The Effect of COVID-19 on Altruism: Evidence From Tipping Behavior at goPuff *

Srikar Katta and Regina Ruane
Data Science Institute, Temple University

February 7, 2021

Abstract

Selfless acts of charity are of key importance to policymakers, as it can ensure the efficient allocation of resources to groups in need. If the public is helping certain populations, then government interventions can be used to help populations with less public support. This is especially important during the COVID-19 pandemic where almost everyone is in need of aid and support. However, estimating the causal effect in this scenario poses a difficult challenge because of the lack of untreated groups typically needed to estimate causal effects. Instead, we use time series prediction techniques to forecast a counterfactual and estimate the treatment effect. Using data on tipping behavior from the digital delivery service goPuff, we show that COVID-19 caused an increase in altruistic behavior.

Keywords: COVID-19, altruism, tipping behavior, counterfactual prediction, causal inference

*We would like to thank Ken Alston, Adam Whitaker, and Cory Haberkern from goPuff for providing us with data. Any errors are our own. All correspondences should be sent to Srikar Katta at srikar@temple.edu.

1 Introduction

People often make decisions with regard for others' well being, known as altruism. From individuals acting kind randomly to countries hosting refugees from war-torn countries, altruism exists all around. From a policy perspective, understanding altruistic behavior can guide government regulation: can private organizations/individuals effectively handle the redistribution of income and supply of goods or is the best method for societal growth government regulation (Kennett, 1980)? This is of key importance in times of crisis – such as pandemics – as it can ensure the efficient allocation of resources; groups supported by the public may need less support than the unsupported. Understanding the effect of COVID-19 on altruism can redefine government aid and spending plans: are individuals helping each other or does the government need to intervene to help *everyone*?

In traditional causal-inference lab experiments, researchers treat some units (e.g., people or animals) and do not treat others. So for example, if studying the effect of aspirin on headaches, researchers offer aspirin (i.e., the treatment) to some participants, known as treated units, and not to others, known as untreated units. They would then compare the frequency in headaches between these two groups to find the causal effect (Rubin, 2005). However, due to COVID-19's wide-reaching impacts, *everything* globally was effected by the virus – from lifestyles to international relations – so there are no untreated units or variables could be used to infer causality. Using forecasting techniques and pre-pandemic observations, we can predict the potential outcomes had the virus not happened (i.e., a counterfactual) and infer causal effects. While this is similar to regression discontinuity design frameworks, we are able to estimate beyond immediate treatment effects.

While altruism as a whole is very difficult to measure, a prime example of altruism is tipping behavior: people voluntarily pay more than the minimum price for a service to thank someone. I use data on tipping behavior from the Philadelphia-based, digital delivery service goPuff from May 1, 2019 to May 1, 2020 as a proxy for altruism. I also evaluate different time series forecasting models to identify the best counterfactual predictor and measure the effect of COVID-19 on altruism.

2 Related Literature

Altruism in COVID-19 has been explored by a few researchers, discovering that differences in altruism might dictate different government responses. For instance, Alfaro et al. (2020) found that stringency measures matter less in more altruistic communities, suggesting that increased altruism may indicate more relaxed government regulations. Additionally, Rieger et al. (2020) found that triggering altruism leads to a greater willingness to be vaccinated. Taken together, these studies suggest that changes in altruism can redefine government responses to the pandemic. However, little research has been conducted to evaluate whether COVID-19 actually impacted altruistic behavior.

A key concern for measuring altruism is the lack of a viable proxy. While some studies utilize survey methods (e.g., Vieira et al. (2020)), survey methods lack the robustness that "real world" examples of altruism may suggest. For instance, researchers often use kidney donations as a standard for altruism, as it is truly a selfless act (Marsh et al., 2014). However, collecting such data can be quite difficult, which is why we propose using tipping behavior at goPuff as a proxy for altruistic behavior.

The traditional motivators of tipping behavior are altruism, reward, and duty (Lynn, 2015; Ayres et al., 2004). In fact, Jacob et al. (2013) discovered that exposure to altruism prior to tipping significantly increased tipping behavior, evidence to suggest that altruism constitutes a large portion of the drivers of tipping behavior. However, the other two motivators may suggest that tipping behavior is not a completely selfless act as reward and duty are both rational drivers of the behavior. But tipping behavior may be a viable proxy for impure altruism – the idea that altruistic behavior itself is guided by positive feelings from acting benevolently.

