Modeling Monthly Birth/Death by using Sample Paradata from the Current Employment Statistics Survey October 2013

Jeremy Oreper[1]
[1]Bureau of Labor Statistics, 2 Massachusetts Avenue NE, Washington, DC 20212

Abstract

In the Current Employment Statistics (CES) survey, the monthly employment change as a result of out of sample births and deaths of establishments is modeled using an ARIMA time series. Five years of historical data, derived from the State Unemployment Insurance (UI) counts of net birth/death employment, serve as the inputs to the model. This model lacks any information current to the time that estimates are calculated, and as a result errors can increase when birth/death changes in a different pattern than the previous five years. To overcome these limitations a model was specified using paradata. The number of respondents able to report for the current month by the first deadline was combined into a modified Lotka–Volterra equation. This model has the advantage of using data contemporaneous with the production of the first release of estimates. The estimate of the birth/death using first release paradata and birth/death residuals from 2003-2012 are used in two year increments to forecast the monthly birth/death values for March 2006 through March 2012.

Key Words: Paradata, Birth/Death, Lotka-Volterra, ARIMA, Current Employment Statistics

1. Introduction

1.1 Current Employment Statistics Paradata

The Current Employment Statistics (CES) Program[1] is a monthly establishment survey that releases estimates of employment, hours, and earnings of workers on nonfarm payrolls. The survey collects data from establishments monthly; however not all data is available by the first release of the estimates. CES first preliminary estimates are published each month approximately three weeks after the reference period. Estimates are then revised twice before being held constant until the annual benchmarking process. Second preliminary estimates for a given month are published the month following the initial release, and final sample-based estimates are published two months after the initial release. If data is received after the final release or 'closing' it is not used but is still recorded.

It is desirable to receive as much of the data as possible by the first closing. Inevitably, some respondents will report their data in time for latter closings, after 3rd closing, or not at all. CES studied administrative variables that might allow prioritization of prompting

[1] The Current Employment Statistics (CES) program is a monthly survey of about 145,000

and assigning resources towards those establishments most likely to be either non-respondents or late respondents.

The initial investigation into the prediction of respondent responsiveness was confounded by the impact of employment changes in the reported micro data. When an establishment has large changes in their reported employment there are also changes in their reporting behavior. Respondents that formerly were consistent first closing reporters would stop reporting or suddenly report at latter closings when reported employment suddenly changes. Administrative variables explain changes to the timing of response only so much as they were correlated to reported employment changes. The observed relationship was investigated from the position that the connection between responsiveness and employment change was an indirect one. Changes in responsiveness translated to large macro employment changes through both being concurrent due to a third variable, the stress of the respondent themselves. The value of this linkage is that paradata describing responsiveness of establishments can be used to improve the accuracy of the estimates.

1.2 What is Birth-Death?

The monthly employment estimates have two components. A weighted link-relative estimator uses the sample trend in the cell to move the previous level to the current-month estimated level and a second smaller model-based component is used to account for the net employment difference between business births and deaths not captured by the sample. The sample based estimates account for most of the monthly change in employment. However, the sample is drawn yearly from the Quarterly Census of Employment and Wages (QCEW)[2], and as a result there are some establishments opening or closing outside the frame in between sample rotations. The process of establishment opening and closing is also known as establishment births and deaths.

A two component correction is used to account for the net change to the employment level from births and deaths. First, establishments do not have their employment removed from an estimate when they stop reporting and are instead updated by the monthly change of their industry. Estimates are produced by calculating a sample link based on the continuing units that is applied the prior month's employment estimate of a given industry cell. Units that do not respond in the current month due to either non-response or establishment death will have their employment contribution to the cell advanced by the sample link, an implicit form of imputation. This is done because research[3] for 1995-2007 showed that the employment changes from births and from death came very closely to canceling each other out. As a result, any employment that is lost to a business death is being immediately replaced by an offsetting birth by not allowing the employment contribution from the dead unit to drop out from the estimation cell. The second step is to model the small residual net employment change from the offset in births and deaths called the net birth/death. To account for this birth/death residual, an auto-regressive

[2] The Quarterly Census of Employment and Wages (QCEW) program publishes a quarterly count of employment and wages reported by employers covering 98 percent of U.S. jobs, available at the county, MSA, state and national levels by industry. To access QCEW data see http://www.bls.gov/cew/.

