To: Dr. Gregory Macfarlane

From: Dr. Gregory Macfarlane

Subject: Validation Lab Solution

Date: April 3, 2014

In our labs thus far, we have treated the outputs of the model as "truth." This is not really a good assumption; for starters, we know that all of the underlying models have confidence intervals. It would be responsible to extend these confidence intervals to the ridership and volume estimates, but this would triple or quadruple the number of calculations required. Also, the assumptions that we make to simplify the model may introduce bias. For instance, the transfer penalty may inflate the ridership of long bus routes at the expense of shorter ones.

Whenever we start using a newly-assembled travel demand model, we need to *validate* that the numbers coming from it represent reality at some level. If they do not, we can *calibrate* the model to better approximate observed conditions.[1] Data for validation can be obtained from departments of transportation, transit agencies, custom traffic counts, or even the underlying surveys used to construct the model. While we should be careful about circular validation, going back to the NHTS or other surveys can identify any distortions that your assumptions introduced in the intervening steps.

1 Lab

Because the most recent year for which data is available is 2007, you will need to run a 2007 base scenario. You may as well get this started now.

1.1 Highway Volumes

UDOT maintains records of average annual daily traffic for important roads in the state, and makes maps of them publicly available on its website at `http://www.udot.utah.gov/main/f?p=100:pg:0::::V,T:,528`. Luckily for you, the planners at WFRC and MAG have already included the AADT for the most important links as fields in the highway network file (`AWDT07`, etc.). Analyzing the fit between the model and the UDOT Counts is simply an issue of comparing the two vectors in R.

The data you need is in the `COMPARE_.....CSV` file. Basically, what we want to do is compare the `ModelAWDT` and `ObsAWDT` fields. Because the UDOT counts aren't available for every link, however, you need to restrict your analyis to links where $AWDT07 > 0$.

```
Highways <- read.csv("COMPARE_MODEL_OBS_VOL_BY_AADT_ID.CSV")
Highways <- Highways[ Highways$ObsAWDT != 0,]
```

Create a plot of AWDT versus the forecasted two-way volume. Include the average trend line. What is the slope of this trend line? What should this slope be? What is the correlation coefficient? Are these good enough for you?

Separate the data into the functional types listed below, and provide a plot of each type's observed and forecasted volumes. Which type has the best fit, and the most realistic slope? Which type has the worst? What does this say about the accuracy of the model in general?

- Freeways
- Multilane Highways
- Principal Arterials

[1] Calibration can introduce its own error, however. Relying on an over-calibrated model weakens the behavioral theory that should govern forecasts.

```
modelall <- lm(ModelAWDT ~ ObsAWDT, data = Highways)

AWDTall <- ggplot(data = Highways, aes(y=ModelAWDT, x=ObsAWDT)) + geom_point()
AWDTall <- AWDTall +  annotate("text", x = 50000, y = 150000,
                                   label = lm_eqn(modelall))
AWDTall +  stat_smooth(method="loess") + theme_bw() +
  xlab("Observed") + ylab("Actual")
```

