Legislative Productivity of the 115th Congress of the United States of America

The Wharton School at the University of Pennsylvania

Abstract

This study aims to analyze the count data via NBD models obtained from the number of bills each member of the 115th Congress enacted into law that he/she/they was a primary sponsor of. The study also looks at categorizations based on party affiliation in both the House of Representatives and the Senate, and performs a chi-squared goodness of fit test to evaluate its overall fit. In addition, multiple methods such as the MLE, MoM and Means and Zeros are utilized. Reasonable inferences across the two data sets – the House of Representatives and the Senate – are made via the parameters provided by the NBD model and the distribution of lambdas across groups are presented. It is the author's view that this data will be extremely valuable for a broad range of cases, from constituents wanting to know the legislative productivity of their representative, from opponents wanting to challenge incumbents based on these metrics, from social activist groups wanting to know the best politician to reach out and from businesses wanting to increase the effectiveness of their lobbying efforts.

Keywords: NBD Model; Congress; Legislation; Fader

1 Introduction

This paper focuses on the legislative productivity at an individual level of the 115th Congress as measured by the number of laws enacted by representatives. The data for this study was extracted from GovTrack.us, an independent non-governmental website that tracks legislation in the United States Congress. Specifically, for the 115th Congress, the number of laws enacted by each member of both chambers – the House of Representatives and the Senate – were used for this study. It is important to note that representatives have other responsibilities apart from enacting legislation such as overseeing bureaucracy, attending to their constituents, holding town hall meetings, raising campaign funds as well as running for political office (Frantzich, 1979). However, enacting laws is one of the few quantifiable metrics. Therefore, this study and data are not a reflection of the value of each representative, and only take into account the number of bills that each representative was a primary sponsor of that became law.

2 Background

Information regarding the legislative productivity of Representatives is critically important to activist groups as well as protest organizations. Oftentimes, these organizations are

severely limited in their time and resources, and hence it can be difficult to decide which congress person would be most effective in helping them aside from their very own representative. While this might not be as big of a problem for large organizations who gain significant media attention, the issues raised by small organizations are just as important and need to be addressed. The author of this article has also worked with above mentioned organizations – albeit not at the federal level – and has witnessed the lengthy process to reach a representative in hopes of passing a bill. On another note, businesses will also find this information valuable as it will allow them to more effectively lobby for bills that will support their activities and industries.

3 Note about the Data

The data from the Senate and the House of Representatives will be treated as two different data sets. Although it may seem like these data sets can be combined into one larger data set, the legislative process for a bill to become law is widely different in the Senate and the House of Representatives. A few of the main differences include the size of the chamber, the filibuster rule, simple majority rule, term lengths, subcommittee voting methods, etc. Henceforth, it is most appropriate to treat these as two separate datasets.

4 Questions

This paper will aim to answer the following questions :

- How well can the NBD model capture a dataset political in nature?
- How varied is the heterogeneity across each chamber?
- How is the above affected by which party controls each chamber of Congress?
- How does the propensity of representatives to enact laws differ from across parties?
- Which party and chamber have representatives who have the highest propensity to enact a law present?
- Which party and chamber have most representatives around the lowest propensity to enact a law?

5 Method and Models

5.1 *Count Data*

There are three main types of datasets that are to be considered: time data, choice data, and count data. This study does not measure the discrete-time duration until a representative passes a law, or a certain number of laws, and hence it does not represent time data. Furthermore, since there is no binding upper bound on the number of laws that each representative can introduce, and thus this is not a choice data. This data is in the form of count data, where the count signifies the number of laws enacted by each representative.

5.2 *Poisson and NBD*

The below image illustrates the different models we can choose from (taken from Bruce Hardie's Webpage; citation at the end):

Summary of Probability Models

Phenomenon	Individual-level	Heterogeneity	Model
Counting	Poisson	gamma	NBD
Timing	exponential	gamma	Pareto II
Discrete timing (or counting)	geometric	beta	BG
Choice	binomial	beta	BB

Figure 1: Probability Models

As established above, this is a count data in nature and therefore there are two main choices of probability models: the Poisson and the NBD.

5.2.1 *The Poisson*

Let us define a random variable Q that denotes the relationship between the representatives and the number of laws enacted. The Poisson model assumes that this random variable Q follows a Poisson distribution with parameter lambda. There are certain limitations that come from this mode:

- Does not account for heterogeneity: This Poisson model does not allow for each individual representative to have their own lambda, and instead only models Q with respect to all lawmakers. Therefore, it is restricted in that it does not account for differences across lawmakers.
- Equidispersion: One of the key properties of the Poisson distribution is that its mean equals its variance. Although this is an incredibly useful property, it heavily restricts the effectiveness of this distribution as most natural datasets are hyper dispersed in nature. As follows, so are the two datasets presented in this paper.

For the above reason, we choose not to run the Poisson distribution as it will not aid in answering our proposed questions.

Value	All Senators	All House of Representatives
mean	4.171	1.406
variance	13.657	2.533

5.2.2 *The NBD*

We will run the Negative Binomial probability distribution (NBD) on our data sets. By doing so, the following assumptions are accepted :

- The individual-level behavior of each representative as it relates to enacting laws follows a Poisson distribution with parameter lambda
- Heterogeneity in the above-mentioned lambdas follows a gamma distribution with shape parameter r and a scale parameter alpha.
- Stationarity: The lambda parameter of each representative does not change during the observed time period.
- Independence: The lambdas of each representative are independent of each other.

Furthermore, since the 115th Congress has a fixed time period of two years, the number of opportunities for each representative to pass a law are considered to be the same for the purposes of this model. As such, we will use a unit time period NBD.

