Winter, 2007 (Assignment Wk 6)

Comparison of Poisson Regression Model and Negative Binomial Regression Model: Population Ecology of Ethnic Newspapers in New York, 1725-1990

Jae-Woo Kim

1. Introduction

Newspapers, historically, have shown large variations in the founding and mortality rates (Carroll, 1987). This is why students of population ecology have examined environmental effects on selection processes with the data of newspapers. The goal of this study is 1) to examine what factors affect the founding rate of ethnic newspapers in New York, one of cities with high racial and ethnic diversity in the US, and 2) to better understand methodological issues that might arise when handling count outcome variables in the study of organizational ecology[1].

2. Theory

According to the density-dependent model of legitimation and competition (Hannan and Carroll, 1992) in the population ecology of organizations, the change in the number of an organizational population (N) with time (i.e., the change in the difference between the births and the deaths with time) is expressed by the sigmoid curve:

$dN/dt=aN-bN^2$ [Equation 1]

This formula is actually the same with the Lotka-Volterra equation, given that K is the carrying capacity (i.e., the non-zero steady state of the population) and r is the intrinsic

[1] My original plan was ambitious: Comparing one model to another model where the founding rate of political newspapers is included. Please, read Epilogue for this.

growth rate (i.e., the speed with which the population grows in the absence of resource constraints):

$$dN/dt=rN[(K-N)/K]=rKN-(r/K)N^2 \text{ [Equation 2]}$$

Comparing the two equations, it turns out that a and b in Equation 1 match rK and r/K in Equation 2. To get the big picture of social processes, the key point is that the effect of density is curvilinear because population increases have different consequences for legitimation and competition, depending on whether the density is low or high. The early range of density legitimates the organizational form itself and helps increase the founding rate, but this effect decreases with time because another addition no longer carry a significant legitimating effect. As the density increases further, while passing the point of inflection (i.e., it turns out that N is a/b at that point), the competitive process begins to operative over the legitimation process, which has the effect of discouraging additional births (Hannan and Carroll, 1989; Singh and Lumsden, 1990; Russell and Hanneman, 2002).

Next, following the institutional context, one should take into account the change in the institutional environments such as political turbulence and regulation policies, rather than technological environments that might be more critical to some other populations (e.g. corporations), although it cannot take effect in a vacuum (e.g. with concomitant change in the existing resource mobilization or dependence structure). For example, Blau and Elman (2002) hypothetically proposed that the founding rate of political newspaper is influenced by the structure of patronage possibility. Specifically saying, centralized patronage stifles the establishment of political newspapers, while both decentralized patronage and the absence of patronage encourage it. As another instance, Carroll (1987: Chapter 6) introduced some variables about the political environment such as political turbulences and conventional political activities in the models.

Relatedly, this is not exactly the institutional change, but the density of ethnic community over time is another important factor that might affect the founding rate of newspapers in general and ethnic newspapers in particular. Similarly, Russell and

Hanneman (2002) considered the lagged effects of immigration on the formation of worker cooperatives in Israel. As long as ethnic newspapers are concerned, it could be hypothetically proposed that the founding rate of ethnic newspapers will increase after large waves of immigration in the sense that such newspapers provide a mechanism of social control in ethnic communities and help to accelerate assimilation (Blau et. al., 1998).

Another major approach in population ecology literatures is based on population dynamics. As previous researches suggested, the cyclical patterns of organizational foundings over time might be due to the effects of prior organizational foundings and failures on the availability of resources. First, the disbanding of an existing organization may create free-floating resources which could be reassembled into new organizations. There is an upper limit, however, since an even larger number of deaths would signal an environment unfavorable to potential entrepreneurs, which thereby discourage foundings. Taken together, it turns out that there is a curvilinear relationship between current foundings and prior failures. Second, it could be proposed that the effect of prior foundings on current foundings is curvilinear, based on the similar logic (Delacroix and Carroll, 1983; Singh and Lumsden, 1990: 164)[2]

[2] As Russell and Hanneman(2002: 336-339) discussed it carefully, unlike the theory of density dependence, the population dynamics theory gives no consideration to the consequences of population changes for processes of legitimation. It confines its attention to their implications for processes of competition alone. They suggested that the frequent incidence of disconfirmatory results can be attributable to the negative consequences of recent deaths for legitimation. In this aspect, population ecology should incorporate the institutional approach in it more sophisticatedly. On the other hand, I still think that the number of births or deaths alone can neither guarantee the contagion (i.e, the occurrence of an event affects the rate of subsequent occurrence (Hannan and Carroll, 1992: 78-79)) nor explain different manifestation of its effect. In this aspect, the model of population dynamics in particular and the study of population ecology in general should have paid more attention to social network theory.

