Response Time Improvement from a New Rescue: Culver City Fire Department

Jacob Kohlhepp

April 8, 2019

Executive Summary

This report estimates one additional rescue would have reduced the response time of 584 calls by more than 3 minutes between July 1, 2017 and June 30, 2018. To get this estimate, we use 4,013 vehicle-response times for Rescue 41 and 43 and assume the new rescue would be placed at Station 2. We model travel response time using driving distance, location within Culver City, response district, building type, day of the week, and weather conditions as explanatory variables. We estimate this model with median regression. Using this model, we predict changes in response time among the 1,282 calls that would have been assigned to this additional rescue. According to these predictions, 584 calls (45%) would see response time savings of more than 3 minutes. **In general, 90% (1,153 calls) would see response time improvements of 51 seconds or more.** Many of the calls were estimated to have much more dramatic improvements.[1] All of these reductions are driven in our model by a decrease in the driving distance between the call address and the responding rescue. In this report, we explain the methods behind this calculation, including a break down of the calls underlying the result. We also report summary statistics, and additional findings from the median regression model.

The Data

Culver City Call and Unit Data

We received the data in two pieces on August 17, 2018 from Mrs. Cara Flores. The first file[2] is a data set where each observation is a unique call. We use this data set for call specific information, including the address, when the call was received by dispatch, and the location type. The second file[3] is a data set where each observation is a unit responding to a specific call. We use this data for response and travel times, and to determine which units were assigned to which calls.

After receiving the data, we performed the following data cleaning steps, which were suggested by Mrs. Cara Flores:

1. Removed all observations in the unit data where a unit responded to the same call multiple times.
2. Removed all observations in the unit data where “UNIT ID” did not begin with a 4.
3. Removed all observations in the call data where “REPORT NO” was “NA.”
4. Kept only observations with “CAD TYPE” of E, R, T, or BC.
5. Removed all test calls with “CALL KEY” of the form “FS———”.

Finally, we removed approximately 20 “CALL KEY”s that were not present in both data sets (either had a call without a unit or a unit without a call).

We also performed these cleaning steps to make our data more similar to the data used for accreditation purposes:

1. Removed calls that were more than 3 standard deviations from the mean.
2. Removed mutual aid calls.

[1] Response time savings are listed for each individual call in the Appendix.
[2] File name is “FCARSOWNER_FCARSCALL2017-18.xlsx”
[3] File name is “FCARSOWNER_FCARSCALLUNIT2017-18.xlsx”

3. Kept only calls in the relevant categories, defined as the list of categories sent to us by Mrs. Cara Flores in October 2018.

After this preliminary cleaning, we created a data set of the first arriving unit for each call. We then developed an interactive website displaying summary statistics for all types of response times across many categories.[4]

For the new rescue analysis, we created a separate data set containing only the rescues (identified as UNIT IDs R41 and R43) but all of the calls for these rescues (so we included calls where the rescue was not the first arriving unit). We also removed calls that occured within 100 meters of the station of the responding rescue, because these calls did not have any meaningful travel time. In the raw unit data, there were 5,729 observations associated with Rescue 41 and 43. After all processing, there are 4,013.

We performed one final (but very important) modification to the Culver City data. We were given the assigned station of each rescue by Cara Flores, and from this we could assign a preliminary "Starting Address" for each rescues. However, looking at the data revealed many instances where a rescue responded to a call very soon after the close of a prior call. So soon, in many cases, that it seemed unreasonable to assume the rescue returned to the station before responding. To partially remedy this problem, we assigned the starting location as the address of the previous call when the previous call's "CLEAR TIME" was within 2 minutes of the next call's "DISPATCH TIME." This amounted to starting address modifications for 542 unit-calls.

External Data Sources

We relied on several additional data sources to create the variables that went into our final model. For calculating driving distance, we used Google Map's Distance API.[5]. We pulled the fastest driving distance between the starting address[6] and end point, captured in the variable "LOCATION." We pulled these distances estimated as of 3:00 AM on December 20, 2018 with setting "avoid highways." We used these specifications to avoid traffic distortions, and because we could not pull historical driving distances. As we discuss later, we pulled these again for the calls assigned to the new rescue. When the driving distance was 0, we replaced the distance with the minimum non-zero distance among all other calls.

We also used weather data from the National Oceanic and Atmospheric Administration Record of Climatological Observations.[7]. We used data from the Santa Monica and Culver City Stations, which is publically available upon request. We explored variables related to the daily minimum and maximum tempature and amount of precipitation. When the Culver City Station had a data gap, we used records from Santa Monica. Only data on precipitation was used in our final model.

Lastly, we experimented with a data set on federal holidays, but this data was not used in our final model.[8].

