Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

Small-Area Estimation With State–Space Models Subject to Benchmark Constraints

Danny Pfeffermann & Richard Tiller

To cite this article: Danny Pfeffermann & Richard Tiller (2006) Small-Area Estimation With State–Space Models Subject to Benchmark Constraints, Journal of the American Statistical Association, 101:476, 1387-1397, DOI: 10.1198/016214506000000591

To link to this article: http://dx.doi.org/10.1198/016214506000000591

Published online: 01 Jan 2012.

Submit your article to this journal

Article views: 144

View related articles

Citing articles: 20 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=uasa20

Download by: [Board of Governors of the Federal Reserve System] **Date:** 23 October 2017, At: 10:31

Small-Area Estimation With State–Space Models Subject to Benchmark Constraints

Danny PFEFFERMANN and Richard TILLER

This article shows how to benchmark small-area estimators, produced by fitting separate state–space models within the areas, to aggregates of the survey direct estimators within a group of areas. State–space models are used by the U.S. Bureau of Labor Statistics (BLS) for the production of all of the monthly employment and unemployment estimates in census divisions and the states. Computation of the benchmarked estimators and their variances is accomplished by incorporating the benchmark constraints within a joint model for the direct estimators in the different areas, which requires the development of a new filtering algorithm for state–space models with correlated measurement errors. The new filter coincides with the familiar Kalman filter when the measurement errors are uncorrelated. The properties and implications of the use of the benchmarked estimators are discussed and illustrated using BLS unemployment series. The problem of small-area estimation is how to produce reliable estimates of area (domain) characteristics and compute their variances when the sample sizes within the areas are too small to warrant the use of traditional direct survey estimates. This problem is commonly handled by borrowing strength from either neighboring areas and/or from previous surveys, using appropriate cross-sectional/time series models. To protect against possible model breakdowns and for consistency in publication, the area model–dependent estimates often must be benchmarked to an estimate for a group of the areas, which it is sufficiently accurate. The latter estimate is a weighted sum of the direct survey estimates in the various areas, so that the benchmarking process defines another way of borrowing strength across the areas.

KEY WORDS: Autocorrelated measurement errors; Generalized least squares; Recursive filtering; Sampling errors; Unemployment.

Downloaded by [Board of Governors of the Federal Reserve System] at 10:31 23 October 2017

1. INTRODUCTION

The U.S. Bureau of Labor Statistics (BLS) uses state–space time series models for the production of the monthly employment and unemployment estimates in the 9 census divisions (CDs) and the 50 states and the District of Columbia that make them up. The models are fitted to the direct sample estimates obtained from the Current Population Survey (CPS). Using models is necessary because both the state CPS samples and the CD samples are too small to produce reliable direct estimates, which is a typical "small-area estimation" problem. The coefficient of variation (CV) of the direct unemployment estimates varies from about 4.5% to 8.5% for the CDs and from about 8% to about 16% for the states. The use of time series models significantly reduces the variances of the estimators by borrowing information from past estimates; see the application in Section 5. (For a recent review of small-area estimation methods, see Pfeffermann 2002. Section 6 considers the use of time series models.) [Rao (2003) provides a detailed methodological account of this topic.]

The state–space models are fitted independently for each CD and state, and combine a model for the true population values with a model for the sampling errors. The direct survey estimates are the sums of these two unknown components. The published estimates are the differences between the direct estimates and the estimates of the sampling errors, as obtained under the combined model. At the end of each calendar year, the monthly model-dependent estimates are modified so that their annual mean equals the corresponding mean of the direct CPS estimates. The purpose of the benchmarking is to provide protection against possible model failure. However, this benchmarking procedure has two major shortcomings:

1. The mean annual CPS estimates are still unreliable, particularly for states, because the monthly estimates are highly correlated due to large-sample overlaps between different months (see Sec. 2).
2. The benchmarking is retrospective**,** occurring at the end of each year after that the monthly model-dependent estimates have already been published, and hence they provide no protection to the real time estimates. (The benchmarked estimates are used for trend estimation.)

In this article we study a solution to the benchmarking problem that addresses the two shortcomings of the current BLS procedure. The proposed benchmarking consists of two steps:

a. Fit the model jointly to all the $D = 9$ CDs, each month adding the constraint

$$\sum_{d=1}^{D} w_{dt}\hat{Y}_{dt}^{\text{model}} = \sum_{d=1}^{D} w_{dt}\hat{Y}_{dt}^{\text{cps}}, \qquad t = 1, 2, \ldots, \tag{1}$$

where $\hat{Y}_{dt}^{\text{model}}$ and $\hat{Y}_{dt}^{\text{cps}}$ define the model-dependent estimator and the direct CPS estimator at time t, and the w_{dt}'s are fixed weights.

b. Fit the model jointly to all of the states in any given CD, each month adding the constraint

$$\sum_{s\in d} w_{st}\hat{Y}_{st}^{\text{model}} = \hat{Y}_{dt}^{\text{bench}}, \qquad d = 1, \ldots, D, t = 1, 2, \ldots, \tag{2}$$

where $\hat{Y}_{dt}^{\text{bench}}$ is the benchmarked estimator for CD d at time t as obtained in step a.

The justification for incorporating the constraints (1) is that the direct CPS estimators, which are unreliable in single CDs (notably in the smaller ones), can be trusted when averaged over different CDs. Note in this respect, that the sampling errors of the direct estimates, that are highly correlated within a CD are independent between CDs. As illustrated in the application in Section 5, the benchmarked CD estimators are much

Danny Pfeffermann is Professor of Statistics, Hebrew University of Jerusalem, Jerusalem 91905, Israel, and Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, SO17 1BJ, U.K. (E-mail:*msdanny@huji.ac.il*). Richard Tiller is Senior Mathematical Statistician, Statistical Methods Staff, Office of Employment and Unemployment Statistics, Bureau of Labor Statistics, Washington, DC 20212 (E-mail: *Tiller.Richard@bls.gov*). The authors thank the editor, an associate editor, and the reviewer for constructive comments and suggestions.

In the Public Domain
Journal of the American Statistical Association
December 2006, Vol. 101, No. 476, Applications and Case Studies
DOI 10.1198/016214506000000591

Downloaded by [Board of Governors of the Federal Reserve System] at 10:31 23 October 2017

more accurate than the corresponding CPS estimates, thus justifying benchmarking the state estimates to the benchmarked CD estimates as in (2). The basic idea behind using either of these constraints is that if all of the direct CPS estimates jointly increase or decrease due to some sudden external shocks that are not accounted for by the model, then the benchmarked estimators will reflect this change much faster than the model-dependent estimators obtained by fitting separate models for each CD and state. This property is strikingly illustrated in the application in Section 5 using actual CD unemployment series. Note also that by incorporating the constraints (1) and (2), the benchmarked estimators in a given month "borrow strength" both from past data and cross-sectionally, unlike the model-dependent estimators in current use that borrow strength only from the past. This property is reflected by the reduced variance of the benchmarked estimators, which is shown in Section 5.

Remark 1. Unlike in classical benchmarking problems that use external (independent) data (e.g., census figures or estimates from another large sample) for the benchmarking process, the model-dependent estimates on the left side of (1) are benchmarked to a weighted mean of the direct CPS estimates, which are the input data for the models. A similar problem exists with the benchmarking of the state estimates in (2). External independent data to which the monthly CD or state estimates can be benchmarked are not available even for isolated months. (See Hillmer and Trabelsi 1987; Doran 1992; Durbin and Quenneville 1997 for benchmarking procedures to external data sources in the context of state–space modeling.)

An important question underlying use of the constraints (1) and (2) is defining the weights $\{w_{dt}\}$ and $\{w_{st}\}$. Possible definitions for the weights $\{w_{dt}\}$ in (1) are

$$w_{1dt} = \frac{1}{D}, \qquad w_{2dt} = \frac{N_{dt}}{\sum_{d=1}^{D} N_{dt}},$$
$$w_{3dt} = \frac{1/\operatorname{var}_{dt}^{\text{cps}}}{\sum_{d=1}^{D}[1/\operatorname{var}_{dt}^{\text{cps}}]}, \tag{3}$$

where N_{dt} and $\operatorname{var}_{dt}^{\text{cps}}$ denote the total size of the labor force and the variance of the direct CPS estimate in CD d at month t. Using the weights w_{1dt} is appropriate when the direct estimates are totals; using w_{2dt} is appropriate when the direct estimates are proportions. Therefore, using these weights guarantees that the CD estimates each month add up to the corresponding national CPS estimate. By using similar weights in (2), the benchmarked state estimates add up to the benchmarked CD estimate in each of the CDs, and hence also to the national CPS estimate, thus satisfying publication-consistency requirements. Using w_{3dt} minimizes the variance of the benchmark $\sum_{d=1}^{D} w_{dt}\hat{Y}_{dt}^{\text{cps}}$ but does not satisfy publication consistency.

The proposed solution of combining the individual CD or state models into a joint state–space model with built-in benchmark constrains intensifies the computations. The dimension of the state vector in the separate models is 29 (see the next section), implying that by fitting the model jointly to, say, the 9 CDs, the dimension of the joint state vector would be 261. Fitting state–space models of this size puts heavy demands on CPU time and memory. This creates problems in a production environment in which all of the CD and state estimates (and numerous other estimates) must be produced and published each month soon after the new sample data become available. A possible solution to this problem studied herein that also allows computing the variances of the benchmarked estimators is to include the sampling errors as part of the error terms in the model measurement equation, instead of the current practice of fitting a linear time series model for the sampling errors and including them in the state vector. Implementing this solution reduces the dimension of each of the separate state vectors by half, because the sampling errors make up 15 elements of the state vector (see the next section). As explained in Section 2, including the sampling errors in the measurement equation does not change the model.

Using this solution, however, does induce autocorrelated errors in the measurement equation of the state–space model, because as mentioned earlier, the sampling errors are highly correlated over time. This raises the need to develop a recursive filtering algorithm with good statistical properties for state–space models with autocorrelated measurement errors. Although originally motivated by computational considerations, the development of such an algorithm is of general interest because the common practice of fitting a linear time series model for the measurement errors when they are autocorrelated is not always practical. See the next section for the sampling error model approximation used by the BLS. To the best of our knowledge, no such algorithm has been studied previously. Pfeffermann and Burck (1990) likewise added constraints of the form (1) to a state–space model and developed an appropriate recursive filtering algorithm, but their model did not contain autocorrelated sampling errors, so the measurement errors are independent cross-sectionally and over time.

Implementation of the proposed benchmarking procedure, which is the primary focus of this article, therefore requires solving three problems:

1. Develop a general recursive filtering algorithm for state–space models with correlated measurement errors.
2. Incorporate the benchmark constraints and compute the corresponding benchmarked estimates (estimates of employment or unemployment measures in the present application).
3. Compute the variances of the benchmarked estimators.

We emphasize with regard to the third problem that the constraints in (1) and (2) are imposed only for computing the benchmarked estimators, not when computing the variances of the benchmarked estimators, which account for all of the sources of variability, including the errors in the benchmark equations. As explained later, it would have been necessary to develop a new filtering algorithm for computing the correct variances, even if the sampling errors were left in the state vector.

Remark 2. An alternative and simpler way of enforcing the benchmark constraints is by pro-rating the separate model-dependent estimates each month, such that for CDs, for example,

$$\hat{Y}_{dt}^{\text{pro}} = \hat{Y}_{dt}^{\text{model}}\left(\sum_{d=1}^{D} w_{dt}\hat{Y}_{dt}^{\text{cps}} \Big/ \sum_{d=1}^{D} w_{dt}\hat{Y}_{dt}^{\text{model}}\right). \tag{4}$$

Downloaded by [Board of Governors of the Federal Reserve System] at 10:31 23 October 2017

Using (4) again satisfies $\sum_{d=1}^{D} w_{dt}\hat{Y}_{dt}^{\text{pro}} = \sum_{d=1}^{D} w_{dt}\hat{Y}_{dt}^{\text{cps}}$ and does not require changing the current BLS modeling practice. This benchmarking method is often used in small-area estimation applications that use cross-sectional models (see, e.g., Fay and Herriot 1979; Battese, Harter, and Fuller 1988; Rao 2003). However, the properties of using (4) are unclear, and it does not lend itself to simple parametric estimation of the variances of the pro-rated estimators. Variance estimation is an essential requirement for any benchmarking procedure. For the case of a single state–space model (with no benchmarking), Pfeffermann and Tiller (2005) developed bootstrap variance estimators that account for estimation of the model parameters with bias of correct order, but extension of this resampling method to the pro-rated predictors defined by (4) is not straightforward and requires a double-bootstrap procedure that is not practical in the present context because of the very intensive computations involved. Another simple benchmarking procedure is a difference adjustment of the form

$$\hat{Y}_{dt}^{\text{dif}} = \hat{Y}_{dt}^{\text{model}} + \left(\sum_{d=1}^{D} w_{dt}\hat{Y}_{dt}^{\text{cps}} - \sum_{d=1}^{D} w_{dt}\hat{Y}_{dt}^{\text{model}}\right). \tag{5}$$

But this procedure inflates the variances of the benchmarked estimators and has the further drawback of adding the same amount to each of the model-dependent estimators, irrespective of their magnitude.

