Predicting the Effect of Business Births and Deaths on the Current Employment Statistics Survey: Using Sample Information to Minimize Coverage Error

Chris Grieves, Steve Mance, and Collin Witt[1]
U.S. Bureau of Labor Statistics, 2 Massachusetts Avenue NE, Washington, DC 20212

Abstract

The Current Employment Statistics (CES) program uses a model to account for the bias in monthly payroll employment estimates arising from establishment births and deaths that fall outside the survey frame. Actual birth-death values derived from administrative counts are available with a substantial lag but must be predicted for the current month's estimation. The existing CES net birth-death model relies on the bias from business births and deaths to follow a consistent, seasonal pattern, characterized by an ARIMA process. This has broken down during extreme changes in the labor market necessitating interventions in the model during the COVID-19 recession and recovery. Previous research showed the ARIMA models can be improved by including covariates available coincident with the survey. This paper explores several modelling frameworks that use information from the CES survey to predict birth-death values, substantially outperforming previously examined models over the period from 2007–2020, covering the Great Recession through the pandemic. Forecast combination techniques are also examined and compared with predictions from the individual frameworks.

Key Words: Nowcasting, birth-death model, forecast combination

1. Introduction

The Current Employment Statistics (CES) program at the U.S. Bureau of Labor Statistics (BLS) produces nonfarm payroll employment, hours, and earnings data each month at the national level, for all 50 states, the District of Columbia, Puerto Rico, and about 450 metropolitan areas and divisions. National level data are typically released on the first Friday following the end of the reference month, representing one of the timeliest Principal Federal Economic Indicators, and are closely watched by policy makers, markets, and others as a coincident measure of the U.S. economy.

CES data are produced using a monthly establishment survey but include an adjustment—the net birth-death forecast—due to the survey's inability to capture new businesses ("births") and difficulty capturing closures ("deaths") in real time. BLS can calculate the "actual birth-death value"—the adjustment needed to offset this coverage error—but only at a lag of 10-12 months behind the current reference period. BLS uses these actual birth-death values as input to a seasonal ARIMA time series model forward to the current month. This captures the seasonal pattern in birth-death values but does a poor job reflecting business cycle variation since it does not include any real-time information.

Using concurrent information to predict the actual birth-death values can be considered a "nowcasting" problem. Previous work by Battista (2013) showed that covariates derived from the CES survey or other

[1] Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics.

timely economic indicators added to the ARIMA models would have improved predictive accuracy during the Great Recession (2007-09). CES used a modified version of this method during the steep downturn of the COVID-19 pandemic, but found its impact to be insufficient, and made temporary adjustments to the survey estimator to capture the effect of short-term business closures and re-openings.

We find that predictions of the actual birth-death values can be substantially improved by using information in the CES survey without making changes to the survey estimator. We simulate 44 individual prediction approaches over the period 2007-2020, many of which give good out-of-sample results, and which typically encompass the true birth-death values. This holds over all phases of the business cycle including the Great Recession, the ensuing recovery/expansion, and the extreme events of the pandemic. We would not have known the best *ex post* approach *ex-ante* and all likely suffer from some degree of model misspecification. Therefore, we consider forecast combination and find relatively simple techniques to work well.

2. Coverage Error from Business Births and Deaths

The monthly CES survey contains about 122,000 businesses and government agencies representing approximately 666,000 establishments reporting positive employment in the current and prior month. CES cannot capture births since they are not on the survey frame, the BLS Longitudinal Database (LDB), which is derived from the universe of Unemployment Insurance (UI) tax records, available at a lag of several months. Deaths are likewise difficult to capture since permanent closures generally result in nonresponse, and it is not practicable to determine the status of all nonrespondents during the short timeline needed for production. Some closures are reported—often temporary ones, or closures of establishments that were part of a multiple worksite employer.

The CES Net Birth-Death model was introduced as part of a probability-based redesign of the CES survey undertaken in the 1990s (Mueller 2006), replacing a bias-adjustment model that accounted for births, deaths, and other sources of bias in the previous quota sample. Nonrespondents and those businesses that report a closure are implicitly imputed with the sample weighted link relative (equation 1), which CES uses to estimate relative employment growth among continuing businesses, calculated within each estimating cell (detailed industry[2]) at time ***t***:

Equation 1: Weighted Link Relative

$$WLR_t = \frac{\sum_{i=1}^{n} w_i * ae_{i,t}}{\sum_{i=1}^{n} w_i * ae_{i,t-1}}$$

Where **ae** represents the number of all employees at establishment ***i*** in all ***n*** responding establishments with $ae_t > 0$ and $ae_{t-1} > 0$, and ***w*** is the survey weight associated with each reporter.

This eliminates the need to determine the employment status of nonrespondents and accounts for most business births.

[2] Basic estimating cells comprise an industry or set of industries defined by the North American Industry Classification (NAICS) code, often at the 6-digit NAICS level of detail, sometimes broken out by Census region.

The historical bias from applying this procedure (i.e., the amount it under- or over-accounts for births) can be determined using data from the LDB: the survey estimator is applied to the population from an initialization period t=0, representing a March benchmark month, through t=24. The population version of the link relative (also calculated at the estimating cell level) is shown in equation 2:

Equation 2:Population Link Relative

$$LR_t^{POP} = \frac{\sum_{i=1}^{N} ae_{i,t}}{\sum_{i=1}^{N} ae_{i,t-1}}$$

For all $\boldsymbol{N}$ population establishments with $\boldsymbol{ae_t} > 0$, $\boldsymbol{ae_{t-1}} > 0$, and $\boldsymbol{ae_{t=0}} > 0$. The establishment must have positive employment in the initialization month of the frame, t=0, which removes any births occurring since the starting period. No weights are applied since all population establishments with positive employment in the current, previous, and initialization months are included.

Next, the Continuous Plus Imputed (CIMP) series from t=1 to t=24 is derived by applying the population link relative to population employment level from t=0:

Equation 3: Continuous Plus Imputed Series

$$CIMP_t = AE_{t=0}^{LDB} \prod_{i=1}^{t} LR_{t=i}^{POP}$$

The difference between LDB employment and the CIMP represents the cumulative birth-death error from the initialization month. The actual birth-death values are then calculated as the over-the-month change in this residual:

Equation 4: Birth-Death calculated from LDB

$$BD_t = (1 - B)(AE_t^{LDB} - CIMP_t)$$

Where $\boldsymbol{B}$ represents the backshift (or lag) operator.[3]

Two full years (24 months) of historical birth-death values are calculated from each initialization so that birth-death values can be derived for the appropriate age of the sample used in rotation in CES. (In a given quarter, the sample frame used for some industries is a year older than in other industries.) Several years of BD values calculated from months 1-12 following the initialization period are chained together to form a "Year 1" time series, while values from months 13-24 are chained to form a "Year 2" series. In practice, there is usually little difference between the Year 1 and Year 2 values; Year 2 tends to be slightly more positive than Year 1 in each calendar month/industry.

Historical values of BD_t must be forecast up to 12 months beyond their end date for use in estimation. Consider that BD_t can be characterized as following a seasonal ARIMA process z_t[4] with mean βX_t:

[3] $B^n \times X_t = X_{t-n}$

[4] Seasonal ARIMA models of the type examined in this paper are explained in the Reference Manual for X13-ARIMA-SEATS, Chapter 4: RegARIMA Modelling Capabilities, available at: https://www2.census.gov/software/x-13arima-seats/x-13-data/documentation/docx13as.pdf

Equation 5: Birth-Death as RegARIMA process

$$BD_t = \beta X_t + z_t$$

Where X_t is a matrix of covariates available at time t and β is a vector of coefficients.

Historically, CES has not included covariates in the model (i.e., X_t is null[5]), and has assumed a seasonal integrated moving average model for z_t:

Equation 6: Birth-Death as Seasonal IMA

$$BD_t = z_t = \Theta(B^{12})a_t/(1 - B^{12})$$

Where a_t are white noise.

Putting everything together, CES employment estimates are created by setting a benchmark population employment level and applying the weighted link relative plus birth-death adjustment for successive months:

Equation 7: CES All Employees Estimator

$$AE_t = AE_{t=0}^{POP} \prod_{i=1}^{t} WLR_{t=i} + BD_{t=i}$$

The problems addressed in this paper can be considered as determining what to include in X_t, estimating β, and forecasting z_t.[6] We assume that the historical values of BD_t derived from the LDB are measured precisely and that our task lies in their prediction. There are sources of nonsampling error in both QCEW and CES that challenge this assumption. However, the birth-death values applied to CES data provide well-centered benchmark revisions when compared to QCEW and improved prediction of the birth-death values could substantially reduce the size of the benchmark revisions.

