ISSN 1725-4825

WORKING PAPERS AND STUDIES

2005 EDITION

Small area estimation with stochastic benchmark constraints: teory and practical application in US labor statistics

COPYRIGHT
Luxembourg: Office for Official Publications of the European Communities, 2005

ISSN 1725-4825
ISBN 92-79-01325-4
Cat. No. KS-DT-05-028-EN-N

EUROPEAN COMMISSION

eurostat

EURO INDICATORS

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server (http://europa.eu.int).

Luxembourg: Office for Official Publications of the European Communities, 2005

ISSN 1725-4825
ISBN 92-79-01325-4

Small Area Estimation with Stochastic Benchmark Constraints: Theory and Practical Application in US Labor Statistics

Danny Pfeffermann (msdanny@mscc.huji.ac.i), Department of Statistics, Hebrew University, Israel and University of Southampton, U.K. , Richard Tiller (tiller_r@bls.gov), and Sharon Brown (Brown.Sharon@bls.gov), Bureau of Labor Statistics, Washington, DC.

Workshop on Frontiers in Benchmarking Techniques and Their Application to Official Statistics

Luxembourg, 7-8 April 2005

Abstract

The Bureau of Labor Statistics uses state-space time series models to produce small area labor force estimates from the monthly Current Population Survey (CPS). For each area, a separate model is fitted that combines a classical components model for the unobservable population values with a model for the sampling errors. The use of the model produces estimators with much smaller variances than the survey estimates, but the models can be slow to adapt to external shocks to the economy if they occur towards the end of the series.

To protect the area models from this type of bias, we constrain each month the model estimates to add up to the aggregate direct CPS estimate in a set of areas for which the latter is sufficiently accurate. This is done by combining individual area models into a multivariate state-space model and adding the benchmark constraints to the joint observation equation.

Because the CPS design involves partial replacements of sample units each month, the survey errors are autocorrelated. The conventional approach to handling autocorrelated survey errors requires a very high order state-space system with large computational demands. We develop new algorithms for handling autocorrelated measurement errors with a lower order system. These algorithms produce concurrent and historical estimates that satisfy the benchmarks, along with their variances.

Our approach differs from conventional benchmarking. First, model dependent estimates are often benchmarked to external data that is only recorded at few time points, resulting in retrospective benchmarking with potentially large revisions to the published data. In our approach benchmarking takes place concurrently so that the estimates adapt to external shocks as they occur. Second, we use benchmarks that are not independent of the input data for the models (the sum of the area estimates in the present application). The empirical illustrations use CPS series, to illustrate benchmarking during 2001 when the U.S. economy experienced major shocks.

We describe the estimating approach used in the United States for developing official labor force estimates for States and large areas. Actual estimating issues associated with this approach and which led to the decision to conduct research on concurrent benchmarking are discussed. Empirical illustrations include the development of official labor force estimates for 2001 using the interim benchmarking approach implemented by BLS with official labor force estimates for January 2005.

1. Introduction

Among the important economic data developed by the United States Bureau of Labor Statistics (BLS), unemployment estimates for States and local areas are viewed as key indicators of local economic conditions. These estimates are produced by State workforce agencies under the Federal-State cooperative Local Area Unemployment Statistics (LAUS) program. Currently, monthly estimates of employment, unemployment, and the unemployment rate are prepared for more than 7,000 areas—regions, divisions, all States and the District of Columbia, metropolitan and small labor market areas, counties, cities of 25,000 population or more, and all cities and towns in New England regardless of population.[i] The LAUS estimates are used by a number of agencies in the United States to allocate more than $40 billion in Federal funds to States and areas for a variety of socioeconomic programs. State and local governments use the estimates for planning and budgetary purposes and as determinants of need for local services and programs. The LAUS estimates are one of the timeliest subnational economic measures, as the State labor force estimates are released by BLS five weeks after the reference week and just two weeks after the national estimates. In operating the LAUS program, the BLS is responsible for concepts and definitions, technical procedures, and review, analysis and publication of estimates. The State agencies are responsible for the production of the estimates and analysis and dissemination of the data to their own customers.

As the principal fact-finding agency for the Federal Government in the broad field of labor economics and statistics, BLS strives to ensure that its programs satisfy a number of criteria: relevance to social and economic issues, timeliness in reflecting today's rapidly changing economic conditions, accuracy and consistently high statistical quality, and impartiality. With estimates for January 2005, the LAUS program has completed a redesign that includes the introduction of real-time benchmarking in current estimation, an approach that is on the frontier of benchmarking methods and applications to official statistics. These improvements to LAUS methodology further the BLS mission to provide the best data possible on a timely basis.

1.1 Estimating employment and unemployment for States and large areas

A key element of the Bureau's approach to subnational labor force estimation is to ensure that these estimates are comparable to the official concepts and measures of the labor force used in the Current Population Survey (CPS). The CPS is the monthly survey of households that is designed to provide reliable monthly labor force estimates for the nation. To support reliability of subnational estimates, the CPS employs a State-based sample design. The State design constraint ensures that the survey sample in a State is large enough so that there is no more than an 8 percent Coefficient of Variation (CV) on the annual average level of unemployment when the unemployment rate is 6 percent. (For comparison, the national reliability standard is a 1.9 percent CV on the monthly level.)

A hierarchy of estimation methods is used to produce the 7,000 estimates covered by the LAUS program, based in large part on the availability and quality of data from the CPS. The strongest estimating method—signal-plus-noise univariate models for current estimation and annual average CPS benchmarks—has been employed for 48 States and the District of Columbia, and for four large areas—New York City, the Los Angeles metropolitan area, balance of New York State, and balance of California as far back as 1989. While not reliable enough to use directly, the monthly CPS values are integral to the signal-plus-noise estimation.

As the official source of labor force statistics, the CPS provides the State series with the benchmark measure for purposes of current estimation and periodic adjustment. In order to ensure comparability across States, the annual average employment and unemployment levels from the CPS were used as the benchmarks for the modeled LAUS estimates. (In subsequent discussion, the term "State" will be used to refer to all modeled areas and the term "previous" will signify the estimating approach used through 2004.)

In general, the previous method of model estimation resulted in an overestimate of employment and an underestimate of unemployment and the unemployment rate in States as compared to the national CPS estimates. . This is shown in the following charts which depict the differences in the LAUS sum-of-State and independent national CPS estimates from January 2000 to December 2004.

Chart 1. LAUS Sum-of-States minus CPS Employment, Not Seasonally Adjusted, January 2000 - December 2004

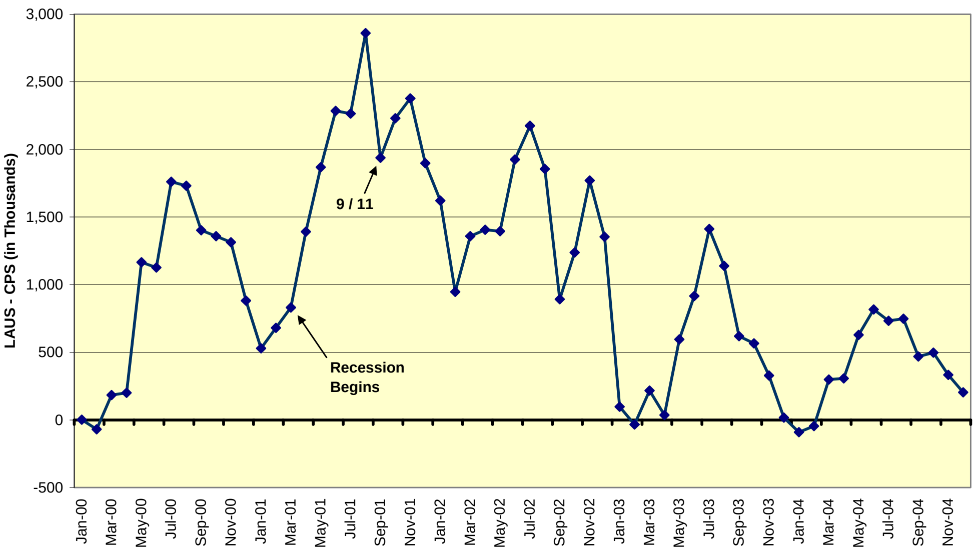


Chart 1 describes the relationship between the sum of State model-based estimates and the independent CPS-based estimate of employment for the nation. With the exception of four months, the LAUS model-based sum-of-State estimate was higher than the CPS national estimate. The overestimation reached its highest levels in 2001. During that year, the nation went into a recession starting in March and experienced the terrorist attacks of September. The State LAUS employment overestimation reached a peak of nearly 2.9 million in August 2001.