Another difficulty with such a study is measuring the effects of COVID-19. Because of the lack of untreated units, COVID-19-related causal inference poses a significant challenge. However, many researchers have worked around it, utilizing approaches similar to ours. For instance, Shino and Binder (2020) use a regression discontinuity design (RDD) to estimate the effects of government responses to political rallying. Additionally, Dang and Trinh (2020) use RDD to estimate the effects of the pandemic on air quality globally. However, these approaches only allow for immediate treatment effects and require an exogenously determined shock date. To overcome this, we use time series techniques to forecast a counterfactual to estimate treatment effects.

3 Methods

3.1 Estimating Causality

Before exploring the data and detailing modeling approaches, it is best to understand how we estimate causality, an approach very similar to Regression Discontinuity in Time (RDiT) (Hausman and Rapson, 2017)). Suppose Y_{it} represents an observed series of interest for $i = 1, \ldots, N$ units and $t = 1, ..., T$ time periods. Let $Y_{it}(0)$ represent what would have happened if unit i at time t was not treated and $Y_{it}(1)$ represent what would have happened if it was treated. Even though we cannot observe $Y_{it}(0)$ and $Y_{it}(1)$ simultaneously, both are necessary to discover the average causal effect on the treated units in the treatment period, defined as $\tau = \mathbb{E}_{it}[Y_{it}(1) - Y_{it}(0)|Y_{it} = Y_{it}(1)]$.

In traditional causal inference settings, we use data from untreated outcomes after the treatment date, T_0, to estimate $Y_{it}(0)$ (Rubin, 2005). However, if *all* N units are treated after T_0, then utilizing untreated units is no longer viable. To overcome this challenge, RDiT utilizes pre-treatment observations (i.e., $Y_{it}(0)$ when $t < T_0$) to predict $\hat{Y}_{it}(0)$, which should yield valid counterfactual estimates because pre-treatment values have not been impacted by the treatment. In the traditional RDiT setting, we use covariates to predict the counterfactual. However, due to COVID-19's wide reach, identifying unaffected variables is quite difficult, especially in the digital delivery sector after stay-at-home orders. To over come this challenge, we compare different forecasting techniques that utilize *only* historical observations to identify the best counterfactual predictor. We can then estimate the ATT using $\hat{Y}_{it}(0)$ for $T_0 \leq t \leq T$:

$$\hat{\tau} = \frac{\sum_{i=1}^{N} \sum_{t=T_0}^{T} \left(Y_{it}(1) - \hat{Y}_{it}(0)\right)}{N(T - T_0 + 1)}. \tag{1}$$

3.2 Data

We used tipping behavior data collected by the digital delivery service goPuff from May 1, 2019 to May 1, 2020 across the 92 US metropolitan regions that goPuff is active in. Due to goPuff's data sharing policies, we were only allowed to receive data from 50,000 unique customers over this year, which required data selection decision-making prior to actually observing the dataset. Because there is little-to-no pre-COVID-19-data on the tipping behavior of people who joined/increased activity during the COVID-19-period to predict a counterfactual, we filtered out any customers who were not very active (by goPuff's standards) before March 1, 2020 (start of COVID-19 period). Afterwards, we received all of the orders placed by a random sample of 50,000 customers. After receiving the data, we also removed any observations that had no revenue (perhaps because of data collection issues) and observations in which the amount tipped exceeded the amount paid because those are rare and outliers. One important consideration about goPuff's service is that it does not require drivers or customers to share demographic information, so such details cannot be used in our analyses.

On the goPuff app, aftering confirming purchases, the customer is presented with options to tip the delivery driver \$1, \$2, \$3, or none (see Figure 1). Because people are presented with these as default options instead of the traditional tip price in percentage, we tested to see if the amount tipped is associated with revenue, discovering a \$0.09 increase in tips for every extra dollar in order cost (p-value < 0.01; see Table 1). Because of this, all subsequent references to "tips" considers tips as a percentage of revenue rather than a raw dollar amount.