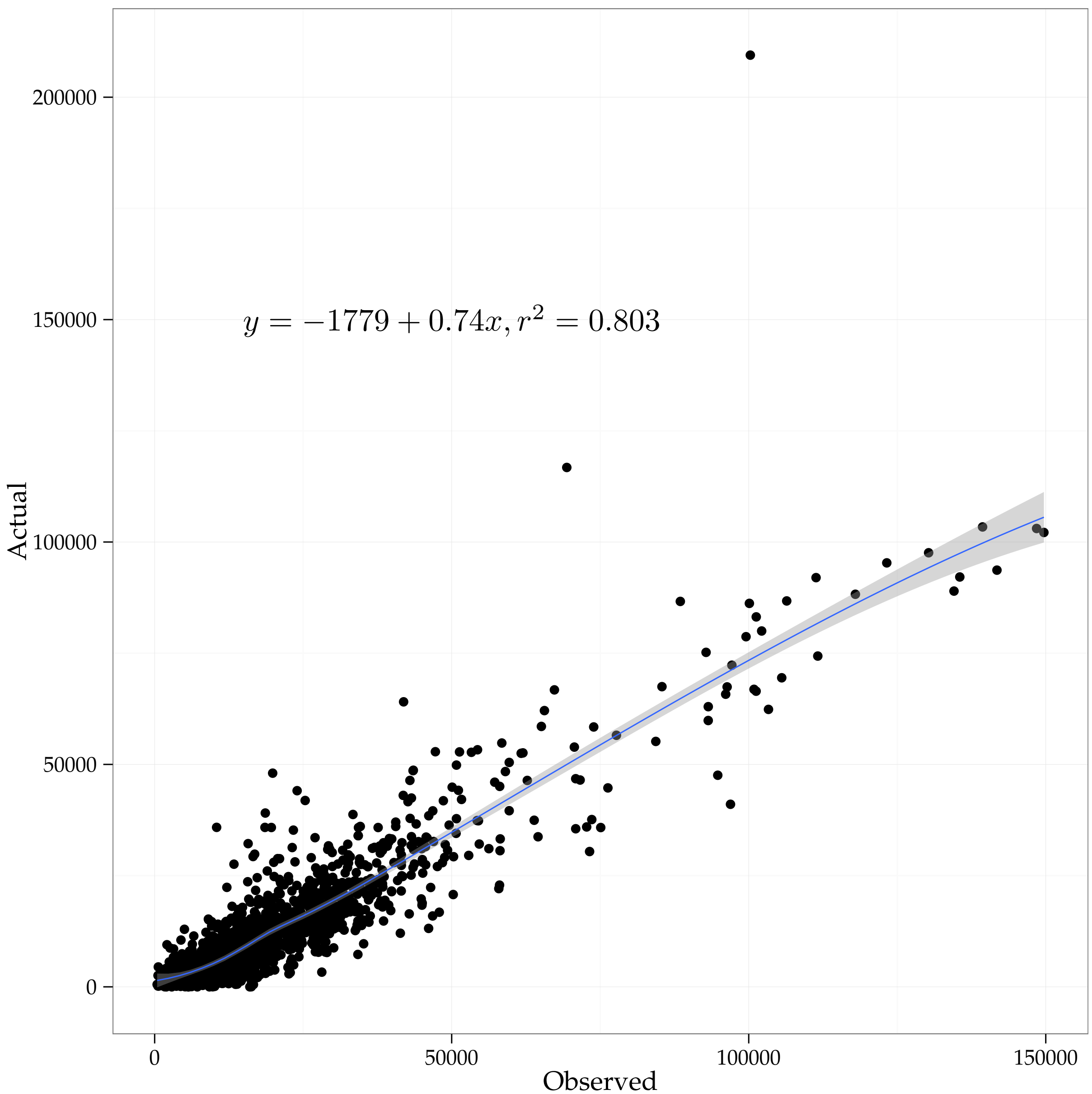

Figure 1: Plot of forecasted versus actual volumes.

- Principal Rural Highways

```
types <- c("Freeways", "Multilane", "Arterials", "Rural")
freeways.ft <- c(31, 32)
multi.ft <- 11
arterials.ft <- c(2, 42)
rural.ft <- 22
types.no <- c(freeways.ft, multi.ft, arterials.ft, rural.ft)
T_Highways <- Highways[ Highways$FT %in% types.no,]
T_Highways$FT <- ifelse(T_Highways$FT == 32, 31, T_Highways$FT)
T_Highways$FT <- ifelse(T_Highways$FT == 42, 2, T_Highways$FT)
T_Highways$FTf <- factor(T_Highways$FT, labels=c("Arterials", "Multilane",
                                                  "Rural", "Freeways"))

modelslist <- list()

for(i in types){
  modelslist[[i]] <- lm(ModelAWDT ~ ObsAWDT,
                        data = T_Highways[T_Highways$FTf == i,])
}
```

2 Transit Ridership

Validating transit ridership is a little more difficult, because UTA is not required to make route-level ridership data available to the public. Also, the data they have is not entirely trustworthy (in my opinion). I have attached a spreadsheet to this assignment showing what UTA uses internally, and what they would probably give to the MPO for validation purposes. I have also included a `BusRoutes.csv` file to spare you the effort of macro-processing the spreadsheet. You should note that the route number needs to be separated from the region prefix in the model.

```
BusesRidership <- read.csv("BusRoutes.csv")
TransitInfo <- read.dbf("~/Desktop/SaltLakeCity/V7_103759/tdmcourse_models/4_ModeChoice/Mo/Boardings,

TransitInfo$MNAME <- gsub("[^0-9]", "", TransitInfo$NAME)

TransitInfo <- merge(TransitInfo, BusesRidership, by.x = "MNAME", by.y = "Route")
```

