IMPROVING MODELS TO ESTIMATE BIAS IN PAYROLL EMPLOYMENT ESTIMATES

George Stamas, Diem-Tran Kratzke, and Kirk Mueller
Bureau of Labor Statistics
George Stamas, 2 Massachusetts Ave. N.E., Washington, D.C. 20212

KEY WORDS: Kalman filter, revisions, time series

The Current Employment Statistics (CES) survey is a monthly payroll survey of more than 380,000 non-agricultural business establishments. It provides estimates of employment, hours, and earnings by industry for the Nation, States and Metropolitan Areas. Employment estimates are revised annually in a process called benchmarking, reconciling estimates to counts of employment available from administrative records of the Unemployment Insurance programs (UI) (BLS, 1988). Revisions made from benchmarking have been large, occasionally, even for total payroll employment at the national level. The objective of this research is to improve CES estimates to reduce benchmark revisions in a manner that can be applied in a production environment.

Background

The estimator for employment in the CES program is called the link relative estimator. The link, the ratio of the reported current month's employment, e_c, to the reported previous month's employment, e_p, is multiplied by the estimated previous month's employment, $\hat{E}_p$, to estimate the current month employment, $\hat{E}_c$. Only the sample units that appear in both months are used in computing the link.

$$\hat{E}_c = (e_c/e_p) \cdot \hat{E}_p$$

Employment estimates for a given month are revised four different times, the last revision is the benchmark revision. The benchmark employment level in March is taken from the Covered Employment and Wages (ES-202) program which summarizes quarterly data for workers covered by the States' UI programs. Benchmark revisions for other months are smoothed over the 12-month period. Our work deals with the third estimate (second revision). This third-closing estimate is produced at the beginning of the third month following the reference month.

The published employment estimates from the CES survey include an adjustment for bias that results primarily from the inability of the sample to capture employment at new establishments and to reflect losses when establishments close down in a timely fashion. This results in the sample frame bias and the response bias. The establishments are not on the frame to be sampled because of the lag between going into business and showing up on a UI name and address file available to the state's CES operation. This is a sample frame bias. Failure to receive a report from an establishment may be interpreted as only a nonresponse when the establishment has actually gone out of business. This is a response bias.

Methods used to date to adjust for the bias, such as simple averages of three previous benchmark revisions or regression adjusted averages, have not performed consistently.

In a recent article, Neumark and Wascher (1991) have shown that the use of other economic information available in time for the release of preliminary CES estimates can reduce the annual revision in total payroll employment. Although this work was done on the national level, it was hoped that some of the information would be of help in modeling at the industry division level.

In this paper, we will attempt to adjust for third closing bias at the major industry division level. The divisions are Mining; Construction; Manufacturing; Transportation and Public Utilities (TPU); Wholesale Trade; Retail Trade; Finance, Insurance, and Real Estate (FIRE); and Services.

Data

Two series of data are used for the research: the CES published third-closing estimates and the ES-202 data from April 1983 through June 1992. Recall that the published CES estimates were adjusted for bias. We removed the bias adjustment from the third-closing estimates and computed the monthly links. Beginning with third-closing estimates in November 1988, records of the amount of monthly bias added to the estimates are available. Monthly links were computed at the division level by removing the bias from the published employment estimates and dividing the unadjusted estimates by the bias-adjusted previous month's estimates. Prior to November 1988, only the annual amounts of bias were available. Analysts believe that the monthly bias did not vary much during this period. Monthly links were computed by prorating the annual amount of bias and removing it evenly from these monthly links.

We consider the ES-202 data series as the true employment series and computed the ES-202 monthly links after the effects of noneconomic code changes are removed. Noneconomic changes are administrative changes in establishment industry or location assignment. State Employment Security Agencies implement these changes January of each year along with a revised December employment level. To avoid drastic movement in January links caused by this reassignment, we recalculated January ES-202 links using the revised December employment level. Reassignments have less of an impact on CES links because CES links are calculated using only establishments that are in the sample both the current and the previous month. Both series were prior adjusted to remove the effect of strikes.

