PRELIMINARY AND INCOMPLETE: PLEASE DO NOT QUOTE WITHOUT PERMISSION

INFORMATION IN DATA REVISION PROCESSES: PAYROLL EMPLOYMENT AND REAL-TIME MEASUREMENT OF EMPLOYMENT CONDITIONS

Ellis W. Tallman
Federal Reserve Bank of Atlanta, Atlanta, GA
tel: (404)-498-8915
e-mail: ellis.tallman@atl.frb.org.

Peter A. Zadrozny
Bureau of Labor Statistics, Washington, DC
tel: (202)-691-6591
e-mail: zadrozny_p@bls.gov

January 2005

ABSTRACT

We develop an estimated time-series model of revisions of U.S. payroll employment in order to obtain more accurate filtered estimates of the "true" or underlying condition of U.S. employment. Our estimates of "true" employment are filtered, according to an estimated signal-plus-noise (S+N) model, so as to remove serially correlated observation errors. We are motivated by the perception that raw unfiltered employment estimates based on payroll surveys often overestimate true employment in business-cycle downturns and underestimate it in upturns. Our analysis and estimates operate in real time in the sense that they explicitly account for the timing of initial data releases and revisions and do not simply consider a historical sample of the most revised data as is often done. We view each datum as the sum of a true signal value plus an observation error or noise. Accordingly, we estimate a S+N time-series model, in which each true signal value in the sample is observed multiple times as an initial release followed by revisions, such that the signal and noises are generated by separate autoregressive processes. The signal follows a univariate process and the noises follow a vector process whose dimension depends on the number of vintages of observations in the sample. We use payroll employment data from 1969-2003 to estimate by maximum likelihood an S+N model and use the estimated model to obtain filtered estimates of true employment for each period in the sample. Intuitively, the S+N model structure is sufficiently restrictive to allow us to exploit own- and cross-serial correlations in the data to estimate separate models of the signal and the noises and, thereby, to obtain more accurate estimates of true employment than are indicated directly by raw and unfiltered data.

JEL Classification: E44, C53

*Opinions expressed herein are ours, not necessarily those of the Federal Reserve Bank of Atlanta, Federal Reserve System, or Bureau of Labor Statistics. We thank Dean Croushore and the Federal Reserve Bank of Philadelphia for data and Jim Nason and Juan Rubio-Ramirez for comments.

1. Introduction.

1.1. Motivation.

Business journalists as well as monetary and fiscal policymakers pay close attention to the condition of the labor market, a key indicator of economic activity, which gets special attention during business cycle transition periods. In the United States, the Bureau of Labor Statistics (BLS) produces aggregate labor market indicators from two different sources of labor market information: the household survey, the source data for the unemployment rate, and the establishment survey, the source data for the payroll employment statistics. Historically, business analysts have utilized both aggregate labor market indicators to infer the most recent status of the job market. In economic downturns and recoveries, however, the weaknesses of each series are magnified and their signals of whether employment is growing or shrinking can conflict. Recent research by Kitchen (2003) questions the adequacy of the payroll employment data as an effective real-time measure of true employment during business-cycle transitions. He suggests that processing and reporting delays in released payroll data could introduce biases that would thereby hide changes in “true” labor market conditions, relative to timely releases of other data series that are more quickly available.

We study whether information concealed in payroll employment revisions could be exploited to refine real-time payroll employment estimates. We apply the established statistical technique of estimating a signal from noisy observations, such that both the signal and the noises are assumed to be generated by estimated autoregressive processes (cf., Hillmer and Tiao, 1984). In particular, we consider the estimated signal as representing an estimate of true employment, which is observed noisily in real time in an initial release followed by revisions. New and revised payroll information is combined in an estimated time-series model to obtain a presumably more accurate estimate of the signal or true value of aggregate employment. The method should be especially useful in business-cycle transitions, such as in the past two cycles, during which unfiltered estimates of employment have been considered poor indicators of true employment.

The paper first outlines the elements of the payroll employment survey and the institutional explanation for the lag in final payroll reports. Next, we illustrate graphically under and over estimation of payroll employment during recession and recovery periods, along with the magnitude and variability in the revision process. We show that most data revisions occur within 24 months of a benchmark revision, although benchmark revisions following a decadal census can extend beyond 24 months. The graphs show that the signal from revisions may be particularly important during business cycle contractions. Then, we state a Kalman filtering based method for extracting information from the historical revisions. The resulting estimated state-space model produces time-series estimates of the underlying employment conditions, which are measured imperfectly in real time by the payroll employment survey. To date, this approach offers signal estimates of true employment with accuracy comparable to real-time household data. Currently, the model is set up to only to handle month-to-month revisions of the Current Employment Statistics (CES) and does not yet handle the annual benchmark revisions. In future work, we plan also to model benchmark revisions. In such an extension, we could possibly also incorporate other real-time data series, like household employment, into the state-space model to enhance the real-time signal for the employment market.

1.2. Background.

Kitchen (2003) suggests that real-time initial releases of household survey employment provide a noisy yet reliable signal about underlying labor market conditions. For example, his chart 3e shows that the real time household employment measure matches the most recently revised payroll data for 1990-93, the present best estimate of true employment. In contrast, real-time payroll employment data are much different from final revised payroll employment data. Also, revisions in payroll employment reflect notable fluctuations in perceived employment conditions from successive benchmark revisions.

Accepting Kitchen’s conclusion, one might conclude that the household survey produces the more reliable real time signals of current and recent employment conditions. Nevertheless, released BLS payroll data aims to be the best gauge of the employment market. BLS payroll employment data has other properties, such as a larger sample and eventual benchmarking to a population estimate, that make it the preferred indicator in discussions of labor market conditions. Also, there may be unexploited information in the serial correlations of revisions for the payroll employment.

The household employment estimate is taken from the household survey, a statistical sample of 50,000 households. Typically, the survey is revised only in terms of its level, based on the Census of Population. This revision does not affect the data’s turning points, a key point noted by Kitchen. The household survey data are used mainly to produce the unemployment rate and other ratios such as the labor force participation rate and the employment to population ratio. Nevertheless, the statistical sample for household employment has been used to estimate aggregate employment, an estimate that Kitchen considers useful, especially in providing a real-time signal of cyclical business cycle conditions.

The CES survey, typically called the payroll survey, is initially a random sample of firms from a population of firms registered with the Internal Revenue Service that pay unemployment insurance (UI). Using a sample of establishments to estimate the CES payroll employment allows BLS to produce an estimate of payroll employment that is often released with less than a week’s time lag. Each year, BLS releases a benchmark revision of payroll employment that matches population data. U.S. population data, collected with by State Employment Security agencies, includes about 8 million business establishments of all registered firms and more fully reflects the aggregate labor market. The benchmark revision updates payroll employment to incorporate information on the level of the population in March of each year.[1]

---

[1]The payroll survey is separate from the population data. BLS uses population data to calculate business employment dynamics data, which emphasize the sources of changes in employment levels. Although attractive as population data, business employment dynamics data are released as long as three

We can explain why signals from payroll employment data failed to describe accurately the US labor market in the early 1990s because payroll employment data did not account for job losses caused by establishment closures. After several months, BLS investigated non-reporting establishments.[2] In addition, there was difficulty accounting for the level of employment increases caused by creation of new firms, i.e. births of firms. Although Business Employment Dynamics data for 1991-1992 are not easily available for verification, it has been suggested that employment was underestimated in the June 1993 benchmark revision because it failed to account sufficiently for employment created by new firms.