[3] Summary of Birth/Death Regression Variable Research, Nathan Clausen, Victoria Battista U.S. Bureau of Labor Statistics, 2 Massachusetts Ave NE, Washington, DC 20212

integrated moving average (ARIMA) time series model is used to forecast the birth/death employment change left over after imputation.

$$[1]\ Birth/Death\ residual\ =\ Population - Samplebased\ estimate\ +\ Error$$ [4]

2. Modeling Birth/Death Using Paradata

2.1 The Distributed Closing Model

All forecasts, including the net birth/death, have inherent problems. Forecast only use historical values as inputs making it difficult for a model to accurately capture turning points in an employment series. Input data for net birth/death forecasting lags 9 months behind the publication of first closing employment estimates, and as a result will miss any changes that occur in the intervening months. Using paradata presented a possibility for investigating if the changes over time in the number of respondents reporting at a given closing could provide an independent and concurrent supplementary variable for predicting the birth/death residual.

A linear regression, named the Distributed Closing Model was specified that used the count of reporters successfully reporting micro data at a given closing. This model can be thought of as a rough or reduced form of the final specification.

$$[2]\ NBD = \beta_0 + \beta_1 Nmiss + \beta_1 \Delta Nmiss + \beta_3 Late + \beta_4 \Delta Late + u$$

The model was estimated controlling by both Month and Industry. Here NBD is the net birth/death, the dependent variable, with $Nmiss$ being the count of respondents eligible to report that did not report for a month, $\Delta Nmiss$ the over the month change in the number of respondents that did not report, $Late$ is the number of respondents reporting after the first closing of the month, and $\Delta Late$ is the over the month change in the number of respondent reporting after the first closing. The rationale behind this model was that it meant to capture the stress level of both the establishments that were reporting in the survey, and the individuals themselves reporting in each establishment.

When establishments are under stress it is difficult or impossible for them to report their employment counts by the first release of the data. This could be because they are closing down operations, splitting up the company, merging, opening new locations, or performing any number of stressful actions that would strain the ability of their payroll offices to respond to requests for filing their data. The mechanism behind the Distributed Closing Model that connects response behavior with changes in employment depends on the stress level of the personnel in the payroll office to be the common element. Additionally, the individuals who do the reporting are much more likely to be transferred, laid-off, promoted, or otherwise overwhelmed during times of flux and transition and so change their ability or willingness to respond. Establishment births and deaths are inherently disruptive activities and so disruptive shifts in the employment level could be reflected in disruptions to reporting behavior. The Distributed Closing Model connects the churn and upheaval in respondent behavior with changes in establishment births and

[4] For more information on the birth/death residual concepts and methodology, see “Technical notes to establishment survey data,” http://www.bls.gov/ces/#technical.

deaths. Firms that report large batch files through the Electronic Data Information center where automated transmissions deliver data electronically were excluded from the input data to this model because of their lack of a defined point of contact to be stressed.

Table 1: Comparison of Year 1 net birth/death values with Distributed Closing Model Predictions from April 2006 through March 2009

Date	Year 1 net birth/death (Thousands)	Distributed Closing Prediction
06 to 07	1,023	741
07 to 08	540	502
08 to 09	-215	-140
Net Contribution (Apr-Mar)		

This initial model proved to be informative and successful at fitting the actual historical values of the birth/death residual, but it also suffered from depending on information that cannot be forecasted or known contemporaneously with the timing of the first release of estimate. The significance of the Distributed Closing Model was that it serves as a reduced form for a later structural form based on the Lotka-Volterra Predator-Prey Model.

2.2 Limitations of the Distributed Closing Model

The Distributed Closing Model was altered to use the timestamp of when a respondent's data was first received before first closing so as to specify a model that would be ready by the time of first closing. The Distributed Closing Model was able to make successful predictions by measuring the flow of respondents from earlier to later closings and back again, and there was interest in trying to see if a continuous distribution of time stamps leading up to first closing could correlate with net birth/death as well as the variation across counts of respondents by closings.