The Model

We decided to use the statistical technique called median regression to estimate the impact of an additional rescue. We chose this technique for several reasons. First, median regression estimates the effect of an explanatory variable on the median as opposed to the mean of the variable being explained. This seems more appropriate, because the department analyzes percentiles (the median is the 50th percentile) rather than means during accredation. It is also appropriate from a data point of view, because it is robust to travel response time outliers. Additionally, using median regression allows us to avoid making any assumptions about the error term of the regression.

[4]The website is here: www.intrepidinsight.com/shinyapp/culverfd/. It is password protected. The password can be given to any approved parties.

[5]Documentation for the API is available here: https://developers.google.com/maps/documentation/distance-matrix/start

[6]Usually the relevant station address, but in 542 cases the prior call's starting address. See last section for details.

[7]The NDC NOAA portal: https://www.ncdc.noaa.gov/data-access

[8]This data was courtesy of Data World: https://data.world/sudipta/us-federal-holidays-2011-2020

One caveat with this approach is that, as opposed to OLS (or "normal regression") median regression is asking the question "if we increase driving distance by 1 unit, how does this effect the median response time." We are looking at "the typical" call rather than the mean.

To design the model itself, we first selected variables we believed should theoretically be important. We then tried several models, and chose the model with the best Aikake Information Criteria Score that did not throw any warnings during estimation. The equation below represents the model we selected.

$$T_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 Y_i + \beta_4 R_i + \beta_5 L_i + \beta_6 W_i + \beta_7 D_i \times L_i + \beta_8 P_i \times D_i + U_i$$

$$\begin{aligned}
T_i &= \text{Travel Response Time} \\
D_i &= \text{Driving Distance (Google Maps API)} \\
X_i &= \text{X Coordinate (Centered)} \\
Y_i &= \text{Y Coordinate (Centered)} \\
R_i &= \text{Response Districts (Indicators)} \\
L_i &= \text{Location Type (Indicators)} \\
W_i &= \text{Day of the Week (Indicators)} \\
P_i &= \text{Precipitation} \\
U_i &= \text{Random Error}
\end{aligned}$$

The estimates for the coefficients are given in Table 1, along with standard errors and tests of statistical significance:[9]

Estimating Change in Response Time

This section describes how we estimated the change in travel response times.

Selecting Relevant Calls

Calls that occured in Response Districts 21 or 23 are considered to be calls that would have been assigned to the hypothetical new rescue. There were some cases where both rescues responded to the same call. In those cases, it was assumed that the new rescue took the place of the rescue that was farther away.

Updating Driving Distance

For the selected calls, we updated the driving distance by changing the starting address to Station 2 (11252 Washington Blvd, Culver City, CA 90230). Most original starting addresses were previously Station 1 or Station 3, with some starting addresses being the site of the previous call (see the data section for more details).

The Google Maps API driving distances were pulled again using 11252 Washington Blvd, Culver City, CA 90230 as the starting address. The distances were pulled using the "avoid highways" option as of January 3, 2019 at 3:00 AM. Like with the original driving distances, 3:00 AM was chosen because it gives the cleanest measure of shortest driving distance. This is because Google Maps selects routes based on quickest driving time. 3:00 AM is a time when traffic is usually non-existant, so the route chosen will likely be the shortest driving distance. Distances pulled during the workday will likely not be the shortest driving distance but rather routes that avoid traffic. The updated distances are given in Appendix Table A under the column labeled "Distance New."

[9] Estimation was performed with R package "quantreg" at $\tau = 0.5$ with default method "br."

Table 1: Median Regression Results

	Dependent variable:
	travel_resptime
Driving Distance	0.057*** (0.001)
X-Coordinate	0.004*** (0.001)
Y-Coordinate	0.003*** (0.0005)
Response District 21	43.884*** (3.790)
Response District 23	83.738*** (6.059)
Response District 33	45.471*** (5.584)
Location Type I	39.170*** (9.136)
Location Type N	14.552*** (4.474)
Monday	−11.038** (4.338)
Saturday	−8.321* (4.252)
Sunday	−11.430*** (3.667)
Thursday	−9.323** (3.794)
Tuesday	−3.545 (3.999)
Wednesday	−3.114 (3.911)
Distance*Location Type I	−0.012*** (0.003)
Distance*Location Type N	0.0005 (0.002)
Distance*Precipitation	0.005* (0.003)
Constant	52.056*** (5.593)
Observations	4,013
Note:	*p<0.1; **p<0.05; ***p<0.01 Standard Errors are bootstrapped.

Calculating Response Time Reduction and a Confidence Interval

To calculate the change in driving time, we assumed that the only change was the driving distance. This is reasonable because this is a counterfactual analysis, not a predictive analysis. We are asking "what would have happened?" not "what will happen?" A given call during the period either happened during a rainy day or it did not. It either happened on a Wednesday or it did not. It either occured at an apartment building or it did not. Adding another rescue would not have changed these factors.