3. Methods

1) Data

I could not get any other raw data appropriate for the study of population ecology of organizations, one of my interest areas, except this: "Time-series for the Births and Deaths of Newspapers in Baltimore, Boston, Buffalo, New York City, Philadelphia, and Washington DC, 1690-1994." (ICPSR study no.4058) The data were collected from secondary data sources, including multiple sources for newspapers, United States historical censuses, immigration data, wars, recessions, unemployment rates, and election data. For most files, the year is the unity of analysis, which lends itself to Poisson count analyses or event history. The original data are composed of 91 subsets, but I extracted the only part of the data for New York City because I judged it provides the best information of organizational histories during the longest period. I focused only on one organizational species in this study: ethnic newspapers (1725-1990). In Figure 1, I presented the number of births and deaths of ethnic newspapers compared to the number of all types of newspapers in New York.

2) Variables and Measurement

The dependent variable is the number of ethnic newspapers' births every year from 1725 to 1990. With the help of the offset variable log(DEN), I am able to transform the dependent variable (BIRTH) into the logarithm of the founding rate. Drawing upon the model of density dependence, the first group of covariates can be composed of DEN (the total number of newspapers) and DENS (the squared DEN). Next, the model of population dynamics helps to build another group of covariates such as RBIRTH (the number of newspapers founded at the previous year), RBIRTHS (the squared RBIRTH), RDEATH (the number of newspapers that died at the previous year), and RDEATHS (the squared RDEATH). Besides, moving onto the institutional environments, fortunately, a couple of good dummy variables are already given in the original data here and there: DEP(1 for depression periods such as 1836-38, 1858-59, 1873-80, 1893-97, 1929-1940, 1948-49, 1973, and 1976; 0 otherwise) and WAR(1 for wars such as 1812-15, 1846-47, 1860-65, 1898-1900, 1913-20, 1940-45, 1950-53, and 1966-7; 0 otherwise). Lastly, focusing on the nestedness of newspapers in the ethnic community, I employed the following dummy variables. The first one is IMLAW for years of new immigration laws (1 if new laws; 0 otherwise). The next one is OP1(1st peak for old immigrants, 1847-57), OP2(2nd, 1865-74), OP3(3rd, 1880-93), OP4(4th, 1903-08), OP5(5th, 1921-29), OP6(6th, 1949-52), NP1(1st peak for new immigrants, 1899-1914), NP2(2nd, 1920-4), and NP3(3rd, 1978-88)[3].

Lastly, suffice to say here about why an offset variable should be incorporated in the model. For the sake of analysis, instead of the number of an event that happens during a unit period, it is more reasonable to use its rate. By inserting the logarithm of the total number of cases at risk – which fluctuates over time with the inflow and outflow of newspapers - during a unit period on the right side of the equation, I am able to do regression on the logarithm of the founding rate as the dependent variable:

$$Log(E(y))=log(exposure)+a+b_1x_1+\ldots \text{ [Equation 3]}$$

$$Log(E(y)/exposure)=a+b_1x_1+\ldots \text{ [Equation 4]}$$

Finally, I have got the following model:

$$Log(BIRTH)= log(EXPOSURE)+\alpha+\beta_1(DEN)_t+\beta_2(DEN)_t^2+\beta_3(RBIRTH)_{t-1}+\beta_4(RBIRTH)_{t-1}^2+ \beta_5(RDEATH)_{t-1}+\beta_6(RDEARH)_{t-1}^2+\beta_7(DEP)+\beta_8(WAR)+\beta_9(IMLAW)+\beta_{10}(OP1)+\beta_{11}(OP2)+ \beta_{12}(OP3)+\beta_{13}(OP4)+\beta_{14}(OP5)+\beta_{15}(OP6)+\beta_{16}(NP1)+\beta_{17}(NP2)+\beta_{18}(NP3) \text{ [Equation 5]}$$

---

[3] In the original data, each wave of German immigrants into the US was recorded as well, but in a little bit different way: such as 1st peak(1845-56), 2nd peak(1863-75), 3rd peak(1880-93), 4th peak(1901-14), and 5th peak(1920-31). After comparing two frequency tables of dummy variables from each classification, I decided not to use German immigrant waves because the population in my study is not German newspapers but ethnic newspapers.