However, this approach suffered from several problems. The first issue was that when data was missing--not reported by an establishment--it was difficult to determine a value that should be imputed for the date time stamp without distorting the meaning of the other values being inputted to the model. A second issue was that using date time stamps have a variability that was not meaningful. The CES survey references the pay period that includes the twelfth of the month when asking establishments for their employment levels for a given month. Establishments will often not have their employment counts until just before first closing, making it unlikely that they could send data earlier than their pay periods allow. CES sends email and telephone prompts to self respondents in the final week before the first closing deadline while computer assisted telephone calls are scheduled as soon as the reference pay period ends. For any given first closing, the collection period only varies between 10 and 16 days. This means that a large proportion of the data is reported in a compressed period of time that varies more on the calendar than on the stress level of the respondents or changes within an establishment. Later models would avoid this by aggregating responses by the closing that that came in by rather than by the day.

The final limitation of the Distributed Closing Model was the need for $Nmiss$ to be included in the regression, which can only be known after the final employment estimate release for a month. The regression uses the movements of respondents between the different possible categories of response: first closing, later than first closing, and missing. First closing is not included in the specification because of the need to prevent collinearity between the variables, adding together all three categories will always result in them summing together to equal the sample size. $Nmiss$ is crucial to any effective specification of the Distributed Closing Model, rearranging the terms to have first closing counts used, without $Nmiss$, led to poor correlations.

3. Lotka-Volterra Equations and the Establishment Predator-Prey Model

3.1 A Naïve Transformation to Predator-Prey

At this point in the research the focus moved away from determining whether or not there was sufficient proof of concept to look for an improvement on the ARIMA forecast, to whether or not paradata could be used to find a functional form for describing the net birth/death. The Distributed Closing Model was developed from looking at a wide variety of variable forms plugged in through stepwise regressions. This proved to be a productive way of analyzing novel datasets, but this meant that it was unclear what elements of the regression were necessary to describe the process of births and deaths in the establishment population and which were simply placed there by the stepwise process and could be removed with an exact functional form.

The need for a functional form that describes the birth/death process led to Lotka-Volterra Predator-Prey Model. The Predator-Prey function consists of a pair of simultaneous equations that define how a population changes with respect to time. These equations are:

$$[3.1]\ \frac{dx}{dt} = (b - py)x \qquad [3.2]\ \frac{dy}{dt} = (rx - d)y$$

Integrating for the total conserved population in the system yields:

$$[3.3]\ z = b\ln y_{(t)} - py_{(t)} - rx_{(t)} + d\ln x_{(t)} + C$$ [5]

Table 2: Definition of terms in equation [3.3] [6]

x= Population of Prey	y= Population of Predators	z=Total population (x+y)	b=Birth rate of x in the absence of y

d= Death rate of y in the absence of x	r =Growth rate of y in the presence of x	p= Loss rate of x in the presence of y	t=Time
			C=Constant

[5] http://onlinemathcircle.com/wp-content/uploads/2012/03/Lotka-Volterra-Equations.pdf
[6] Frank Hoppensteadt (2006) Predator-Prey Model. Scholarpedia, 1(10):1563.

Initial inspection of the function for calculating the total population shows a similarity of form.

$$[4.1]\ NBD = \beta_0 + \beta_1 Nmiss + \beta_1 \Delta Nmiss + \beta_3 Late + \beta_4 \Delta Late$$

$$[4.2]\ z = b \ln y_{(t)} - \ py_{(t)} - rx_{(t)} + d \ln x_{(t)} + C$$

Furthermore, the use of the Predator-Prey Model is appealing due to its explicit inclusion of birth and death affects in determining the size of a population with respect to time, approximately the same type of problem as estimating the births/deaths of establishments. The different form of the Predator-Prey Model, introducing non linear effects through the use of logarithms in place of over the month changes, creates the possibility of being able to form a regression without the use of $Nmiss$. The inclusion of exponential effects replaces the need for $Nmiss$.