2.1 Relationship between Birth-Death and CES Sample

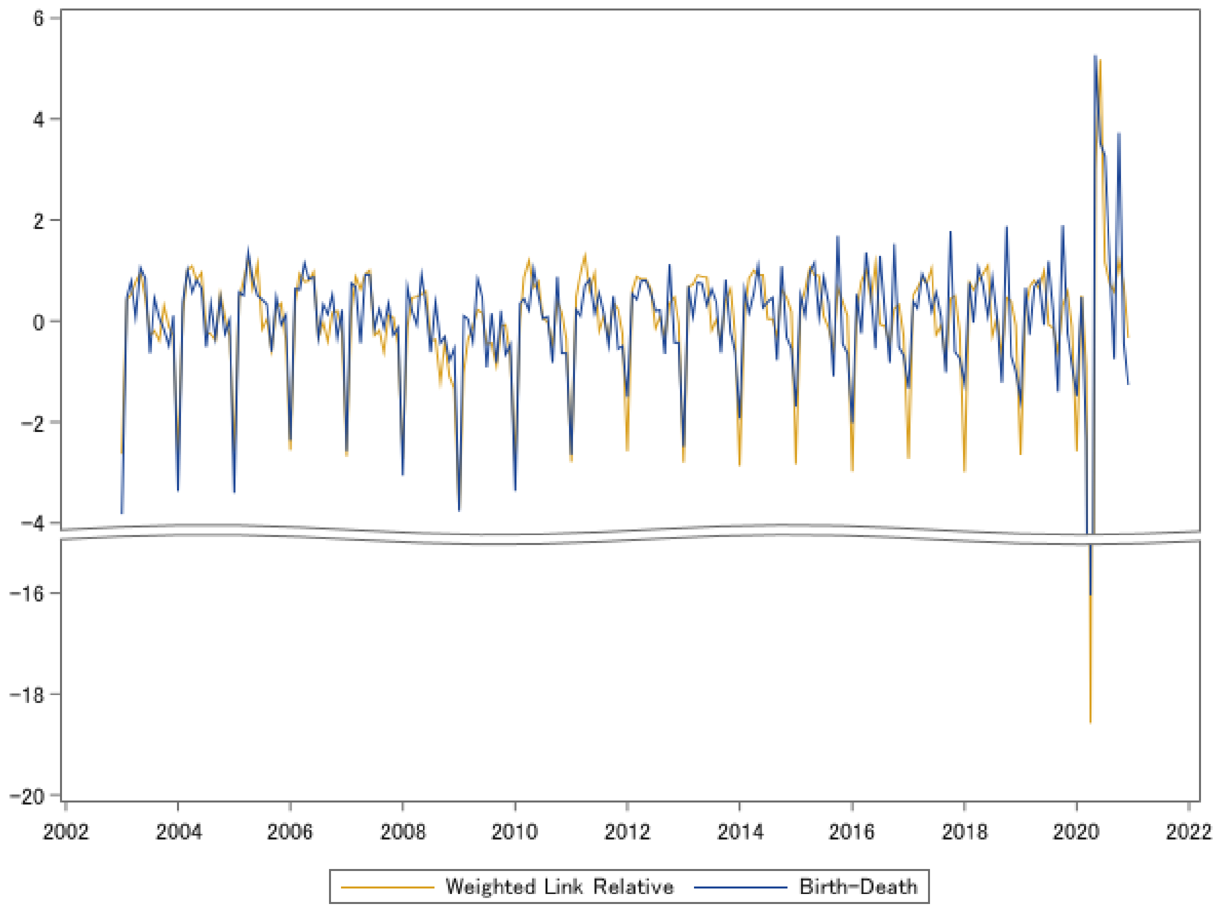
Both the actual birth-death values and the CES sample weighed link relative exhibit seasonal, cyclical, and irregular time series characteristics. Figure 1 shows both time series of Year 1 birth-death values and the weighted link relative at the total private level, standardized by the mean and standard deviation of each series prior to 2020. Both series have somewhat similar seasonal patterns, and both have a similar cyclical decline during the Great Recession (Dec. 2007 – Jun. 2009). The degree to which the sample link and birth-death share similar seasonality varies considerably by component industry. They are closely linked in construction and leisure and hospitality, where seasonality is primarily weather-driven and there are many small businesses but have little in common in industries such as education services, with many larger establishments, and seasonality arising from administrative effects

[5] The exception is when additive (point) outliers are included in the X matrix, but these are not known for the forecast period. Additive outliers are also described in the X-13ARIMA-SEATS chapter on RegARIMA.

[6] In some models we do allow for nonlinear relationships in which $BD_t = f(X_t) + z_t$.

(e.g., the school year). The extreme negative and positive values both series exhibited in 2020 were of comparable magnitude and timing.

Figure 1: Total Private Birth-Death (Year 1) and Weighted Link Relative - Standardized



At a very high level, both birth-death and the sample demonstrate similar cyclical patterns. Table 1 shows annual sums of actual birth-death Year 1 values and the average sample link relative at the total private level. Both showed similar downturns in 2008 and 2009 during the Great Recession, although the sample recovered earlier than the birth-death values in 2010-2011.

Table 1: Year 1 Birth-Death (annual total, thousands) and Weighted Link Relative (annual average)

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
B-D (thousands)	755	278	18	474	606	964	803	880	993	823	908	911	879	75
Weighted Link Relative	1.000	0.997	0.996	1.001	1.001	1.001	1.001	1.002	1.001	1.001	1.001	1.001	1.001	0.995

Note that the weighted link relative never accounts for more than a couple tenths of a percentage point of annual growth, while actual net birth-death can account for close to 1 million jobs, or about 7 tenths of a percent. (Total private annual employment ranged between about 115 million and 130 million over this period.) This aligns with Haltiwanger (2013), which demonstrated that most net job creation stemmed from growth in new and young businesses. Much of this growth happens in businesses too new for the CES sample to capture. However, information in the CES sample can help predict net job growth among new businesses and business closures.

3. Predicting Actual Birth-Death Values: Simulations

We simulated birth-death predictions for April 2007 – December 2020 for 3 baseline models, 44 individual forecasting models, and 6 types of forecast combinations. We accounted for rotation of Year 1/Year 2 values according to CES sample rotation.

We considered five categories of approaches for predicting birth-death values:

1. Baseline ARIMA/RegARIMA
2. Extended RegARIMA
3. Partially cross-sectional
4. Recurrent Neural Networks
5. Forecast combinations

Each will be broadly described in this section.

All models made use of a subset of shared set of covariates based on the CES sample. These included:

- The sample link relative (WLR_t). In some cases, we used the link relative at a higher level of industry detail to model at the detailed level. In many cases we took the natural log of the link relative, which often gave better diagnostics, but gave similar results in most cases.
- A measure of reported drops-to and returns-from zero employment in the CES sample.
- Seasonal dummy variables.
- Other information about the CES sample, such as the proportion of the sample below a certain employment size threshold.
- Detailed information about the CES sample including the amount of employment by establishment size increasing or decreasing within a certain range.

We produced our models at up to 3 different levels of industry aggregation, each of which we considered to be a separate model. CES needs basic industry (up to 6-digit NAICS level) detailed birth-death values in production. However, we also ran simulations at the 3-digit NAICS and super sector (generally 2-digit) level, which can serve as a control for the sum of basic level predictions. (This sort of reconciliation is currently done to produce birth-death forecasts at the state and metro area level.) For our results, we sum to the super sector or total private and evaluate at that level.

3.1 Baseline ARIMA/RegARIMA

We replicated the seasonal ARIMA models that CES has used since the introduction of the net birth-death models in 2003. We used automatic outlier selection in SAS PROC X13 (SAS 2018) instead of outliers selected during a manual review by BLS analysts, so the forecasts in our replication did not perfectly match those created in production, but the overall results were similar enough to use as a baseline.

We also replicated the approach from Battista (2013) and a variant that was used in production in 2020. This method adds the sample links for each super sector[7] to the RegARIMA model estimated at the basic level. When the model was applied in 2020, a considerable number of series exhibited negative coefficients for the relationship between the sample link and the net birth-death value—two series that presumably should move together rather than in opposition. In those cases, and others where modeling diagnostics were unacceptable, the seasonal difference was removed from the ARIMA part of the model and seasonal dummy variables were added. This usually resulted in predictions very similar to the baseline ARIMA, so we substituted with those forecast results under the same conditions. With all simulations, in the rare cases where models failed to converge, we substituted with the baseline ARIMA.

3.2 Extended RegARIMA

We considered it preferable to use established, easily interpretable models. To that end, we extended the use of RegARIMA models. We lengthened our input series as far as possible and used the sample link calculated for the same domain we modeled birth-death. Some variants included a measure of reported drops-to and returns-from zero as a covariate or added lags of the covariates.

3.3 Partially Cross-Sectional

Many of the methods we explored use the following 2-step approach. In the first step, we ignored the time series nature of the data and estimated a cross sectional regression relating birth-death to the covariates. The residuals from this first step were autocorrelated, so as a second step, we forecasted the residuals with a seasonal ARIMA model.