The benchmarking procedure developed in this article overcomes the problems mentioned with respect to procedures (4) and (5) and is not more complicated once an appropriate computer program is written. Note in this respect, that any other "built-in" benchmarking procedure would require the development of a new filtering algorithm, even if the sampling errors were left in the state vector. This is so because the benchmark errors that need to be accounted for when computing the variance of the benchmarked estimator are correlated concurrently and over time with the sampling errors. As mentioned earlier, available algorithms for benchmarking state–space models are not suitable for handling benchmarks of the form (1) or (2).

In what follows, we focus on benchmarking the CD estimates (the first step in the benchmarking process), but we comment on the adjustments required when benchmarking the state estimates in the second step. Section 2 presents the BLS models in current use. Section 3 develops the recursive filtering algorithm for state–space models with correlated measurement errors and discusses its properties. Section 4 shows how to incorporate the benchmark constraints and compute the variances of the benchmarked estimators. Section 5 describes the application of the proposed procedure using the unemployment series, with special attention to the year 2001, when the World Trade Center was attacked. Section 6 concludes by discussing some remaining problems in the application of this procedure.

2. BUREAU OF LABOR STATISTICS MODEL IN CURRENT USE

In this section we consider a single CD, and hence we drop the subscript d from the notation. The model used by the BLS combines a model for the true CD values (total unemployment or employment) with a model for the sampling errors. The model is discusses in detail, including parameter estimation and model diagnostics in Tiller (1992). Next, we provide a brief description of the model, to aid understanding of the developments and applications in subsequent sections.

2.1 Model Assumed for Population Values

Let y_t denote the direct CPS estimate at time t and let Y_t denote the corresponding population value, so that $e_t = y_t - Y_t$ is the sampling error. The model is

$$\begin{aligned}
&Y_t = L_t + S_t + I_t, \qquad I_t \sim \mathrm{N}(0,\sigma_I^2);\\
&L_t = L_{t-1} + R_{t-1} + \eta_{Lt}, \qquad \eta_{Lt} \sim \mathrm{N}(0,\sigma_L^2),\\
&R_t = R_{t-1} + \eta_{Rt}, \qquad \eta_{Rt} \sim \mathrm{N}(0,\sigma_R^2);\\
&S_t = \sum_{j=1}^{6} S_{j,t};\\
&S_{j,t} = \cos\omega_j S_{j,t-1} + \sin\omega_j S_{j,t-1}^* + \nu_{j,t}, \qquad \nu_{j,t} \sim \mathrm{N}(0,\sigma_S^2),\\
&S_{j,t}^* = -\sin\omega_j S_{j,t-1} + \cos\omega_j S_{j,t-1}^* + \nu_{j,t}^*, \qquad \nu_{j,t}^* \sim \mathrm{N}(0,\sigma_S^2),\\
&\omega_j = \frac{2\pi j}{12}, \qquad j = 1,\ldots,6.
\end{aligned} \tag{6}$$

The model defined by (6) is known in the time series literature as the basic structural model (BSM), with L_t, R_t, S_t, and I_t defining the trend level, slope, seasonal effect, and "irregular" term operating at time t. The error terms I_t, η_{Lt}, η_{Rt}, ν_{jt}, and ν_{jt}^* are independent white noise series. The model for the trend approximates a local linear trend, whereas the model for the seasonal effect uses the classical decomposition of the seasonal component into six subcomponents, $S_{j,t}$, that represent the contribution of the cyclical functions corresponding to the six frequencies (harmonics) of a monthly seasonal series. The added noise permits the seasonal effects to evolve stochastically over time but in such a way that the expectation of the sum of 12 successive seasonal effects is 0. (See Harvey 1989 for discussion of the BSM.)

Remark 3. The future state models (but not the CD models) will be bivariate, with the other dependent variable in each state being "the number of persons in the state receiving unemployment insurance benefits" when modeling the unemployment estimates, and "the number of payroll jobs in business establishments" when modeling the employment estimates. These two series are census figures with no sampling errors. The benchmarking will be applied only to the unemployment and employment estimates.

2.2 Model Assumed for the Sampling Errors

The CPS variance of the sampling error varies with the level of the series. Denoting $v_t^2 = \operatorname{var}(e_t)$, the model assumed for the standardized residuals $e_t^* = e_t/v_t$ is AR(15), which 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 overlap underlying the CPS sampling design. By this design, households selected to the sample are surveyed for 4 successive months, are left out of the sample for the next 8 months, and then are surveyed again for 4 more months. The AR(2) model accounts for the autocorrelations

Downloaded by [Board of Governors of the Federal Reserve System] at 10:31 23 October 2017

arising from the fact that households dropped from the survey are replaced by households from the same "census tract." The reduced ARMA representation of the sum of the two processes is ARMA(2, 17), which is approximated by an AR(15) model.

The separate models holding for the population values and the sampling errors are cast into a single state–space model for the observations $\mathbf{y}_t$ (the CPS estimates) of the form

$$y_t = \mathbf{Z}_t\boldsymbol{\alpha}_t^*, \qquad \boldsymbol{\alpha}_t^* = \mathbf{T}\boldsymbol{\alpha}_{t-1}^* + \boldsymbol{\eta}_t^*,$$
$$E(\boldsymbol{\eta}_t^*) = \mathbf{0}, E(\boldsymbol{\eta}_t^*\boldsymbol{\eta}_t^{*\prime}) = \mathbf{Q}^*, \quad (7)$$

with the first equation to the left defining the *measurement* (observation) equation and the second equation defining the *state* (transition) equation. The state vector, $\boldsymbol{\alpha}_t^*$, consists of the trend level, L_t; the slope, R_t; the 11 seasonal coefficients, $S_{j,t}, S_{k,t}^*, j = 1, \dots, 6, k = 1, \dots, 5$; the irregular term, I_t; and the concurrent and 14 lags of the sampling errors—a total of 29 elements. The sampling errors and the irregular term are included in the state vector, so that there is no residual error term in the observation equation. Note in this respect, that including the sampling errors in the measurement equation as described in Section 1 and pursued later instead of the current practice of including them in the state equation does not change the model holding for the observations as long as the model fitted for the sampling errors reproduces their autocovariances.

The parameters indexing the model are estimated separately for each CD in two steps. In the first step, the AR(15) coefficients and residual variance are estimated by solving the corresponding Yule–Walker equations. (The sampling error variance and autocorrelations are estimated externally by the BLS.) In the second step, the remaining model variances are estimated by maximizing the model likelihood, with the AR(15) model parameters held fixed at their estimated values (see Tiller 1992 for details).

The monthly employment and unemployment estimates published by the BLS are obtained under the model (7) and the relationship $Y_t = y_t - e_t$ as

$$\hat{Y}_t = y_t - \hat{e}_t = \hat{L}_t + \hat{S}_t + \hat{I}_t, \quad (8)$$

where $\hat{L}_t$, $\hat{S}_t$, and $\hat{I}_t$ denote the estimated components at time t, as obtained by application of the Kalman filter (Harvey 1989).

3. FILTERING OF STATE–SPACE MODELS WITH AUTOCORRELATED MEASUREMENT ERRORS

In this section we develop a recursive filtering algorithm for state–space models with autocorrelated measurement errors. By this, we mean an algorithm that updates the most recent predictor of the state vector every time that a new observation becomes available. This filter is required for implementing the benchmarking procedure discussed in Section 1 (see Sec. 4), but, as already mentioned, it is general and can be used for other applications of state–space models with autocorrelated measurement errors. In Section 3.2 we discuss the properties of the proposed filter.

3.1 Recursive Filtering Algorithm

Consider the following linear state–space model for a (possibly vector) time series $\mathbf{y}_t$:

$$\textit{Measurement equation}: \quad \mathbf{y}_t = \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{e}_t,$$
$$E(\mathbf{e}_t) = \mathbf{0}, E(\mathbf{e}_t\mathbf{e}_t') = \boldsymbol{\Sigma}_{tt}, E(\mathbf{e}_\tau\mathbf{e}_t') = \boldsymbol{\Sigma}_{\tau t}; \quad (9a)$$
$$\textit{Transition equation}: \quad \boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t,$$
$$E(\boldsymbol{\eta}_t) = \mathbf{0}, E(\boldsymbol{\eta}_t\boldsymbol{\eta}_t') = \mathbf{Q}, E(\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-k}') = \mathbf{0}, k > 0. \quad (9b)$$

It is also assumed that $E(\boldsymbol{\eta}_t\mathbf{e}_\tau') = \mathbf{0}$ for all t and τ. (The restriction to time-invariant matrices $\mathbf{T}$ and $\mathbf{Q}$ is for convenience; extension of the filter to models that contain fixed effects in the observation equation is straightforward.) Clearly, what distinguishes this model from the standard linear state–space model is that the measurement errors $\mathbf{e}_t$ (the sampling errors in the BLS model) are correlated over time. Note in particular that unlike the BLS model representation in (7), where the sampling errors and the irregular term are part of the state vector so that there are no measurement errors in the observation equation, the measurement (sampling) errors now are featured in the observation equation. The recursive filtering algorithm developed herein takes into account the autocovariance matrices $\boldsymbol{\Sigma}_{\tau t}$.

At Time 1. Let $\hat{\boldsymbol{\alpha}}_1 = (\mathbf{I} - \mathbf{K}_1\mathbf{Z}_1)\mathbf{T}\hat{\boldsymbol{\alpha}}_0 + \mathbf{K}_1\mathbf{y}_1$ be the filtered state estimator at time 1, where $\hat{\boldsymbol{\alpha}}_0$ is an initial estimator with covariance matrix $\mathbf{P}_0 = E[(\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0)(\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0)']$ and $\mathbf{K}_1 = \mathbf{P}_{1|0}\mathbf{Z}_1'\mathbf{F}_1^{-1}$ is the "Kalman gain." We assume for convenience that $\hat{\boldsymbol{\alpha}}_0$ is independent of the observations. The matrix $\mathbf{P}_{1|0} = \mathbf{T}\mathbf{P}_0\mathbf{T}' + \mathbf{Q}$ is the covariance matrix of the prediction errors $\hat{\boldsymbol{\alpha}}_{1|0} - \boldsymbol{\alpha}_1 = \mathbf{T}\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_1$, and $\mathbf{F}_1 = \mathbf{Z}_1\mathbf{P}_{1|0}\mathbf{Z}_1' + \boldsymbol{\Sigma}_{11}$ is the covariance matrix of the innovations (one-step-ahead prediction errors) $\boldsymbol{\nu}_1 = \mathbf{y}_1 - \hat{\mathbf{y}}_{1|0} = \mathbf{y}_1 - \mathbf{Z}_1\hat{\boldsymbol{\alpha}}_{1|0}$. Because $\mathbf{y}_1 = \mathbf{Z}_1\boldsymbol{\alpha}_1 + \mathbf{e}_1$,

$$\hat{\boldsymbol{\alpha}}_1 = (\mathbf{I} - \mathbf{K}_1\mathbf{Z}_1)\mathbf{T}\hat{\boldsymbol{\alpha}}_0 + \mathbf{K}_1\mathbf{Z}_1\boldsymbol{\alpha}_1 + \mathbf{K}_1\mathbf{e}_1. \quad (10)$$

