

**RESEARCH WORKING PAPER SERIES**

**NEW BUSINESS APPLICATIONS DURING THE COVID-19 PANDEMIC**

Katherine Stevens Andersen<sup>I</sup> and Hayden Murray<sup>II</sup>

May 2022

**ABSTRACT**

During 2020 and into 2021, the number of business applications filed each month experienced a great deal of volatility, deviating sharply from historical values. There was a sharp drop in applications in April 2020, followed by a spike in applications in July 2020. In this paper, we examine trends in new business applications over the last two decades using seasonally adjusted data from the U.S. Census Bureau's Business Formation Statistics database. We apply a model that predicts the expected number of business applications for each month in 2020 based on historical data and compares these predictions to the actual number of applications filed in 2020. The results of this analysis indicate that the number of business applications filed in 2020 differed from the values predicted by the model to a statistically significant degree for the period between March and April 2020 and the period between June and December 2020.

Keywords: entrepreneurship, startups, new business applications, predictions, COVID-19

Acknowledgements: Alyse Freilich, Thom Goff, Travis Howe, Jessica Looze, Erin Kenney

*Research Working Papers have not necessarily been peer-reviewed and are made available by the Ewing Marion Kauffman Foundation to share research and encourage discussion. The views and findings expressed herein are those of the authors and do not reflect the official views of the Kauffman Foundation.*

<sup>I</sup> consultant, Ewing Marion Kauffman Foundation

<sup>II</sup> program officer, Ewing Marion Kauffman Foundation





## INTRODUCTION

There is a significant body of research that examines changes in the number of new business applications filed over time and their relationship to employment growth and general economic activity (Bayard et al., 2018; Asturias et al., 2021). Multiple studies, in fact, have found positive correlations, suggesting that new business applications can serve as a signal for business formation and economic activity more broadly.

In 2020 and 2021, during the COVID-19 pandemic, the number of business applications filed each month experienced a great deal of volatility, deviating sharply from historical values. As business applications are often used as an indicator for the general health of the economy, this paper examines these data more closely. We begin with a discussion of overall trends in new business applications. Specifically, we use seasonally adjusted data on businesses that apply for an Employer Identification Number (EIN) from the U.S. Census's Business Formation Statistics database. We then use a model that predicts the expected number of business applications for each month in 2020 based on historical data and compare these predictions to the actual number of applications filed in 2020. The results of this analysis suggest that the recent increase in business applications is historically unprecedented.

### Highlights:

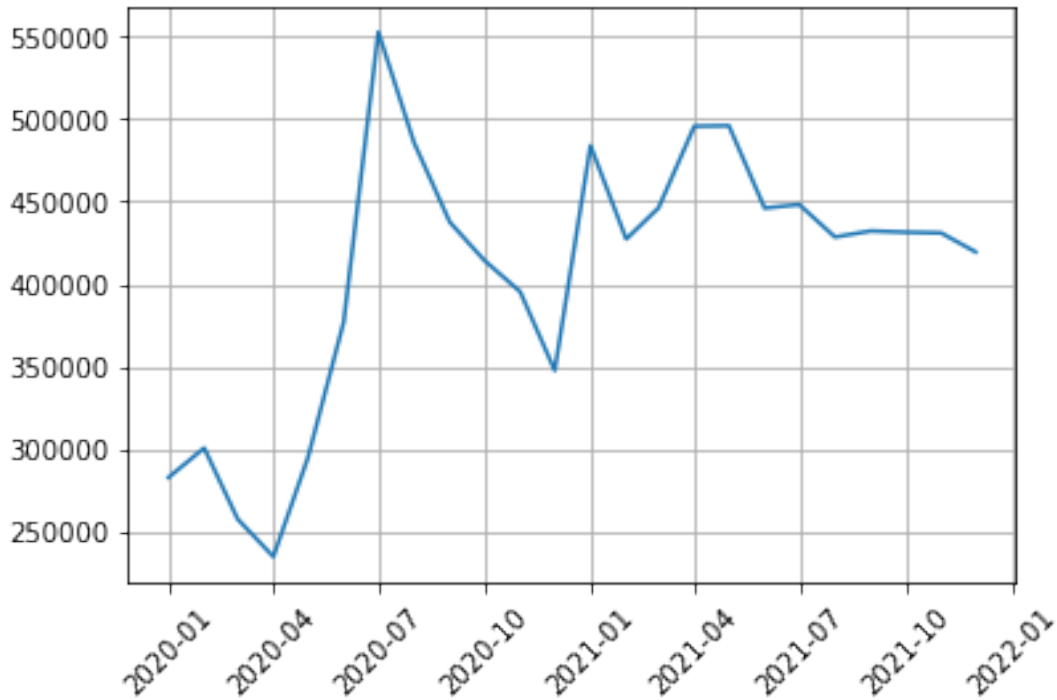
- Between February and April 2020, the number of business applications fell by 21.9%.
- Business applications increased 135% from April to July, and dropped by 37.2% from July to December.
- Overall, applications were up 15.6% in December 2020, compared to where they were in February of that year.
- The number of business applications filed each month in 2020 differed from the values predicted by the model to a statistically significant degree for the period between March and April 2020 and the period between June and December 2020.
- The number of monthly new business applications continued to fluctuate throughout 2021.