3) Statistical Technique

A. Poisson Regression Model

Count variables are often treated as though they are continuous and the linear regression model is applied. But, using it for count outcomes can result in inefficient, inconsistent, and biased estimates. The Poisson regression model is the most basic model among alternatives (Scott, 1997: 217). This is why students of population ecology of organizations usually employ the Poisson regression model to examine the effect of independent variables on the founding rate of an organizational population. To put it another way, the Poisson process "serves as a natural baseline model for stochastic arrival processes." (Hannan and Carroll, 1992)

However, two issues - which are intertwined with each other - about its assumptions arise when doing social researches in general and organizational ecology study in particular. The first assumption is called equi-dispersion[4], which means that the conditional mean of the outcome should be equal to the conditional variance. The Poisson regression model rarely fits in practice since in most applications the conditional variance is greater than the conditional mean. If the mean structure is correct, but there is overdispersion, the estimates are consistent, but inefficient. Also, the standard errors will be biased downward, which makes z-values more significant (Scott, 1997: 230). "Either unobserved heterogeneity or positive contagion can generate overdisperson." (Hannan and Carroll, 1992: 79) In a word, "Failure to account for heterogeneity among individual in the rate of a count variable leads to overdispersion in the marginal distribution." (Scott, 1997: 221) In this situation, the negative binomial regression model can be suggested instead, since it allows the variance to exceed the mean by incorporating an error term in the Poisson regression model as I will discuss how exactly it can obviate this problem. More importantly, the negative binomial regression model is enough as long as the overdispersion problem is concerned, but not any more for another issue.

[4] For some other assumptions and properties of the Poisson distribution, see Scott(1997: 218-9). Also, he suggested doing the zero-inflated regression model especially when there are many 0s in the data, which is also related to the assumption of equi-disperson.

Another issue is none other than the assumption of independent events in the Poisson distribution, which means that when an event occurs it does not affect the probability of the event occurring in the future (Scott, 1997: 219). This criterion is also hard to meet in the reality because it is frequently that "the occurrence of an event affects the rate of subsequent occurrences." (Hannan and Carroll, 1992: also see p.241) This issue boils down to the problem of autocorrelation. "Another potentially important complication has so far escaped attention in research on organizational foundings – autocorrelation. It seems unlikely that the founding rate in one year is independent of the founding rate in the previous year, as is assumed implicitly in ML estimation of Poisson regressions and negative binomial regressions." (Hannan and Carroll, 1992: 80) The problem does not come from the model of population dynamics I discussed in the section of theory. Rather, the nature of time-series data entails the possibility of autocorrelation. However, as Hannan and Carroll (1992: 242) pointed out, this issue is not restricted to time-series analysis. Cross-sectional data of actors are subject to spatial autocorrelation: cross-sectional autocorrelation.

B. Negative Binomial Regression Model

In a sense, the Poisson regression model can be regarded as a special case of the negative binomial regression model (Hannan and Carroll, 1992). Despite this commonality, how to obviate the problem of overdispersion in the negative binomial regression model, then? As Scott (1997: 230-231) elucidates it, in the Poisson regression model, variation in a is introduced through only observed heterogeneity, which implies that different values of x result in different values of a (i.e., average here), but all individuals with the same x have the same a (Equation 6). In the negative binomial model, besides variation in x among individuals, unobserved heterogeneity is introduced by ε (Equation 7).

$a_i=E(y_i|x_i)=\exp(x_i\beta)$ [Equation 6]

$\tilde{a}_i=E(y_i|x_i)=\exp(x_i\beta+ \varepsilon_i)=a_i\exp(\varepsilon_i)$ [Equation 7]

However, unfortunately, as long as the problem of autocorrelation is concerned, the negative binomial regression model is not so helpful. It will be able to redress this problem, but not completely. Hannan and Carroll (1992, See Appendix B) argued that what is called "Quasi-likelihood Estimation" can be an alternative method to tackle problems caused by both overdisperson and autocorrelation.

4. Results

A. Poisson Regression Model

Compared to the deviance, it is better to use the likelihood ratio statistic (LRS) to test the global goodness of fit, since the deviance does not follow a chi-square distribution. That is the difference between the log likelihood of a baseline model including only the intercept and the log likelihood of a full model. Keeping in mind the concept of deviation, given -2(log likelihood of a baseline model) is -1072.90 (df=265) and -2(log likelihood of a full model) is -716.18 (df=247), it turned out that the LRS is 356.72 (df=18), which is statistically significant at the alpha level of 0.01. In other words, at least one of β is not 0. According to Table 1, the full model can be expressed as follows:

$$\text{Log(founding rate)}= -7.044+0.0432(DEN)_t-0.00157(DEN)_t^2+0.112(RBIRTH)_{t-1}-0.0128(RBIRTH)_{t-1}^2-0.0656(RDEATH)_{t-1}+0.00808(RDEARH)_{t-1}^2+0.365(DEP)-0.0662(WAR)-0.239(IMLAW)+1.483(OP1)+1.356(OP2)+0.452(OP3)-0.416(OP4)-0.543(OP5)-0.101(OP6)+0.649(NP1)+0.702(NP2)-0.764(NP3)$$

Turning to the local goodness of fit, one can interpret coefficients as we do in logistic regression models. For example, the log of the founding rate of ethnic newspapers increases by 0.0432 as their number (DEN) increases by a unit, controlling for all other explanatory variables. However, the incidence density ratio (IDR) helps to better understand their meanings in the substantive sense since the IDR corresponds to the odds ratio in logistic regression models. Relatedly, $(e^b-1)*100$ indicates the percent change in the expected founding rate with each one unit increase in a given explanatory variable. For instance, it could be said that the expected founding rate increases by the percentage of 4.4% with one unit increase in the number of ethnic newspapers.