Chart 2. LAUS Sum-of-States minus CPS Unemployment, Not Seasonally Adjusted, January 2000 - December 2004

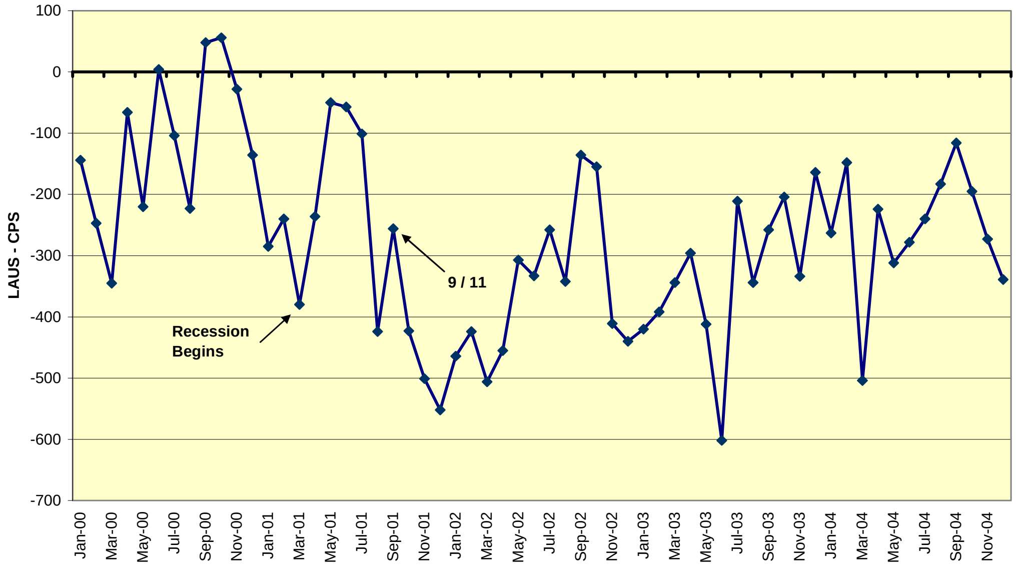


Chart 2 depicts the relationship between the sum of State model-based estimates of unemployment and the national estimate of unemployment. For this labor force measure, the model has, with the exception of three months, consistently underestimated unemployment relative to the national CPS measure. Consistent, large monthly underestimates of unemployment began in 2001. The average monthly difference was greatest in 2002 (-352,600), while the largest monthly difference (-602,000) occurred in June 2003.

Chart 3. LAUS Sum-of-States minus CPS Unemployment Rate, Not Seasonally Adjusted, January 2000 - December 2004

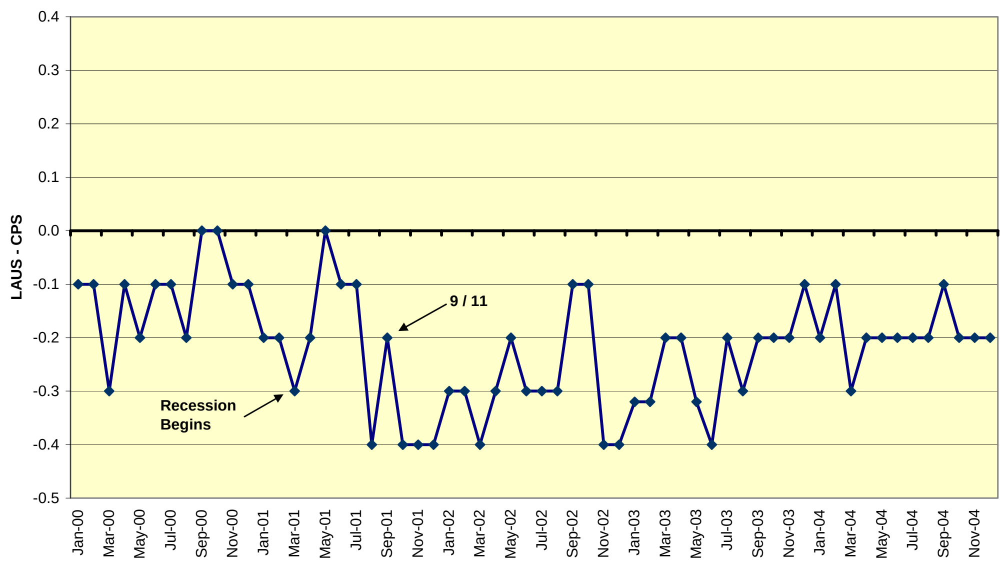


Chart 3 describes the relationship between the unemployment rate developed using the sum of State LAUS estimates and the CPS national measure. For nearly the entire period, the sum of State estimate falls below the independent national jobless rate. (In two months, the rates were identical.) While for many months, the difference is -0.2 percentage point or less, it is important to note that the direction of the difference is consistent. Starting in 2001, sum-of-States differences of -0.3 point and greater were reported with increasing frequency. In 2001, differences of -0.3 percentage point or more were recorded for five months, in 2002, for nine months, and in 2003, for five months.

To address the over- and underestimation associated with current model-based estimates, the model-based estimates of employment and unemployment were benchmarked to the respective annual average estimates from the CPS. However, the use of annual average State CPS benchmarks created other problems. It reintroduced sampling error into the series and resulted in significant end-of-year revisions in a large number of States, caused economic anomalies that are an artifact of the benchmarking approach, distorted seasonality in the previous year so that analysis is impaired, and often missed shocks to the economy. (A detailed discussion of these issues follows.)

BLS was provided with resources to improve the methods used to develop State and area LAUS estimates, including upgrading and enhancing the modeling approach, extending it to more areas, and incorporating decennial updates to procedures, data inputs, and geography. As part of this major LAUS Program Redesign, BLS implemented an innovative alternative to model benchmarking that is part of improved monthly model-based estimation. This alternative addresses longstanding issues related to accuracy and end-of-year revision, and also enhances the analytical capability of the estimates.

The new method of estimation ensures that State estimates add to the national estimates of employment and unemployment each month (real-time benchmarking). In doing so, the benchmark has been changed from annual average State-level estimates of employment and unemployment to monthly national estimates of these measures, and is part of current monthly estimation. In this way, economic changes will be reflected in the State estimates on a real-time basis, and end-of-year revisions will be significantly smaller.

1.2 Previous Modeling and Benchmarking Procedure

In 1989, time series models were first implemented in 39 small States and the District of Columbia for developing labor force estimates. In 1996, the time series approach to sample survey data was also extended to large States; thus, all States and the District of Columbia employed the time series methodology. The purpose of the approach was to reduce the high variability in monthly CPS estimates due to small sample sizes.

A signal-plus-noise univariate form of the model was used, with the monthly CPS sample estimate described as the sum of the true labor force value (signal) and sampling error (noise). Two models—one for the employment/population ratio and one for the unemployment rate—were developed for each State. In estimating the signal, the employment/population ratio model used the statewide monthly estimate of workers on nonfarm payrolls and intercensal population data, while the unemployment rate model used counts of unemployment insurance claimants who file for the CPS reference week and nonfarm payroll data. Each model had a trend, seasonal, and irregular component, as well as the regression component. An important feature of the model was the use of the Kalman filter to update regression coefficients and trend and seasonal terms when gradual structural changes occur. The signal term allowed the extraction of noise from the CPS time series data, thus providing a better estimate of the true value. The error term of the model reflected unique sampling error characteristics of the CPS, outliers, and irregular movements in the underlying true series. Seasonal adjustment was performed externally, with the application of X-11 ARIMA software to the unadjusted estimates.

Because of the potential for bias in the models and to ensure comparability in the estimates across all States, each year the monthly estimates of employment and unemployment were benchmarked to the respective CPS annual averages. (Also as part of annual benchmarking, the model inputs were revised as necessary, and the models were re-estimated and smoothed in an iterative process that allowed each observation to benefit from all observations in the series.) The primary external impetus for benchmarking to the CPS annual averages was to address the use of the estimates in distributing Federal funds. Beyond addressing this legislative use, benchmarking to the CPS was viewed as appropriate given the role of the CPS in providing the conceptual standard for the program.

The goal assigned to the statistical benchmarking procedure was twofold: (1) to ensure that the annual average of the final benchmarked series equaled the CPS annual average and (2) to preserve the monthly pattern of the model series as much as possible. In practice, the two goals were conflicting, and some changes to the monthly pattern were necessary to meet the first goal. The particular approach used is the Denton method.

The Denton method combined a constraint feature (relating to goal 1) and a feature that maintained the monthly pattern of the original series (goal 2). The specific routine sought to minimize the percent differences (squared) in the model/benchmarked series estimates from month-to-month. The method was used because of the overall modeling goal of accuracy of the month-to-month changes. The method was applied to three years in pairs of years, to minimize discontinuities within the benchmark period.

1.3 Issues with Previous Benchmark Procedures

An annual average CPS benchmark was employed in the LAUS program since 1974, and the Denton method of benchmarking since 1989. The Denton method is a mechanical procedure that does not take into account the properties of the time series models and ignores the survey error. As a result, no reliability measures were available for the benchmarked estimates.

While achieving the specific goals of ensuring comparability of estimates across States and addressing potential bias in the models, a number of methodological and analytical issues surfaced in the current estimation/benchmark procedures. These included reintroduction of sampling error to monthly estimates, discontinuities between December benchmarked and January model estimates, impaired comparability of data over the year, and inability to address, on a timely basis, “shocks” to the model such as the September 11 terrorist attacks and the onset of the economic recession.

<u>Reintroducing sampling error</u>

Despite the State-based sample design of the CPS, the State samples are fairly small (averaging about 950 households in small States and 2,200 in large States) and the resultant annual averages contain a significant degree of sampling error. On the other hand, the previous model did a very good job of removing error from the current CPS estimates. The noise component of the signal-plus-noise model is a sophisticated measure of the error in the CPS related to the unique aspects of the CPS sample design, as well as outliers and variance. Thus, the previous model estimate of the signal was viewed as a good estimate of the true labor force value. Because the variance of the model was less than the sampling error of the annual average CPS, by using the CPS annual average State employment and unemployment levels as the point benchmarks, the current method puts variability back into the monthly estimates.

