

Gabbies on Streets Sensitive to Supply and Demand— Data Analysis on Street Price of Gabapentin

Siqi Fu, Yunran Chen, Lingxi Song

10/10/2019

Introduction

Gabapentin is a medication used to treat certain types of seizures and nerve pain. It is first approved for use by FDA in 1993 and has been approved as a generic medication in the USA on *September 23th, 2014*. This drug is ranked 4th most popular prescribed medication in the United States, according to GoodRx¹. The euphoric effects of Gabapentin make it also popular on the streets under the name “Gabbies”. Various survey shows it is increasingly being abused and misused for recreation (getting high)².

We are interested in the fluctuation of price (price per mg) for Gabapentin on streets. Specifically, we are interested in exploring which factors have potential effect on price; how do these factors take effects; whether there are regional variances of the price; whether a time trend or a seasonal effect exists; whether a policy take effects on the price; etc. Our exploration on these questions suggests the price is influenced by the policy and dosage strength significantly (with p-value less than 0.005³), and is not influenced by the time, region, bulk purchase, source and primary reason for purchase. This results are well explained by the supply and demand theory (See Conclusion section for further details).

We build a mixed effect model on a dataset from website StreetRx⁴, which includes 1578 self-reported street price (per mg) records by anonymous users from 2013/02/16 to 2019/3/30 in the United States. Other potential predictors are provided on the dataset, including the date of purchase, location of purchase (state, region), source of information, dosage strength in mg of the units purchased, whether it is a bulk purchase, and the reason for purchase.⁵

Exploratory Data Analysis

We consider logarithm transformation for price per mg (ppm) as our response variable, since the price is positive and the logarithm ppm follows normal distribution approximately. Figure 1 shows the trend of $\log_{10} \text{ppm}$ over time (by quarters). The third quarter of 2014 acts as a special time point. A significant drop of price happens at this quarter, and the price remains relatively stable at a relative lower level since then. No significant seasonal effect is shown. We marked 7 outliers (3.5 standard deviation from the mean) which do not show some special patterns with respect to state and other predictors. Figure 2 suggests there is a heterogeneity across states, but there seems no heterogeneity across four USA regions. Figure 3 further indicates there is no significant correlation across states geographically. The dosage strength in mg of the units purchased seems to pose a nonlinear effect on the price (see Figure 4). As the dosage strength increase (except for 350 point), the price decrease and the decrease rate slower down. See Figure 10 in the Appendix for the summary of other predictors without significant effect on the price.

Exploratory data analysis suggests including an indicator for purchase date after the third quarter of 2014, which indicates whether policy that Gabapentine can be considered as a generic medication in the United States takes effects. The results also suggests considering nonlinear effect of dosage strength, and a heterogeneity of states in our model. Figure 5 suggests states with less sample size tends to have more variance, indicating a hierarchical model is needed to allow for borrowing information across states⁶.

¹Official website of GoodRx: <<https://www.goodrx.com/drug-guide>>

²Information of Gabapentin comes from wikipedia page: <<https://en.wikipedia.org/wiki/Gabapentin>>

³We consider the p-value threshold to be 0.005 as suggested by an article in nature: Benjamin D J, Berger J O, Johannesson