We constructed a third series, the bias series, by taking the difference between the ES-202 monthly links and the third-closing monthly links. For simplicity, we will refer to the ES-202 monthly link series as the employment (link) series, the third-closing monthly links as the sample (link) series, and their difference as the (link) bias series.

We also have series of potential explanatory variables available: the change in the Index of Leading Economic Indicators, including several of its component parts, CPS employment, initial claims for unemployment benefits, sample ratio (ratio of seasonally adjusted sample link to the previous month's seasonally adjusted sample link), the change in average number of hours of production and non-supervisory workers, and quarterly birth data .

Model

The basic model for y_t^S, the sample link, is

$$y_t^S = y_t^E + y_t^B,$$

where y_t^E is the employment link at time t, and y_t^B is the link bias at time t.

The employment link and the link bias series are modeled separately with state space models as described in Harvey (1989). The relationship above will be used to update the state vector every time a new sample link is available. From there, the true employment link is forecasted.

For each industry division, a basic structural model is used to fit the employment link series and the link bias series. The model is set up explicitly in terms of components that have direct interpretation. The employment link is written as

$$y_t^E = \mu_t^E + \gamma_t^E + \varepsilon_t^E$$

where μ_t^E is a local linear trend

$$\mu_t^E = \mu_{t-1}^E + \beta_{t-1}^E + \upsilon_t^E \ ,$$

$$\beta_t^E = \beta_{t-1}^E + \xi_t^E \ ,$$

and γ_t^E is a local seasonal pattern

$$\sum_{j=1}^{11} \gamma_{t-j}^E = \omega_t^E \ .$$

In state space form this is expressed as follows.

$$y_t^E = (z^E)' \cdot \alpha_t^E + \varepsilon_t^E \qquad (1)$$

$$\alpha_t^E = T^E \cdot \alpha_{t-1}^E + \eta_t^E \qquad (2)$$

where

$$\alpha_t^E = [\ \mu_t^E \quad \beta_t^E \quad \gamma_t^E \quad \gamma_{t-1}^E \quad \cdots \quad \gamma_{t-10}^E\ ]' \ ,$$

$$z^E = [\ 1 \quad 0 \quad 1 \quad 0 \quad \cdots \quad 0\ ]',$$

$$\eta_t^E = [\ \upsilon_t^E \quad \xi_t^E \quad \omega_t^E \quad 0 \quad \cdots \quad 0\ ]',$$

and

$$T^E = T = \begin{vmatrix} 1 & 1 & 0\cdots0 \\ 0 & 1 & 0\cdots0 \\ 0 & 0 & -1\cdots-1 \\ 0 & 0 & \quad\quad 0 \\ \vdots & \vdots & I \quad \vdots \\ 0 & 0 & \quad\quad 0 \end{vmatrix}_{13x13}$$

We assume that the error ε_t^E is a serially uncorrelated disturbance distributed $N(0,h^E)$, and η_t^E is a vector of serially uncorrelated disturbances distributed $N(0,Q^E)$. The disturbances are uncorrelated with each other for all time periods and are uncorrelated with the initial state vector α_0^E, which is assumed to be distributed $N(a_0^E, P_0^E)$.

The model could be extended to include explanatory variables,

$$y_t^E = \mu_t^E + \gamma_t^E + (x_t)' \cdot \delta_t + \varepsilon_t^E$$

where the vector x_t contains the explanatory variables, the vector δ_t contains the unknown parameters associated with them, and

$$\delta_t = \delta_{t-1} + \nu_t .$$

ν_t is a white-noise disturbance vector with a positive semi-definite covariance matrix.

