Hierarchical and Empirical Bayes Estimation of All Employee Links

Malay Ghosh
Department of Statistics, University of Florida, Gainesville, FL 32611
So Young Sohn
Department of Operations Research, Naval Postgraduate School, Monterey, CA 93943

ABSTRACT
The Bureau of Labor Statistics provides monthly sample-based estimates of employment in non-agricultural establishments. The source of these data is the Current Employment Survey (CES), also known as the 790 survey. Often, however, the CES estimates are based only on a few samples. This phenomenon reduces the reliability of these estimates, and generates the need for composite estimation, or borrow strength from other sources. We have provided in this paper hierarchical and empirical Bayes methods to produce such composite estimates. The methods are applied to the analysis of several data sets. The composite estimates are found to improve on the CES estimates most of the time.

Key Words and Phrases: Current Employment Survey, unemployment insurance administrative records, composite estimates, hierarchical Bayes, empirical Bayes, metropolitan statistical areas.

1. INTRODUCTION

The Bureau of Labor Statistics (BLS) provides monthly sample-based estimates of employment, hours and earnings of workers employed in non-agricultural establishments. Among these, the main component is employment, and the BLS provides monthly all employee (AE) estimates for a number of industrial divisions like construction, mining, durable manufacturing, non-durable manufacturing, services, wholesale trade, retail trade, etc. These figures are provided for several metropolitan statistical area (MSA's) within every state. Occasionally, these numbers are available at finer levels of classification, for example, industries classified by size, from which the aggregate figures are easy to obtain, The source of these data is the Current Employment Survey (CES), also known as the 790 survey. These employment figures can be checked against the employment counts obtained once a year from the Unemployment Insurance (UI) administrative records (the ES-202-data), since typically the latter come closer to the truth.

Sensible procedures for finding the CES estimates are given in Madow (1981), Royall (1981) or West (1984). Often, however, the CES estimates are based only on a few samples. This phenomenon reduces the reliability of the estimates, and generates the need for composite estimation, or "borrow strength" from the other sources. The current practice is to assume that employment counts are known at the base month, say month zero, so that it suffices to estimate the "links," a link being the ratio of all employee counts in a particular month to the corresponding figure in the previous month. Our target is therefore, to obtain on a monthly basis, composite estimates of the links for different industrial divisions and different MSA's within the states.

For finding the composite link for a particular month, the current practice adapted by the states is to take the CES estimate of the link as well as the average of the links for the same month from the previous year or previous three years' UI data. Then an ad hoc weighted average of the CES link and the average UI link is taken subject to the constraint that for at least 75% of state-wide- basic-cell employment estimates (i.e., establishments within industrial divisions) the CES links receive at least 85% weight. Barring this constraint, each state has the freedom to choose its own weights, which may change arbitrarily from one month to the next as well as for the different MSA's and different industrial divisions.

Clearly, the above method lacks any scientific rationale, and depends solely on experts' opinion to assess the weights. Needless to say that there are situations where such an assessment by experienced individuals may lead to composite links which are much closer to truth than the raw CES links. In the absence of such an expert guess, however, these estimates are bound to be in error, and the BLS has felt the need to formalize a procedure which provides adaptive composite links, or those where the choice of weights is dictated by the data at hand based on some valid statistical method.

In section 2 of this article, we have introduced hierarchical Bayes (HB) and empirical Bayes (EB) procedures which meet this need. Each method produces composite estimates of links which are weighted averages of the CES links, and certain regression estimates. For a given month and a given industrial division, the proposed composite estimate for a certain MSA borrows strength from other MSA's in the state. We have derived these estimators, and the associated standard errors based on a slight generalization of the work of Lindley and Smith (1972), Morris (1981) or Ghosh (1989).