The primary motivation for this approach follows from the fact that (outside of the COVID-19 pandemic) most of the variation in birth-death values and the CES sample is seasonal, that the seasonality is likely cointegrated, and that this simple approach gives good out-of-sample results. Our intuition follows Barsky and Miron (1989), who found close links between seasonality and the business cycle. We think that in many industries, the same causes drive seasonal and business cycle fluctuations in both the sample and birth-death. Seasonal differencing or fitting seasonal dummies may result in model misspecification if the series are cointegrated. In future work, we would like to thoroughly investigate the possible seasonal cointegrating relationships.

The secondary motivation for this approach was that, in many detailed series, sampling error and other noise obscures the relationship between birth-death and the sample link. Pooling information across all series was easily handled in this approach but could also be done with multivariate time series models.

We tried many slightly different approaches within this partially cross-sectional framework. At its simplest, we estimated a univariate linear regression with the sample link. Variations included adding sets of seasonal dummy variables. We had some success fitting an initial regression, finding some structure in the residuals by calendar month, then constructing clustered seasonal dummies for the final stage 1 regression[8]. (E.g., we may form one intercept for Jan., Apr., and Jun., and another for the other 9 months.) This seemed to strike a balance between setting a full set of seasonal dummies and ignoring seasonality all together. We also had some success in the pooled approach constructing many covariates

[7] Super sectors are high-level industry combinations, mostly defined at the 2-digit NAICS level of detail, such as construction, retail trade, and leisure and hospitality.

[8] This was done by fitting conditional inference trees in the R package **party** (Hothorn et al. 2023).

out of the CES sample based on establishment size. The intuition there was that the amount of growth or loss among smaller worksites would be closely linked to births and deaths, but the relationship would attenuate among larger businesses.

3.4 Recurrent Neural Networks

Neural networks have been explored extensively recently in many areas including time series forecasting. One such python package created for this purpose is GlutonTS (Alexandrov 2019). We used the DeepAR (Deep Autoregressive) model from the GluonTS in our research. A DeepAR model is an autoregressive recurrent neural network that is trained on multiple series at once to learn a global structure that allows probabilistic forecasts (Salinas 2020). The model allows for the use of covariates and multi-horizon forecasts. We used the same covariates as the other methods in this model.

The motivation for this method is to experiment with newer machine learning methods for time series forecasting at BLS. We applied this method with the default settings with moderate success, however more experimentation with these models will be needed.

3.5 Forecast Combination

Combinations of economic forecasts have long been found to perform well compared with individual forecasts, dating back to at least Bates and Granger (1969), who proposed that information available to expert forecasters may not be available to the individual forecaster. Forecast combination can also serve to make predictions more robust to model misspecification bias—our models are at best rough approximations to the true data generating process—which serves as our motivation for combining forecasts to predict birth-death (Timmermann 2006).

We considered two different sets of candidate forecasts for combination:

- S_1: All forecasting models, excluding baseline approaches
- S_2: Partially cross-sectional forecasts without full seasonal dummies ($S_2 \subseteq S_1$)

The reason for the second set is that these approaches did not seem to suffer from the same biases as other approaches but individually could be noisy and flawed.

With each candidate set we investigated three forecast combination approaches for our birth-death simulations. The first forecast combination approach is a simple average of the set of candidate forecasts ($f_i \in S_j$):

Equation 7: Simple Average Forecast Combination

$$\widehat{BD}_t^j = N^{-1} \sum_i^N f_{i,t}$$

Very often simple averages empirically outperform forecast combinations that attempt to find optimal weights, which has become known as the "forecast combination puzzle" (Steel 2020).

The second approach is from Bates and Granger (1969), which weights based on the error variance of the individual forecasts:

Equation 8: Bates-Granger Forecast Combination

$$\widehat{BD}_t^j = \sum_i^N w_i^{BG} f_{i,t}\, , w_i^{BG} = \frac{\hat{\sigma}_i^{-2}}{\sum_{j=1}^N \hat{\sigma}_j^{-2}}$$

We estimated the variances based on prior out-of-sample forecast error. (Our first year of simulations used equal weights.)

The third approach uses a sophisticated subset averaging developed by Diebold and Shin (2019), who found the best approach in an empirical study was to incorporate principles of selection and regularization, discard most forecasts, and use equal weights for the rest. Their approach searches for the "best average" combination of N ≤ K candidate forecasts based on prior out-of-sample results. This can be very computationally expensive[9] and we set K=5.

4. Results

We evaluated two different out-of-sample error measures. The first (M1) was a straightforward root mean square error measure of the predicted birth-death values evaluated at the super sector level, based on an information set containing actual birth-death values 10-12 months behind the end of the nowcasting horizon, for each method (*j*). E.g., we evaluated Jan.-Mar. 2017 predictions with birth-death actuals through Mar. 2016 and covariates through Mar. 2017. This mirrors the information CES has when producing the initial monthly employment estimates. Results are presented by year as well as across all years since data users care about performance across the business cycle.

Equation 9: RMSE Monthly Birth-Death

$$M1_j = \sqrt{N_t^{-1} \sum_{t=1}^{N_t} \left(\widehat{BD}_{t,j} - BD_t\right)^2}$$

The second measure (M2) relates to the total error in our post-benchmark re-estimates. After setting a March benchmark level, CES updates birth-death values for the following 9 months in addition to producing initial forecasts for Jan.-Mar. of the following year. We sum the 12 months of predictions and actuals to calculate a root mean square error measure evaluated at the total private level. Years are designated by the year of the last month being predicted. (E.g., "2015" covers errors in the summed birth-death predictions over Apr. 2014 - Mar. 2015.)

Equation 10: RMSE 12-Mo. Post-Benchmark

$$M2_j = \sqrt{N_{year}^{-1} \sum_{year=1}^{N_{year}} \left(\sum_{m=1}^{12} \left(\widehat{BD}_{t,j} - BD_t\right)\right)^2}$$

Summing by year across all industries is done because errors around business cycle turning points have often been correlated across industries for several months in a row. For example, the production birth-death forecasts covering the worst of the Great Recession were too high broadly across industries over

---

[9] With 44 forecasts and K=5 this requires 1,086,008 combinations evaluated for each super sector, each quarter.

several months since few new businesses formed during a financial crisis while many closed; forecasts were then too low during the ensuing recover/expansion.

Due to the large number of forecasting models, we present the baseline models, the forecast combinations, and results for the five best, median, and worst-performing individual model (evaluated across all years). Results for all individual methods can be found in the appendix.

The version of Battista (2013) with modifications made to the RegARIMA models in 2020 was the best baseline approach, which we compared against the alternatives we investigated.

Under Metric 1 (Table 2), our best *ex post* individual forecast (N28)—which included zeros and clustered seasonal intercepts in the regression—cut the RMSE nearly in half compared to that baseline. Most of this gain was achieved in 2020. Nearly all individual methods outperformed the best baseline. The simple average and Bates-Granger forecast combinations using the subset of methods (S_2) performed nearly as well as the best *ex post* individual. These methods did not perform quite as well as the best *ex post* forecast in 2020 but tended to perform somewhat better in other years. The combinations performed well more consistently than any individual approach. Simple averaging and Bates-Granger averaging using the full set of individuals (S_1) performed only slightly worse than using the subset. The Diebold-Shin approach did not perform as well as the other combinations, driven by underperformance in 2020. With the extreme volatility that year, a strategy to discard most forecasts and average the rest did not work as well.

Table 2 - RMSE, Super Sector Monthly Birth-Death, Thousands (M1)

	Method	Overall	2008	'09	'10	'11	'12	'13	'14	'15	'16	'17	'18	'19	'20
Baselines	Production (ARIMA)	27.1	6.9	7.7	6.2	5.4	8.2	4.3	4.4	6.2	7.2	6	4.9	5	95.3
	RegARIMA - Battista 2013	23.2	5.7	6.7	5.8	5.8	8.2	4.3	4.5	6.3	7.3	5.8	5.1	4.9	81.1
	RegARIMA - Battista 2013 (modified)	20.1	5.3	6.4	5.8	5.7	8.2	4.3	4.5	5.9	7.3	5.9	5.1	4.8	69.6
Combinations	Simple Avg. (S1)	12.8	5.2	6.1	6.1	6.2	8	4.4	4.3	6.2	7.1	5.7	5.1	5.1	41.3
	Simple Avg. (S2)	10.9	4.7	6.1	6.6	6.8	8.1	4.6	4.4	6.2	7.1	5.7	5.4	5.1	33.2
	Bates-Granger (S1)	12.5	5.2	6.2	6.1	6.3	8	4.4	4.3	6.2	7.1	5.7	5.2	5.1	40.2
	Bates-Granger (S2)	10.9	4.7	6.2	6.7	7	8.1	4.7	4.4	6.2	7.2	5.8	5.4	5.1	33.4
	Diebold-Chen (S1)	15.4	5.2	6.1	6.4	6.5	8.3	4.6	4.4	6.4	7.1	5.7	5.1	5	51.6
	Diebold-Chen (S2)	13.4	4.7	6.2	6	6.3	8.3	4.8	4.5	6.3	7.1	5.7	5.2	5.1	43.8
Individual Forecasts	Rank 1 Individual (N28)	10.7	4.7	6.8	9.3	8.8	8.1	5.2	4.6	6.7	7.3	6.1	5.6	5.5	30.7
	Rank 2 Individual (N31)	11	5	7.4	9.9	9.7	8.2	5.9	4.8	7.3	7.2	6.2	5.6	5.5	31.2
	Rank 3 Individual (N08)	11.1	4.7	6.6	6.2	6.8	8.1	4.5	4.5	6.4	7.2	6.3	5.3	5.5	34.1
	Rank 4 Individual (N11)	11.4	4.7	6.8	6.2	6.9	8.1	4.5	4.5	6.4	7.3	6.2	5.4	5.5	35.2
	Rank 5 Individual (N09)	11.8	4.7	6.4	6.1	6.7	8	4.5	4.5	6.3	7.2	6.1	5.4	5.5	37
	Rank 22 (Median) Individual (B28)	16.2	5.8	6.6	9.2	7.9	8.4	5.3	4.6	7	7.6	5.7	5.7	5.3	53.5
	Rank 44 (Worst) Individual (RNN)	25.1	8.6	10.4	7.9	6.9	8.1	6.8	5.6	7.5	7.8	7.7	7	8.9	86.4