6 Pre-Analysis Discussion

We must first recognize that there is a meaningful categorization for both datasets based on party affiliation. Moreover, due to the nature of Congress, the controlling party in each house has significant power over the number and type of laws that are passed through it. It is also critical to note that members of different parties are more likely to disagree with each other than members of the same party, and this is reflected in the nature and sponsors of the bills that eventually become law. This, along with the needed minimum number of votes being at least 50 percent of the members for each chamber, provides a case that the distribution for the majority party would be more heterogeneous in nature. As the majority party controls the bills that pass through, their members would have a broad range of lambdas from those who do not focus on legislation to those interested in passing multiple laws. In contrast, the minority party would be more homogeneous, since only those members who hold views that enable them to reach across the aisle would have a slightly higher propensity, and therefore a slightly higher lambda than minority members who are not able to so. Adding on, the member with the most number of laws enacted is far more likely to be of the majority party. Although the range of lambdas is not a direct indicator of heterogeneity, it does influence the variance, which in turn is one of the factors that contribute to heterogeneity. In our NBD model terms, this equates to a higher r-value for the minority party as compared to the majority party. We can also reasonably conclude that since the majority party controls the agenda of laws, the set of lambdas of the minority party would be smaller than the set of lambdas of the majority party. Therefore, in terms of our model parameters, we would expect a higher alpha for the controlling party. Finally,

for the above-mentioned reasons, we anticipate that there will be enough differences in the nature of representatives based on the party to warrant extra parameters for a better overall fit. That is, we expect that running separate NBD models for each chamber based on the party will provide a better overall fit than running a pooled NBD model for each chamber.

During the 115th Congress, Republicans held the majority in both the Senate and the House of Representatives. In summary, we expect the following :

- NBD models divided by party will provide an overall better fit than a pooled model for both chambers of Congress.
- The NBD model ran on the dataset of Republicans will have a lower r and a higher alpha than the NBD model ran on the dataset of Democrats for both chambers.

7 Analysis

The NBD model was run using the data in its most granular form and hence no data was lost. The presented histograms were compiled by aggregating the data but not right censoring in order to display the data in its full form and visually aid the reader. Finally, the data was aggregated into bins to carry out the chi-squared goodness of fit test as well as right-censored at the 5+ mark to meet the recommendation that 80% of expected values should be larger than 5. Truncated, shifted, or inflated NBD models were not used for this analysis.

7.1 *The United States Senate*

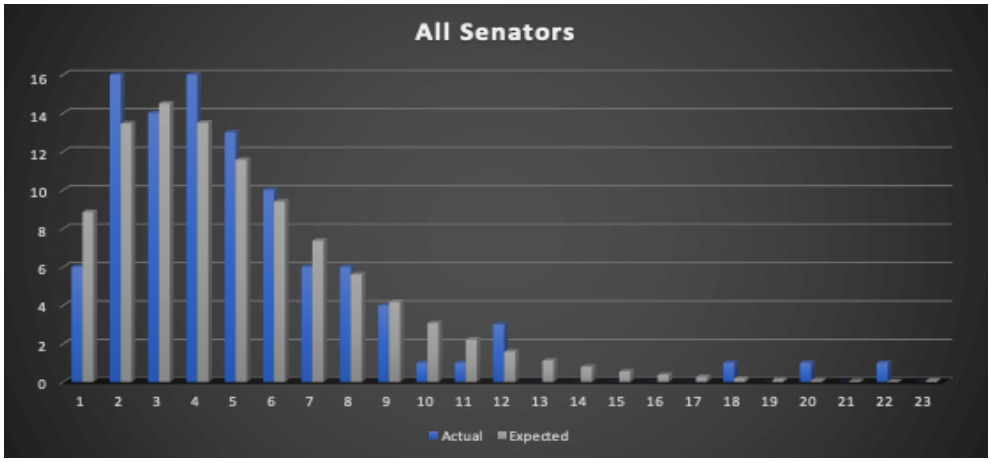

Figure 2: NBD on All Senators

r	2.393
alpha	0.574
p-value	$p > 0.2$

The r-value of 2.39 is evidence that there is quite a bit of heterogeneity among United States Senators and the alpha value of 0.57 indicates that the distribution of lambdas is not that broad. Moreover, the general verdict of the chi-squared fitness test is that our NBD model is a well fit for our data (p is greater than 0.2). Visual representation in the form of a histogram allows for qualitative conclusions to be made. With only a few exceptions, the observed and expected values seem to agree quite well for most values and there is not a pattern of over or under mapping the observed data points.

7.1.1 *Robustness*

The following three graphs (presented in the next two pages) represent the distribution of λs from the *gamma* distribution with parameters from their respective methods.

Method	MLE	MoM	Mean and Zeros
r	2.393	1.834	3.745
alpha	0.574	0.439	0.897

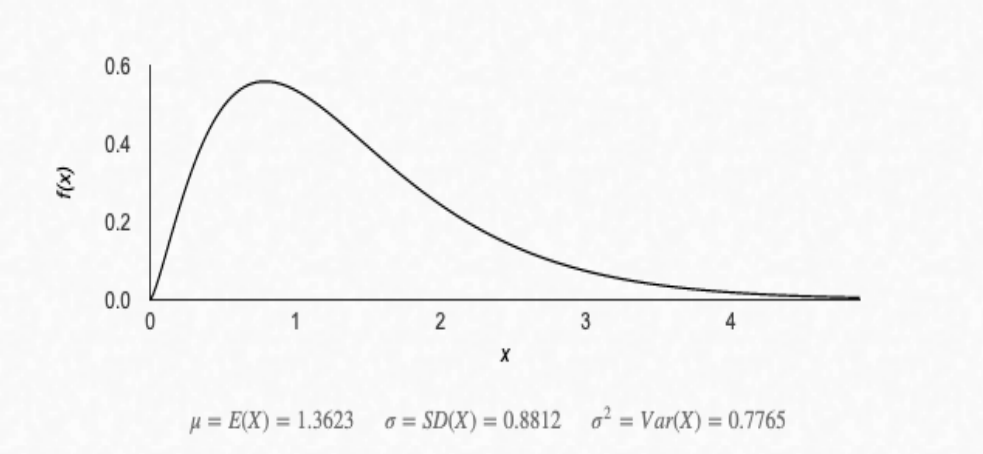

Figure 3: Distribution of λs based on parameters obtained via the Maximum Likelihood Estimation