When the model includes explanatory variables, the measurement equation (1) and the transition equation (2) of the state space remain the same. However the measurement vector z^E and the state vector α_t^E change slightly to include additional information as

$$z^E = \begin{bmatrix} 1 & 0 & 1 & 0 & \cdots & 0 & x_t' \end{bmatrix}' ,$$

$$\alpha_t^E = [\mu_t^E \quad \beta_t^E \quad \gamma_t^E \quad \gamma_{t-1}^E \quad \cdots \quad \gamma_{t-10}^E \quad \delta_t{}']' ,$$

and the transition matrix becomes $T^E = \begin{bmatrix} T & 0 \\ 0 & I \end{bmatrix}$.

The bias link series is modeled as a basic structural model, similarly to the employment link model above, with no explanatory variables. In state space form, all equations of the basic structural model stay the same.

Models of the employment link and the link bias could be combined in one state space form, relating their relationship to the sample link. We refer to this as a two-equation model. We could write the sample link in state space form as

$$y_t^S = (z^S)' \cdot \alpha_t^S + \varepsilon_t^S ,$$

$$\alpha_t^S = T^S \cdot \alpha_{t-1}^S + \eta_t^S ,$$

where

$$z^S = \begin{bmatrix} z^E \\ z^B \end{bmatrix}, \alpha_t^S = \begin{bmatrix} \alpha_t^E \\ \alpha_t^B \end{bmatrix}, \eta_t^S = \begin{bmatrix} \eta_t^E \\ \eta_t^B \end{bmatrix}, \; T^S = \begin{bmatrix} T^E & 0 \\ 0 & T^B \end{bmatrix}$$

and $\varepsilon_t^S = \varepsilon_t^E + \varepsilon_t^B$.

We would like to estimate α_t^S given all the information available up to and including time $t-1$. Let a_{t-1}^S be the optimal estimator of α_{t-1}^S based on all information available up to and including time $t-1$, with mean square error P_{t-1}^S. We estimate α_t^S at time $t-1$ by

$$a_{t/t-1}^S = T^S \cdot a_{t-1}^S ,$$

with MSE

$$P_{t/t-1}^S = T^S \cdot P_{t-1}^S \cdot (T^S)' + Q^S .$$

Once the observation y_t^S is available at time t, we can update the state vector α_t^S by a_t^S with MSE P_t^S using the usual updating equations of the filter.

The employment link at time t could then be forecast to be

$$y_t^E = \begin{bmatrix} z^E & 0 \end{bmatrix}' \cdot a_t^S .$$

This approach is similar to Harvey (1984) for data revisions, although he models the components as autoregressive processes. Coomes (1988) applies a similar approach to CES area estimates. Sommers and Stamas (1991) apply this model, without explanatory variables, to CES data for selected two digit SICs.

Alternatively, the sample link could be used as an explanatory variable in the extended model for the employment link, without using information from the link bias series. In this paper, this is referred to as a single equation model.

Estimation

In the two-equation model, the employment link and the link bias are modeled separately. In this section the superscripts indicating employment or bias series will be dropped. Each model is initialized with a_0 as a vector of zeros and $P_0 = \kappa I$ where κ is a large number and I is an identity matrix. Maximum likelihood is used to estimate the hyperparameters using the prediction error decomposition method. The hyperparameters of Q are expressed relative to h. We assume the covariance of the level and the seasonal component are zero, making Q a diagonal matrix of variances for each component. The variance h is a scalar representing the variance of the disturbance term in the measurement equation. The diagonal elements of the matrix Q/h that have not already been assumed zero are selected to minimize L.

$$L = \sum_{t=d+1}^{T} \log f_t + (T-d)\log(1/T-d)\Big(\sum_{t=d+1}^{T} v_t^2 / f_t\Big)$$

where $f_{t/t-1}^S = (z^S)' \cdot P_{t/t-1}^S \cdot z^S + h^S$, T is the number of observations used in the estimation of the parameters, d is the number of rows in the transition matrix, and $v_t = y_t - \hat{y}_{t/t-1}$.

