

[1]Mean Absolute Percentage Error (MAPE) is defined as $\frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$. sMAPE has been selected in the past for evaluating the submissions of the M3 and M4 Competitions.

[2]The vast majority of studies on combining multiple forecasts have dealt with point forecasting, even though point forecasts (without associated measures of uncertainty) provide insufficient information for decision-making.

CASCADE
ANALYTICS

models perform better than any other single approach. Other examples include the combination of forecasts from a particular family of models (see, for example [15]) and simple or weighted combinations across multiple different families [23, 20].

Decomposition-based methods include the temporal hierarchies as described above, and the theta method where the seasonally-adjusted signal is separated into long and short-term components [1]. In the same category we could place bagging-related approaches [4, 22], where the original series is used to produce several bootstrap series; these bootstraps are forecasted independently (using the same or different models), and, finally, the forecasts from all bootstraps are re-combined. Rapach et al. (2010) is an example of using an array of different predictors, and [7] take approach (c) to a logical (frequentist) conclusion and compute and average over forecasts using all possible subsets of models from a group of predictors.

We work with approach (a), combining the forecasts from univariate and multivariate models described in 3.1. Typically, the combined forecast is constructed as a linear combination of the individual forecasts [21, 6, 23] which can be written

$$Y_{k,t} = \sum_{k=1}^{K} \omega_k \hat{y}_{k,t} = \boldsymbol{\omega}_t' \hat{\mathbf{y}}_t \tag{3.2}$$

where $\hat{\mathbf{y}}_t$ is the column vector of one-step-ahead forecasts at time t (y_{t+1}) produced by the k-th forecasting method, and $\boldsymbol{\omega}_t'$ is the column vector of weights for the set of K forecasting methods. The weights $\omega_{k,t}$ will generally depend on the historical accuracy of base forecasts. Therefore when forecasting at time t, we use all observations prior to to estimate both base forecast model parameters and forecast weights. In the methods described below, weights obtained in forecasting at time t are also utilized in forecasting multiple steps ahead.

In a seminal paper, [11] set out the foundations of optimal forecast combinations under symmetric and quadratic loss functions. They showed that under mean squared error (MSE) loss the optimal weights can be estimated through an ordinary least squares (OLS) regression of the target variable on a vector of forecasts plus an intercept to account for model bias. If the loss function differs from MSE, then the computation of optimal weights may require methods other than a simple OLS regression, An intercept term is often included following the suggestion by [11] to ensure that the bias of the forecast is optimally determined. This formulation is expressed as,

$$Y_{k,t} = \omega_0 + \boldsymbol{\omega}_t' \hat{\mathbf{y}}_t + \epsilon_{t+1} \tag{3.3}$$

Assuming that the regression coefficients in (3.1) are constant, the intercept and weights can be estimated by OLS using an expanding window of the data. It is perhaps more common to use equal weights in the construction of forecasts. Many empirical studies have found that it is difficult to produce more precise forecasts than those generated by such equal weighted combinations [10]. [5] argues that this equal weighting of component forecasts is often the best strategy in this context. This is still true three decades later and is referred to as the "forecast combination puzzle," a term coined by [27]. A rigorous attempt to explain why simple average weights often outperform more sophisticated forecast combination techniques is provided in a simulation study

CASCADE
ANALYTICS

by [25], who ascribe this surprising empirical finding to the effect of finite-sample error in estimating the combination weights.

The simple average method of the forecast combination can be written

$$Y_{k,t} = \frac{1}{k} \sum_{k=1}^{K} f(k) \tag{3.4}$$

where $f(k) = \hat{y}_{k,t}$.

No information is required about the precision of the forecasts or about the dependence among the forecasts to generate a simple average. The treatment of assigning each forecast the same weight, however, implies that the forecasts are treated as being exchangeable [6].

In recent years considerable attention has moved towards the use of probabilistic forecasts [10, 14, 19] as they enable a rich assessment of forecast uncertainties. When working with probabilistic forecasts, issues such as diversity among individual forecasts can be more complex and less understood than combining point forecasts [24].

Simplifying the notation so that $\tilde{Y} = Y_{k,t}$ and $Y_k = \hat{y}_t$, as well as ignoring the time subscripts. The generalized form of combination can be written as

$$
\begin{aligned}
\tilde{Y} &= \sum_{k}^{K} \omega_k Y_k, \qquad Y_k \sim N(\mu_k, \sigma_k^2), \\
\text{where} \quad \tilde{Y} &\sim N(\tilde{\mu}, \tilde{\sigma}^2), \\
\tilde{\mu} &= \sum_{k=1}^{N} \omega_k \mu_k, \\
\tilde{\sigma}^2 &= \sum_{k=1}^{N} \omega_k^2 \sigma_i^2 + 2 \sum_{1 \le k} \sum_{<j \le N} \omega_k \omega_j Cov(Y_k, Y_j) \\
&= \sum_{k=1}^{N} \omega_k^2 \sigma_k^2 + 2 \sum_{1 \le i} \sum_{<j \le N} \omega_i \omega_j \rho_{ij} \sigma_k \sigma_j
\end{aligned}
\tag{3.5}
$$

ρ_{kj} is the correlation between individual residuals.