## BUSINESS APPLICATIONS DURING THE COVID-19 PANDEMIC<sup>1</sup>

As the COVID-19 pandemic took hold across the U.S. between February and April 2020, business applications fell by 21.9%, from 300,705 in February to 234,838 in April. This was followed by a 135% increase in new business applications from April to July, followed by a 37.2% decrease between July and December. The largest month-over-month decrease during 2020 was from February to March, with new business applications dropping 14.1% from 300,705 to 257,673. The largest month-over-month increase in 2020 followed soon after, with a 46.6% increase from June to July, from 377,081 to 552,748.

Volatility continued throughout 2021. Figure 1 depicts monthly business applications between January 2020 and December 2021.

**Figure 1** Monthly new business applications, 2020–2021



## BUSINESS APPLICATIONS PRIOR TO THE COVID-19 PANDEMIC

In the year before the COVID-19 pandemic began, we saw much less volatility in the number of new business applications each month. In January 2019, for example, there were 277,317 new business applications. Applications then increased 8.3% from January to March, decreased 4.9% between March and May, increased 1.7% from May

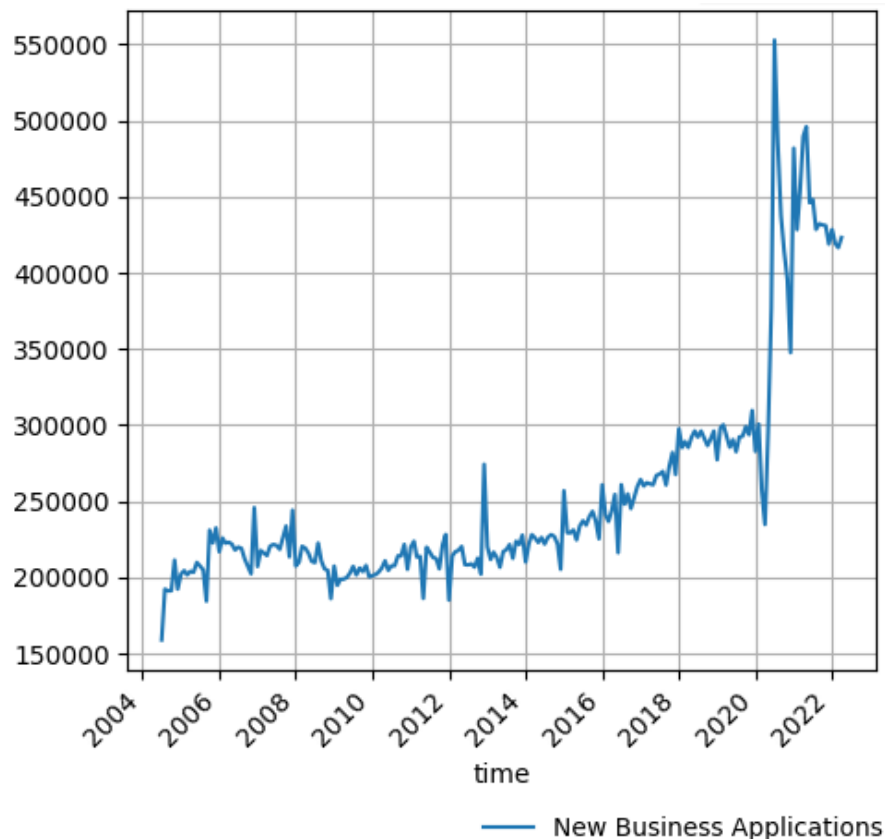
to June, decreased 2.7% from June to July, increased 5.8% from July to October, decreased 1.8% October to November, and decreased 5.4% November to December.

A consideration of the differences in the standard deviation (SD) of new business applications between 2019 and 2020 illustrates this lower level of volatility in 2019. For 2019, the SD of new business applications was 8,580.9. By contrast, the SD of new business applications in 2020 was 96,831.9 (or a coefficient of variation (COV) of 0.27) – the highest for all years of data. For reference, when we calculated the SD and COV of the monthly data for 2020 and compared it to prior years for which we had complete monthly data, the SD was more than 10 times the average for 2005-2019 (and 4.6 times the next highest year, 2012) and the COV was about 6.5 times the average and 2.7 times 2012’s value (see Figure A-1 in the appendix).<sup>2</sup>

Figure 2 illustrates these longer-term trends, presenting monthly business applications from July 2004 – the first month for which data are available – through December 2021. While there has been, overall, a moderate increase in the seasonally adjusted monthly applications over the last two decades, the monthly shifts in new business applications in 2019 are consistent with historical trends.



**Figure 2** Monthly new business applications, 2004–2021



A visual inspection of the data suggests then that the spikes and fluctuations in business application levels in 2020 were unprecedented. Without a formal model however, we cannot confirm or quantify these deviations from the expected values based on historical trends. Moreover, we cannot identify the specific months in which shifts deviated significantly from expectations.

## METHODOLOGY

In order to verify that business application levels in 2020 were, indeed, incommensurate with the values we would expect based on historical trends and to quantify these irregularities, we developed a model to predict monthly business applications for 2020-2021 based on historical values. We adopted an Autoregressive Integrated Moving Average (ARIMA) model, which uses past values of data (in this case, business applications) to predict future values.<sup>3</sup> The model uses seasonally adjusted monthly business applications data for 2004-2019 to produce predictions for 2020.



We developed our model through a process of cross-validation, adjusted as appropriate for time-series applications. For every possible set of model hyper-parameters<sup>4</sup> (within certain bounds<sup>5</sup>), we first generated predictions for each year between 2010 and 2019 using all of the data for the prior years. Comparing these predictions to the actual values for each year, we then generated a score indicating how well that model performed.<sup>6</sup> For example, we would take a model, say of order (2,1, 2) and no trend, and fit it on data from 2004-2009 to predict 2010. We would then fit it on data from 2004-2010 to predict 2011. We would repeat this process for each year. We then averaged these scores to create an overall score for each model. We selected the model – from across dozens of models – with the best overall score to be our final model.