Recently, BLS attempted to overcome the inherent inability to account for net new firm changes in real-time by estimating a model of net employment changes over the past 5 years using the Business Employment Dynamics data. Although these methods should improve payroll estimates by reducing their error, it remains likely that lags in accounting for net changes in firms will still persist, although perhaps less than in the past. In effect, models of net job changes are unlikely to account fully for net changes in employment at turning points.[3]

The data revisions that followed the 1990-91 recession are likely to have reinforced the perception that raw payroll employment is an unreliable real-time signal of true employment during business cycle transitions. Figure 1 displays peak-to-peak payroll employment in the period after the 1990-1991 recession. The three lines report three different vintages of employment: May 1992, May 1993, and June 1993. The June 1993 vintage includes a benchmark revision that indicates that previously reported employment undercounted creation of new jobs.

---

quarters after payroll data are published and released. For example, although the payroll survey for April 2004 is available in early May 2004, at that same time the business employment dynamics data are available only through the second quarter of 2003.

[2]BLS currently uses a sampling procedure to overcome this problem. The exact wording on the BLS website is: "Effectively, business deaths are not included in the sample-based link portion of the estimate, and the implicit imputation of their previous month's employment is assumed to offset a portion of the employment associated with births. There is an operational advantage associated with this approach as well. Most firms will not report that they have gone out of business; rather, they simply cease reporting and are excluded from the link, as are all other non-respondents. As a result, extensive follow-up with monthly non-respondents to determine whether a company is out-of-business or simply did not respond is not required."

[3]Until March 26, 2003, there was no indication that net new job creation from firm births less deaths had changed measurably following the March 2003 benchmark revision. In addition, net changes in employment levels from net firm births less deaths followed a different pattern in the 1993-94 recoveries.

Presumably, ex-post, the revision reduced uncertainty about the state of the US economy in 1991-1993. The May 1993 data incorporates a benchmark revision that took place in June 1992 and reflects a downward revision relative to May 1992 payroll employment. Much of that downward revision was offset by the June 1993 benchmark revision.

Nevertheless, the sizes of the revisions in Figure 1 are small. Even revisions of 1 million jobs represent less than one percent of the employment levels of about 110 million persons. The purpose of this study is to exploit the temporal correlations of revisions in order to obtain more accurate employment estimates, regardless of the average sizes and directions of individual revisions.

1.3. The Process of Revising Payroll Employment Data.

As mentioned above, the revision of payroll employment has two components. The first revision process occurs monthly at the same frequency as the initial releases, consisting of correcting for the delayed responses of surveyed firms. The initial release of a previous month's payroll employment includes information from about 65 percent of surveyed firms. The first revision, released after another month passes, includes about 80 percent of surveyed firms. After a third month passes, employment is further revised to include information from about 90 percent of surveyed firms. For example, in May of a year BLS releases an initial estimate of April employment, a first revision of March employment, and a second revision of February employment.

The second revision process of payroll employment reflects less frequent, annual and decadal, revisions introduced by benchmarking payroll employment to an estimate of "population" employment. The benchmarking process matches the level of estimated payroll employment to the March level of the population data for the previous year. The difference between survey- and population-based estimates is divided evenly between months before and after the March benchmark. Benchmarking may produce large changes in the level of estimated employment, which would often be insignificant as month-to-month changes.

The CES payroll survey is a large sample that covers about 37 percent of establishments of US firms. Although a large sample indicates small sampling errors, benchmarking introduces information that may be overlooked either by sample composition relative to establishment population changes or by information from late survey submissions defined as the last 10 percent of the survey.

Non-sampling errors arise in payroll employment from three sources: (1) coverage or composition of the sample, (2) response of the last 10 percent of the survey respondents, and (3) processing errors mentioned by BLS, but not considered in this discussion until now. From the perspective of real time business cycle analysis, one error in payroll employment is not including employment of newly created establishments. This fact is known and BLS has advanced in overcoming this gap. As noted above, this gap in payroll employment becomes more prominent at business cycle troughs. However, probability weight sampling and forecasting net new jobs from births and deaths of establishments reduces this gap and improves estimates.

Benchmarking matches the level of payroll employment based on the CES survey to the estimated employed population the previous March. The difference between the most recent benchmark and previous March estimates is spread evenly or linearly over the previous 12 months.[4] This technique attributes one twelfth of the level difference to each of the prior 12 months (see the BLS website named "Benchmark Article"). Sometimes, the benchmark revisions significantly change estimated employment up to 24 months earlier. Decadal revisions that incorporate information in the decadal Census can significantly change estimates of employment even further back.

If we focus mainly on the survey-response error rate, then, we can limit the analysis to the first three data releases, the initial release and two subsequent monthly revisions. Such an approach greatly simplifies the analysis, although it limits the revision information that the study considers. As a first pass of the method, we consider only two revisions on the presumption that including additional revisions would not significantly alter estimates of true employment, although, in the future,

[4]See BLS web page on employment for a description of the linear "wedge back" procedure.

we shall consider adding revisions going further back, including benchmark revisions.[5]

2. Signal-Plus-Noise Model.

2.1. Structure of the Model.

The S+N model described here extends the S+N model of three noisy observations per period on a variable, considered by Chen and Zadrozny (2001), to the general case of any number of m noisy observations per period on a variable. Throughout, the following three sets of terms are synonymous: "observations", "data", and "estimates"; "true" and "signal"; and "observation errors" and "noises".

Let $y_t = (y_{1,t}, \ldots, y_{m,t})^T$ and $u_t = (u_{1,t}, \ldots, u_{m,t})^T$ denote $m\times1$ vectors of observations and their unobserved observation errors or noises, at vintages v = 1, ..., m, on unobserved true values or signals of a scalar variable, denoted y^*_t, in sample periods t = 1, ..., T (superscript T denotes vector or matrix transposition). Thus, $y_{v,t}$ denotes a so-called vintage v observation in period t on y^*_{t-v+1} made with noise $u_{v,t}$. For now, to simplify notation, we assume that there are no missing observations in any sample period t, so that in every period the observation vector y_t contains observations on all vintages 1, ..., m.