Table 1: Associating Tips and Revenue

	Dependent variable: Tips (\$)
Cost of goods (\$)	0.094***
	(0.0002)
Constant	0.155***
	(0.004)
Observations	664,880
R^2	0.361
Adjusted R^2	0.361
Note: HC1 std. errors	*p<0.1; **p<0.05; ***p<0.01

In order to predict a trustworthy COVID-period counterfactual, the pre-COVID time series need little missing data. However, in over 90% of the data, customers did not order daily or even weekly (average number of pre-COVID orders was 13), limiting our ability to use time series methods for counterfactual prediction. To circumvent the data paucity problem, we averaged tips by metro region and day. We then filtered out observations in metro regions with less than 15 observations for more than 36 days (10% of the data), as that would suggest an insufficient amount of data to make strong claims about tipping behavior for those days. Afterwards, only eleven metro regions remained (see Figure 2). Because people tend to repeat similar orders on the same day of the week (e.g., people order alcoholic beverages on Friday night every week), missing data was interpolated using the daily average percent tip for that metro region from seven days before.

3.3 Counterfactual Modeling and Inference

After all data filtering and aggregation, the dataset was compressed from 600,000 observations to about 4,000 observations. Due to the limited number of observations, we used a cross validation technique made specifically for time-series (Hyndman and Athanasopoulos, 2018). In the first iteration, we used observations from May 1, 2019 until October 30, 2019 as my training dataset and observations from October 30, 2019 to December 30, 2019 as the validation dataset. In the second iteration, we added one day to the training set and shifted the validation set over one day; so, we used observations from May 1, 2019 until October 31, 2019 as my training dataset and observations from October 31, 2019 to December 31, 2019 as the validation dataset. We repeated this process until the validation set bled into the COVID-19 period (after March 1, 2020). See section A.1 for more details.

We compared three forecasting techniques to use for predicting a counterfactual. To establish a baseline, we used the last observation from the training set for each region as the prediction (naive model). We also tested Vector Autoregression (VAR), a statistical technique that captures the relationship between different quantities over time. Considering that each regions' time series is similar, lagged values of tipping behavior from other regions were used as features. We increased the number of lags incrementally until the residuals were uncorrelated with a Portmanteau test (Hyndman and Athanasopoulos, 2018). We also considered neural network autoregression (NNAR), a feed forward neural network with one hidden layer that utilizes lagged values as inputs, with a sum of squares objective function. In the hidden layer, the number of nodes was equal to the floor of half the number of inputs, and the hidden layer used a sigmoid activation function, which does not fall susceptible to the vanishing gradient problem since there is only one hidden layer. See section A.2 for more details on these methods. Predicting N values of quantity y, denoted as $\hat{y}$, models were then evaluated using mean absolute percent error (MAPE) and root mean squared error (RMSE):

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \text{ and } RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{2}$$

After discovering the best model, counterfactual forecasts were made for the COVID-period of March 1, 2020 to May 1, 2020 for each region. First, we stacked predicted and actual COVID-period values, added an indicator for treatment, and ran the following regression with HC1 standard errors to fix for heteroskedasticity, where β_1 represents the percent impact of COVID-19[1]:

$$log(Tips(\%)) \sim \beta_0 + \beta_1\mathbb{I}(Observed) + \varepsilon. \tag{3}$$

4 Results

NNAR had lower MAPE and RMSE scores than the other models (see Table 2). Unsurprisingly, the naive model performed the poorest with a MAPE of 14.35% and RMSE of 0.018. VAR was in-between both. I believe NNAR's performance is due to the regularization term

[1]A clarifying note: Percent-impact refers to the percent change in tips, which are themselves a percentage of order cost. For example, a 25% impact would suggest an increase in average daily tips from 10% to 12.5% and not a shift from 10% to 35%.

that keeps weights small and the fact that inputs were already between 0 and 1. Both of these help prevent overfitting in neural networks, which is explains why NNAR was able to perform better than VAR.