At Time 2. Let $\hat{\boldsymbol{\alpha}}_{2|1} = \mathbf{T}\hat{\boldsymbol{\alpha}}_1$ define the predictor of $\boldsymbol{\alpha}_2$ at time 1 with covariance matrix $\mathbf{P}_{2|1} = E[(\hat{\boldsymbol{\alpha}}_{2|1} - \boldsymbol{\alpha}_2)(\hat{\boldsymbol{\alpha}}_{2|1} - \boldsymbol{\alpha}_2)']$. An unbiased predictor $\hat{\boldsymbol{\alpha}}_2$ of $\boldsymbol{\alpha}_2$ [i.e., $E(\hat{\boldsymbol{\alpha}}_2 - \boldsymbol{\alpha}_2) = \mathbf{0}$] based on $\hat{\boldsymbol{\alpha}}_{2|1}$ and the observation $\mathbf{y}_2$ is the generalized least squares (GLS) predictor in the regression model

$$\begin{pmatrix} \mathbf{T}\hat{\boldsymbol{\alpha}}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I} \\ \mathbf{Z}_2 \end{pmatrix}\boldsymbol{\alpha}_2 + \begin{pmatrix} \mathbf{u}_{2|1} \\ \mathbf{e}_2 \end{pmatrix}, \qquad \mathbf{u}_{2|1} = \mathbf{T}\hat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_2, \quad (11)$$

that is,

$$\hat{\boldsymbol{\alpha}}_2 = \left[(\mathbf{I}, \mathbf{Z}_2')\mathbf{V}_2^{-1}\begin{pmatrix} \mathbf{I} \\ \mathbf{Z}_2 \end{pmatrix}\right]^{-1}(\mathbf{I}, \mathbf{Z}_2')\mathbf{V}_2^{-1}\begin{pmatrix} \mathbf{T}\hat{\boldsymbol{\alpha}}_1 \\ \mathbf{y}_2 \end{pmatrix}, \quad (12)$$

where

$$\mathbf{V}_2 = \mathrm{var}\begin{pmatrix} \mathbf{u}_{2|1} \\ \mathbf{e}_2 \end{pmatrix} = \begin{bmatrix} \mathbf{P}_{2|1} & \mathbf{C}_2 \\ \mathbf{C}_2' & \boldsymbol{\Sigma}_{22} \end{bmatrix} \quad (13)$$

and $\mathbf{C}_2 = \mathrm{cov}[\mathbf{T}\hat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_2, \mathbf{e}_2] = \mathbf{T}\mathbf{K}_1\boldsymbol{\Sigma}_{12}$ [follows from (9a) and (10)]. Note that $\mathbf{V}_2$ is the covariance matrix of the errors $\mathbf{u}_{2|1}$ and $\mathbf{e}_2$ and not of the predictors $\mathbf{T}\hat{\boldsymbol{\alpha}}_1$ and $\mathbf{y}_2$. As discussed later and proved in Appendix A, the GLS predictor $\hat{\boldsymbol{\alpha}}_2$ is the best linear unbiased predictor (BLUP) of $\boldsymbol{\alpha}_2$ based on $\mathbf{T}\hat{\boldsymbol{\alpha}}_1$ and $\mathbf{y}_2$, with covariance matrix

$$E[(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_2)(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_2)'] = \left[(\mathbf{I}, \mathbf{Z}_2')\mathbf{V}_2^{-1}\begin{pmatrix} \mathbf{I} \\ \mathbf{Z}_2 \end{pmatrix}\right]^{-1} = \mathbf{P}_2. \quad (14)$$

Downloaded by [Board of Governors of the Federal Reserve System] at 10:31 23 October 2017

At Time t. Let $\hat{\boldsymbol{\alpha}}_{t|t-1} = \mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1}$ define the predictor of $\boldsymbol{\alpha}_t$ at time $t-1$ with covariance matrix $E[(\hat{\boldsymbol{\alpha}}_{t|t-1} - \boldsymbol{\alpha}_t)(\hat{\boldsymbol{\alpha}}_{t|t-1} - \boldsymbol{\alpha}_t)'] = \mathbf{T}\mathbf{P}_{t-1}\mathbf{T}' + \mathbf{Q} = \mathbf{P}_{t|t-1}$, where $\mathbf{P}_{t-1} = E[(\hat{\boldsymbol{\alpha}}_{t-1} - \boldsymbol{\alpha}_{t-1}) \times (\hat{\boldsymbol{\alpha}}_{t-1} - \boldsymbol{\alpha}_{t-1})']$. Set the random coefficients regression model

$$\begin{pmatrix} \mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1} \\ \mathbf{y}_t \end{pmatrix} = \begin{pmatrix} \mathbf{I} \\ \mathbf{Z}_t \end{pmatrix} \boldsymbol{\alpha}_t + \begin{pmatrix} \mathbf{u}_{t|t-1} \\ \mathbf{e}_t \end{pmatrix}, \qquad \mathbf{u}_{t|t-1} = \mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1} - \boldsymbol{\alpha}_t, \tag{15}$$

and define

$$\mathbf{V}_t = \operatorname{var}\begin{pmatrix} \mathbf{u}_{t|t-1} \\ \mathbf{e}_t \end{pmatrix} = \begin{bmatrix} \mathbf{P}_{t|t-1} & \mathbf{C}_t \\ \mathbf{C}_t' & \boldsymbol{\Sigma}_{tt} \end{bmatrix}. \tag{16}$$

The covariance matrix $\mathbf{C}_t = \operatorname{cov}[\mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1} - \boldsymbol{\alpha}_t, \mathbf{e}_t]$ is computed as follows. Let $[\mathbf{I}, \mathbf{Z}_j']\mathbf{V}_j^{-1} = [\mathbf{B}_{j1}, \mathbf{B}_{j2}]$ where $\mathbf{B}_{j1}$ contains the first q columns of $[\mathbf{I}, \mathbf{Z}_j']\mathbf{V}_j^{-1}$ and $\mathbf{B}_{j2}$ contains the remaining columns, with $q = \dim(\boldsymbol{\alpha}_j)$. Define $\mathbf{A}_j = \mathbf{T}\mathbf{P}_j\mathbf{B}_{j1}, \tilde{\mathbf{A}}_j = \mathbf{T}\mathbf{P}_j\mathbf{B}_{j2}, j = 2, \ldots, t-1, \tilde{\mathbf{A}}_1 = \mathbf{T}\mathbf{K}_1$. Then

$$\begin{aligned} \mathbf{C}_t &= \operatorname{cov}[\mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1} - \boldsymbol{\alpha}_t, \mathbf{e}_t] \\ &= \mathbf{A}_{t-1}\mathbf{A}_{t-2}\cdots\mathbf{A}_2\tilde{\mathbf{A}}_1\boldsymbol{\Sigma}_{1t} + \mathbf{A}_{t-1}\mathbf{A}_{t-2}\cdots\mathbf{A}_3\tilde{\mathbf{A}}_2\boldsymbol{\Sigma}_{2t} \\ &\quad + \cdots + \mathbf{A}_{t-1}\tilde{\mathbf{A}}_{t-2}\boldsymbol{\Sigma}_{t-2,t} + \tilde{\mathbf{A}}_{t-1}\boldsymbol{\Sigma}_{t-1,t}. \end{aligned} \tag{17}$$

The GLS predictor of $\boldsymbol{\alpha}_t$ based on $\mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1}$ and $\mathbf{y}_t$, and the covariance matrix of the prediction errors are obtained from (15) and (16) as

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_t &= \left[(\mathbf{I}, \mathbf{Z}_t')\mathbf{V}_t^{-1} \begin{pmatrix} \mathbf{I} \\ \mathbf{Z}_t \end{pmatrix} \right]^{-1} (\mathbf{I}, \mathbf{Z}_t')\mathbf{V}_t^{-1} \begin{pmatrix} \mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1} \\ \mathbf{y}_t \end{pmatrix}; \\ \mathbf{P}_t &= E[(\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t)(\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t)'] = \left[(\mathbf{I}, \mathbf{Z}_t')\mathbf{V}_t^{-1} \begin{pmatrix} \mathbf{I} \\ \mathbf{Z}_t \end{pmatrix} \right]^{-1}. \end{aligned} \tag{18}$$

Alternatively, the predictor $\hat{\boldsymbol{\alpha}}_t$ can be written (App. C) as

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_t = \mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1} &+ (\mathbf{P}_{t|t-1}\mathbf{Z}_t' - \mathbf{C}_t) \\ &\times [\mathbf{Z}_t\mathbf{P}_{t|t-1}\mathbf{Z}_t' - \mathbf{Z}_t\mathbf{C}_t - \mathbf{C}_t'\mathbf{Z}_t' + \boldsymbol{\Sigma}_{tt}]^{-1} \\ &\times (\mathbf{y}_t - \mathbf{Z}_t\mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1}). \end{aligned} \tag{19}$$

Written in this way, the predictor of $\boldsymbol{\alpha}_t$ at time t is seen to equal the predictor of $\boldsymbol{\alpha}_t$ at time $t-1$, plus a correction factor that depends on the magnitude of the innovation (one-step-ahead prediction error) when predicting $\mathbf{y}_t$ at time $t-1$.

3.2 Properties of the Filtering Algorithm

Assuming known model parameters, the recursive GLS filter defined by (18) or (19) has the following properties:

1. At every time point t, the filter produces the BLUP of $\boldsymbol{\alpha}_t$ based on the predictor $\hat{\boldsymbol{\alpha}}_{t|t-1} = \mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1}$ from time $t-1$ and the new observation $\mathbf{y}_t$. The BLUP property means that $E(\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t) = \mathbf{0}$ and $\operatorname{var}[\mathbf{d}'(\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t)] \le \operatorname{var}[\mathbf{d}' \times (\hat{\boldsymbol{\alpha}}_t^L - \boldsymbol{\alpha}_t)]$ for every vector coefficient $\mathbf{d}'$ and any other linear unbiased predictor of the form $\hat{\boldsymbol{\alpha}}_t^L = \mathbf{L}_1\hat{\boldsymbol{\alpha}}_{t|t-1} + \mathbf{L}_2\mathbf{y}_t + \mathbf{l}$ with general fixed matrices $\mathbf{L}_1$ and $\mathbf{L}_2$ and vector $\mathbf{l}$. See Appendix A for proof of this property.
2. When the measurement errors are independent, the GLS filter coincides with the familiar Kalman filter (Harvey 1989); see Appendix B for the proof. Thus the recursive GLS filter can be viewed as an extension of the Kalman filter for the case of correlated measurement errors.
3. Unlike the Kalman filter, which assumes independent measurement errors and therefore yields the BLUP of $\boldsymbol{\alpha}_t$ based on all the individual observations $\mathbf{y}_{(t)} = (\mathbf{y}_1', \ldots, \mathbf{y}_t')'$ (i.e., the BLUP out of all of the linear unbiased predictors of the form $\sum_{i=1}^t \mathbf{J}_{it}\mathbf{y}_i + \mathbf{j}_t$ for general fixed matrices $\mathbf{J}_{it}$ and vector $\mathbf{j}_t$), the filter (18) yields the BLUP of $\boldsymbol{\alpha}_t$ based on $\hat{\boldsymbol{\alpha}}_{t|t-1}$ and the new observation $\mathbf{y}_t$.

(The predictor $\hat{\boldsymbol{\alpha}}_{t|t-1}$ is itself a linear combination of all the observations until time $t-1$, but the values of the matrix coefficients are fixed by the previous steps of the filter, see also Remark 4.)

Computing the BLUP of $\boldsymbol{\alpha}_t$ under correlated measurement errors out of all possible unbiased predictors that are linear combinations of all the individual observations in $\mathbf{y}_{(t)}$ with arbitrary coefficients [or the minimum mean squared error predictor, $E(\boldsymbol{\alpha}_t|\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1; \boldsymbol{\lambda})$, under normality of the model error terms where $\boldsymbol{\lambda}$ defines the model parameters] requires joint modeling of $\mathbf{y}_{(t)}$ for every time t. For long series, the computations become very heavy and generally are not practical in a production environment that requires routine runs of many series with high-dimensional state vectors in a short time. Empirical evidence so far suggests that the loss in efficiency from using the GLS algorithm instead of the BLUP based on all of the individual observations is mild. Section 5 provides an empirical comparison of the two filters.

