Where Are We Now?

Real-Time Estimates of the Macro Economy

Martin D. D. Evans

Georgetown University

and the NBER

March 2005

Abstract

This paper describes a method for calculating daily real-time estimates of the current state of the U.S. economy. The estimates are computed from data on scheduled U.S. macroeconomic announcements using an econometric model that allows for variable reporting lags, temporal aggregation, and other complications in the data. The model can be applied to find real-time estimates of GDP, inflation, unemployment or any other macroeconomic variable of interest. In this paper I focus on the problem of estimating the current level of and growth rate in GDP. I construct daily real-time estimates of GDP that incorporate public information known on the day in question. The real-time estimates produced by the model are uniquely-suited to studying how perceived developments the macro economy are linked to asset prices over a wide range of frequencies. The estimates also provide, for the first time, daily time series that can be used in practical policy decisions.

Introduction

Information about the current state of real economic activity is widely dispersed across consumers, firms and policymakers. While individual consumers and firms know the recent history of their own decisions, they are unaware of the contemporaneous consumption, saving, investment and employment decisions made by other private sector agents. Similarly, policymakers do not have access to accurate contemporaneous information concerning private sector activity. Although information on real economic activity is collected by a number of government agencies, the collection, aggregation and dissemination process takes time. Thus, while U.S. macroeconomic data are released on an almost daily basis, the data represent official aggregations of past rather than current economic activity.

The lack of timely information concerning the current state of the economy is well-recognized among policymakers. This is especially true in the case of GDP, the broadest measure of real activity. The Federal Reserve's ability to make timely changes in monetary policy is made much more complicated by the lack of contemporaneous and accurate information on GDP. The lack of timely information concerning macroeconomic aggregates is also important for understanding private sector behavior, and in particular the behavior of asset prices. When agents make trading decisions based on their own estimate of current macroeconomic conditions, they transmit information to their trading partners. This trading activity leads to the aggregation of dispersed information and in the process affects the behavior of asset prices. Evans and Lyons (2004a) show that the lack of timely information concerning the state of the macroeconomy can significantly alter the dynamics of exchange and interest rates by changing the trading-based process of information aggregation.

This paper describes a method for estimating the current state of the economy on a continual basis using the flow of information from a wide range of macroeconomic data releases. These real-time estimates are computed from an econometric model that allows for variable reporting lags, temporal aggregation, and other complications that characterize the daily flow of macroeconomic information. The model can be applied to find real-time estimates of GDP, inflation, unemployment or any other macroeconomic variable of interest. In this paper, I focus on the problem of estimating GDP in real time.

The real-time estimates derived here are conceptually distinct from the real-time *data* series studied by Croushore and Stark (1999, 2001), Orphanides (2001) and others. A real-time data series comprises a set of *historical* values for a variable that are known on a particular date. This date identifies the vintage of the real-time data. For example, the March 31st. vintage of real-time GDP data would include data releases on GDP growth up to the forth quarter of the previous year. This vintage incorporates current revisions to earlier GDP releases but does not include a contemporaneous estimate of GDP growth in the first quarter. As such, it represents a *subset* of public information available on March 31st. By contrast, the March 31st. real-time estimate of GDP growth comprises an estimate of GDP growth in the first quarter based on information available on March 31st. The real-time estimates derived in this paper use an information set that spans the history of data releases on GDP and 18 other macroeconomic variables.

A number of papers have studied the problem of estimating GDP at a monthly frequency. Chow and Lin (1971) first showed how a monthly series could be constructed from regression estimates using monthly data related to GDP and quarterly GDP data. This technique has been subsequently integrated into VAR forecasting procedures (see, for example, Robertson and Tallman 1999). More recently, papers by Lui and Hall (2000) and Mariano and Murasawa (2003) have used state space models to combine quarterly GDP

1

data with other monthly series. The task of calculating real-time estimates of GDP growth has also been addressed by Clarida, Kitchen and Monaco (2002). They develop a regression-based method that uses a variety of monthly indicators to forecast GDP growth in the current quarter. The real-time estimates are calculated by combining the different forecasts with a weighting scheme based on the relative explanatory power of each forecasting equation.

I differ from this literature by modelling the growth in GDP as the quarterly aggregate of an unobserved *daily* process for real economy-wide activity. The model also specifies the relationship between GDP, data releases on GDP growth, and data releases on a set of other macroeconomic variables in a manner that accommodates the complex timing of releases. In particular, I incorporate the variable reporting lags that exist between the end of each data collection period (i.e., the end of a month or quarter) and the release day for each variable. This is only possible because the model tracks the evolution of the economy on a daily basis. An alternative approach of assuming that GDP aggregates an unobserved monthly process for economy-wide activity would result in a simpler model structure (see, Lui and Hall 2000 and Mariano and Murasawa 2003), but it could not accommodate the complex timing of data releases. The structure of the model also enables me to compute real-time estimates of GDP as the solution to an inference problem. In practice, I obtain the real-time estimates as a by-product of estimating the model. First, the model parameters are estimated by (qausi) maximum likelihood using the Kalman Filter algorithm. The real-time estimates are then obtained by applying the algorithm to the model evaluated at the maximum likelihood estimates.

My method for computing real-time estimates has several noteworthy features. First, the estimates are derived from a single fully-specified econometric model. As such, we can judge the reliability of the real-time estimates by subjecting the model to a variety of diagnostic tests. Second, a wide variety of variables can be computed from the estimated model. For example, the model can provide real-time forecasts for GDP growth for *any* future quarter. It can also be used to compute the precision of the real-time estimates as measured by the relevant conditional variance. Third, the estimated model can be used to construct high frequency estimates of real-economic activity. We can construct a daily series of real-time estimates for GDP growth in the current quarter, or real-time estimates of GDP produced in the current month, week, or even day. Fourth, the method can incorporate information from a wide range of economic indicators. In this paper I use the data releases for GDP and 18 other macroeconomic variables, but the set of indicators could be easily expanded to include many other macroeconomic series and financial data. Extending the model in this direction may be particular useful from a forecasting perspective. Stock and Watson (2002) show that harnessing the information in a large number indicators can have significant forecasting benefits.

The remainder of the paper is organized as follows. Section 1 describes the inference problem that must be solved in order to compute the real-time estimates. Here I detail the complex timing of data collection and macroeconomic data releases that needs to be accounted for in the model. The structure of the econometric model is presented in Section 2. Section 3 covers estimation and the calculation of the real-time estimates. I first show how the model can be written in state space form. Then I describe how the sample likelihood is constructed with the use of the Kalman Filter. Finally, I describe how various real-time estimates are calculated from the maximum likelihood estimates of the model. Section 4 presents the model estimates and specification tests. Here I compare the forecasting performance of the model against a survey of GDP estimates by professional money managers. These private estimates appear comparable to the model-based

estimates even though the managers have access to much more information than the model incorporates. Section 5 examines the model-based real-time estimates. First, I consider the relation between the real-time estimates and the final GDP releases. Next, I compare alternative real-time estimates for the level of GDP and examine the forecasting power of the model. Finally, I study how the data releases on other macro variables are related to changes in GDP at a monthly frequency. Section 6 concludes.

1 Real-Time Inference

My aim is to obtain high frequency real-time estimates on how the macro economy is evolving. For this purpose, it is important to distinguish between the arrival of information and data collection periods. Information about GDP can arrive via data releases on any day t. GDP data is collected on a quarterly basis. I index quarters by τ and denote the last day of quarter τ by $\text{Q}(\tau)$, with the first, second and third months ending on days $\text{M}(\tau,1)$, $\text{M}(\tau,2)$ and $\text{M}(\tau,3)$ respectively. I identify the days on which data is released in two ways. The release day for variable $\varkappa$ collected over quarter τ is $\text{R}_\varkappa(\tau)$. Thus, $\varkappa_{\text{R}(\tau)}$ denotes the value of variable $\varkappa$, over quarter τ, released on day $\text{R}_\varkappa(\tau)$. The release day for monthly variables is identified by $\text{R}_\varkappa(\tau,i)$ for $i=1,2,3$. In this case, $\varkappa_{\text{R}(\tau,i)}$ is the value of $\varkappa$, for month i in quarter τ, announced on day $\text{R}_\varkappa(\tau,i)$. The relation between data release dates and data collection periods is illustrated in Figure 1.

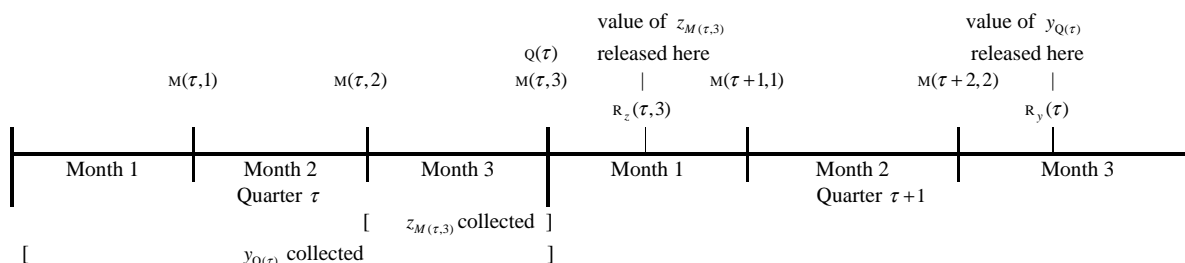**Figure 1**: Data collection periods and release times for quarterly and monthly variables. The reporting lag for "final" GDP growth in quarter τ, $y_{\text{Q}(\tau)}$, is $\text{R}_y(\tau)-\text{Q}(\tau)$. The reporting lag for the monthly series $z_{\text{M}(\tau,j)}$ is $\text{R}_z(\tau,j)-\text{M}(\tau,j)$. for $j=1,2,3$.

The Bureau of Economic Analysis (BEA) at the U.S. Commerce Department releases data on GDP growth in quarter τ in a sequence of three announcements: The "advanced" growth data are released during the first month of quarter $\tau+1$; the "preliminary" data are released in the second month; and the "final" data are released at the end of quarter $\tau+1$. The "final" data release does not represent the last official word on GDP growth in the quarter. Each summer, the BEA conducts an "annual" or comprehensive revision that generally lead to revisions in the "final" data values released over the previous three years. These revisions incorporate more complete and detailed micro data than was available before the "final" data release date.[1]

Let $x_{\text{Q}(\tau)}$ denote the log of real GDP for quarter τ ending on day $\text{Q}(\tau)$, and $y_{\text{R}(\tau)}$ be the "final" data

---

[1] For a complete description of BEA procedures, see Carson (1987), and Seskin and Parker (1998).

released on day $\mathrm{R}_y(\tau)$. The relation between the "final" data and actual GDP growth is given by

$$y_{\mathrm{R}(\tau)} = \Delta^{\mathrm{Q}} x_{\mathrm{Q}(\tau)} + v_{\mathrm{R}(\tau)}, \tag{1}$$

where $\Delta^{\mathrm{Q}} x_{\mathrm{Q}(\tau)} \equiv x_{\mathrm{Q}(\tau)} - x_{\mathrm{Q}(\tau-1)}$ and $v_{\mathrm{A}(\tau)}$ represents the effect of the future revisions (i.e., the revisions to GDP growth made after $\mathrm{R}_y(\tau)$). Notice that equation (1) distinguishes between the end of the reporting period $\mathrm{Q}(\tau)$, and the release date $\mathrm{R}_y(\tau)$. I shall refer to the difference $\mathrm{R}_y(\tau) - \mathrm{Q}(\tau)$ as the reporting lag for quarterly data. (For data series $\varkappa$ collected during month i of quarter τ, the reporting lag is $\mathrm{R}_\varkappa(\tau, i) - \mathrm{M}(\tau, i)$.) Reporting lags vary from quarter to quarter because data is collected on a calendar basis but announcements are not made on holidays and weekends. For example, "final" GDP data for the quarter ending in March has been released between June 27 th. and July 3 rd.

Real-time estimates of GDP growth are constructed using the information in a specific information set. Let Ω_t denote an information set that only contains data that is publicly known at the end of day t. The real-time *estimate* of GDP growth in quarter τ is defined as $E[\Delta^{\mathrm{Q}} x_{\mathrm{Q}(\tau)}|\Omega_{\mathrm{Q}(\tau)}]$, the expectation of $\Delta^{\mathrm{Q}} x_{\mathrm{Q}(\tau)}$ conditional on public information available at the end of the quarter, $\Omega_{\mathrm{Q}(\tau)}$. To see how this estimate relates to the "final" data release, y, I combine the definition with (1) to obtain

$$y_{\mathrm{R}(\tau)} = E\left[\Delta^{\mathrm{Q}} x_{\mathrm{Q}(\tau)}|\Omega_{\mathrm{Q}(\tau)}\right] + E\left[v_{\mathrm{R}(\tau)}|\Omega_{\mathrm{Q}(\tau)}\right] + \left(y_{\mathrm{R}(\tau)} - E\left[y_{\mathrm{R}(\tau)}|\Omega_{\mathrm{Q}(\tau)}\right]\right). \tag{2}$$

The "final" data released on day $\mathrm{R}_y(\tau)$ comprises three components; the real-time GDP growth estimate, an estimate of future data revisions, $E\left[v_{\mathrm{R}(\tau)}|\Omega_{\mathrm{Q}(\tau)}\right]$, and the real-time forecast error for the data release, $y_{\mathrm{R}(\tau)} - E\left[y_{\mathrm{R}(\tau)}|\Omega_{\mathrm{Q}(\tau)}\right]$. Under the reasonable assumption that $y_{\mathrm{R}(\tau)}$ represents the BEA's unbiased estimate of GDP growth, and that $\Omega_{\mathrm{Q}(\tau)}$ represents a subset of the information available to the BEA before the release day, $E\left[v_{\mathrm{R}(\tau)}|\Omega_{\mathrm{Q}(\tau)}\right]$ should equal zero. In this case, (2) becomes

$$y_{\mathrm{R}(\tau)} = E\left[\Delta^{\mathrm{Q}} x_{\mathrm{Q}(\tau)}|\Omega_{\mathrm{Q}(\tau)}\right] + \left(y_{\mathrm{R}(\tau)} - E\left[y_{\mathrm{R}(\tau)}|\Omega_{\mathrm{Q}(\tau)}\right]\right). \tag{3}$$

Thus, the data release $y_{\mathrm{R}(\tau)}$ can be viewed as a noisy signal of the real-time estimate of GDP growth, where the noise arises from the error in forecasting $y_{\mathrm{R}(\tau)}$ over the reporting lag. By construction, the noise term is orthogonal to the real-time estimate because both terms are defined relative to the same information set, $\Omega_{\mathrm{Q}(\tau)}$. The noise term can be further decomposed as

$$y_{\mathrm{R}(\tau)} - E\left[y_{\mathrm{R}(\tau)}|\Omega_{\mathrm{Q}(\tau)}\right] = \left(E\left[y_{\mathrm{R}(\tau)}|\Omega_{\mathrm{Q}(\tau)}^{\mathrm{BEA}}\right] - E\left[y_{\mathrm{R}(\tau)}|\Omega_{\mathrm{Q}(\tau)}\right]\right) + \left(y_{\mathrm{R}(\tau)} - E\left[y_{\mathrm{R}(\tau)}|\Omega_{\mathrm{Q}(\tau)}^{\mathrm{BEA}}\right]\right), \tag{4}$$

where Ω_t^{BEA} denotes the BEA's information set. Since the BEA has access to both private and public information sources, the first term on the right identifies the informational advantage conferred on the BEA at the end of the quarter $\mathrm{Q}(\tau)$. The second term identifies the impact of new information the BEA collects about $x_{\mathrm{Q}(\tau)}$ during the reporting lag. Since both of these terms could be sizable, there is no a priori reason to believe that real-time forecast error is always small.