When considering the impact of COVID-19 on tipping behavior, the regression results show an 8% increase in average daily tips across the eleven regions (p-value < 0.01), suggesting COVID-19 caused an increase in altruistic behavior (see Table 3 and Figure 3). These results fit in with the theory that altruism is "born of suffering," a belief that hardships motivate people to act kinder towards one another (Vollhardt, 2009). Such behavior has also been discovered in times of war, where those exposed to greater violence are more responsive to refugee distress (Hartman and Morse, 2015). However, these results contradict survey research that found no change in altruistic behavior caused by COVID-19 at a population level (Vieira et al., 2020). That said, Vieira et al. (2020) was aimed at studying the effect of increased threat from COVID-19 (i.e., increase in regional cases) on altruism, which may explain the different conclusions. Perhaps the impact of COVID-19 on altruism was limited only to the presence of COVID-19 itself in regions and not the number of cases.

4.1 Placebo Test

While the technique used to estimate the causal effect is seemingly intuitive, it is also underutilized because it cannot capture changes due to other interventions. In most policy-related research, there are usually untreated units that can help capture these potential differences; however, this is certainly not the case here. To further prove the validity of this technique, I perform the same set of analyses but with a new 61-day intervention period and a randomly selected start date of October 14, 2019 to see if this study's finds are truly significant or just spurious.

4.1.1 Placebo Test Methods

Using the cross validation technique outlined in section A.1, we evaluate the same three models. In the first iteration, I used observations from May 1, 2019 until June 15, 2019 as my training dataset and observations from June 15, 2019 to August 8, 2019 as the validation dataset (61 day validation period). In the second iteration, I added one day to the training set and shifted the validation set over one day; so, I used observations from May 1, 2019 until June 16, 2019 for my training set and observations from June 16, 2019 to August 9, 2019 as the validation dataset. I repeated this process until the test set bled into the intervention period starting October 14, 2019. The naive baseline, VAR, and NNAR were compared using MAPE and RMSE again.

After discovering the best model, counterfactuals were predicted for the intervention period for each region using March 1, 2019 to October 14, 2019 as the training period. Causal estimates were derived using the regression outlined in equation 3.

4.1.2 Placebo Test Results

Again, NNAR had lower MAPE and RMSE scores than the other models (see Table 2), with the naive prediction as the worst model. Additionally, the $\mathbb{I}(Observed)$ coefficient was not significant with a p-value greater than 0.05 (see table 2 and Figure A.2), suggesting that there was no treatment effect found in this time period, which makes sense considering there were no national interventions in this time period (to my knowledge). This null result is further evidence for this technique's validity, which should be taken advantage of during COVID-19 to guide policy responses.

Table 2: Comparing Forecasting Model Performance For COVID-19 and the Placebo Tests

Test	Model	MAPE	RMSE
COVID-19	Naive	14.358	0.018
	VAR	11.258	0.014
	NNAR	10.473	0.013
Placebo	Naive	14.358	0.018
	VAR	11.258	0.014
	NNAR	10.473	0.013

Table 3: Comparing Actual and Predicted Tips (%) During COVID-19 (Left) and During Placebo Intervention (Right)

	Dep. var.: log(Tips (%))	
	COVID-19 Test	Placebo Test
$\mathbb{I}$(Actual)	0.08^{***}	0.002
	(0.001)	(0.008)
Constant	0.155^{***}	-2.310^{***}
	(0.015)	(0.005)
Note:		*p<0.1; **p<0.05; ***p<0.01

5 Conclusion

This study discovered that COVID-19 *increased* altruistic behavior slightly. Importantly, these discoveries can be applied to various policy domains, especially with regards to relief packages and government aid. Because communities have been found to be more altruistic, governments may be able to reduce the amount of aid needed for gig-economy workers like goPuff delivery drivers and instead focus their efforts on helping unemployed workers and small businesses with no delivery programs.

Also, these results should also be taken cautiously and not applied to the greater public. Because this study utilized data from goPuff (whose target audience is the Millenial generation), these results may not generalize to the greater public, including big businesses, who have the most resources to help the economy recover from COVID-19.

Additionally, there may be some controversy around interpreting causality with this approach. One might suggest the increase in average tips was not a result of COVID-19 but rather a reaction to increased activity because businesses were shut down during this period. However, the root cause of these shut downs was still COVID-19, so as long as the root cause of any other intervention was COVID-19, then these findings can still be causal. Considering COVID-19 greatly impacted everything/everywhere, this is quite plausible.