The reliability criterion for the State CPS sample is an 8 percent or less Coefficient of Variation on the annual average level of unemployment when the unemployment rate is 6 percent. This relates to a 90 percent confidence interval of ± 0.8 percentage point on the annual average unemployment rate in a typical State. Each year, some number of States experienced significant benchmark revisions that were related to the random nature of sampling error.

The last time a retrospective annual average benchmark was used in the LAUS program was in 2004, when monthly estimates were benchmarked to 2003 State CPS estimates. For 2003, the benchmark revision for 10 States was 0.5 percentage point or more, with the maximum at 0.8 percentage point. (See table below.) Underscoring the random nature of the CPS variance and its reflection in the State benchmark revision, six of the States with large unemployment rate revisions to the 2003 CPS annual averages did not have significant revisions in the prior year. As long as the LAUS estimates were benchmarked to the CPS annual average, each year a small group of States were expected to experience large noneconomic revisions in the series.

Discontinuity between December benchmarked and January model estimates: the Endpoint Effect

Under the previous methodology, the previous year's December level—the endpoint of the benchmarking—reflected the adjustment to the CPS annual average and the sampling error that it contained, while the January estimate was model-based. December-January is a very seasonal period with predictable changes in employment in many States. Depending upon the size and direction of the employment benchmark revision in the State, the December-January employment change may not have reflected economic reality. Rather, it was an artifact of the benchmarking method. In the past, procedures were instituted that maintained the December-January model relationship for employment (the November endpoint), but they created serious distortion in the historical series.

Impaired analysis over the year

Regardless of whether the endpoint was moved to November (to preserve the December-January change) or kept at December, the ability to analyze over-the-year change in labor force series was compromised in a number of States each year. With a November endpoint, the difference between the annual average of the model series and the CPS was forced into eleven months, causing the series to rotate around August. This distortion in the series affected analysis of the labor force data over time. Even with the December endpoint, comparisons of modeled to benchmarked estimates provided spurious results, depending on the size of the benchmark revision in the State.

Addressing "shocks" to the series: Sum of States versus National Estimates

In the previous methodology, the State model estimates were developed independent of the national CPS. Although the monthly State CPS input data sum to the national measures, the sum of the State model estimates generally did not equal the national CPS estimates. To evaluate model performance, each month the sum of the State model estimates was compared to the national CPS estimates. Until 2001, the difference between the sum of State model estimates and the national CPS was well within sampling error of the national estimates. In 2001, significant deviations occurred in the sum of States versus national CPS measures in a number of months, specifically March, August, and October-December. Economic shocks to the national economy related to the onset of the recession and to the September 11 terrorist attacks occurred in these months. These economic shocks were not reflected in the State model estimates because the model viewed the increase in the State CPS unemployment in these periods as related to sampling error. Most evident was the post-September 11 period, exacerbating the economic recession, and continuing into 2002. The inability of the previous methodology to provide protection for economic shocks negatively impacted the use of the estimates in federal fund allocation and in labor market analysis.

1.4 Improved Method: Models with Real-time Benchmarking

As part of the LAUS redesign, the signal-plus-noise univariate models of the unemployment rate and the employment-population ratio were replaced with improved models. The new models are also signal-plus-noise models, where the signal is a bivariate model of the unemployment or employment levels. The unemployment insurance claims and nonfarm payroll employment inputs themselves are modeled, as well as their interaction with the appropriate CPS series. The resultant correlations provide important information for understanding and analyzing monthly model estimates. Seasonal, trend, and irregular components are developed for each modeled estimate. Seasonal adjustment occurs within the model structure through the removal of the seasonal component. The models produce reliability measures for the seasonally adjusted and not seasonally adjusted series, and on over-the-month change.

The new bivariate models incorporate a major change in the approach to benchmarking and the benchmarking process. Rather than continue with an annual average State benchmark applied retrospectively that reintroduces sampling error to the historical monthly estimates, the new approach uses a reliable real-time monthly national benchmark for controlling current State model estimates of employment and unemployment. In this process, benchmarking is part of the monthly State model estimation process, rather than a once-a-year retrospective adjustment.

Sections 2 and 3 provide the theoretical discussion of benchmarking by constrained estimation, while Section 4 describes empirical examples. The model-based approach discussed in these sections is still under development for practical program use. As an interim approach, external benchmark adjustment was evaluated and found to be adequate. Thus, with estimates for January 2005, the Bureau of Labor Statistics introduced real-time benchmarking as part of monthly State model estimation of employment and unemployment. Section 5 provides a discussion of the evaluation of the models with real-time benchmarking through pro rata adjustment and their implementation in the development of official State labor force estimates.

2. BLS Unobserved Component Models

Models have been developed independently for each of the CPS unemployment and employment estimates in all 50 States and the District of Columbia. The model employed by BLS is discussed in detail, including hyper-parameter estimation and model diagnostics in Tiller (1992). In this section we provide a brief description.

Let the direct survey estimator (CPS) for the s^{th} State series, $y_{s,t}$ be represented as the sum of two independent processes, the true population values, $Y_{s,t}$, and the errors, $e_{s,t}$, arising from sampling only a portion of the total population,

$$y_{s,t} = Y_{s,t} + e_{s,t} . \tag{1}$$

We combine a time series model of the population with a model of the survey error developed independently from the survey design. Each model is described below.

2.1 Model of the Area Population values

The model for $Y_{s,t}$ is specified as a classical time series decomposition where L_t, R_t, S_t and I_t represent respectively the trend level, slope, seasonal effects, and irregular component,

$$Y_{s,t} = L_{s,t} + S_{s,t} + I_{s,t} . \tag{2}$$

The model for each component is given below.

$$L_{s,t} = L_{s,t-1} + R_{s,t-1} + \eta_{L_s t}, \quad \eta_{L_s t} \sim NID\left(0, \sigma^2_{\eta_{L_s}}\right)$$

$$R_t = R_{t-1} + \eta_{L_s t}, \qquad \eta_{R_s t} \sim NID\left(0, \sigma^2_{\eta_{R_s}}\right)$$

$$S_{s,t} = \sum_{j=1}^{6} S_{s,j,t}$$

$$S_{s,j,t} = \cos\omega_j S_{s,j,t-1} + \sin\omega_j S^*_{s,j,t-1} + \nu_{s,j,t} \ ; \quad \nu_{s,j,t} \sim N(0, \sigma^2_s)$$

$$S^*_{s,j,t} = -\sin\omega_j S_{s,j,t-1} + \cos\omega_j S^*_{s,j,t-1} + \nu^*_{s,j,t} \ ; \ \nu^*_{s,j,t} \sim N(0, \sigma^2_s)$$

$$\omega_j = 2\pi, \quad j/12, \quad j = 1,\ldots,6$$

$$I_{s,t} = \nu_{I_s}, \quad \nu_{I_s} \sim NID\left(0, \sigma^2_{I_s}\right)$$

The model defined by (2) is known in the time series literature as the Basic Structural Model (BSM). The disturbances $\eta_{L_s t}, \eta_{R_s t}, \nu_{s,j,t}$, and ν_{I_s} are independent white noise series.

2.2 Model of the Survey Errors

The CPS survey's complex design induces heteroscedasticiy and autocorrelation in the survey errors,

$$e_{s,t} \sim N\left(0, \sigma^2_{s,t}\right); \quad E\left(e_{s,\tau} e_{s,t}\right) = \sigma_{s,\tau t} .$$

These errors are heteroscedastic because their variances depend on the population levels and change with redesigns. Autocorrelations at long lags are induced by the rotating panel design and sample replacement policy. Households selected to the sample are surveyed for 4 successive months, dropped from the sample for the next 8 months and then surveyed again for 4 more months. Households permanently dropped from the survey are replaced by households from the same ‘census tract’.

We model the variance-covariance structure of the CPS in linear stochastic form since it can be easily put into a state-space form described below. Later on we handle survey error differently within the context of benchmarking.

The model specified for the sampling error is AR(15) with a changing variance,

$$e_{s,t} = \gamma_{s,t} e^*_{s,t} \tag{3}$$

where,

$$e^*_{s,t} \simeq \sum_{i=1}^{15} \phi_{s,i} e^*_{s,t-i} + \nu_{e^*_s t}, \quad \gamma_{s,t} = \sigma_{e_s,t} \big/ \sigma_{e^*_s} \quad .$$

The AR(15) is used as an approximation to the sum of an MA(15) process and an AR(2) process. The MA(15) process accounts for the autocorrelations implied by the sample rotation scheme. The AR(2) process accounts for the autocorrelations arising from the replacement of households permanently dropped from the survey. The reduced ARMA representation of the sum of the two processes is ARMA(2,17).

The variances and covariances of the survey errors are provided by design information independent of the time series modeling of the population series.

2.3 State Space formulation

A general approach to optimally predicting the unobserved components of a time series corrupted by measurement error is to put the model into a state-space form and then apply the Kalman filter and smoother algorithms.

The state-space form consists of two equations:

$$\begin{aligned} & y_t = Z_t \alpha_t + e_t; \quad E(e_t) = 0, \quad E(e_\tau e_t{}') = \begin{cases} \Sigma_{tt} & \tau = t \\ 0 & \tau \neq t \end{cases} \\ & \alpha_t = T\alpha_{t-1} + \eta_t \ ; \quad \eta_t \sim NIID\left(0, Q\right) \end{aligned} \tag{4}$$

where Z_t and T are known non-stochastic matrices, and η_t a vector containing the white noise disturbances of the component models.