To compute real-time estimates of GDP, we need to characterize the evolution of Ω_t and describe how inferences about $\Delta^{\mathrm{Q}} x_{\mathrm{Q}(\tau)}$ can be calculated from $\Omega_{\mathrm{Q}(\tau)}$. For this purpose, I incorporate the information contained in the "advanced" and "preliminary" GDP data releases. Let $\hat{y}_{\mathrm{R}(\tau)}$ and $\tilde{y}_{\mathrm{R}(\tau)}$ respectively denote

the values for the "advanced" and "preliminary" data released on days $\text{R}_{\hat{y}}(\tau)$ and $\text{R}_{\tilde{y}}(\tau)$ where $\text{Q}(\tau) < \text{R}_{\hat{y}}(\tau) < \text{R}_{\tilde{y}}(\tau)$. I assume that $\hat{y}_{\text{R}(\tau)}$ and $\tilde{y}_{\text{R}(\tau)}$ represent noisy signals of the "final" data, $y_{\text{R}(\tau)}$:

$$\hat{y}_{\text{R}(\tau)} \;=\; y_{\text{R}(\tau)} + \tilde{e}_{\text{R}(\tau)} + \hat{e}_{\text{R}(\tau)}, \tag{5}$$

$$\tilde{y}_{\text{R}(\tau)} \;=\; y_{\text{R}(\tau)} + \tilde{e}_{\text{R}(\tau)}, \tag{6}$$

where $\tilde{e}_{\text{R}(\tau)}$ and $\hat{e}_{\text{R}(\tau)}$ are independent mean zero revision shocks. $\tilde{e}_{\text{R}(\tau)}$ represents the revision between days $\text{R}_{\hat{y}}(\tau)$ and $\text{R}_y(\tau)$ and $\hat{e}_{\text{R}(\tau)}$ represents the revision between days $\text{R}_{\hat{y}}(\tau)$ and $\text{R}_{\tilde{y}}(\tau)$. The idea that the provisional data releases represent noisy signals of the "final" data is originally due to Mankiw and Shapiro (1986). It implies that the revisions $\tilde{e}_{\text{R}(\tau)}$ and $\tilde{e}_{\text{R}(\tau)} + \hat{e}_{\text{R}(\tau)}$ are orthogonal to $y_{\text{R}(\tau)}$. I impose this orthogonality condition when estimating the model. The specification of (5) and (6) also implies that the "advanced" and "preliminary" data releases represent unbiased estimates of actual GDP growth. This assumption is consistent with the evidence reported in Faust et. al (2000) for US data releases between 1988 and 1997. (Adding non-zero means for $\tilde{e}_{\text{R}(\tau)}$ and $\hat{e}_{\text{R}(\tau)}$ is a straightforward extension to accommodate bias that may be present in different sample periods.)[2]

The three GDP releases $\{\hat{y}_{\text{R}(\tau)}, \tilde{y}_{\text{R}(\tau)}, y_{\text{R}(\tau)}\}$ represent a sequence of signals on actual GDP growth that augment the public information set on days $\text{R}_{\hat{y}}(\tau)$, $\text{R}_{\tilde{y}}(\tau)$ and $\text{R}_y(\tau)$. In principle we could construct real-time estimates based only on these data releases as $E[\Delta^{\text{Q}} x_{\text{Q}(\tau)} | \Omega^y_{\text{Q}(\tau)}]$, where Ω^y_t is the information set comprising data on the three GDP series released on or before day t :

$$\Omega^y_t \equiv \left\{ \hat{y}_{\text{R}(\tau)}, \tilde{y}_{\text{R}(\tau)}, y_{\text{R}(\tau)} : \text{R}(\tau) < t \right\}.$$

Notice that these estimates are only based on data releases relating to GDP growth *before* the current quarter because the presence of the reporting lags exclude the values of $\hat{y}_{\text{R}(\tau)}$, $\tilde{y}_{\text{R}(\tau)}$, and $y_{\text{R}(\tau)}$ from $\Omega_{\text{Q}(\tau)}$. As such, these candidate real-time estimates exclude information on $\Delta^{\text{Q}} x_{\text{Q}(\tau)}$ that is available at the end of the quarter. Much of this information comes from the data releases on other macroeconomic variables, like employment, retail sales and industrial production. Data for most of these variables are collected on a monthly basis[3], and as such can provide timely information on GDP growth. To see why this is so, consider the data releases on Nonfarm Payroll Employment, z. Data on z for the month ending on day $\text{M}_z(\tau, j)$ is

---

[2] It is also possible to accommodate Mankiw and Shapiro's "news" view of data revisions within the model. According to this view, provisional data releases represent the BEA's best estimate of $y_{\text{R}(\tau)}$ at the time the provision data is released. Hence $\tilde{y}_{\text{R}(\tau)} = E[y_{\text{R}(\tau)} | \Omega^{\text{BEA}}_{\text{R}_{\tilde{y}}(\tau)}]$ and $\hat{y}_{\text{R}(\tau)} = E[y_{\text{R}(\tau)} | \Omega^{\text{BEA}}_{\text{R}_{\hat{y}}(\tau)}]$. If the BEA's forecasts are optimal, we can write $y_{\text{R}(\tau)} = \tilde{y}_{\text{R}(\tau)} + \tilde{w}_{\text{R}(\tau)}$, and $y_{\text{R}(\tau)} = \hat{y}_{\text{R}(\tau)} + \hat{w}_{\text{R}(\tau)}$ where $\tilde{w}_{\text{R}(\tau)}$ and $\hat{w}_{\text{R}(\tau)}$ are the forecast errors associated with $E[y_{\text{R}(\tau)} | \Omega^{\text{BEA}}_{\text{R}_{\tilde{y}}(\tau)}]$ and $E[y_{\text{R}(\tau)} | \Omega^{\text{BEA}}_{\text{R}_{\hat{y}}(\tau)}]$ respectively. We could use these equations to compute the projections of $\tilde{y}_{\text{R}(\tau)}$ and $\hat{y}_{\text{R}(\tau)}$ on $y_{\text{R}(\tau)}$ and a constant:

$$\tilde{y}_{\text{R}(\tau)} \;=\; \tilde{\beta}_0 + \tilde{\beta} y_{\text{R}(\tau)} + \tilde{\varepsilon}_{\text{R}(\tau)},$$

$$\hat{y}_{\text{R}(\tau)} \;=\; \hat{\beta}_0 + \hat{\beta} y_{\text{R}(\tau)} + \hat{\varepsilon}_{\text{R}(\tau)}.$$

The projection errors $\tilde{\varepsilon}_{\text{R}(\tau)}$ and $\hat{\varepsilon}_{\text{R}(\tau)}$ are orthogonal to $y_{\text{R}(\tau)}$ by construction so these equations could replace (5) and (6). The projection coefficients, $\tilde{\beta}_0, \tilde{\beta}, \hat{\beta}_0$ and $\hat{\beta}$, would add to the set of model parameters to be estimated. I chose not to follow this alternative formulation because there is evidence that data revisions are forecastable with contemporaneous information (Dynan and Elmendorf, 2001). This finding is inconsistent with the "news" view if the BEA makes rational forecasts. Furthermore, as I discuss below, a specification based on (5) and (6) allows the optimal (model-based) forecasts of "final" GDP to closely approximate the provisional data releases. The model estimates will therefore provide us with an empirical perspective on the "noise" and "news" characterizations of data revisions.

[3] Data on inital unemployment claims are collected week by week.

released on $R_z(\tau, j)$, a day that falls between the 3'rd and the 9'th of month $j+1$ (as illustrated in Figure 1). This reporting lag is much shorter than the lag for GDP releases but it does exclude the use of employment data from the 3r'd month in estimating real-time GDP. However, insofar as employment during the first two months is related to GDP growth over the quarter, the values of $z_{R(\tau,1)}$ and $z_{R(\tau,2)}$ will provide information relevant to estimating GDP growth at the end of the quarter.

The real-time estimates I construct below will be based on data from the three GDP releases and the monthly releases of other macroeconomic data. To incorporate the information from these other variables, I decompose quarterly GDP growth into a sequence of daily increments:

$$\Delta^{Q} x_{Q(\tau)} = \sum_{i=1}^{D(\tau)} \Delta x_{Q(\tau-1)+i}, \tag{7}$$

where $D(\tau) \equiv Q(\tau) - Q(\tau-1)$ is the duration of quarter τ. The daily increment Δx_t represents the contribution on day t to the growth of GDP in quarter τ. If x_t were a stock variable, like the log price level on day t, Δx_t would identify the daily growth in the stock (e.g. the daily rate of inflation). Here $x_{Q(\tau)}$ denotes the log of the flow of output over quarter τ so it is not appropriate to think of Δx_t as the daily growth in GDP. I will examine the link between Δx_t and daily GDP in Section 3.3 below.

To incorporate the information contained in the i'th. macro variable, z^i, I project $z^i_{R(\tau,j)}$ on a portion of GDP growth

$$z^i_{R(\tau,j)} = \beta_i \Delta^{M} x_{M(\tau,j)} + u^i_{M(\tau,j)}, \tag{8}$$

where $\Delta^{M} x_{M(\tau,j)}$ is the contribution to GDP growth in quarter τ during month j:

$$\Delta^{M} x_{M(\tau,j)} \equiv \sum_{i=M(\tau,j-1)+1}^{M(\tau,j)} \Delta x_i.$$

β_i is the projection coefficient and $u^i_{M(\tau,j)}$ is the projection error that is orthogonal to $\Delta^{M} x_{M(\tau,j)}$. Notice that equation (8) incorporates the reporting lag $R_z(\tau,j) - M_z(\tau,j)$ for variable z which can vary in length from month to month.

The real-time estimates derived in this paper are based on a information set specification that includes the 3 GDP releases and 18 monthly macro series; $z^i = 1, 2, ...18$. Formally, I compute the end-of-quarter real-time estimates as

$$E[\Delta^{Q} x_{Q(\tau)} | \Omega_{Q(\tau)}], \tag{9}$$

where $\Omega_t = \Omega^{\tilde{z}}_t \cup \Omega^y_t$ with $\Omega^{\tilde{z}}_t$ denoting the information set comprising of data on the 18 monthly macro variables that has been released on or before day t :

$$\Omega^{\tilde{z}}_t \equiv \bigcup_{i=1}^{21} \left\{ z^i_{R(\tau,j)} : R(\tau, j) < t \text{ for } j = 1, 2, 3 \right\}.$$

The model presented below enables us to compute the real-time estimates in (9) using equations (1), (5), (6), (7), and (8) together with a time-series process for the daily increments, Δx_t. The model will also

enable us to compute *daily* real-time estimates of quarterly GDP, and GDP growth:

$$x_{\mathrm{Q}(\tau)|i} \equiv E[x_{\mathrm{Q}(\tau)}|\Omega_i] \tag{10}$$

$$\Delta^{\mathrm{Q}}x_{\mathrm{Q}(\tau)|i} \equiv E[\Delta^{\mathrm{Q}}x_{\mathrm{Q}(\tau)}|\Omega_i]. \tag{11}$$

for $\mathrm{Q}(\tau-1) < i \leq \mathrm{Q}(\tau)$. Equations (10) and (11) respectively identify the real-time estimate of log GDP, and GDP growth in quarter τ, based on information available on day i during the quarter. $x_{\mathrm{Q}(\tau)|i}$ and $\Delta^{\mathrm{Q}}x_{\mathrm{Q}(\tau)|i}$ incorporate real-time forecasts of the daily contribution to GDP in quarter τ between day i and $\mathrm{Q}(\tau)$. These high frequency estimates are particularly useful in studying how data releases affect estimates of the current state of the economy, and forecasts of how it will evolve in the future. As such, they are uniquely suited to examining how data releases affect a whole array of asset prices.

2　The Model

The dynamics of the model center on the behavior of two partial sums:

$$s_t^{\mathrm{Q}} \equiv \sum_{i=\mathrm{Q}(\tau)+1}^{\min\{\mathrm{Q}(\tau),t\}} \Delta x_i, \tag{12}$$

$$s_t^{\mathrm{M}} \equiv \sum_{i=\mathrm{M}(\tau,j-1)+1}^{\min\{\mathrm{M}(\tau,j),t\}} \Delta x_i. \tag{13}$$

Equation (12) defines the cumulative daily contribution to GDP growth in quarter τ, ending on day $t \leq \mathrm{Q}(\tau)$. The cumulative daily contribution between the start of month j in quarter τ and day t is defined by s_t^{M}. Notice that when t is the last day of the quarter, $\Delta^{\mathrm{Q}}x_{\mathrm{Q}(\tau)} = s_{\mathrm{Q}(\tau)}^{\mathrm{Q}}$ and when t is the last day of month j, $\Delta^{\mathrm{M}}x_{\mathrm{M}(\tau,j)} = s_{\mathrm{M}(\tau,j)}^{\mathrm{M}}$. To describe the daily dynamics of s_t^{Q} and s_t^{M}, I introduce the following dummy variables:

$$\lambda_t^{\mathrm{M}} = \begin{cases} 1 & \text{if } t = \mathrm{M}(\tau,j)+1, \text{ for } j=1,2,3, \\ 0 & \text{otherwise}, \end{cases}$$

$$\lambda_t^{\mathrm{Q}} = \begin{cases} 1 & \text{if } t = \mathrm{Q}(\tau)+1, \\ 0 & \text{otherwise}. \end{cases}$$

Thus, λ_t^{M} and λ_t^{Q} take the value of one if day t is the first day of the month or quarter respectively. We may now describe the daily dynamics of s_t^{Q} and s_t^{M} with the following equations:

$$s_t^{\mathrm{Q}} = (1 - \lambda_t^{\mathrm{Q}})s_{t-1}^{\mathrm{Q}} + \Delta x_t, \tag{14}$$

$$s_t^{\mathrm{M}} = (1 - \lambda_t^{\mathrm{M}})s_{t-1}^{\mathrm{M}} + \Delta x_t. \tag{15}$$

The next portion of the model accommodates the reporting lags. Let $\Delta^{\mathrm{Q}(j)}x_t$ denote the quarterly growth in GDP ending on day $\mathrm{Q}(\tau-j)$ where $\mathrm{Q}(\tau)$ denotes the last day of the most recently completed quarter and

7

$t \geq \text{Q}(\tau)$. Quarterly GDP growth in the last (completed) quarter is given by

$$\Delta^{\text{Q}(1)} x_t = \left(1 - \lambda_t^{\text{Q}}\right) \Delta^{\text{Q}(1)} x_{t-1} + \lambda_t^{\text{Q}} s_{t-1}^{\text{Q}}. \tag{16}$$

When t is the first day of a new quarter, $\lambda_t^{\text{Q}} = 1$, so $\Delta^{\text{Q}(1)} x_{\text{Q}(\tau)+1} = s_{\text{Q}(t)}^{\text{Q}} = \Delta^{\text{Q}} x_{\text{Q}(\tau)}$. On all other days, $\Delta^{\text{Q}(1)} x_t = \Delta^{\text{Q}(1)} x_{t-1}$. On some dates the reporting lag associated with a "final" GDP data release is more than one quarter, so we will need to identify GDP growth from two quarters back, $\Delta^{\text{Q}(2)} x_t$. This is achieved with a similar recursion:

$$\Delta^{\text{Q}(2)} x_t = \left(1 - \lambda_t^{\text{Q}}\right) \Delta^{\text{Q}(2)} x_t + \lambda_t^{\text{Q}} \Delta^{\text{Q}(1)} x_{t-1}. \tag{17}$$