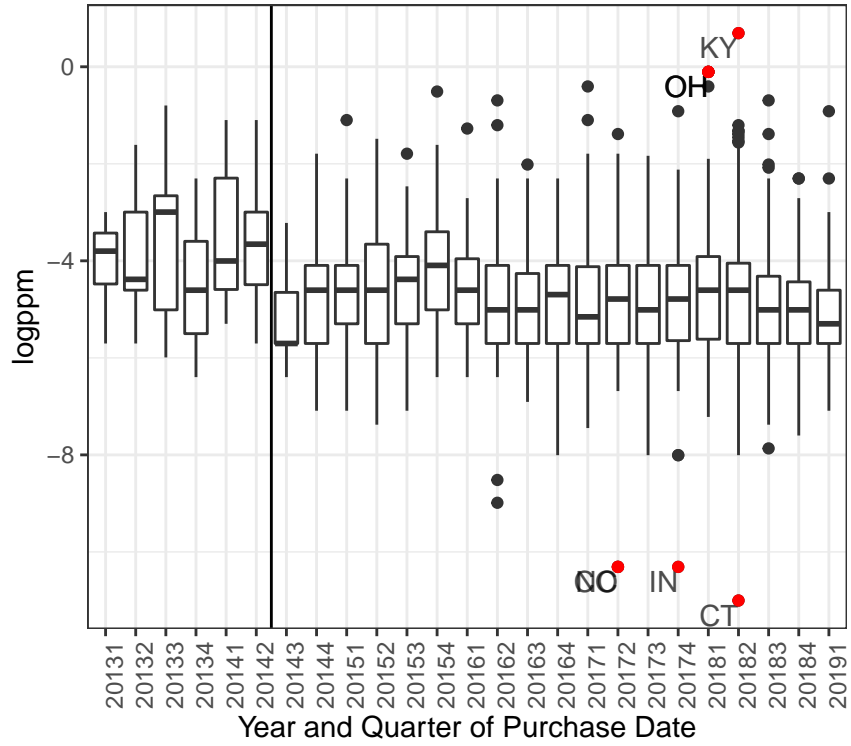


Figure 1: Price Drop on 2014Q3

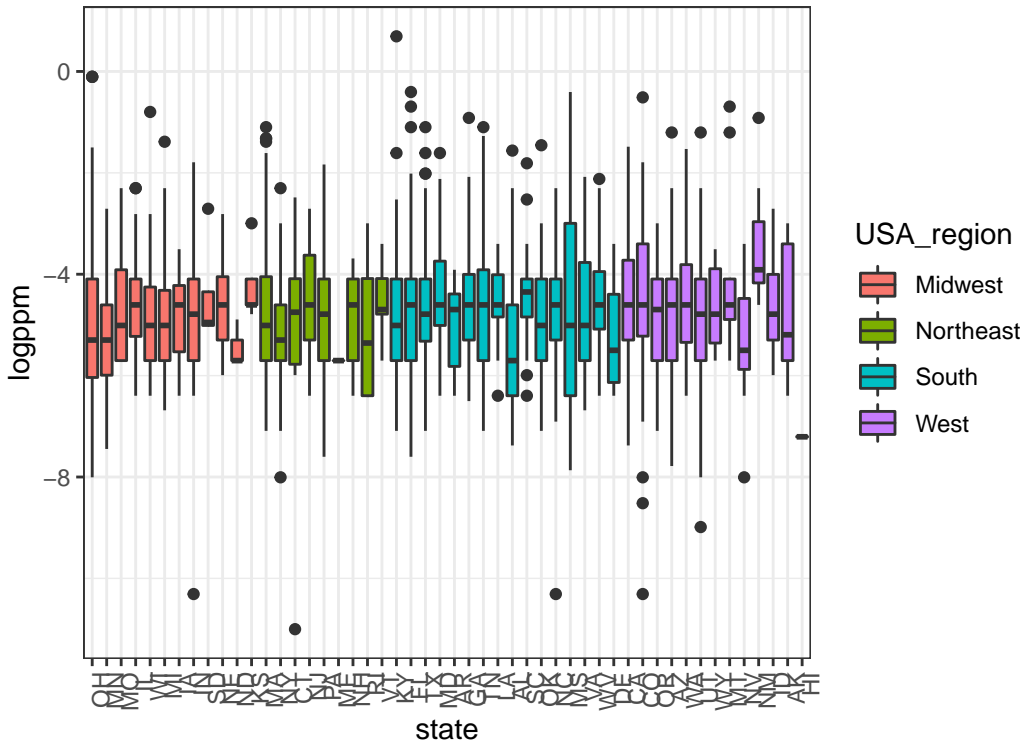


Figure 2: Average logppm per State (colored by USA_region)