The first equation is the observation equation that represents the observed data as a linear combination of the unobserved components or state variables included in the vector α_t plus measurement error. The second equation is the transition equation that describes the evolution of the state vector, α_t , as a first-order vector autoregressive process

While the transition equation may appear to be restrictive, a surprisingly wide range of models can be transformed to an AR form by constructing artificial state variables. For all processes that can be given a state-space representation, the KF algorithm provides a simple unified approach to prediction and estimation.

For our application, the restriction that measurement error, $e_{s,t}$, be uncorrelated with its previous values requires modification of the observation equation. When measurement error is correlated, as is the case with CPS survey error, the usual practice is to remove it from the observation equation and add it to the state vector. This is possible provided the measurement error can be represented by a model that has a state-space form. Our CPS survey error model satisfies this requirement. Therefore, the observation equation in (5) below has no measurement error.

The separate models holding for the population values and the sampling errors for the s^{th} area are included in the transition equation. The resulting state vector, $\alpha_{s,t}$, contains the trend level, the slope, 11 seasonal components, the irregular term and the 15 lags of the sampling errors, a total of 29 elements referred to as the order of the state vector (q_s).

$$y_{s,t} = Z_{s,t}\alpha_{s,t}$$
$$\underbrace{\alpha_{s,t}}_{q_s \times 1} = T_s + \alpha_{s,t-1} + \eta_{s,t}, \quad E\left(\eta_{s,t}\eta_{s,t}{}'\right) = Q_s, \quad q_s = 29 \tag{5}$$

Usually we are interested in some linear combination of the state vector, such as the trend or non-seasonal component, commonly referred to as the “signal”, which we represent as,

$$y_{s,t,\mathbf{model}} = z_{s,t}\alpha_{s,t} \,. \tag{6}$$

Given the sample data ($y_{s,1}, \ldots, y_{s,n}$), the problem is to predict the state vector. The predictor, $\hat{\alpha}_{s,t|n}$, is the conditional expectation given the data with covariance matrix $P_{s,t|n}$,

$$\hat{\alpha}_{s,t|n} = E\left(\alpha_{s,t} \left| y_{s,n}\right.\right), \quad P_{s,t|n} = E\left[\left(\alpha_{s,t} - \hat{\alpha}_{s,t|n}\right)^2 \left| y_{s,n}\right.\right]. \tag{7}$$

E denotes the expectation operator and n indexes the latest period for which data are available. The value we predict may refer to the present ($t = n$), past ($n > t$) or future ($n < t$).

The predicted signal and variance are given by

$$\hat{y}_{s,t|n,\mathbf{model}} = z_{s,t}\,\hat{\alpha}_{s,t|n}, \quad Var\left(\hat{y}_{s,t|n,\mathbf{model}}\right) = z_{s,t}P_{s,t|n}z'_{s,t} \,. \tag{8}$$

These estimators have the property of Best Linear Unbiased Predictor (BLUP) given their respective information sets.

2.3 Filtering

Filtering is a real time process that predicts the signal and noise for the latest available observation and is implemented with the Kalman filter (KF) algorithm. For $t = n$, the most recent observation, the KF computes the predictor, $\hat{\alpha}_{s,t|t}$ and its covariance $P_{s,t|t}$, from just last period values $\hat{\alpha}_{s,t|t-1}$ and $P_{s,t|t-1}$ and the current observation y_t.

$$\begin{aligned} \hat{\alpha}_{s,t|t} &= T_s \hat{\alpha}_{s,t-1|t-1} + G_{s,t}\left(y_{s,t} - Z_{s,t} T_s \hat{\alpha}_{s,t-1|t-1}\right) \\ P_{s,t|t} &= \left(T_s P_{s,t|t-1} - G_{s,t} Z_{s,t} P_{s,t|t-1}\right) T_s' + Q_s \end{aligned} \tag{9}$$

where,

$$F_{s,t} = Z_{s,t} P_{s,t|t-1} Z_{s,t}', \quad G_{s,t} = T_s P_{s,t|t-1} Z_{s,t}'$$

2.4 Smoothing

Predicting past values of the state vector where n > *t* using all of the available data before and after *t* is referred to as "smoothing". While there are a number of different smoothing algorithms the most relevant one for our purposes is the fixed point smoother (Harvey, 1989). Given a fixed time point, *d*, we add $\alpha_{s,d}$ to the state vector at time $t > d$. This yields the augmented state vector,

$$\alpha_{s,t}^{d} = \left(\alpha_{s,t}', \alpha_{s,d}'\right)'$$

of order $q_s^d = 2q_s$ and the augmented state space system,

$$\begin{aligned} y_{s,t} &= Z_{s,t}^{d} \alpha_{s,t}^{d} \\ \alpha_{s,t}^{d} &= T_{s,t}^{d} \alpha_{s,t-1}^{d} + \eta_{s,t}^{d} \end{aligned} \tag{10}$$

where,

$$T_s^d = Diag\left(T_s, \underbrace{I}_{q_s \times q_s}\right), \quad Z_{s,t}^d = \left[Z_{s,t}, \quad \underbrace{0}_{s \times q_s}\right], \quad \eta_{s,t}^d = \begin{bmatrix} \eta_t \\ \underbrace{0}_{q_s \times 1} \end{bmatrix}, \quad V\left(\eta_t^d\right) = Q_s^d = Diag\left(Q_s, \underbrace{0}_{q_s \times q_s}\right)$$

Note the state vector $\alpha_{s,d}$ is fixed for $t > d$.

Applying the KF to the augmented model yields the smoothed estimator, $\hat{\alpha}_{s,d|n}$, which is the last q_s rows of $\hat{\alpha}_{s,n|n}^{d}$ and its variance $P_{s,d|n}$ which is the square submatrix of order q_s in the lower right hand corner of $P_{s,n|n}^{d}$. Iterating the process by varying *d* produces the smoothed estimator for all points in the observation set.

3. Benchmarking

While modeling survey data is a cost effective way of reducing the effects of small samples, it raises the question of how to protect against model breakdowns. Monitoring prediction errors in real time is a common practice for detecting model breakdowns, but even when large prediction errors are detected it is rarely possible to determine the appropriate model re-specification until additional data become available. Therefore, it is desirable to have a "built-in mechanism" to ensure the robustness of the estimators when the model fails to hold.

Our proposed solution is to modify the model dependent estimators such that they add to a weighted sum of the direct survey estimates at each time point. We do this with a joint model of a group of States subject to the linear constraint that the model dependent estimates at the area level, $\hat{y}_{s,t,\text{model}}$, add to a weighted sum over areas of the direct survey estimator, $\hat{y}_t^{bmk}$, as shown below.

$$\sum_{s=1}^{S} w_{s,t}\hat{y}_{s,t,\text{model}} = \sum_{s=1}^{S} w_{s,t} y_{s,t} \tag{11}$$

There are a number of options available for setting the weights. Given that our direct survey estimators are totals, we could set the weights to one. For numerical reasons we may scale the data to prevent computational difficulties from having very large variances for the aggregates. When the survey estimators are proportions the weights could be the area's relative labor force share.

The justification for incorporating the constraints is that the direct CPS estimators, $y_{s,t}$, which are unreliable in single a area, can be trusted when aggregated over areas. The use of the constraint provides real time protection against widespread external shocks. For example, if all the direct CPS estimates in the same group jointly increase or decrease due to some sudden external effect not accounted for by the model, the benchmarked estimators will reflect this change much faster than the model dependent estimators obtained by fitting the model separately in each of the States. This property is illustrated very strikingly in the empirical part of this paper.

In addition to protection from real time shocks, there are additional benefits to imposing the constraint. An important feature is the potential for reducing the variances of the model based area estimates. That is, the constraint provides a way of using information cross-sectionally to improve the time series models originally fit independently to the individual areas. While this type of interdependence is limited it does allow borrowing strength across areas without having to deal with the complexity of cross-sectional modeling. Moreover, even more complicated multivariate models would be vulnerable to external shocks. Of course, there is a risk of increasing variability when imposing the constraint on models that may be working well, but in this case the constraint should not lead to large adjustments.

Another useful feature of the constraint is that aggregation consistency is assured in the publication of sub national and national estimates. It is a common practice for statistical agencies to publish area estimates produced by indirect methods and the direct survey estimates at the national level. The former will not in general sum to the latter. Our constraint eliminates this publication issue.

To implement our proposed solution to benchmarking we combine the area models into one joint model and add the benchmark constraint to the joint observation equation. However filtering and smoothing with this constraint is not straightforward. The benchmark, a weighted average of the area survey estimates, is stochastic and correlated with its own past values and with current and past values of the area model predictions. In contrast, other model based approaches to benchmarking assume highly reliable external (independent) data exist to control the more noisy survey series. (See, Hillmer and Trabelsi, 1987 ; and Durbin and Quenneville, 1997 for benchmarking to external data sources in the context of state-space modeling.)

Doran (1992) considered constraints in a state space model when external data sources are available in the form of $R_t\tilde{\alpha}_t = r_t$ where the *r's* are non-stochastic. He shows that the ordinary KF and smoothing algorithms will satisfy the constraint. Our constraint, however, is stochastic. Pfeffermann and Burck (1990) consider the stochastic case and provide a modified Kalman filter that satisfies the constraint and adjusts the model variances to account for the errors in the constraint. Their approach, however, assumes the measurement errors are independent across-areas and over time.