The r values obtained using alternative methods such as MoM and Means and Zeroes are not incredibly close to that of the MLE. Specifically, MoM results in a lower r-value as well as a lower alpha value. On the other hand, Means and Zeroes result in a higher r-value and a higher alpha value. The inaccuracy of the MoM method can be attributed to the long right tail of our data, which would affect both the mean as well as the variance in some capacity. Indeed from the above graphs of the gamma distribution of lambdas obtained by the respective method's parameters, it can be seen that the MoM method ranges the lambdas at an upper bound of 3, whereas the more precise MLE method places it beyond 4. Additionally, the MoM method places the bulk of the interior mode towards the 0.5 mark, whereas the MLE disagrees and places it around the 1 mark. Overall, the MoM method overestimates the senators who have a Poisson wheel with a lower lambda

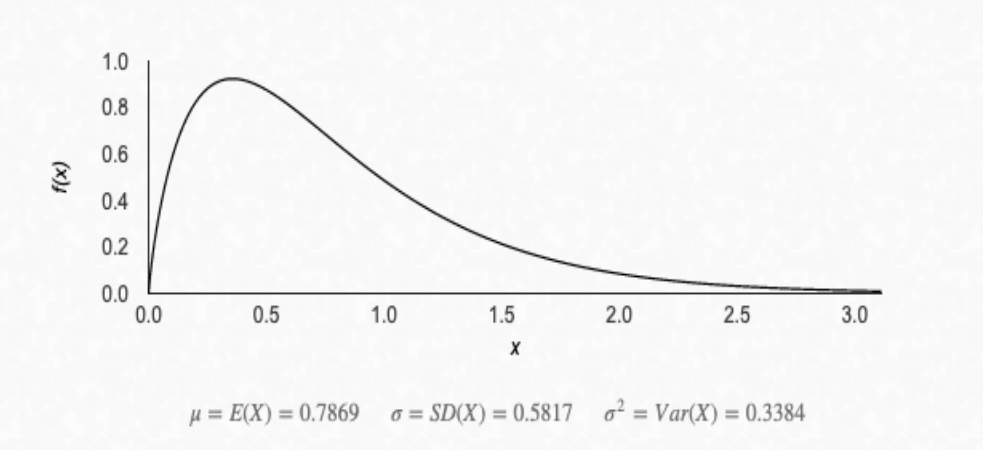

Figure 4: Distribution of λs based on parameters obtained via the Method of Moments

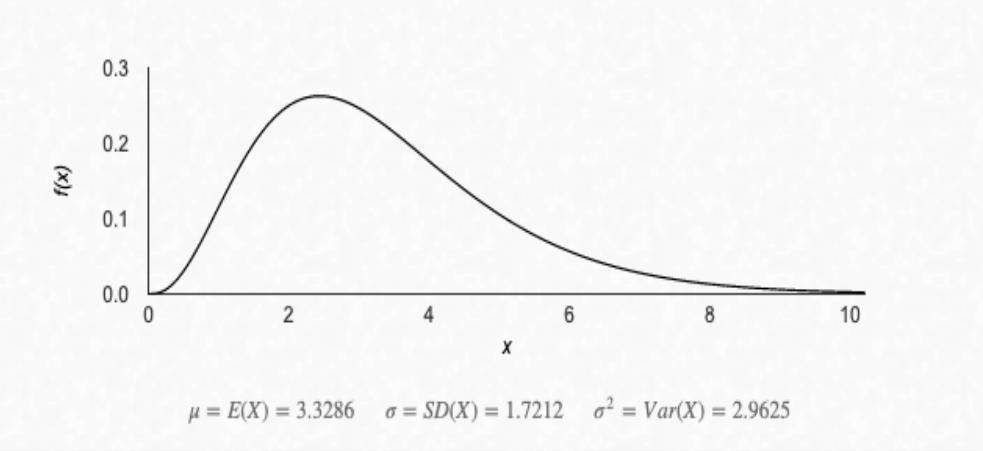

Figure 5: Distribution of λs based on parameters obtained via the Means and Zeroes Method

(lambda less than 0.5), underestimates the representatives who have a very high lambda (lambda more than 3), and reports a more heterogeneous population. Shifting our focus to the other method, the inaccuracy of the Means and Zeroes demonstrates that NBD expects far fewer congressmen and congresswomen with lambdas around the 0 mark and puts the interior mode farther along the x-axis when only allowed access to the P(x=0) and the mean. Furthermore, this method also dramatically increases the range of our lambdas, going as high as 10 as displayed in the provided graph. Once again, the reason for this inaccuracy can be attributed in part to our right tail. However, it also signifies that there might be external factors that cannot be captured by our NBD model for the P(X=0) cell. In summary, the Means and Zeroes under reports on the heterogeneity of the Senators as well as incorrectly places a few senators with very high lambdas.