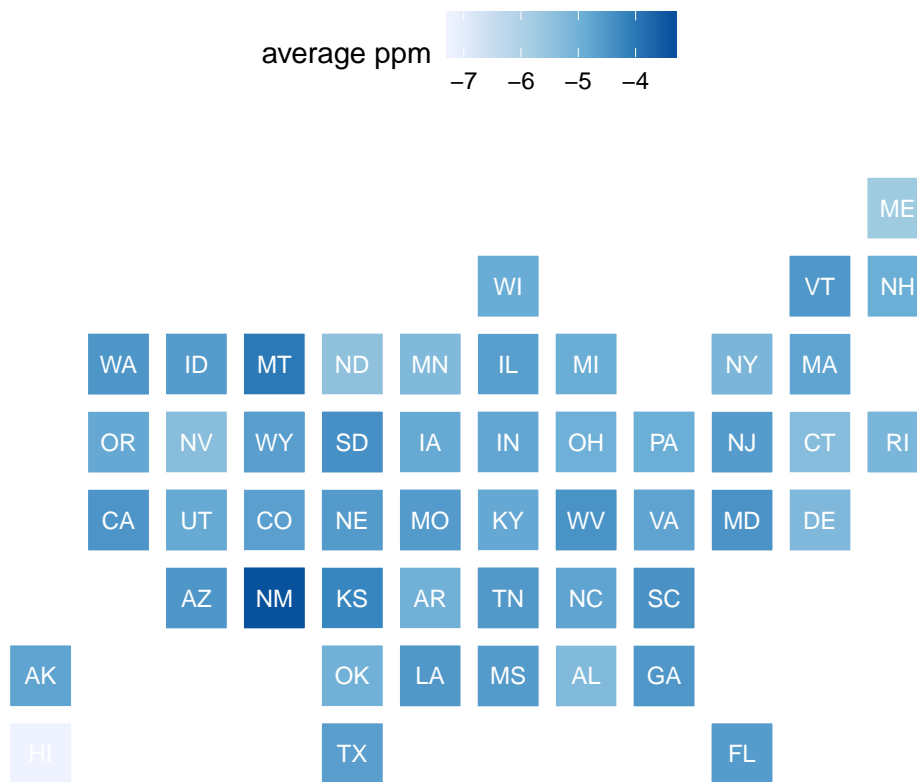


Figure 3: Average logppm per State Geographically

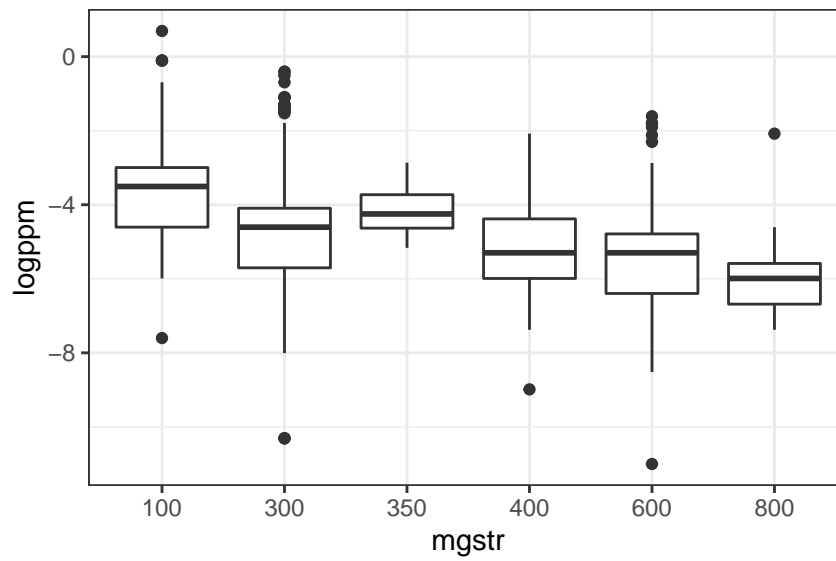


Figure 4: EDA for dosage strength (mgstr)