We point out technical problems with the conventional filtering and smoothing algorithms and propose a new filtering algorithm designed to handle correlated measurement error. We then use this new algorithm to produce filtered and smoothed predictions that satisfy the constraints and corrects the estimated variances to reflect errors in the constraints.

3.1 Joint Model

The models for the survey estimates for *S* areas, $y_t = \left(y_{1t}, \ldots, y_S \right)'$ are combined into a joint model of the following form,

$$\underbrace{y_t}_{S\times 1} = Z_t \alpha_t; \quad \underbrace{\alpha_t}_{q\times 1} = T\alpha_{t-1} + \eta_t, \quad q = \sum_{s=1}^{S} q_s, \quad E\left(\eta_t \eta_t'\right) = Q = Diag\left(\sigma_{\eta_1}^2, \ldots, \sigma_{\eta_S}^2\right) \tag{12}$$

where

$$\alpha_t = (\alpha_{1t}'...\alpha_{St}')', \quad \eta_t = (\eta_{1t}'...\eta_{St}')', \quad Z_t = \mathrm{I}_S \otimes z_{st}, \quad T_t = \mathrm{I}_S \otimes T_{st}$$

and $\sigma_{\eta_s}^2$ is a vector of the white noise variance for the s^{th} area model.

Note *Q* is diagonal since we ignore cross-sectional correlations between corresponding components of the state vectors in different areas. The order of the joint model, *q*, is the sum of the orders of the area models.

We now want to constrain the predicted signals across areas to add to the benchmark value, y_t^{Bmk} for each *t*. We begin by augmenting the vector y_t by the benchmark y_t^{bmk} and adding a corresponding row vector to Z_t, $z_t^{bmk} = \left[w_{1t} z_{1t}, ..., w_{St} z_{St} \right]$.

$$\tilde{y}_t = \begin{bmatrix} y_t \\ y_t^{bmk} \end{bmatrix} = \begin{bmatrix} Z_t \\ z_t^{bmk} \end{bmatrix} \alpha_t \tag{13}$$

A major drawback to this approach is that the order of the state vector, $q = 29S$, becomes very large if there are more than a few areas. Also, we have almost 30 years of data for each series. In real time as a new observation becomes available we need quick execution. To produce historical estimates requires fixed point smoothing, which doubles the dimension of the state vector and requires processing all the observations.

Another problem is that we need to account for error in the benchmark. That is

$$y_t^{bmk} = Y_t + e_t^{bmk}$$

where,

$$Y_t = \sum\nolimits_{s=1}^{S} w_s Y_{s,t}, \quad e_t^{bmk} = \sum\nolimits_{s=1}^{S} w_s e_{s,t}$$

The most direct way to include a stochastic benchmark in the system is to add it to the error vector in the measurement equation. Of course, this violates the assumption of uncorrelated measurement error and introduces the further complication that the area survey errors in the transition equation are no longer independent of the benchmark.

Our approach is to deal directly with the problem of a state-space system with autocorrelated errors in the measurement equation. This allows us to reduce the size of the state vector by moving the area survey errors from the transition equation back to the observation equation. There is also another benefit which we discuss later. Since the order of the survey error model for each area is 15 this reduces the size of the state vector by one-half. Because the KF is no longer BLUP we develop a new filtering algorithm that accounts for correlated measurement errors and has desirable statistical properties. Our approach, therefore, requires solving three major problems:

1. develop filtering and smoothing algorithms for state-space models with correlated measurement errors,

2. incorporate the benchmark constraints and compute the benchmarked state estimates,

3. compute the variances of the benchmarked estimators.

3.2 GLS Filtering with Correlated Measurement Errors

We propose a filtering algorithm to predict state vector estimates conditional on just the state vector prediction updated from the previous time point and the most recent observation. The algorithm is based on a stochastic GLS result from Pfeffermann (1984). To illustrate this estimator we consider a single series and then later apply it to our benchmarking problem.

For a single series we have the following observation equation with survey error,
where now the state vector is of reduced order, $q_s^\dagger = q_s - 15$.

$$y_{s,t} = Z_{s,t}\alpha_{s,t}^\dagger + e_{s,t}; \quad E(e_{s,t}) = 0, \quad E(e_{s,\tau}e_{s,t}') = \Sigma_{s,\tau t} \tag{14}$$

We formulate our problem in terms of a random coefficient regression model of the form

$$\begin{pmatrix} T\hat{\alpha}_{s,t|t-1}^\dagger \\ y_{s,t} \end{pmatrix} = \begin{pmatrix} I_{q_s} \\ Z_{s,t} \end{pmatrix}\alpha_{s,t}^\dagger + \begin{pmatrix} u_{s,t|t-1} \\ e_{s,t} \end{pmatrix} \tag{15}$$

where

$$u_{s,t|t-1} = T\hat{\alpha}_{s,t|t-1}^\dagger - \alpha_{s,t|}^\dagger .$$

Let

$$y_{s,t}^* = \begin{pmatrix} T_s\hat{\alpha}_{s,t|t-1}^\dagger \\ y_{st} \end{pmatrix}; \quad X_{s,t}^* = \begin{pmatrix} I \\ Z_{s,t} \end{pmatrix}; \quad \upsilon_{s,t} = \begin{pmatrix} u_{s,t|t-1} \\ e_{s,t} \end{pmatrix}$$

then an unbiased predictor $\hat{\alpha}_{s,t|t}^\dagger$ of $\alpha_{s,t}^\dagger$, $E(\hat{\alpha}_{s,t|t}^\dagger - \alpha_{s,t}^\dagger) = 0$, based on $\hat{\alpha}_{s,t|t-1}^\dagger$ and $y_{s,t-1}$ is given by stochastic GLS (Pfeffermann, 1984)

$$\hat{\alpha}_{s,t|t}^\dagger = \left(X_{s,t}^{*}{}'V_{s,t}^{-1}X_{s,t}^*\right)^{-1} X_{s,t}^{*}{}'V_{s,t}^{-1}y_{s,t}^* \tag{16}$$

with variance

$$\mathrm{Var}\left(\hat{\alpha}^{\dagger}_{s,t|t}\right) = P_{s,t|t} = \left(X^{*}_{s,t}{}'V^{-1}_{s,t}X^{*}_{s,t}\right)^{-1}.$$

$V_{s,t}$ is the covariance matrix of the state prediction errors with the survey errors,

$$V_{s,t} = Var\begin{pmatrix} u_{s,t|t-1} \\ e_{s,t} \end{pmatrix} = \begin{pmatrix} P_{s,t|t-1} & C_{s,t} \\ C'_{s,t} & \Sigma_{s,tt} \end{pmatrix} \tag{17}$$

where

$$P_{s,t|t-1} = Var\left(\hat{\alpha}^{\dagger}_{s,t|t-1} - \alpha^{\dagger}_{s,t}\right); \quad \Sigma_{s,tt} = Var\left(e_{s,t}\right); \quad C_{s,t} = Cov\left[\left(\hat{\alpha}^{\dagger}_{s,t|t-1} - \alpha^{\dagger}_{s,t}\right), e_{s,t}\right].$$

Under classical filtering $C_{s,t}$ contains all zeroes. Non-zero values reflect the fact that predictions of the state vector from past data will depend on the current observation, y_t, because the latter contains an autocorrelated error component.

The way we implement this estimator is to modify the Kalman filtering recursions in (9) for time $t > 1$

$$P_{s,t} = P_{s,t|t-1} - G_{s,t-1}\left(Z_{s,t}P_{s,t|t-1} - C_{s,t}{}'\right)$$

where,

$$F_{s,t} = Z_{s,t}P_{s,t|t-1}Z'_{s,t} - Z_{s,t}C_{s,t} - C_{s,t}{}'Z_{s,t}{}' + \Sigma_{s,tt}, \tag{18}$$

$$G_{s,t-1} = \left(P_{s,t|t-1}Z_{s,t}{}' - C_{s,t}\right)F_{s,t}{}^{-1}$$

and add a parallel recursion for $C_{s,t}$,

$$C_{s,t} = \sum_{j=1}^{t-1} c_{s,jt}\Sigma_{s,jt}$$

where $C_{s,t}$ is a linear combination of the measurement error lag covariances. See Pfefferman and Tiller (2004) for computation of the coefficients $c_{s,jt}$ and the complete proofs of the following properties.

1. At each time point t, the GLS filter produces the BLUP of α_t based on the predictor $\hat{\alpha}_{s,t|t-1} = T_s\hat{\alpha}_{s,t-1}$ from the previous period and the new observation $y_{s,t}$.
2. Unlike the KF, which is BLUP for the class of predictors constructed from any linear combination of the past observations and $y_{s,t}$, when measurement errors are independent, the GLS filter is BLUP for a more restricted class of predictors based on a fixed linear combination of the past observations implicit in the one step ahead state predictor.
3. When the measurement errors are independent, the GLS filtering algorithm coincides with the familiar Kalman filter since $C_{s,t} = 0$ in equation. (see Harvey, 1989).

When measurement errors are autocorrelated and can not be represented by a specific time series model, it is impossible to construct a recursive algorithm that is BLUP for the entire set of observations. Even when the KF assumptions are appropriate, the loss in efficiency using the GLS instead of the KF in the reduced state space system appears to be small (see Pfeffermann and Tiller, 2004).