The first attempt was to try and run a regression where NBD is z and the regression Betas estimate the coefficients of the Predator-Prey Model. $First$ in the equation below represents the count of respondents reporting by first closing.

$$[5]\ NBD = \beta_0 + \beta_1 First + \beta_1 \ln First + \beta_3 Late + \beta_4 \ln Late$$

This model worked approximately as well as the original Distributed Closing Model for fitting historical values, but it too did not produce accurate forecasts. The lagged release of the net birth/death means that any model must be able to calculate coefficients using historical values that can be used in forecasting current net birth/death. This form also suffered from being adopted from a function that calculates z, which is a population and not a net change of population. The result is that $NBD = z_t - z_{t-1}$ with the z_{t-1} being lumped in with the linear intercept term β_0, making forecasting of the intercept both difficult and crucial to the strength of the regression.

In equation [5] the net birth/death is estimated using five years of actual birth/death value. The values being used in the regression are all counts that are not adjusted for sample size. The CES sample does not change rapidly or drastically, but there are adjustments made to it that can accumulate over time. This means that the interpretation of the values for the input variables can differ widely over time without some adjustment to account for the sample size changes. It is similar to the original problem of how to include the count of missing respondents in the regression before it can be known which are missing. It is not possible to use a registry of administrative data because this is also generated from inputs by the QCEW, which are lagged behind CES by exactly as much as the net birth/death itself. To address the problem of correcting for sample size a new paradata metric was created to supplement the existing variables, called a synthetic denominator.

3.3 Additional Paradata: The Synthetic Denominator

Each year the CES survey enrolls new establishments to replace the ones that will be dropped as part of sample rotation in order to reduce respondent burden. However, it can take up to a year to have respondents at the new establishments initiated, and often several months between first contact and collection of data items. Some newly selected firms will refuse to participate, others will enroll but drop out shortly afterwards, and a portion will go out of business. In addition, respondent fatigue will cause others who are continuing units to drop out as well. This irregular process of establishments entering and exiting the sample makes it unwise to use a simple monthly attrition or enrollment metric to control for the sample size. By the end of the enrollment year for establishments new to the survey, the interviewers are required to have contacted all the respondents they were assigned. These new units will be merged into the estimates alongside the continuing units in the year following their enrollment.

The synthetic denominator takes advantage of the end of year cut off for contacting respondents to calculate an approximation of what the total number of respondents could be. All of the respondents that reported at least once successfully in either the enrollment sample or the continuing sample during the year prior to the year being forecast for net births and deaths are counted towards the synthetic denominator. This functions as a high water mark for what the maximum number of respondents could be for a given month, given that sample attrition is slowly reducing the number of respondents able and willing to report.

3.2 Establishment Predator-Prey Model

The final model is:

$$[6]\ NBD = \beta_0 + \beta_1(\ln First) + \beta_2\left(\frac{First}{Synthetic\ Denominator}\right)$$

The model in equation [6] performed the best under both back casting and forecasting. The addition of the synthetic denominator improved the fit. Specifications that followed the predator prey format strictly were neither the best performing nor the simplest. The reason for why the above model improves on the basic predator and prey specification is because the population of establishments, while competitive and full of births and deaths, does not have two separated groups where one behaves strictly as prey and the other strictly as a predator. All establishments can behave like predators, prey, both, or neither at any time.

The original function for z, the total population, can be transformed into equation [6] by replacing x and y with a single population variable E, representing all sample predator or prey establishments, and setting z to represent the total population of establishments in existence that are predators or prey, constituting some sub-group of the original total population z. Both activities contain elements that promote births and deaths and combining them allows estimating their net effect. The size of the population of establishments E that are potentially predators or prey in the sample is represented by the count reporting at first closing divided by the synthetic denominator, $E = \frac{First}{Synthetic\ Denominator}$.

$$[7.1]\ z = b\ln y_{(t)} - py_{(t)} - rx_{(t)} + d\ln x_{(t)} + C$$

$$[7.2]\ z = b\ln E_{(t)} - pE_{(t)} - rE_{(t)} + d\ln E_{(t)} + C$$

$$[7.3]\ NBD \sim z = (d+b)\ln E_{(t)} - (\mathrm{p}+\mathrm{r})\mathrm{E}_{(\mathrm{t})} + \mathrm{C}$$

Now this definition of net birth/death is cast into the form of a linear regression so we can estimate the parameters d,b,p, and r.

$(d+b)=\beta_1$ is the net population change from naturally occurring births and deaths. $(p+r)=\beta_2$ is the net population change due to establishment predation and avoidance of predation.