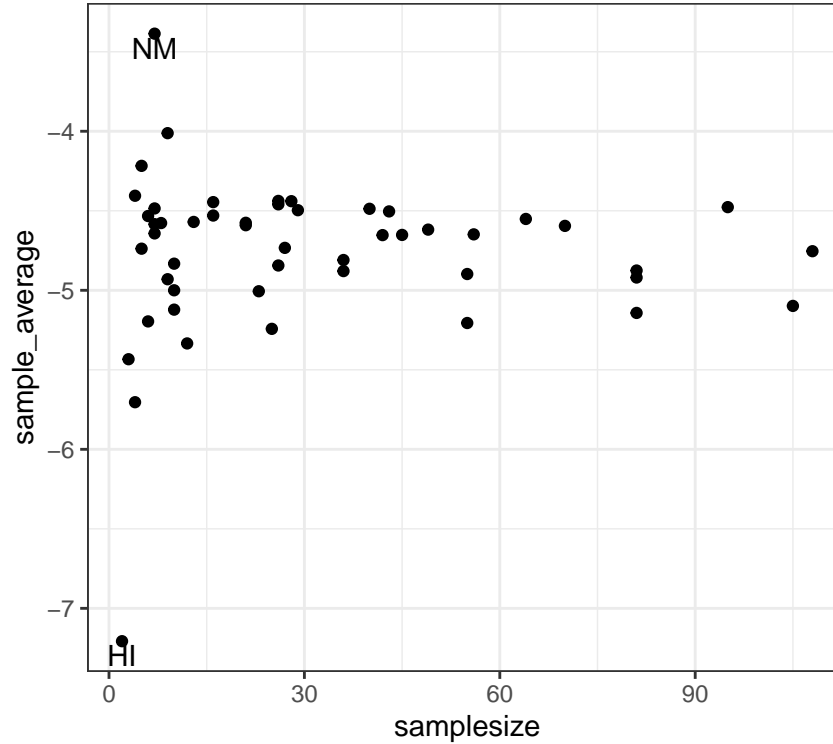


Figure 5: Sample Size vs Average logppm per State

Model

Based on the insights from exploratory data analysis, we include state as a random effect, and include both purchase time and dosage strength into our model. Then we conduct variable selection and compare different models based on likelihood ratio test with p-value threshold 0.005 and BIC.

The main challenge is how to incorporate time effect and the dosage strength since they both show nonlinear effect on the price. For the time effect, we compare three possible predictor: purchase date, year and quarter of purchase date, the indicator of purchase date greater than 2014/09/23 (the date when policy takes effect nationwide), and indicator of purchase date greater than the third quarter of 2014 (the year and quarter when when policy takes effect nationwide). For the first two continuous predictors, we compare spline and polynomial transformation. For the dosage strength effect, we compare include it as a continuous variable and as a categorical variable. For the continuous dosage strength effect, we compare logarithm, spline and polynomial transformation. We compare the aforementioned models based on likelihood ratio test (LRT), BIC criterion, residual plot and QQ-plot. *The best model includes the logarithm of dosage strength ($\log(\text{mgstr})$) and the indicator of purchase date greater than the third quarter of 2014 (referred as $I(\text{policy})$ thereafter).*

To explore whether others factors have potential effects on the price, we consider a full model with logarithm of dosage strength ($\log(\text{mgstr})$), the indicator of policy ($I(\text{policy})$) and include all other potential predictors such as USA region, source, bulk purchase, primary reason. We conduct LRT and check BIC for each predictor to check whether including it will bring significant difference to the model. For the potential predictors with multiple categories, we also test whether different ways to merge the categories will bring better model performance. Table 1 shows the result of the LRT and BIC. *Based on the testing result, we conclude that other predictors except for the dosage strength and policy do not bring significant effect nor improve the fitting*

M, et al. Redefine statistical significance[J]. Nature Human Behaviour, 2018, 2(1): 6.

⁴<https://streetrx.com/>

⁵See the Appendix for the distribution of each predictor and data preprocess procedure.

⁶Labelled data are 2 standard deviation away from the total mean

performance of model. This conclusion is consistent with results from exploratory data analysis.