CASCADE
ANALYTICS

4 Results

5 Conclusion

6 References

[1] V. Assimakopoulos and K. Nikolopoulos. The theta model: A decomposition approach to forecasting. *International Journal of Forecasting*, 16(4):521–530, 2000. The M3- Competition.

[2] D. K. Barrow and N. Kourentzes. Distributions of forecasting errors of forecast combinations: Implications for inventory management. *International Journal of Production Economics*, 177:24–33, 2016.

[3] J. M. Bates and C. W. J. Granger. The combination of forecasts. *Operational Research Quarterly*, 20(4):451, Dec. 1969.

[4] C. Bergmeir, R. J. Hyndman, and J. M. Benítez. Bagging exponential smoothing methods using stl decomposition and box–cox transformation. *International Journal of Forecasting*, 32(2):303–312, 2016.

[5] R. T. Clemen. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4):559–583, 1989.

[6] R. T. Clemen and R. L. Winkler. Combining economic forecasts. *Journal of Business & Economic Statistics*, 4(1):39–46, 1986.

[7] G. Elliott, A. Gargano, and A. Timmermann. Complete subset regressions with large-dimensional sets of predictors. *Journal of Economic Dynamics and Control*, 54(1):86–110, 2015.

[8] G. Elliott, C. W. J. Granger, and A. Timmermann, editors. *Handbook of Economic Forecasting*. Number 24 in Handbooks in Economics. Elsevier North-Holland, Amsterdam; Boston, 1st ed edition, 2013. OCLC: ocm70063087.

[9] G. Elliott and A. Timmermann. Economic forecasting. *Journal of Economic Literature*, 46(1):3–56, March 2008.

[10] W. P. Gaglianone and L. R. Lima. Constructing optimal density forecasts from point forecast combinations. *Journal of Applied Econometrics*, 29(5):736–757, 2014.

[11] C. W. J. Granger and R. Ramanathan. Improved methods of combining forecasts. *Journal of Forecasting*, 3(2):197–204, 1984.

[12] R. J. Hyndman and Y. Khandakar. Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software*, 27(3):1–22, 2008.

[13] R. J. Hyndman, A. B. Koehler, R. D. Synder, and S. Grose. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3):439–454, 2002.

[14] G. Kapetanios, J. Mitchell, S. Price, and N. Fawcett. Generalised density forecast combinations. *Journal of Econometrics*, 188(1):150–165, 2015.

[15] N. Kourentzes, D. Barrow, and F. Petropoulos. Another look at forecast selection and combination: Evidence from forecast pooling. *International Journal of Production Economics*, 209:226–235, Mar. 2019.

[16] S. Makridakis and M. Hibon. The m3 competition: Results, conclusions and implications. *International Journal of Forecasting*, 16(4):451–479, 2000.

[17] S. Makridakis, E. Spiliotis, and V. Assimakopoulos. The m4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1):54–74, 2020.

[18] S. Makridakis and R. L. Winkler. The combination of forecasts. *Management Science*, 29(9):987–996, 1983.

[19] G. M. Martin, R. Loaiza-Maya, W. Maneesoonthorn, D. T. Frazier, and A. Ramírez-Hassan. Optimal probabilistic forecasts: When do they work? *International Journal of Forecasting*, 38(1):384–406, 2022.

[20] P. Montero-Manso, G. Athanasopoulos, R. J. Hyndman, and T. S. Talagala. Fforma: Feature-based forecast model averaging. *International Journal of Forecasting*, 36(1):86–92, 2020. M4 Competition.

[21] P. Newbold and C. W. J. Granger. Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society. Series A (General)*, 137(2):131–165, 1974.

[22] F. Petropoulos, R. J. Hyndman, and C. Bergmeir. Exploring the sources of uncertainty: Why does bagging for time series forecasting work? *European Journal of Operational Research*, 268(2):545–554, July 2018.

[23] F. Petropoulos and I. Svetunkov. A simple combination of univariate models. *International Journal of Forecasting*, 36(1):110–115, Jan. 2020.

[24] R. Ranjan and T. Gneiting. Combining probability forecasts. *Journal of the Royal Statistical Society*, 72(1):72–91, 2010.

[25] J. Smith and K. F. Wallis. A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, 71(3):331–355, 2009.

[26] J. Stock and M. Watson. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. Technical Report w6607, National Bureau of Economic Research, Cambridge, MA, June 1998.

[27] J. H. Stock and M. W. Watson. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6):405–430, Sept. 2004.

[28] X. Wang, R. J. Hyndman, F. Li, and Y. Kang. Forecast combinations: An over 50-year review. *International Journal of Forecasting*, 2022.

[29] R. L. Winkler and S. Makridakis. The combination of forecasts. *Journal of the Royal Statistical Society*, 146(2):150–157, 1983.