We have applied the methods of Section 2 to several actual data sets. One such analysis is presented in Section 3. The link data pertain to the "manufacturing" industry for six metropolitan statistical areas (MSA's) namely (1) Charlottesville, (2) Danville, (3) Lynchburg, (4) Norfolk-Virginia Beach-Newport News, (5) Richmond-Petersburg, and (6) Roanoke in the state of Virginia. The estimates are compared with the corresponding UI estimates, since typically the latter come much closer to the truth. It turns out that the HB and the EB estimates are almost always closer to the UI estimates than the corresponding CES estimates. The ad hoc estimates may occasionally come closer to the truth than the HB or the EB estimates but fail miserably on other occasions, and definitely perform worse than the HB and the EB estimates on an average. Also, we reiterate that experts' guess to find adjusted CES estimates is more art than science, and our method should have a more universal appeal than any ad hoc adjustment scheme.

2. HB AND EB ESTIMATION

We fix a month, say k, and also an industry, say i. Let Z_{kij} denote the AE total for the ith industry in month k in the MSA j based on the CES data. Define $Y_{kij} = Z_{kij}/Z_{k-1,ij}$ $(j = 1, ..., m)$, when m denotes the number of MSA's. Also, let u^l_{kij} denote the UI figure for the ith industry in month k for the jth MSA l years previous to the current year; $l = 1, ..., p$ (l=1 referring to the previous year). Define $x^l_{kij} = u^l_{kij}/u^l_{k-1,ij}$. Since the month and the industry are held fixed, we shall omit the suffixes k and i from now on. Define $\mathbf{Y}_j = (Y_1, .., Y_m)^T$, $\mathbf{x}_j^T = (x^1, .., x^p)^T$, $j = 1, .., m$. Also, let θ_j denote the true AE link for the jth MSA. The following hierarchical model is used:

I. $Y_1, .., Y_m, S \mid \theta_1, .., \theta_m, \beta, \sigma^2, \tau^2$ are mutually independent with $Y_j \sim N(\theta_j, \sigma^2)$ and $S \sim \sigma^2\chi^2_\nu$;

II. $\theta_1, .., \theta_m, \mid \beta, \sigma^2, \tau^2$ are mutually independent with $\theta_j \mid N(\mathbf{x}_j^T\beta, \tau^2)$;

III. Marginally, β, σ^2, and τ^2 are mutually independent with $\beta \sim uniform(R^p)$, σ^2 has pdf $f(\sigma^2) \propto (\sigma^2)^{-1}I_{(0,\infty)}(\sigma^2)$, τ^2 has pdf $f(\tau^2) \propto (\sigma^2)^{-1}I_{(0,\infty)}(\tau^2)$, where I denote the usual indicator function.

The random variable S is usually taken as $\sum_{j=1}^m \sum_{l=1}^n (Y_j^l - \bar{Y}_j)^2$, where $Y_j^1, .., Y_j^n$ denote the CES links for the MSA j based on years 1,...,n respectively in the past (year 1 referring to the previous year), and $\bar{Y}_j = \sum_{l=1}^n Y_j^l/n$, $j = 1, .., m$. In our actual applications, n is small, say 2 or 3. The present model works well when for a given industrial division and a fixed month, the variability is similar among the CES links of the different MSA's over a certain number of years. The model remains unaffected even if there is considerable variation among the links over different months and different industrial divisions. However, in the event of sudden changes (e.g., oil crisis, crisis in the automobile industry, etc.) which may affect the AE links of a certain industry within a bigger industrial division, one may need a change in the present model.

Our objective is to find $E(\theta_j \mid y, s)$ and $V^{1/2}(\theta_j \mid y, s)$. The present HB analysis provides a more detailed information about the posterior distribution of θ from which the posterior means and s.d.'s can be calculated.

Before, stating the theorem, we need to introduce a few notations. Write $\theta = (\theta_1, .., \theta_m)^T$, $X^T = (x_1, .., x_m)$, $P_x = X(X^TX)^{-1}X^T$, assuming rank $(X) = p$. Also, let $\sigma^2 = R^{-1}$ and $\tau^2 = (\Lambda R)^{-1}$, $B = \Lambda/(1 + \Lambda)$. Also, Z denotes Gamma$(\alpha, \gamma)$ variable. Then the following theorem can be proved.