First, the result indicates that the density dependence operates in the population of ethnic newspapers, although the negative effect of the square of their number is not that statistically significant. Regarding the model of population dynamics, the result cannot reject that there must be the positive effect of recent births and the negative effect of its square. Although the coefficients of recent deaths and its square are not statistically significant, the result interestingly shows the negative effect of recent death on the founding rate (in terms of legitimation effect) and the positive effect of the square of recent death (in terms of competition effect). Using the IDR, for example, a year with economic depressions has the founding rate higher by 44.08% compared to a year without them. Also, the founding rate decreased by 21.28% at a year when the new immigration law was made compared to the reference group.

Focusing on the effect of immigration, the first three peaks for old immigrants statistically significantly affect the founding rate of ethnic newspapers. This is not that surprising when taking into account the fact that German newspapers covered about 53.2% out of 190 all ethnic newspapers for old immigrants, and about 26.7% out of 378 for new immigrants. Besides, both of two peaks for new immigrants from 1899 to 1914 and from 1920 to 1924 also significantly contributed to the increase in the founding rate of ethnic newspapers. However, none of immigration waves after the mid-1920s influenced the increase in the founding rate. How to explain the negative effect of the rest of dummy variables on the founding rate (e.g. OP4, OP5, OP6, and NP3), then? From the codebook and Blau et.al. (1998), I found out that the restrictive immigration law was also enacted in those periods: Such as 1903 and 1907 for OP4; 1921, 1924, and 1927 for OP5; 1952 for OP6; 1978 for NP3. Nevertheless, one question still remain because such negative effect owing to the enactment of immigration laws did not appear for OP3 although they were passed three times (1882, 1885, and 1891) in that period.

To test the equi-dispersion assumption, the [deviance/df] presented by SAS is useful. If this is larger than 2, given the baseline is 1, one can suspect the over-dispersion. It seems that the variance ($4.032=2.008^2$) is much larger than the mean (1.421) as it is shown in Table 1. STATA provides the Likelihood-ratio test of alpha when the null hypothesis tells that $\alpha=0$ (Long and Freese, 2003: 269-270). Since G2 is 12.25 ($p<0.01$), which implies that there is a significant evidence of over-dispersion, negative binomial regression models are preferred in order to allow the variance to exceed the mean while maintaining the Poisson structure.

B. Negative Binomial Regression Model

In the same way as above, I presented the LRSs for different nested models and a full model in Table 3. All of them are statistically significant at the alpha level of 0.01, which means that at least one of coefficients in each model is not zero. The full model (Model 7 in Table 3) takes the form like:

$$\text{Log(founding rate)}= -7.093+0.0464(\text{DEN})_t-0.000225(\text{DEN})_t^2+0.150(\text{RBIRTH})_{t-1}-0.0149(\text{RBIRTH})_{t-1}^2-0.0770(\text{RDEATH})_{t-1}+0.00999(\text{RDEARH})_{t-1}^2+0.373(\text{DEP})-0.0438(\text{WAR})-0.260(\text{IMLAW})+1.420(\text{OP1})+1.278(\text{OP2})+0.461(\text{OP3})-0.338(\text{OP4})-0.435(\text{OP5})-0.139(\text{OP6})+0.654(\text{NP1})+0.643(\text{NP2})-0.755(\text{NP3})$$

The overall result in the negative binomial regression full model gives pretty much the same impression as the result from the Poisson regression full model, although the former is not generous to statistical significance.

5. Questions for this week

First, one of difficulties is how to make a new dataset of the number of political newspapers alive at each year as well as the number of political newspapers that were born and died at its previous year, given the information of the years of births and deaths of "political" newspapers. This is the main reason I could not achieve my original plan.

Second, how to deal with the lagged effect on the birth at the first year? The births at the previous year might affect the births at the next year according to the model of population dynamics, but this logic cannot be applied to the births at the starting time point.

Third, Figure 3 seems to imply that the negative binomial regression model is better fit in the data than the Poisson regression model, especially at the range closer to 0. However, although the pseudo-R2 cannot be the same as R2 in OLS, the pseudo-R2 provided by the Poisson regression model is much bigger. (0.3325>0.2016 for two full models in Table 2 and 3) How can I explain these contradictory results? Another, what about the global goodness of fit in the full model? (356.72 for Poisson, while 177.7 for Negative binomial) Is this also contradictory to Figure 3?