3.3 GLS Filtering with Benchmarking

Given that we have a filtering algorithm for correlated measurement error we now redefine the joint observation equation (12) as follows,

$$\tilde{y}_t = \tilde{Z}_t \underbrace{\tilde{\alpha}_t}_{\tilde{q}\times 1} + \tilde{e}_t; \quad \tilde{\alpha}_t = \left(\alpha^{\dagger}_{1,t}, \ldots, \alpha^{\dagger}_{S,t}\right), \quad \tilde{q} = \sum_{s=1}^{S} q^{\dagger}_s \tag{19}$$

where the observation vector contains the area survey estimators and the benchmark value as before, but now we include a measurement error vector containing the corresponding survey errors for each area series and for the benchmark.

$$\tilde{e}_t = (e_{1t} \ldots e_{St}, \sum\nolimits_{s=1}^{S} w_{st} e_{st})'$$

We now must account for a complex set of correlations in the measurement errors.

$$E(\tilde{e}_\tau \tilde{e}_t\,') = \tilde{\Sigma}_{\tau t} = \begin{bmatrix} \Sigma_{\tau t} & h_{\tau t} \\ h_{t\tau}' & \nu_{\tau t} \end{bmatrix} \tag{20}$$

where,

$$\Sigma_{\tau t} = Diag\left(\sigma_{1\tau t} \ldots \sigma_{S\tau t}\right);\ \sigma_{s\tau t} = Cov\left(e_{s\tau}, e_{st}\right)$$

$$\nu_{\tau t} = \sum\nolimits_{s=1}^{S} w_{s\tau} w_{st} \sigma_{s\tau t} = Cov\left(\sum\nolimits_{s=1}^{S} w_{s\tau} e_{s\tau}, \sum\nolimits_{s=1}^{S} w_{st} e_{st}\right)$$

$$h_{\tau t} = (h_{1\tau t} \ldots h_{S\tau t})';\ h_{s\tau t} = w_{st} \sigma_{s\tau t} = Cov\left(e_{s\tau}, \sum\nolimits_{s=1}^{S} w_{st} e_{st}\right)$$

On the diagonal of $\tilde{\Sigma}_{\tau t}$ are the within area lag autocovariances, $\Sigma_{\tau t}$ and the benchmark covariance, $\nu_{\tau t}$. Since the State samples are drawn independently, $\Sigma_{\tau t}$ is diagonal. The off diagonals, $h_{\tau t}$ and its transpose, are the correlations between the area survey errors and the error in the benchmark (a weighted sum of the S area survey errors).

While our aggregate benchmark contains error we nevertheless want the weighted sum of the area model predictions to exactly match the benchmark value. That is we have a binding constraint,

$$\sum\nolimits_s w_s \hat{y}_{s,t,\mathbf{model}} = \sum\nolimits_s w_s y_{s,t} = y_t^{bmk}.$$

Under the true model , $y_{s,t} = z_{s,t}\alpha_{s,t} + e_{s,t}$ and the benchmark can be written as

$$y_t^{bmk} = Y_t + e_t^{bmk}$$

where

$$Y_t = \sum_s w_s z_{s,t} \alpha_{s,t}.$$

Therefore, by requiring the sum of the model estimates to exactly match y_t^{bmk} we are ignoring e_t^{bmk} which is equivalently to setting the variance of the benchmark error and its covariance with the area survey errors to zero,

$$Var\left(\sum_{s=1}^{S} w_{s,t} e_{s,t}\right) = Cov\left(e_{s,t}, \sum_{s=1}^{S} w_{s,t} e_{s,t}\right) = 0\,. \tag{21}$$

This is implemented in the GLS algorithm (3) by replacing the variance matrix $\tilde{\Sigma}_{tt}$ of the observation equation by the pseudo variance matrix,

$$\tilde{\Sigma}_{tt}^{*} = \begin{bmatrix} \Sigma_{tt} & , & 0_{(S)} \\ 0_{(S)}\,' & , & 0 \end{bmatrix} \tag{22}$$

and setting the last column of C_t^{bmk} matrix to zero,

$$C_t^{bmk} = Cov\left(\tilde{u}_{t|t-1}, \tilde{e}_t\right), \quad \tilde{u}_{t|t-1} = \tilde{T}\tilde{\alpha}_{t-1}^{bmk} - \alpha_t \tag{23}$$

which gives

$$C_{t,0}^{bmk} = Cov\left(\tilde{u}_{t|t-1}, \tilde{e}_{t,0}\right), \quad \tilde{e}_{t,0} = \left(e_{1,t}, \ldots, e_{st}, 0\right). \tag{24}$$

Application of the algorithm with the imposed benchmark constraints yields the benchmarked predictor, $\tilde{\alpha}_{t|t}^{bmk}$

$$\begin{aligned} \tilde{\alpha}_t^{bmk} &= \tilde{\alpha}_{t|t-1}^{bmk} + \tilde{G}_t\left(\tilde{y}_t - \hat{Z}_t\tilde{\alpha}_{t|t-1}^{bmk}\right) \\ P_{t|t}^{*} &= P_{t|t-1}^{bmk} - \tilde{G}_t\left[\tilde{Z}_t P_{t|t-1}^{bmk} - C_{t,0}^{bmk\prime}\right] \end{aligned} \tag{25}$$

where

$$\begin{aligned} \tilde{F}_t &= \tilde{Z}_t P_{t|t-1}^{bmk}\tilde{Z}_t\,' - \tilde{Z}_t C_{t,0}^{bmk} - C_{t,0}^{bmk\prime}\tilde{Z}_t\,' + \Sigma_{tt}^{*}, \\ \tilde{G}_t &= \left(P_{t|t-1}^{bmk}\tilde{Z}_t' - C_{t,0}^{bmk}\right)\tilde{F}_t^{-1} \end{aligned}$$

The system of equations form a pseudo model in the sense that the variance covariance matrix of the state vector, $P_{t|t}^{*}$ fails to account for the errors in the benchmark and its correlations with the area survey errors.

Having implemented the benchmark as an exact linear constraint in the observation equation, we now wish to take account of the stochastic nature of the benchmark when computing the V/C matrix for the predicted state vector. Following Pfeffermann and Burck (1990), we modify the recursion for updating the V/C matrix as follows,

$$Var(\tilde{\alpha}_t^{bmk} - \tilde{\alpha}_t) = g_t P_{t|t-1}^{bmk} g_t' + h_t\tilde{\Sigma}_{tt}h_t' + g_t C_t^{bmk} h_t' + h_t C_t^{bmk\prime} g_t' \tag{26}$$

where,

$$\tilde{\alpha}_{t|t}^{bmk} - \tilde{\alpha}_t = g_t\left(\tilde{\alpha}_{t|t-1}^{bmk} - \tilde{\alpha}_t\right) + h_t\tilde{e}_t$$

$$g_t = P_{t|t}^{*}B_{1t}, \quad h_t = P_{t|t}^{*}B_{2t}, \quad B_t = X_t^{*\prime}V_t^{-1} = \left(\underbrace{B_{1t}}_{q\times q}, \underbrace{B_{2t}}_{q\times S+1}\right) \text{ see (15).}$$

3.4 GLS Smoothing

Clearly, as new data accumulate it is desirable to modify past predictors produced by the filtering algorithm, which is particularly useful for trend estimation. To predict the state vector at each time point using all the available data we develop a GLS version of the fixed point smoother as described in (10). This requires doubling the order of the system. Even though this type of smoother has smaller memory requirements than other smoothing methods it is computationally demanding. Fortunately, smoothing is not executed in real time. In practice, BLS performs smoothing at the end of each year to update the latest filtering results.

We augment the state vector with α_d, as in (10), and add a second constraint to the measurement equation to compute the fixed point benchmarked state predictor, $\tilde{\alpha}_d^{bmk}$. Specifically we add the benchmark value at time d to $\tilde{y}_t$, an extra row to $\tilde{Z}_t$ and the corresponding benchmark error to the measurement error vector, $\tilde{e}_t$.

$$
\begin{aligned}
&\underbrace{\tilde{y}_t^d}_{(S+2)\times 1} = \tilde{Z}_t^d \tilde{\alpha}_t^d + \tilde{e}_t^d; \quad E\left(\tilde{e}_t^d\right) = 0, \quad E\left(\tilde{e}_\tau^d \tilde{e}_t^{d\prime}\right) = \tilde{\Sigma}_{\tau t}^d, \quad t, \tau > d \\
&\underbrace{\tilde{\alpha}_t^d}_{2q^\dagger \times 1} = T^d \tilde{\alpha}_{t-1}^d + \tilde{\eta}_t^d; \quad E\left(\tilde{\eta}_t^d \tilde{\eta}_t^{d\prime}\right) = \tilde{Q}^d = Diag\left(\tilde{Q}^d, \underbrace{0}_{q\times q}\right) \\
&\tilde{y}_t^d = \left(\tilde{y}_t', y_d^{bmk\prime}\right)'; \ \tilde{e}_t^d = \left(\tilde{e}_t', e_d^{bmk\prime}\right)', \ \ \tilde{Z}_t^d = Diag\left[\tilde{Z}_t, \left(w_{1t}z_{1t}, \ldots, w_{St}z_{St}\right)\right]
\end{aligned}
\tag{27}
$$