THEOREM 1. Consider the hierarchical model given in (I)-(III) with $m \geq p + 4$. Then,

(i) $\theta \mid R = r, B = b, Y = y, S = s \sim N[(1 - b)y + bP_xy, r^{-1}((1 - b)I_m + bP_x)]$;

(ii) $R \mid B = b, Y = y, S = s \sim Gamma[(0.5(by^T(I_m - P_x)y + s), 0.5(m + \nu - p - 2)]$;

(iii) $B \mid Y = y, S = s$ has pdf $f(b \mid y, s) \propto b^{(m-p-4)/2}(1 + by^T(I_M - P_x)y/s)^{-(m+\nu-p-2)/2}I_{[0,1]}(b)$;

(iv) $E(\theta \mid y, s) = [1 - E(B \mid y)]y + E(B \mid y)P_x y$;

(v) $V(\theta \mid y, s) = V(B \mid y, s)(I_m - P_x)yy^T(I_m - P_x) + [s + y^T(I_m - P_x)y](m + \nu - p - 4)^{-1}[(1 - E(B \mid y, s))I_m + E(B \mid y, s)P_x]$.

In Morris (1981), the model given in (I) - (III) was considered, with the exception of known σ^2, so that no prior distribution needed to be assigned to σ^2. Accordingly, the posterior distribution of B given Y=y and S=s was different from what is given here.

In an EB formulation, we assume (I) and (II) of the model, but not (III). First, we find the posterior distribution of θ under (I) and (II) as

$$\theta \mid y, s \sim N[(1-b)y + bX\beta, \sigma^2(1-b)I_m]. \quad (1)$$

Since, marginally $Y \sim N(X\beta, (\sigma^2 + \tau^2)I_m)$ i.e., $N(X\beta, (\sigma^2 b^{-1} I_m)$ using the method of maximum likelihood, β is estimated by $\hat{\beta} = (X^T X)^{-1} X^T Y$ and $(\sigma^2 b^{-1})^{-1} = b\sigma^{-2}$ is estimated by $(m - rank(X) - 2)/(Y^T(I_m - P_x)Y) = (m - p - 2)/(Y^T(I - P_x)Y)$. Finally, σ^2 is estimated by $S/(\nu + 2)$. Accordingly, one estimates b by

$$\hat{b} = min[1, (S/(\nu+2))(m-p-2)/(Y^T(I-P_x)Y)). \quad (2)$$

The EB estimator of θ is thus

$$\hat{\theta}^{EB} = (1-\hat{b})y + \hat{b}P_x y. \quad (3)$$

The MSE matrix associated with $\hat{\theta}^{EB}$ is $(S/(\nu + 2))(1-\hat{b})I_m$ which is typically an underestimate. This is because the EB procedure as opposed to the HB procedure, does not incorporate the uncertainty due to estimation of in computing the MSE matrix.

3. DATA ANALYSIS

We consider the six MSA's labeled (1) - (6) (in the introduction) within the state of Virginia. Also, the months January through December are coded as 1,...,12. Denote by $e^k_{j,CES}$, $e^k_{j,UI}$, , $e^k_{j,HB}$ $e^k_{j,EB}$, $e^k_{j,VA}$ the CES estimate, the UI estimate, HB estimate, EB estimate, and the adjusted Virginia estimate for the jth MSA and the kth month respectively. The standard errors associated with the HB and EB estimators are denoted $s^k_{j,HB}$'s and $s^k_{j,EB}$'s respectively. The numbers are provided for all six MSA's and all the 12 months. In our formulas for the HB and EB estimators derived in Section 2, we have chosen n=2 and p=1, that is we have used previous two years' CES data from each MSA to estimate S and only the previous year's UI estimate in the regression model. We have also tried larger values of n and p. But the resulting estimates have, in general, drifted further away from the UI estimates than the (2,1) combination of (n,p).