Another, I captured the small portion of my data in Figure 2 to show you how I calculated the number of newspapers at risk at each time point. But, what is the best way of setting an offset variable in the model?

Lastly, when it comes to one question at the last class, the Tobit model cannot be used for count outcomes, as far as I know. His question does not make sense from the beginning. I could not understand what you said that we can compare two results from the Tobit model and the Poisson model. It seems to me that Scott (1997) suggested what is called the zero-truncated poisson model or negative binomial regression model. I checked it, and it turned out that none of them is supported by SAS, unlike STATA.

6. Epilogue: What if two populations

When extending the Lotka-Volterra equation to two populations that compete with each other, one can get the following formula, which basically implies that the presence of the second (or first) population reduces the carrying capacity for the first (or second) population from K_1 to K_1-$\alpha_{12}N_2$ (or K_2 to K_2- $\alpha_{21}N_1$).

$$\frac{dN_1}{dt} = r_1 N_1 \left(\frac{K_1 - \alpha_{12} N_2 - N_1}{K_1} \right) \text{ [Equation 8]}$$

$$\frac{dN_2}{dt} = r_2 N_2 \left(\frac{K_2 - \alpha_{21} N_1 - N_2}{K_2} \right) \text{ [Equation 9]}$$

Here, two alphas are called "competition coefficient," which can be defined by the probability that a member of Population 1 will encounter a member of Population 2 at a particular resource position averaged over all resource positions divided by the probability that it will encounter a member of its own population at each position. In this sense, the competition coefficient corresponds to the concept of structural equivalence in social network analysis (DiMaggio, 1986; Burt, 1992)[5].

Following the logic behind Equation 8 and 9, I could add another variable representative of the competition among two populations in the extended forms of Equation 3 for one population, given two populations such as political newspapers (coded 1 below) and ethnic newspapers (2 coded below). When the dependent variables are the number of political newspapers founded at each year in the first model (BIRTH_1) and the number of ethnic newspapers in the second model (BIRTH_2), and offset variables are log(DEN_1) and log(DEN_2), the same covariates in Equation 3 can be used: 1) DEN_1 and DEN_2; 2) DENS_1 and DENS_2; 3) RBIRTH_1 and RBIRTH_2; 4) RBIRTHS_1 and RBIRTHS_2; 5) RDEATH_1 and RDEATH_2; 6) RDEATHS_1 and RDEATHS_2; 7) DEP, WAR,

---

[5] For example, "two organizations operate in the same niche to the extent that they recruit the same kinds of people. Such organizations are structurally equivalent with respect to the defined resource segments." (Ibid: 211) In this aspect, it could be suggested that the niche overlap in terms of structural equivalence adversely affects the founding rate by decreasing a in Equation 1. However, Hannan and Carroll (1992: 52-3) brought up one problem with the structural equivalence-based approach. "This proposal has the advantage of directing attention to patterns of dependence on environments including other organizational populations. We wonder, however, whether this approach can provide temporarily stable classifications of forms. It seems likely that small changes in ties in a network, resulting perhaps from the demise of one of the earlier organizations, will cause numerical clustering or blockmodeling procedures to yield quite different block structures. That is, it is likely that such a procedure is sensitive to relatively small perturbations in the observable data. If so, this approach will have difficulty identifying *enduring* bases of unit character of populations." Whatever it is, first of all, there are no such variables available in my data. Even if I took advantages of the information of political and ethnic affiliation of newspapers, the problem of temporal order still remains (i.e., the founding of newspapers followed their affiliation?). Rather, I would like to suggest using that information for event history analysis.