Based on aforementioned model comparison, the best model includes logarithm of dosage strength ($\log(\text{mgstr})$), policy effective date ($I(\text{policy})$) and a random effect for states. We further check whether an interaction is needed between $\log(\text{mgstr})$ and $I(\text{policy})$; and if they are significant, whether a random effect is needed for $\log(\text{mgstr})$ and $I(\text{policy})$ across states. We consider LRT and BIC for model comparison. Table 1 suggests *there is no interaction and random effects of $\log(\text{mgstr})$ and $I(\text{policy})$ needed.*

Our best performance model is as follows:

$$\log(\text{ppm})_{ij} = \beta_0 + b_{0j} + \beta_1 I(\text{policy})_{ij} + \log(\text{mgstr})_{ij} + \epsilon_{ij}$$

where i refer to individual record i , j refer to the state j . $I(\text{policy})$ indicates purchase date after the second quarter of 2014, $\log(\text{mgstr})$ represents the logarithm of dosage strength in mg of the units purchased. We assume $\epsilon_{ij} \sim N(0, \sigma^2)$ (i.i.d), $b_{0j} \sim N(0, \tau^2)$ (i.i.d) and ϵ_{ij} is independent of b_{0j} for any i, j .

Table 1: Diagnosis Table for Potential Predictor

Variable	df	p_value	BIC	BIC.full.	BIC.best.
USA_region	3	0.0185872	4909.542	4931.734	4802.573
Source	4	0.1288671	4899.558	4931.734	4802.573
Bulk Purchase	1	0.8085181	4899.558	4931.734	4802.573
I(Policy)	1	0.0001294	4936.878	4931.734	4802.573
$\log(\text{Dosage Strength})$	1	0.0000000	5288.313	4931.734	4802.573
$\log(\text{Dosage Strength}):I(\text{Policy})$	1	0.7409263	4811.215	4931.734	4802.573

Estimation and Inference

Table 2 shows the estimations of our model parameters. Holding other factors consistent, the policy, which considering Gabapentine as a generic medicine, leads to a drop of the price. Specifically, we are 95% confident that after the policy takes effect nationwide, the price is expected to decrease by around 30.5% to 59.9%. Increasing the dosage strength in mg per unit by 10% will expect to decrease the price by 8.2% to 9.8%. It may be because that as the dosage strength in mg increase, the size of tablet will decrease, which will decrease the cost of production which leading to the decrease of price.

Additionally, we found that there is heterogeneity across different states. We are 95% confident that the variance of logarithm of price among different states are around 0.038 to 0.201. But there are no significant heterogeneity across states.

Table 2: Estimation and 95% Confidence Interval

	Estimate	Std. Error	2.5 %	97.5 %	p-value of LRT
Intercept	1.592	0.296	1.012	2.174	
I(Policy)	-0.64	0.14	-0.915	-0.365	5.432e-06
$\log(\text{Dosage Strength})$	-0.991	0.048	-1.086	-0.897	1.302e-83
State	0.015	0.123	0.038	0.201	1
Residual	1.201	1.096	1.057	1.135	

We further applied our model to rank states based on the price controlling the policy and dosage effect. Table 3 and Table 4 that have the highest price per mg, and the lowest price per mg. Based on our prediction, Hawaii has the highest price per mg, and Tennessee State has the lowest price per mg.