Table 1 provides all the estimates as well as the standard errors associated with the HB and the EB estimates for the manufacturing industry in Virginia. The HB and the EB estimates are clearly superior to the CES estimates for nearly every month, but the dominance, holds for most of the months in all the MSA's. We have prepared Table 2 showing the average (over the eight MSA's) MSE's (the UI figures being treated as truth) of the CES, HB, EB and the adjusted Virginia estimates for each of the 12 months. The HB and the EB estimates perform better than the CES estimates nearly all the time, and better than the adjusted Virginia estimates on most of the occasions. A further average over all the 12 months provides an overall average MSE of .00015559 for the HB estimates, .00016457 for the EB estimates, .00016927 for the adjusted Virginia estimates and .00017857 for the CES estimates. Thus the HB estimates achieve an overall MSE reduction of 5.46in comparison with the adjusted Virginia estimates, and 12.87comparison with the CES estimates.

A similar phenomenon occurs when one computes the average (over the MSA's) mean relative errors (MRE's) for different months. For the HB estimate, the MRE is defined by $\mid UI - HB \mid / \mid UI \mid$. Similarly MRE's for EB and CES estimates are defined. Once again, the HB and the EB estimates have a clear-cut superiority over the CES as well as the adjusted Virginia estimates on most of the occasions. With the exception of months 5,6, and 10, the HB or the EB estimates have usually smaller MRE than the corresponding CES estimates, although in months 6 and 10, the HB estimates are only slightly inferior to the adjusted Virginia estimates. The HB and the EB estimates have improved MRE's over the adjusted Virginia estimate only half the time, but in this respect, on an average, the HB estimates improve on the adjusted Virginia estimates by 12.49Virginia estimates by 10.39

ACKNOWLEDGMENTS

We wish to thank Kenneth Shipp, George Stamas, Christopher Johnson, Stephen Woodruff and John Sommers for their many valuable discussions.

REFERENCES

Ghosh, M. (1989), "Hierarchical and Empirical

Bayes Multivariate Estimation, "Technical Report No. 330, University of Florida, Department of Statistics.

Lindley, D.V., and Smith, A.F.M. (1972), "Bayes Estimates for the Linear Model" (with discussion), Journal of the Royal Statistical Society, B, 1-41.

Madow, W.G. (1981), "Role of Models in Sample Design and Estimation on the 790 Revision, "Bureau of Labor Statistics Report.

Morris, C. (1981), "Parametric Empirical Bayes Confidence Intervals", Scientific Inference, Data Analysis and Robustness, eds. G.E.P. Box, T. Leonard, and C.F. Jeff Wu, Academic Press, pp. 25-50.

Royall, R.M. (1981), "Study of Role of Probability Models in 790 Survey Design and Estimation, "Bureau of Labor Statistics Report.

West, S.A. (1981), "Linear Models for Monthly All Employment Data," Bureau of Labor Statistics Report.

Table 2. The MSE's and MRE's For Different Estimates

k	MSE($c^k_{\cdot,HB}$)	MSE($c^k_{\cdot,EB}$)	MSE($c^k_{\cdot,VA}$)	MSE($c^k_{\cdot,CES}$)	MRE($c^k_{\cdot,HB}$)	MRE($c^k_{\cdot,EB}$)	MRE($c^k_{\cdot,VA}$)	MRE($c^k_{\cdot,CES}$)
1	.00005926	.00005031	.00060262	.00007728	0.004923	0.004922	0.018531	0.005766
2	.00003960	.00003242	.00005038	.00006085	0.005545	0.004705	0.006122	0.006298
3	.00006765	.00011999	.00000525	.00008779	0.007100	0.009671	0.001611	0.007637
4	.00010523	.00009406	.00005912	.00011812	0.007726	0.007565	0.007244	0.008173
5	.00003198	.00003951	.00003172	.00002569	0.005257	0.005659	0.003956	0.004523
6	.00003063	.00004280	.00005026	.00004463	0.005180	0.005645	0.005947	0.005110
7	.00012692	.00013016	.00024597	.00015888	0.009547	0.008713	0.013903	0.010179
8	.00067443	.00068089	.00044153	.00069163	0.016047	0.016055	0.014524	0.016067
9	.00061673	.00066347	.00035369	.00071871	0.013660	0.013882	0.011545	0.014424
10	.00002306	.00003574	.00004825	.00002072	0.004072	0.004556	0.005913	0.004022
11	.00004551	.00004495	.00004415	.00007337	0.004374	0.004335	0.003803	0.005949
12	.00004613	.00004057	.00009833	.00006546	0.005259	0.005107	0.008247	0.006212
AVG	.00015559	.00016457	.00016927	.00017857	.0073909	.0075679	.0084455	.0078635