Equations (14), (16) and (17) enable us to define the link between the daily contributions to GDP growth Δx_t, and the three GDP data releases $\{\hat{y}_t, \tilde{y}_t, y_t\}$. Let us start with the "advanced" GDP data releases. The reporting lag associated with these data is always less than one quarter, so we can combine (1) and (5) with the definition of $\Delta^{\text{Q}(1)} x_t$ to write

$$\hat{y}_t = \Delta^{\text{Q}(1)} x_t + \upsilon_{\text{R}(\tau)} + \tilde{e}_{\text{R}(\tau)} + \hat{e}_{\text{R}(\tau)}. \tag{18}$$

It is important to recognize that (18) builds in the variable reporting lag between the release day, $\text{R}_{\hat{y}}(\tau)$, and the end of the last quarter $\text{Q}(\tau)$. The value of $\Delta^{\text{Q}(1)} x_t$ dose not change from day to day after quarter ends, so the relation between the data release and actual GDP growth is unaffected by within-quarter variations in the reporting lag. The reporting lag for the "preliminary" data are also always less than one quarter. Combining (1) and (6) with the definition of $\Delta^{\text{Q}(1)} x_t$ we obtain

$$\tilde{y}_t = \Delta^{\text{Q}(1)} x_t + \upsilon_{\text{R}(\tau)} + \tilde{e}_{\text{R}(\tau)}. \tag{19}$$

Data on "final" GDP growth is release around the end of the following quarter so the reporting lag can vary between one and two quarters. In cases where the reporting lag is one quarter,

$$y_t = \Delta^{\text{Q}(1)} x_t + \upsilon_{\text{R}(\tau)}, \tag{20}$$

and when the lag is two quarters,

$$y_t = \Delta^{\text{Q}(2)} x_t + \upsilon_{\text{R}(\tau)}. \tag{21}$$

I model the links between the daily contributions to GDP growth and the monthly macro variables in a similar manner. Let $\Delta^{\text{M}(i)} x_t$ denote the monthly contribution to quarterly GDP growth ending on day $\text{M}(\tau, j - i)$, where $\text{M}(\tau, j)$ denotes the last day of the most recently completed month and $t \geq \text{M}(\tau, j)$. The contribution GDP growth in the last (completed) month is given by

$$\Delta^{\text{M}(1)} x_t = \left(1 - \lambda_t^{\text{M}}\right) \Delta^{\text{M}(1)} x_{t-1} + \lambda_t^{\text{M}} s_{t-1}^{\text{M}}, \tag{22}$$

and the contribution from $i \ (> 1)$ months back is

$$\Delta^{\text{M}(i)} x_t = \left(1 - \lambda_t^{\text{M}}\right) \Delta^{\text{M}(i)} x_t + \lambda_t^{\text{M}} \Delta^{\text{M}(i-1)} x_{t-1}. \tag{23}$$

These equations are analogous to (16) and (17). If t is the first day of a new month, $\lambda_t^{\text{M}} = 1$, so $\Delta^{\text{M}(1)}x_{\text{M}(\tau,j)+1} = s_{\text{M}(\tau,j)}^{\text{Q}} = \Delta^{\text{M}}x_{\text{M}(\tau,j)}$ and $\Delta^{\text{M}(i)}x_{\text{M}(\tau,j)+1} = \Delta^{\text{M}(i-1)}x_{\text{M}(\tau,j)}$, for $j = 1, 2, 3$. On all other days, $\Delta^{\text{M}(i)}x_t = \Delta^{\text{M}(i)}x_{t-1}$. The $\Delta^{\text{M}(i)}x_t$ variables link the monthly data releases, z_t^i, to quarterly GDP growth. If the reporting lag for macro series i is less than one month, the value released on day t can be written as

$$z_t^i = \beta_i \Delta^{\text{M}(1)}x_t + u_t^i. \tag{24}$$

In cases where the reporting lag is two months,

$$z_t^i = \beta_i \Delta^{\text{M}(2)}x_t + u_t^i. \tag{25}$$

As above, both equations allow for a variable within-month reporting lag, $\text{R}_{z^i}(\tau,j) - \text{M}_{z^i}(\tau,j)$.

Equations (24) and (25) accommodate all the monthly data releases I use except for the index of consumer confidence, $i = 18$. This series is released *before* the end of the month in which the survey data are collected. These data are potentially valuable for drawing real-time inferences because they represent the only monthly release before $\text{Q}(\tau)$ that relates to activity during the last month of the quarter. I incorporate the information in the consumer confidence index ($i = 18$) by projecting z_t^{18} on the partial sum s_t^{M} :

$$z_t^{18} = \beta_{18} s_t^{\text{M}} + u_t^{18}. \tag{26}$$

To complete the model we need to specify the dynamics for the daily contributions, Δx_t. I assume that

$$\Delta x_t = \sum_{i=1}^{k} \phi_i \Delta^{\text{M}(i)}x_t + e_t, \tag{27}$$

where e_t is an i.i.d. $N(0, \sigma_e^2)$ shock. Equation (27) expresses the growth contribution on day t as a weighted average of the monthly contributions over the last k (completed) months, plus an error term. This specification has two noteworthy features. First, the daily contribution on day t only depends on the history of Δx_t insofar as it is summarized by the monthly contributions, $\Delta^{\text{M}(i)}x_t$. Thus, forecast for Δx_{t+h} conditional $\left\{\Delta^{\text{M}(i)}x_t\right\}_{i=1}^{k}$ are the same for horizons h within the current month. The second feature of (27) is that the process aggregates up to a AR(k) process for $\Delta^{\text{M}}x_{\text{M}(\tau,j)}$ at the monthly frequency. As I shall demonstrate, this feature enables use to compute real-time forecasts of future GDP growth over monthly horizons with comparative ease.

3   Estimation

Finding the real time estimates of GDP and GDP growth requires a solution to two related problems. First, there is a pure inference problem of how to compute $E[x_{\text{Q}(\tau)}|\Omega_i]$ and $E[\Delta^{\text{Q}}x_{\text{Q}(\tau)}|\Omega_i]$ using the quarterly signalling equations (18) - (21), the monthly signalling equations (24) - (26), and the Δx_t process in (27), given values for all the parameters in these equations. Second, we need to estimate these parameters from the three data releases on GDP and the 18 other macro series. This problem is complicated by the fact that individual data releases are irregularly spaced, and arrive in a non-syncronized manner: On some days there

is one release, on others there are several, and on some there are none at all. In short, the temporal pattern of data releases is quite unlike that found in standard time-series applications.

The Kalman Filtering algorithm provides a solution to both problems. In particular, given a set of parameter values, the algorithm provides the means to compute the real-time estimates $E[x_{Q(\tau)}|\Omega_i]$ and $E[\Delta^Q x_{Q(\tau)}|\Omega_i]$. The algorithm also allows us to construct a sample likelihood function from the data series, so that the model's parameters can be computed by maximum likelihood. Although the Kalman Filtering algorithm has been used extensively in the applied time-series literature, its application in the current context has several novel aspects. For this reason, the presentation below concentrates on these features.[4]

3.1 The State Space Form

To use the algorithm, we must first write the model in state space form comprising a state and observation equation. For the sake of clarity, I shall present the state space form for the model where Δx_t depend only on last month's contribution (i.e., $k = 1$ in equation (27)). Modifying the state space form for the case where $k > 1$ is straightforward.

The dynamics described by equations (14) - (17), (22), (23) and (27) with $k = 1$ can be represented by the matrix equation:

$$
\begin{bmatrix}
s_t^Q \\
\Delta^{Q(1)} x_t \\
\Delta^{Q(2)} x_t \\
s_t^M \\
\Delta^{M(1)} x_t \\
\Delta^{M(2)} x_t \\
\Delta x_t
\end{bmatrix}
=
\begin{bmatrix}
1 - \lambda_t^Q & 0 & 0 & 0 & 0 & 0 & 1 \\
\lambda_t^Q & 1 - \lambda_t^Q & 0 & 0 & 0 & 0 & 0 \\
0 & \lambda_t^Q & 1 - \lambda_t^Q & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 - \lambda_t^M & 0 & 0 & 1 \\
0 & 0 & 0 & \lambda_t^M & 1 - \lambda_t^M & 0 & 0 \\
0 & 0 & 0 & 0 & \lambda_t^M & 1 - \lambda_t^M & 0 \\
0 & 0 & 0 & 0 & \phi_1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
s_{t-1}^Q \\
\Delta^{Q(1)} x_{t-1} \\
\Delta^{Q(2)} x_{t-1} \\
s_{t-1}^M \\
\Delta^{M(1)} x_{t-1} \\
\Delta^{M(2)} x_{t-1} \\
\Delta x_{t-1}
\end{bmatrix}
+
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
e_t
\end{bmatrix},
$$

or, more compactly

$$
\mathbb{Z}_t = \mathbb{A}_t \mathbb{Z}_{t-1} + \mathbb{V}_t. \tag{28}
$$

Equation (28) is known as the state equation. In traditional time-series applications, the state transition matrix $\mathbb{A}$ is constant. Here elements of $\mathbb{A}_t$ depend on the quarterly and monthly dummies, λ_t^Q and λ_t^M and so is time-varying.

Next, we turn to the observation equation. The link between the data releases on GDP and elements of the state vector are described by (18), (19), (20) and (21). These equations can be rewritten as

$$
\begin{bmatrix}
\hat{y}_t \\
\tilde{y}_t \\
y_t
\end{bmatrix}
=
\begin{bmatrix}
0 & \mathrm{QL}_t^1(\hat{y}) & \mathrm{QL}_t^2(\hat{y}) & 0 & 0 & 0 & 0 \\
0 & \mathrm{QL}_t^1(\tilde{y}) & \mathrm{QL}_t^2(\tilde{y}) & 0 & 0 & 0 & 0 \\
0 & \mathrm{QL}_t^1(y) & \mathrm{QL}_t^2(y) & 0 & 0 & 0 & 0
\end{bmatrix}
\mathbb{Z}_t
+
\begin{bmatrix}
1 & 1 & 1 \\
0 & 1 & 1 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
\hat{e}_t \\
\tilde{e}_t \\
v_t
\end{bmatrix}, \tag{29}
$$

where $\mathrm{QL}_t^i(\varkappa)$ denotes a dummy variable that takes the value of one when the reporting lag for series $\varkappa$ lies between $i - 1$ and i quarters, and zero otherwise. Thus, $\mathrm{QL}_t^1(y) = 1$ and $\mathrm{QL}_t^2(y) = 0$ when "final" GDP

---

[4] For a textbook introduction to the Kalman Filter and its uses in standard time-series applications, see Harvey (1989) or Hamilton (1994).

for the first quarter is released before the start of the third quarter, while $\mathrm{QL}_t^1(y) = 0$ and $\mathrm{QL}_t^2(y) = 1$ in case where the release in delayed until the third quarter. Under normal circumstances, the "advance" and "preliminary" GDP data releases have reporting lags that are less than a month. However, there was one occasion in the sample period where all the GDP releases were delayed so that the $\mathrm{QL}_t^i(\varkappa)$ dummies are also needed for the $\hat{y}_t$ and $\tilde{y}_t$ equations.

The link between the data releases on the monthly series and elements of the state vector are described by (24) - (26). These equations can be written as

$$z_t^i = \left[\begin{array}{ccccccc} 0 & 0 & 0 & \beta_i\mathrm{ML}_t^0(z^i) & \beta_i\mathrm{ML}_t^1(z^i) & \beta_i\mathrm{ML}_t^2(z^i) & 0 \end{array} \right] \mathbb{Z}_t + u_t^i, \tag{30}$$

for $i = 1, 2, ..18$. $\mathrm{ML}_t^i(\varkappa)$ is the monthly version of $\mathrm{QL}_t^i(\varkappa)$. $\mathrm{ML}_t^i(\varkappa)$ is equal to one if the reporting lag for series $\varkappa$ lies between $i - 1$ and i months ($i = 1, 2$), and zero otherwise. $\mathrm{ML}_t^0(\varkappa)$ equals one when the release day is before the end of the collection month (as is the case with the Index of Consumer Confidence). Stacking (29) and (30) gives

$$\left[\begin{array}{c} \hat{y}_t \\ \tilde{y}_t \\ y_t \\ z_t^1 \\ \vdots \\ z_t^{18} \end{array} \right] = \left[\begin{array}{ccccccc} 0 & \mathrm{QL}_t^1(\hat{y}) & \mathrm{QL}_t^2(\hat{y}) & 0 & 0 & 0 & 0 \\ 0 & \mathrm{QL}_t^1(\tilde{y}) & \mathrm{QL}_t^2(\tilde{y}) & 0 & 0 & 0 & 0 \\ 0 & \mathrm{QL}_t^1(y) & \mathrm{QL}_t^2(y) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_1\mathrm{ML}_t^0(z^1) & \beta_i\mathrm{ML}_t^1(z^1) & \beta_1\mathrm{ML}_t^2(z^1) & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \beta_{18}\mathrm{ML}_t^0(z^{18}) & \beta_{18}\mathrm{ML}_t^1(z^{18}) & \beta_{18}\mathrm{ML}_t^2(z^{18}) & 0 \end{array} \right] \mathbb{Z}_t + \left[\begin{array}{c} \hat{e}_t + \tilde{e}_t + v_t \\ \tilde{e}_t + v_t \\ v_t \\ u_t^1 \\ \vdots \\ u_t^{18} \end{array} \right],$$

or

$$\mathbb{X}_t = \mathbb{C}_t \mathbb{Z}_t + \mathbb{U}_t. \tag{31}$$

This equation links the vector of *potential* data releases for day t, $\mathbb{X}_t$, to elements of the state vector. The elements of $\mathbb{X}_t$ identify the value that would have been released for each series given the current state, $\mathbb{Z}_t$, if day t was in fact the release day. Of course, on a typical day, we would only observe the elements in $\mathbb{X}_t$ that corresponding to the actual releases that day. For example, if data on "final" GDP and monthly series $i = 1$ are released on day t, we would observe the values in the 3'rd. and 4'th rows of $\mathbb{X}_t$. On days when there are no releases, none of the elements of $\mathbb{X}_t$ are observed.

The observation equation links the data releases for day t to the state vector. The vector of actual data releases for day t, $\mathbb{Y}_t$, is related to the vector of *potential* releases by

$$\mathbb{Y}_t = \mathbb{B}_t \mathbb{X}_t,$$

where $\mathbb{B}_t$ is a $n \times 7$ selection matrix that "picks out" the $n \geq 1$ data releases for day t. For example, if data on monthly series $i = 1$ is released on day t, $\mathbb{B}_t = [\begin{array}{ccccccc} 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array}]$. Combining this expression with (31) gives the observation equation:

$$\mathbb{Y}_t = \mathbb{B}_t \mathbb{C}_t \mathbb{Z}_t + \mathbb{B}_t \mathbb{U}_t. \tag{32}$$

Equation (32) differs in several respects from the observation equation specification found in standard time-series applications. First, the equation only applies on days for which at least one data release takes

11

place. Second, the link between the observed data releases and the state vector varies through time via $\mathbb{C}_t$ as $\mathrm{QL}_t^i(\varkappa)$ and $\mathrm{ML}_t^i(\varkappa)$ change. These variations arise because the reporting lag associated with a given data series change from release to release. Third, the number and nature of the data releases varies from day to day (i.e., the dimension of $\mathbb{Y}_t$ can vary across consecutive data-release days) via the $\mathbb{B}_t$ matrix. These changes may be a source of heteroskedasticity. If the $\mathbb{U}_t$ vector has a constant covariance matrix Ω_u, the vector of noise terms entering the observation equation will be heteroskedastic with covariance $\mathbb{B}_t \Omega_u \mathbb{B}_t'$.