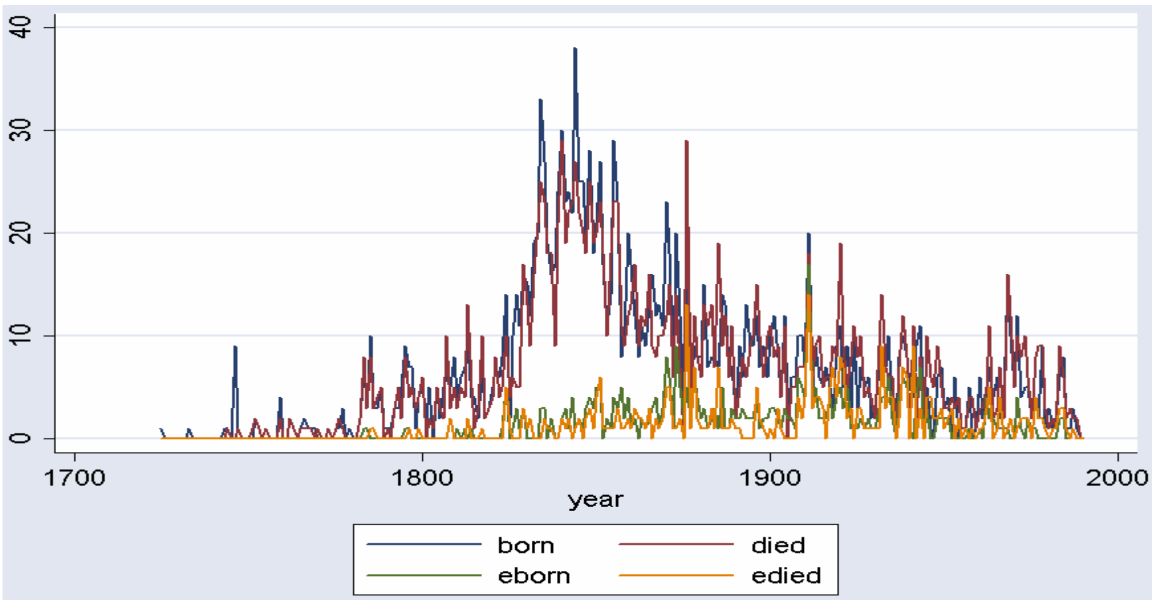
IMLAW, OP1 to OP6, and NP1 to NP3 for both models. Finally, the two models can be expressed as follows. Only one difference here is that I inserted DEN_2 in the first model and DEN_1 in the second model, in order to check whether there is the competition effect among two populations (Recall Equation 8 and 9).

$$\text{Log(BIRTH\_1)} = \log(\text{exposure\_1})+\alpha+\beta_1(\text{DEN\_1})_t+\beta_2(\text{DEN\_1})_t^2+\beta_3(\mathit{DEN\_2})_t+ \beta_4(\text{RBIRTH\_1})_{t-1}+\beta_5(\text{RBIRTH\_1})_{t-1}^2+ \beta_6(\text{RDEATH\_1})_{t-1}+ \beta_7(\text{RDEARH\_1})_{t-1}^2+\beta_8(\text{DEP})+\beta_9(\text{WAR})+\beta_{10}(\text{IMLAW})+\beta_{11}(\text{OP1})+\beta_{12}(\text{OP2})+ \beta_{13}(\text{OP3})+\beta_{14}(\text{OP4})+\beta_{15}(\text{OP5})+\beta_{16}(\text{OP6})+\beta_{17}(\text{NP1})+\beta_{18}(\text{NP2})+\beta_{19}(\text{NP3}) \quad \text{[Equation 10]}$$

$$\text{Log(BIRTH\_2)} = \log(\text{exposure\_2})+\alpha+\beta_1(\text{DEN\_2})_t+\beta_2(\text{DEN\_2})_t^2+\beta_3(\mathit{DEN\_1})_t+ \beta_4(\text{RBIRTH\_2})_{t-1}+\beta_5(\text{RBIRTH\_2})_{t-1}^2+ \beta_6(\text{RDEATH\_2})_{t-1}+ \beta_7(\text{RDEARH\_2})_{t-1}^2+\beta_8(\text{DEP})+\beta_9(\text{WAR})+\beta_{10}(\text{IMLAW})+\beta_{11}(\text{OP1})+\beta_{12}(\text{OP2})+ \beta_{13}(\text{OP3})+\beta_{14}(\text{OP4})+\beta_{15}(\text{OP5})+\beta_{16}(\text{OP6})+\beta_{17}(\text{NP1})+\beta_{18}(\text{NP2})+\beta_{19}(\text{NP3}) \quad \text{[Equation 11]}$$

7. Appendix

<Figure 1> Births and Deaths of Ethnic Newspapers (eborn, edied) and All Types of Newspapers (born, died): 1725-1990



<Table 1> Descriptive Statistics

Variable	N	Mean	Std.Dev.	Min	Max
BIRTH	266	1.421	2.008	0	17
DEN	266	26.109	26.109	0	88
RBIRTH	266	1.421	2.008	0	17
RDEATH	266	1.312	2.071	0	14