Through this process of cross-validation, we determined that the model with the greatest predictive ability was of order (1,0,1), trend ‘t’, and without seasonality (note that the data was already seasonally adjusted). The formula for such a model is as follows:

$$Y_t = \beta t + \varphi Y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$

where  $Y_t$  is the predicted number of new business applications at time ‘t’,  $t$  is the time period,  $\varepsilon_t$  is the error term at time ‘t’, and  $\beta$ ,  $\varphi$ , and  $\theta$  are parameters estimated by the model. The error term,  $\varepsilon_t$ , is a normally distributed variable with mean 0 and variance  $\sigma^2$ .

Once fit on the data, our model becomes:<sup>1</sup>

$$Y_t = 639.57t + 0.99Y_{t-1} - 0.75\varepsilon_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, 11758.44^2)$$

As a check on our model, we graphed a few aspects of the residuals to ensure that there were no obvious biases or missing elements. While this check does not guarantee a “good” model, issues in these graphs could allow us to identify a “bad” model. As

---

<sup>1</sup> Please note that this equation is specific to our particular model. There is no clean and concise way to represent the general form of an ARIMA model.

desired, our graph of the residuals was mean 0 and displayed no clear patterns.<sup>7</sup> Furthermore, our autocorrelation plot displayed no autocorrelation in the residuals of our model.<sup>8</sup> In other words, our residuals are independent of one another in time, as desired. Autocorrelation would indicate that there is some missing information in the model. See figures A-4 and A-5 in the appendix for additional notes on the model.

## **DEVIATION BETWEEN 2020 MONTHLY BUSINESS APPLICATION NUMBERS AND PREDICTED VALUES**

Table 1 (below) displays the results. When the actual value of business applications falls within the 95% prediction interval (displayed in the “Predicted Interval” column), we classify the result as “expected.” Results that fall below the interval are classified as “low,” and those that fall above the interval are classified as “high.”

**Table 1** New business applications, 2020-2021<sup>9</sup>

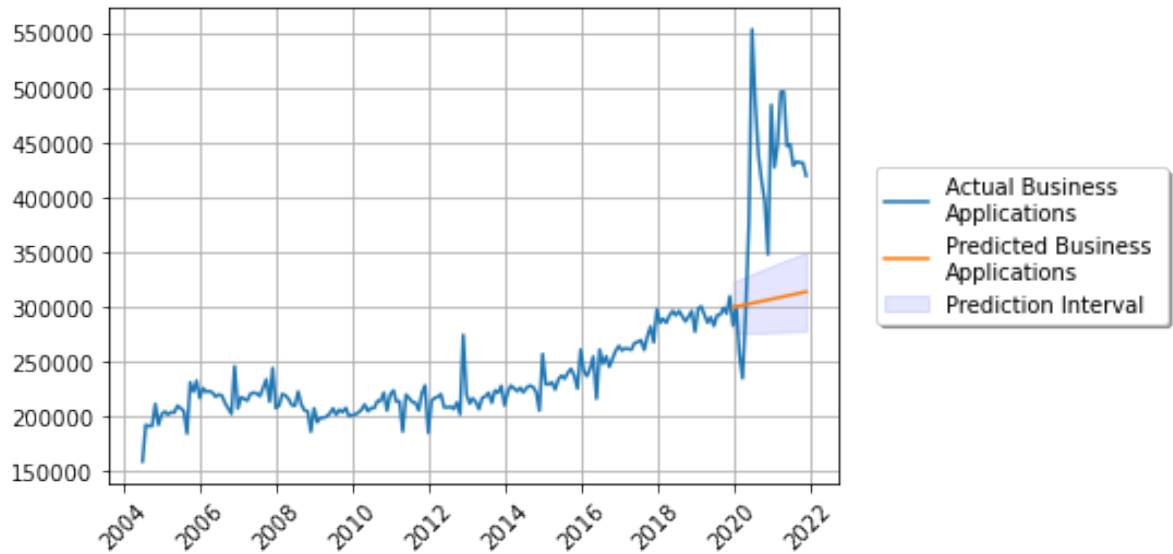
	<b>Predicted Interval</b>	<b>Actual value</b>	<b>Classification</b>
January 2020	276,443.22 – 322,535.31	282,802	Expected
February 2020	276,373.93 – 323,838.53	300,705	Expected
March 2020	276,324.10 – 325,122.30	257,673	Low
April 2020	276,292.18 – 326,388.16	234,838	Low
May 2020	276,276.81 – 327,637.47	294,957	Expected
June 2020	276,276.80 – 328,871.44	377,081	High
July 2020	276,291.09 – 330,091.11	552,748	High
August 2020	276,318.74 – 331,297.42	485,701	High
September 2020	276,358.90 – 332,491.23	437,527	High
October 2020	276,410.91 – 333,673.69	414,335	High
November 2020	276,473.92 – 334,844.69	395,397	High
December 2020	276,547.40 – 336,005.24	347,651	High
January 2021	276,630.66 – 337,155.40	483,857	High
February 2021	276,723.41 – 338,296.65	427,256	High
March 2021	276,825.06 – 339,429.00	446,171	High
April 2021	276,935.19 – 340,552.88	495,621	High
May 2021	277,053.40 – 341,668.68	495,976	High
June 2021	277,179.32 – 342,776.78	446,039	High
July 2021	277,312.61 – 343,877.52	448,197	High
August 2021	277,452.95 – 344,971.21	428,523	High
September 2021	277,600.05 – 346,058.14	432,170	High
October 2021	277,753.64 – 347,138.60	431,427	High
November 2021	277,913.44 – 348,212.84	431,017	High
December 2021	278,079.24 – 349,281.10	419,467	High