3.2   The Kalman Filter and Sample Likelihood Function

Equations (28) and (32) describe a state space form which can be used to find real-time estimates of GDP in two steps. In the first, I obtain the maximum likelihood estimates of the model's parameters. The second step calculates the real-time estimates of GDP using the maximum likelihood parameter estimates. Below, I briefly describe these steps noting where the model gives rise to features that are not seen in standard time-series applications.

The parameters of the model to be estimated are $\theta = \{\beta_1, \ldots, \beta_{21}, \phi_1, \ldots, \phi_k, \sigma_e^2, \sigma_{\tilde{e}}^2, \sigma_{\hat{e}}^2, \sigma_v^2, \sigma_1^2, \ldots, \sigma_{18}^2\}$ where σ_e^2, $\sigma_{\tilde{e}}^2$, $\sigma_{\hat{e}}^2$, and σ_v^2 denote the variances of e_t, $\tilde{e}_t$, $\hat{e}_t$ and v_t respectively. The variance of u_t^i is σ_i^2 for $i = 1, \ldots, 18$. For the purpose of estimation, I assume that all variances are constant, so the covariance matrices for $\mathbb{V}_t$ and $\mathbb{U}_t$ can be written as Σ_v and Σ_u respectively. The sample likelihood function is built up recursively by applying the Kalman Filter to (28) and (32). Let n_t denote the number of data releases on day t. The sample log likelihood function for a sample spanning $t = 1, \ldots T$ is

$$\mathcal{L}(\theta) = \sum_{t=1, n_t > 0}^{T} \left\{ -\frac{n_t}{2} \ln(2\pi) - \frac{1}{2} \ln|\omega_t| - \frac{1}{2} \eta_t' \omega_t^{-1} \eta_t \right\}, \tag{33}$$

where η_t denotes the vector of innovations on day t with $n_t > 0$, and ω_t is the associated conditional covariance matrix. The η_t and ω_t sequences are calculated as functions of θ from the filtering equations:

$$\mathbb{Z}_{t|t} = \mathbb{A}_t \mathbb{Z}_{t-1|t-1} + \mathbb{K}_t \eta_t, \tag{34a}$$

$$\mathbb{S}_{t+1|t} = \mathbb{A}_t \left(I - \mathbb{K}_t \mathbb{B}_t \mathbb{C}_t \right) \mathbb{S}_{t|t-1} \mathbb{A}_t' + \Sigma_v, \tag{34b}$$

where

$$\eta_t = \mathbb{Y}_t - \mathbb{B}_t \mathbb{C}_t \mathbb{A}_t \mathbb{Z}_{t-1|t-1}, \tag{35a}$$

$$\mathbb{K}_t = \mathbb{S}_{t|t-1} \mathbb{C}_t' \mathbb{B}_t' \omega_t^{-1}, \tag{35b}$$

$$\omega_t = \mathbb{B}_t \mathbb{C}_t \mathbb{S}_{t|t-1} \mathbb{C}_t' \mathbb{B}_t' + \mathbb{B}_t \Sigma_u \mathbb{B}_t', \tag{35c}$$

if $n_t > 0$, and

$$\mathbb{Z}_{t|t} = \mathbb{A}_t \mathbb{Z}_{t-1|t-1}, \tag{36a}$$

$$\mathbb{S}_{t+1|t} = \mathbb{A}_t \mathbb{S}_{t|t-1} \mathbb{A}_t' + \Sigma_v, \tag{36b}$$

when $n_t = 0$. The recursions are initialized with $\mathbb{S}_{1|0} = \Sigma_v$ and $\mathbb{Z}_{0|0}$ equal to a vector of zeros. Notice that

(34) - (36) differ from the standard filtering equations because the structure of the state-space form in (28) and (32) changes via the $\mathbb{A}_t$, $\mathbb{C}_t$ and $\mathbb{B}_t$ matrices. The filtering equations also need to account for the days on which no data is released.

As in standard applications of the Kalman Filter, we need to insure that all the elements of θ are identified. Recall that equation (1) includes a error term v_t to allow for annual revisions to the "final" GDP data that take place after the release day $\text{R}_y(\tau)$. The variance of v_t, σ_v^2, is not identified because the state space form excludes data on the annual revisions. Rather than amend the model to include these data, I impose the identifying restriction: $\sigma_v^2 = 0$.[5] This restriction limits the duration of uncertainty concerning GDP growth to the reporting lag for the "final" GDP release. In Section 5, I show that most of the uncertainty concerning GDP growth in quarter τ is resolved by the end of the first month in quarter $\tau + 1$, well before the end of the reporting lag. Limiting the duration of uncertainty does not appear unduly restrictive.

3.3 Calculating the Real-Time Estimates of GDP

Once the maximum likelihood estimates of θ have been found, the Kalman Filtering equations can be readily used to calculate real-time estimates of GDP. Consider, first, the real-time estimates at the end of each quarter $\Delta^{\text{Q}} x_{\text{Q}(\tau)|\text{Q}(\tau)}$. By definition, $\mathbb{Z}_{t|j}$ denotes the expectation of $\mathbb{Z}_t$ conditioned on data released by the end of day j, $E[\mathbb{Z}_t|\Omega_j]$. Hence, the real-time estimate of quarterly GDP growth are given by

$$\Delta^{\text{Q}} x_{\text{Q}(\tau)|\text{Q}(\tau)} = E[s_{\text{Q}(\tau)}^{\text{Q}}|\Omega_{\text{Q}(\tau)}] = h_1 \hat{\mathbb{Z}}_{\text{Q}(\tau)|\text{Q}(\tau)}, \tag{37}$$

for $\tau = 1, ...,$ where h_i is a vector that selects the i'th. element of $\mathbb{Z}_t$. $\hat{\mathbb{Z}}_{t|t}$ denotes the value of $\mathbb{Z}_{t|t}$ based on the MLE of θ computed from (34) - (36). The Kalman Filter allows us to study how the estimates of $\Delta^{\text{Q}} x_{\text{Q}(\tau)}$ change in the light of data releases after the quarter has ended. For example, the sequence $\Delta^{\text{Q}} x_{\text{Q}(\tau)|t} = h_2 \hat{\mathbb{Z}}_{t|t}$, for $\text{Q}(\tau) < t \leq \text{Q}(\tau + 1)$, shows how data releases between the end of quarters τ and $\tau + 1$ change the real-time estimates $\Delta^{\text{Q}} x_{\text{Q}(\tau)}$.

We can also use the model to find real-time estimates of GDP growth before the end of the quarter. Recall that quarterly GDP growth can be represented as the sum of daily increments:

$$\Delta^{\text{Q}} x_{\text{Q}(\tau)} = \sum_{i=1}^{\text{D}(\tau)} \Delta x_{\text{Q}(\tau-1)+i}. \tag{8}$$

Real time estimates of $\Delta^{\text{Q}} x_{\text{Q}(\tau)}$ based on information Ω_t where $\text{Q}(\tau - 1) < t \leq \text{Q}(\tau)$ can be found by taking conditional expectations on both sides of this equation:

$$\Delta^{\text{Q}} x_{\text{Q}(\tau)|t} = E[s_t^{\text{Q}}|\Omega_t] + \sum_{h=1}^{\text{Q}(\tau)} E[\Delta x_{t+h}|\Omega_t]. \tag{38}$$

---

[5] In principle, the state space form could be augmented to accommodate the revision data, but the resulting state vector would have 40-odd elements because revisions can take place up to three years after the "final" GDP data is released. Estimating such a large state space system would be quite challenging. Alternatively, one could estimate σ_v^2 directly from the various vintages of "final" growth rates for each quarter, and then compute the maximum likelihood estimates of the other parameters conditioned on this value.

The first term on the right hand side is the real-time estimate of the partial sum s_t^{Q} defined in (12). Since s_t^{Q} is the first element in the state vector $\mathbb{Z}_t$, a real-time estimate of s_t^{Q} can be found as $E[s_t^{\mathrm{Q}}|\Omega_t] = h_1\hat{\mathbb{Z}}_{t|t}$. The second term in (38) contains real-time *forecasts* for the daily increments over the remaining days in the month. These forecasts can be easily computed from the process for the increments in (27):

$$E[\Delta x_{t+h}|\Omega_t] = \sum_{i=1}^{k} \phi_i E[\Delta^{\mathrm{M}(i)} x_t|\Omega_t]. \tag{39}$$

Notice that the real-time estimates of $\Delta^{\mathrm{M}(i)} x_t$ on the right hand side are also elements of the state vector $\mathbb{Z}_t$ so the real-time forecasts can be easily found from $\hat{\mathbb{Z}}_{t|t}$. For example, for the state space form with $k = 1$ described above, the real-time estimates can be computed as $\Delta^{\mathrm{Q}} x_{\mathrm{Q}(\tau)|t} = \left[h_1 + h_5\hat{\phi}_1(\mathrm{Q}(\tau) - t) \right] \hat{\mathbb{Z}}_{t|t}$, where $\hat{\phi}_1$ is the MLE of ϕ_1.

The model can also be used to calculate real-time estimates of log GDP, $x_{\mathrm{Q}(\tau)|i}$. Once again it is easiest to start with the end of quarter real-time estimates, $x_{\mathrm{Q}(\tau)|\mathrm{Q}(\tau)}$. Iterating on the identity $\Delta^{\mathrm{Q}} x_{\mathrm{Q}(\tau)} \equiv x_{\mathrm{Q}(\tau)} - x_{\mathrm{Q}(\tau-1)}$, we can write

$$x_{\mathrm{Q}(\tau)} = \sum_{i=1}^{\tau} \Delta^{\mathrm{Q}} x_{\mathrm{Q}(i)} + x_{\mathrm{Q}(0)}. \tag{40}$$

Thus, log GDP for quarter τ can be written as the sum of quarterly GDP growth from quarters 1 to τ, plus the initial log level of GDP for quarter 0. Taking conditional expectations on both sides of this expression gives

$$x_{\mathrm{Q}(\tau)|\mathrm{Q}(\tau)} = \sum_{i=1}^{\tau} \mathbb{E}[\Delta^{\mathrm{Q}} x_{\mathrm{Q}(i)}|\Omega_{\mathrm{Q}(\tau)}] + \mathbb{E}[x_{\mathrm{Q}(0)}|\Omega_{\mathrm{Q}(\tau)}]. \tag{41}$$

Notice that the terms in the sum on the right hand side are *not* the real-time estimates of GDP growth. Rather, they are current estimates (i.e. based on $\Omega_{\mathrm{Q}(\tau)}$) of past GDP growth. Thus, we cannot construct real-time estimates of log GDP by simply aggregating the real-time estimate of GDP growth from the current and past quarters.

In principle, $x_{\mathrm{Q}(\tau)|\mathrm{Q}(\tau)}$ could be found using estimates of $\mathbb{E}[\Delta^{\mathrm{Q}} x_{\mathrm{Q}(i)}|\Omega_{\mathrm{Q}(\tau)}]$ computed from the state-space form with the aid of the Kalman Smoother algorithm (see, for example, Hamilton 1994). An alternative approach is to apply the Kalman Filter to a modified version of the state-space form:

$$\mathbb{Z}_t^a = \mathbb{A}_t^a \mathbb{Z}_t^a + \mathbb{V}_t^a, \tag{42a}$$

$$\mathbb{Y}_t = \mathbb{B}_t \mathbb{C}_t^a \mathbb{Z}_t^a + \mathbb{B}_t \mathbb{U}_t, \tag{42b}$$

where

$$\mathbb{Z}_t^a \equiv \begin{bmatrix} \mathbb{Z}_t \\ x_t \end{bmatrix}, \qquad \mathbb{A}_t^a \equiv \begin{bmatrix} \mathbb{A}_t & \mathbf{0} \\ h_7 & 1 \end{bmatrix}, \qquad \mathbb{V}_t^a = \begin{bmatrix} I_7 \\ h_7 \end{bmatrix} \mathbb{V}_t, \text{ and } \mathbb{C}_t^a \equiv \begin{bmatrix} \mathbb{C}_t & \mathbf{0} \end{bmatrix}.$$

This modified state-space form adds the cumulant of the daily increments, $x_t \equiv \sum_{i=1}^{t} \Delta x_i + x_{\mathrm{Q}(0)}$, as the eight element in the augmented state vector $\mathbb{Z}_t^a$. At the end of the quarter when $t = \mathrm{Q}(\tau)$, the cumulant is equal to $x_{\mathrm{Q}(\tau)}$. So a real-time end-of-quarter estimate of log GDP can be computed as $x_{\mathrm{Q}(\tau)|\mathrm{Q}(\tau)} = h_8^a \hat{\mathbb{Z}}_{\mathrm{Q}(\tau)|\mathrm{Q}(\tau)}^a$, where $\hat{\mathbb{Z}}_{t|t}^a$ is the estimate of $\mathbb{Z}_t^a$ derived by applying that Kalman Filter to (42) and h_i^a is a vector that picks out

14

the i'th. element of $\mathbb{Z}_t^a$.

Real-time estimates of log GDP in quarter τ based on information available on day $t < \mathrm{Q}(\tau)$ can be calculated in a similar fashion. First, we use (7) and the definition of x_t to rewrite (40) as

$$x_{\mathrm{Q}(\tau)} = x_t + \sum_{h=t+1}^{\mathrm{Q}(\tau)} \Delta x_h.$$

As above, the real-time estimate is found by taking conditional expectations:

$$x_{\mathrm{Q}(\tau)|t} = \mathbb{E}[x_t|\Omega_t] + \sum_{h=t+1}^{\mathrm{Q}(\tau)} \mathbb{E}[\Delta x_h|\Omega_t]. \tag{43}$$

The real-time estimate of log GDP for quarter τ, based on information available on day $t \leq \mathrm{Q}(\tau)$ comprises the real time estimate of x_t and the sum of the real-time forecasts for Δx_{t+h} over the remainder of the quarter. Notice that each component on the right hand side was present in the real-time estimates discussed above, so finding $x_{\mathrm{Q}(\tau)|t}$ involves nothing new. For example, in the $k=1$ case, $x_{\mathrm{Q}(\tau)|t} = [h_8^a + h_5^a \hat{\phi}_1 (\mathrm{Q}(\tau) - t)]\hat{\mathbb{Z}}_{t|t}^a$.

To this point I have concentrated on the problem of calculating real-time estimates for GDP and GDP growth measured on a quarterly basis. We can also use the model to calculate real-time estimates of output flows over shorter horizons, such as a month or week. For this purpose, I first decompose quarterly GDP into its daily components. These components are then aggregated to construct estimates of output measured over any horizon.