Create a plot comparing the bus ridership with that predicted in your model. This time, do a thorough analysis on which routes are over or under predicted. Does the transit comparison show any traits that the highway links did not? What factors lead the model to heavily skew a prediction for a transit route? What does this say about the faith you should place in the model outputs?

```
AWDTmatch <- ggplot(data = T_Highways, aes(y=ModelAWDT, x=ObsAWDT, col=FTf)) +
  geom_point(size=6, alpha=0.5)
AWDTmatch +  stat_smooth(method=lm) + theme_bw() +
  facet_grid(.~ FTf, scales="free") +
  xlab("Observed") + ylab("Actual")
```

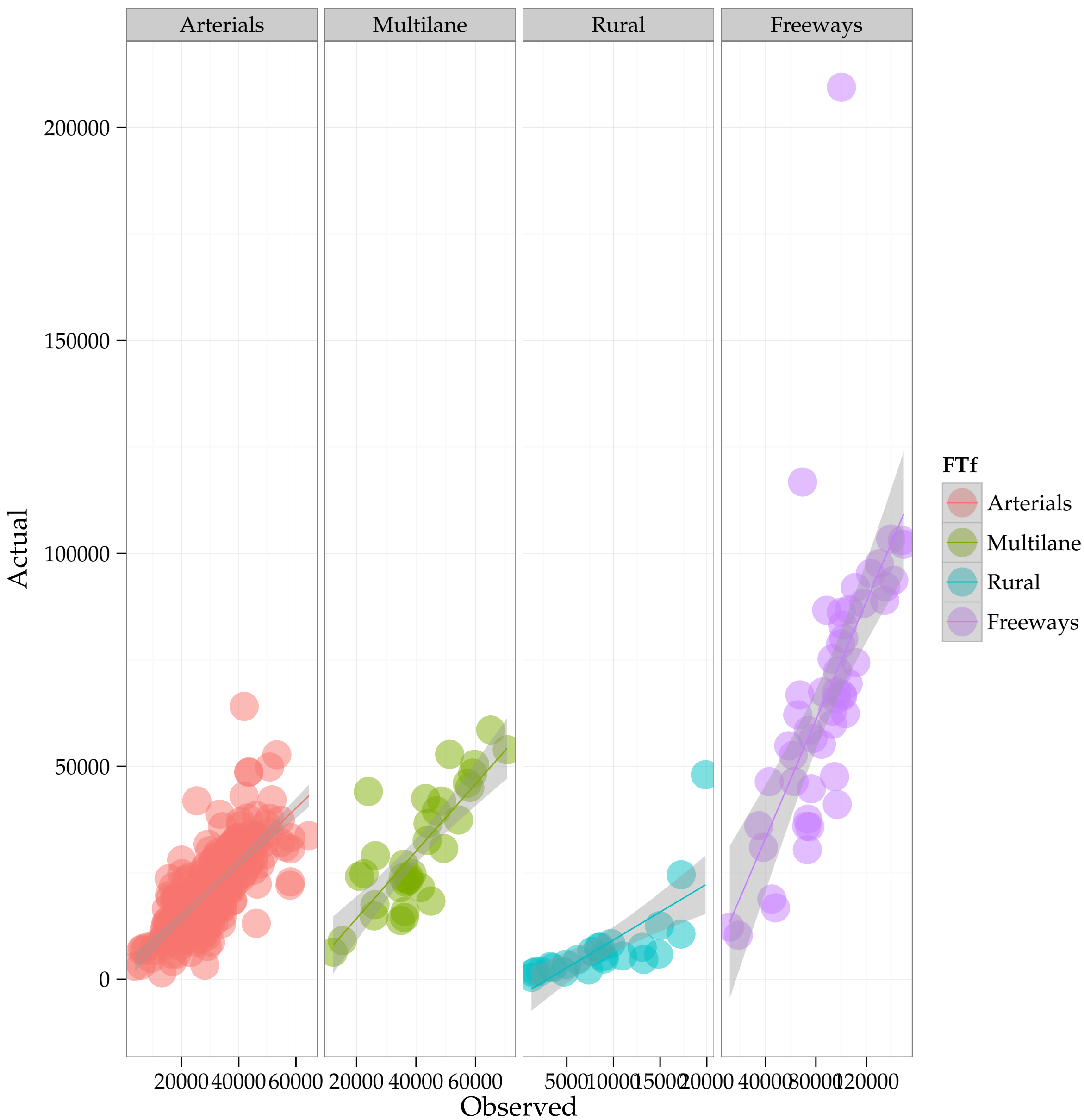

Figure 2: Plot of forecasted versus actual volumes, by functional type.