Table 1. Showing the Different Estimates and the Standard Errors for the HB and EB Estimates

j	k	$e^k_{j,UI}$	$e^k_{j,CES}$	$e^k_{j,VA}$	$e^k_{j,HB}$	$s^k_{j,HB}$	$e^k_{j,EB}$	$s^k_{j,EB}$
1	1	0.98889	0.98802	1.00000	0.99038	.0075329	0.99208	.0045661
2	1	0.98817	0.98655	0.99401	0.98892	.0074930	0.99062	.0045013
3	1	0.99138	0.99574	0.99149	0.99427	.0076628	0.99322	.0049304
4	1	0.97959	0.99958	1.02242	0.99723	.0074061	0.99554	.0043681
5	1	0.99208	0.99287	0.98113	0.99261	.0074737	0.99243	.0047519
6	1	0.99000	0.99641	0.95122	0.99575	.0079810	0.99528	.0054595
1	2	1.00000	0.98616	0.98864	0.99100	0.012819	0.99691	.0052127
2	2	0.99401	0.99176	1.00000	0.99256	0.014212	0.99354	.0097165
3	2	0.99565	1.00300	1.00000	1.00078	0.012906	0.99807	.0064121
4	2	1.00595	0.99620	1.00292	0.99614	0.012497	0.99606	.0053917
5	2	1.00319	1.00338	1.00160	0.99985	0.012859	0.99555	.0059185
6	2	0.99495	0.99935	0.98462	0.99951	0.014176	0.99972	.0096597
1	3	0.97753	0.96330	0.97701	0.96512	0.047692	0.96710	0.038856
2	3	0.98193	0.97266	0.98193	0.98118	0.038312	0.99042	0.016924
3	3	1.00873	1.00680	1.00429	0.99829	0.038269	0.98905	0.016782
4	3	0.99408	0.99871	0.99708	0.99798	0.039411	0.99719	0.021220
5	3	1.00159	1.00234	1.00000	0.99693	0.038156	0.99105	0.017084
6	3	1.00508	0.99050	1.00521	0.99482	0.040425	0.99950	0.023660
1	4	1.01149	0.99479	1.02353	0.99500	0.045722	0.99521	0.039343
2	4	1.00000	0.99914	1.00613	1.00037	0.036634	1.00161	0.020615
3	4	1.00433	0.99968	0.99573	1.00020	0.035467	1.00072	0.017307
4	4	1.00446	1.00103	1.01170	1.00077	0.035303	1.00050	0.016798
5	4	0.99682	1.00060	1.00160	1.00072	0.035576	1.00084	0.017641
6	4	0.98485	1.00438	0.98964	1.00255	0.035481	1.00072	0.017307
1	5	0.98864	0.99456	0.98851	0.99765	.0093424	0.99935	.0059787
2	5	1.01227	1.01089	1.01829	1.00770	.0094039	1.00595	.0060360
3	5	0.99569	0.99992	0.99571	1.00064	.0086828	1.00103	.0054617
4	5	1.00741	1.00434	1.00867	1.00318	.0086492	1.00254	.0053850
5	5	1.00319	0.99453	0.99201	0.99668	.0092268	0.99786	.0060060
6	5	1.00000	1.00386	1.00524	1.00225	.0095177	1.00137	.0064343
1	6	1.00000	0.99916	1.01163	1.00295	0.020253	1.00734	0.014064
2	6	0.99394	0.99590	1.00000	0.99860	0.020809	1.00173	0.015326
3	6	1.00866	1.00590	1.00862	1.00566	0.017695	1.00537	0.008068
4	6	1.01471	1.01141	1.01146	1.00822	0.017637	1.00453	0.007516
5	6	1.00954	0.99911	1.00000	1.000150	0.017650	1.00427	0.007721
6	6	1.00513	1.01672	1.01042	1.01127	0.017863	1.00495	0.007561