Remark 4. For general covariance matrices $\boldsymbol{\Sigma}_{\tau t}$ between the measurement errors, it is impossible to construct a recursive filtering algorithm that is a linear combination of the predictor from the previous time point and the new observation and is BLUP out of all unbiased predictors of the form $\sum_{i=1}^t \mathbf{J}_{it}\mathbf{y}_i + \mathbf{j}_t$. To see this, consider a very simple example of three observations y_1, y_2, and y_3 with common mean μ and variance σ^2. If the three observations are independent, then the BLUP of μ based on the first two observations is $\bar{y}_{(2)} = (y_1 + y_2)/2$ and the BLUP based on the three observations is $\bar{y}_{(3)} = (y_1 + y_2 + y_3)/3 = (2/3)\bar{y}_{(2)} + (1/3)y_3$. The BLUP $\bar{y}_{(3)}$ is the Kalman filter predictor for time 3. Suppose, however, that $\operatorname{cov}(y_1, y_2) = \operatorname{cov}(y_2, y_3) = \sigma^2\rho_1$ and $\operatorname{cov}(y_1, y_3) = \sigma^2\rho_2 \ne \sigma^2\rho_1$. The BLUP of μ based on the first two observations is again $\bar{y}_{(2)} = (y_1 + y_2)/2$, but the BLUP of μ based on the three observations in this case is $\bar{y}_{(3)}^{\text{opt}} = ay_1 + by_2 + ay_3$, where $a = (1-\rho_1)/(3 - 4\rho_1 + \rho_2)$ and $b = (1 - 2\rho_1 + \rho_2)/(3 - 4\rho_1 + \rho_2)$. Clearly, because $a \ne b$, the predictor $\bar{y}_{(3)}^{\text{opt}}$ cannot be written as a linear combination of $\bar{y}_{(2)}$ and y_3. For example, if $\rho_1 = .5$ and $\rho_2 = .25$, then $\bar{y}_{(3)}^{\text{opt}} = .4y_1 + .2y_2 + .4y_3$.

4. INCORPORATING THE BENCHMARK CONSTRAINTS

4.1 Joint Modeling of Several Concurrent Estimates and Their Weighted Mean

In this section we jointly model the concurrent CD estimates and their weighted mean. In Section 4.2 we show how to incorporate the benchmark constraints and compute the variances of the benchmarked predictors. We follow the BLS modeling practice and assume that the state vectors and the measurement errors are independent between the CDs. Section 6 considers an

Downloaded by [Board of Governors of the Federal Reserve System] at 10:31 23 October 2017

extension of the joint model that allows for cross-sectional correlations between corresponding components of the state vectors operating in different CDs.

Suppose that the models underlying the CD estimates are as in (9) with the correlated measurement (sampling) errors included in the observation equation. In what follows we add the subscript d to all of the model components to distinguish between the CDs. The (direct) estimates y_{dt} and the measurement errors e_{dt} are now scalars, and $\mathbf{Z}_{dt}$ is a row vector (hereinafter denoted as $\mathbf{z}'_{dt}$). Let $\tilde{\mathbf{y}}_t = (y_{1t}, \ldots, y_{dt}, \sum_{d=1}^{D} w_{dt}y_{dt})'$ define the concurrent estimates for the D CDs and their weighted mean [the right side of the benchmark equation (1)]. The corresponding vector of measurement errors is $\tilde{\mathbf{e}}_t = (e_{1t}, \ldots, e_{Dt}, \sum_{d=1}^{D} w_{dt}e_{dt})'$. Let $\mathbf{Z}_t^* = \mathbf{I}_D \otimes \mathbf{z}'_{dt}$ (a block-diagonal matrix with $\mathbf{z}'_{dt}$ as the dth block), $\tilde{\mathbf{T}}_t = \mathbf{I}_D \otimes \mathbf{T}$, $\tilde{\mathbf{Z}}_t = \left[\begin{smallmatrix} \mathbf{Z}_t^* \\ w_{1t}\mathbf{z}'_{1t} \cdots w_{dt}\mathbf{z}'_{Dt} \end{smallmatrix}\right]$, $\tilde{\boldsymbol{\alpha}}_t = (\boldsymbol{\alpha}'_{1t}, \ldots, \boldsymbol{\alpha}'_{Dt})'$, and $\tilde{\boldsymbol{\eta}}_t = (\boldsymbol{\eta}'_{1t}, \ldots, \boldsymbol{\eta}'_{Dt})'$. By (9) and the independence of the state vectors and measurement errors between CDs, the joint model holding for $\tilde{\mathbf{y}}_t$ is

$$\tilde{\mathbf{y}}_t = \tilde{\mathbf{Z}}_t\tilde{\boldsymbol{\alpha}}_t + \tilde{\mathbf{e}}_t,$$
$$E(\tilde{\mathbf{e}}_t) = \mathbf{0}, \qquad E(\tilde{\mathbf{e}}_\tau\tilde{\mathbf{e}}'_t) = \tilde{\boldsymbol{\Sigma}}_{\tau t} = \begin{bmatrix} \boldsymbol{\Sigma}_{\tau t} & \mathbf{h}_{\tau t} \\ \mathbf{h}'_{t\tau} & \nu_{\tau t} \end{bmatrix}, \tag{20a}$$

$$\tilde{\boldsymbol{\alpha}}_t = \tilde{\mathbf{T}}\tilde{\boldsymbol{\alpha}}_{t-1} + \tilde{\boldsymbol{\eta}}_t;$$
$$E(\tilde{\boldsymbol{\eta}}_t) = \mathbf{0}, \qquad E(\tilde{\boldsymbol{\eta}}_t\tilde{\boldsymbol{\eta}}'_t) = \mathbf{I}_D \otimes \mathbf{Q}_d = \tilde{\mathbf{Q}}, \tag{20b}$$
$$E(\tilde{\boldsymbol{\eta}}_t\tilde{\boldsymbol{\eta}}'_{t-k}) = \mathbf{0}, \qquad k > 0,$$

where

$$\boldsymbol{\Sigma}_{\tau t} = \operatorname{diag}[\sigma_{1\tau t}, \ldots, \sigma_{D\tau t}); \qquad \sigma_{d\tau t} = \operatorname{cov}[e_{d\tau}, e_{dt}),$$
$$\nu_{\tau t} = \sum_{d=1}^{D} w_{d\tau}w_{dt}\sigma_{d\tau t} = \operatorname{cov}\left[\sum_{d=1}^{D} w_{d\tau}e_{d\tau}, \sum_{d=1}^{D} w_{dt}e_{dt}\right];$$
$$\mathbf{h}_{\tau t} = (h_{1\tau t}, \ldots, h_{D\tau t})';$$
$$h_{d\tau t} = w_{dt}\sigma_{d\tau t} = \operatorname{cov}\left[e_{d\tau}, \sum_{d=1}^{D} w_{dt}e_{dt}\right].$$

Remark 5. The model (20a)–(20b) is the same as the separate models defined by (9a)–(9b) for the different CDs. Adding the model for $\sum_{d=1}^{D} w_{dt}y_{dt}$ to the observation equation provides no new information.

4.2 Recursive Filtering With Correlated Measurements and Benchmark Constraints

To compute the benchmarked predictors, we apply the recursive GLS filter (19) to the joint model defined by (20a)–(20b), setting the variance of the benchmarked errors, $\sum_{d=1}^{D} w_{dt}e_{dt}$, to 0. The idea behind this procedure is as follows. By (8) and (20), the model-dependent predictor of the signal (true population value) in CD d at time t takes the form $\hat{\mathbf{Y}}_{dt}^{\text{model}} = \mathbf{z}'_{dt}\hat{\boldsymbol{\alpha}}_{dt}$. Thus the benchmark constraints (1) can be written as

$$\sum_{d=1}^{D} w_{dt}\mathbf{z}'_{dt}\hat{\boldsymbol{\alpha}}_{dt} = \sum_{d=1}^{D} w_{dt}y_{dt}, \qquad t = 1, 2, \ldots. \tag{21}$$

In contrast, under the model (20),

$$\sum_{d=1}^{D} w_{dt}y_{dt} = \sum_{d=1}^{D} w_{dt}\mathbf{z}'_{dt}\boldsymbol{\alpha}_{dt} + \sum_{d=1}^{D} w_{dt}e_{dt}, \tag{22}$$

and so a simple way of satisfying the benchmark constraints is by imposing $\sum_{d=1}^{D} w_{dt}y_{dt} = \sum_{d=1}^{D} w_{dt}\mathbf{z}'_{dt}\boldsymbol{\alpha}_{dt}$ or, equivalently, by setting

$$\operatorname{var}\left[\sum_{d=1}^{D} w_{dt}e_{dt}\right] = \operatorname{cov}\left[e_{dt}, \sum_{d=1}^{D} w_{dt}e_{dt}\right] = 0,$$
$$d = 1, \ldots, D, t = 1, 2, \ldots. \tag{23}$$

It is important to emphasize that using (23) is just a convenient technical way of forcing the constraints and hence computing the benchmarked predictors. In Appendix D we show how to compute the true variances of the benchmarked predictors, accounting in particular for the errors $\sum_{d=1}^{D} w_{dt}e_{dt}$ in the right side of (22) [and no longer imposing (23)]. The imposition of (23) in the GLS filter (19) is implemented by replacing the covariance matrix $\tilde{\boldsymbol{\Sigma}}_{tt}$ of the measurement equation (9a) by the matrix $\tilde{\boldsymbol{\Sigma}}_{tt}^* = \left[\begin{smallmatrix} \boldsymbol{\Sigma}_{tt} & \mathbf{0}_{(D)} \\ \mathbf{0}'_{(D)} & 0 \end{smallmatrix}\right]$, where $\mathbf{0}_{(D)}$ is the null vector of order D, and setting the elements of the last column of the matrix $\mathbf{C}_t^{bmk} = \operatorname{cov}[\tilde{\mathbf{T}}\tilde{\boldsymbol{\alpha}}_{t-1}^{bmk} - \tilde{\boldsymbol{\alpha}}_t, \tilde{\mathbf{e}}_t]$ to 0.

Application of the GLS filter (19) to the model (9), with the benchmark constraints imposed by application of (23), yields the vector of CD benchmarked estimators

$$\begin{aligned} \mathbf{Z}_t^*\tilde{\boldsymbol{\alpha}}_t^{bmk} &= \mathbf{Z}_t^*\big\{\tilde{\mathbf{T}}\tilde{\boldsymbol{\alpha}}_{t-1}^{bmk} + (\mathbf{P}_{t|t-1}^{bmk}\tilde{\mathbf{Z}}'_t - \mathbf{C}_{t,0}^{bmk}) \\ &\quad \times \big[\tilde{\mathbf{Z}}_t\mathbf{P}_{t|t-1}^{bmk}\tilde{\mathbf{Z}}'_t - \tilde{\mathbf{Z}}_t\mathbf{C}_{t,0}^{bmk} - \mathbf{C}_{t,0}^{bmk\prime}\tilde{\mathbf{Z}}'_t + \tilde{\boldsymbol{\Sigma}}_{tt}^*\big]^{-1} \\ &\quad \times (\tilde{\mathbf{y}}_t - \tilde{\mathbf{Z}}_t\tilde{\mathbf{T}}\tilde{\boldsymbol{\alpha}}_{t-1}^{bmk})\big\}, \end{aligned} \tag{24}$$

where $\mathbf{Z}_t^* = \mathbf{I}_D \otimes \mathbf{z}'_{dt}$, $\mathbf{P}_{t|t-1}^{bmk} = \tilde{\mathbf{T}}\mathbf{P}_{t-1}^{bmk}\tilde{\mathbf{T}}' + \tilde{\mathbf{Q}}$, $\mathbf{P}_{t-1}^{bmk} = E[(\tilde{\boldsymbol{\alpha}}_{t-1} - \tilde{\boldsymbol{\alpha}}_{t-1}^{bmk})(\tilde{\boldsymbol{\alpha}}_{t-1} - \tilde{\boldsymbol{\alpha}}_{t-1}^{bmk})']$, and, by defining $\tilde{\mathbf{e}}_{t,0} = (e_{1t}, \ldots, e_{Dt}, 0)'$, $\mathbf{C}_{t,0}^{bmk} = \operatorname{cov}[\tilde{\mathbf{T}}\tilde{\boldsymbol{\alpha}}_{t-1}^{bmk} - \tilde{\boldsymbol{\alpha}}_t, \tilde{\mathbf{e}}_{t,0}]$.

Remark 6. The matrix $\mathbf{P}_{t|t-1}^{bmk} = E[(\tilde{\mathbf{T}}\tilde{\boldsymbol{\alpha}}_{t-1}^{bmk} - \tilde{\boldsymbol{\alpha}}_t)(\tilde{\mathbf{T}}\tilde{\boldsymbol{\alpha}}_{t-1}^{bmk} - \tilde{\boldsymbol{\alpha}}_t)']$ is the true prediction error covariance matrix. See Appendix D for the computation of $\mathbf{P}_t^{bmk} = E[(\tilde{\boldsymbol{\alpha}}_t - \tilde{\boldsymbol{\alpha}}_t^{bmk})(\tilde{\boldsymbol{\alpha}}_t - \tilde{\boldsymbol{\alpha}}_t^{bmk})']$ and $\mathbf{C}_t^{bmk} = \operatorname{cov}[\tilde{\mathbf{T}}\tilde{\boldsymbol{\alpha}}_{t-1}^{bmk} - \tilde{\boldsymbol{\alpha}}_t, \tilde{\mathbf{e}}_t]$ under the model (9) without the constraints. The matrix $\mathbf{P}_t^{bmk}$ accounts for the variability of the state vector components and the variances and covariances of the sampling errors (defining the matrices $\tilde{\boldsymbol{\Sigma}}_{tt}$ and $\mathbf{C}_t^{bmk}$).