β_0 represents a dummy variable for each month of the year used to control for persistent calendar effects.

$$[7.4]\ NBD = \beta_0 \ + \beta_1 \ln E - \ \beta_2 E$$

$$[7.5]\ NBD = \beta_0 \ + \beta_1 \ln\left(\frac{First}{Synthetic\ Denominator}\right) - \ \beta_2\left(\frac{First}{Synthetic\ Denominator}\right)$$

In the β_1 term the synthetic denominator is dropped because it merges into the intercept term β_0 when the logarithm is changed to:

$$[8]\ \beta_1 \ln\left(\frac{First}{Synthetic\ Denominator}\right) = \beta_1(\ln First - \ln Synthetic\ Denominator)$$

The synthetic denominator is not a constant, but it only changes once per year. Approximating $-\beta_1 \ln Synthetic\ Denominator$ as part of the intercept simplifies the regression and improves the forecast. The synthetic denominator is also a strongly backwards looking variable that directly measures what happened the year before and only indirectly explains the current month being estimated. This means that the synthetic denominator may serve as a good approximation of the sample size for the β_2 term which estimates exponential effects, but is too coarse for use in the β_1 term that estimates linear effects.

$$[9]\ NBD = \beta_0 + \beta_1 \ln First - \ \beta_2\left(\frac{First}{Synthetic\ Denominator}\right)$$

This is the Establishment Predator-Prey Model, a predatory prey situation where the x and y populations are combined into one population that can perform either role. The model functions as a linear approximation of the conserved population value of a Predator-Prey system in long term equilibrium. Furthermore, because it is a linear approximation it is important that it be recalculated as frequently as the data can support using only data that is most relevant to the time period we are approximating.

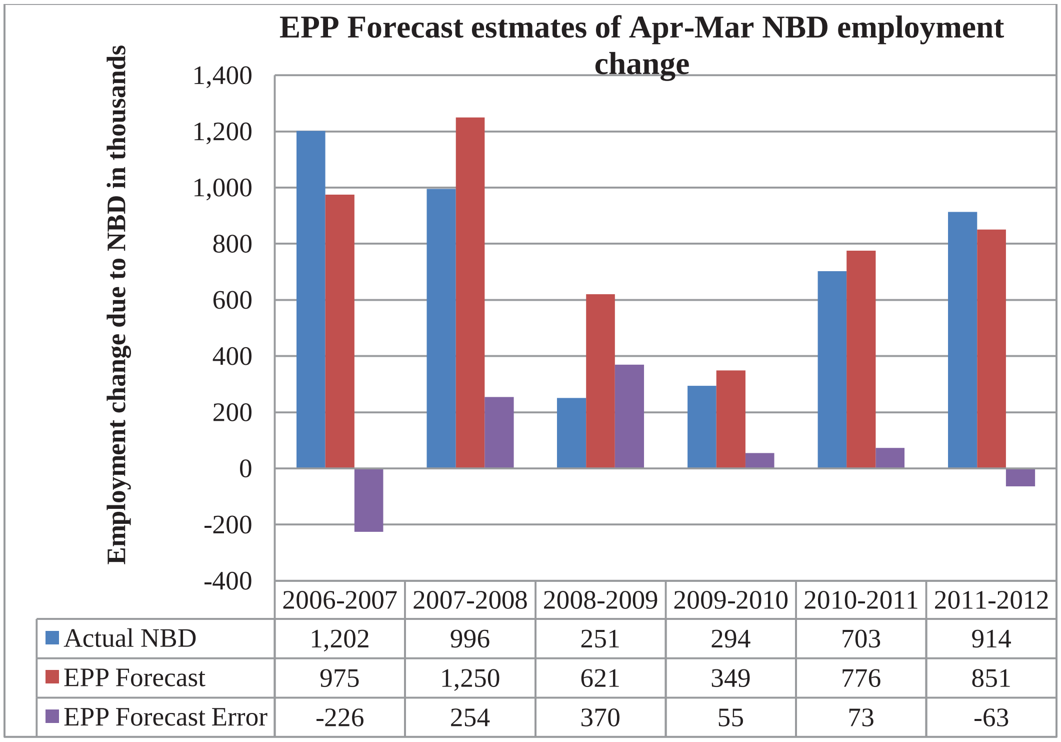
4. Results of Forecasting an Establishment Predator-Prey (EPP) Model

4.1 Forecast Results

To test whether or not the model serves as an effective representation of the employment changes from net birth/death, it must be able to accurately estimate the value of NBD nine months ahead of when the actual values will be known from the QCEW. Each forecast used eight quarters to calculate regression coefficients that were then applied with a nine month gap to predict one quarter of NBD values. For example: 2004 q1-q4 and 2005 q1-q4 were used to estimate a forecast of 2006 q4; and 2004 q2-q4 , 2005 q1-q4, and 2006 q1 were used to estimate a forecast of 2007 q1.