For all periods t = 1, ... T, the observations, signals, and noises are linked by the scalar observation equations $y_{v,t} = y^*_{t-v+1} + u_{v,t}$ or equivalently by the $m\times1$ vector observation equation

$$(2.1) \qquad y_t = (1, \ldots, L^{m-1})^T y^*_t + u_t,$$

where L denotes the lag operator whose multiplication of a variable moves it back one period. It simplifies notation to allow the first element, $y_{1,t}$, in the observation vector, y_t, to be contemporaneous with

[5] Harvey and Chung (2000) employ the state-space form and a Kalman filter to estimate the underlying change in unemployment in the UK. Their application exploits the structure of the sampling design for the data series to aid the design and estimation of the time-series model. In addition, their application also introduces an additional data series that is released in a more timely manner to improve the forecast accuracy of the estimated model.

the first element, y_t^*, in the vector of signals $(1, ..., L^{m-1})^T y_t^*$. Only relative lagged positions of elements of $(1, ..., L^{m-1})^T y_t^*$ are relevant, so that $(1, ..., L^{m-1})^T$ could be multiplied by any positive or negative power of L and its inverse merged into a redefinition of y_t^*. Whether y_t^* represents a current, past, or future signal depends on the application. We proceed as if y_t^* represents a current signal.

In an application, when a model is estimated using maximum likelihood, some or all elements of y_t could be missing in some sample periods. However, this causes no problems if, as we propose, the likelihood function is formed using an appropriate missing-data version of the Kalman filter (MDKF). We could algebraically describe the correct handling of missing-data with an appropriate MDKF (Zadrozny, 1988, 1990) and this algebra could be implemented variously. When some observations are missing in a period, the rows of equation (2.1) with missing values are deleted and the standard non-missing-data Kalman filter (Anderson and Moore, 1979) is correspondingly reduced. For example, if all vintages are observed in periods 2, ..., T, but only vintages 2, ..., m-1 are observed in period 1, then, in period 1, equation (2.1) becomes $(y_{2,1}, ..., y_{m-1,1})^T = (L, ..., L^{m-2})^T y_t^* + (u_{2,1}, ..., u_{m-2,1})^T$ and, in periods 2, ..., T, equation (2.1) is unchanged.

The data may be visualized in the following T×m data matrix:

Table 1: Data Matrix Indexed by Sample Periods and Vintages.

$y_{v,t}$, v = vintage, t = sample period						
$y_{1,1}$	$y_{2,1}$	$y_{3,1}$	$y_{4,1}$	$y_{5,1}$	...	$y_{m,1}$
$y_{1,2}$	$y_{2,2}$	$y_{3,2}$	$y_{4,2}$	$y_{5,2}$	...	$y_{m,2}$
$y_{1,3}$	$y_{2,3}$	$y_{3,3}$	$y_{4,3}$	$y_{5,3}$	...	$y_{m,3}$
$y_{1,4}$	$y_{2,4}$	$y_{3,4}$	$y_{4,4}$	$y_{5,4}$	...	$y_{m,4}$
$y_{1,5}$	$y_{2,5}$	$y_{3,5}$	$y_{4,5}$	$y_{5,5}$	...	$y_{m,5}$
...	...	...	...	...	...	...
$y_{1,T}$	$y_{2,T}$	$y_{3,T}$	$y_{4,T}$	$y_{5,T}$	...	$y_{m,T}$

In the table, row 1 contains period 1 observations $y_1 = (y_{1,1}, ..., y_{m,1})^T$

on signals y_1^*, ..., y_{2-m}^* at vintages 1, ..., m, row 2 contains period 2 observations $y_2 = (y_{1,2}, ..., y_{m,2})^T$ on signals y_2^*, ..., y_{3-m}^* at vintages 1, ..., m, and so forth; column 1 contains T first observations $(y_{1,1}, ..., y_{1,T})^T$ on y_1^*, ..., y_T^*, column 2 contains T second observations $(y_{2,1}, ..., y_{2,T})^T$ on y_0^*, ..., y_{T-1}^*, and so forth. Going from left to right and top to bottom in table 1, namely, in the order $y_{1,1}$, ..., $y_{m,1}$, ..., $y_{1,T}$, ..., $y_{m,T}$, is a natural order for inputting data from a storage file into a program for estimating the model.

We assume signals, y_t^*, are generated by the scalar autoregressive moving-average model,

(2.2) $$y_t^* = \alpha_1 y_{t-1}^* + ... + \alpha_{p_1} y_{t-p_1}^* + \varepsilon_t^* + \beta_1 \varepsilon_{t-1}^* + ... + \beta_{q_1} \varepsilon_{t-q_1}^*,$$

denoted ARMA(p_1,q_1), with scalar autoregressive coefficients, α_1, ..., α_{p_1}, scalar moving-average coefficients, β_1, ..., β_{q_1}, and scalar disturbance, ε_t^*, distributed normally, independently, identically, with mean zero, and constant variance σ_ε^2 or $\varepsilon_t^* \sim NIID(0,\sigma_\varepsilon^2)$.

We make the following *basic assumptions* on the parameters of signal model (2.2): (i) ARMA degrees p_1 and q_1 are finite and nonnegative integers, such that min(p_1,q_1) = 1, but max(p_1,q_1) could be <, =, or > m, the maximum number of observations per period; (ii) model (2.2) could be stationary or nonstationary but is invertible, which means that any complex number, λ, which satisfies $1 + \beta_1\lambda + ... + \beta_{p_1}\lambda^{p_1} = 0$ also satisfies $|\lambda| > 1$; and, (iii) $\sigma_\varepsilon^2 > 0$.

Similarly, we assume noises, u_t, are generated by the m$\times$1 vector autoregressive moving-average model,

(2.3) $$u_t = A_1 u_{t-1} + ... + A_{p_2} u_{t-p_2} + \eta_t + B_1 \eta_{t-1} + ... + B_{q_2} \eta_{t-q_2},$$

denoted VARMA(p_2,q_2), with $m \times m$ autoregressive coefficient matrices, A_1, ..., A_{p_2}, $m \times m$ moving-average coefficient matrices, B_1, ..., B_{q_2}, and $m \times 1$ disturbance vector, η_t, distributed NIID(0,Ω) and independently of ε_t^*.

We make the following *basic assumptions* on the parameters of noise model (2.3): (iv) VARMA degrees p_2 and q_2 are finite and nonnegative integers, such that min(p_2,q_2) = 1, but max(p_2,q_2) could be <, =, or > m; (v) model (2.3) is stationary, which means that any complex number, λ, which satisfies $\det[I_m - A_1\lambda - \ldots - A_{p_2}\lambda^{p_2}] = 0$ also satisfies $|\lambda| > 1$, where $\det[\cdot]$ denotes the determinant of a square matrix and I_m denotes the $m \times m$ identity matrix; (vi) model (2.3) is invertible, which means that any complex number, λ, which satisfies $\det[I_m + B_1\lambda + \ldots + B_{q_2}\lambda^{q_2}] = 0$ also satisfies $|\lambda| > 1$; and, (vii) Ω is positive definite, which is denoted by $\Omega > 0$.