7.1.2 χ^2 *LRT Test*

As discussed previously, there is an appropriate covariate based on party affiliation. As none of the models – data based only on Republicans and Democrats or the together pooled model – were provided with extra parameters, and they together operate on the same data, we can conclude that these models are nested. The chi-squared LRT was performed as presented below :

$$2 * (LL_{\text{Republican Senators}} + LL_{\text{Democrat Senators}} - LL_{\text{Pooled}})$$

LRT	24.902
df	2
p-value	$p < 0.001$

The results of the LRT ($p < 0.001$) test signal that our increased parameters are in fact justified and provide a better overall model fit. Adding on, this provides statistical support that the makeup of Republican Senators is in fact different from that of Democrat Senators.

Below are the summary statistics as well as the distribution of lambdas when an NBD model was fit for Senators from each party separately.

Value	Senate Republicans	Senate Democrats
r	2.576	6.406
alpha	0.475	2.135
p-value	$p > 0.2$	$p > 0.2$

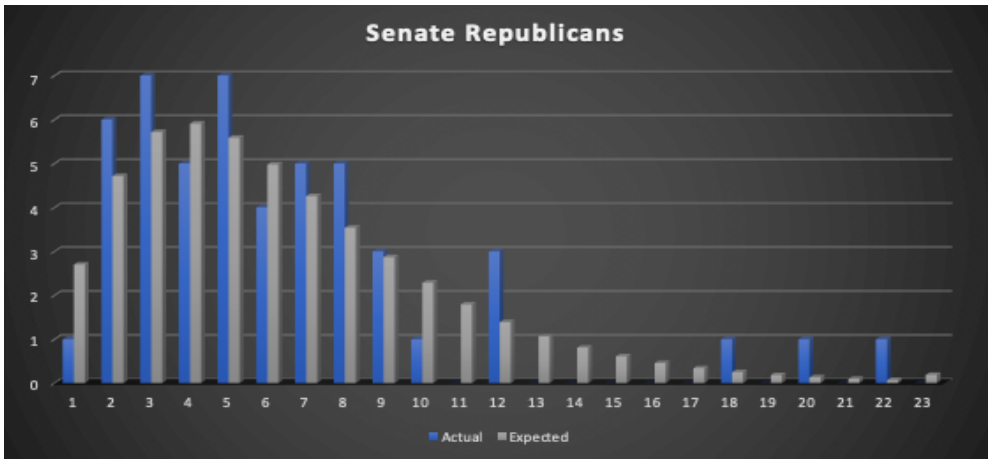

Figure 6: NBD on Senate Republicans

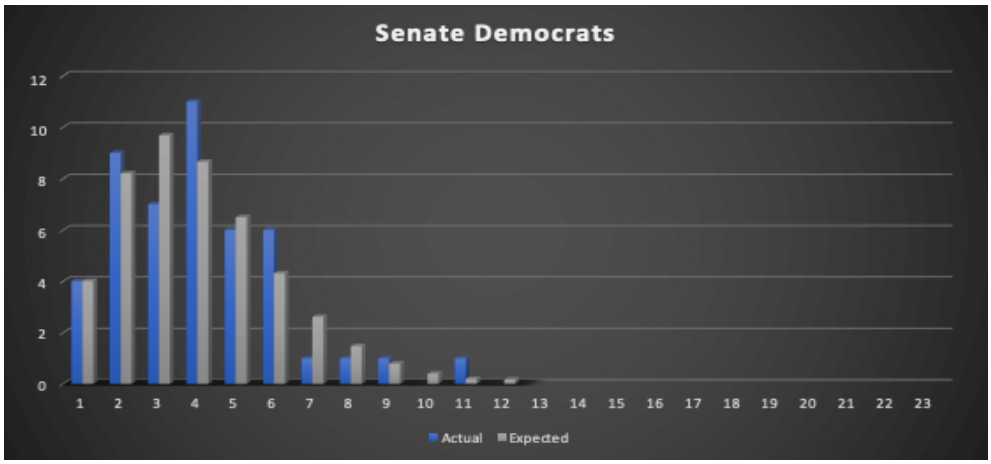

Figure 7: NBD on Senate Democrats

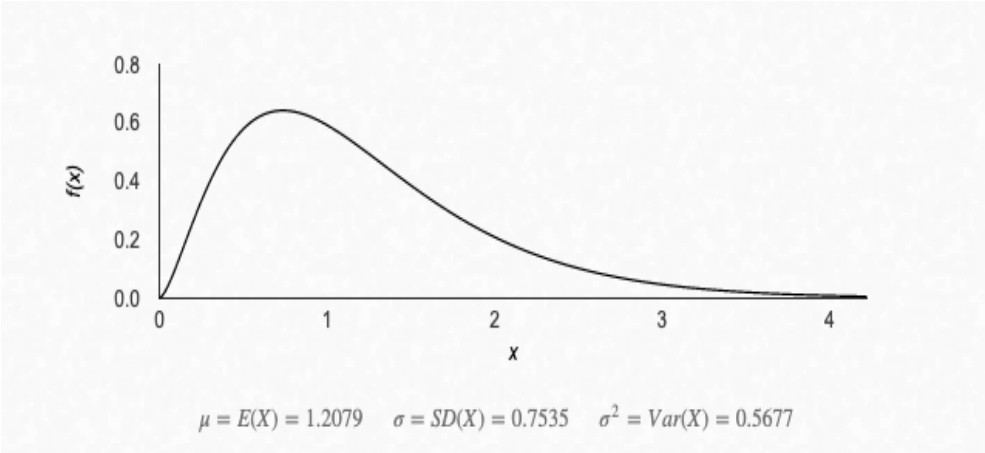

Figure 8: Distribution of λs for Senate Republicans

From the graphs, it is clear that interior mode for the Republican Senators is closer to the center whereas for the Democratic Senators it seems to move towards the left corner. Therefore, as shown by the observed data as well as the pre-analysis discussion, the r value of the Republican senators is lower than that of their fellow senators. As such, the GOP Senators have a higher heterogeneity. From the above distribution of lambdas based on the gamma parameters as resulted by the NBD models, it is also determined that the range of lambdas among the Democrat Senators is in fact higher than that of Republican Senators, a contradiction to our pre-analysis discussion. A few possible explanations for this could lie in the ideology of these parties, with Democrats pushing for more change as it relates to gun laws, minimum wage laws etc. Another possible explanation is that democrats tend to have younger representatives, who are more likely to pass bills and push for change. However, it is worthy to acknowledge that these are just a few possible explanations, and more concrete conclusions can be drawn from analyzing the datasets from previous versions of Congress before the 115th one.