In the future, researchers should consider other avenues of altruism, especially with regards to charitable behavior resulting from COVID-19. A particularly interesting question may be, "Did COVID-19 increase altruistic behavior for non-COVID-19-prevention related efforts." While many large businesses donated ventilators and masks to hospitals, understanding if these large businesses also increased their support of other charitable organizations could help guide government responses.

Figures

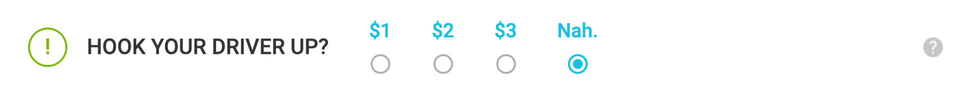

Figure 1: Screen presented to goPuff customers offering people the opportunity to tip drivers.

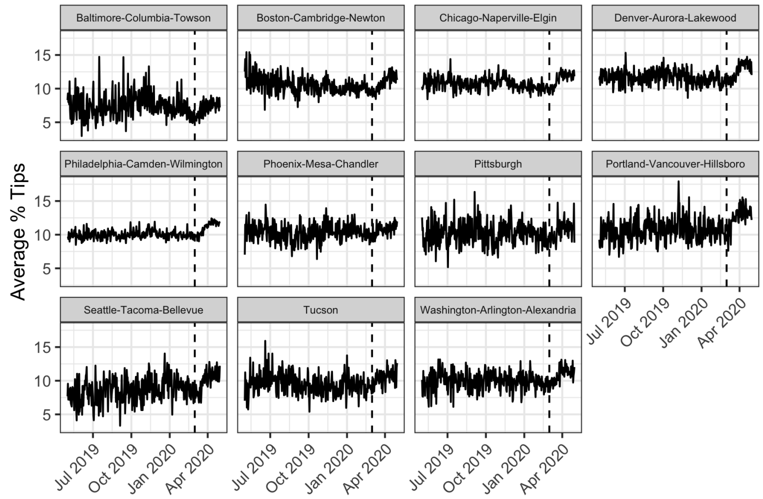

Figure 2: Average daily tips (as a percentage of order cost) per region from May 1, 2019 to May 1, 2020. The dashed line represents the start of COVID-19 – March 1, 2020. After cleaning, only these eleven regions had sufficient data (less than 36 days with less than 15 orders). Even upon visual inspection, it is obvious to recognize a positive shift in average tips after COVID-19 began, especially in the Philadelphia-Camden-Wilmington region.

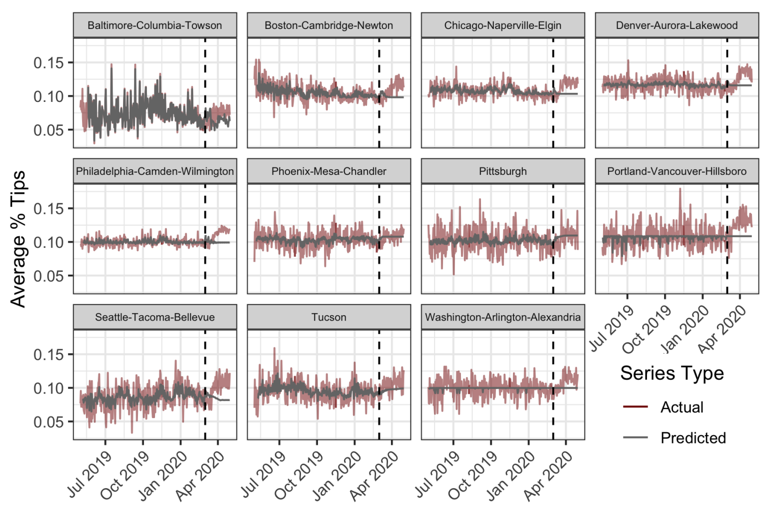

Figure 3: Counterfactual forecasts were developed using NNAR with March 1, 2019 to February 29, 2020 as training set and March 1, 2020 to May 1, 2020 as test set (denoted by dashed line). During COVID-19, average percent tipped *increased* by about 8%, suggesting increased altruism.