Thus, we have made *basic assumptions* (i)-(vii) on S+N model (2.1)-(2.3). In (ii), we assume signal model (2.2) is nonstationary, but, in (v), we assume noise model (2.3) is stationary, so that any observed nonstationarity arises in the signal model. In practice, we expect any observed nonstationarity can be accounted for by unit autoregressive roots in the signal model.

Equations (2.2)-(2.3) purposely have no constant terms. In the absence of constant terms, if true model (2.2) is stationary, equations (2.2)-(2.3) imply that $Ey_t = 0$. Thus, we assume that all data have been normalized before being used to estimate a model. This is the simpler way to proceed because it avoids estimating the constant terms jointly with the other ARMA and VARMA parameters. Strictly, when the true model and the data are nonstationary, the means of the data do not exist, but we ignore this subtlety and always normalize the data before estimating a model. In essence, the normalization can be viewed as only a temporary translation and rescaling of the data to facilitate estimation, so that the estimated model can be transformed back to the form of the original unnormalized data. In section 2.3, we further discuss estimation of mean values when we discuss identification of parameters.

2.2 State-Space Representation of the Signal-Plus-Noise Model.

We now restate equations (2.1)-(2.3) as an single overall state-space representation in terms of the state vector x_t.

Let $y_t^{**} = (y_{1,t}^{**}, \ldots, y_{s_1,t}^{**})^T = (y_t^*, \ldots, y_{t-r_1+1}^*, \varepsilon_t^*, \ldots, \varepsilon_{t-q_1+1}^*)$ be an s_1x1 state vector for true model (2.2), where $s_1 = r_1 + q_1$ and $r_1 = \max(m,p_1)$. In terms of y_t^{**}, observation equation (2.1) is

(2.4) $$y_t = My_t^{**} + u_t,$$

where $M = [I_m, 0_{m\times(s_1-m)}]$ and $0_{m\times(s_1-m)}$ denotes the $m\times(s_1\text{-}m)$ zero matrix. When $m \geq p_1$ and $q_1 = 0$, as in the application in section 3, $M = I_m$.

Let $\varepsilon_t^{**} = (\varepsilon_{1,t}^*, \ldots, \varepsilon_{s_1,t}^*)^T$ denote an $s_1\times1$ disturbance vector whose first element $\varepsilon_{1,t}^* \equiv \varepsilon_t^*$, the disturbance in true model (2.2), and whose remaining elements are "almost" identically equal to zero in the following sense. Theoretically, we would like the second to last elements of ε_t^{**} to be identically equal to zero. Practically, ε_t^{**} is part of the overall state-space representation to which a Kalman filter will be applied and we can generally guarantee the filter's numerical accuracy only if the covariance matrix of ε_t^{**} is positive definite. Thus, we assume that $\varepsilon_t^{**} \sim NIID(0,\Sigma_\varepsilon^{**})$, where $\Sigma_\varepsilon^{**} = \sigma_\varepsilon^2 e_{1,s_1} e_{1,s_1}^T + \delta I_{s_1}$, $e_{1,s_1} = (1, 0, \ldots, 0)^T$ is the $s_1\times1$ vector with one in position one and zeroes elsewhere, and δ is a small positive number, small enough not to noticeably affect Kalman filtering with the overall state-space representation, but large enough to guarantee the numerical accuracy of the filtering. [Note to Ellis: think carefully about the role of δ in the estimation. The gist is to use it as a trick to guarantee positive definite covariance matrix for the signal process error. Look at the estimate of the value from the empirical estimates of the model.]

Then, the following state equation in y_t^{**} incorporates true model (2.2) as its first element,

$$(2.5) \qquad y_t^{**} = A_1^* y_{t-1}^{**} + B_0^* \varepsilon_t^{**}, \quad A_1^* = \begin{bmatrix} A_{1,11}^* & A_{1,12}^* \\ A_{1,21}^* & A_{1,22}^* \end{bmatrix},$$

$$A_{1,11}^* = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & . & . & . & \alpha_{r_1} \\ 1 & 0 & 0 & . & . & . & 0 \\ 0 & 1 & . & . & & & . \\ . & 0 & . & . & . & & . \\ . & & . & . & . & . & . \\ . & & & . & . & . & 0 \\ 0 & . & . & . & 0 & 1 & 0 \end{bmatrix}, \quad A_{1,12}^* = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & . & . & . & \beta_{q_1} \\ 1 & 0 & 0 & . & . & . & 0 \\ 0 & 1 & . & . & & & . \\ . & 0 & . & . & . & & . \\ . & & . & . & . & . & . \\ . & & & . & . & . & 0 \\ 0 & . & . & . & 0 & 1 & 0 \end{bmatrix},$$

$$A_{1,21}^* = 0_{q_1 \times r_1}, \quad A_{1,22}^* = \begin{bmatrix} 0 & 0 & 0 & . & . & . & 0 \\ 1 & 0 & 0 & . & . & . & 0 \\ 0 & 1 & . & . & & & . \\ . & 0 & . & . & . & & . \\ . & & . & . & . & . & . \\ . & & & . & . & . & 0 \\ 0 & . & . & . & 0 & 1 & 0 \end{bmatrix}, \quad B_0^* = \begin{bmatrix} e_{1,r_1} e_{1,r_{1_1}}^T & 0_{r_{1_1} \times q_1} \\ e_{1,q_1} e_{1,r_1}^T & 0_{q_1 \times q_1} \end{bmatrix},$$

where $A_{1,11}^*$, $A_{1,12}^*$, and $A_{1,22}^*$, have the dimensions $r_1 \times r_1$, $r_1 \times q_1$, and $q_1 \times q_1$.

Stacking equations (2.4), (2.5), and (2.3) on top of each other implies

(2.6)

$$\begin{bmatrix} I_m & -M & -I_m \\ 0_{s_1 \times m} & I_{s_1} & 0_{s_1 \times m} \\ 0_{m \times m} & 0_{m \times s_1} & I_m \end{bmatrix} \begin{bmatrix} y_t \\ y_t^{**} \\ u_t \end{bmatrix} = \begin{bmatrix} 0_{m \times m} & 0_{m \times s_1} & 0_{m \times m} \\ 0_{s_1 \times m} & A_1^* & 0_{s_1 \times m} \\ 0_{m \times m} & 0_{m \times s_1} & A_1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-1}^{**} \\ u_{t-1} \end{bmatrix} + \quad ...$$

$$+ \begin{bmatrix} 0_{m \times m} & 0_{m \times s_1} & 0_{m \times m} \\ 0_{s_1 \times m} & 0_{s_1 \times s_1} & 0_{s_1 \times m} \\ 0_{m \times m} & 0_{m \times s_1} & A_{p_2} \end{bmatrix} \begin{bmatrix} y_{t-p_2} \\ y_{t-p_2}^{**} \\ u_{t-p_2} \end{bmatrix}$$