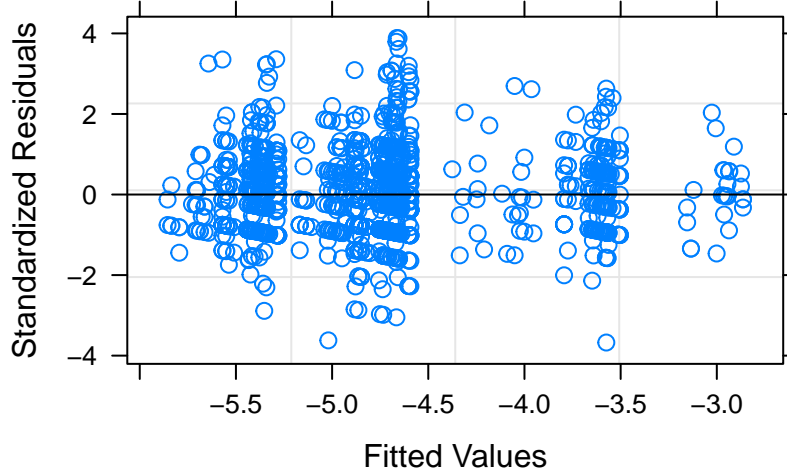


Figure 6: Residual Plot

Table 3: Top 5 States with Lowest Average Price (Median)

Region	State	Observed Mean	Estimation	5% CI	95% CI
West	HI	0.001	0.004	0.001	0.019
West	AK	0.006	0.005	0.001	0.019
South	MS	0.007	0.005	0.001	0.022
South	DE	0.004	0.006	0.001	0.023
Northeast	RI	0.005	0.006	0.001	0.024

Table 4: Top 5 States with Highest Average Price (Median)

Region	State	Observed Mean	Estimation	5% CI	95% CI
Midwest	KS	0.01	0.026	0.006	0.098
West	NM	0.02	0.010	0.002	0.045
West	CA	0.01	0.010	0.002	0.040
West	MT	0.01	0.010	0.002	0.040
South	TN	0.01	0.010	0.002	0.040

Model Diagnosis

During the model comparison with other alternatives, our model has the lowest BIC, and each term we included is statistically significant (p-value less than 0.005)⁷. From the residual plot (Figure 6), we found that all the residuals are normally distributed around zero. From the qqplot as shown in Figure 7, the overall residuals satisfy the normal assumption and only several points at the tail slightly violate the normal assumption.

Sensitivity Analysis

We conducted sensitivity analysis to check the influence of the outliers⁸ and influential points⁹ on our

⁷We include non-significant random effect term for we want to borrow information across states and make inference on states.

⁸Defined as the data with logarithm of price outside 3.5 standard deviation from the mean.

⁹Defined as the data with cook distance greater than $4/n$, where n is the sample size.