Innovations, the differences between the predicted and the realized value, are incorporated into the likelihood function after d observations. Minimization was achieved using the method, with scaling, of Broyden, Fletcher, Goldfarb and Shanno as programmed in GAUSS386i (1993). Relative variances are returned to the original Q, nonrelative form, and the filter is reprocessed, returning estimates of a_T and P_T to initialize the process estimating employment through the sample link model.

Evaluating the models

Models of the employment link and the link bias series are tested for serial correlation and normality of the residuals (prediction errors in the updating equation). The method of testing the standardized residuals as given in Harvey (1989) is used for serial correlation. The statistic to test the significance of the first P residual autocorrelation, Q^*, takes the form described in Harvey (1989, equation 5.4.7). We have chosen two values for P, 5 and 12.

The statistic used to test the normality assumption can be found in Harvey (1989, equation 5.4.12). The associated null hypothesis is that the errors are normally distributed.

We have 110 months of historical ES-202 data available from April 1983 through June 1992. We would like to evaluate the predictive ability of a model based on fifteen months of forecasts, from April 1991 through June 1992. To simulate the actual data availability of the two series, we divide the fifteen months of forecasts into five three-month periods. The first period is April 1991 through June 1991. ES-202 data for a particular quarter is not available until six months later. Thus we fit the model using ES-202 data available through September 1990. Then forecasts are made for October 1990 through June 1991, but only the last three months of forecast errors are evaluated, as the first six months in the projection would always fall into the period before the reference month. The last three months of forecast errors correspond to forecast errors of the seventh, eighth, and ninth month for each period. When additional ES-202 data is available through December 1990, forecasts are then made for January 1991 through September 1991, and only the last three months of forecast errors are evaluated. The process is continued until all five periods are evaluated, each period has three months of forecast errors. This constitutes a total of 15 months of forecast errors. The forecast error or the prediction error is defined as the absolute difference between the true employment link and the forecasted employment link. The mean prediction error (MPE) is defined as the average of these prediction errors over the 15 months.

Another measure of error is the extrapolitive sum of squares (ESS) which is the sum of the squared errors for all of the observations in the forecast period. After fitting models through March 1991, we calculated the ESS for the April 1991 to March 1992 period, divided by 12 and called it the mean extrapolated squared error (MESE). We compare these with the sample estimates.

To compare the model based forecast with the benchmark revision in March, models are also forecasted either 9 months or 12 months ahead to give a March 1992 forecast. These forecasts for March 1992 are compared with the sample based estimates (CES third closing estimates without the bias adjustment) and the benchmark revisions.

Results

Employment and bias models were run for total private and the eight major industry divisions. The two time series models considered are the one that includes level, trend, and seasonal components (LTS) and the one that includes level and seasonal components (LS). For employment models, explanatory variables were also added. Normality and serial correlation are tested on all models fitted from April 1983 through December 1990.

Testing the underlying assumptions. Tables 1 and 2 present the diagnostic results for some of the employment and bias models. The assumptions that the standardized residuals are normally distributed and serially uncorrelated appear to be violated in many of the division level results. The distribution of residuals show a high level of kurtosis which is often an indicator of outliers. However, efforts to identify outliers by large residual values and remove them lead to other observations being identified as outliers. Outliers can impact heavily on the ability of a model to forecast. Because most outliers are at the first month of a quarter, we suspect that they are related to administrative changes in the ES-202 program. Also affecting normality could be errors associated with creating unbiased sample links by removing constant amounts of bias for each month in a 12-month period early in the series. It also possible that this absence of normality in the residuals is a result of what is, comparatively speaking, a small sample size.

The results from these two tests suggest that, except in limited cases, we have misspecified the model or that we have data problems. Despite the evidence of serial correlation and the lack of support for normality, we proceeded to compare forecasts from each of these models across the industry divisions.