$$+ \begin{bmatrix} \varsigma_t \\ B_0^* \varepsilon_t^{**} \\ u_t \end{bmatrix} + \begin{bmatrix} 0_{m \times m} & 0_{m \times s_1} & 0_{m \times m} \\ 0_{s_1 \times m} & 0_{s_1 \times s_1} & 0_{s_1 \times m} \\ 0_{m \times m} & 0_{m \times s_1} & B_1 \end{bmatrix} \begin{bmatrix} \varsigma_{t-1} \\ B_0^* \varepsilon_{t-1}^{**} \\ u_{t-1} \end{bmatrix} + \ ... \ + \begin{bmatrix} 0_{m \times m} & 0_{m \times s_1} & 0_{m \times m} \\ 0_{s_1 \times m} & 0_{s_1 \times s_1} & 0_{s_1 \times m} \\ 0_{m \times m} & 0_{m \times s_1} & B_{q_2} \end{bmatrix} \begin{bmatrix} \varsigma_{t-q_2} \\ B_0^* \varepsilon_{t-q_2}^{**} \\ u_{t-q_2} \end{bmatrix},$$

where ζ_t is an $m\times1$ disturbance vector with a role analogous to the second to last elements of ε_t^{**}. In theory, we want ζ_t to be identically equal to zero, but, in practice, we want its covariance matrix to have some positive definiteness. Thus, we assume $\zeta_t \sim NIID(0,\Sigma_\zeta)$, where $\Sigma_\zeta = \delta I_m$ and δ is a small positive number, such that this δ can be identical to the previously introduced one in Σ_ε^{**}.

Let $z_t = (y_t^T,\ y_t^{**T},\ u_t^T)^T$ be a $(2m+s_1)\times1$ vector and let $\xi_t = ((\zeta_t + MB_0^*\varepsilon_t^{**} + \eta_t)^T,\ B_0^*\varepsilon_t^{**T},\ \eta_t^T)^T$ be its $(2m+s_1)\times1$ innovation vector. The inverse of the leading matrix in equation (2.6) is just the same matrix with the minus signs deleted. Thus, we can write equation (2.6) equivalently as

$$(2.7)\qquad z_t = F_1 z_{t-1} + \ldots + F_{p_2} z_{t-p_2} + \xi_t + G_1\xi_{t-1} + \ldots + G_{q_2}\xi_{t-q_2},$$

$$F_i = \begin{bmatrix} 0_{m\times m} & MA_i^* & A_i \\ 0_{s_1\times m} & A_i^* & 0_{s_1\times m} \\ 0_{m\times m} & 0_{m\times s_1} & A_i \end{bmatrix},\quad G_j = \begin{bmatrix} 0_{m\times m} & 0_{m\times s_1} & B_j \\ 0_{s_1\times m} & 0_{s_1\times s_1} & 0_{s_1\times m} \\ 0_{m\times m} & 0_{m\times s_1} & B_j \end{bmatrix},$$

where $A_i^* = 0_{s_1\times s_1}$ for $i > 1$, $F_i = 0_{(2m+s_1)\times(2m+s_1)}$ for $i > p_2$, $G_j = 0_{(2m+s_1)\times(2m+s_1)}$ for $j > q_2$, $\xi_t \sim NIID(0,\Sigma_\xi)$, and

$$(2.8)\qquad \Sigma_\xi = \begin{bmatrix} \delta I_m + MB_0^*(\sigma_\varepsilon^2 e_{1,s_1} e_{1,s_1}^T + \delta I_{s_1})B_0^{*T}M^T + \Omega & MB_0^*(\sigma_\varepsilon^2 e_{1,s_1} e_{1,s_1}^T + \delta I_{s_1})B_0^{*T} & \Omega \\ B_0^*(\sigma_\varepsilon^2 e_{1,s_1} e_{1,s_1}^T + \delta I_{s_1})B_0^{*T}M^T & B_0^*(\sigma_\varepsilon^2 e_{1,s_1} e_{1,s_1}^T + \delta I_{s_1})B_0^{*T} & 0_{s_1\times m} \\ \Omega & 0_{m\times s_1} & \Omega \end{bmatrix}.$$

As desired, Σ_ξ is positive definite ($\Sigma_\xi > 0$) and, when $m \geq r_1$ and $q_1 = 0$, Σ_ξ reduces to

$$(2.9)\qquad \Sigma_\xi = \begin{bmatrix} \delta I_m + (\sigma_\varepsilon^2+\delta)e_{1,m}e_{1,m}^T + \Omega & (\sigma_\varepsilon^2+\delta)e_{1,m}e_{1,m}^T & \Omega \\ (\sigma_\varepsilon^2+\delta)e_{1,m}e_{1,m}^T & (\sigma_\varepsilon^2+\delta)e_{1,m}e_{1,m}^T & 0_{m\times m} \\ \Omega & 0_{m\times m} & \Omega \end{bmatrix}.$$

Let $x_t = (x_{1,t}^T, \ldots, x_{r_2,t}^T)^T$ denote the $n\times1$ overall state vector, partitioned into r_2 subvectors $x_{i,t}$ of dimension $\nu\times1$, where $r_2 = \max(p_2,q_2+1)$ and $\nu = 2m+s_1$, so that $n = r_2\cdot\nu$. Following Ansley and Kohn (1983), VARMA representation (2.7)-(2.8) has the state-space representation, with the observation equation

$$(2.10)\qquad y_t = Hx_t,$$

with no observation error, where $H = [I_m, 0_{m\times(n-m)}]$, and the state equation

$$(2.11)\qquad x_t = Fx_{t-1} + G\xi_t,$$

$$F = \begin{bmatrix} F_1 & I_\nu & 0_{\nu\times\nu} & . & . & . & 0_{\nu\times\nu} \\ F_2 & 0_{\nu\times\nu} & . & . & & & . \\ . & . & . & . & . & & . \\ . & . & & . & . & . & . \\ . & . & & & . & . & 0_{\nu\times\nu} \\ . & . & & & & . & I_\nu \\ F_{r_2} & 0_{\nu\times\nu} & . & . & . & . & 0_{\nu\times\nu} \end{bmatrix}, \quad G = \begin{bmatrix} I_\nu \\ G_1 \\ . \\ . \\ . \\ G_{r_2-1} \end{bmatrix},$$

where $\xi_t \sim \mathrm{NIID}(0,\Sigma_\xi)$ and Σ_ξ is given by equation (2.8).

The objective is to estimate y_t^*. To do this using the Kalman filter, we need to know where y_t^* is in x_t. Backwards recursive substitution in equation (2.11), in the order $x_{r_2,t}, \ldots, x_{1,t}$, shows that $x_{1,t} = z_t = (y_t^T, y_t^{**T}, u_t^T)^T$ and y_t^* is the first element of y_t^{**}. Thus, because y_t is $m\times1$, y_t^* is element $m+1$ of x_t. This holds in general, regardless of the values of m, p_1, p_2, q_1, and q_2.