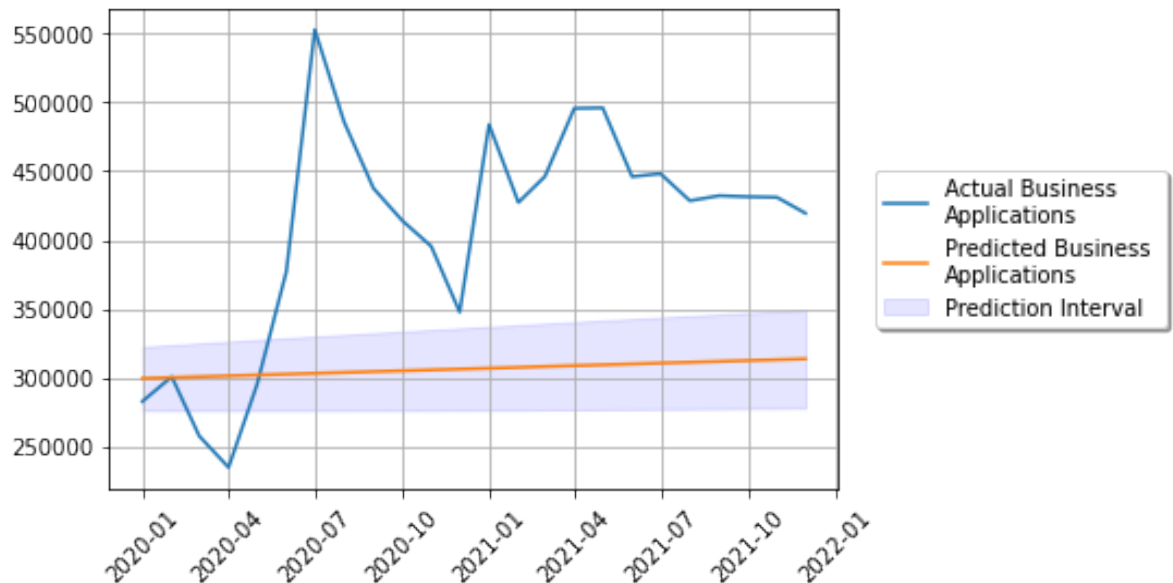


Based on this analysis, business applications grew rapidly from an April 2020 trough to a June spike, (see flip from low to high classification, above). Every month in 2020 after March – with the exception of May– was outside of the prediction interval, suggesting that business applications during these months were historically unprecedented. Figures 3 and 4 present the prediction and actual values of business applications, with Figure 4 focusing specifically on the years affected by the pandemic, 2020 and 2021.

**Figure 3** Predicted vs. actual business applications, 2004-2021



**Figure 4** Predicted vs. actual business applications, 2020-2021

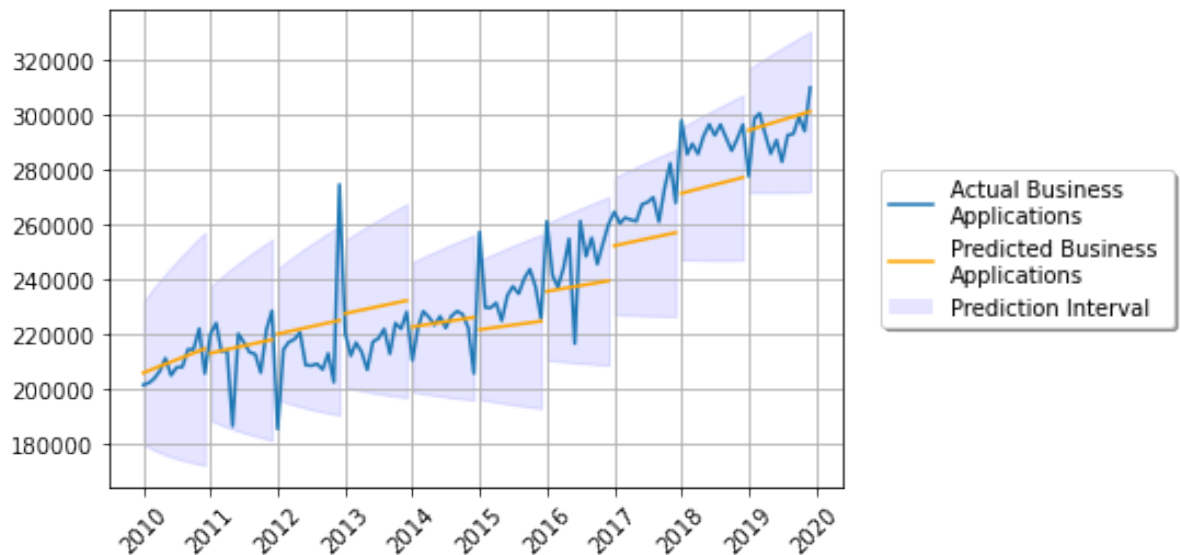




As shown in these figures, there was a dip in March-April 2020 that was outside of the prediction interval, followed by a spike in July 2020. While the number of business applications in May 2020 was technically within the prediction interval, this overlap is simply a function of its timing as the number of business applications grew rapidly between the dip in April and the spike in July. It seems then that it falls inside the prediction interval largely by happenstance. Thus, we can conclude that the business application data after March 2020 exhibit historically unprecedented values, particularly after June 2020.