Table 3: RMSE, Total Private Post-Benchmark 12-Month Summed Birth-Death, Thousands (M2)

	Method	Overall	2008	'09	'10	'11	'12	'13	'14	'15	'16	'17	'18	'19	'20
Baselines	Production (ARIMA)	293	179	723	345	157	433	251	222	108	115	149	38	65	257
	RegARIMA - Battista 2013	225	104	486	355	150	297	176	203	72	101	168	55	54	225
	RegARIMA - Battista 2013 (modified)	214	76	416	373	136	301	187	209	62	101	180	63	56	197
Combination	Simple Avg. (S1)	201	142	398	348	129	252	123	171	67	95	188	58	73	206
	Simple Avg. (S2)	175	106	238	354	210	204	103	151	49	82	191	65	83	176
	Bates-Granger (S1)	199	142	395	342	131	250	124	170	66	98	189	59	73	201
	Bates-Granger (S2)	174	106	222	347	222	200	105	151	48	85	192	65	83	171
	Diebold-Shin (S1)	179	142	226	339	114	259	130	165	68	90	185	39	73	233
	Diebold-Shin (S2)	167	106	175	327	123	239	150	150	55	75	176	47	77	220
Individual Forecasts	Rank 1 Individual (B60)	156	125	105	19	38	278	229	97	61	138	25	83	243	246
	Rank 2 Individual (S09)	161	121	155	309	213	206	88	150	39	101	185	55	105	161
	Rank 3 Individual (S08)	167	98	43	347	291	182	87	141	29	101	189	61	102	152
	Rank 4 Individual (S15)	169	142	167	328	117	244	106	135	22	34	113	9	160	267
	Rank 5 Individual (N11)	172	44	97	415	227	179	58	135	44	103	230	80	46	145
	Rank 22 (Median) Individual (B28)	219	50	479	389	183	193	142	166	105	124	240	125	22	146
	Rank 44 (Worst) Individual (B27)	279	201	685	389	70	326	214	209	131	164	209	92	29	235

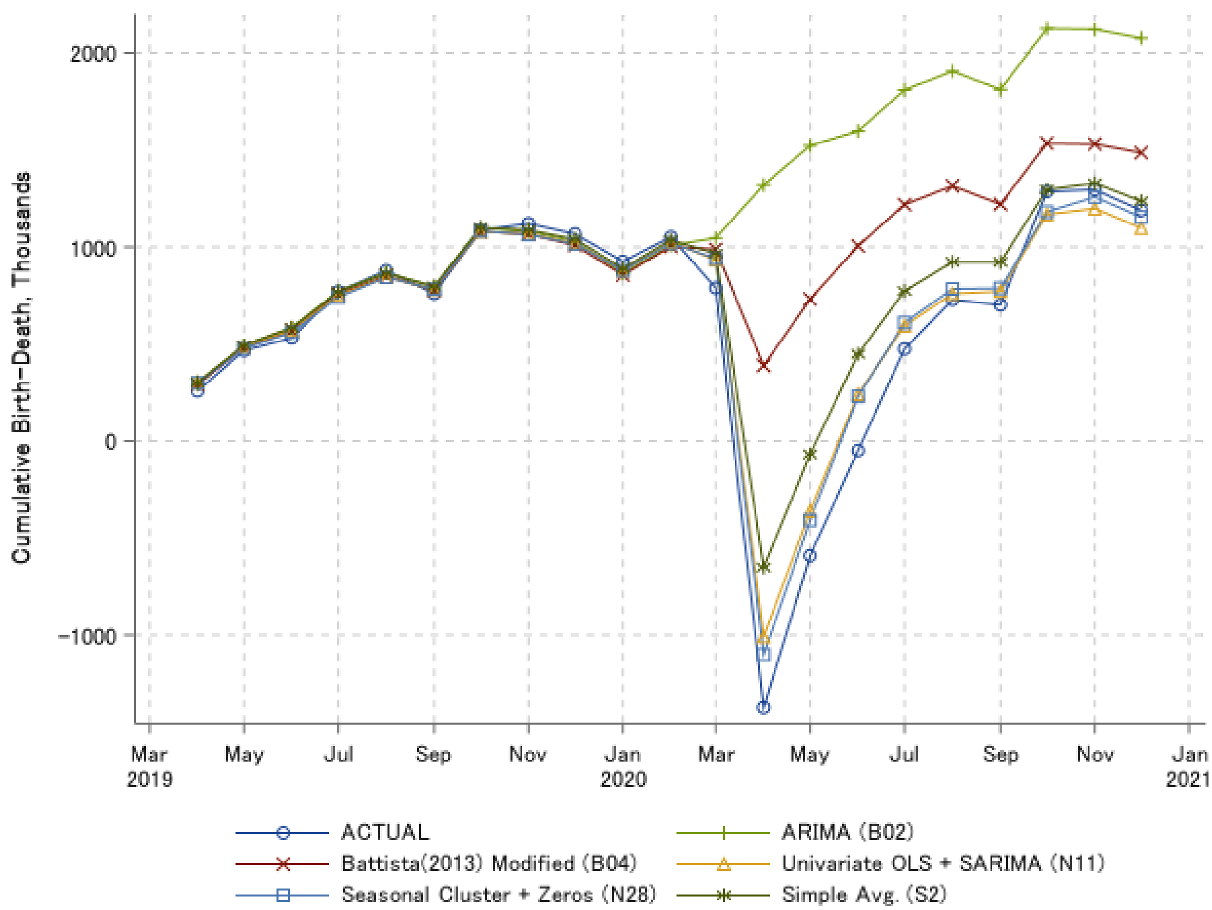
Under Metric 2 (Table 3), the best individual methods outperformed the best baseline by over 25 percent, although the best baseline slightly outperformed the median individual method. The forecast combinations using the subset (S_2) once again performed comparably to the best *ex post* individual. In this case, the Diebold-Shin averaging performed best. However, due to the way we set up the simulations, M2 does not capture much of the pandemic period when Diebold-Shin performed poorly.

If we shifted the window for M2 to sum the 12 months following each December, instead of following the March benchmark, we would better capture the volatility of 2020. In that alternative construction, Diebold-Shin performs comparably to the other combination techniques overall since it does worse in 2020. Overall, our individual methods and forecast combinations tend to show more gain with this alternate window. We must note that, had we evaluated the M2 at the super sector level instead of total private, we would not have seen a substantial gain against the baseline. Both these alternative metrics are reported in the appendix.

Among the individual methods, one of the very simplest partially cross-sectional approaches (designated N11) ranks among the top 5 in both metrics. Its performance is comparable to the better forecast combinations. The most complicated partially cross-sectional approach (B60) performs best on the second metric and has the practical benefit of being modeled at the basic level. The various RegARIMA approaches we tried—adding additional covariates and lags—generally performed worse than the partially cross-sectional models but slightly better than the baselines, at least on Metric 1. The DeepAR approach fared comparably to the baselines. We only tried one approach in this class of models, without contemporaneous information, and we think that it may have more success under different formulations, or as the second stage in a partially cross-sectional approach.

Figure 2 displays cumulative birth-death actual values and predicted values for 5 selected approaches at the Total Private level from April 2019 – December 2020. The seasonal ARIMA model does not incorporate any concurrent information and predicts birth-death values consistent with historical trends—mostly positive. The modified version of RegARIMA models from Battista (2013) improves considerably in capturing some of the downturn in Mar.-Apr. 2020. The simple average of S_2 provides substantially more gain, and the very simple partially-cross-sectional model (N11) and best *ex post* under M1 (N28) predictions track even closer to the actual birth-death values.