Let d_t denote the log of output on day t. Since GDP for quarter τ is simply the aggregate of daily output over the quarter,

$$x_{\mathrm{Q}(\tau)} \equiv \ln \left(\sum_{i=1}^{\mathrm{D}(\tau)} \exp \left(d_{\mathrm{Q}(\tau-1)+i} \right) \right), \tag{44}$$

where $\mathrm{D}(\tau) \equiv \mathrm{Q}(\tau) - \mathrm{Q}(\tau-1)$ is the duration of quarter τ. Equation (44) describes the exact nonlinear relation between log GDP for quarter τ and the log of daily output. In principle, we would like to use this equation and the real-time estimates of $x_{\mathrm{Q}(\tau)}$ to identify the sequence for d_t over each quarter. Unfortunately, this is a form of nonlinear filtering problem that has no exact solution. Consequently, to make any progress we must work with either an approximate solution to the filtering problem, or a linear approximation of (44). I follow the second approach by working with a first order Taylor approximation to (44) around the point where $d_t = x_{(\tau)} - \ln \mathrm{D}(\tau)$:

$$x_{\mathrm{Q}(\tau)} \cong \frac{1}{\mathrm{D}(\tau)} \sum_{i=1}^{\mathrm{D}(\tau)} \left\{ d_{\mathrm{Q}(\tau-1)+i} + \ln \mathrm{D}(\tau) \right\}. \tag{45}$$

Combining (45) with the identity $\Delta^{\mathrm{Q}} x_{\mathrm{Q}(\tau)} \equiv x_{\mathrm{Q}(\tau)} - x_{\mathrm{Q}(\tau-1)}$ gives

$$\Delta^{\mathrm{Q}} x_{\mathrm{Q}(\tau)} \cong \frac{1}{\mathrm{D}(\tau)} \sum_{i=1}^{\mathrm{D}(\tau)} \left\{ d_{\mathrm{Q}(\tau-1)+i} - \left(x_{\mathrm{Q}(\tau-1)} - \ln \mathrm{D}(\tau) \right) \right\}. \tag{46}$$

This expression takes the same form as the decomposition of quarterly GDP growth in (7) with $\Delta x_t \cong \left\{ d_t - x_{\mathrm{Q}(\tau-1)} + \ln \mathrm{D}(\tau) \right\} / \mathrm{D}(\tau)$. Rearranging this expression gives us the following approximation for log

daily output

$$d_t \cong x_{Q(\tau-1)} + D(\tau)\Delta x_t - \ln D(\tau). \tag{47}$$

According to this approximation, all the within-quarter variation in the log of daily output is attributable to daily changes in the increments Δx_t. Thus changes in x_t within each quarter provide an approximate (scaled) estimate of the volatility in daily output.

The last step is to construct the new output measure based on (47). Let x_t^h denote the log flow of output over h days ending on day t: $x_t^h \equiv \ln\left(\sum_{i=0}^{h-1}\exp\left(d_{t-i}\right)\right)$. As before, I avoid the problems caused by the nonlinearity in this definition by working with a first order Taylor approximation to x_t^h around the point where $d_t = x_t^h - \ln h$. Combining this approximation with (47) and taking conditional expectations gives

$$x_{t|t}^h \cong \frac{1}{h}\sum_{i=1}^{h-1}\left\{E[x_{Q(\tau_{t-i}-1)}|\Omega_t] + D(\tau_{t-i})E[\Delta x_{t-i}|\Omega_t] - \ln D(\tau_{t-i}) + \ln h\right\}, \tag{48}$$

where τ_t denotes the quarter in which day t falls. Equation (48) provides us with an approximation for the real-time estimates of x_t^h in terms that can be computed from the model. In particular, if we augment the state vector to include $x_{Q(\tau_{t-i}-1)}$ and Δx_{t-i} for $i = 1, ..., h-1$, and apply the Kalman filter to the resulting modified state-space, the estimates of $E[x_{Q(\tau_{t-i}-1)}|\Omega_t]$ and $E[\Delta x_{t-i}|\Omega_t]$ can be constructed from $\hat{\mathbb{Z}}_{t|t}^a$.

4 Empirical Results

4.1 Data

The macroeconomic data releases used in estimation are from International Money Market Services (MMS). These include real-time data on both expected and announced macro variables. I estimate the model using the three quarterly GDP releases and the monthly releases on 18 other variables from 4/11/93 until 6/30/99. In specification tests described below, I also use market expectations of GDP growth based on surveys conducted by MMS of approximately forty money managers on the Friday of the week before the release day. Many earlier studies have used MMS data to construct proxies for the news contained in data releases (see, for example, Urich and Watchel 1984, Balduzzi et al. 2001, and Andersen et al. 2003). This is the first paper to use MMS data in estimating real time estimates of macroeconomic variables.

The upper panel of Table 1 lists the data series used in estimation. The right had columns report the number of releases and the range of the reporting lag for each series during the sample period. The lower panel shows the distribution of data releases. The sample period covers 1682 workdays (i.e. all days excluding weekends and national holidays).[6] On approximately 55% of these days there was a least one data release. Multiple data releases occurred much less frequently; on approximately 16% of the workdays in the sample. The were no occasions when more that four data releases took place.

---

[6]Although economic activity obviously takes place on weekend and holidays, I exclude these days from the sample for two reasons. First, they contain no data releases. This means that the contribution to GDP on weekends and holidays must be exclusively derived from the dynamics of (27). Second, by including only workdays we can exactly align the real-time estimates with days on which US financial markets were open. This feature will be very helpful in studying the relation between the real-time estimates and asset prices.

Table 1: Data Series (4/11/93 - 6/30/99)

	Release	Obs.	Reporting Lag
Quarterly	Advanced GDP	26	1-2 Months
	Preliminary GDP	25	2-3 Months
	Final GDP	26	3-4 Months
Monthly			
Real Activity	Nonfarm Payroll Employment	78	3 - 9 days
	Retail Sales	78	12-15 days
	Industrial Production	78	15-18 days
	Capacity Utilization	78	15-18 days
	Personal Income	76	30-33 days
	Consumer Credit	78	33-40 days
Consumption	Personal Consumption Expenditures	76	30-33 days
	New Home Sales	77	27-33 days
Investment	Durable Goods Orders	77	24-29 days
	Construction Spending	77	31-34 days
	Factory Orders	76	29-35 days
	Business Inventories	78	38-44 days
Government	Government Budget Deficit	78	15-21 days
Net Exports	Trade Balance	78	44-53 days
Forward-looking	Consumer Confidence Index	78	-8-0 days
	NAPM index	78	0-6 days
	Housing Starts	77	14-20 days
	Index of Leading Indicators	78	27-45 days
Distribution of Data Releases			
Releases per Day	Fraction of Sample	Observations	
0	45.48%	765	
1	38.76%	652	
2	10.46%	176	
3	4.34%	73	
4	0.95%	16	
> 0	54.52%	917	

The release data were transformed in two ways before being incorporated in the model. First, I subtracted the sample mean from each of the GDP releases. This transformation implies that the real-time estimates presented below are based on the assumption that long-run GDP growth remained constant over the sample period. If the span of my data were considerably longer, I could identify how the long-run rate of GDP growth has varied by estimating a modified form of the model that replaced (27) with a process that decomposed Δx_t into short and long run components. I leave this extension of the model for future work.

The second transformation concerns the monthly data. Let $\tilde{z}^i_{\text{R}(\tau,j)}$ denote the raw value for series i released on day $t = \text{R}(\tau, j)$. The model incorporates transformed series $z^i_{\text{R}(\tau,j)} = \left(\tilde{z}^i_{\text{R}(\tau,j)} - \bar{z}^i\right) - \alpha_i\left(\tilde{z}^i_{\text{R}(\tau,j-1)} - \bar{z}^i\right)$ where $\bar{z}^i$ is the sample mean of $\tilde{z}^i$. Recall that in the model the monthly series provide noisy signals on the monthly contribution to GDP growth (see equations (24) - (26)). Quasi differencing in this manner allows each of the raw data series to have a differing degree of persistence than the monthly contribution to GDP growth without

inducing serial correlation in the projection errors shown in (24) - (26). The degree of quasi-differencing depends on the α_i parameters which are jointly estimated with the other model parameters.

4.2   Estimates and Diagnostics

The maximum likelihood estimates of the model are reported in Table 2. There are a total of 63 parameters in the model and all are estimated with a great deal of precision. T-tests based on the asymptotic standard errors (reported in parenthesis) show that all the coefficient are significant at the 1% level. Panel A of the table shows the estimated parameters of the daily contribution process in (27). Notice that the reported estimates and standard errors are multiplied by 25. With this scaling the reported values for the ϕ_i parameters represent the coefficients in the time-aggregated AR(6) process for $\Delta^{\mathrm{M}} x_{\mathrm{M}(\tau,j)}$ in a typical month (i.e. one with 25 workdays). I shall examine the implications of these estimates for forecasting GDP below.

Panel B reports the estimated standard deviations of the difference between the "advanced" and "final" GDP releases, $\omega_a \equiv \hat{y}_t - y_t$, and the difference between the "preliminary" and "final" releases $\omega_p \equiv \tilde{y}_t - y_t$. According to equations (5) and (6) of the model, $\mathbb{V}(\omega_a) = \mathbb{V}(\omega_p) + \mathbb{V}(\hat{e}_{\mathrm{R}(\tau)})$ so the standard deviation of ω_a should be at least as great as that of ω_p. By contrast, the estimates in panel B imply that $\mathbb{V}(\omega_a) < \mathbb{V}(\omega_p)$.[7] This suggests that revisions the BEA made between releasing the "preliminary" and "final" GDP data were negatively correlated with the revision between the "advanced" and "preliminary" releases. It is hard to understand how this could be a feature of an optimal revision process within the BEA. However, it is also possible that the implied correlation arises simply by chance because the sample period only covers 25 quarters.

Estimates of the parameters linking the monthly data releases to GDP growth are reported in Panel C. The first column shows that there is considerable variation across the 18 series in the estimates of α_i. In all cases the estimates of α_i are statistically significant indicating that the quasi-differenced monthly releases are more informative about GDP growth than the raw series. The α_i estimates also imply that the temporal impact of a change in growth varies across the different monthly series. For example, changes in GDP growth will have a more persistent effect on the Consumer Confidence Index ($\hat{\alpha}_{15} = 0.977$) than on Nonfarm Payroll Employment ($\hat{\alpha}_1 = 0.007$). The β_i estimates reported in the third column show that 12 of the 18 monthly releases are pro-cyclical (i.e. positively correlated with contemporaneous GDP growth). Recall that all the coefficients are significant at the 1% level so the β_i estimates provide strong evidence that all the monthly releases contain incremental information about current GDP growth beyond that contained in past GDP data releases.

The standard method for assessing the adequacy of a model estimated by the Kalman Filter is to examine the properties of the estimated filter innovations, $\hat{\eta}_t$ defined in (35a) above. If the model is correctly specified, all elements of the innovation vector $\hat{\eta}_t$ should be uncorrelated with any elements of Ω_{t-1}, including past innovations. To check this implication, Table 3 reports the autocorrelation coefficients for the innovations associated with each data release. For example, the estimated innovation associated with the "final" GDP release for quarter τ on day $\mathrm{R}(\tau)$, is $\eta^y_{\mathrm{R}(\tau)} \equiv y_{\mathrm{R}(\tau)} - \mathbb{E}[y_{\mathrm{R}(\tau)}|\hat{\Omega}_{\mathrm{R}(\tau)-1}]$. For the quarterly releases, the table shows the correlation between $\eta^y_{\mathrm{R}(\tau)}$ and $\eta^y_{\mathrm{R}(\tau-n)}$ for $n = 1$ and 6. In the case of monthly release i, the

---

[7] To check robustness, I also estimated the model with the $\mathbb{V}(\omega_a) = \mathbb{V}(\omega_p) + \mathbb{V}(\hat{e}_{\mathrm{R}(\tau)})$ restriction imposed. In this case the MLE of $\mathbb{V}(\hat{e}_{\mathrm{R}(\tau)})$ is less than 0.0001.

Table 2: Model Estimates

A: Process for Δx_t

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	σ_e
estimate[**]	−0.384	0.296	0.266	-0.289	-0.485	0.160	3.800
standard error[**]	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.010)

B: Quarterly Data Releases

		$\mathbb{V}(\omega_i)$	sdt$(\omega_i)^*$
$i = a$	Advanced GDP Growth	0.508	(0.177)
p	Preliminary GDP Growth	1.212	(0.312)

C: Monthly Data Releases

		α_i	sdt$(\alpha_i)^*$	β_i	sdt$(\beta_i)^*$	σ_i	sdt$(\sigma_i)^*$
$i =1$	Nonfarm Payroll Employment	0.007	(0.218)	0.656	(0.301)	0.932	(0.171)
2	Retail Sales	-0.047	(0.282)	0.285	(0.136)	0.381	(0.082)
3	Industrial Production	-0.028	(0.145)	0.189	(0.090)	0.229	(0.035)
4	Capacity Utilization	0.924	(0.088)	0.125	(0.114)	0.382	(0.020)
5	Personal Income	-0.291	(0.219)	0.038	(0.126)	0.227	(0.040)
6	Consumer Credit	0.389	(0.300)	-0.160	(0.966)	2.961	(0.494)
7	Personal Cons. Expenditures	-0.405	(0.206)	0.133	(0.074)	0.111	(0.029)
8	New Home Sales	0.726	(0.170)	-0.011	(0.171)	0.473	(0.071)
9	Durable Goods Orders	-0.224	(0.258)	0.989	(0.753)	1.999	(0.413)
10	Construction Spending	0.312	(0.197)	-0.135	(0.233)	0.655	(0.123)
11	Factory Orders	-0.194	(0.288)	0.997	(0.489)	-0.856	(0.306)
12	Business Inventories	0.128	(0.277)	-0.019	(0.061)	0.228	(0.032)
13	Government Budget Deficit	-0.359	(0.418)	-0.992	(1.423)	3.262	(0.508)
14	Trade Balance	0.819	(0.189)	0.361	(0.602)	1.585	(0.344)
15	Consumer Confidence Index	0.977	(0.076)	0.208	(0.136)	-0.482	(0.084)
16	NAPM index	0.849	(0.115)	-0.008	(0.047)	0.151	(0.024)
17	Housing Starts	0.832	(0.175)	0.002	(0.026)	0.071	(0.014)
18	Index of Leading Indicators	0.107	(0.240)	0.212	(0.077)	0.231	(0.033)

Notes: "*" and "**" indicate that the estimate or standard error is multiplied by 100 and 25 respectively.

innovation is $\eta^i_{\text{R}(\tau,j)} \equiv z^i_{\text{R}(\tau,j)} - \mathbb{E}[z^i_{\text{R}(\tau,j)}|\hat{\Omega}_{\text{R}(\tau,j)-1}]$ and the table shows the correlation between $\eta^i_{\text{R}(\tau,j)}$ and $\eta^i_{\text{R}(\tau,j-n)}$ for $n = 1$ and 6. Under the BPQ(j) headings the table also reports p-values computed from the Box-Pierce Q statistic for joint significance of the correlations from lag 1 to j. Overall, there is little evidence of serial correlation in the innovations. Exceptions arise only in the case of "preliminary" GDP at the six quarter lag, and in the cases of consumer credit and business inventories at the six month lag.