<Table 2> Poisson Regression: Saturated Model with Explanatory Variables

Variable	b	std.error	IDR	95% CI	
DEN	.0431714***	.0104815	1.044117	1.022886	1.065788
DENS	-.0001573	.0001361	.9998427	.999576	1.000109
RBIRTH	.1122291*	.0594527	1.118769	.9957132	1.257033
RBIRTHS	-.0127634**	.005061	.9873177	.9775725	.99716
RDEATH	-.0656476	.0653622	.9364609	.8238596	1.064452
RDEATHS	.0080807	.0059955	1.008113	.9963365	1.02003
DEP					
No				Reference	
Yes	.365219**	.1694367	1.440829	1.033684	2.00834
WAR					
No				Reference	
Yes	-.0662231	.1650072	.9359221	.6773067	1.293284
IMLAW					
No				Reference	
Yes	-.2392201*	.1399027	.7872416	.5984428	1.035603
OP1					
No				Reference	
Yes	1.482943***	.2329145	4.405891	2.791107	6.954902
OP2					
No				Reference	
Yes	1.355845***	.2215701	3.88004	2.513248	5.99014
OP3					
No				Reference	
Yes	.45201**	.2263128	1.571468	1.008481	2.448743
OP4					
No				Reference	
Yes	-.4160552	.3340305	.6596439	.3427538	1.269512
OP5					
No				Reference	
Yes	-.5431355	.3505445	.5809239	.292237	1.154791
OP6					
No				Reference	
Yes	-.1007117	.4285671	.9041937	.3903593	2.094394
NP1					
No				Reference	
Yes	.6491888***	.2416363	1.913988	1.19195	3.073409
NP2					
No				Reference	
Yes	.7024469*	.3725949	2.018686	.9725579	4.190078
NP3					
No				Reference	
Yes	-.7640829	.473953	.4657609	.1839643	1.179214
Intercept	-7.044				
Pseudo R2	0.3325				
LRS	356.72				
Df	248				
Sample size	266				

<Table 3> Negative Binomial Regression: Comparison of Nest Models and a Full Model

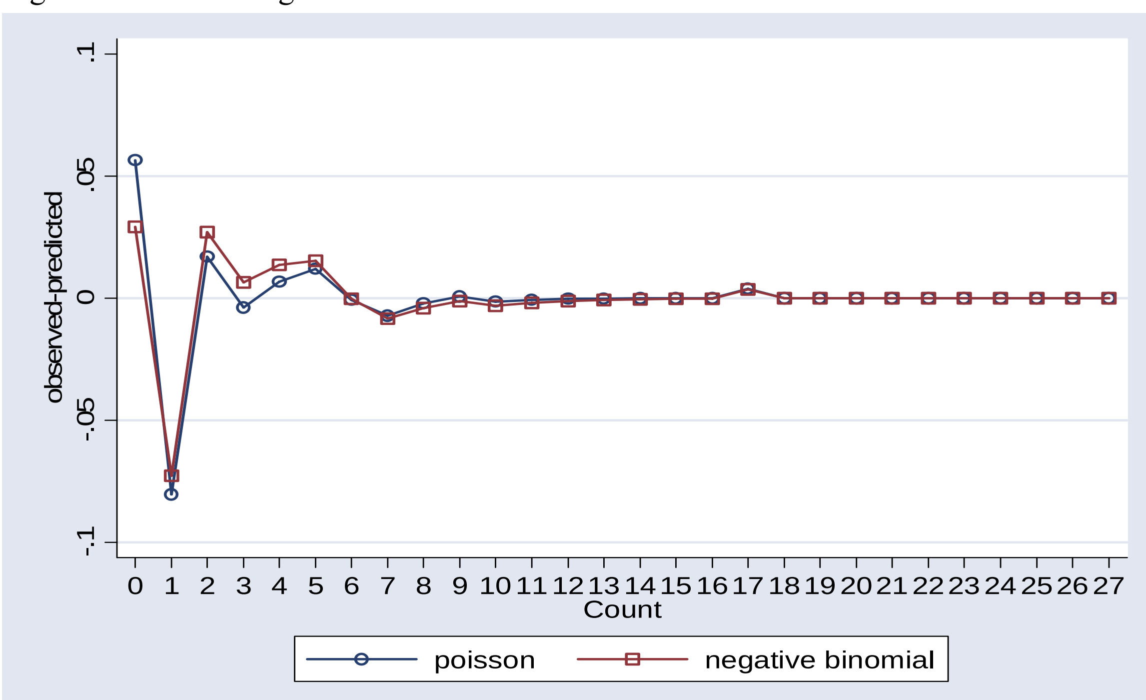
	Model 1		Model 2		Model 3		Model 4	
Variable	b	SE	b	SE	b	SE	b	SE
DEN	.0675***	.0105	.0526***	.0107	.0644***	.0108	.0531***	.0107
DENS	-.000517***	.00014	-.000422***	.000140	-.000520***	.000144	-.000428***	.00014
RBIRTH			.329***	.0666			.354***	.0791
RBIRTS			-.0176***	.00554			-.0208***	.00761
RDEATH					.127	.0784	-.0529	.0910
RDEATS					-.00748	.00760	.00647	.0101
DEP								
WAR								
IMLAW								
OP1								
OP2								
OP3								
OP4								
OP5								
OP6								
NP1								
NP2								
NP3								
Intercept Pseudo R2 LRS Df	-6.897 .118 104.2 264		-7.066 .149 131.25 262		-6.947 0.122 107.76 262		-7.056 0.150 131.68 260	
	Model 5		Model 5		Model 6		Model 7	
Variable	b	SE	b	SE	b	SE	b	SE
DEN	.0624***	.0120	.0381***	.0109	.0627***	.0112	.0464***	.0126
DENS	-.000490***	.000148	-.0000729	.000148	-.000607***	.000159	-.000225	.000166
RBIRTH	.333***	.0787	.247***	.0716	.293***	.0795	.150**	.0729
RBIRTHS	-.0210***	.00750	-.0203***	.00658	-.0173**	.00752	-.0150**	.00640
RDEATH	-.0693	.0906	-.122	.0815	-.00140	.0903	-.0770	.0804
RDEATHS	.00845	.00998	.0135	.00848	.00330	.00978	.00999	.00803
DEP	.00123	.202					.372*	.214
WAR	-.00543	.202					.0438	.207
IMLAW	-.426**	.180					-.260	.174
OP1			1.398***	.284			1.420***	.284
OP2			1.271***	.286			1.278***	.281
OP3			.289	.262			.461*	.279
OP4			-.123	.367			-.338	.427
OP5			-.458	.325			-.435	.417
OP6			-.325	.498			-139	.317
NP1					.531*	.289	.654**	.463
NP2					.0728	.446	.643	.506
NP3					-1.166**	.514	-.755	.169
Intercept Pseudo R2 LRS Df	-7.064 0.156 137.5 257		-7.119 0.1854 163.4 254		-7.053 0.160 140.8 257		-7.093 0.2016 177.7 248	