This quarterly rotation process matches the quarterly releases of updates to the QCEW, and ensures that exactly twenty-four months of training data is used to estimate a forecast, adding more quarters almost always made for a worse forecast. The regression is a linear approximation of the EPP function, and longer time frames will introduce information to the regression that will no longer be well approximated by a linear relationship, and as a result, harm the forecast. As a result, the forecast using the EPP will be a series of linear approximations of the EPP model at a give time frame that is recalculated with each new quarter, a series of forecasted slices rather than a time series.

Figure 1: Result of forecasting the EPP Model March 2006 – April 2012



	2006-2007	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012
🟦 Actual NBD	1,202	996	251	294	703	914
🟥 EPP Forecast	975	1,250	621	349	776	851
🟪 EPP Forecast Error	-226	254	370	55	73	-63

These forecasts were estimated controlling for *Industry* like in the earlier Distributed Closing Model but not for *Month* because of the inclusion of month dummy variables for the intercept term.

4.2 Two versus Three Years of Training Data

It is important to note that there is a weakness to taking a short period of training data to make forecasts. Reviewing the individual *Industry* forecasts reveals that the short training frame can sometimes lead to extreme prediction. The one case of this that was found in the above time period was for the construction industry in the fourth quarter of 2010.

Table 3: Comparison of forecasts with two versus three year training period for construction in quarter 4 2010

Year	Month	EPP Forecast Error Two Year Training (Thousands)	EPP Forecast Error Three Year Training (Thousands)
2010	October	77	-2
2010	November	91	11
2010	December	100	19

These extremely large errors, considering they are three months of only one industry, are caused by two unlikely events converging. The first is that according to the rotation schedule of quarters used in the two year training data, quarter one of 2008 through quarter four of 2009 are used to forecast quarter four of 2010. This was the time period of the recession and there were particularly dramatic movements in the NBD during this frame. Additionally, there were anomalies in the processing and collection of data that affected the values of the variables used to estimate NDB values. Nothing was done to alter or correct these estimates in the results for the EPP forecasts because recessions and anomalies in the collection process are factors that will come up from time to time indefinitely and must be accounted for when evaluating the effectiveness of a forecast.

Included in the table is a column for what the forecast error would have looked like for construction for this time frame if a 3 year instead of 2 year training data set were used. The additional data points dramatically reduce the error; however this comes at the expense of losing some of the responsiveness of the regression by adding more data to the training set, particularly at a turning point.

Table 4: Comparison of forecasts with two versus three year training period for all industries

	EPP 2 year Forecast Error	EPP 3 year Forecast Error
2007-2008	254	172
2008-2009	370	711
2009-2010	55	116
2010-2011	73	-57
2011-2012	-63	-38
Average Error	163	219
Variance of the error	38,239	78,021

Note: 2006-2007 values not included because of insufficient data to estimate the EEP 3 Year Forecast in that time period.

The EPP Three Year Forecast has a higher average error, 219 versus 163, and higher variance, 78,021 versus 23,795, for April 2007 to March 2012 time period. Largely the EPP Three Year Forecast deteriorated because of having twice as much error forecasting the recession months from April 2008 to March 2009. The addition of an extra year of data to the training set does smooth out the surge in the error in construction for the fourth quarter of 2010, but it results in forecasts that are less accurate in all time periods, especially the turning point at 2008-2009. As a result, the EPP Forecast with two years of training data was chosen to be the final specification for forecasting the net birth/death.

4.3 Comparisons with the current forecast of net birth/death

In order to understand the relative strength or weakness of the EPP Model it is important to compare its performance with the existing X12-ARIMA based forecast.