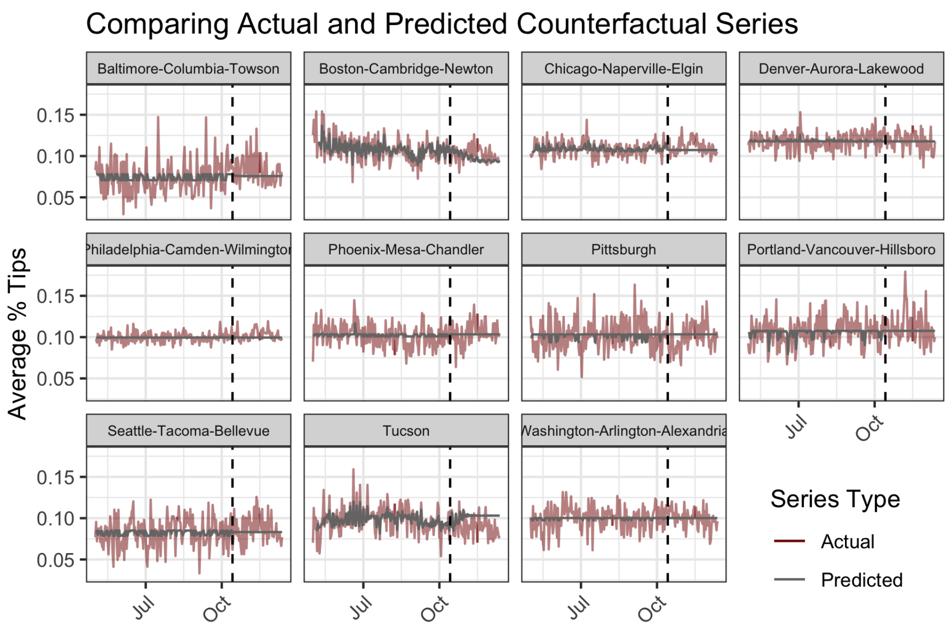

Figure 4: Counterfactual forecasts were developed using NNAR with March 1, 2019 to October 14, 2019 as training set and October 14, 2019 to December 14, 2020 as test set (denoted by dashed line). During the treatment period, average percent tipped did not increase, further validating this technique for estimating causal effects.

A Appendix

A.1 Time Series Cross Validation

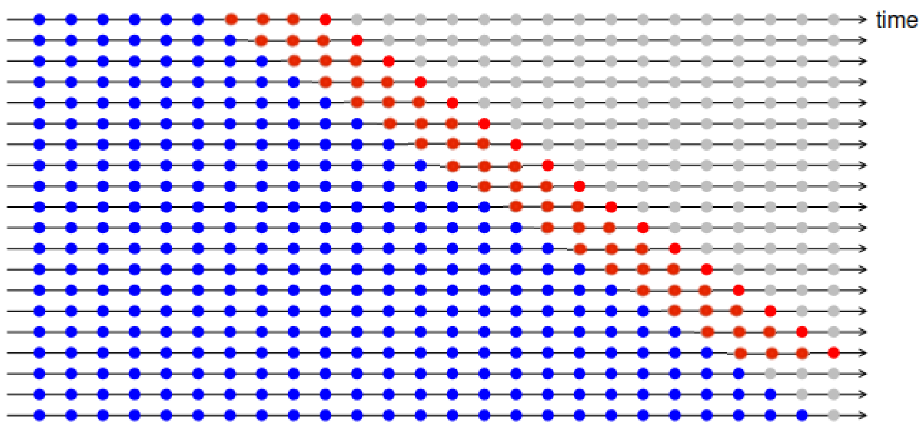

Figure A.1: Suppose we have a time series Y_t with observations for time $t = 1, \ldots, T$. We first use $Y_1, \ldots, Y_{T_1}$ (where $t < T_1 < T$) to predict $Y_{T_1+1}, \ldots, Y_{T_1+k}$. In the next step, we use $Y_1, \ldots, Y_{T_1}, Y_{T_1+1}$ to predict $Y_{T_1+2}, \ldots, y_{T_1+1+k}$. We repeat this $T - T_1 + k + 1$ times. The blue points represent the training data while the red points represent a k step forecast. This image was adapted from (Hyndman and Athanasopoulos, 2018).