One thing that is not in the model but that would have changed is personnel. Adding another rescue would require different people manning the vehicle. These drivers likely would have responded to calls differently than the personnel on Rescues 41 and 43. However, we do not have any way to know who would have manned this new rescue on any given day, and we have no reason to believe they would have been slower or faster. So this is left as an important caveat or weakness of the model.

The calculation of the travel time change from adding another rescue comes directly from our model. It boils down to three formulas based on the type of call:

For calls at Location Type "C":

$$\Delta\text{Travel Time (Seconds)} = \Delta\text{Driving Distance} \times 0.057 + \text{Inches Rain} \times \Delta\text{Driving Distance} \times 0.005$$

For calls at Location Type "I":

$$\Delta\text{Travel Time (Seconds)} = \Delta\text{Driving Distance} \times 0.046 + \text{Inches Rain} \times \Delta\text{Driving Distance} \times 0.005$$

For calls at Location Type "N":

$$\Delta\text{Travel Time (Seconds)} = \Delta\text{Driving Distance} \times 0.058 + \text{Inches Rain} \times \Delta\text{Driving Distance} \times 0.005$$

Note that $\Delta\text{Driving Distance} = \text{New Travel Distance} - \text{Old Travel Distance}$. When the new rescue is closer to the call this number will be negative, meaning that negative numbers correspond to quicker response times. From these formulas, the reader can see that for a given reduction in travel distance and no rain, calls with Location Type "N" will see the largest reduction in travel time. Because these formulas vary by "Location Type" and "Precipitation", we have included these variables as columns in Appendix A-1.

Using these formulas, we can calculate the predicted change in travel time for each of the 1,282 calls. These 1,282 results are listed for each call in the final column of Appendix Table A-1. Here are some summary statistics:

Table 1: Summary of Travel Time Changes from a New Rescue

Statistic	Value
Count (calls)	1282
Travel Time Reduced (calls)	1270
Travel Time Increased (calls)	12
Mean (seconds)	-146.81
Median (seconds)	-172.56

Table 2: Percentiles of 1,282 Travel Time Changes

	Percentiles (sec.)
0%	-391.12
10%	-205.36
20%	-198.89
30%	-190.60
40%	-189.00
50%	-172.56
60%	-141.06
70%	-102.74
80%	-94.79
90%	-51.19
100%	122.20

As can be seen in the previous two tables, the vast majority of calls saw large reductions in travel time, with only a handful seeing travel time increases. We believe the most relevant numbers are the median (-172.56 seconds), the 10th percentile (-205.36 seconds), and the 90th percentile (-51.19 seconds). These numbers give a good estimate of the travel time reductions of typical calls in the data. They capture the variety of travel time reductions across the data - variety that is driven mainly by driving distance changing by different amounts. Some calls saw massive reductions in the driving distance - meaning that the old rescue had to travel much farther to arrive on scene.

Although these estimates are very large and therefore very compelling, it is important to quantify the uncertainty associated with them. They are generated from a model, and this model is not perfect. Travel time even between two points using the same route has a random aspect to it - there is always a chance the driver may encounter different obstacles or face different traffic conditions. A confidence interval captures this uncertainty - it is an upper and lower bound on the estimates. A 95% confidence interval for the median travel time change is [-173.06,-172.06]. That is we are 95% confident that the true reduction is in this range.[10]

This range is rather small, which is a reflection of the fact that we have many observations. It also fails to capture the randomness associated with individual predictions. A prediction interval betters captures both the uncertainty of the individual response time predictions and the estimation error in the model. A rough 95% prediction interval for the median response time change (-172.56 seconds) is: [-163.72,-182.00].[11] This can be interpretted as: with 95% probability, a call with these characteristics would see a reduction in response time in this range.

Additional Insight from the Model

Although the primary purpose of the model was to analyze the impact of another rescue on response times, it is also useful for observing patterns in the data. By looking at our model rather than the raw statistics, we can isolate the impact of one factor from another: we can say that one factor was associated with faster or slower response times *with all else constant.*

For example, it is the case that the raw median travel time on Wednesday is 240 seconds, while on Thursday it is 236 On the surface, this suggests rescues respond slower on Wednesdays than Thursdays. However, after accounting for all the other factors in our model, we find the opposite: response times are typically faster on Wednesdays than Thursdays. This reversal means that there is something else driving the raw difference. And indeed, the conditions were different: more rain fell on Wednesdays, and there were more calls at far away locations on Wednesdays.

[10] This method assumes normality and uses the Continuous Mapping Theorem to calculate standard errors from the variance-covariance matrix, which was estimated from bootstrapping.