For comparison, the graph below displays the results if we were to repeat this exercise for each of the years between 2010 and 2020, predicting values for each year based on all the previous years' data (i.e., the 2010 prediction is based on data from 2004-2009, and the 2011 prediction is based on data from 2004-2010). For the years prior to the pandemic, the actual values are almost entirely within the prediction intervals for each year. While there are a few cases in which the actual value falls outside of the prediction interval, these cases are single data points that remain close to the prediction interval. By contrast, there are several consecutive months during 2020 in which the actual number of business applications fall significantly outside the prediction intervals.

**Figure 5** Year-by-year predictions



## CONCLUSION

Based on the results of our model, we can conclude that the changes in the number of applications we saw during the first two years of the COVID-19 pandemic were unprecedented. While the analysis cannot comment on the cause for this change, it is not entirely surprising that there was a significant dip in business applications in the first full month of the pandemic as communities throughout the U.S. were experiencing the



initial shocks of the pandemic – including business and school closures that were profoundly shaping everyday lives. One possible explanation for the sharp increase in applications later in the year is that many people may have been considering entrepreneurial pursuits as a way to compensate for lost income, to replace jobs that were lost, or to capitalize on new opportunities. Keep in mind that this is purely speculation – our analysis, which can only identify the existence of unprecedented volatility, defers to future research to identify the causes.

Specifically, future research will help shed greater light on the drivers of the volatility in the number of new business applications during this period, as well as the motivations of these new business owners, the trajectories of the businesses that were founded, and the extent to which this cohort of new businesses might be qualitatively different from previous cohorts. Such research may also have a variety of implications for policy, depending on the findings. For example, future research may determine the degree to which business applications can or cannot be used as a proxy for the health of the economy during periods of pandemic. In exploring this and other research questions, we will be able to better understand and respond to economic conditions during future periods of uncertainty.



## ABOUT THE DATA

Data for this study are from the Business Formation Statistics (BFS) of the U.S. Census Bureau (<https://www.census.gov/programs-surveys/bfs.html>). The BFS variables used in the calculation of these measures are *business applications* (BA), representing actual values from years 2005-2021. It is worth noting that these data were pulled in January 2022 and are subject to revision by the U.S. Census Bureau in future releases. For more information on the BFS variables, see Bayard et al., 2018 and <https://www.census.gov/econ/bfs/methodology.html>

### Comments on the Model

The results of this model rely on a few assumptions. First, it assumes that the chosen structure is the “right” structure to model business applications. Since ARIMA is specifically designed for modeling a single variable across time based on its own past values, it is a natural fit for our study. It also has the advantage over linear regression that it accounts for the fact that the number of business applications in years that are near each other in time will be more closely related than those in years that are further from each other. Finally, our prediction intervals rely on the assumption that the residuals – or the differences between the predicted values and actual values – are normally distributed. This assumption was easily verified by graphing the residuals.<sup>10</sup>

It is important to note that ARIMA models, in general, allow for a seasonal component. As the data we fed into the model, however, was already seasonally adjusted by the Census,<sup>11</sup> it is not surprising that our best-performing model did not have seasonal components. Furthermore, please note that this model does not identify the reasons for the deviation from expected values. It only confirms that the deviation in business applications data in 2020 from expected values cannot simply be explained by past seasonal patterns or trends and must reflect some new factor that is not captured by the model.

## REFERENCES

- Asturias, J., Dinlersoz, E., Haltiwanger, and J., Hutchinson, R. (2021). “Business Applications as Economic Indicators.” U.S. Census Bureau, Center for Economic Studies, Working Paper CES-21-09.
- Bayard, K., Dinlersoz, E., Dunne, T., Haltiwanger, J., Miranda, J., and Stevens, J. (2018). “Early-stage business formation: An analysis of applications for employer identification numbers.” NBER Working Paper No. 24364.
- Ewing Marion Kauffman Foundation (2020). “Fewer New Businesses Have Become Employers.” *Trends in Entrepreneurship* No. 11. Kansas City, Missouri.



## APPENDIX

**Figure A.1** Mean, standard deviation (SD), and coefficient of variation (COV), 2005-2021

	<b>Mean</b>	<b>SD</b>	<b>COV</b>
2005	209,114.1	13,561.95	0.064854
2006	219,624.1	10,732.74	0.048869
2007	221,291.3	9,803.36	0.044301
2008	210,324.5	9,584.90	0.045572
2009	202,559.8	4,133.02	0.020404
2010	208,317.3	5,991.42	0.028761
2011	214,594.9	10,727.36	0.049989
2012	214,690.1	20,903.87	0.097368
2013	217,623.3	5,860.99	0.026932
2014	222,237.3	7,164.73	0.032239
2015	235,250.6	8,860.92	0.037666
2016	247,960.9	12,727.24	0.051328
2017	266,390.5	6,389.30	0.023985
2018	291,585.0	4,333.60	0.014862
2019	292,875.8	8,580.97	0.029299
2020	365,117.9	96,831.87	0.265207
2021	448,810.1	27,455.77	0.061175