A.2 The Vector Autoregression Model

The vector autoregression model (VAR) is a statistical time series model that allows for the prediction of multiple co-dependent time series.[2] Suppose we have two time series X_t and Y_t with time $t = 1, \ldots, T$. In a simple autoregressive model, we use historical observations of X_t to predict future values using linear regression,

$$X_t = \beta_0 + \gamma_0 X_{t-1} + \ldots + \gamma_k X_{t-k} + \varepsilon, \tag{4}$$

where β_0 is the intercept and ε is the error term. However, suppose X_t and Y_t are correlated with one another. Then, it would make sense to use historical observations of Y_t to predict X_t as well,

$$X_t = \beta_0 + \gamma_0 X_{t-1} + \ldots + \gamma_k X_{t-k} + \eta_0 Y_{t-1} + \ldots + \eta_k Y_{t-l} + \varepsilon, \tag{5}$$

[2]All information on VAR comes from (Hyndman and Athanasopoulos, 2018)

where β_0 and ε are intercept and error terms again. Extending this one step further, if we wanted to forecast Y_t as well, then we could run a multivariate regression,

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{bmatrix} \beta^1_{00} & \beta^1_{01} \\ \beta^1_{10} & \beta^1_{11} \end{bmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \ldots + \begin{bmatrix} \beta^k_{00} & \beta^k_{01} \\ \beta^k_{10} & \beta^k_{11} \end{bmatrix} \begin{pmatrix} X_{t-k} \\ Y_{t-k} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}, \tag{6}$$

where the coefficients in the first row of the matrix help predict X_t, the coefficients in the second row of the matrix help predict Y_t, and the superscript on the coefficient represents for which lag that coefficient applies to. This is essentially just an extended linear regression. While there are many ways to decide how many lagged terms to use, one technique is to identify the errors are correlated with itself one time period ago (i.e., $Correlation(\varepsilon_t, \varepsilon_{t-1}) \neq 0$), known as serial correlation. Serial correlation suggests that there is some relationship between the data at time t and time $t-k-1$, where k is the number of lagged terms specified in the regression. So adding another lagged term should improve forecast accuracy.

A.2.1 Stationarity

One key assumption of VAR is that time series have a constant mean and variance over time, known as stationarity. We can statistically test for this using the Dickey-Fuller test, whose null hypothesis is that data are non-stationary (Hyndman and Athanasopoulos, 2018). As seen in Table 4, the data for each metro region before COVID-19 starts have a p-value less than 0.01, suggesting that we can reject the null and say that data are stationary. So, VAR is a viable option.

Table 4: Dickey-Fuller Test for Stationarity Results For Pre-COVID-19 Time Series

Metro Region	Test Statistic	Number of Lagged Terms in Test	P-value
Baltimore-Columbia-Towson	4.602	6	< 0.010
Boston-Cambridge-Newton	4.990	6	< 0.010
Chicago-Naperville-Elgin	4.736	6	< 0.010
Denver-Aurora-Lakewood	5.316	6	< 0.010
Philadelphia-Camden-Wilmington	4.377	6	< 0.010
Phoenix-Mesa-Chandler	5.157	6	< 0.010
Pittsburgh	4.878	6	< 0.010
Portland-Vancouver-Hillsboro	6.598	6	< 0.010
Seattle-Tacoma-Bellevue	4.597	6	< 0.010
Tucson	5.293	6	< 0.010
Washington-Arlington-Alexandria	5.240	6	< 0.010

A.3 NNAR

Neural network autoregression (NNAR) is a one-hidden-layer, feed forward neural network specialized for time series models in which we use lagged values of the time series to forecast future observations. We train the neural network the same as we do a normal feed-forward network; however, in the prediction step, predicted values are also used to forecast other values. For example, suppose we have a time series Y_t with $t = 1, \ldots, T$ time periods. Suppose we want to predict values from $T_1 + 1$, such that $t < T_1 < T$, to T (i.e., predict $Y_{T_1+1}, \ldots, Y_T$). Then, we train the NNAR on $Y_1, \ldots, Y_{T_1}$ and update the inputs over time for prediction (Hyndman and Athanasopoulos, 2018).