Note again that the benchmarking filter developed in this article and particularly the covariance matrix $\mathbf{P}_t^{bmk}$ is different from the state–space benchmarking filters developed in other works, where the series observations are benchmarked to external (independent) data sources. For example, Doran (1992) considered benchmark constraints (possibly a set) of the form $\mathbf{R}_t\tilde{\boldsymbol{\alpha}}_t = r_t$, where the r_t's are constants. In our case the benchmarks $\sum_{d=1}^{D} w_{dt}y_{dt}$ are random and depend heavily on the sum $\sum_{d=1}^{D} w_{dt}\mathbf{z}'_{dt}\boldsymbol{\alpha}_{dt}$. Another difference between the present filter and the other filters developed in the context of state–space modeling is the accounting for correlated measurement errors in the present filter.

4.3 Simulation Results

To test whether the proposed benchmarking algorithm with correlated measurement errors as developed in Section 4.2 performs properly, we designed a small simulation study. The study consists of simulating 10,000 time series of length 45 for each of 3 models and computing the simulation variances of

Downloaded by [Board of Governors of the Federal Reserve System] at 10:31 23 October 2017

the benchmark prediction errors $(\tilde{\boldsymbol{\alpha}}_t^{bmk} - \tilde{\boldsymbol{\alpha}}_t)$, which we compare with the true variances under the model $\mathbf{P}_t^{bmk} = E[(\tilde{\boldsymbol{\alpha}}_t^{bmk} - \tilde{\boldsymbol{\alpha}}_t)(\tilde{\boldsymbol{\alpha}}_t^{bmk} - \tilde{\boldsymbol{\alpha}}_t)']$ as defined by (D.3) in Appendix D. We also compare the simulation covariances, $\mathrm{cov}[\tilde{\mathbf{T}}\tilde{\boldsymbol{\alpha}}_{t-1}^{bmk} - \tilde{\boldsymbol{\alpha}}_t, \tilde{\mathbf{e}}_t]$, with the true covariances, $\mathbf{C}_t^{bmk}$, under the model as defined by (D.4). The models used for generating the three sets of time series have the general form

$$\begin{aligned} \mathbf{y}_{dt} &= \boldsymbol{\alpha}_{dt} + \mathbf{e}_{dt}, \\ \boldsymbol{\alpha}_{dt} &= \boldsymbol{\alpha}_{d,t-1} + \boldsymbol{\eta}_{dt}, \\ \mathbf{e}_{dt} &= \boldsymbol{\varepsilon}_{dt} + .55\boldsymbol{\varepsilon}_{d,t-1} + .30\boldsymbol{\varepsilon}_{d,t-2} + .10\boldsymbol{\varepsilon}_{d,t-3}, \\ & \quad d = 1, 2, 3, \end{aligned} \tag{25}$$

where $\{\boldsymbol{\eta}_{dt}\}$ and $\{\boldsymbol{\varepsilon}_{dt}\}$ are independent white noise series. The random-walk variances for the three models are $\mathrm{var}(\boldsymbol{\eta}_{dt}) = (.01, .88, 1.2)$. The corresponding measurement error variances are $\mathrm{var}(\mathbf{e}_{dt}) = (.30, .08, 1.21)$, with autocorrelations $\mathrm{corr}(\mathbf{e}_t, \mathbf{e}_{t-1}) = .53$, $\mathrm{corr}(\mathbf{e}_t, \mathbf{e}_{t-2}) = .25$, and $\mathrm{corr}(\mathbf{e}_t, \mathbf{e}_{t-3}) = .07$ (same autocorrelations for the three models). The benchmark constraints are

$$\hat{\alpha}_{1t} + \hat{\alpha}_{2t} + \hat{\alpha}_{3t} = y_{1t} + y_{2t} + y_{3t}, \qquad t = 1, \ldots, 45. \tag{26}$$

Table 1 shows for each of the three series the values of the true variance $p_{d45} = E(\alpha_{d45}^{bmk} - \alpha_{d45})^2$ and covariance $c_{d45} = \mathrm{cov}[\alpha_{d44}^{bmk} - \alpha_{d45}, e_{d45}]$ for the last time point ($t = 45$), along with the corresponding simulation variance $p_{d45}^{\mathrm{Sim}} = \sum_{k=1}^{10{,}000}(\alpha_{d45}^{bmk(k)} - \alpha_{d45}^{(k)})^2/10^4$ and covariance $c_{d45}^{\mathrm{Sim}} = \sum_{k=1}^{10{,}000}(\alpha_{d44}^{bmk(k)} - \alpha_{d45}^{(k)})e_{d45}^{(k)}/10^4$, where the superscript k indicates the kth simulation.

The results presented in Table 1 show a close fit even for the third model where both the random-walk variance and the measurement error variance are relatively high. These results support the theoretical expressions.

4.4 Benchmarking of State Estimates

We considered so far (and again in Sec. 5) the benchmarking of the CD estimates but as discussed in Section 1, the second step of the benchmarking process is to benchmark the state estimates to the corresponding benchmarked CD estimate obtained in the first step, using the constraints (2). This step is carried out similarly to the benchmarking of the CD estimates, but it requires changing the computation of the covariance matrices $\tilde{\boldsymbol{\Sigma}}_{\tau t} = E(\tilde{\mathbf{e}}_\tau \tilde{\mathbf{e}}_t')$ [see (20a)] and hence the computation of the covariances $\mathbf{C}_t^{bmk} = \mathrm{cov}[\tilde{\mathbf{T}}\tilde{\boldsymbol{\alpha}}_{t-1}^{bmk} - \tilde{\boldsymbol{\alpha}}_t, \tilde{\mathbf{e}}_t]$ [see (D.4)] and the variance matrices $\mathbf{P}_t^{bmk} = E[(\tilde{\boldsymbol{\alpha}}_t^{bmk} - \tilde{\boldsymbol{\alpha}}_t)(\tilde{\boldsymbol{\alpha}}_t^{bmk} - \tilde{\boldsymbol{\alpha}}_t)']$ [see (D3)].

Table 1. Theoretical and Simulation Variances and Covariances Under the Model (25) and Benchmark (26) for Last Time Point (t = 45), 10,000 Simulations

	Variances		Covariances	
Series	Theoretical p_{d45}	Simulation p_{d45}^{sim}	Theoretical c_{d45}	Simulation c_{d45}^{sim}
1	.274	.276	.039	.041
2	1.122	1.119	.615	.614
3	.337	.344	.063	.068

The reason why the matrices $\tilde{\boldsymbol{\Sigma}}_{\tau t}$ have to be computed differently when benchmarking the states is that in this case the benchmarking is carried out by imposing the constraints

$$\sum_{s \in d} w_{st}\mathbf{z}_{st}'\boldsymbol{\alpha}_{st} = \mathbf{z}_{dt}'\tilde{\boldsymbol{\alpha}}_{dt}^{bmk} = \mathbf{z}_{dt}'\tilde{\boldsymbol{\alpha}}_{dt} + \mathbf{z}_{dt}'(\tilde{\boldsymbol{\alpha}}_{dt}^{bmk} - \tilde{\boldsymbol{\alpha}}_{dt}), \tag{27}$$

so that the benchmarking error is now $\mathbf{e}_{dt}^{bmk} = \mathbf{z}_{dt}'(\tilde{\boldsymbol{\alpha}}_{dt}^{bmk} - \tilde{\boldsymbol{\alpha}}_{dt})$, which has a different structure than the benchmarking error $\sum_{d=1}^D w_{dt}e_{dt}$ underlying (1) [see (22)]. The vectors $\tilde{\mathbf{e}}_t$ now have the form $\tilde{\mathbf{e}}_t^d = (e_{1t}^d, \ldots, e_{St}^d, e_{dt}^{bmk})$, where e_{st}^d is the sampling error for state s belonging to CD d. The computation of the matrices $\tilde{\boldsymbol{\Sigma}}_{\tau t} = E(\tilde{\mathbf{e}}_t^d \tilde{\mathbf{e}}_t^{d\prime})$ is carried out by expressing the errors $\tilde{\mathbf{l}}_{dt}^{bmk} = \tilde{\boldsymbol{\alpha}}_{dt}^{bmk} - \tilde{\boldsymbol{\alpha}}_{dt}$ when benchmarking the CDs as linear combinations of the concurrent and past CD sampling errors $\tilde{\mathbf{e}}_j = (e_{1j}, \ldots, e_{Dj}, \sum_{d=1}^D w_{dj}e_{dj})'$ and the concurrent and past model residuals $\tilde{\boldsymbol{\eta}}_j = \tilde{\boldsymbol{\alpha}}_j - \tilde{\mathbf{T}}\tilde{\boldsymbol{\alpha}}_{j-1}$ [see (20b)], $j = 1, 2, \ldots, t$. Note that the CD sampling error is a weighted average of the state sampling errors,

$$e_{dj} = y_{dj} - Y_{dj} = \sum_{s \in d} w_{sj}(y_{sj} - Y_{sj}) = \sum_{s \in d} w_{sj}e_{sj}^d. \tag{28}$$

5. ESTIMATION OF UNEMPLOYMENT IN CENSUS DIVISIONS WITH BENCHMARKING

In what follows we apply the benchmarking methodology developed in Section 4.2 for estimating total unemployment in the nine CDs for the period January 1998–December 2003. Very similar results and conclusions are obtained when estimating the CD total employment. The year 2001 is of special interest because it was affected by the start of a recession in March and the attack on the World Trade Center in September. These two events provide a good test for the performance of the benchmarking method.

It was mentioned in Section 3 that the loss in efficiency from using the recursive GLS filter (19) instead of the optimal predictors based on all the individual observations (CPS estimates in our case) is mild (no benchmarking in either case). Table 2 shows for each division the means and standard deviations (STDs) of the monthly ratios between the STD of the GLS and the STD of the optimal predictor when predicting the total unemployment Y_t and the trend levels L_t [see (6)]. As can be seen, the largest loss in efficiency, measured by the means, is 3%.

Next, we combined the individual division models into the joint model (20). The benchmark constraints are as defined

Table 2. Means and STDs (in parentheses) of Monthly Ratios Between the STD of the GLS Predictor and the STD of the Optimal Predictor, 1998–2003

Division	Prediction of unemployment	Prediction of trend
New England	$1.03_{(.002)}$	$1.02_{(.002)}$
Middle Atlantic	$1.02_{(.002)}$	$1.02_{(.002)}$
East North Central	$1.00_{(.001)}$	$1.00_{(.001)}$
West North Central	$1.02_{(.002)}$	$1.02_{(.002)}$
South Atlantic	$1.02_{(.001)}$	$1.02_{(.001)}$
East South Central	$1.00_{(.001)}$	$1.00_{(.001)}$
West South Central	$1.02_{(.001)}$	$1.01_{(.001)}$
Mountain	$1.03_{(.002)}$	$1.03_{(.002)}$
Pacific	$1.02_{(.001)}$	$1.02_{(.001)}$

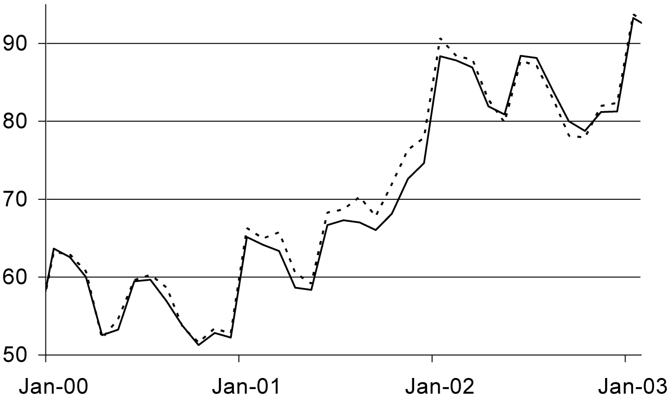

Figure 1. Monthly Total Unemployment, National CPS, and Sum of Unbenchmarked Division Model Estimates (numbers in 100,000) (······ CPS; —— DivModels).