Table 5: Cumulative Forecast Error relative to historical values predicted by the current X12-ARIMA (in thousands)[7]

	2006-07	**2007-08**	**2008-09**	**2009-10**	**2010-11**	**2011-12**	**Abs. Avg.**
Establishment Predator-Prey	-226	254	370	55	73	-63	174
Current Forecast	-145	-184	489	129	-234	-374	259
Error Reduction	-55.9%	-38.0%	24.3%	57.4%	68.8%	83.2%	32.8%

Note: Negative error reduction values signify there was an increase in the error

The Establishment Predator-Prey Model, during this time period, improves on the Current Forecast, reducing the average error from 259 to 174. It is important point out that the improvements by the EPP Model over the Current Forecast are not uniform. In earlier time periods the EPP Model performs worse than the Current Forecast. However, there is a steady pace of improvement across the entire 2006-2012 time frame with the EPP Model reducing the Error compared to the Current Forecast by 83.2% by the 2011-12 period. This period of improvement coincides with substantial efforts to improve the data collection process that began in 2008 and have continued every year since. The substantial impact of these improvements can be seen in the increase in the Collection Rate over the period following 2008. It is the improvements to the quality of the micro data that make it possible for the EPP Model to converge on a more accurate solution than the Current Forecast.

[7] For historical birth/death residual values , see "Historical Net Birth/Death Adjustments," http://www.bls.gov/web/empsit/cesbdhst.htm

Table 6: Average Collection Rate at First Closing per year 2003-2013[8]

2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013 as of June
65.0%	68.9%	64.9%	66.9%	65.9%	68.6%	73.3%	70.4%	71.1%	72.5%	77.2%

5. Conclusion: The Foundational Importance of Data Collection

The Establishment Predator Prey model is not only a functional model for estimating the net birth/death but also a demonstration of the value of using paradata within its own structural model independent of the underlying data that is of interest. In this case, response rates across different years of a sample were arranged in a modified Predator-Prey function to successfully give a more accurate estimate of the birth/death residual.

The results during this time period are encouraging, but any conclusion made from it is constrained by extraordinary events that occur within the limited available time frame. The period from 2006-2012 contains the volatility from the recent recession (December 2007- June 2009) within its range, and this leaves unanswered the question as to whether or not the success of the EPP model is in part dependent on the recession/recovery behavior of the net birth/death. With the addition of more data, further research should demonstrate these promising results are applicable to a wider range of economic conditions.

Furthermore, there were simultaneous changes made to the collection process that altered the quality and composition of the micro data improving the paradata's precision and appear to have as a consequence substantially improved the ability of the EPP Model to generate accurate forecasts. Paradata is always generated from measures of the collection process itself and will always be dependent on the quality of that process.

Further research will need to focus on better understanding what about the data collection process leads to improvements, or degradations, of the paradata measures that were used in the EPP Model. Paradata measures are taken from some midpoint within the entire survey process, and are as a result sensitive to not only final error, but also intermediate errors that may be offset by the final output but present in stages leading up. A systematic accounting of the total survey error would help to define the dependencies and variability of paradata measures and thus help extend the foundations of the Establishment Predator-Prey Model.

[8] For CES collection rate values, see "CES Registry Receipts by Release," http://www.bls.gov/web/empsit/cesregrec.htm

Acknowledgements

I want to thank Greg Erkens and Michael Sverchov for their feedback and review of this paper. I also want to recognize those who have helped and supported my work throughout the development of this paper and its ideas: Victoria Battista, Edmond Cheng, Louis Harrell, Julie Hatch Maxfield, and Kenneth Robertson.

Any opinions expressed in this paper are those of the author and do not constitute policy of the Bureau of Labor Statistics.

References

Frank Hoppensteadt (2006) Predator-Prey model. Scholarpedia, 1(10):1563.

http://onlinemathcircle.com/wp-content/uploads/2012/03/Lotka-Volterra-Equations.pdf

“Technical notes to establishment survey data” http://www.bls.gov/ces/#technical.

“Historical Net Birth/Death Adjustments,” http://www.bls.gov/web/empsit/cesbdhst.htm

“CES Registry Receipts by Release,” http://www.bls.gov/web/empsit/cesregrec.htm