Panel A of Table 4 compares model-based forecasts for "final" GDP against the provisional data releases. Under the Data Revision columns I report the mean and mean squared error (M.S.E.) for the data revisions associated with the "advanced" and "preliminary" data releases (i.e., $y_{\text{R}(\tau)} - \hat{y}_{\text{R}(\tau)}$ and $y_{\text{R}(\tau)} - \tilde{y}_{\text{R}(\tau)}$). The mean and M.S.E. for the different between "final" GDP, $y_{\text{R}(\tau)}$, and the estimates of $\mathbb{E}[y_{\text{R}(\tau)}|\Omega_{\text{DR}(\tau)}]$ where

Table 3: Model Diagnostics

Innovation Autocorrelations		ρ_1	BPQ(1)	ρ_6	BPQ (6)
Quarterly Releases					
	Advanced GDP	0.058	(0.766)	-0.061	(0.889)
	Preliminary GDP	-0.364	(0.069)	-0.034	(0.012)
	Final GDP	0.001	(0.996)	-0.172	(0.729)
Monthly Releases					
$i = 1$	Nonfarm Payroll Employment	-0.023	(0.841)	0.051	(0.902)
2	Retail Sales	0.005	(0.966)	-0.028	(0.789)
3	Industrial Production	0.005	(0.963)	0.003	(0.981)
4	Capacity Utilization	-0.029	(0.800)	0.147	(0.885)
5	Personal Income	-0.069	(0.687)	0.057	(0.770)
6	Consumer Credit	-0.091	(0.422)	0.310	(0.040)
7	Personal Consumption Expenditures	0.122	(0.477)	-0.021	(0.427)
8	New Home Sales	-0.219	(0.084)	-0.056	(0.220)
9	Durable Goods Orders	-0.094	(0.418)	-0.121	(0.650)
10	Construction Spending	0.064	(0.699)	0.131	(0.798)
11	Factory Orders	-0.161	(0.327)	-0.113	(0.483)
12	Business Inventories	-0.068	(0.552)	0.339	(0.000)
13	Government Budget Deficit	-0.091	(0.421)	-0.137	(0.100)
14	Trade Balance	-0.203	(0.077)	0.087	(0.578)
15	Consumer Confidence Index	0.047	(0.678)	-0.111	(0.624)
16	NAPM index	-0.067	(0.556)	-0.017	(0.639)
17	Housing Starts	-0.160	(0.161)	-0.127	(0.518)
18	Index of Leading Indicators	0.021	(0.850)	0.043	(0.525)

Notes: ρ_i denotes the sample autocorrelation at lag i. p-values are calculated for the null hypothesis of $\rho_i = 0$.

$\text{DR}(\tau)$ denotes the date of the day of either the "advanced" and "preliminary" release (i.e., $\text{R}_{\hat{y}}(\tau)$ or $\text{R}_{\tilde{y}}(\tau)$) are reported under the Model columns. Note that the provisional data are part of the information set $\Omega_{\text{DR}(\tau)}$ used to compute the model-based forecasts. If $\hat{y}_{\text{R}(\tau)}$ and $\tilde{y}_{\text{R}(\tau)}$ are close to the best forecasts of "final" GDP on days $\text{R}_{\hat{y}}(\tau)$ and $\text{R}_{\tilde{y}}(\tau)$, there should be little difference between the data revisions and the estimated forecast errors, $y_{\text{R}(\tau)} - \mathbb{E}[y_{\text{R}(\tau)}|\Omega_{\text{DR}(\tau)}]$.[8] The table shows that the mean and M.S.E. of revision errors based

---

[8] The model does not impose the assumption that the provisional data are equal to the best forecast of "final" GDP on day $\text{DR}(\tau)$. Rather, equation (34a) of the Kalman Filter implies that the model forecasts for "final" GDP on these days are given by

$$\mathbb{E}[y_{\text{R}(\tau)}|\Omega_{\text{DR}(\tau)}] = \mathbb{E}[y_{\text{R}(\tau)}|\Omega_{\text{DR}(\tau)-1}] + \mathbb{K}^y_{\text{DR}(\tau)}\left(\bar{y}_{\text{R}(\tau)} - \mathbb{E}[\bar{y}_{\text{R}(\tau)}|\Omega_{\text{DR}(\tau)-1}]\right) + \mathbb{K}^{\varkappa}_{\text{DR}(\tau)}\left(\varkappa_{\text{R}(\tau)} - \mathbb{E}[\varkappa_{\text{R}(\tau)}|\Omega_{\text{DR}(\tau)-1}]\right)$$

where $\bar{y}_{\text{R}(\tau)} = \{\hat{y}_{\text{R}(\tau)}, \tilde{y}_{\text{R}(\tau)}\}$ and $\varkappa_{\text{R}(\tau)}$ denotes the vector of other data releases on $\text{DR}(\tau)$. $\mathbb{K}^y_{\text{DR}(\tau)}$ and $\mathbb{K}^{\varkappa}_{\text{DR}(\tau)}$ are elements of the Kalman Gain matrix on release days. Inspection of this equation reveals that $\mathbb{E}[y_{\text{R}(\tau)}|\Omega_{\text{DR}(\tau)}] = \bar{y}_{\text{R}(\tau)}$ if $\mathbb{K}^y_{\text{DR}(\tau)} = 1$ and $\mathbb{K}^{\varkappa}_{\text{DR}(\tau)}$ is a vector of zeros. These conditions cannot be exactly satisfied when there is any noise in (5) and (6), but they could hold approximately if the noise variance is small relative the variance of other shocks. Under these circumstances, the provisional

on the "preliminary" data releases are comparable to those based on the model forecasts. In the case of the "advance" releases, by contrast, the mean revision error is roughly two-and-a-half times the size of the forecast error. This finding suggests that the "advanced" releases contain some "noise" and do not represent the best forecasts for "final" GDP that can be computed using publicly available data. It is also consistent with the regression findings reported by Dynan and Elmendorf (2001).

Table 4: Forecast Comparisons

A:	Data Revision		Model	
	Mean	M.S.E.	Mean	M.S.E.
Advanced	0.246	(0.446)	0.090	(0.441)
Preliminary	0.038	(0.067)	0.040	(0.066)
Combined	0.142	(0.257)	0.065	(0.254)

B:	M.M.S.		Model	
	Mean	M.S.E.	Mean	M.S.E
In-sample				
Advanced	0.729	(1.310)	0.190	(1.407)
Preliminary	0.160	(0.249)	0.096	(0.418)
Final	0.042	(0.062)	0.080	(0.395)
Combined	0.310	(0.540)	0.122	(0.740)
Out-of-sample				
Advanced	0.985	(1.464)	0.380	(1.500)
Preliminary	0.046	(0.178)	0.178	(0.801)
Final	-0.015	(0.057)	0.099	(0.208)
Combined	0.338	(0.566)	0.219	(0.836)

Panel B of Table 4 compares the forecasting performance of the model against the survey responses collected by MMS. On the Friday before each scheduled data release, MMS surveys approximately forty professional money managers on their estimate for the upcoming release. Panel B compares the median estimate from the surveys against the real-time estimate of GDP growth implied by the model on survey days. For example, in the first row under the MMS columns I report the mean and M.S.E. for the difference between $y_{\mathrm{R}(\tau)}$, and the median response from the survey conducted on the last Friday before the "advanced" GDP release on day $\mathrm{S}(\tau)$. The mean and M.S.E of the difference between $y_{\mathrm{R}(\tau)}$ and the estimate of $\mathbb{E}[y_{\mathrm{R}(\tau)}|\Omega_{\mathrm{S}(\tau)}]$ derived from the model are reported under the Model columns. As above, all the survey and model estimates are compared against the value for the "final" GDP release. This means that the forecasting horizon, (i.e. the difference between $\mathrm{R}(\tau)$ and $\mathrm{S}(\tau)$) falls from approximately 11 weeks in the first row, to five weeks in the second, and less than one week in the third. The fourth row reports the mean and M.S.E. at all three horizons.

The upper portion of Panel B compares the survey responses against model-based forecasts computed from parameter estimates reported in Table 2. These estimates are derived from the full data sample and

---

data could closely approximate the model's forecasts for "final" GDP (i.e., $\mathbb{E}[y_{\mathrm{R}(\tau)}|\Omega_{\mathrm{DR}(\tau)}] \simeq \bar{y}_{\mathrm{R}(\tau)}$).

so contain information that was not available to the money mangers at the time they were surveyed. The lower portion of Panel B reports on a pseudo out-of-sample comparison. Here the model-based forecasts are computed from model estimates obtained from the first half of the sample (4/11/93 - 3/31/96). These estimates are then used to compute model-based forecasts of "final" GDP on the survey days during the second half of the sample (4/01/96 - 6/30/99). The table compares the mean and M.S.E. of these out-of-sample forecasts against the survey responses during this latter period.

The table shows that both the mean and M.S.E. associated with both the survey and model forecasts fall with the forecasting horizon. In the case of the in-sample statistics, the median survey response provides a superior forecast than the model in terms of mean and M.S.E at short horizons. Both the mean and M.S.E. are smaller for survey responses than the model forecasts in the third row. Moving up a row, the evidence is ambiguous. The model produces a smaller mean but larger M.S.E. than the median survey. In the first row, the balance of the evidence favors the model; the M.S.E. is slightly higher but the mean is much lower than the survey estimates. This general pattern in repeated in the out-of-sample statistics. The strongest support for the model again comes from a comparison of the survey and model forecasts conducted one week before the "advanced" release. In this case the mean forecast error from the model is approximately 60 per smaller than the mean survey error. Over shorter forecasting horizons, the survey measures dominate the model-based forecasts.

Overall, these forecast comparisons provide rather strong support for the model. It is clear that the in-sample comparisons use model estimates based in part on information that was not available the money managers at the time. But it is much less clear whether this puts the managers at a significant informational disadvantage. Remember that the money managers had access to private information and other contemporaneous data that is absent from the model. Moreover, we are comparing a model-based forecast against the median forecast from a forty manager survey. In the out-of-sample comparisons, the informational advantage clearly lies with the managers. Here the model forecast are based on a true subset of the information available to managers so the median forecast from a forty manager survey should outperform the model. This is what we see when the forecasting horizon is less than five weeks. At longer horizons, the use of private information imparts less of a forecasting advantage to money managers.[9] In fact, the results suggest that as we move the forecasting day back towards then end of quarter τ, both the in and out-of-sample model-based forecasts outperform the surveys. When the forecasting day is pushed all the way back to the end of the quarter, the model-based forecast gives us the real-time estimates of GDP growth. Thus, the results in panel B indicate that real-time estimates derived from the model should be at least comparable to private forecasts based on much richer information sets.

---

[9]This advantage might be further reduced if each of the model forecasts were computed using parameter estimates that utilized all the data available on the survey date rather than a single set of estimates using data from the first half of the sample. A full-blown real-time forecasting exercise of this kind would be computationally demanding because the model would have to be repeatedly estimated, but it should also give superior model-based forecasts. For this reason, the out-of-sample exercise undertaken here probably understates the true real-time forecasting potential of the model.

5 Analysis

This section examines the model estimates. First, I consider the relation between the real-time estimates and the "final" GDP releases. Next, I compare alternative real-time estimates for the level of GDP and examine the forecasting power of the model. Finally, I study how the monthly releases are related to changes in GDP at a monthly frequency.

5.1 Real-time Estimates over the Reporting Lag

Figure 2 allows us to examine how the real-time estimates of GDP growth change over the reporting lag. The green line plots the "final" GDP growth for quarter τ released on day $\text{R}_y(\tau)$. The blue line plots the real-time estimates of the GDP growth last month $\Delta^{\text{Q}} x_{\text{Q}(\tau)|t}$ where $\text{Q}(\tau) < t \leq \text{R}_y(\tau)$ for each quarter. The vertical dashed portion represent the discontinuity in the series at the end of each quarter (i.e. on day $\text{Q}(\tau)$).[10] Several features of the figure stand out. First, the real-time estimates vary considerably in the days immediately after the end of the quarter. For example, at the end of 1994, the real-time estimates of GDP growth in the forth quarter change from approximately 1.25% to 2.25% and then to 1.5% in the space of a few days. Second, in many cases there is very little difference between the value for "final" GDP and the real-time estimate immediately prior to the release (i.e., $y_{\text{R}(\tau)} \simeq \Delta^{\text{Q}} x_{\text{Q}(\tau)|\text{R}(\tau)-1}$). In these cases, the "final" release contains no new information about GDP growth that was not already inferred from earlier data releases. In cases where the "final" release contains significant new information, the blue plot "jumps" to meet the green plot on the release day.

Figure 3 provides further information on the relation between the real-time estimates and the "final" data releases. Here I plot the variance of $\Delta^{\text{Q}} x_{\text{Q}(\tau)}$ conditioned on information available over the reporting lag; $\mathbb{V}(\Delta^{\text{Q}} x_{\text{Q}(\tau)}|\Omega_{\text{Q}(\tau)+i})$ for $0 < i \leq \text{R}_y(\tau)-\text{Q}(\tau)$. Estimates of this variance are identified as the 2'nd. diagonal element in $\mathbb{S}_{t+1|t}$ obtained from the Kalman Filter evaluated at the maximum likelihood estimates. Figure 3 plots the sample average of $\mathbb{V}(\Delta^{\text{Q}} x_{\text{Q}(\tau)}|\Omega_{\text{Q}(\tau)+i})$ together with a 95% confidence band. Although the path for the conditional variance varies somewhat from quarter to quarter, the narrow confidence band shows that the average pattern displayed in the figure is in fact quite representative of the variance path seen throughout the sample.

Figure 3 clearly shows how the flow of data releases during the reporting lag provides information on $\Delta^{\text{Q}} x_{\text{Q}(\tau)}$. In the first 20 days or so the variance falls by approximately 25% as information from the monthly releases provides information on the behavior of GDP during the previous month (month 3 of quarter τ). The variance then falls significantly following the "advanced" GDP release. The timing of this release occurs between 19 and 23 working days after the end of the quarter so the averaged variance path displayed by the figure spreads the fall across these days. Thereafter, the variance falls very little until the end of the reporting lag when the "final" value for GDP growth is released.[11] This pattern indicates that the "preliminary" GDP release provides little new information about GDP growth beyond that contained in the "advanced" GDP

---

[10]In cases where the reporting lag is less than one quarter, the discontinuity occurs (a couple of days) after $\text{R}_y(\tau)$ so the end of each solid blue segment meets the turning point in the green line identifying the "final" GDP release. When the reporting lag is longer than one quarter, there is a horizontal gap between the end of a blue segment and the next turning point on the green line equal to $\text{R}_y(\tau)-\text{Q}(\tau+1)$ days.

[11]Although the variance falls immediately to zero on the day of the release, the averaged variance falls to zero over several days in the figure because the reporting lag varies from 58 to 65 workdays in the sample.
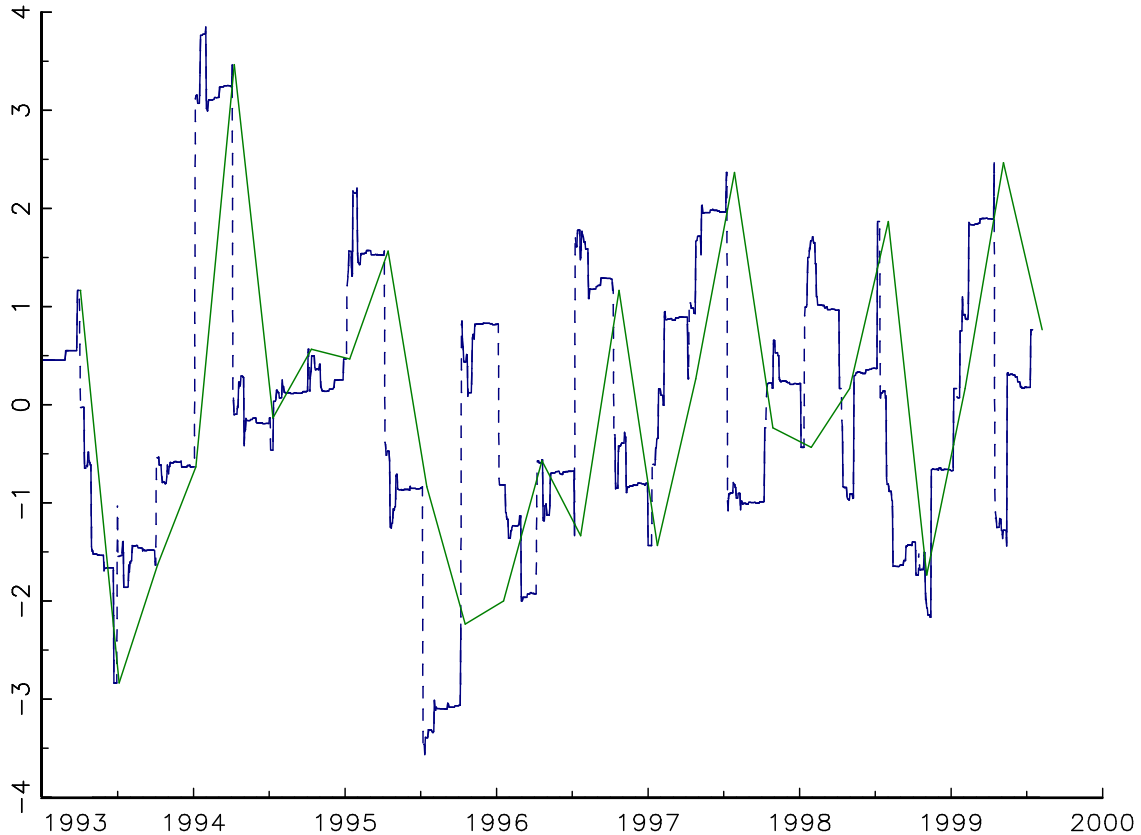
Figure 2: Real-time esimtates of quarterly GDP growth $\Delta^{\mathrm{Q}} x_{\mathrm{Q}(\tau)|t}$ where $\mathrm{Q}(\tau) < t \leq \mathrm{R}_y(\tau)$ (blue line), and "final" releases for GDP growth (green line).

release and the monthly data. The figure also shows that most of the uncertainty concerning GDP growth in the last quarter is resolved well before the day when the "final" data is released.