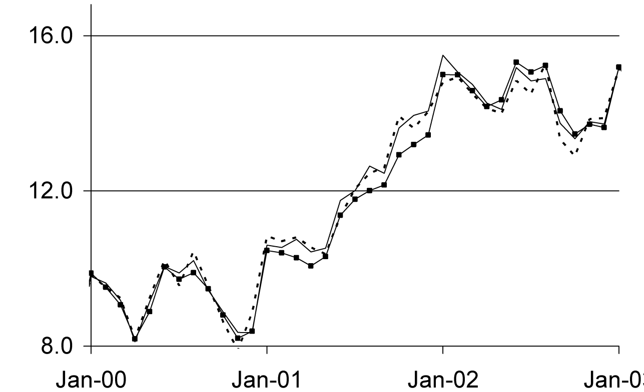

Figure 3. CPS, Benchmarked, and Unbenchmarked Monthly Estimates of Total Unemployment, South Atlantic Division (numbers in 100,000) (······ CPS; —— BMK; —■— UnBMK).

in (1) with $w_{dt} = 1$, such that the model-dependent predictors of the CD total unemployment are benchmarked to the CPS estimator of national unemployment. The CV of the latter estimator is 2%.

Figure 1 compares the monthly sum of the model-dependent predictors in the nine divisions without benchmarking, with the corresponding CPS national unemployment estimate. In the first part of the observation period, the sum of the model predictors is close to the corresponding CPS estimate. In 2001, however, there is evidence of systematic model underestimation, which, as explained earlier, results from the start of a recession in March and the attack on the World Trade Center in September. The bias of the model-dependent predictors is further highlighted in Figure 2, which plots the differences between the two sets of estimators. As can be seen, all of the differences from March 2001 to April 2002 are negative, and in some months the absolute difference is larger than twice the standard error (SE) of the CPS estimator. Successive negative differences are already observed in the second half of 2000, but they are much smaller in absolute value than the differences occurring after March 2001.

Figures 3–5 show the direct CPS estimators, the unbenchmarked predictors, and the benchmarked predictors for three out of the nine census divisions. We restrict the graph to the period January 2000–January 2003, to better illuminate how the benchmarking corrects in real time the underestimation of the unbenchmarked predictors in the year 2001. Similar corrections are observed for the other divisions except New England (not shown), where the benchmarked predictors have a slightly larger positive bias than the unbenchmarked predictors. This is explained by the fact that unlike in the other eight divisions, in this division the unbenchmarked predictors in 2001 are actually higher than the CPS estimators.

Another important conclusion that can be reached from Figures 3–5 is that imposing the benchmark constraints in regular periods affects the predictors only very mildly. This is expected because in regular periods, the benchmark constraints are approximately satisfied under the correct model even without imposing them directly. In fact, the degree to which the unbenchmarked predictors satisfy the constraints can be viewed as a model diagnostic tool. To further illustrate this point, Table 3 gives for each of the nine CDs the means and STDs of the monthly ratios between the benchmarked predictor and the unbenchmarked predictor separately for 1998–2003 excluding 2001 and also for 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%, demonstrating that in normal periods, the effect of benchmarking on the separate model-dependent predictors is indeed very mild.

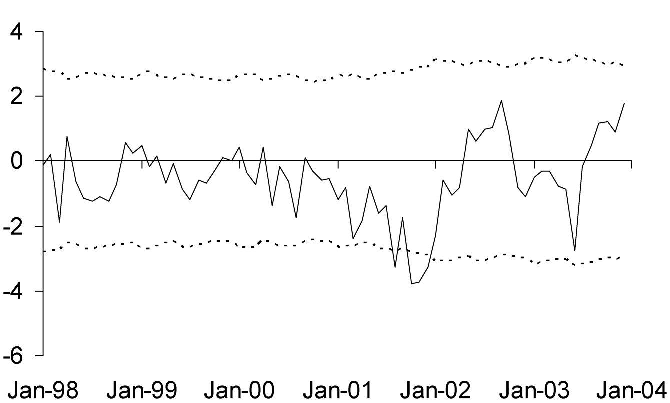

Figure 2. Monthly Total Unemployment, Difference Between Sum of Unbenchmarked Division Model Estimates, and National CPS (numbers in 100,000) [—— difference; ······ + (−)2× SE(CPS)].

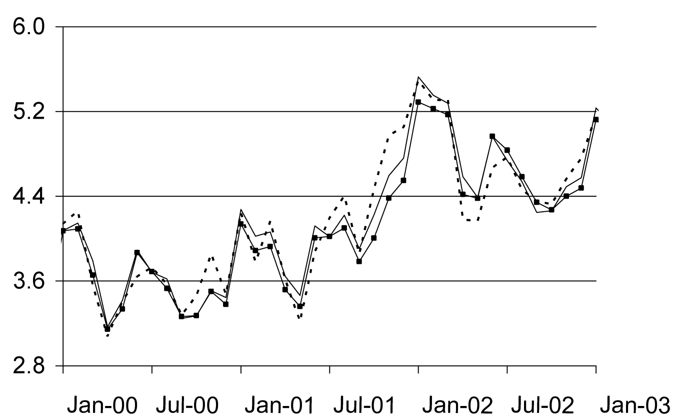

Figure 4. CPS, Benchmarked, and Unbenchmarked Monthly Estimates of Total Unemployment, East South Central Division (numbers in 100,000) (······ CPS; —— BMK; —■— UnBMK).

Downloaded by [Board of Governors of the Federal Reserve System] at 10:31 23 October 2017

Downloaded by [Board of Governors of the Federal Reserve System] at 10:31 23 October 2017

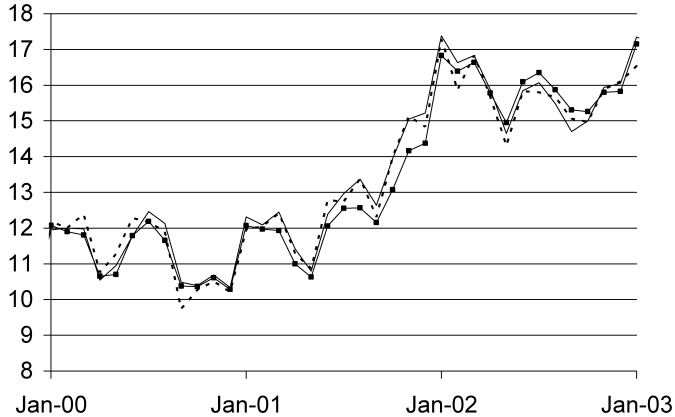

Figure 5. CPS, Benchmarked, and Unbenchmarked Monthly Estimates of Total Unemployment, Pacific Division (numbers in 100,000) (- - - - - - CPS; ——— BMK; —■— UnBMK).

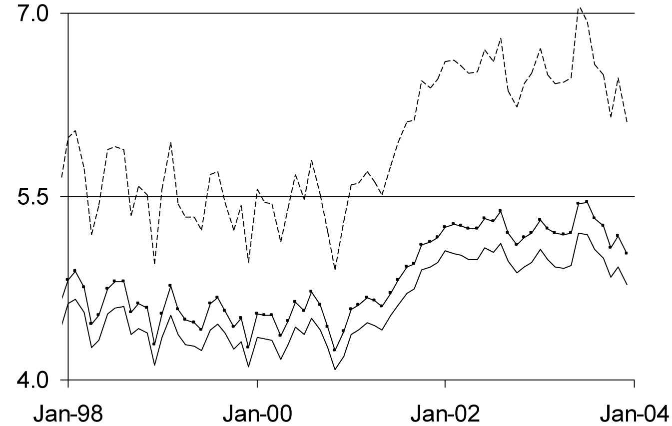

Figure 6. STD of CPS, Benchmarked, and Unbenchmarked Estimates of Total Monthly Unemployment, South Atlantic Division (numbers in 10,000) (------- CPS; ——— BMK; —•— UnBMK).

Finally, in Section 1 we mentioned that by imposing the benchmark constraints, the predictor in any given area "borrows strength" from other areas. The effect of this property is demonstrated by comparing the STDs of the benchmarked predictors with the STDs of the unbenchmarked predictors. Figures 6–8 show the two sets of STDs, along with the STDs of the direct CPS estimators, for the same three divisions as in Figures 3–5. Note that the STDs of the CPS estimators are with respect to the distribution of the sampling errors over repeated sampling, whereas the STDs of the two other predictors also account for the variability of the state vector components. As can be seen, the STDs of the benchmarked predictors are systematically lower than the STDs of the unbenchmarked predictors for all of the months, and both sets of STDs are much lower than the STDs of the corresponding CPS estimators. This pattern repeats itself in all of the divisions. Table 4 further compares the STDs of the benchmarked and unbenchmarked predictors. The mean ratios in Table 4 show gains in efficiency of up to 15% in some of the divisions using benchmarking. The ratios are very stable throughout the years, despite the fact that the STDs of the two sets of the model-based predictors vary between months due to changes in the STDs of the sampling errors; see also Figures 6–8.

6. CONCLUDING REMARKS AND OUTLINE OF FURTHER RESEARCH

Agreement of small-area model–dependent estimators with the direct sample estimate in a "large area" that contains the small areas is a common requirement for statistical agencies producing official statistics. (See, e.g., Fay and Herriot 1979; Battese et al. 1988; Pfeffermann and Barnard 1991; Rao 2003 for proposed modifications to meet this requirement.) The present article has shown how this requirement can be implemented using state–space models. As explained in Section 1, benchmarking constraints of the form (1) or (2) cannot be imposed within the standard Kalman filter, but instead require the development of a filter that produces the correct variances of the benchmarked estimators under the model. The filter developed in this article has the further advantage of being applicable in situations where the measurement errors are correlated over time. The GLS filter developed for the case of correlated measurement errors (without incorporating the benchmark constraints) produces the BLUP of the state vector at time t out of all of the predictors that are linear combinations of the state predictor from time $t-1$ and the new observation at time t. When the measurement errors are independent over time, the GLS filter is the same as the Kalman filter. An important feature of the proposed benchmarking procedure illustrated in Section 5 is that by jointly modeling a large number of areas and imposing the benchmark constraints, the benchmarked predictor in any given area borrows strength from other areas, resulting in reduced variance.

Table 3. Means and STDs (in parentheses) of Ratios Between the Benchmarked and Unbenchmarked 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	$1.00_{(.017)}$	$1.04_{(.020)}$

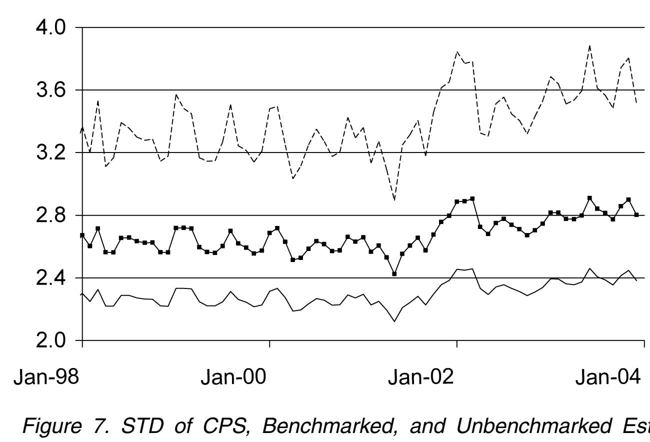

Figure 7. STD of CPS, Benchmarked, and Unbenchmarked Estimates of Total Monthly Unemployment, East South Central Division (numbers in 10,000) (------- CPS; ——— BMK; —■— UnBMK).

Downloaded by [Board of Governors of the Federal Reserve System] at 10:31 23 October 2017

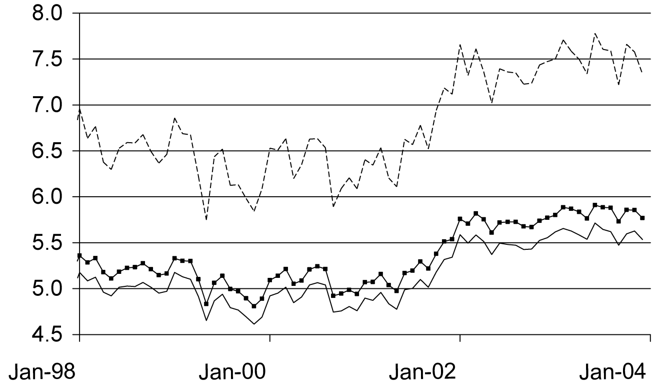

Figure 8. STDs of CPS, Benchmarked, and Unbenchmarked Estimates of Total Monthly Unemployment, Pacific Division (numbers in 10,000) (------- CPS; ——— BMK; —▪— UnBMK).