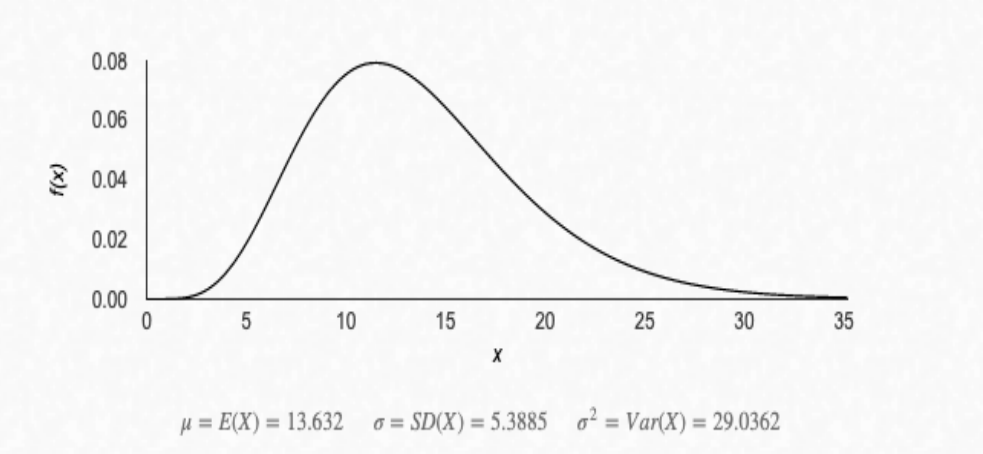

Figure 9: Distribution of λs for Senate Democrats

7.2 *The House of Representatives*

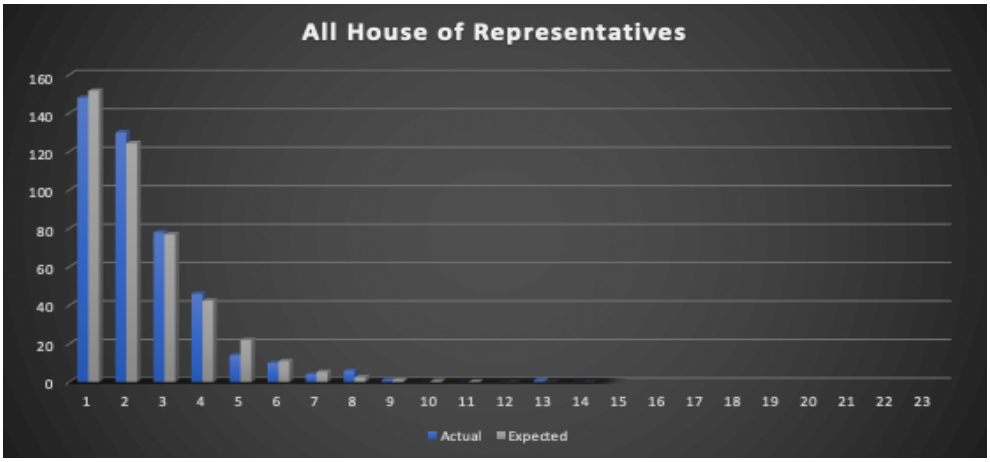

Figure 10: NBD on All House of Representatives

r	1.967
alpha	1.399
p-value	$p > 0.2$

The NBD models gives us a r-value of 1.96 and an alpha value of about 1.3. In comparison to the overall Senate, the house of representatives has a lower r value and therefore is more heterogeneous. In terms of alpha, the house has a higher alpha than the Senate and thus the distribution of lambdas is more broad. A possible explanation for this could lie in the makeup of the house – which has over 4 times the population of the Senate and is known to be far more diverse. Together these factors could lead to a broader range of lambdas as reflected in the alpha value. The chi-squared goodness of fit test also signifies that the fit of the model is quite well ($p > 0.2$). Graphically speaking, the graph of the

house looks better than that of the Senate, with almost all observed and expected values agreeing about the values at each cell. A possible reason for this could be that the number of bins in terms of observed data are fewer in the House when compared to the Senate, although they hold more values in each bin.

7.2.1 *Robustness*

The following three graphs (presented in the next two pages) represent the distribution of λs from the *gamma* distribution with parameters from their respective methods.