5.2 Real-Time Estimates of GDP and GDP Growth

Figure 4 compares the real-time estimates of log GDP against the values implied by the "final" GDP growth releases. The red line plots the values of $x_{\mathrm{Q}(\tau)|t}$ computed from (43):

$$x_{\mathrm{Q}(\tau)|t} = \mathbb{E}[x_t|\Omega_t] + \sum_{h=1}^{\mathrm{Q}(\tau)-t} \mathbb{E}[\Delta x_{t+h}|\Omega_t]. \tag{44}$$

Recall that $x_{\mathrm{Q}(\tau)}$ represents log GDP for quarter τ, so $x_{\mathrm{Q}(\tau)|t}$ includes forecasts for Δx_{t+h} over the remaining days in the quarter when $t < \mathrm{Q}(\tau)$. To assess the importance of the forecast terms, Figure 4 also plots $\mathbb{E}[x_t|\Omega_t]$ in blue. This series represents a naive real-time estimate of GDP since it assumes $\mathbb{E}[\Delta x_{t+h}|\Omega_t] = 0$ for $1 \leq h \leq \mathrm{Q}(\tau) - t$. The dashed green line in the figure plots the cumulant of the "final" GDP releases $\sum_{i=1}^{\tau} y_{\mathrm{R}(i)}$ with a lead 60 days. This plot represents an ex post estimate of log GDP based on the "final"
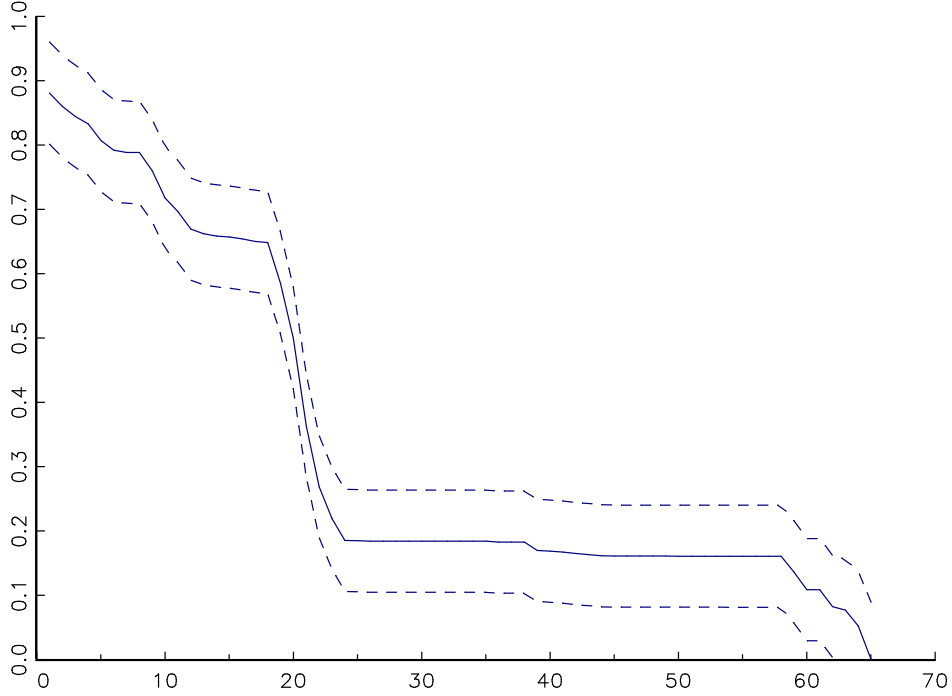
Figure 3: The solid line is the sample average of $\mathbb{V}(\Delta^{\mathrm{Q}}x_{\mathrm{Q}(\tau)}|\Omega_{\mathrm{Q}(\tau)+i})$ for $0 < i \leq \mathrm{R}_y(\tau) - \mathrm{Q}(\tau)$, and the dashed lines denote the 95% confident band. The horizontal axis marks the the number of days i past the end of quarter $\mathrm{Q}(\tau)$.

data releases. The vertical steps identify the values for "final" GDP growth 60 days before the actual release day.

Figure 4 displays three notable features. First, both sets the real-time estimates display a much greater degree of volatility than the cumulant series. This volatility reflects how inferences about current GDP change as information arrives in the form of monthly data releases during the current quarter and GDP releases referring to growth in the previous quarter. The second noteworthy feature concerns the relation between the real-time estimates. The vertical difference between the red and blue plots represents the contribution of the Δx_{t+h} forecasts to $x_{\mathrm{Q}(\tau)|t}$. As the figure shows, these forecasts contributed significantly to the real-time estimates in 1996 and 1997, pushing the real-time estimates of $x_{\mathrm{Q}(\tau)}$ well below the value for $\mathbb{E}[x_t|\Omega_t]$. The third noteworthy feature of the figure concerns the vertical gap between the red and green plots. This represents the difference between the real-time estimates and an ex post estimate of log GDP based on the "final" GDP releases. This gap should be insignificant if the current level of GDP could be precisely inferred from contemporaneously available data releases. Figure 3 shows this to be the case during the third and forth quarters of 1995. During many other periods, the real-time estimates were much less precise.
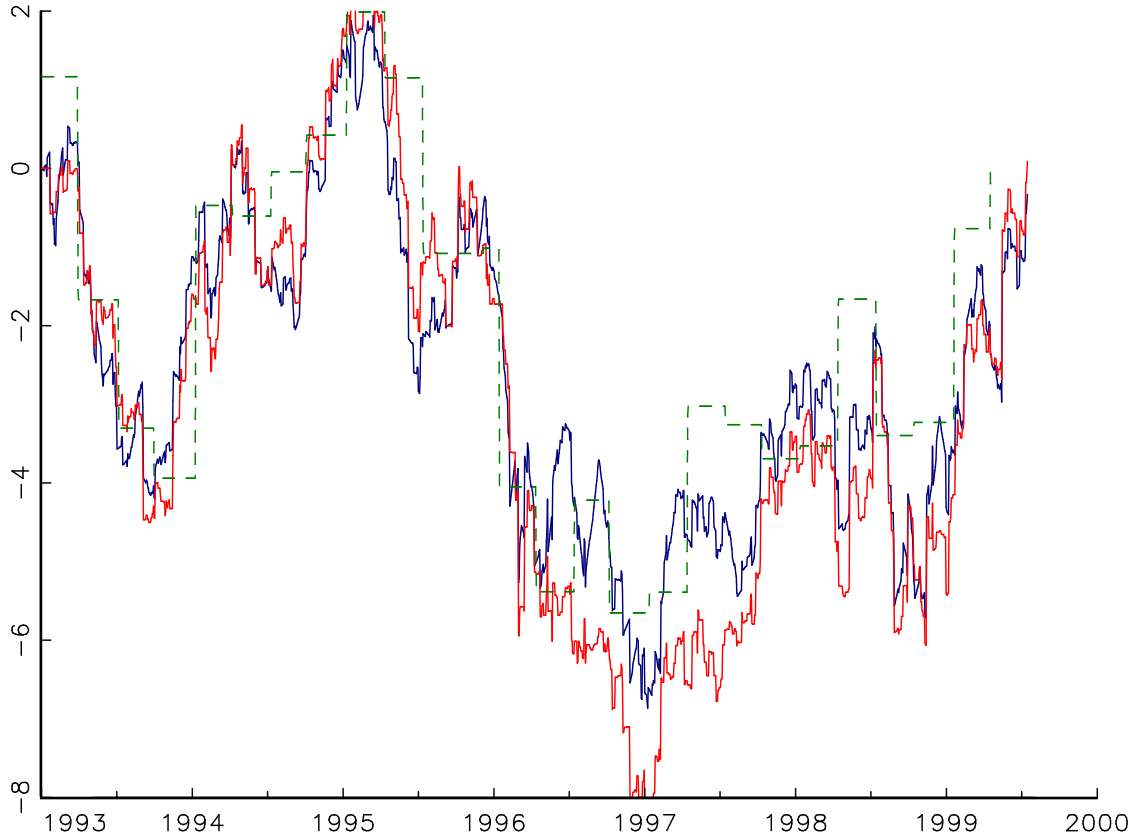
25

Figure 4: Real-time estimates of log GDP $x_{\mathrm{Q}(\tau)|t}$ (red line), $E[x_t|\Omega_t]$ (blue line) and cumulant of "final" GDP releases (green dashed line).

5.3 Forecasting GDP Growth

The model estimates in Table 2 show that all of the ϕ_i coefficients in the daily growth process are statistically significant. I now examine their implications for forecasting GDP growth. Consider the difference between the real time estimates of x_{t+m} and x_{t+n} based on information available on days $t+m$ and $t+n$, where $m > n$:

$$x_{t+m|t+m} - x_{t+n|t+n} = \sum_{h=t+n+1}^{t+m} \mathbb{E}\left[\Delta x_h|\Omega_t\right] + \left(x_{t+m|t+m} - x_{t+m|t}\right) - \left(x_{t+n|t+n} - x_{t+n|t}\right). \tag{49}$$

This equation decomposes the difference in real-time estimates of x_t into forecasts for Δx_t over the forecast horizon between $t+n$ and $t+m$, the revision in the estimates of x_{t+m} between t and $t+m$ and x_{t+n} between t and $t+n$.

Since both revision terms are uncorrelated with elements of Ω_t, we can use (49) to examine how the predictability of Δx_t implied by the model estimates translates into predictability for changes in $x_{t|t}$. In

26

particular, after multiplying both sides of (49) by $x_{t+m|t+m} - x_{t+n|t+n}$ and taking expectations, we obtain

$$\mathbb{R}^2(m,n) = \frac{\sum_{h=t+n+1}^{t+m} \mathbb{CV}\left(\mathbb{E}\left[\Delta x_h | \Omega_t\right], x_{t+m|t+m} - x_{t+n|t+n}\right)}{\mathbb{V}\left(x_{t+m|t+m} - x_{t+n|t+n}\right)}.$$

This statistic measures the contribution of Δx_t forecasts to the variance of $x_{t+m|t+m} - x_{t+n|t+n}$. If most of the volatility in $x_{t+m|t+m} - x_{t+n|t+n}$ is due to the arrival of new information between t and $t + m$ (i.e., via the revision terms in (49)), the $\mathbb{R}^2(m,n)$ statistic should be close to zero. Alternatively, if the the daily growth process is highly forecastable, much less of the volatility in $x_{t+m|t+m} - x_{t+n|t+n}$ will be attributable to news and the $\mathbb{R}^2(m,n)$ statistic will be positive.

Table 5: Forecasting Real-time GDP

m	$\mathbb{R}^2(m,n)$	(sdt.)
	Monthly ($n = m - 20$)	
20	0.196	(0.015)
40	0.186	(0.014)
60	0.180	(0.013)
80	0.163	(0.013)
100	0.138	(0.012)
120	0.125	(0.011)
140	0.097	(0.010)
160	0.061	(0.007)
180	0.032	(0.006)
200	0.035	(0.006)
220	0.027	(0.006)
240	0.005	(0.006)
	Quarterly: ($n = m - 60$)	
60	0.144	(0.008)
120	0.079	(0.006)
180	0.031	(0.003)
240	0.006	(0.002)

The table reports estimates of $R^2(m, m - 20)$ computed as the slope coefficient δ_m from the regression $\sum_{h=t+m-19}^{t+m} E\left[\Delta x_h | \Omega_t\right] = \delta_m \left(x_{t+m|t+m} - x_{t+m-20|t+m-20}\right) + \xi_{t+m}$ computed in daily data from the maximum likelihood estimates of $x_{t+m|t+m} - x_{t+m-20|t+m-20}$ and $E\left[\Delta x_h | \Omega_t\right]$. OLS standard errors are reported in the right hand column.

Table 5 reports estimates for $\mathbb{R}^2(m,n)$ for various forecasting horizons. The estimates are computed as the slope coefficient in the regression of $\sum_{h=t+n+1}^{t+m} \mathbb{E}\left[\Delta x_h | \Omega_t\right]$ on $x_{t+m|t+m} - x_{t+n|t+n}$ in daily data. The table also reports OLS standard errors in parenthesis.[12] The upper panel shows how predictable the monthly

---

[12] These statistics only approximate to the true standard errors for two reasons. First, the regressor is computed from the maximum likelihood estimates and so contians some sampling error. Second, there is no correction for MA process induced by the overlapping forecast horizons in the regression residuals.

changes in the real-time estimates of x_t are for horizons m of 20 to 240 days. The estimated process for Δx_t implies a reasonable high degree of predictability: the $\mathbb{R}^2$ estimates fall from approximately 20 to 10 per cent as horizon rises from 20 to 140 workdays. Estimates of $\mathbb{R}^2(m, n)$ for quarterly changes in the real-time estimates are reported in the lower panel of the table. These are somewhat smaller, but again clearly indicate the presence of some predictability.