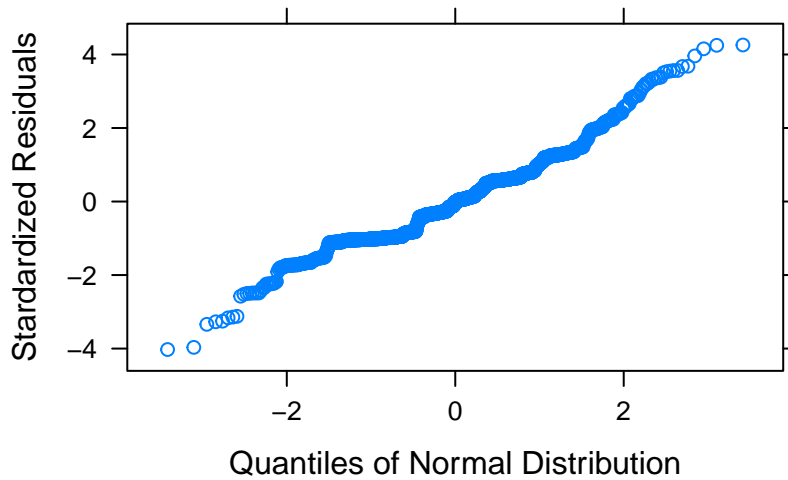


Figure 7: QQ plot for Residuals

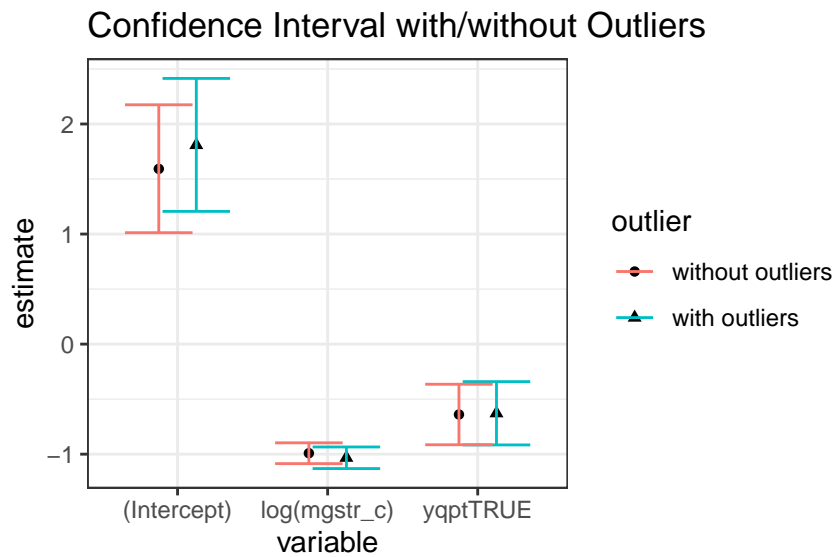


Figure 8: confidence interval of fixed effects with/without outliers

model. Specifically, we compare the model estimation with or without outliers, then compare the model estimation with or without influential points (excluding the outliers in the analysis).

Figure 8 shows the coefficients confidence intervals are similar, and only the intercept confidence interval are relatively different. Since we would like to focus on exploring the which factors will significantly affect the price and make inference based on the majority of data, we exclude the outliers.

Also, compared the two confidence intervals of coefficients as shown in Figure 9, we did not found significant difference. So we decided to keep the influential points.

Conclusion

Our results can fit the data well as shown by model diagnosis and comparison among other possible models.

Furthermore, our model can be perfectly explained from supply and demand aspect in economics. After FDA consider Gabapentine as a generic medication in the United States, the supply of Gabapentine increase

Confidence Interval with/without influential points

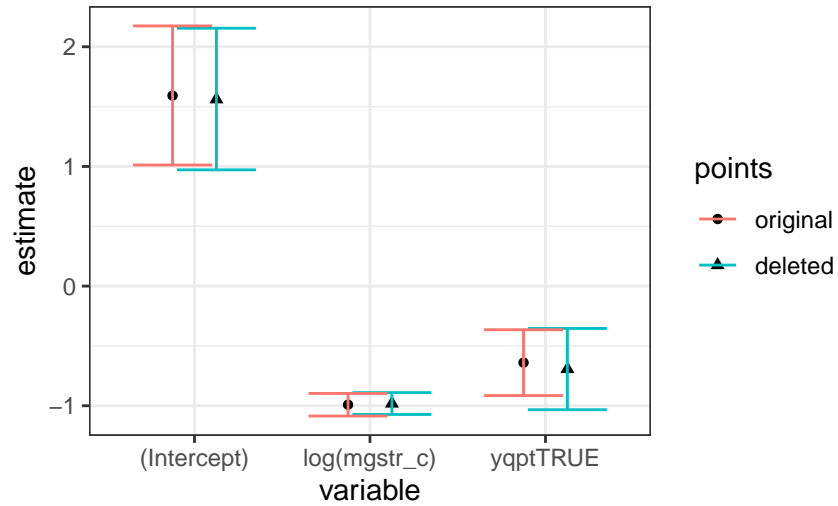


Figure 9: confidence interval of fixed effects with/without influential points

leading to a drop of the price on streets. Such policy is effective for the whole country; therefore, it does not show significant heterogeneous effect across distinct states. After the approval of FDA, Gabapentine is listed as one of the most popular prescribed medications, which ensures the supply on the streets market. As a result, the price of Gabapentine is relatively stable over time. Since the price depends on the supply and demand balance, the primary reason of buyer, the source of information, bulk purchase do not have significant connection to supply and demand on streets market, resulting in insignificant effects on the drug price. In addition, the convenience of transportation across states and especially the emerging Internet pharmacy, the streets market in distinct states can be considered as connected nationwide, leading to insignificant heterogeneity across states.

It is worth noting that our model suggest increasing dosage strength in mg per unit makes negative effect on drug price, which may be explained by the cost-effectiveness from a producer's aspect. A increase in dosage strength in mg per unit may lead to a smaller size drug per unit, which may have less cost for producer.