Overall state-space representation (2.10)-(2.11) is inefficient because x_t and ξ_t could be made much smaller by eliminating zeroes in the coefficient matrices F and G. However, we use this representation because it has the structure required by the FORTRAN estimation program that we use to estimate the S+N model and to estimate the true values y_t^*. Because computers operate so quickly today, it is not worth the trouble to rewrite the estimation program in terms of a more concise version of representation (2.10)-(2.11).

2.3. Identification of Structural Parameters.

We now state assumptions for identifying the structural parameters of S+N model (2.1)-(2.3) and prove that they are sufficient for this purpose. Unless the parameters are identified, they cannot be estimated uniquely. We now discuss identification of the structural parameters in their order of estimation, first, the non-mean-value parameters and, then, the mean-value parameters.

2.3.1. Identification of Non-mean-value Parameters.

Let θ and ϕ, respectively, denote vectors of structural and reduced-form non-mean-value parameters. To estimate θ by maximum likelihood, we must define θ so that it can vary in an open set. We can directly include the true ARMA parameters and the noise VARMA coefficients in θ, but cannot directly include the noise disturbance-covariance matrix, Ω, in θ, because, being symmetric, Ω's upper and lower triangular elements would duplicate each other. Similarly, because Ω must be positive definite, we cannot just include its upper- or lower-triangular elements in θ. Thus, we reparameterize Ω to its lower-triangular Cholesky factor, $\Omega^{1/2}$, which satisfies $\Omega^{1/2}\Omega^{T/2} = \Omega$, and define $\theta = (\alpha_1, \ldots, \alpha_{p_1}, \beta_1, \ldots, \beta_{q_1}, \sigma_\varepsilon^2, \mathrm{vec}(A_1)^T, \ldots, \mathrm{vec}(A_{p_2})^T, \mathrm{vec}(B_1)^T, \ldots, \mathrm{vec}(B_{q_2})^T, \mathrm{vech}(\Omega^{1/2})^T)^T$, where $\mathrm{vec}(\cdot)$ denotes the columnwise vectorization of a matrix and $\mathrm{vech}(\cdot)$

denotes the columnwise vectorization of the lower-triangular part of a matrix, including its principal diagonal.

State-space form (2.10)-(2.11) always has the VARMA form

$$(2.12) \qquad y_t = \Phi_1 y_{t-1} + \ldots + \Phi_{r_2} y_{t-r_2} + \xi_{1,t} + \Theta_1 \xi_{1,t-1} + \ldots + \Theta_{r_2} \xi_{1,t-r_2},$$

where $\xi_{1,t}$, the first $m \times 1$ subvector of ξ_t, is the innovation vector of z_t in VARMA form (2.7). The covariance matrix of $\xi_{1,t}$ is the (1,1) block of Σ_ξ in equation (2.8), namely, $\Sigma_{\xi_1} = \delta I_m + M B_0^*(\sigma_\varepsilon^2 e_{1,s_1} e_{1,s_1}^T + \delta I_{s_1}) B_0^{*T} M^T + \Omega$. Let $R_1 R_1^T = \Sigma_{\xi_1}$, where R_1 is lower triangular. Then, we define $\phi = (\text{vec}(\Phi_1)^T, \ldots, \text{vec}(\Phi_{r_2})^T, \text{vec}(\Theta_1)^T, \ldots, \text{vec}(\Theta_{r_2})^T, \text{vech}(R_1)^T)^T$.

A model maps each admissible structural parameter values to one or more reduced-form parameter values. We denote this mapping by $\phi = f(\theta) \in X_\phi \subset \boldsymbol{R}^{\dim(\phi)}$, for $\theta \in X_\theta \subset \boldsymbol{R}^{\dim(\theta)}$. If $f(\cdot)$ maps each admissible value of $\theta \in X_\theta$ to a unique value of $\phi \in X_\phi \subset \boldsymbol{R}^{\dim(\phi)}$, then, $f(\cdot)$ can be inverted uniquely as $\theta = g(\phi) \equiv f^{-1}(\phi) \in X_\theta$, for $\phi \in X_\phi$, and $\theta \in X_\theta$ is identified in terms of $\phi \in X_\phi$. If $\theta \in X_\theta$ is identified in terms of $\phi \in X_\phi$ and $\dim(\theta) < \dim(\phi)$, then, θ is over identified in terms of ϕ; otherwise, if θ is identified in terms of ϕ and $\dim(\theta) = \dim(\phi)$, then, θ is just identified in terms of ϕ. Priestley (1981, pp. 801-804) states Hannan's (1976) sufficient conditions for identifying VARMA parameters -- the reduced-form parameters here -- in terms of theoretical covariances of observed variables. Strictly, identification concerns theoretical quantities, but, in practice, we estimate parameters using sample covariances. Presumably, sufficient conditions, such as stationarity, also hold so that sample covariances converge in probability to their theoretical counterparts as the number of sampling periods goes to infinity. It remains for us to state and verify sufficient conditions for identifying the S+N structural parameters in terms of the reduced-form VARMA parameters.

For simplicity, we discuss identification in the special case of $m = 3$, $p_1 = 2$, $p_2 = 1$, and $q_1 = q_2 = 0$. The general case of any m, p_1, p_1,

q_1, and q_2 follows similarly. In this special case, structural AR coefficients are mapped to reduced-form AR coefficients according to

(2.13) $\Phi_1 = \alpha_1 I_3 + A_1$, $\Phi_2 = \alpha_2 I_3 - \alpha_1 A_1$, $\Phi_3 = -\alpha_2 A_1$.

2.3.2. Identification of Mean-value Parameters.

To be completed.

3. Application.

We apply the signal plus noise model to payroll employment data in a real-time data exercise. The initial application exploits the institutional features of the BLS employment release, namely that the employment data estimates each measured data period three times in sequence, an initial release and two subsequent revisions. In each month, the new information consists of the initial release and the revisions of the two prior months of payroll data. For concreteness, suppose we have the first release estimate of March 2001 and then have the first revision of February 2001, and the second revision of January 2001. These three measures comprise an observation of the data in our analysis. We employ observations of this form – three observations per observation – for the sample period November 1964 to October 2004. The interpretation of the Data matrix table is analogous to the description of Table 1 above.

Table 2: Data Matrix For Application

$y_{v,t}$, v = vintage, t = sample period		
$Y_{1,1964:11}$	$y_{2,1964:10}$	$y_{3,1964:09}$
$Y_{1,1964:12}$	$y_{2,1964:11}$	$y_{3,1964:10}$
$y_{1,1965:1}$	$y_{2,1964:12}$	$y_{3,1964:11}$
$Y_{1, 1965:2}$	$y_{2, 1965:1}$	$y_{3,1964:12}$
$y_{1,1965:3}$	$y_{2,1965:2}$	$y_{3,1965:1}$
…	…	…
$y_{1,2004:10}$	$y_{2,2004:09}$	$y_{3,2004:08}$

With three releases per period as "data," we estimate a signal plus noise model assuming that the signal is an AR(3) and that the noise process is a VAR(1). We allow the covariance matrix of the disturbance terms in the measurement error equation to be unrestricted; that is, we need to estimate only 6 parameters for the variance-covariance matrix [(3 * 4)/2].[6] In this case, the number of structural parameters that we estimate is 19 (3 for the AR, 1 for the standard error of the signal process, 9 for the VAR parameters, and 6 for the covariance matrix of the disturbance terms in the measurement error process).