The V/C matrix for $\tilde{e}_t$ now contains an additional row and column to account for the variance of y_d^{bmk} and its covariances with the other survey errors and the benchmark at time t,

$$
\tilde{\Sigma}_{\tau t}^d = \begin{bmatrix} \tilde{\Sigma}_{tt} & & k_{\tau d} \\ & & v_{\tau d} \\ k_{\tau d}' & v_{d\tau} & v_{dd} \end{bmatrix}
$$

where,

$$
\begin{aligned}
&\underbrace{k_{\tau,d}}_{S\times 1} = \left[k_{1,\tau,d} \cdots k_{S,\tau,d}\right], \quad k_{s,d,\tau} = Cov\left(e_{s,\tau}, e_d^b\right) = Cov\left(e_{s,\tau}, \sum_s e_{s,d}\right) \\
&v_{\tau d} = Cov\left(e_\tau^b e_d^b\right) = \mathrm{Cov}\left(\sum_s w_s e_{s,\tau}, \sum_s w_s e_{s,d}\right) \\
&v_{dd} = Var\left(e_d^b\right) = \mathrm{Cov}\left(\sum_s w_s e_{sd}, \sum_s w_s e_{sd}\right).
\end{aligned}
$$

Applying the GLS filter to this augmented model yields the GLS smoothed predictor, $\tilde{\alpha}_d^{bmk}$, which is the last q rows of $\tilde{\alpha}_t^{d,bmk}$, computed in the recursions below.

$$\underbrace{\tilde{\alpha}_t^{d,bmk}}_{2q\times 1} = \tilde{\alpha}_{t|t-1}^{d,bmk} + \tilde{G}_t^d\left(\tilde{y}_t^d - \hat{Z}_t^d\tilde{\alpha}_{t|t-1}^{d,bmk}\right)$$

$$P_{t|t}^{d,*} = P_{t|t-1}^{d,bmk} - \tilde{G}_t^d\left[\tilde{Z}_t^d P_{t|t-1}^{d,bmk} - C_{t,0}^{d,bmk\prime}\right] \qquad (28)$$

where,

$$\tilde{F}_t^d = \tilde{Z}_t^d P_{t|t-1}^{bmk}\tilde{Z}_t^{d\prime} - \tilde{Z}_t^d C_{t,0}^{d,bmk} - C_{t,0}^{d,bmk\prime}\tilde{Z}_t^{d\prime} + \Sigma_{tt}^{d,*}$$

$$\tilde{G}_t^d = \left(P_{t|t-1}^{d,bmk}\tilde{Z}_t^{d\prime} - C_{t,0}^{d,bmk}\right)\left(\tilde{F}_t^d\right)^{-1}$$

We have two binding constraints which we enforce by setting the variances and covariances related to the two benchmark values to zeroes.

$$\underbrace{\Sigma_{tt}^{d,*}}_{(S+2)\times(S+2)} = Diag\left(\tilde{\Sigma}_{tt}, 0, 0\right)$$

$$C_{t,0}^{d,bmk} = Cov\left(\tilde{u}_{t|t-1}^d, \tilde{e}_{t,0}^d\right), \quad \tilde{e}_{t,0}^d = \left(e_{1,t}, \ldots, e_{st}, 0, 0\right)$$

Finally, we adjust the pseudo V/C matrix, $P_{t|t}^{d,*}$ computed above to account for the benchmark errors in the same way as (26). For more details consult Pfeffermann and Tiller (2005).

4. Empirical Illustrations for Model Based Benchmarking

The empirical results presented below refer to the monthly CPS series of unemployment in the 9 Census divisions of the U.S.A. for the period January 1998-December 2003. The benchmark constraints are defined such that the model dependent predictors of the Divisions' unemployment are benchmarked to the total national unemployment. The CV of the CPS estimator of the total national unemployment is 2%, which is viewed as sufficiently precise. The year 2001 is of special interest for illustration since it is affected by the start of a recession in March and the attack on the World Trade Center in September. These two events provide an excellent test for the performance of the proposed benchmarking procedure. Figure 1 shows the direct CPS estimators, the not benchmarked predictors and the benchmarked predictors for the South Atlantic division. Similar pictures have been obtained for the other divisions. As is evident from the graph, in 2001 there is a systematic model underestimation, which is corrected by the use of the benchmark constraints. The adjustments in other years are relatively small.

In order to illustrate this point further, we show in Table 1 the means and STDs of the monthly ratios between the benchmarked predictor and the not benchmarked predictor for each of the 9 census divisions, separately for 1998-2003 excluding 2001, and for only 2001. As can be seen, in 2001 some of the mean ratios are about 4% but in the other years the means never exceed 1%, showing that in normal times benchmarking has a relatively small affect on the model dependent predictors.

We mentioned earlier that by imposing the benchmark constraints the predictor in any given area "borrows strength" from other areas. Figure 2 shows the standard deviations (STD) of the not benchmarked predictors, the benchmarked predictors and the direct CPS estimators for the same census division. Again similar pictures are obtained for the other divisions; see Pfeffermann and Tiller (2004). As can be seen, the STDs of the benchmarked predictors are somewhat lower than the STDs of the not benchmarked predictors for all the months and both sets of STDs are much lower than the STDs of the CPS estimators. The two sets of STDs are further compared in Table 2 based on the years 1998-2003. Table 2 shows gains in efficiency of up to 15% in some of the divisions by use of benchmarking. Notice that the ratios are very stable throughout the years despite the fact that the STDs of both sets of model predictors change between months due to changes in the STDs of the sampling errors.

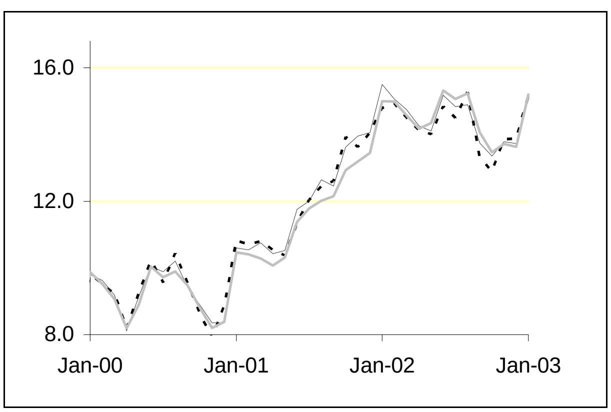


Figure 1. (Total Unemployment,
------ CPS ——— BMK ——— Not BMK
South Atlantic (100,000))

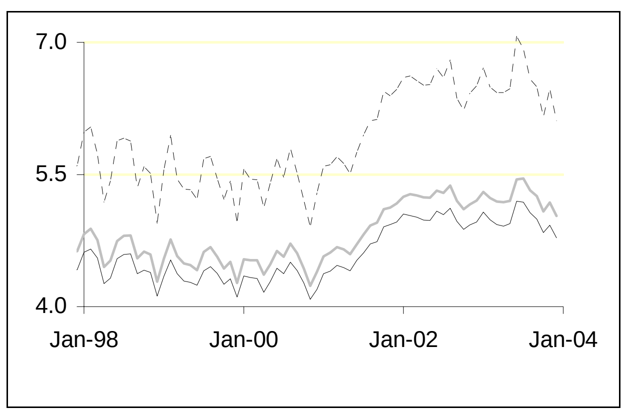


Figure 2. (STDs Total Unemployment,
------ CPS ——— BMK ——— Not BMK
South Atlantic (10,000))

Table 1. Means and STDs (in parentheses) of Ratios Between Benchmarked and Not benchmarked Predictors of Total Unemployment in Census Divisions

Division	1998-2000, 2002-2003	2001
New England	1.01 (.016)	1.03 (.017)
Middle Atlantic	1.00 (.014)	1.03 (.019)
East North Central	1.00 (.013)	1.03 (.013)
West North Central	1.01 (.014)	1.03 (.011)
South Atlantic	1.00 (.016)	1.04 (.016)
East South Central	1.01 (.016)	1.03 (.014)
West South Central	1.00 (.014)	1.04 (.022)
Mountain	1.00 (.011)	1.02 (.011)
Pacific	1.00 (.017)	1.04 (.020)

Table 2. Means and STDs (in parentheses) of Ratios between STDs of Benchmarked and Unbenchmarked Predictors of Total Unemployment in Census Divisions.

Division	Means (STDs)
New England	0.85 (.013)
Middle Atlantic	0.94 (.005)
East North Central	0.96 (.004)
West North Central	0.89 (.008)
South Atlantic	0.96 (.004)
East South Central	0.86 (.007)
West South Central	0.92 (.006)
Mountain	0.88 (.009)
Pacific	0.96 (.004)

5. Implementation of Real-time Benchmarking in US Labor Statistics

In 2005, time series models with real-time benchmarking to monthly national estimates of employment and unemployment was introduced as the official method of developing statistics for States and selected large areas. Under real-time benchmarking, a tiered approach to estimation is used. Model-based estimates are developed for the nine Census divisions that geographically exhaust the United States using univariate signal-plus-noise models. (Census division groupings have long been used to analyze and publish LAUS estimates.) The division models are similar to the State models, but do not use unemployment insurance claims or nonfarm payroll employment as variables. This allows division models to be developed in a very timely manner without sacrificing reliability. The division estimates are benchmarked to the national levels of employment and unemployment on a monthly basis. The benchmarked division model estimate is then used as the monthly benchmark for the States within the division.