Figure 2: Selected Birth-Death Predictions and Actual Values, Total Private



5. Conclusion and future directions

In this paper, we showed that BLS can improve predictions of the actual birth-death values substantially beyond what has been done before using only the CES sample. The largest improvements were surrounding the COVID-19 pandemic recession, when BLS needed to make additional methodological changes to the estimator to reasonably capture business shutdowns and re-openings, but there were also substantial improvements surrounding the Great Recession and the ensuing recovery.

We found that how a model handles seasonality to be critical to its abilities to relate birth-death to the CES sample. Simple models that ignored seasonality and the time series nature of the data in a first step fared quite well. Seasonal differencing and the use of seasonal dummies greatly attenuated the relationship between birth-death and the sample resulting in worse predictions during turning points.

Forecast combination provided good results, although considering the set of which forecasts to combine mattered. Some individual methods performed about as well across metrics as the combinations, although less consistently, and we intend to better understand why. Regardless we think that forecast combination will be useful to make predictions more robust to misspecification bias.

Future work should push to improve the overall birth-death error along the metrics outlined in this paper, but also address additional production and data user needs. We think that it is important to test the possible cointegrating relationship between the CES sample and birth-death and to investigate appropriate models that take this relationship into account. Recurrent Neural Networks remain a promising part of the effort to improve birth-death prediction and should be pursued further. The impact of prediction methods on various quality metrics (such as the basic level benchmark revisions) will need to be fully addressed. Finally, any change to birth-death prediction will need to consider its explainability to data users.

Acknowledgements

We wish to thank Victoria Battista and Chris Manning for useful comments.

6. References

Alexandrov, A., Benidis, K., Bohlke-Schneider, M., Flunkert, V., Gasthaus, J., Januschowski, T., ... & Wang, Y. (2019). Gluonts: Probabilistic time series models in python. *arXiv preprint arXiv:1906.05264*.

Bates, J.M. and C.W. Granger (1969), "The Combination of Forecasts," *OR*, 20(4), https://doi.org/10.2307/3008764.

Barsky, Robert B. and Jeffrey A. Miron (1989), "The Seasonal Cycle and the Business Cycle," *Journal of Political Economy*, 97(3), https://doi.org/10.1086/261614

Battista, Victoria (2013), "Back to the Future: Using Current Regression Variables to Forecast Forward from Historical Net Birth/Death Employment," *JSM Proceedings*, American Statistical Association, https://www.bls.gov/osmr/research-papers/2013/pdf/st130140.pdf.

Bureau of Labor Statistics (2023). "Current Employment Statistics – National", *BLS Handbook of Methods*, https://www.bls.gov/opub/hom/ces/, last accessed September 26, 2023.

Diebold, F.X. and M. Shin (2019), "Machine learning for regularized survey forecast combination: Partially-egalitarian LASSO and its derivatives", *International Journal of Forecasting*, 35 (4), https://doi.org/10.1016/j.ijforecast.2018.09.006.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda (2013), "Who Creates Jobs? Small versus Large versus Young", *The Review of Economics and Statistics*, 95 (2), https://doi.org/10.1162/REST_a_00288

Hothorn, Torsten, Kurt Hornik, Carolin Strobl, and Achim Zeilis (2023), "party: A Laboratory for Recursive Partytioning", R package version 1.3-13. https://cran.r-project.org/package=party.

Mueller, Kirk (2006), "Impact of business births and deaths in the payroll survey", *Monthly Labor Review*. https://www.bls.gov/opub/mlr/2006/05/art4full.pdf.

Timmermann, Allan (2006), “Chapter 4: Forecast Combinations,” *Handbook of Economic Forecasting*, Volume 1, pp. 135-195. Elsevier. https://doi.org/10.1016/S1574-0706(05)01004-9.

Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, *36*(3), 1181-1191.

SAS Institute (2018), “The X13 Procedure”, *SAS/ETS 15.1 User’s Guide*, SAS Institute, Cary, NC. https://support.sas.com/documentation/onlinedoc/ets/151/x13.pdf.

Steel, Mark F.J. (2020), “Model Averaging and Its Use in Economics”, *Journal of Economic Literature*, 58(3), https://doi.org/10.1257/jel.20191385.

U.S. Census Bureau (2023), *X-13ARIMA-SEATS Reference Manual*, Version 1.1, U.S. Census Bureau, Washington, DC. https://www2.census.gov/software/x-13arima-seats/x13as/unix-linux/documentation/docx13ashtml.pdf, last accessed September 25, 2023.

Appendix 1 – RMSE, Super Sector Monthly Birth-Death, Thousands M1)

Rank	Method	Overall	2008	‘09	‘10	‘11	‘12	‘13	‘14	‘15	‘16	‘17	‘18	‘19	‘20
1	N28 (Partially Cross-Sectional)	10.7	4.7	6.8	9.3	8.8	8.1	5.2	4.6	6.7	7.3	6.1	5.6	5.5	30.7
2	Simple Avg. (S2)	10.9	4.7	6.1	6.6	6.8	8.1	4.6	4.4	6.2	7.1	5.7	5.4	5.1	33.2
3	Bates-Granger (S2)	10.9	4.7	6.2	6.7	7	8.1	4.7	4.4	6.2	7.2	5.8	5.4	5.1	33.4
4	N31 (Partially Cross-Sectional)	11	5	7.4	9.9	9.7	8.2	5.9	4.8	7.3	7.2	6.2	5.6	5.5	31.2
5	N08 (Partially Cross-Sectional)	11.1	4.7	6.6	6.2	6.8	8.1	4.5	4.5	6.4	7.2	6.3	5.3	5.5	34.1
6	N11 (Partially Cross-Sectional)	11.4	4.7	6.8	6.2	6.9	8.1	4.5	4.5	6.4	7.3	6.2	5.4	5.5	35.2
7	N09 (Partially Cross-Sectional)	11.8	4.7	6.4	6.1	6.7	8	4.5	4.5	6.3	7.2	6.1	5.4	5.5	37
8	Bates-Granger (S1)	12.5	5.2	6.2	6.1	6.3	8	4.4	4.3	6.2	7.1	5.7	5.2	5.1	40.2
9	Simple Avg. (S1)	12.8	5.2	6.1	6.1	6.2	8	4.4	4.3	6.2	7.1	5.7	5.1	5.1	41.3
10	B08 (Partially Cross-Sectional)	13.1	5.4	6.8	6	5.8	8.5	5	4.5	6.6	7.6	5.7	5.7	5.4	41.9
11	B11 (Partially Cross-Sectional)	13.1	5.1	6.7	6	5.9	8.6	5.1	4.5	6.6	7.6	5.6	5.7	5.5	41.9
12	B60 (Partially Cross-Sectional)	13.4	6.9	6.9	7.5	9	11	7.7	7.3	8.6	10	9.3	9.5	7.7	38.2
13	Diebold-Shin (S2)	13.4	4.7	6.2	6	6.3	8.3	4.8	4.5	6.3	7.1	5.7	5.2	5.1	43.8
14	S35 (Partially Cross-Sectional)	13.6	5.6	6.1	6.6	6.2	8.1	4.9	4.8	6	7	6.1	5.2	5.1	44.2
15	S09 (Partially Cross-Sectional)	13.6	5	6.4	6.3	6.9	8	4.9	4.4	6.2	7.3	5.9	5.3	5.3	44.2
16	B09 (Partially Cross-Sectional)	13.6	5.7	6.9	6	5.8	8.6	5	4.4	6.6	7.5	5.7	5.6	5.3	44.1
17	S29 (Partially Cross-Sectional)	13.8	5.4	6.4	9.3	8.4	8.1	5.1	4.5	6.2	7.3	5.9	5.4	5.3	44.4
18	S08 (Partially Cross-Sectional)	13.9	5.6	7.2	7.1	7.3	8.1	4.9	4.5	6.2	7.4	5.9	5.4	5.3	44.9
19	S12 (Partially Cross-Sectional)	14	6.2	6.6	6.1	6.6	8	4.6	4.4	6.3	7.4	6	5	5.3	45.9
20	S28 (Partially Cross-Sectional)	14.1	6	7.4	10.2	8.7	8.3	5.6	4.6	6.3	7.3	5.8	5.5	5.3	44.7
21	S11 (Partially Cross-Sectional)	14.4	5.9	8.2	7.8	7.5	8.1	5.1	4.6	6.3	7.4	5.9	5.4	5.3	46.8
22	S15 (Partially Cross-Sectional)	15	5.7	6.2	6.6	6.1	8.1	4.8	4.8	6.1	7.2	6.2	5.1	5.2	49.6
23	B12 (Partially Cross-Sectional)	15	5.9	7.2	5.8	5.8	8.4	4.7	4.7	6.7	7.3	5.7	5.4	5.1	49.8
24	N12 (Partially Cross-Sectional)	15	5.4	6.5	5.9	6	8	4.2	4.4	6.4	7.3	6	5	5.4	50.1
25	Diebold-Shin (S1)	15.4	5.2	6.1	6.4	6.5	8.3	4.6	4.4	6.4	7.1	5.7	5.1	5	51.6
26	B29 (Partially Cross-Sectional)	15.6	6.1	6.8	9.1	7.2	8.4	5.4	4.6	6.7	7.6	5.7	5.6	5.4	51.3
27	S31 (Partially Cross-Sectional)	15.9	6.4	7.9	11.1	10.9	8.3	6.3	4.8	6.8	7.4	6	5.6	5.4	51.3
28	B28 (Partially Cross-Sectional)	16.2	5.8	6.6	9.2	7.9	8.4	5.3	4.6	7	7.6	5.7	5.7	5.3	53.5
29	S10 (Partially Cross-Sectional)	16.9	6.9	7	6.1	6.4	7.9	4.6	4.5	6.4	7.3	5.9	4.9	5.2	57.1
30	B10 (Partially Cross-Sectional)	17.4	7	7.6	5.8	5.8	8.6	4.8	4.6	6.7	7.4	5.6	5.4	5.1	58.8
31	N30 (Partially Cross-Sectional)	17.4	6.5	7	8.6	6.3	7.9	4.3	4.5	6.4	7.3	6.2	5	5.3	58.7
32	S27 (RegARIMA)	17.5	6.3	6.4	6.3	6.6	8	4.6	4.5	6.4	7.2	5.9	4.9	5.1	59.6
33	S30 (Partially Cross-Sectional)	17.6	6.8	7	7	6.5	8.1	4.6	4.6	6.5	7.3	5.9	5	5.3	59.5
34	S07 (RegARIMA)	17.8	6.4	6.5	5.9	6.3	7.9	4.6	4.5	6.3	7.2	5.9	4.9	5.1	73.5
35	N10 (Partially Cross-Sectional)	17.8	6.5	7	6.1	5.9	8.2	4.2	4.5	6.3	7.1	6.1	5	5.4	60.7
36	N35 (Partially Cross-Sectional)	17.9	5.8	6.3	6.6	6.3	8.3	4.5	4.7	6.2	6.9	6.1	5.3	5.2	61
37	B30 (Partially Cross-Sectional)	18	6.9	7.6	8.7	7	8.5	4.4	4.6	6.7	7.4	5.7	5.4	5.2	60.7
38	B31 (Partially Cross-Sectional)	18.2	5.6	6.5	9.5	8.8	8.4	5.8	4.7	7.6	7.6	5.8	5.8	5.4	60.9
39	B26 (RegARIMA)	18.8	7.8	7.9	6.8	6.9	8.7	4.4	4.3	6.6	7.3	5.7	5.6	5.1	63.8
40	N27 (RegARIMA)	18.9	6.2	6.7	9.4	6.3	8	4.5	4.5	6.3	7.2	6.1	5	5.3	64.5