Table 1. P-Values for Normality of Standardized Residuals

Division	Employment				Bias	
	LS	LTS	LS+ sample	LTS+ sample	LS	LTS
Total private	.10	.05	.40	.27	.36	.30
Mining	.15	.24	.00	.01	.00	.00
Construction	.08	.34	.06	.07	.95	.75
Manufacturing	.42	.60	.00	.00	.00	.00
Transportation	.00	.00	.00	.00	.00	.00
Wholesale	.63	.65	.61	.17	.68	.30
Retail	.58	.39	.00	.00	.00	.00
Finance	.00	.01	.00	.00	.00	.00
Services	.04	.00	.00	.00	.00	.00

Table 2. P-Values for Serial Correlation in Standardized Residuals -- (5 month, 12 month test)

Division	Employment				Bias	
	LS	LTS	LS+ sample	LTS+ sample	LS	LTS
Total	.02, .02	.02, .02	.18, .08	.08, .14	.36, .13	.19, .09
Mining	.13, .30	.07, .29	.04, .30	.01, .23	.21, .65	.08, .45
Constr.	.67, .31	.40, .36	.00, .00	.00, .00	.01, .01	.01, .03
Mfg.	.66, .94	.43, .89	.03, .00	.01, .00	.07, .00	.04, .00
TPU	.75, 1.00	.74, .99	.59, 96	.49, .95	.81, .98	.76, .97
Whlsale	.47, .84	.33, .77	.05, .19	.07, .30	.13, .24	.16, .32
Retail	.02, .01	.03, .01	.04, .00	.03, .00	.00, .01	.33. .01
FIRE	.03, .05	.01, .04	.00, .00	.00, .00	.00, .00	.00, .00
Services	.36, .11	.21, .10	.28, .90	.18, .94	.57, .91	.74, .89

Evaluating forecasts. Table 3 presents the MPE for some of the one-equation and two-equation models by division. In the two-equation models, the time series components of the bias series are the same as the ones of the employment series. With the exception of construction and TPU, the models do as well or better than sample.

The models that include sample as an explanatory variable appear to do better in total, manufacturing, TPU, and services. Otherwise models with only the time series components are competitive with any of the alternatives.

Due to space constraints, not all models are shown in our tables. However, models with sample ratio or leading indicator as an explanatory variable in addition to sample improved the mean predictive errors slightly. Diagnostics on normality and serial correlation for these models are similar to the ones with only sample as an explanatory variable.

Table 4 presents the MESE for each model by division for two-equation models. Models with sample as an explanatory variable appear to perform marginally better than those without. Of note from this table is the relative performance of the sample as an estimator for employment. The models routinely perform better. This is most likely because the models estimate the seasonal movement in the ES-202 while the sample is known to have a different seasonal pattern.

Table 5 compares hypothetical revisions from the model based forecasts (two-equation models) with the sample based estimates and the published benchmark revision in March 1992 (adjusted for noneconomic code changes). The benchmark revision amounts to the error on a 12-month forecast. Forecasting over 9 months, the model based forecasts generally produce smaller revisions than the sample based estimates; the exceptions are mining, and TPU (and FIRE for LS+sample). With the exception of TPU, the revisions in the model based forecasts compare favorably to the revisions in the published estimates. When we compare the 9-month and 12-month forecasts from the model, we can see large changes in the error for estimating employment for March 1992. In many cases these errors are much larger, while in others they are much smaller. The model based forecasts lack the stability expected. While the revisions still compare favorably with those for the published estimates in many divisions, we find the large increase in error for the services division particularly distressing.

Table 3. Mean prediction error April 1991 - June 1992

Division	LS		LS +sample		LTS		LTS +sample		Sample
	2 Part	1 Part	2 Part	1 Part	2 Part	1 Part	2 Part	1 Part	
Total	.41	1.02	.24	.24	.62	1.16	.17	.17	.62
Mining	.44	1.74	.48	.68	.50	1.81	.46	.63	.73
Constr.	1.39	1.71	1.06	1.15	2.61	3.83	1.13	1.14	.78
Mfg.	.43	1.75	.39	.36	.42	2.24	.31	.29	.48
TPU	1.70	2.21	.94	1.01	1.19	1.89	.55	.61	.39
Whlsale	.37	.72	.41	.50	.42	.77	.26	.31	.98
Retail	.81	.82	.74	.77	1.40	1.31	.76	.77	.75
FIRE	.11	.40	.28	.27	.19	.50	.15	.13	.38
Services	.71	.96	.43	.68	.95	1.27	.26	.45	1.08

Errors in tables 3 are 100X the difference between the estimated ratio and the "true" ratio. They are equivalent

to the percent error in estimating employment levels. The error associated with January 1991 estimates is not included in these averages.

