

**REAL-TIME STATE-SPACE METHOD FOR COMPUTING  
FILTERED ESTIMATES OF FUTURE REVISIONS:  
APPLIED TO MONTHLY CHAINED CPI**

Peter A. Zadrozny

Division of Price Index Number Research  
Bureau of Labor Statistics  
2 Massachusetts Ave., NE, Room 4915  
Washington, DC 20212  
e-mail: zadrozny.peter@bls.gov

July 28, 2007

**ABSTRACT**

The monthly chained consumer price index has two released estimates: initial (ICPI) and final (FCPI). We focus here solely on chained CPI, which is revised, and not on the better known CPI for urban consumers which is not. ICPI is released every month for that month within BLS and for the previous month publicly. FCPI is released every February for all months two years before. The current or real-time estimation problem addressed here is overcoming the 14-25 month delays in releases of FCPI and in every month computing current or filtered estimates of FCPI for that month, based on all available current and past observations on ICPI and FCPI. Thus, we treat ICPI as having no release delay.

We consider two methods for this problem, which we call a regression method and a state-space method. The regression method is attractive in its technical simplicity but does not fully exploit current sample information. Present results using chained CPI data indicate that the state-space method's estimates of FCPI are 20% more accurate in terms of a root mean-squared error (RMSE) measure of accuracy. The regression method regresses current FCPI on current and past ICPI and, then, estimates FCPI as the estimated regression line evaluated at current and past ICPI. More complexly, the state-space method estimates by maximum likelihood a vector autoregressive moving-average (VARMA) model of jointly generated ICPI and FCPI and, then, applies the missing-data Kalman filter (MDKF) to the estimated model in order to compute filtered estimates of FCPI based on all current and past observations.

The estimation of FCPI is complicated by the fact that, whereas ICPI is released every month, FCPI is released in February two calendar years after it occurs. We handle this data complexity by indexing the data *historically* in order to estimate a model and in *real time* in order to compute filtered estimates of FCPI based on an estimated model. In historical indexing, data are indexed by periods in which they occur and, in real-time indexing, data are indexed by periods in which they are released. The historical form data is the more familiar one, is compact, and generally has few or no missing values. The real-time form data is expansive and always has missing values, in the present case, many missing values. For example, here the historical form data has dimension  $N \times 2$  but the real-time form data has dimension  $N \times 50$ , where  $N$  is the number of sample periods. When the MDKF is used to estimate FCPI, it must be traversed only once, so that applying it for this purpose to data in expansive real-time form and an estimated model in correspondingly expansive state-space form is not computationally intractable or burdensome.