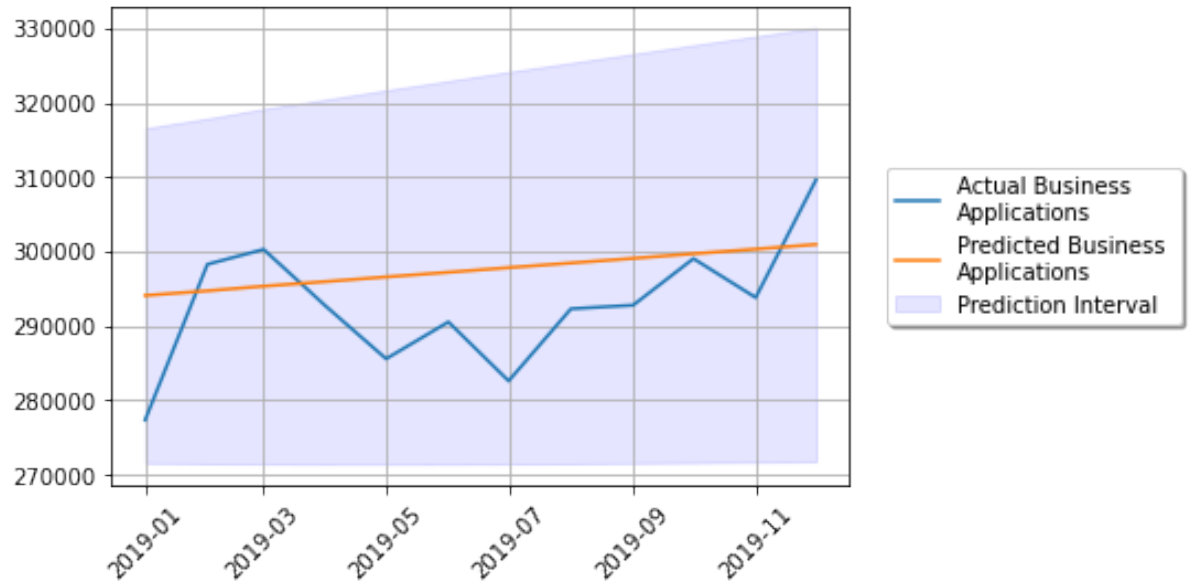


**Figure A.2** Predicted vs. actual new business applications, 2019

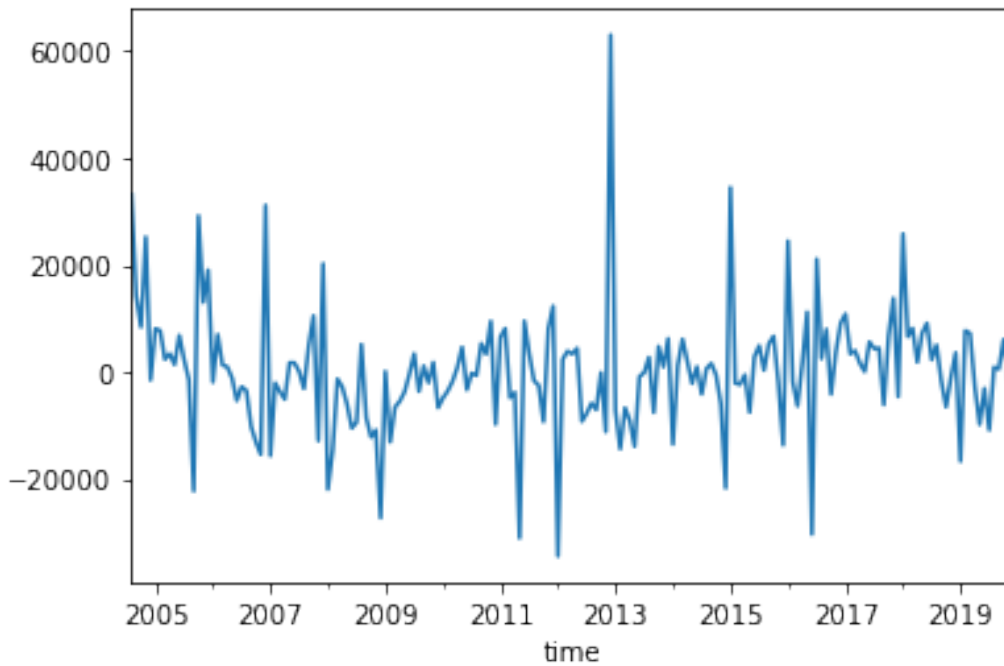
	<b>Predicted value</b>	<b>Actual value</b>	<b>Classification</b>
January 2019	271,548.13 – 316,577.98	277,317	Expected
February 2019	271,487.49 – 317,886.06	298,253	Expected
March 2019	271,446.64 – 319,174.36	300,259	Expected
April 2019	271,423.96 – 320,444.50	292,614	Expected
May 2019	271,418.04 – 321,697.87	285,547	Expected
June 2019	271,427.67 – 322,935.71	290,506	Expected
July 2019	271,451.75 – 324,159.10	282,573	Expected
August 2019	271,489.31 – 325,369.01	292,273	Expected
September 2019	271,539.51 – 326,566.30	292,756	Expected
October 2019	271,601.56 – 327,751.74	299,023	Expected
November 2019	271,674.76 – 328,926.03	293,781	Expected
December 2019	271,758.48 – 330,089.80	309,607	Expected