We display the parameter estimates for a specification using differences from the previous period (consistent within vintages – not across vintages). We are investigating alternative normalizing transformations of the data. Differencing across vintages, though in conflict with most real-time data analysis intuitions, may capture more effectively the correlations among the measurement errors for subsequent formal analysis for the time-series processes. We have also estimated specifications that use differenced logs and they offer similar insights. The sample period is from November 1964 to October 2004, and the forecast comparison periods are from November 1989 to October 2004. Table 3 (below) lists the complete set of parameter estimates for the full sample period model. We emphasize the analysis of the coefficients in the measurement error process below.

TABLE 2: VAR Coefficients For The Measurement Error Process (A Matrix)

$u_t = A_1 u_{t-1} + \eta_t$ in the VAR(1) case, $\mathbf{u}_t = [\ u_{1,t}\ u_{2,t}\ u_{3,t}\ ]`$

$$\begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} u_{1,t-1} \\ u_{2,t-1} \\ u_{3,t-1} \end{bmatrix} + \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \end{bmatrix}$$

[6] Recall that the disturbance vector, η_t, assumed to be distributed NIID(0,Ω) and independently of ε_t^*.

A =	**0.1050457E-01**	**0.7531286**	**0.3105490**
	0.4954400	**-0.6363300**	**-0.1728135**
	0.9205456E-01	**0.7701526**	**-0.1726643**

The VAR coefficients from the measurement error process provide interpretable relationships between the revision errors. For example, the first coefficient (a_{11}) is small (.01) suggesting that there is little correlation between subsequent "first release" measurement errors. This finding is reassuring in that such measurement errors should be unsystematic. In contrast, the coefficient (a_{21}) measuring the relationship between the second release measurement error and the first release error from the prior period (e.g., both measuring the same data period) is relatively large (.495). Note that each measure (today's second release and last month's first release) calculates employment for the same measurement period, so that the sizable coefficient estimate is plausible. Unfortunately, the other large coefficients in the estimation are less clearly interpretable. Specifically, the entire second column of the A matrix relates the measurement error for the second release from the prior month to the measurement errors for all releases in the current period. The coefficient estimates for this column are large for all releases. The interpretation of each is ambiguous. One could forego interpretation and suggest that these coefficients may reflect multicollinearity among the measurement error series. For example, the measurement error for the second release of the prior month has a sizable positive coefficient for the measurement error for the first release of the current month. However, the second release measurement error coefficient relating to this month's second release measurement error is sizably negative. Finally, the second release measurement error last month has a large positive coefficient with respect to the measurement error for the third release for the current month. This coefficient (a_{32}) can be explained in the same way as a_{21} noted above, that is, the measurement of employment in the same month. The other two coefficients in the second column, however, are discomforting. The coefficients of the third column offer a less intuitive interpretation as well. Here again, we may infer the multicollinearity of the measurement errors hinders precise coefficient estimates, and that there strength of correlation is minimal.

Estimates of the model using 5 releases display the same degree of fluctuation in coefficient estimates. We need to look further into this issue.

Initial results suggest that we are capturing some of the correlations in the model

Still thinking about how best to introduce the benchmark process

(clearly the most substantial source of data revision)

Business cycle phase – recession, recovery have largest revisions

[to be completed]

4. Conclusion.

[to be completed]

REFERENCES

Anderson, B.D.O. and J.B. Moore (1979), Optimal Filtering, Englewood Cliffs: Prentice-Hall.

Ansley, C.F. and R. Kohn (1983), "Exact Likelihood of Vector Autoregressive Moving-Average Models, Biometrika 60: 255-265.

Chen, B. and P.A. Zadrozny (2002), "Real-Time Quarterly Signal-Plus-Noise Model for Estimating 'True' GDP," ASA Proceedings of Business and Economic Statistics Section, 2001, Alexandria, VA: American Statistical Association.

Hannan, E.J. (1976), "The Identification and Parameterization of ARMAX and State-Space Forms," Econometrica 4: 713-723.

Harvey, Andrew and Chia-Hua Chung (2000), “Estimating the underlying change in Unemployment in the UK,” Journal of the Royal Statistical Association 163, Part 3: 303-339.

Priestley, M.B. (1981), Spectral Analysis and Time Series, London, UK: Academic Press.

Zadrozny, P.A. (1988), "Gaussian Likelihood of Continuous-Time ARMAX Models when Data Are Stocks and Flows at Different Frequencies," Econometric Theory 4: 109-124.

Zadrozny, P.A. (1990), "Estimating a Multivariate ARMA Model with Mixed-Frequency Data: An Application to Forecasting U.S. GNP at Monthly Intervals," Working Paper No. 90-6, Research Department, Federal Reserve Bank of Atlanta.

Zadrozny, P.A. (1998), "An Eigenvalue Method of Undetermined Coefficients for Solving Linear Rational Expectations Models," Journal of Economic Dynamics and Control 22: 1353-1373.

TABLE 1: DATA DESCRIPTION

Sample consists of first, second and third releases of payroll employment aggregate figures.

Data Period – 1964:November ending in 2004 October.

TABLE 2: VAR Coefficients For The Measurement Error Process (A Matrix)

$u_t = A_1 u_{t-1} + \eta_t$ in the VAR(1) case

$\mathbf{u}_t = [\ u_{1,t}\ u_{2,t}\ u_{3,t}\ ]`$

$$\begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} u_{1,t-1} \\ u_{2,t-1} \\ u_{3,t-1} \end{bmatrix} + \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \end{bmatrix}$$

A =

0.1050457E-01	0.7531286	0.3105490
0.4954400	-0.6363300	-0.1728135
0.9205456E-01	0.7701526	-0.1726643

TABLE 3: Estimates For The Signal Process

Autoregressive parameter estimates for the Signal process

Order	Estimate	Asymptotic SE	T Ratio	Marg. Sign. Level
1	0.4532663	1.238658	0.3659333	0.7144149
2	0.2913920E-01	1.510261	0.1929414E-01	0.9846065
3	0.2308405	0.8777201	0.2630001	0.7925505

Estimated standard error for the signal process

Standard Error	Asymptotic SE	T Ratio	Marg. Sign. Level
0.7567297	0.5468428	1.383816	0.1664149