Table 1. Showing the Different Estimates and the Standard Errors for the HB and EB Estimates (cont'd)

j	k	$e^k_{j,UI}$	$e^k_{j,CES}$	$e^k_{j,VA}$	$e^k_{j,HB}$	$s^k_{j,HB}$	$e^k_{j,EB}$	$s^k_{j,EB}$
1	7	0.98851	0.97914	0.97701	0.98295	0.017019	0.98798	0.007708
2	7	0.99390	1.00789	1.01796	1.00317	0.017167	0.99694	0.007861
3	7	1.00000	0.97550	0.97863	0.97907	0.017951	0.98379	0.010321
4	7	0.99565	0.99839	0.99858	0.99827	0.017104	0.99811	0.008512
5	7	0.99213	0.99513	1.00000	0.99811	0.018285	1.00205	0.011224
6	7	1.00510	0.99782	0.98969	0.99229	0.017839	0.98499	0.009482
1	8	1.03488	1.01894	1.00000	1.01544	0.017329	1.01696	.0070734
2	8	1.00000	0.99711	1.00000	0.99982	0.017489	0.99865	.0071573
3	8	1.00858	1.00455	1.00000	1.00379	0.016855	1.00412	.0070764
4	8	0.99709	0.99031	0.99291	0.99036	0.016820	0.99034	0070739
5	8	1.01270	1.00654	1.00805	1.00613	0.016881	1.00631	.0070861
6	8	1.00508	0.94354	0.96875	0.94544	0.017689	0.94462	.0072396
1	9	1.01124	1.00507	1.00000	1.00620	0.030198	1.00566	0.010106
2	9	1.04908	1.06204	1.04118	1.06119	0.031144	1.06159	0.010272
3	9	1.00851	1.01012	1.00437	1.01115	0.030152	1.01066	0.010099
4	9	0.99416	0.99295	0.99429	0.99494	0.030301	0.99400	0.010110
5	9	1.00157	1.00347	0.99840	1.00496	0.030192	1.00426	0.010100
6	9	1.01010	1.07412	1.05376	1.06934	0.030935	1.07159	0.010123
1	10	1.00000	0.99507	1.01176	0.99748	.0094706	1.00055	.0063728
2	10	0.98830	0.99262	0.99435	0.99521	.0095088	0.99850	.0064122
3	10	1.00844	1.00414	1.00435	1.00204	.0082939	0.99937	.0034236
4	10	1.00147	1.00065	0.99282	1.00019	.0081595	0.99961	.0033306
5	10	1.00000	1.00211	1.00480	1.00087	.0082552	0.99929	.0035404
6	10	0.99500	1.00260	0.99490	1.00140	.0083229	0.99987	.0037662
1	11	1.00000	1.01921	1.00000	1.01489	0.010286	1.01479	.0062931
2	11	1.00000	1.00113	1.00000	1.00072	0.009340	1.00071	.0060664
3	11	1.00000	0.99562	1.00433	0.99484	0.010572	0.99482	.0068835
4	11	0.99853	0.99524	0.99855	0.99821	0.009785	0.99828	.0061669
5	11	1.00469	0.99858	1.00159	0.99980	0.009362	0.99983	.0060496
6	11	1.00000	0.99840	0.98462	0.99971	0.009375	0.99975	.0060528
1	12	1.00000	1.00514	1.01163	1.00124	0.010809	0.99828	.0061852
2	12	0.98817	0.98580	0.98864	0.98627	0.011895	0.98662	.0083313
3	12	1.00418	0.99535	0.99138	0.99372	0.010712	0.99248	.0064956
4	12	0.99266	0.99280	0.98986	0.99384	0.010485	0.99463	.0061852
5	12	1.00623	1.00155	1.00000	1.00095	0.011477	1.00049	.0077347
6	12	1.00000	0.98385	0.98438	0.98847	0.010949	0.99199	.0061852