We are presently studying the benchmarking of the state estimates as outlined in Sections 1 and 4.4. Another important task is the development of a smoothing algorithm that accounts for correlated measurement errors and incorporates the benchmarking constraints. Clearly, as new data accumulate it is desirable to modify past predictors; this is particularly important for trend estimation. An appropriate "fixed-point" smoother has been constructed and is currently being tested.

Finally, the present BLS models assume that the state vectors operating in different CDs or states are independent. It seems plausible, however, that changes in the trend or seasonal effects over time are correlated between neighboring CDs, and even more so between states within the same CD. Accounting for these correlations within the joint model defined by (20) is simple and might further improve the efficiency of the predictors.

APPENDIX A: PROOF OF THE BEST LINEAR UNBIASED PREDICTOR PROPERTY OF THE GENERALIZED LEAST SQUARES PREDICTOR $\hat{\boldsymbol{\alpha}}_t$ [EQ. (18)]

The model holding for $\mathbf{Y}_t = (\hat{\boldsymbol{\alpha}}'_{t|t-1}, \mathbf{y}'_t)'$ is

$$\mathbf{Y}_t = \begin{pmatrix} \hat{\boldsymbol{\alpha}}_{t|t-1} \\ \mathbf{y}_t \end{pmatrix} = \begin{pmatrix} \mathbf{I} \\ \mathbf{Z}_t \end{pmatrix} \boldsymbol{\alpha}_t + \begin{pmatrix} \mathbf{u}_{t|t-1} \\ \mathbf{e}_t \end{pmatrix}, \qquad \mathbf{u}_{t|t-1} = \hat{\boldsymbol{\alpha}}_{t|t-1} - \boldsymbol{\alpha}_t,$$

where $\hat{\boldsymbol{\alpha}}_{t|t-1}$ is the predictor of $\boldsymbol{\alpha}_t$ from time $t-1$. Denote $\mathbf{u}_t = (\mathbf{u}'_{t|t-1}, \mathbf{e}'_t)'$, $\mathbf{X}_t = [\mathbf{I}, \mathbf{Z}'_t]'$. In what follows all of the expectations are over the joint distribution of the observations $\{\mathbf{Y}_t\}$ and the state vectors $\{\boldsymbol{\alpha}_t\}$, so that $E(\mathbf{u}_t) = \mathbf{0}$, $E(\mathbf{u}_t\mathbf{u}'_t) = \mathbf{V}_t$ [see (16)]. A predictor $\boldsymbol{\alpha}^*_t$ is unbiased for $\boldsymbol{\alpha}_t$ if $E(\boldsymbol{\alpha}^*_t - \boldsymbol{\alpha}_t) = \mathbf{0}$.

Table 4. Means and STDs (in parentheses) of Ratios Between STDs of Benchmarked and Unbenchmarked Predictors of Total Unemployment in Census Divisions, 1998–2003

Division	*Means (STDs)*
New England	$.85_{(.013)}$
Middle Atlantic	$.94_{(.005)}$
East North Central	$.96_{(.004)}$
West North Central	$.89_{(.008)}$
South Atlantic	$.96_{(.004)}$
East South Central	$.86_{(.007)}$
West South Central	$.92_{(.006)}$
Mountain	$.88_{(.009)}$
Pacific	$.96_{(.004)}$

Theorem. The predictor $\hat{\boldsymbol{\alpha}}_t = (\mathbf{X}'_t\mathbf{V}_t^{-1}\mathbf{X}_t)^{-1}\mathbf{X}'_t\mathbf{V}_t^{-1}\mathbf{Y}_t = \mathbf{P}_t\mathbf{X}'_t \times \mathbf{V}_t^{-1}\mathbf{Y}_t$ is the BLUP of $\boldsymbol{\alpha}_t$ in the sense of minimizing the prediction error variance out of all the unbiased predictors that are linear combinations of $\hat{\boldsymbol{\alpha}}_{t|t-1}$ and $\mathbf{y}_t$.

Proof. $\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t = \mathbf{P}_t\mathbf{X}'_t\mathbf{V}_t^{-1}(\mathbf{X}_t\boldsymbol{\alpha}_t + \mathbf{u}_t) - \boldsymbol{\alpha}_t = \mathbf{P}_t\mathbf{X}'_t\mathbf{V}_t^{-1}\mathbf{u}_t$, so that $E(\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t) = \mathbf{0}$ and $\operatorname{var}(\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t) = E[(\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t)(\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t)'] = \mathbf{P}_t$. Clearly, $\hat{\boldsymbol{\alpha}}_t$ is linear in $\mathbf{Y}_t$.

Let $\hat{\boldsymbol{\alpha}}^L_t = \mathbf{L}_1\hat{\boldsymbol{\alpha}}_{t|t-1} + \mathbf{L}_2\mathbf{y}_t + \mathbf{l} = \mathbf{L}\mathbf{Y}_t + \mathbf{l}$ be any other linear unbiased predictor of $\boldsymbol{\alpha}_t$ and define $\mathbf{D}_t = \mathbf{L} - \mathbf{P}_t\mathbf{X}'_t\mathbf{V}_t^{-1}$ such that $\mathbf{L} = \mathbf{D}_t + \mathbf{P}_t\mathbf{X}'_t\mathbf{V}_t^{-1}$. Because $\hat{\boldsymbol{\alpha}}^L_t$ is unbiased for $\boldsymbol{\alpha}_t$,

$$\begin{aligned} E(\hat{\boldsymbol{\alpha}}^L_t - \boldsymbol{\alpha}_t) &= E[(\mathbf{D}_t + \mathbf{P}_t\mathbf{X}'_t\mathbf{V}_t^{-1})(\mathbf{X}_t\boldsymbol{\alpha}_t + \mathbf{u}_t) + \mathbf{l} - \boldsymbol{\alpha}_t] \\ &= E[\mathbf{D}_t\mathbf{X}_t\boldsymbol{\alpha}_t + \mathbf{D}_t\mathbf{u}_t + \mathbf{P}_t\mathbf{X}'_t\mathbf{V}_t^{-1}\mathbf{u}_t + \mathbf{l}] \\ &= E[\mathbf{D}_t\mathbf{X}_t\boldsymbol{\alpha}_t + \mathbf{l}] = \mathbf{0}. \end{aligned}$$

Thus for $\hat{\boldsymbol{\alpha}}^L_t$ to be unbiased for $\boldsymbol{\alpha}_t$ irrespective of the distribution of $\boldsymbol{\alpha}_t$, $\mathbf{D}_t\mathbf{X}_t = \mathbf{0}$, and $\mathbf{l} = \mathbf{0}$, which implies that $\operatorname{var}(\hat{\boldsymbol{\alpha}}^L_t - \boldsymbol{\alpha}_t) = \operatorname{var}[(\mathbf{D}_t + \mathbf{P}_t\mathbf{X}'_t\mathbf{V}_t^{-1})(\mathbf{X}_t\boldsymbol{\alpha}_t + \mathbf{u}_t) - \boldsymbol{\alpha}_t] = \operatorname{var}[\mathbf{D}_t\mathbf{u}_t + \mathbf{P}_t\mathbf{X}'_t\mathbf{V}_t^{-1}\mathbf{u}_t]$.

Hence

$$\begin{aligned} \operatorname{var}(\hat{\boldsymbol{\alpha}}^L_t - \boldsymbol{\alpha}_t) &= E[(\mathbf{D}_t\mathbf{u}_t + \mathbf{P}_t\mathbf{X}'_t\mathbf{V}_t^{-1}\mathbf{u}_t)(\mathbf{u}'_t\mathbf{D}'_t + \mathbf{u}'_t\mathbf{V}_t^{-1}\mathbf{X}_t\mathbf{P}_t)] \\ &= \mathbf{D}_t\mathbf{V}_t\mathbf{D}'_t + \mathbf{P}_t\mathbf{X}'_t\mathbf{V}_t^{-1}\mathbf{V}_t\mathbf{D}'_t + \mathbf{D}_t\mathbf{V}_t\mathbf{V}_t^{-1}\mathbf{X}_t\mathbf{P}_t + \mathbf{P}_t \\ &= \mathbf{P}_t + \mathbf{D}_t\mathbf{V}_t\mathbf{D}'_t \\ &= \operatorname{var}(\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t) + \mathbf{D}_t\mathbf{V}_t\mathbf{D}'_t, \quad \text{because } \mathbf{D}_t\mathbf{X}_t = \mathbf{0}. \end{aligned}$$

APPENDIX B: EQUALITY OF THE GENERALIZED LEAST SQUARE FILTER AND THE KALMAN FILTER WHEN THE MEASUREMENT ERRORS ARE UNCORRELATED OVER TIME

Consider the model $\mathbf{y}_t = \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{e}_t, \boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t$, where $\mathbf{e}_t$ and $\boldsymbol{\eta}_t$ are independent white noise series with $\operatorname{var}(\mathbf{e}_t) = \boldsymbol{\Sigma}_t$ and $\operatorname{var}(\boldsymbol{\eta}_t) = \mathbf{Q}_t$. Let $\hat{\boldsymbol{\alpha}}_{t|t-1} = \mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1}$ define the predictor of $\boldsymbol{\alpha}_t$ at time $t-1$, with prediction error variance matrix, $\mathbf{P}_{t|t-1}$, assumed to be positive definite.

GLS setup at time t.

$$\begin{pmatrix} \mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1} \\ \mathbf{y}_t \end{pmatrix} = \begin{pmatrix} \mathbf{I} \\ \mathbf{Z}_t \end{pmatrix} \boldsymbol{\alpha}_t + \begin{pmatrix} \mathbf{u}_{t|t-1} \\ \mathbf{e}_t \end{pmatrix}$$

[see (15)], where now

$$\mathbf{V}_t = \operatorname{var}\begin{pmatrix} \mathbf{u}_{t|t-1} \\ \mathbf{e}_t \end{pmatrix} = \begin{bmatrix} \mathbf{P}_{t|t-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_t \end{bmatrix}$$

[same as (16) except that $\mathbf{C}_t = \mathbf{0}$].

The GLS predictor at time t is

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_t &= \left[(\mathbf{I}, \mathbf{Z}'_t)\mathbf{V}_t^{-1}\begin{pmatrix} \mathbf{I} \\ \mathbf{Z}_t \end{pmatrix}\right]^{-1}(\mathbf{I}, \mathbf{Z}'_t)\mathbf{V}_t^{-1}\begin{pmatrix} \hat{\boldsymbol{\alpha}}_{t|t-1} \\ \mathbf{y}_t \end{pmatrix} \\ &= [\mathbf{P}^{-1}_{t|t-1} + \mathbf{Z}'_t\boldsymbol{\Sigma}_t^{-1}\mathbf{Z}_t]^{-1}[\mathbf{P}^{-1}_{t|t-1}\hat{\boldsymbol{\alpha}}_{t|t-1} + \mathbf{Z}'_t\boldsymbol{\Sigma}_t^{-1}\mathbf{y}_t] \\ &= [\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{Z}'_t\mathbf{F}_t^{-1}\mathbf{Z}_t\mathbf{P}_{t|t-1}][\mathbf{P}^{-1}_{t|t-1}\hat{\boldsymbol{\alpha}}_{t|t-1} + \mathbf{Z}'_t\boldsymbol{\Sigma}_t^{-1}\mathbf{y}_t] \\ &= [\mathbf{I} - \mathbf{P}_{t|t-1}\mathbf{Z}'_t\mathbf{F}_t^{-1}\mathbf{Z}_t]\hat{\boldsymbol{\alpha}}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{Z}'_t\boldsymbol{\Sigma}_t^{-1}\mathbf{y}_t \\ &\quad - \mathbf{P}_{t|t-1}\mathbf{Z}'_t\boldsymbol{\Sigma}_t^{-1}\mathbf{y}_t + \mathbf{P}_{t|t-1}\mathbf{Z}'_t\mathbf{F}_t^{-1}\mathbf{y}_t \\ &= \hat{\boldsymbol{\alpha}}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{Z}'_t\mathbf{F}_t^{-1}(\mathbf{y}_t - \mathbf{Z}_t\hat{\boldsymbol{\alpha}}_{t|t-1}), \\ \mathbf{F}_t &= \mathbf{Z}_t\mathbf{P}_{t|t-1}\mathbf{Z}'_t + \boldsymbol{\Sigma}_t. \end{aligned}$$

The last expression is the same as the Kalman filter predictor (Harvey 1989, eq. 3.2.3a).