41	S26 (RegARIMA)	19.2	6.8	6.3	6.2	6.1	8.3	4.6	4.5	6.5	7	6.1	5	5.2	65.8
42	N07 (RegARIMA)	19.4	6.3	6.9	5.9	5.7	8.1	4.3	4.4	6.3	7.2	6	5	5.3	81.1
43	S06 (RegARIMA)	19.5	6.2	6.5	5.6	6.2	7.9	4.7	4.5	6.5	7	6.1	4.9	5.1	67.1
44	N26 (RegARIMA)	19.6	6.7	6.7	9.5	6	8.1	4.5	4.6	6.2	7.2	6.3	4.9	5.3	67
45	N06 (RegARIMA)	19.6	6.7	6.9	5.8	5.7	8	4.4	4.5	6.3	7.2	6.1	4.9	5.4	67.5
46	B07 (RegARIMA)	19.7	6.5	7.6	6	5.6	8.6	4.7	4.5	6.7	7.3	5.8	5.5	5	67.7
47	B27 (RegARIMA)	19.8	7.5	7.7	6.7	6.5	8.6	4.5	4.4	6.6	7.2	5.7	5.6	5	67.7
48	RegARIMA - Battista 2013 (modified)	20.1	5.3	6.4	5.8	5.7	8.2	4.3	4.5	5.9	7.3	5.9	5.1	4.8	69.6
49	B06 (RegARIMA)	20.3	6.7	7.6	6	5.6	8.5	4.7	4.6	6.7	7.2	6	5.4	5.1	69.9
50	N15 (Partially Cross-Sectional)	20.4	5.7	6.3	6.4	5.8	8.2	4.4	4.7	6.1	7	6.1	5.1	5.3	70.5
51	RegARIMA - Battista 2013	23.2	5.7	6.7	5.8	5.8	8.2	4.3	4.5	6.3	7.3	5.8	5.1	4.9	81.1
52	DeepAR (RNN)	25.1	8.6	10.4	7.9	6.9	8.1	6.8	5.6	7.5	7.8	7.7	7	8.9	86.4
53	Production (ARIMA)	27.1	6.9	7.7	6.2	5.4	8.2	4.3	4.4	6.2	7.2	6	4.9	5	95.3

Appendix 2 - RMSE, Total Private Post-Benchmark 12-Mo. Summed Birth-Death, Thousands (M2)

Rank	Method	Overall	2008	‘09	‘10	‘11	‘12	‘13	‘14	‘15	‘16	‘17	‘18	‘19	‘20
1	B60 (Partially Cross-Sectional)	156	125	105	19	38	278	229	97	61	138	25	83	243	246
2	S09 (Partially Cross-Sectional)	161	121	155	309	213	206	88	150	39	101	185	55	105	161
3	S08 (Partially Cross-Sectional)	167	98	43	347	291	182	87	141	29	101	189	61	102	152
4	Diebold-Shin (S2)	167	106	175	327	123	239	150	150	55	75	176	47	77	220
5	S15 (Partially Cross-Sectional)	169	142	167	328	117	244	106	135	22	34	113	9	160	267
6	N11 (Partially Cross-Sectional)	172	44	97	415	227	179	58	135	44	103	230	80	46	145
7	Bates-Granger (S2)	174	106	222	347	222	200	105	151	48	85	192	65	83	171
8	Simple Avg. (S2)	175	106	238	354	210	204	103	151	49	82	191	65	83	176
9	N08 (Partially Cross-Sectional)	176	80	181	409	181	203	85	147	47	99	221	72	45	164
10	S35 (Partially Cross-Sectional)	177	225	145	340	124	248	105	142	15	31	121	9	160	263
11	Diebold-Shin (S1)	179	142	226	339	114	259	130	165	68	90	185	39	73	233
12	S11 (Partially Cross-Sectional)	180	53	116	369	369	134	72	152	24	110	187	62	104	148
13	N09 (Partially Cross-Sectional)	182	82	248	403	178	211	97	155	57	96	207	80	49	166
14	S29 (Partially Cross-Sectional)	184	143	192	306	357	194	61	165	46	88	196	64	107	166
15	S28 (Partially Cross-Sectional)	186	105	87	337	411	178	43	144	43	89	213	66	104	154
16	N31 (Partially Cross-Sectional)	186	91	113	465	297	119	18	102	64	103	244	83	48	130
17	N28 (Partially Cross-Sectional)	187	126	214	426	261	152	39	148	71	94	224	72	52	151
18	Bates-Granger (S1)	199	142	395	342	131	250	124	170	66	98	189	59	73	201
19	N15 (Partially Cross-Sectional)	200	145	416	271	38	316	139	145	5	24	100	10	133	283
20	S12 (Partially Cross-Sectional)	200	178	421	298	76	271	109	174	57	87	173	19	105	218
21	Simple Avg. (S1)	201	142	398	348	129	252	123	171	67	95	188	58	73	206
22	S31 (Partially Cross-Sectional)	204	43	43	359	537	94	17	142	49	88	226	62	99	127
23	S06 (RegARIMA)	205	185	454	275	137	253	91	164	41	85	176	19	94	249
24	S26 (RegARIMA)	209	317	415	270	141	240	49	174	45	86	182	13	90	250
25	B31 (Partially Cross-Sectional)	211	44	434	393	222	160	138	151	102	122	240	133	14	124