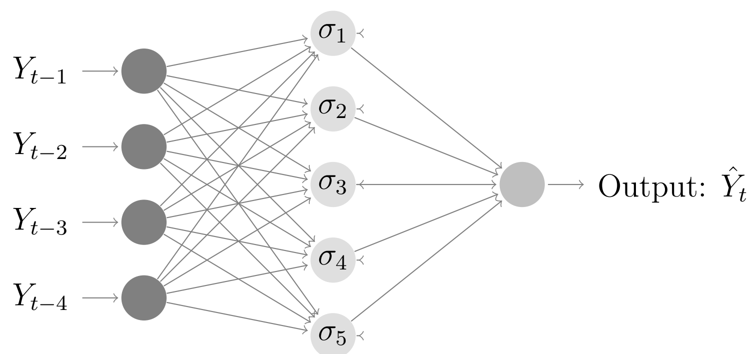

Figure A.2: This is an example of NNAR with four-step-lags as inputs and five hidden nodes in the hidden layer. In each hidden node, the inputs, $Y_{t-1}, \ldots, Y_{t-4}$ are linearly combined and passed through a sigmoid activation function. Then, the outputs of the hidden layer are again linearly combined and $\hat{Y}_t$ is predicted. The network udpates using back-propagation, exactly as we do with a multi-layer perceptron. Similar to VAR, NNAR also produces linear outputs, but it could be thought of as a non-linear regression instead.

References

Alfaro, L., Faia, E., Lamersdorf, N., and Saidi, F. (2020). Social interactions in pandemics: Fear, altruism, and reciprocity. Technical report, National Bureau of Economic Research.

Ayres, I., Vars, F. E., and Zakariya, N. (2004). To insure prejudice: Racial disparities in taxicab tipping. *Yale LJ*, 114:1613.

Dang, H.-A. H. and Trinh, T.-A. (2020). Does the covid-19 lockdown improve global air quality? new cross-national evidence on its unintended consequences. *Journal of Environmental Economics and Management*, 105:102401.

Hartman, A. C. and Morse, B. S. (2015). Wartime violence, empathy, and intergroup altruism: Evidence from the ivoirian refugee crisis in liberia. *May2015. http://cega. beerkely. edu/assets/miscellaneous_file/119_-_HartmanMorseViolenceEmpathy-May_.*

Hausman, C. and Rapson, D. S. (2017). Regression discontinuity in time: Considerations for empirical applications. Working Paper 23602, National Bureau of Economic Research.

Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice.* OTexts.

Jacob, C., Guéguen, N., Ardiccioni, R., and Sénémeaud, C. (2013). Exposure to altruism quotes and tipping behavior in a restaurant. *International Journal of Hospitality Management*, 32:299–301.

Kennett, D. A. (1980). Altruism and economic behavior: Ii private charity and public policy. *American Journal of Economics and Sociology*, 39(4):337–352.

Lynn, M. (2015). Explanations of service gratuities and tipping: Evidence from individual differences in tipping motivations and tendencies. *Journal of Behavioral and Experimental Economics*, 55:65–71.

Marsh, A. A., Stoycos, S. A., Brethel-Haurwitz, K. M., Robinson, P., VanMeter, J. W., and Cardinale, E. M. (2014). Neural and cognitive characteristics of extraordinary altruists. *Proceedings of the National Academy of Sciences*, 111(42):15036–15041.

Rieger, M. O. et al. (2020). Triggering altruism increases the willingness to get vaccinated against covid-19. *Social Health and Behavior*, 3(3):78.

Rubin, D. B. (2005). Causal inference using potential outcomes. *Journal of the American Statistical Association*, 100(469):322–331.

Shino, E. and Binder, M. (2020). Defying the rally during covid-19 pandemic: A regression discontinuity approach. *Social Science Quarterly*, 101(5):1979–1994.

Vieira, J., Pierzchajlo, S., Jangard, S., Marsh, A., and Olsson, A. (2020). Perceived threat and acute anxiety predict increased everyday altruism during the covid-19 pandemic.

Vollhardt, J. R. (2009). Altruism born of suffering and prosocial behavior following adverse life events: A review and conceptualization. *Social Justice Research*, 22(1):53–97.