Further improvements can be done to better model the tail behavior. In our model, we assume a normal distribution for the error term and delete the samples with the price 3.5 standard deviation away from the mean. The sensitivity analysis shows there exists slight difference between the intercept estimation of our model. A heavy tailed t distribution on the error term may be an alternative choice if we want to emphasize the tail behavior of the dataset.

Appendix

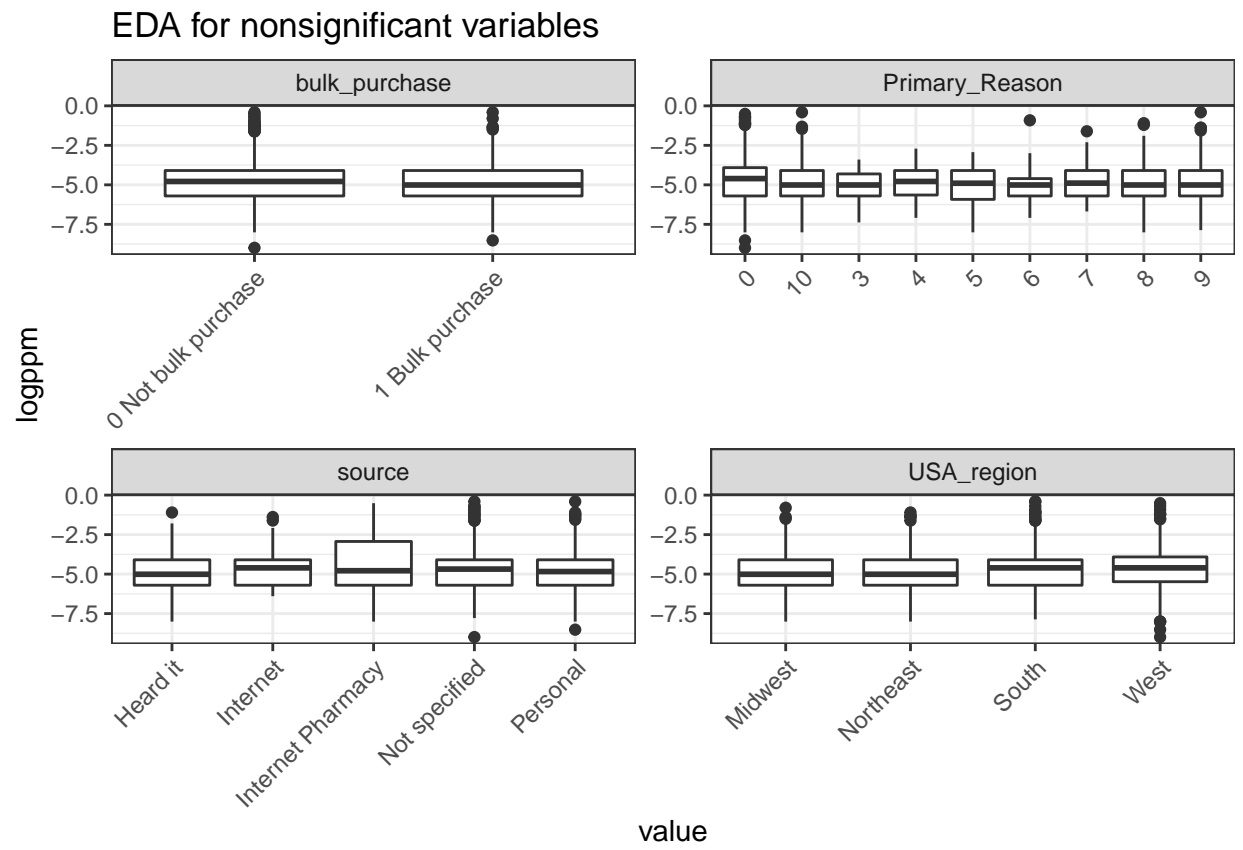


Figure 10: EDA for Nonsignificant Variables