26	B11 (Partially Cross-Sectional)	211	57	408	386	179	216	158	193	105	124	235	139	22	133
27	N35 (Partially Cross-Sectional)	213	261	462	281	5	282	135	149	14	17	108	4	133	280
28	RegARIMA - Battista 2013 (modified)	214	76	416	373	136	301	187	209	62	101	180	63	56	197
29	B28 (Partially Cross-Sectional)	219	50	479	389	183	193	142	166	105	124	240	125	22	146
30	N12 (Partially Cross-Sectional)	220	148	482	352	77	274	130	185	76	89	202	57	59	235
31	B08 (Partially Cross-Sectional)	223	76	459	388	158	239	173	198	109	134	239	134	26	146
32	S07 (RegARIMA)	223	181	521	325	28	304	111	170	60	87	166	20	98	237
33	RegARIMA - Battista 2013	225	104	486	355	150	297	176	203	72	101	168	55	54	225
34	S27 (RegARIMA)	226	244	494	326	69	324	107	167	57	86	166	17	98	239
35	N26 (RegARIMA)	227	261	542	316	43	252	72	188	60	79	164	45	60	250
36	B09 (Partially Cross-Sectional)	227	86	484	386	130	248	184	203	115	134	237	119	28	162
37	B29 (Partially Cross-Sectional)	229	92	516	396	151	227	149	165	109	124	229	121	21	169
38	N06 (RegARIMA)	231	177	566	305	6	299	120	192	52	81	167	48	69	252
39	S10 (Partially Cross-Sectional)	233	205	563	310	58	303	116	177	67	92	164	8	103	243
40	N07 (RegARIMA)	236	182	564	321	62	321	149	191	62	74	163	47	70	253
41	S30 (Partially Cross-Sectional)	240	228	557	324	134	333	115	172	63	92	162	10	101	243
42	N27 (RegARIMA)	241	197	575	343	1	333	117	191	72	72	158	42	68	251
43	N10 (Partially Cross-Sectional)	245	194	596	346	29	315	148	189	68	86	185	52	58	246
44	N30 (Partially Cross-Sectional)	249	208	589	358	93	325	129	194	80	89	185	49	60	252
45	DeepAR (RNN)	253	83	713	292	56	82	184	105	56	5	147	111	143	343
46	B12 (Partially Cross-Sectional)	253	97	581	406	56	301	201	209	135	159	217	96	29	206
47	B26 (RegARIMA)	263	172	670	352	59	304	162	200	103	153	201	80	19	228
48	B06 (RegARIMA)	265	150	654	352	53	311	211	212	143	157	215	77	21	230
49	B07 (RegARIMA)	275	129	681	367	31	343	224	203	156	164	209	92	29	235
50	B30 (Partially Cross-Sectional)	277	119	699	413	29	311	178	214	131	155	207	101	25	230
51	B10 (Partially Cross-Sectional)	277	129	694	397	15	316	212	218	158	155	207	92	31	220
52	B27 (RegARIMA)	279	201	685	389	70	326	214	209	131	164	209	92	29	235
53	Production (ARIMA)	293	179	723	345	157	433	251	222	108	115	149	38	65	257

Appendix 3 - RMSE, Total Private 12-Mo. Summed Birth-Death from December, Thousands (M2 - alternate window)

Rank	Method	Overall	2008	'09	'10	'11	'12	'13	'14	'15	'16	'17	'18	'19	'20
1	B60 (Partially Cross-Sectional)	147	93	13	15	41	459	52	45	1	124	83	146	98	56
2	S15 (Partially Cross-Sectional)	150	181	217	259	11	332	10	14	162	52	3	41	70	13
3	S35 (Partially Cross-Sectional)	154	164	227	263	13	336	8	24	162	57	9	44	71	100
4	N11 (Partially Cross-Sectional)	156	125	310	234	73	278	12	44	212	73	98	56	21	41
5	S09 (Partially Cross-Sectional)	156	178	271	222	56	317	22	46	229	38	52	23	9	16
6	S08 (Partially Cross-Sectional)	160	106	237	295	72	323	6	40	231	46	56	22	3	85
7	N08 (Partially Cross-Sectional)	165	179	346	197	45	306	27	52	213	64	95	55	17	45
8	Diebold-Shin (S2)	169	220	260	176	4	353	45	61	196	13	67	25	2	223
9	Bates-Granger (S2)	169	220	303	244	51	336	31	57	205	37	74	28	6	51
10	Simple Avg. (S2)	170	220	317	234	46	334	33	57	203	35	73	28	3	72
11	S11 (Partially Cross-Sectional)	171	14	168	382	129	303	8	32	236	53	54	20	5	179
12	S29 (Partially Cross-Sectional)	172	174	259	362	80	298	13	54	219	48	67	17	6	30
13	N09 (Partially Cross-Sectional)	173	225	360	188	44	307	36	63	217	51	92	56	15	97
14	S28 (Partially Cross-Sectional)	176	117	227	407	94	299	39	49	227	56	81	9	5	87
15	N28 (Partially Cross-Sectional)	181	220	368	282	86	282	20	68	226	73	90	55	31	4
16	N31 (Partially Cross-Sectional)	183	166	328	354	107	266	66	52	223	72	107	53	30	128
17	N35 (Partially Cross-Sectional)	195	389	342	69	57	379	34	44	155	49	22	28	74	195
18	N15 (Partially Cross-Sectional)	204	371	328	23	77	375	48	24	152	57	10	22	71	335
19	S31 (Partially Cross-Sectional)	206	24	155	535	193	275	66	37	228	74	90	9	22	240
20	S12 (Partially Cross-Sectional)	213	369	364	85	23	350	50	84	212	1	42	17	25	368
21	Bates-Granger (S1)	214	347	377	156	3	362	57	82	214	28	75	37	9	334
22	Simple Avg. (S1)	217	347	380	155	2	361	58	82	212	26	74	37	6	359
23	B11 (Partially Cross-Sectional)	221	322	452	147	35	372	107	103	223	80	125	89	53	245
24	Diebold-Shin (S1)	222	347	281	208	21	357	37	81	216	28	55	24	9	463
25	S26 (RegARIMA)	224	416	301	169	20	304	20	96	207	12	46	18	7	461
26	B31 (Partially Cross-Sectional)	225	314	448	265	63	364	89	95	230	72	125	88	63	225
27	S06 (RegARIMA)	227	433	299	141	8	335	29	84	208	2	43	17	11	454
28	B08 (Partially Cross-Sectional)	236	357	472	125	33	386	125	105	225	82	130	90	55	314
29	B28 (Partially Cross-Sectional)	237	344	469	217	34	374	103	100	233	78	130	84	50	311
30	B09 (Partially Cross-Sectional)	245	380	492	103	19	391	124	114	230	79	131	79	57	339
31	B29 (Partially Cross-Sectional)	248	386	503	193	5	384	103	111	229	71	124	86	46	327
32	S07 (RegARIMA)	253	446	403	64	51	365	46	85	215	5	41	16	16	524
33	S27 (RegARIMA)	254	416	382	111	80	362	39	81	218	6	43	15	15	558
34	N12 (Partially Cross-Sectional)	258	402	419	93	16	367	77	85	216	28	86	46	8	562
35	N06 (RegARIMA)	274	514	374	12	0	391	63	86	198	17	82	47	2	597
36	S30 (Partially Cross-Sectional)	274	458	401	180	84	364	52	88	212	2	33	14	19	616
37	S10 (Partially Cross-Sectional)	275	491	419	84	48	365	57	93	217	10	36	14	23	606
38	N26 (RegARIMA)	278	514	335	72	11	360	28	100	197	14	66	49	1	663
39	RegARIMA - Battista 2013	283	410	338	159	2	399	108	92	212	25	60	70	52	703
40	N07 (RegARIMA)	290	499	454	9	43	408	92	82	204	21	74	53	5	638

41	N27 (RegARIMA)	292	504	426	79	101	388	65	88	209	24	65	44	5	664
42	N10 (Partially Cross-Sectional)	299	515	449	65	42	395	93	87	211	7	79	45	5	681
43	N30 (Partially Cross-Sectional)	299	525	438	167	94	383	75	97	213	18	70	48	5	668
44	B12 (Partially Cross-Sectional)	303	433	558	64	33	428	140	158	262	70	107	77	49	607
45	DeepAR (RNN)	311	572	622	131	61	232	215	97	196	155	35	33	65	590
46	RegARIMA - Battista 2013 (modified)	323	328	360	147	16	408	113	93	213	40	66	60	44	924
47	B06 (RegARIMA)	337	516	558	25	24	435	132	158	240	56	112	72	45	765
48	B10 (Partially Cross-Sectional)	339	523	591	24	52	445	156	163	261	60	116	69	45	725
49	B30 (Partially Cross-Sectional)	339	517	580	100	96	430	139	148	243	66	113	75	31	752
50	B26 (RegARIMA)	343	581	525	14	79	404	120	128	233	45	97	62	50	801
51	B07 (RegARIMA)	345	535	578	16	55	458	144	142	249	57	116	72	40	769
52	B27 (RegARIMA)	351	608	579	36	89	432	136	143	244	54	95	74	41	765
53	Production (ARIMA)	384	558	587	200	177	492	142	135	226	13	72	61	7	924

Appendix 4 - RMSE, Super Sector Post-Benchmark 12-Mo. Summed Birth-Death, Thousands