APPENDIX C: COMPUTATION OF $\hat{\alpha}_t$ [EQ. (19)]

Consider the GLS predictor

$$\hat{\boldsymbol{\alpha}}_t = \left[(\mathbf{I}, \mathbf{Z}'_t)\mathbf{V}_t^{-1}\begin{pmatrix}\mathbf{I}\\ \mathbf{Z}_t\end{pmatrix}\right]^{-1}(\mathbf{I}, \mathbf{Z}'_t)\mathbf{V}_t^{-1}\begin{pmatrix}\mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1}\\ \mathbf{y}_t\end{pmatrix}$$

[see (18)], where $\mathbf{V}_t = \begin{bmatrix}\mathbf{P}_{t|t-1} & \mathbf{C}_t\\ \mathbf{C}'_t & \boldsymbol{\Sigma}_t\end{bmatrix}$. A familiar result on the inverse of a partitioned matrix applied to the matrix $\mathbf{V}_t$ yields the expressions

$$(\mathbf{I}, \mathbf{Z}'_t)\mathbf{V}_t^{-1} = \big[(\mathbf{P}_{t|t-1}^{-1} + \mathbf{P}_{t|t-1}^{-1}\mathbf{C}_t\mathbf{H}_t\mathbf{C}'_t\mathbf{P}_{t|t-1}^{-1} - \mathbf{Z}'_t\mathbf{H}_t\mathbf{C}'_t\mathbf{P}_{t|t-1}^{-1}), \\ (-\mathbf{P}_{t|t-1}^{-1}\mathbf{C}_t\mathbf{H}_t + \mathbf{Z}'_t\mathbf{H}_t)\big] \quad \text{(C.1)}$$

and

$$\mathbf{P}_t = \left[(\mathbf{I}, \mathbf{Z}'_t)\mathbf{V}_t^{-1}\begin{pmatrix}\mathbf{I}\\ \mathbf{Z}_t\end{pmatrix}\right]^{-1} \\ = \big[\mathbf{P}_{t|t-1}^{-1} + (\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1}\mathbf{C}_t)\mathbf{H}_t(\mathbf{Z}_t - \mathbf{C}'_t\mathbf{P}_{t|t-1}^{-1})\big]^{-1}, \quad \text{(C.2)}$$

where $\mathbf{H}_t = [\boldsymbol{\Sigma}_t - \mathbf{C}'_t\mathbf{P}_{t|t-1}^{-1}\mathbf{C}_t]^{-1}$. It follows from (C.1) and (C.2) that

$$\hat{\boldsymbol{\alpha}}_t = \mathbf{P}_t(\mathbf{I}, \mathbf{Z}'_t)\mathbf{V}_t^{-1}\begin{pmatrix}\mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1}\\ \mathbf{y}_t\end{pmatrix} \\ = \mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1} + \mathbf{P}_t(\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1}\mathbf{C}_t)\mathbf{H}_t(\mathbf{y}_t - \mathbf{Z}_t\mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1}). \quad \text{(C.3)}$$

Computing the matrix in the second row of (C.2) using a standard matrix inversion lemma (Harvey 1989, p. 108) and substituting in the second row of (C.3) yields, after some algebra, equation (19).

APPENDIX D: COMPUTATION OF $\mathbf{P}_t^{bmk} = \mathrm{var}(\tilde{\boldsymbol{\alpha}}_t^{bmk} - \tilde{\boldsymbol{\alpha}}_t)$ AND $\mathbf{C}_t^{bmk} = \mathrm{cov}[\tilde{\mathbf{T}}\tilde{\boldsymbol{\alpha}}_{t-1}^{bmk} - \tilde{\boldsymbol{\alpha}}_t, \tilde{\mathbf{e}}_t]$

The benchmarked predictor $\tilde{\boldsymbol{\alpha}}_t^{bmk}$ defined by (24) can be written as

$$\tilde{\boldsymbol{\alpha}}_t^{bmk} = [\mathbf{I} - (\mathbf{P}_{t|t-1}^{bmk}\tilde{\mathbf{Z}}'_t - \mathbf{C}_{t,0}^{bmk})\mathbf{R}_t^{-1}\tilde{\mathbf{Z}}_t]\tilde{\mathbf{T}}\tilde{\boldsymbol{\alpha}}_{t-1}^{bmk} \\ + (\mathbf{P}_{t|t-1}^{bmk}\tilde{\mathbf{Z}}'_t - \mathbf{C}_{t,0}^{bmk})\mathbf{R}_t^{-1}\tilde{\mathbf{y}}_t, \quad \text{(D.1)}$$

where $\mathbf{R}_t = \tilde{\mathbf{Z}}_t\mathbf{P}_{t|t-1}^{bmk}\tilde{\mathbf{Z}}'_t - \tilde{\mathbf{Z}}_t\mathbf{C}_{t,0}^{bmk} - \mathbf{C}_{t,0}^{bmk\prime}\tilde{\mathbf{Z}}'_t + \tilde{\boldsymbol{\Sigma}}_{tt}^*$. Substituting $\tilde{\mathbf{y}}_t = \tilde{\mathbf{Z}}_t\tilde{\boldsymbol{\alpha}}_t + \tilde{\mathbf{e}}_t$ [see (20a)] in (D.1) and decomposing $\tilde{\boldsymbol{\alpha}}_t = [\mathbf{I} - (\mathbf{P}_{t|t-1}^{bmk}\tilde{\mathbf{Z}}'_t - \mathbf{C}_{t,0}^{bmk})\mathbf{R}_t^{-1}\tilde{\mathbf{Z}}_t]\tilde{\boldsymbol{\alpha}}_t + (\mathbf{P}_{t|t-1}^{bmk}\tilde{\mathbf{Z}}'_t - \mathbf{C}_{t,0}^{bmk})\mathbf{R}_t^{-1}\tilde{\mathbf{Z}}_t\tilde{\boldsymbol{\alpha}}_t$ yields

$$\tilde{\boldsymbol{\alpha}}_t^{bmk} - \tilde{\boldsymbol{\alpha}}_t = [\mathbf{I} - (\mathbf{P}_{t|t-1}^{bmk}\tilde{\mathbf{Z}}'_t - \mathbf{C}_{t,0}^{bmk})\mathbf{R}_t^{-1}\tilde{\mathbf{Z}}_t](\tilde{\mathbf{T}}\tilde{\boldsymbol{\alpha}}_{t-1}^{bmk} - \tilde{\boldsymbol{\alpha}}_t) \\ + (\mathbf{P}_{t|t-1}^{bmk}\tilde{\mathbf{Z}}'_t - \mathbf{C}_{t,0}^{bmk})\mathbf{R}_t^{-1}\tilde{\mathbf{e}}_t. \quad \text{(D.2)}$$

Denote $\mathbf{G}_t = \mathbf{I} - (\mathbf{P}_{t|t-1}^{bmk}\tilde{\mathbf{Z}}'_t - \mathbf{C}_{t,0}^{bmk})\mathbf{R}_t^{-1}\tilde{\mathbf{Z}}_t$ and $\mathbf{K}_t = (\mathbf{P}_{t|t-1}^{bmk}\tilde{\mathbf{Z}}'_t - \mathbf{C}_{t,0}^{bmk})\mathbf{R}_t^{-1}$. Then, by (D.2), the variance matrix of the prediction error $(\tilde{\boldsymbol{\alpha}}_t^{bmk} - \tilde{\boldsymbol{\alpha}}_t)$ under the model (20) is

$$\mathbf{P}_t^{bmk} = E[(\tilde{\boldsymbol{\alpha}}_t^{bmk} - \tilde{\boldsymbol{\alpha}}_t)(\tilde{\boldsymbol{\alpha}}_t^{bmk} - \tilde{\boldsymbol{\alpha}}_t)'] \\ = \mathbf{G}_t\mathbf{P}_{t|t-1}^{bmk}\mathbf{G}'_t + \mathbf{K}_t\tilde{\boldsymbol{\Sigma}}_{tt}\mathbf{K}'_t + \mathbf{G}_t\mathbf{C}_t^{bmk}\mathbf{K}'_t + \mathbf{K}_t\mathbf{C}_t^{bmk\prime}\mathbf{G}'_t, \quad \text{(D.3)}$$

where $\mathbf{P}_{t|t-1}^{bmk} = E[(\mathbf{T}\tilde{\boldsymbol{\alpha}}_{t-1}^{bmk} - \tilde{\boldsymbol{\alpha}}_t)(\mathbf{T}\tilde{\boldsymbol{\alpha}}_{t-1}^{bmk} - \tilde{\boldsymbol{\alpha}}_t)'] = \tilde{\mathbf{T}}\mathbf{P}_{t-1}^{bmk}\tilde{\mathbf{T}}' + \tilde{\mathbf{Q}}$, $\tilde{\boldsymbol{\Sigma}}_{tt} = E(\tilde{\mathbf{e}}_t\tilde{\mathbf{e}}'_t)$ [eq. (20a)], and $\mathbf{C}_t^{bmk} = \mathrm{cov}[\tilde{\mathbf{T}}\tilde{\boldsymbol{\alpha}}_{t-1}^{bmk} - \tilde{\boldsymbol{\alpha}}_t, \tilde{\mathbf{e}}_t]$. The matrix $\mathbf{C}_t^{bmk}$ is computed similarly to (17) using the chain

$$\mathbf{C}_t^{bmk} = \mathbf{A}_{t-1}^{bmk}\mathbf{A}_{t-2}^{bmk}\cdots\mathbf{A}_2^{bmk}\tilde{\mathbf{A}}_1^{bmk}\tilde{\boldsymbol{\Sigma}}_{1t} + \mathbf{A}_{t-1}^{bmk}\mathbf{A}_{t-2}^{bmk}\cdots\mathbf{A}_3^{bmk}\tilde{\mathbf{A}}_2^{bmk}\tilde{\boldsymbol{\Sigma}}_{2t} \\ + \cdots + \mathbf{A}_{t-1}^{bmk}\tilde{\mathbf{A}}_{t-2}^{bmk}\tilde{\boldsymbol{\Sigma}}_{t-2,t} + \tilde{\mathbf{A}}_{t-1}^{bmk}\tilde{\boldsymbol{\Sigma}}_{t-1,t}, \quad \text{(D.4)}$$

where $\mathbf{A}_j^{bmk} = \tilde{\mathbf{T}}\mathbf{G}_j$ and $\tilde{\mathbf{A}}_j^{bmk} = \tilde{\mathbf{T}}\mathbf{K}_j$, with $\mathbf{G}_j$ and $\mathbf{K}_j$ defined as earlier.

[Received August 2004. Revised February 2006.]

REFERENCES

Battese, G. E., Harter, R. M., and Fuller, W. A. (1988), "An Error Component Model for Prediction of County Crop Areas Using Survey and Satellite Data," *Journal of the American Statistical Association*, 83, 28–36.

Doran, H. E. (1992), "Constraining Kalman Filter and Smoothing Estimates to Satisfy Time-Varying Restrictions," *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.

Fay, R. E., and Herriot, R. (1979), "Estimates of Income for Small Places: An Application of James–Stein Procedures to Census Data," *Journal of the American Statistical Association*, 74, 269–277.

Harvey, A. C. (1989), *Forecasting Structural Time Series With the Kalman Filter*, Cambridge, U.K.: 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.

——— (2002), "Small Area Estimation: New Developments and Directions," *International Statistical Review*, 70, 125–143.

Pfeffermann, D., and Barnard, C. H. (1991), "Some New Estimators for Small Area Means With Application to the Assessment of Farmland Values," *Journal of Business & Economic Statistics*, 9, 73–83.

Pfeffermann, D., and Burck, L. (1990), "Robust Small-Area Estimation Combining Time Series and Cross-Sectional Data," *Survey Methodology*, 16, 217–237.

Pfeffermann, D., and Tiller, R. B. (2005), "Bootstrap Approximation to Prediction MSE for State–Space Models With Estimated Parameters," *Journal of Time Series Analysis*, 26, 893–916.

Rao, J. N. K. (2003), *Small Area Estimation*, New York: Wiley.

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.

Downloaded by [Board of Governors of the Federal Reserve System] at 10:31 23 October 2017