Table 4. Mean Extrapolitive Squared Error, April 1991-March 1992					
Division	**LS**	**LTS**	**LS+ sample**	**LTS+ sample**	**Sample**
Total	1	2	1	1	12
Mining	2	2	3	3	20
Constr.	10	35	15	15	63
Mfg.	2	3	2	2	5
TPU	14	8	6	4	8
Whlsale	4	3	3	2	13
Retail	5	10	3	4	16
FIRE	1	1	2	2	12
Services	2	3	3	3	20

Extrapolitive sum of squares presented is a 12 month mean multiplied by 1,000,000

Table 5. Benchmark revision compared with the sample and published estimates, March 1992								
Division	**Model 9 months**			**Model 12 months**			**Samp**	**Adj Bmk**
	LS	**LS+S**	**LTS+S**	**LS**	**LS+S**	**LTS+S**		**Rev.**
Total	-.10	.39	.19	-.17	.56	.34	-.37	.15
Mining	-.67	-.30	.30	-.30	-.34	.54	-.25	-.31
Constr.	-1.13	-.55	-1.07	-.83	.81	-.26	1.27	1.58
Mfg.	.33	.73	.44	-.45	1.01	.81	.72	.72
TPU	.67	.93	-.15	2.00	1.03	.00	-.29	.39
Whlsale	-.07	.22	-.07	.45	.92	.41	-1.06	-.55
Retail	-.18	-.16	-.50	-.91	-.02	-.39	-.55	-.67
FIRE	.21	.59	.21	.24	.54	.10	.36	-.86
Services	-.15	.11	-.20	.59	.87	.21	-1.49	-.09

Revision is expressed as a percent of the benchmark level, 100 X (estimate - Benchmark) / Benchmark

CONCLUSIONS

In general, the two-equation models perform as well as or better than one equation models. It also appears that explanatory variables help the models. More work is required in model specification and data development. Given results for normality, we must either develop a means of identifying and adjusting for outliers or resort to a more robust form of the Kalman filter.

Contributing to the uncertainty in modeling employment and bias, the ES-202 employment series is not intended to be an economic time series. Our outlier search has indicated that the series is noisy. That program has been going through a period of continued improvements, and the impact of those changes is not well understood. These changes hide the true employment picture to some extent. The question remains as to how to better use this data in time series modeling. We may resort to modeling to the final benchmarked CES series.

REFERENCES

APTECH Systems Inc., GAUSS386i, Version 3.1.4, 1993.

Bureau of Labor Statistics Bulletin 2285, BLS Handbook of Methods (1988) Chapter 2, Employment, Hours and Earnings from the Establishment Survey, 13-27.

Coomes, Paul, "Forecasting the Present: Regional Employment by Industry," presented at the 1988 Annual Meeting of the Kentucky Economic Association, September 23, 1988.

Harvey, A. C., "A Unified View of Statistical Forecasting Procedures," Journal of Forecasting, Vol. 3, 1984, pp. 245-275.

Harvey, Andrew C., Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, 1989.

Neumark, David, and William L. Wascher, "Can We Improve Upon Preliminary Estimates of Payroll Employment Growth?," Journal of Business & Economic Statistics, Vol. 9, No. 2, 1991, pp. 197-205.

Sommers, John, and Stamas, George, "Correcting for Bias in Payroll Estimates", 1991 Proceedings of the Business and Economic Statistics Section, American Statistical Association, pp. 298-303.