Method	MLE	MoM	Mean and Zeros
r	1.967	1.754	2.177
alpha	1.399	1.247	1.548

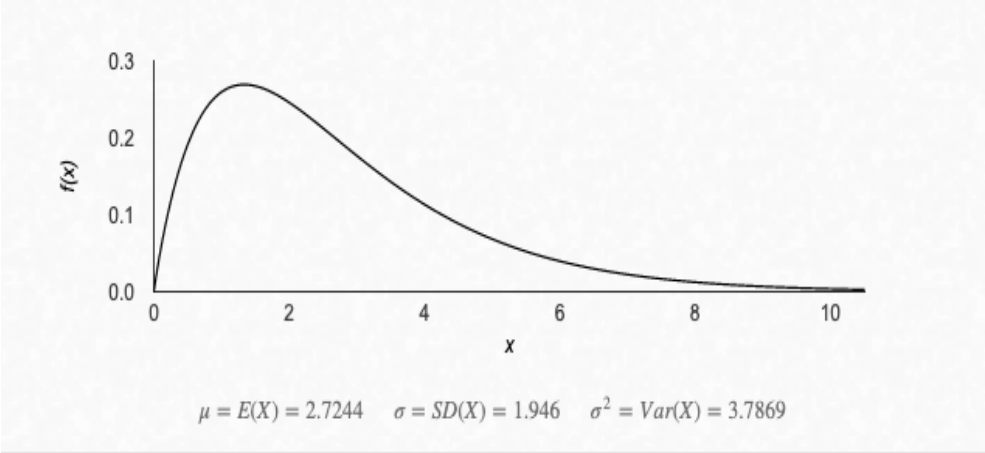

Figure 11: Distribution of λs based on parameters obtained via the Maximum Likelihood Estimation

Similar to the Senate, the MoM methods under reports both the r and the alpha value. For the Senate, we attributed this error to the long right tail. While a right tail does exist in this case, it is not as prominent for the House. Instead, we can notice that most of the observed values are segregated towards the left end of the graph, giving us a reason to believe that for the House our mean value might be the cause of inaccuracy. When comparing the distribution of lambdas based on the parameters from both these methods, it is clear that the most of the data seems to lie around the 2 mark. However, the graph from the parameters by the MoM has a smaller range of lambdas, with an upper bound of around 8 whereas the more accurate MLE has it beyond 10. Once again, similar to the Senate, the means and zeroes method over estimates both the r and the alpha values. Just like the MoM and MLE, this method also places most of the lambdas around the 2 mark, yet reports an even higher range of lambdas than both MoM and MLE. However, this method also reports far fewer data points at each of the x axis values, leading us to conclude that P(x=0) might be where the inaccuracy rises. This is because both the MoM

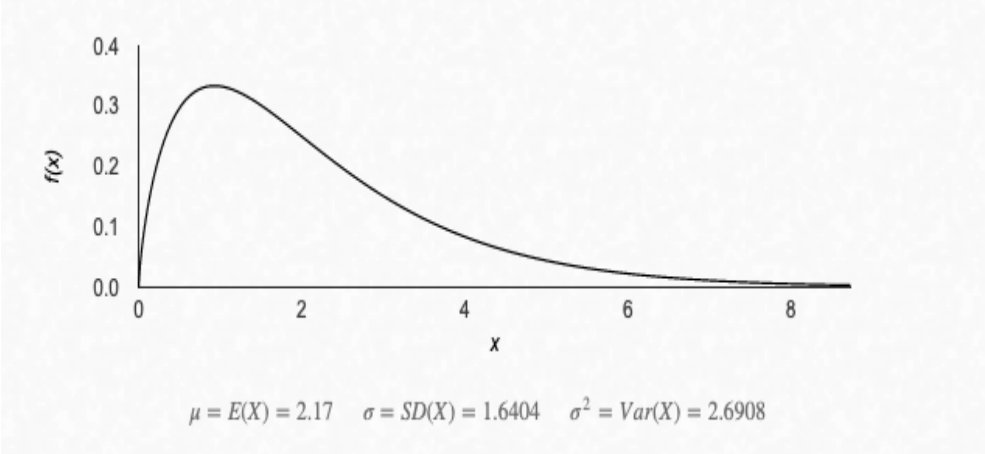

Figure 12: Distribution of λs based on parameters obtained via the Method of Moments

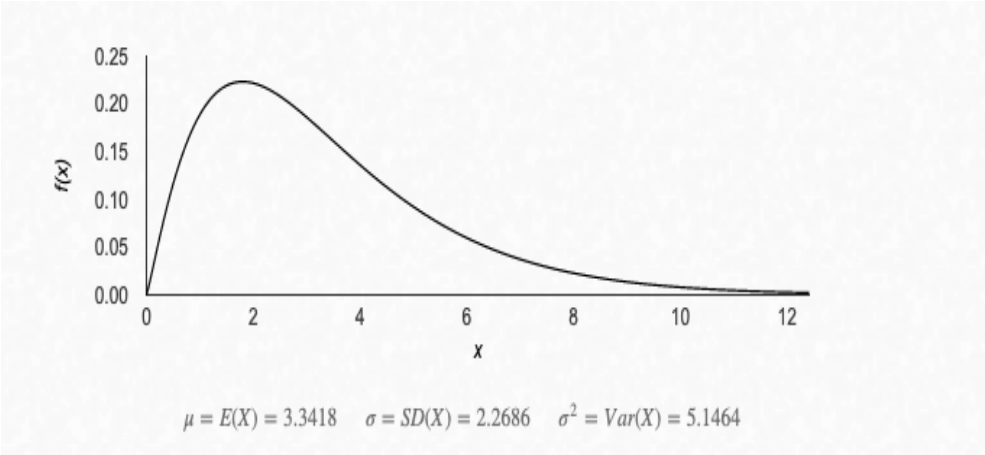

Figure 13: Distribution of λs based on parameters obtained via the Means and Zeroes Method

and means and zeroes use mean as one of there two data points, yet the MoM has the approximately the same y-axis scale as the MLE.

7.2.2 χ^2 *LRT Test*

Similar to the Senate, the chi-squared LRT was performed as presented below :

$$2 * (LL_{\text{Republican Senators}} + LL_{\text{Democrat Senators}} - LL_{\text{Pooled}})$$

LRT	44.328
df	2
p-value	$p < 0.001$

Just like the LRT of with the Senate data, the results of the LRT for the House ($p < 0.01$) also convey that extra parameters are in fact justified and provide a better overall model fit. Moreover, this once again provides statistical support that the makeup of House Republicans is in fact different from that of House Democrats.