TABLE 3: FORECAST ERRORS FROM THE ESTIMATED MODEL

Summary Statistics

Series	Obs	Mean	Std Error	Minimum	Maximum
FIRSTREL	177	100.83	175.75	-415.00	705.00
JULY04RELEASE	177	129.01	178.23	-361.00	506.00
UNDERLYING	177	67.83	96.51	-205.29	266.30
UNDERLYERR	176	60.32	132.12	-300.36	460.49
MEASERR	177	28.18	104.26	-321.00	291.00
RELEASERR	176	32.36	147.44	-342.84	730.49

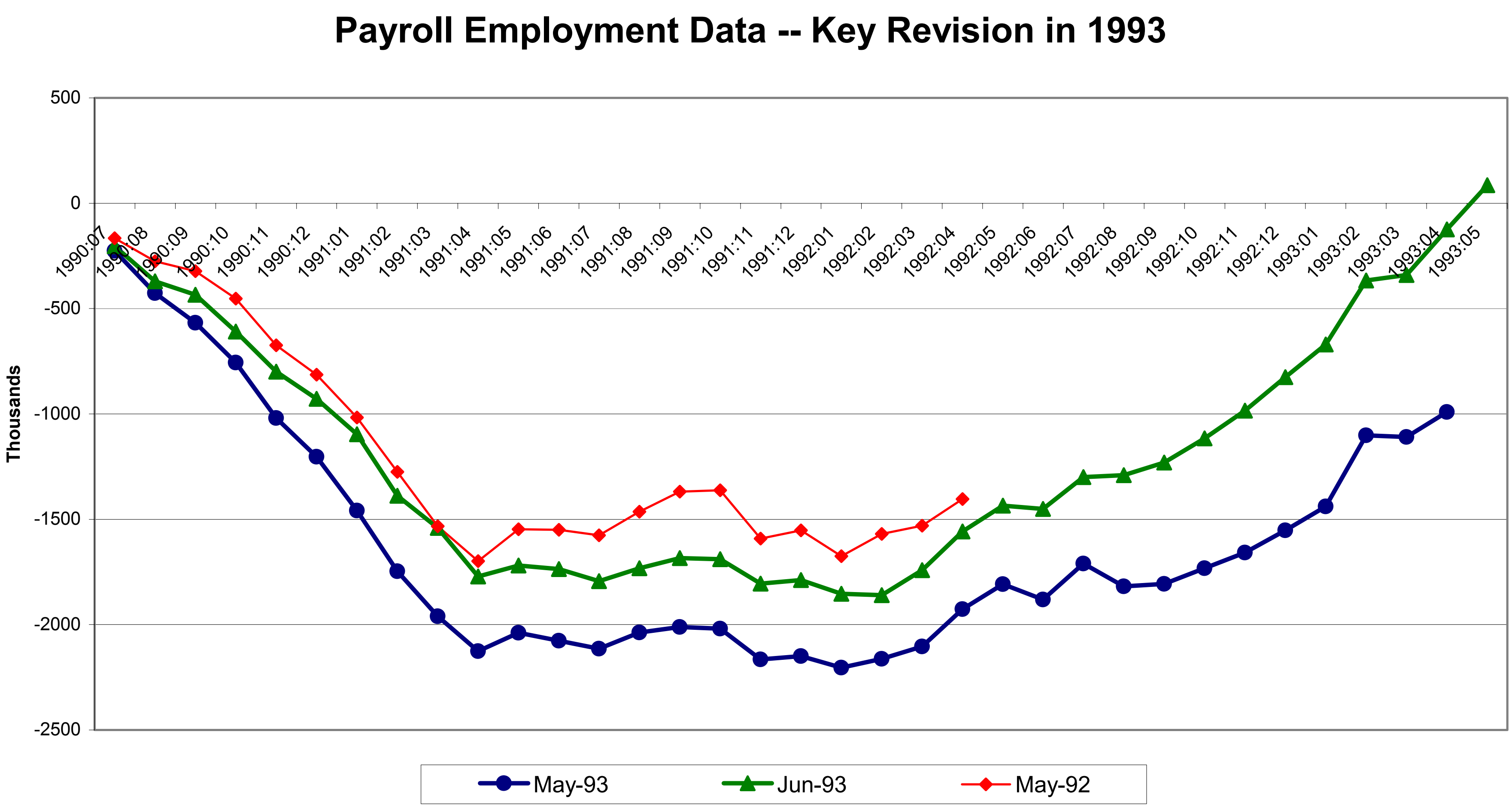
	Mean error	RMSE	rho(1)	rho(2)
Final less first release	32.36	107.0	-.08	0.09
Final-underlying	60.32	124.3	.22	.34

APPENDIX 1: ESTIMATED PARAMETERS FOR FULL SAMPLE (1964:11-2004:10)

MAXIMUM LIKELIHOOD ESTIMATES, ASYMPTOTIC STANDARD ERRORS,T-RATIOS,
AND MARGINAL SIGNIFICANCE LEVELS FOR THE 19 STRUCTURAL PARAMETERS

NO.	PARAM. EST.	ASY. STD. ERR.	T RATIO	MARG. SIGN. LEVEL
1	0.4532663	1.238658	0.3659333	0.7144149
2	0.2913920E-01	1.510261	0.1929414E-01	0.9846065
3	0.2308405	0.8777201	0.2630001	0.7925505
4	0.1050457E-01	1.390194	0.7556190E-02	0.9939711
5	0.4954400	1.290463	0.3839243	0.7010346
6	0.9205456E-01	0.7594781	0.1212077	0.9035266
7	0.7531286	2.737687	0.2750967	0.7832420
8	-0.6363300	1.924749	-0.3306041	0.7409435
9	0.7701526	1.282707	0.6004119	0.5482318
10	0.3105490	2.291353	0.1355309	0.89219
11	-0.1728135	1.101437	-0.1568982	0.8753251
12	-0.1726643	1.051789	-0.1641625	0.8696032
13	0.7567297	0.5468428	1.383816	0.1664149
14	0.3485250	0.6253198	0.5573548	0.5772851
15	0.6274749E-01	0.6276029	0.9997960E-01	0.9203605
16	0.3583206E-01	0.2635071	0.1359814	0.8918360
17	0.1728637	0.4403241	0.3925829	0.6946276
18	-0.7561716E-01	0.4597150	-0.1644870	0.8693478
19	0.1552349	0.3130848	0.4958238	0.6200187

Figure 1



Payroll Employment Data -- Key Revision in 1993
500
0
-500
-1000
-1500
-2000
-2500
Thousands
1990:07
1990:08
1990:09
1990:10
1990:11
1990:12
1991:01
1991:02
1991:03
1991:04
1991:05
1991:06
1991:07
1991:08
1991:09
1991:10
1991:11
1991:12
1992:01
1992:02
1992:03
1992:04
1992:05
1992:06
1992:07
1992:08
1992:09
1992:10
1992:11
1992:12
1993:01
1993:02
1993:03
1993:04
1993:05
May-93
Jun-93
May-92

Figure 2: Preliminary Forecast Comparison