The proposed approach for adjusting the State model estimates to the division model is by constrained estimation (described in Section 3). Under this approach, a constraint is imposed on the estimates such that the State estimates must equal the division estimates. The model-based approach imposes a constraint in the estimation process so that the models produce estimates that satisfy the constraint. (Section 4 describes the tests using division models.) This approach is still under development for practical program use. As an interim approach, external benchmark adjustment was evaluated and found to be adequate. Thus, the new official method uses the distribution of State model estimates within the division as the basis for the monthly benchmark adjustment to the division-level estimates. In the interim pro rata approach, the relative shares of each State's model estimate to its division total are preserved by the monthly benchmark adjustment, and the absolute size of the adjustment to a State's monthly model estimates is directly related to the size of the model estimate. Supporting this, the monthly State model estimation will discount extreme monthly CPS values and therefore avoid affecting the monthly benchmarking adjustment.

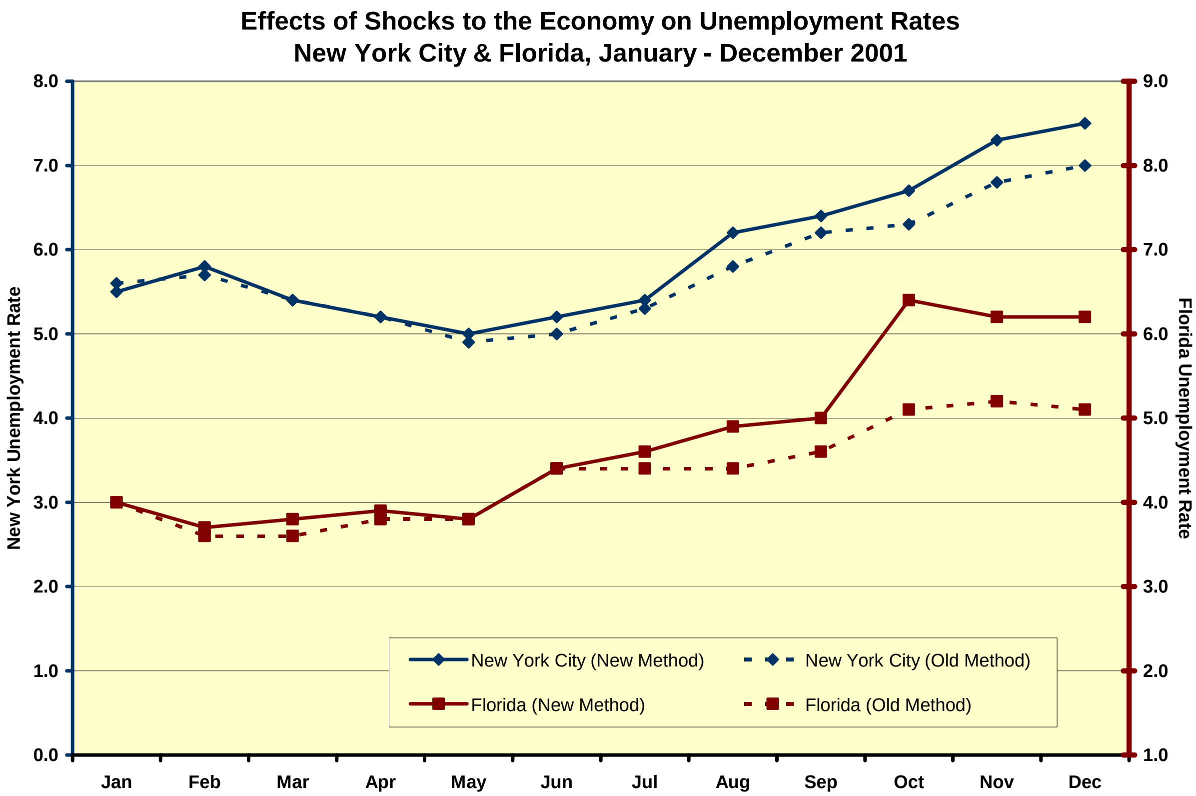
5.1 Evaluation of Real-time Benchmarking

In order to determine whether real-time benchmarking should be used in the official methodology for developing employment and unemployment estimates for States and large areas, comparison and evaluation of the estimates from the new models with real-time benchmarking and the previous models with the retrospective State annual average benchmark was undertaken. In addition, the performance of the proposed method was evaluated in real-time during a one-year dual estimation period.

The evaluation of the new models versus the previous method focused on the 2000-2004 period. By design, the monthly discrepancy between the sum-of-States estimate and the national estimate of employment and unemployment was eliminated. Also by design, December-January discontinuities and distortions in prior year estimates associated with the previous method were removed.

The comparison of methods for 2001 was particularly important. The events in 2001—the onset of the recession and the September terrorist attacks—underscored the inability of the previous method to reflect the labor market impact of these events on a timely basis and led to the consideration of alternatives to the annual benchmark. Therefore, the comparison of methods for 2001 labor force estimation provides actual examples of how real-time benchmarking addresses shocks to the economy.

For the first six months of 2001, the difference between the national unemployment rate and the rate developed by aggregating State estimates was small, reaching 0.3 percentage point in March. In the latter half of the year, as the recession began to impact unemployment and the Nation reacted to the events of September 11, a different picture emerged. Sum-of-States to the Nation unemployment rate differences reached 0.4 percentage point in August and December, with the October and November differences at 0.3 percentage point. Of course, the new models with real-time benchmarking preclude such differences. Of great interest is the performance of the new method in New York City, one of the sites of the terrorist attacks, and Florida, one of the major locations for vacations in the United States.



In the LAUS program, estimates for New York City are developed in the same way as State estimates. For the first four months of 2001, the unemployment rate for New York City using the new method was virtually the same as the previous method. The onset of the recession, combined with the impact of terrorist attack, was recorded in the new method, as the higher rates were registered using real-time benchmarking, with the difference reaching 0.5 percentage point in November and December.

Jobless rate comparisons for Florida are similar, but more extreme. Differences in jobless rates developed using the two approaches were noted in the last half of 2001 when, in reaction to the terrorist attacks, vacation travel declined. By October, the new model unemployment rate was 1.3 percentage points higher than the previous method, and differences remained at 1.0+ percentage points through the end of the year.

The impact of the recession on the unemployment rates of States with significant manufacturing employment was also better measured by the new method with real-time benchmarking. Such States include Michigan and North Carolina.

As part of evaluation, a Dual Estimation Period (DEP) was conducted from February to December 2004 so that proposed methodology and operational systems could be reviewed in a real-time environment and the impact on estimation evaluated. During the DEP period, it was clear that the new models were addressing issues of the previous models with a retrospective State benchmark. In brief, the new models with real-time benchmarking produced somewhat higher estimates of unemployment and the rate and lower estimates of employment, addressing the consistent under- and over-estimation described in Section 1, Charts 1-3. A slight increase in month-to-month volatility was seen in the unemployment series. Operating systems and estimation processes were also successfully tested in a real-time environment.

5.2 Implementation in official statistics

Based on the statistical and empirical evaluation of the estimates developed using models with real-time benchmarking and our experiences during the 2004 DEP, the BLS determined that the new method of developing employment and unemployment at the State level was more accurate than the previous procedure and would be implemented with estimates for January 2005. The full historical series for the States going back to 1976 was replaced with estimates developed using the new models and real-time benchmarking to the national estimates.

Annual historical benchmarking will still continue for State estimates but would be greatly altered. The updating of model inputs, model re-estimation, and incorporation of updated independent population controls will be performed each year, as well as adjustment of the revised State model estimates to the national CPS employment and unemployment levels each month. However, the impact on the historical series of these benchmark activities is considered to be fairly small, especially in comparison with annual revisions using the previous methodology.

The introduction of bivariate models with real-time benchmarking is viewed as one of the most significant methodological changes to be introduced in the LAUS program. The new estimation approach will ensure additivity of the State estimates to national estimates on a monthly basis, thus addressing the timely reflection of economic events and reducing the expected size of the annual revision to the series. For the first time, estimated standard errors for seasonally adjusted estimates will be provided, in addition to reliability measures for the not seasonally adjusted series and on over-the-month change.

References

Doran, H. E. (1992), “Constraining Kalman Filter and Smoothing Estimates to Satisfy Time-Varying Restrictions,” *The Review of Economics and Statistics*, 74, 568-572.

Durbin, J., and Quenneville, B. (1997), “Benchmarking by State Space Models,” *International Statistical Review*, 65, 23-48.

Harvey, A. C. (1989), Forecasting Structural Time Series with the Kalman Filter. *Cambridge: Cambridge University Press*.

Hillmer, S. C., and Trabelsi, A. (1987), “Benchmarking of Economic Time Series,” *Journal of the American Statistical Association,* 82, 1064-1071.

Pfeffermann, D. (1984), “On Extensions of the Gauss-Markov Theorem to the Case of Stochastic Regression Coefficients. *Journal of the Royal Statistical Society*, Ser. B, 46, 139-148.

Pfeffermann, D. and Burck L.,(1990), “Robust Small Area Estimation Combining Time Series and Cross-Sectional Data”, *Survey Methodology*, Vol. 16, No.2, pp. 217-237.

Pfeffermann, D. and Tiller, R. L.,(2004), “Small Area Estimation with Autocorrelated Observations and Stochastic Benchmark Constraints”, presented at the Joint Statistical Meetings, Toronto, August 8-12, 2004.

Pfeffermann, D. and Tiller, R. L.,(2005), “Model-Based Seasonal Adjustment of Survey Series Subject to Benchmark Constraints with a State-Space Smoothing Algorithm”, to be presented at the Joint Statistical Meetings, Minneapolis, August 7-11, 2005.

Tiller, R. B. (1992), „Time Series Modeling of Sample Survey Data from the U.S. Current Population Survey,” *Journal of Official Statistics*, 8, 149-166.