Below are the summary statistics as well as the distribution of lambdas when an NBD model was fit for Representatives from each party separately.

Value	House Republicans	House Democrats
r	3.067	1.807
alpha	1.678	1.996
p-value	$0.01 < p < 0.2$	$p > 0.2$

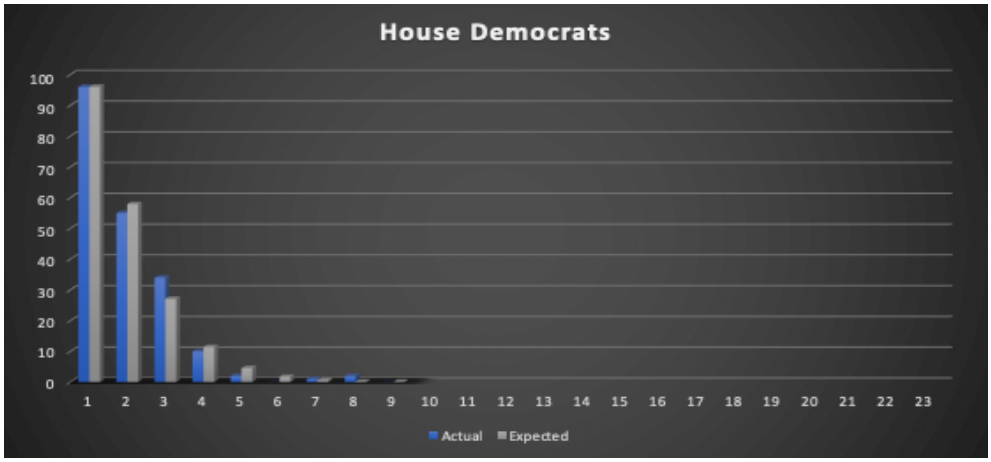

Figure 14: NBD on House Democrats

The parameters from these NBD models demonstrate that House Democrats have a higher alpha and a lower r value, and therefore are more heterogeneous than House Republicans. In fact, this is the opposite of the results from the Senate, in which the Senate Republicans were more heterogeneous. However, both chambers were controlled by Republicans leading us to incline in the direction that heterogeneity might not be affected by whether a party is the majority or the minority, at least for the 115th Congress. The graph of the distribution of lambdas for both parties displays that while most House Democrats seem to be around the x=2 mark, House Republicans have a higher average of around 4 mark. Moreover, House Republicans also seem to have a higher range with an upper bound beyond 15 whereas the Democrats have an upper bound around 13. Focusing on the histogram of these models, there seems to be quite an excess of observed data points around the x=2 cell for Republicans that the NBD falls short for. Although adding a spike at this cell would likely improve the log likelihood, we choose not to implement this inflated model as there is no good justification, or story, to warrant it. Visually speaking, the fit for the House Democrats seems to be better. Indeed, the chi-squared test agrees with this

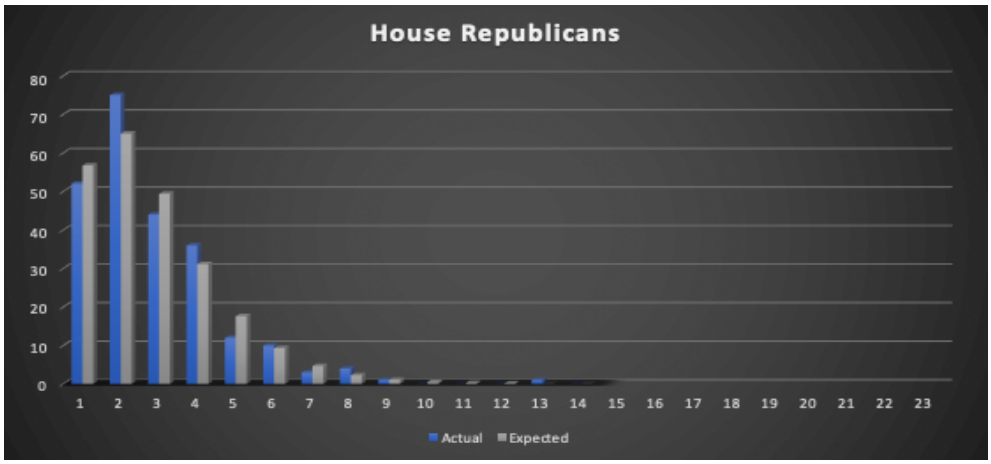

Figure 15: NBD on House Republicans

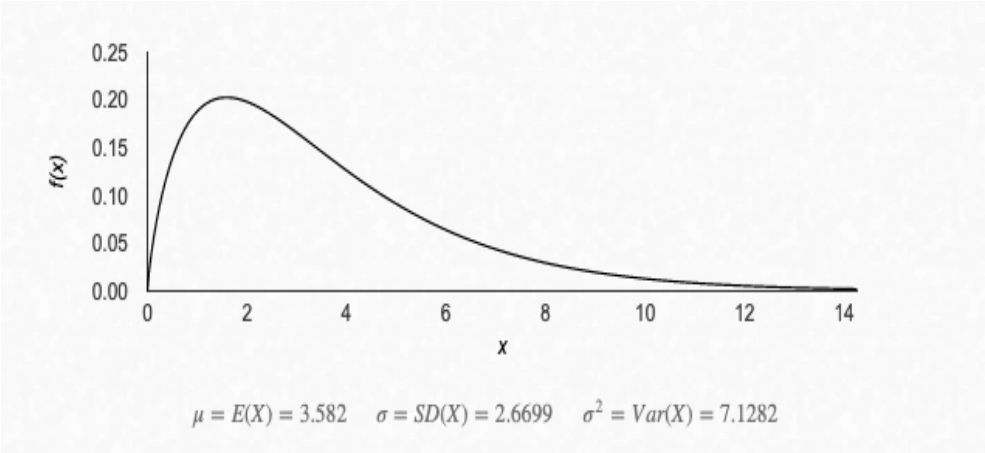

Figure 16: Distribution of λs for House Democrats

observation (p is greater than 0.2). On the other hand, although the fit graphically is acceptable for House Republicans, it doesn't meet the mark for the chi-squared ($p < 0.2$). With regards to the distribution of lambdas (on this page and the next page), although the range of lambdas is approximately the same, the mean of the lambdas is farther to the right for House Republicans and so is the concentration of lambdas.