**Figure A.3** Predicted vs. actual new business applications, 2019 (graphed)

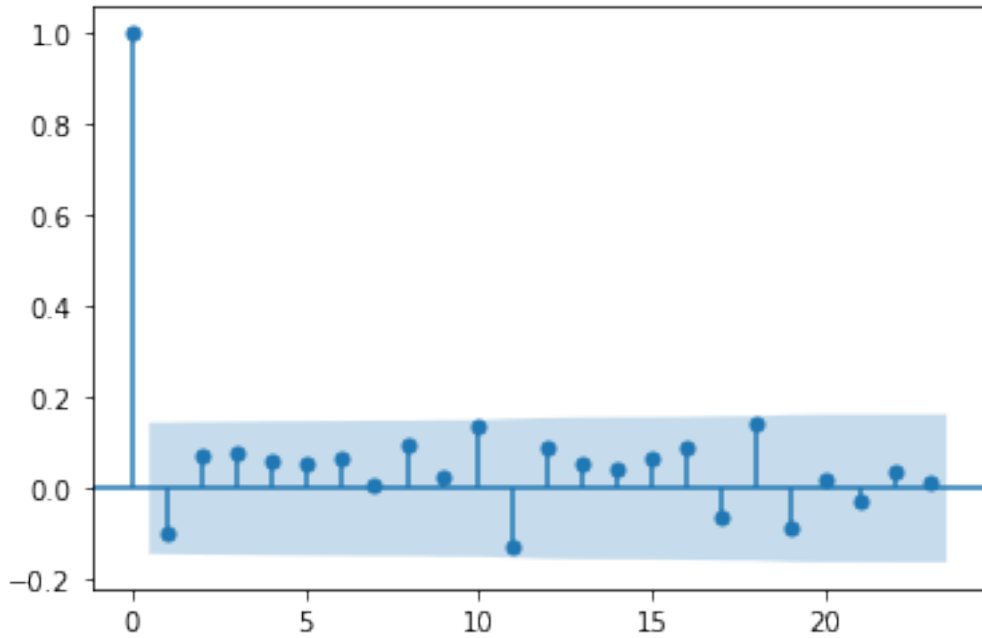


**Figure A.4** Model residuals

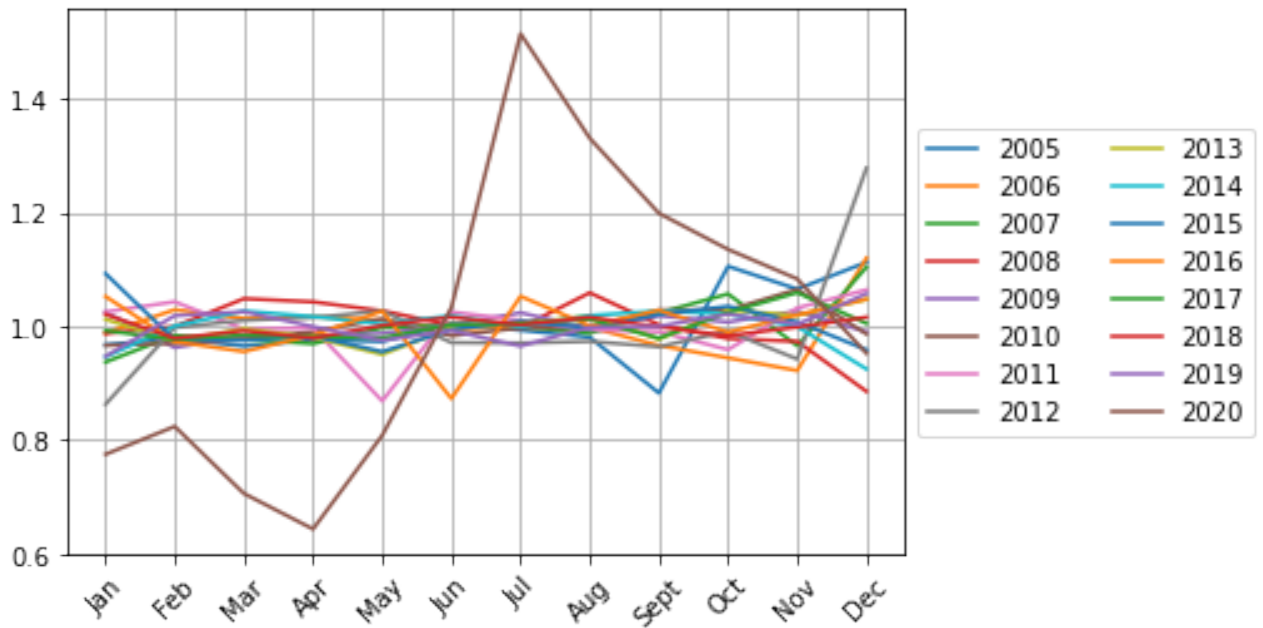




**Figure A.5** Autocorrelation plot of model residuals

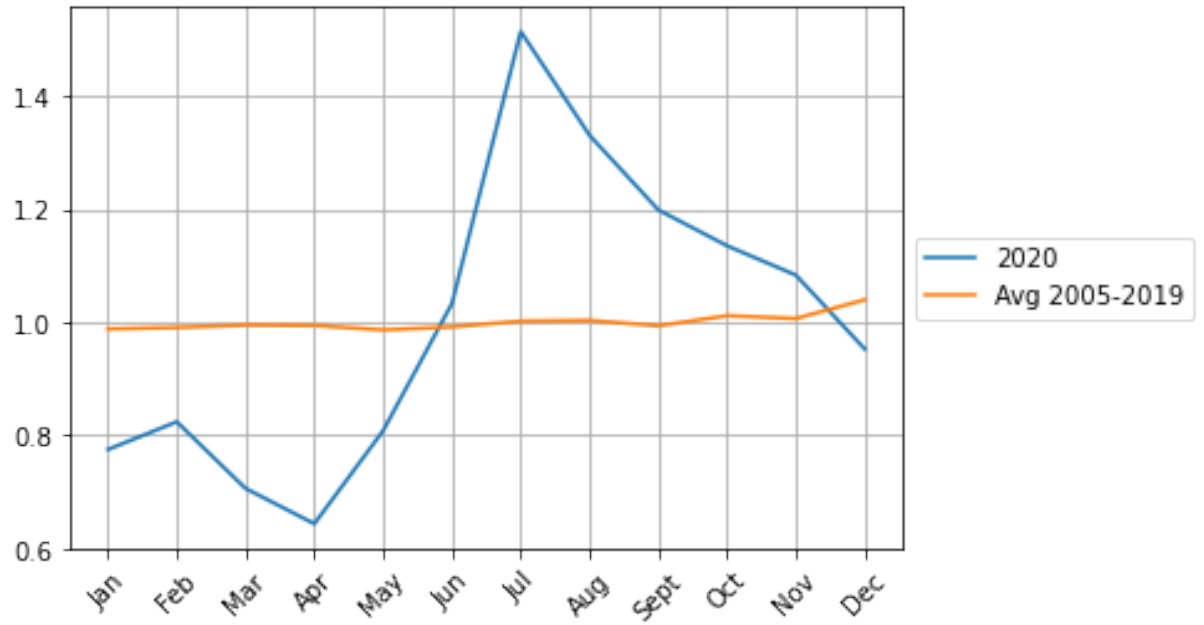


**Figure A.6** Monthly new business applications as percentage of annual average (by year)

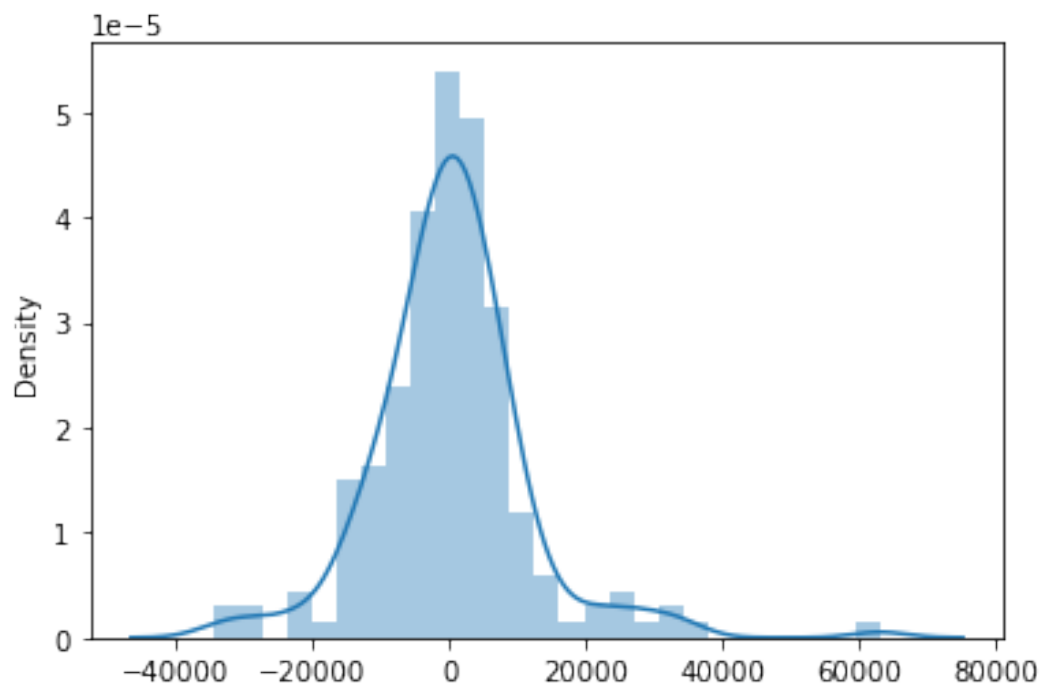




**Figure A.7** Monthly new business applications as percentage of annual average (2020, Avg 2005-2019)



**Figure A.8** Distribution of model residuals





---

<sup>1</sup> “New business applications” refers to the number of applications for an employer identification number (EIN) for a given month. In this brief, we use the Census’s seasonally adjusted version of the data, from July 2004–December 2021.

<sup>2</sup> See Appendix: Figure A.1 for a full table of coefficient-of-variations and standard deviations between 2005 and 2021.

<sup>3</sup> Using a process of cross-validation, we determined that the model with the greatest predictive ability was of order (1,0,1), trend ‘t’, and no seasonality. (Please note that data was already seasonally adjusted.)

<sup>4</sup> This refers to the “order” (p,d,q) of the model, the trend, and seasonal order (P,D,Q). For example, a model could be of order (2,1,2), trend ‘ct’, and seasonal order (1,1,1).

<sup>5</sup> Our search grid went over: 0, 1, 2 for each of p, d, and q. It went over trends: None, ct (constant + trend), t (trend without a constant).

<sup>6</sup> This score was evaluated using Root Mean Squared Error.

<sup>7</sup> See Appendix: Figure A.4.

<sup>8</sup> See Appendix: Figure A.5.

<sup>9</sup> For comparison, see 2019 predictions in Appendix: Figures A.2 and A.3.

<sup>10</sup> See Appendix: Figure A.8

<sup>11</sup> The effectiveness of the Census seasonal adjustments is supported by graphs of monthly business applications. See Appendix: Figures A.6 and A.7.