* For the sake of comparison, I did not include some other measures such as $\chi 2$s, IDRs, and 95% CIs.

<Figure 2> How to Calculate an Offset Variable at Each Year (e.g. 1822-1842)

	A	B	C	D
1	year	birth	death	exposure
2	1822	0	0	378
3	1823	1	0	378
4	1824	5	5	377
5	1825	0	1	377
6	1826	1	0	378
7	1827	3	0	377
8	1828	1	0	374
9	1829	3	3	373
10	1830	0	1	373
11	1831	0	2	374
12	1832	0	0	376
13	1833	0	0	376
14	1834	3	0	376
15	1835	3	2	373
16	1836	0	1	372
17	1837	2	1	373
18	1838	0	0	372
19	1839	0	0	372
20	1840	2	2	372
21	1841	3	1	372
22	1842	1	2	370

<Figure 3> Fitting the Data: Comparison of a Poisson Regression Full Model and a Negative Binomial Regression Full Model



* I followed the steps provided by Scott and Freese (2003: 283-284)

8. References

Blau, Judith et. al.. 1998. “Ethnic Buffer Institutions: The Immigrant Press, New York City, 1820-1984.” *Historical Social Research* 23(3).

Blau, Judith and Cheryl Elman. 2002. “The Institutionalization of US Political Parties: Patronage Newspapers.” *Sociological Inquiry* 72(4).

Carroll, Glenn. 1987. *Publish and Perish: The Organizational Ecology of Newspaper Industries.* JAI Press, Inc..

Delacroix, Jacques and Glenn Carroll. 1983. “Organizational Foundings: An Ecological Study of the Newspaper Industries of Argentina and Ireland.” *Administrative Science Quarterly* 28.

DiMaggio, Paul. 1986. “Structural Analysis of Organizational Fields: A Blockmodel Approach.” In *Research in Organizational Behavior*. Edited by Barry Staw and Lawrence Cummings. JAI Press.

DiMaggio, Paul and Walter Powell. 1983. “The Iron Cage Revisited: Institutional Isomorphism.” *American Sociological Review*. 48.

Hamilton, Lawrence. 2004. *Statistics with STATA: Updated for Version 8*. Brooks-Cole/ Thompson Learning.

Hannan, Michael and John Freeman. 1989. *Organizational Ecology*. Harvard University Press.

Hannan, Michael and Glenn Carroll. 1992. *Dynamics of Organizational Populations: Density, Legitimation, and Competition.* Oxford University Press.

Long, Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Sage Publications.

Long, Scott and Jeremy Freese. 2003. *Regression Models for Categorical Dependent Variables Using STATA.* A Stata Press Publication.

Meyer, John and Brian Rowan. 1977. “Institutionalized Organizations: Formal Structure as Myth and Ceremony.” *American Journal of Sociology*. 83(2).

Rusell, Raymond and Robert Hanneman. 2002. “The Role of Institutional Processes in the Formation of Worker Cooperatives in Israel, 1924-1990.” In *Advancing Socio-Economics: An Institutional Perspective*. Edited by J. Rogers Hollingsworth et.al. Rowman and Littlefield Publishers, Inc..

Singh, Jitendra and Charles Lumsden. 1990. “Theory and Research in Organizational Ecology.” *Annual Review of Sociology* 16.