8 Post-Analysis Discussion

In this section, we aim to answer the questions proposed earlier:

With most NBD models ran in this study having a good visually fit as well as high p-value from the chi squared goodness of fit test, it can be said that the NBD models can capture datasets political in nature quite well. From the above analysis, it has been seen that the House of Representatives is more heterogeneous in nature, as well has a broader range of representatives with varying propensities. This difference could be attributed to the fact that the make up of the members of the House is more diverse in nature. It

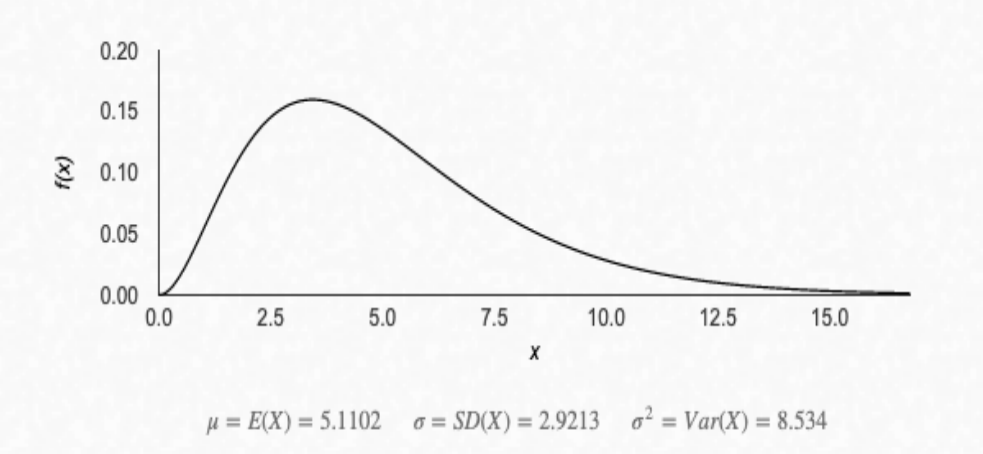

Figure 17: Distribution of λs for House Republicans

does not seem like the controlling party of either chamber has an observable effect on the propensity's of members to pass a law. As it relates to the 115th Congress, although both chambers were controlled by the GOP, the democrats were more heterogeneous in nature in the House while the Republicans were more heterogeneous in the Senate. There is no correlation to alpha as well. On average, republicans seem to be more heterogeneous than Democrats across both chambers, although a limitation of this conclusion is that the republicans controlled both chambers for the 115h Congress. Henceforth, a more concrete conclusion for this question can only be reached after analyzing data from when Democrats controlled either the Senate or the House. From the distribution of lambdas figures, we can reach the conclusion that for this Congress, Senate democrats is the group where members who have the highest propensity to enact a law are present. It can also be seen that Senate Republicans is where the most number of representatives with the lowest propensity to enact law are present.

It is important to note that the above conclusion have been reached by analyzing the 115th Congress. It is not appropriate to extrapolate these findings to make generalized statements about parties and chambers of congress or to other past or future versions of Congress.

9 Limitations of this Study

One of the main limitations of this study was that data for members who identify as Independents in the House of Representatives was not available on the website from where the data was obtained. Therefore, this is not a thorough analysis of the entire House of Representatives and only covers members from the two major parties – democrats and republicans. Another limitation is that due to the nature of politics, it is not likely that the lambdas of each representative remained constant throughout the 2 year period of the 115th Congress. It is reasonable to infer that politicians are more likely to pass bills near election season, and hence our assumption of stationarity is extremely limited. Furthermore, this analysis only considered categorizing representatives based on party. There are quite a

few other covariates that are present in this data, such as political ideology (moderate vs extreme), home state, time served, wealth to name a few. A more rich analysis could be achieved by looking at these divisions.

10 A Note to Future Researchers

While this study only looked at the number of laws enacted by each representative, there are a number of other types of count data that could perhaps be more valuable such as the number of bills introduced by each representative, number of bills introduced that were able to pass through sub-committee, number of votes or sponsors obtained for bills that can be considered bipartisan. Currently, it is difficult for ordinary citizens to obtain data that can provide them with a good picture of the legislative effectiveness of members of Congress. Potential future studies could explore the above mentioned count data, among many other options. It is the author's view that this data will be extremely valuable for a broad range of cases, from constituents wanting to know the legislative productivity of their representative, from opponents wanting to challenge incumbents based on these metrics, from social activist groups wanting to know the best politician to reach out to and from businesses wanting to increase the effectiveness of their lobbying efforts.

Acknowledgements

Frantzich, S. (1979). Who makes our laws? The legislative effectiveness of members of the U. S. Congress. Legislative Studies Quarterly, 4(3), 409. doi:10.2307/439582

Bruce Hardie's home page. (n.d.). Retrieved from https://www.brucehardie.com/

Gamma distribution Applet/Calculator. (n.d.). Retrieved from https://homepage.divms.uiowa.edu/ mbognar/applets/gamma.html