Rank	Method	Overall	2008	‘09	‘10	‘11	‘12	‘13	‘14	‘15	‘16	‘17	‘18	‘19	‘20
1	Diebold-Shin (S2)	24.3	21.8	33.8	39.2	16.8	26.4	17.5	13.6	7.7	16.9	22.2	16.2	20.5	39.4
2	Simple Avg. (S2)	25	21.8	37.8	39.9	22.9	27.1	13.8	13.9	7.8	19	28.5	17	20.3	32.1
3	Diebold-Shin (S1)	25.3	22.2	38.6	39.9	15.7	27.9	16.3	14.9	8.7	18.4	23.4	15.8	20.3	40
4	Bates-Granger (S2)	25.4	21.8	40.5	39.6	23.9	27	13.9	13.7	7.8	19.5	28.7	17.3	20.6	32
5	S09 (Partially Cross-Sectional)	25.6	21.8	38.7	36.1	30.9	31.6	14.2	14.9	7.9	20.2	31.9	15.9	19.3	27.3
6	S15 (Partially Cross-Sectional)	25.8	22	37.7	38.6	23.3	30.8	16.3	13.6	8.1	14.9	21.5	15.4	23.4	42.5
7	Simple Avg. (S1)	26.1	22.2	46.4	38.4	17.6	28.4	14.3	15.4	8.6	19.6	27.1	16.8	19.8	36.5
8	Bates-Granger (S1)	26.3	22.2	47.7	38	17.3	28.4	14.3	15.3	8.5	19.8	27.2	17	20	36.2
9	RegARIMA - Battista 2013 (modified)	26.6	15.6	45.3	42	19.7	29.3	20.3	19.2	8.2	21.8	23.6	16.6	16.8	38.6
10	B31 (Partially Cross-Sectional)	26.8	19.5	49	40.9	23.5	28	14.2	13.7	12.1	22.2	31.2	21	19.6	26.7
11	S35 (Partially Cross-Sectional)	26.8	36.1	36.7	38.8	23.7	30.5	16.8	13.7	7.9	15	21.9	15.2	23.5	41.2
12	N09 (Partially Cross-Sectional)	27	22.7	36.3	47.9	25.7	30.9	14.8	14.9	9.7	20.5	31.4	17.8	20	33.3
13	N15 (Partially Cross-Sectional)	27.2	25	51.1	33.3	20	33.8	15.4	14	6.7	13	16.9	17.9	22.7	44.4
14	B28 (Partially Cross-Sectional)	27.5	20.7	52.6	40.9	20.7	28.5	14	14.9	11.9	21.8	31.5	20.4	20.8	28.8
15	S29 (Partially Cross-Sectional)	27.5	31.3	39.5	36.1	41.6	28.9	14.8	15.2	7.9	20.1	32	16.7	19.5	28.5
16	B11 (Partially Cross-Sectional)	27.5	19.9	47.8	41.4	24.3	25.8	18.1	18.6	11.7	22.9	31.7	20.9	19.2	32.6
17	N08 (Partially Cross-Sectional)	27.7	25.6	35.4	50.1	25.5	32.1	15.4	14.6	9.8	20.9	32.7	18	19.9	33.5
18	N11 (Partially Cross-Sectional)	28	23.5	35.6	51.7	28.6	32	16.3	13.9	10.1	21.6	32.8	17.9	19.7	32.5
19	RegARIMA - Battista 2013	28.1	16.8	54.1	41	21.4	29.2	19.4	18.7	9	21.7	22.2	16.3	16.9	41.9
20	S06 (RegARIMA)	28.2	23.5	56.8	31.7	29.8	30.5	13.5	15.3	8.2	18.5	27.5	14.9	18.7	39.5
21	S12 (Partially Cross-Sectional)	28.2	27.2	55.3	32.7	29.6	33.5	14.9	16	9.8	18.7	28.7	15.1	19.8	33.2
22	N28 (Partially Cross-Sectional)	28.2	28.2	34.9	49.8	31.3	32.6	13.5	14.4	11.1	20.5	33.2	17.5	20.8	32.1
23	B29 (Partially Cross-Sectional)	28.2	20.4	55.4	41.9	19.6	28.1	14.6	15	11.8	22.2	30.7	19.9	20.7	33.3
24	B08 (Partially Cross-Sectional)	28.2	19.2	51.8	42.1	23.6	26.4	19.8	19	11.9	22.7	32.2	20.4	19.8	32.2
25	S26 (RegARIMA)	28.5	37.5	51.7	31.7	30.8	27	15.1	15.8	6.9	17.7	27.6	15.4	18.7	40

26	S27 (RegARIMA)	28.6	29.5	54.4	36.5	23.4	32.8	14.7	15.8	10.2	19.6	27	14.8	18.7	39.7
27	S07 (RegARIMA)	28.6	25.2	57.6	36	24.1	32.4	15.4	16	10.2	19	26.9	14.8	18.8	39
28	N35 (Partially Cross-Sectional)	28.8	41.2	54.1	31	21	31.1	16.7	14.2	6.1	13.8	19.4	16.3	23.3	43.1
29	B09 (Partially Cross-Sectional)	28.8	19.9	53.3	42.1	22.3	27.5	20.4	19.7	12.2	22.8	31.9	19.8	19.9	34.9
30	N31 (Partially Cross-Sectional)	28.9	26.7	36.2	53	32.9	32.8	15.2	13.3	11	21.7	34.4	17.7	20.1	31.6
31	N12 (Partially Cross-Sectional)	29	24.1	55.6	40.1	21.6	33.2	15	17.2	11.1	19.1	28.2	17.2	20.2	39.7
32	N26 (RegARIMA)	29.1	32.4	60.5	35.7	18.3	26	13.4	17.8	7.5	17.8	24.7	18	19.3	43.1
33	S08 (Partially Cross-Sectional)	29.1	24.9	50.5	44.2	41.3	29.4	15	14.1	7.9	20.7	32.3	16.3	18.9	27.6
34	N07 (RegARIMA)	29.6	26.9	62.5	36.6	15.6	32.1	16.2	17.7	8.6	17.4	23.1	18.2	20.8	43.3
35	N06 (RegARIMA)	29.8	24.7	66	34.7	16	30.5	14.2	18.4	7.6	18	24.7	18.2	20.3	43.4
36	N27 (RegARIMA)	29.9	26.6	63.6	37.7	19.1	31.6	14.1	17.8	9.2	17.8	22.9	18.2	20.9	43
37	S28 (Partially Cross-Sectional)	29.9	30	50.7	40.8	47.3	27.6	16.9	13.5	7.9	20.2	32.9	17.2	19	28.5
38	S10 (Partially Cross-Sectional)	30.6	30.5	66.5	34.2	28.3	33.6	15	16.5	10.6	19	26.7	15.3	19.6	37.6
39	B12 (Partially Cross-Sectional)	31	22	63	43.7	20.1	30.5	20.5	19.2	14.6	25.7	27.6	18.8	20	40.8
40	N10 (Partially Cross-Sectional)	31.1	27.7	67.8	39.6	20.4	33.8	16	17.7	9	18.7	25	17.3	20.6	41.5
41	S30 (Partially Cross-Sectional)	31.1	31.8	66.9	35.5	29.9	34.7	13.7	16	10.3	19.4	26.4	15.2	19.5	38.5
42	S11 (Partially Cross-Sectional)	31.4	22.3	59.5	47.6	48	27.3	15.5	15	7.8	21.4	32.2	16.6	18.8	27.9
43	N30 (Partially Cross-Sectional)	31.5	28.5	67.5	40.7	23.8	33.5	14.4	17.6	9.6	18.9	27	17.2	20.3	42.2
44	S31 (Partially Cross-Sectional)	31.5	29.6	52.5	41.5	56.6	26.2	17.9	13.9	9.3	22.9	33.8	18	18.6	26.8
45	B26 (RegARIMA)	32.2	28.7	71.2	37.6	19.5	33.2	14.8	18.7	11.4	24.1	26.5	19	20.7	44.2
46	B06 (RegARIMA)	32.3	24.9	69.6	39.7	20.7	30.6	21.3	19.8	15.6	24.3	27.2	18.3	20.8	44.6
47	B60 (Partially Cross-Sectional)	32.5	32.8	45	27.9	28.1	25.4	23.6	20.2	24.5	34	26.9	38.7	42.3	40.7
48	B07 (RegARIMA)	33	24.8	72.2	40.9	18	33.9	22	18.2	16.1	25.3	26.1	19.3	20.2	44.4
49	B30 (Partially Cross-Sectional)	33.1	24.2	75.2	44.9	15.3	30.9	16.6	19.4	14.3	25.2	26.4	19.5	20.6	43
50	B10 (Partially Cross-Sectional)	33.1	23.9	73.3	43.7	20.8	31.2	21	19.8	16.8	25.2	26.5	18.6	19	43.2
51	B27 (RegARIMA)	33.3	30.7	73.5	41.2	18.5	32	20.5	18.8	13.8	25.3	26.1	19.3	20.2	44.4
52	Production (ARIMA)	33.8	27.3	75.8	37.6	26.5	40.6	25.1	20.4	12.1	23.2	19.7	16	17.2	44
53	DeepAR (RNN)	34.4	30.3	77.8	38.9	23.7	16.6	24.4	15.5	17	22.3	32.4	18.3	21.7	51