[11] This prediction interval is obtained in an atypical way: by calling the "predict.rq" function with all other variables except driving distance set to 0. The "percentile" option is used, meaning the interval is generated using a boostrap algorithm.

In this section, we will go over some of the other effects the model describes. Many of these are useful rules of thumb: they can give the captain or battalion chief an idea of how to mentally estimate median response times under certain conditions. They may also help identify areas for improvement. The following sections go through the different findings in detail, but this table provides a summary:

Table 3: Summary of Variable Impacts on Response Time (from Model)

Variable	Impact on Median (seconds)
Driving Distance (meters)	+0.05748 per meter
Rain (in.)	+0.005 per meter
Call on Tuesday	-5.20 (baseline Sunday)
Call on Friday	-7.64 (baseline Sunday)
Rep. Dist. 21	+43.88 (baseline Dist. 11)
Rep. Dist. 23	+83.74 (baseline Dist. 11)
Rep. Dist. 33	+45.47 (baseline Dist. 11)
Location Type I	+39.17-0.0119 per meter (baseline Type C)
Location Type N	+14.55+0.0005 per meter (baseline Type C)
X-Coordinate	+0.0044 per unit (baseline is average X-coordinate)
Y-Coordinate	+0.0026 per unit (baseline is average Y-coordinate)

Weather

We tested several variables related to the weather, including daily minimum and maximum temperature. The only weather variable that our model selection methods kept in the end was precipitation. In our model, it is multiplied by driving distance, and the resulting variable is just below the threshold of statistical significance at the 5% level. The estimated coefficient is 0.005. As one would expect, the coefficient is positive. This implies that one additional inch of rain is associated with a 0.005 second increase in response time per meter. Rain increases response times in a way that is proportional to the driving distance: calls farther away see more dramatic increases, while calls close by see small increases.

To put this in perspective, the median call had a driving distance of 2458. Under our model, if it rained one inch that day, we would expect this response time to have been 12.29 seconds longer.

Dates and Times

We investigated several variables related to dates and times to see if there were seasonal patterns in the data. The final model incorporates only variables for the day of the week. Because traffic patterns and call volume depend heavily on the day of the week, it is not suprising our model selection methods kept this variable. Our model reports the following associations between day of the week and travel response time. Note all effects for days of the week use Friday as a baseline:

	Difference (Seconds)
Monday	-11.04
Saturday	-8.32
Sunday	-11.43
Thursday	-9.32
Tuesday	-3.54
Wednesday	-3.11

It is worth noting median travel response times on Monday, Sunday, and Thursday, with all other factors constant, are 8-11 seconds faster than on Friday. This result is also statistically significant, meaning that we can reject there being no difference with 95% certainty.

Response District

Because the purpose of our model was to estimate the impact of an additional rescue, we did not include a fixed effect for the two rescues that already exist. Instead, we included variables for which response district each call originated from, in the hopes that this captured some of the shift level effects (who was driving the rescue, etc). This variable also represents some of the area-specific traffic conditions (number of intersections, road size, etc) that we cannot explicitly put in our model.

In our model, typical calls in response districts 21, 23 and 33 are associated with much longer travel times than those occuring in district 11: 44, 84, and 46 seconds longer respectively. All of these effects are statistically significant.

Location Type

Interestingly, the variable Location Type, which appears to capture the type of building the call occured at, was an important piece of our final travel response time model. The reference group (or baseline) is Location Type "C", meaning any effects are relative to calls at Location Type C. This variable is included both as a fixed indicator but also as a multiplicative term with distance. This means the type of building matters not only in a fixed way (typical calls at "N" Location Types have slower or faster response times), but also in a way that varies by distance. For example, calls at locations of type "I" that are very close by are associated with longer response times than calls at location type "N" or "C" that are of similar distances away. However, as the driving distances rises to a mile or more, this difference diminishes.

Conclusion

This paper outlines the methods and model underlying Intrepid Insight's analysis of Culver City Fire Department travel response times. Our main goal is to quantify the benefit of adding an additional rescue. We accomplish this by using our model to perform a counterfactual analysis, which estimates the impact an additional rescue would have had on past calls. Our aggregate results point to large travel response time savings from adding another rescue: in the 2017-2018 data, these savings would have been 172 seconds or more (almost 3 minutes) for 50% of the calls assigned to the rescue. Our use of a median regression model isolates the effect of driving distance from other factors. This allows us to more accurately assess the marginal benefit of adding another vehicle, with all else constant. As an added feature, the model provides insight into the impact of weather, day of the week, location, and response district on travel response time.

Note: Appendix Table A-1 is excluded in this public version of the report in order to maintain the confidentiality of the call-level data. The data can be produced upon request with permission from Culver City Fire Department.