The results in Table 5 related to the change in the real-time estimates of x_t rather than GDP growth. Recall from (43) that the real-time estimate of log GDP on day t in quarter τ is $x_{Q(\tau_t)|t} = \mathbb{E}[x_t|\Omega_t] + \sum_{h=t+1}^{Q(\tau_t)} \mathbb{E}[\Delta x_h|\Omega_t]$, so the growth in quarterly GDP between quarters τ and $\tau + 1$ based on the real-time estimates available at $t + m$ and $t + n$ (where $t + n < Q(\tau) < t + m < Q(\tau + 1)$) is

$$x_{Q(\tau+1)|t+m} - x_{Q(\tau)|t+n} = \left(x_{t+m|t+m} - x_{t+n|t+n}\right) + \sum_{h=t+m+1}^{Q(\tau+1)} \mathbb{E}[\Delta x_h|\Omega_{t+m}] - \sum_{h=t+n+1}^{Q(\tau)} \mathbb{E}[\Delta x_h|\Omega_{t+n}]. \quad (50)$$

This equation shows that changes in the real-time estimates of x_t are only one component of the estimated quarterly growth in GDP. Moreover, since the second and third terms on the right hand side of (50) will generally be correlated with the first term, the results in Table 5 may an unreliable guide to the predictability of GDP growth.
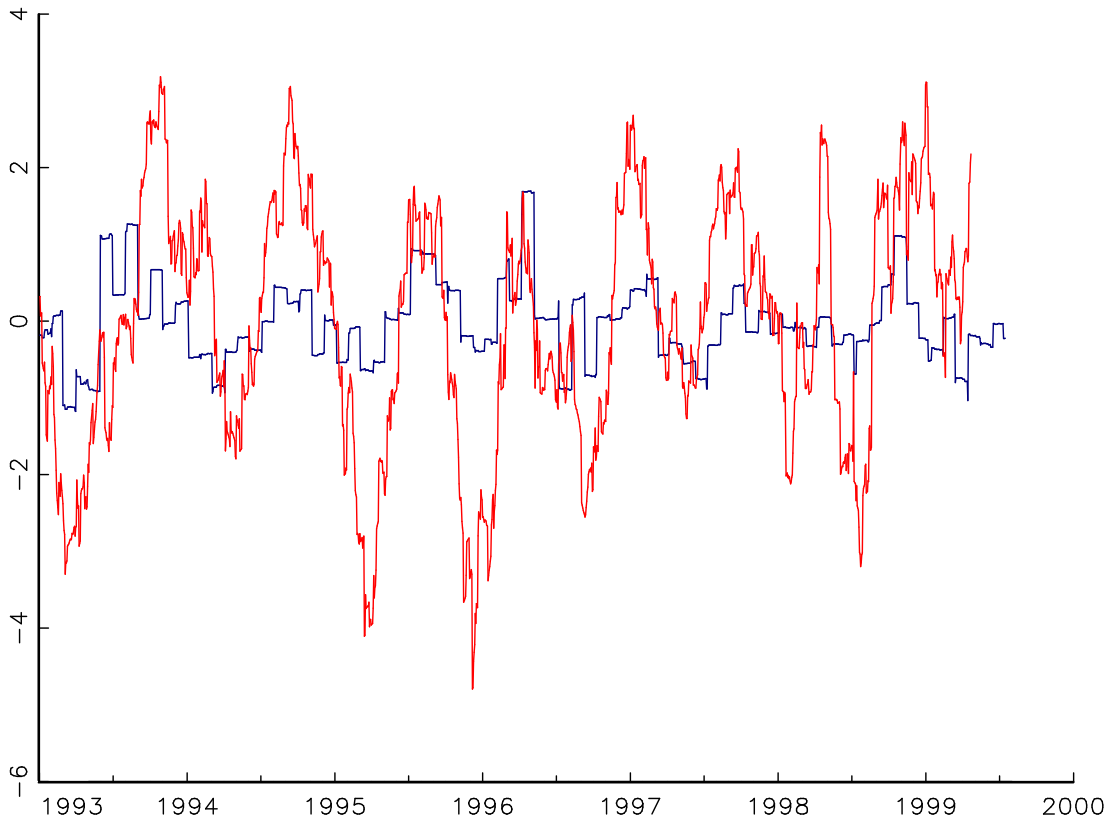

Figure 5: Real-time estimates of the quarterly growth in GDP, $x_{Q(\tau+1)|t+60} - x_{Q(\tau)|t}$ (red line), and ex ante forecasts, $\mathbb{E}[x_{Q(\tau+1)} - x_{Q(\tau)}|\Omega_t]$ (blue line).

28

We can examine the predictability of GDP growth by combining (49) and (50) to give

$$x_{Q(\tau+1)|t+m} - x_{Q(\tau)|t+n} = \mathbb{E}[x_{Q(\tau+1)} - x_{Q(\tau)+1}|\Omega_t] + \varsigma_{t+m},$$

where ς_{t+m} is an error term that depends on news that arrives between t and $t+m$. This equation shows that GDP growth should be predictable in the model because the process for Δx_t implies that $\mathbb{E}[x_{Q(\tau+1)} - x_{Q(\tau)+1}|\Omega_t]$ changes over the sample. Figure 5 plots the estimates of $\mathbb{E}[x_{Q(\tau+1)} - x_{Q(\tau)+1}|\Omega_t]$ and $x_{Q(\tau+1)|t+m} - x_{Q(\tau)|t+n}$ for $m = 60$ and $n = 0$. Clearly, news contributes significantly to the volatility of GDP growth since the $\mathbb{E}[x_{Q(\tau+1)} - x_{Q(\tau)+1}|\Omega_t]$ series is much less volatile than $x_{Q(\tau+1)|t+m} - x_{Q(\tau)|t+n}$. In fact, only 6 per cent of the variance in $x_{Q(\tau+1)|t+m} - x_{Q(\tau)|t+n}$ can be attributable to $\mathbb{E}[x_{Q(\tau+1)} - x_{Q(\tau)+1}|\Omega_t]$ over the sample. While this implies a rather modest degree of predictability, close inspection of Figure 5 suggests that predicting the future direction of GDP growth may be a little more successful. Indeed, the model estimates of $\mathbb{E}[x_{Q(\tau+1)} - x_{Q(\tau)+1}|\Omega_t]$ correctly predict the direction of GDP growth 59 per cent of the time.

For perspective on these forecasting results, I also estimated an AR(2) model for quarterly GDP growth using the sequence of "final" GDP releases. This model provides a simple time series forecast for GDP growth (approximately) one quarter ahead, based on the two most recent releases. In contrast to the results presented above, estimates of the AR(2) model do not indicate that quarterly GDP growth is at all predictable: The coefficients are small and statistically insignificant and the R^2 statistic is only 3 per cent. These results are hardly surprising. Remember that the sample mean was removed from all the GDP growth releases so we are attempting to forecast future *deviations* in GDP growth. Figure 2 shows that these deviations display little serial correlation. Thus, it is not surprising that the history of GDP releases has little forecasting power. These observations also point to the significance of the forecasting results displayed in Table 5 and Figure 5. In particular, they show that monthly data releases contain information that is useful for both estimating the current state of the GDP and for forecasting its future path. This aspect of the model ties in with recent research that harnesses the information in a large number of indicators for forecasting (see, for example, Stock and Watson 2002). The results from these studies suggest that the real-time forecasting performance of this model may be further enhanced by addition of other macroeconomic and financial indicators.

5.4 Monthly Estimates of GDP Growth

One of the unique features of the model is its ability to provide us with high frequency estimates of log GDP and GDP growth. I now examine how the monthly data releases relate to the changing real-time estimates of log GDP. My aim is to provide a simply description of the complex inference problem solved by the model regarding the current state of GDP.

Let t and $t+20$ be workdays in quarters τ_0 and τ_1 with $t \leq Q(\tau_0)$ and $t+20 \leq Q(\tau_1)$. (Note that τ_0 and τ_1 can refer to the same quarter.) I consider two regression models. The first relates the monthly change in the real-time estimates of log GDP to all 18 monthly releases:

$$x_{Q(\tau_1)|t+20} - x_{Q(\tau_0)|t} = \sum_{i=1}^{18} a_i(r_{t+20}^i - r_t^i) + \zeta_{t+20}, \tag{51}$$

where r_t^i denotes the last value released for series i on day t. Thus, the value of r_t^i remains that same from

day to day unless t is the day on which a data release for series i takes place. This means that $r^i_{t+20} - r^i_t$ identifies the change in the latest value for series i released during the 20 workdays ending on $t + 20$. The second model relates the change in the real-time estimates to each monthly release separately:

$$x_{Q(\tau_1)|t+20} - x_{Q(\tau_0)|t} = b_i(r^i_{t+20} - r^i_t) + \zeta_{t+20}, \qquad (52)$$

for $i = 1, 2, ..18$. Notice that the regressors in both models are available on a daily bases. Estimates of the a_i and b_i coefficients therefore summarize how the complex inference imbedded in the model relate to observable changes in the information set comprised of monthly data releases.

Table 6: Monthly Indicator Estimates

	Data Release	a_i	std(a_i)	b_i	std(b_i)	R^2
$i =1$	Nonfarm Payroll Employment	0.224*	(0.033)	0.330*	(0.050)	0.233
2	Retail Sales	0.520*	(0.074)	0.838*	(0.099)	0.186
3	Industrial Production	0.745*	(0.116)	1.047*	(0.142)	0.233
4	Capacity Utilization	-0.006	(0.034)	0.066	(0.057)	0.010
5	Personal Income	-0.110	(0.132)	0.163	(0.224)	0.003
6	Consumer Credit	0.001	(0.012)	-0.002	(0.019)	0.000
7	Personal Consumption Expenditures	0.377*	(0.182)	0.603*	(0.277)	0.033
8	New Home Sales	0.016	(0.069)	0.181*	(0.088)	0.039
9	Durable Goods Orders	0.013	(0.021)	0.064*	(0.027)	0.030
10	Construction Spending	-0.050	(0.051)	-0.102	(0.060)	0.015
11	Factory Orders	0.097*	(0.038)	0.160*	(0.043)	0.066
12	Business Inventories	-0.013	(0.152)	0.289	(0.247)	0.008
13	Government Budget Deficit	-0.008	(0.011)	-0.017	(0.016)	0.005
14	Trade Balance	-0.071*	(0.021)	-0.068*	(0.030)	0.042
15	Consumer Confidence Index	0.068*	(0.028)	0.060	(0.037)	0.021
16	NAPM index	-0.001	(0.145)	0.514*	(0.216)	0.046
17	Housing Starts	-1.330*	(0.421)	0.705	(0.546)	0.012
18	Index of Leading Indicators	0.164	(0.179)	0.476*	(0.199)	0.038

Notes: The table reports the OLS estimates of a_i and b_i from equations (51) and (52). Both equations are estimated at the daily frequency, and the standard errors are corrected for the MA(19) process induced by the overlapping data. The right hand column reports the R^2 statistic from estimating equation (52). A "*" denotes significance at the 5 per cent level.

Table 6 reports the estimates of equations (51) and (52). The left hand columns show that the a_i estimates are statistically significant for data on nonfarm payroll, retail sales, industrial production, personal consumption, factory orders, the trade balance, the index of consumer confidence and housing starts. Each of these data releases provides significant incremental information about the change in the real-time estimates. The 18 monthly releases account for approximately 57 per cent of the variance in $x_{Q(\tau_1)|t+20} - x_{Q(\tau_0)|t}$. This

means that more than 40 per cent of the variation in the real-time estimates is not captured by the simple linear specification. The right hand columns report the results from estimating equation (52). The most noteworthy aspects of these estimates can be seen for $i = 1, 2$, and 3. Changes in both nonfarm payroll and retail sales appear strongly linked to changes in the real-time estimates. Based on the R^2 statistics (in the right hand column), these variables account for 23 and 19 per cent of the variance in the real-time estimates. The results for $i = 3$ provide some justification for the frequent use of industrial production as a monthly proxy for GDP growth. The estimate for b_3 indicates the real-time estimates of log GDP change approximately one-for-one with change in industrial production between releases. Notice, however, that the R^2 statistic is only 0.233. Industrial production does not account for most of the variance in the real-time estimates.

6   Conclusion

In this paper I have presented a method for estimating the current state of the economy on a continual basis using the flow of information from a wide range of macroeconomic data releases. These real-time estimates were computed from an econometric model that allows for variable reporting lags, temporal aggregation, and other complications that characterize the daily flow of macroeconomic information. The model can be applied to find real-time estimates of GDP, inflation, unemployment or any other macroeconomic variable of interest.

In this paper I focused on the problem of estimating GDP in real time. This application of the estimation procedure should be of particular interest to policymakers concerned with the lack of timely information about economy-wide real activity. The real-time estimates I calculate have several noteworthy features: First, the estimates of log GDP display a good deal of high-frequency volatility. This volatility reflects how inferences about current GDP change as information arrives in the form of monthly data releases during the quarter. Second, the gaps between the real-time estimates and ex post GDP data are on occasion both persistent and significant. These findings suggest that the ex post data should not be viewed as a close approximation to what was known at the time. Third, the model estimates reveal that the monthly data releases contain information that is useful for forecasting its future path of GDP. Finally, my comparison of the real-time estimates with the monthly data series shows that standard proxies for real actively at the monthly frequency capture only a fraction of the variance in the real-time estimates.

These findings give but a flavor of the uses to which real-time estimates could be put. One obvious topic for the future concerns the ability of the real-time estimates and forecasts to identify turning points in the business cycle. This issue could be readily addressed if the model were re-estimated over a much longer time-span than was undertaken here. The use of real-time estimates may also bring new perspective to the link between asset prices and macroeconomic fundamentals. Evans and Lyons (2004a) use the methods described here to construct real-time estimates for GDP, inflation, and money supplies for the U.S. and Germany. With the aid of these estimates, they then show that foreign exchange transactions contain significant information about the future path of fundamentals. Since transaction flows also exert a very strong influence on exchange rate dynamics in the short run, this finding points to a much stronger link between fundamentals and exchange rates than pervious research has uncovered. It remains to be seen whether the use of real-time estimates will similarly illuminate the links between macro fundamentals and

other asset prices.

References

Andersen, T. G., T. Bollerslev, F. Diebold and C. Vega (2003), " Mirco Effect of Macro Announcements: Real-Time Price Discovery in Foreign Exchange", *American Economic Review*, 93(1), March, pp. 38-62.

Balduzzi, P., E. J. Elton and T. C. Green (2001), " Economic News and Bond Prices: Evidence from the U.S. Treasury Market", *Journal of Financial and Quantitative Analysis,* 36(4), December, pp. 523-43.

Carson, C. S. (1987), "GNP: An Overview of Source Data and Estimation Methods", *Survey of Current Business*, 67(7) July, pp.103-26.

Chow, G. and A. Lin (1971), "Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Time Series", *Review of Economics and Statistics,* 53, pp.372-75.

Clarida, R., J. Kitchen and R. Monaco (2002), "Real-Time Forecasting in Practice: A Description of the U.S. Treasury's Real-time GDP Forecasting System", mimeo, Office of Economic Policy, U.S. Department of the Treasury.

Croushore, D. and T. Stark (1999), "A Real-Time Data Set for Macroeconomists", Federal Reserve Bank of Philadelphia, working paper 99-4, June.

Croushore, D. and T. Stark (2001), ""A Real-Time Data Set for Macroeconomists", *Journal of Econometrics* 105, November, pp. 111-130.

Dynan, K. E. and D. W. Elmendorf (2001), "Do Provisional Estimates of Output Miss Economic Turning Points?", mimeo, Federal Reserve Board.

Evans, M.D.D. and R. K. Lyons (2004a), "A New Micro Model of Exchange Rate Dynamics", NBER working paper No. 10379, March.

Evans, M.D.D. and R. K. Lyons (2004b), "Do Currency Transaction Flows Forecast Fundamentals?", mimeo, March, (available at http://www.georgetown.edu /faculty/evansm1 /wpapers_files /flowsnber.pdf).

Faust, J., Rogers, J. and J. Wright (2000), "New and Noise in G-7 GDP Announcements", Federal Reserve Board International Finance Discussion Paper No. 2000-690.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

Harvey, A. C. (1989), *Forecasting Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, U.K.

Lui, H. and S. G. Hall (2000), " Creating High Frequency National Accounts with State Space Modelling: A Monte Carlo Experiment", mimeo, Imperial College, London, U.K.

Mariano, R. and Y. Murasawa (2003), "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series", *Journal of Applied Econometrics ,* 18, pp.427-443.

Mankiw, N. and M. Shapiro (1986), "News or Noise: An Analysis of GNP Revisions", *Survey of Current Business*, 66(5), pp. 20-25.

Orphanides, A. (2001), "Monetary Rules Based on Real-Time Data", *American Economic Review.*

Robertson, J. and E. Tallman (1999), "Vector Autoregressions: Forecasting and Reality", *Federal Reserve Bank of Atlanta Economic Review,* QI.

Seskin, E. P. and R. P. Parker (1998), "A Guide to the NIPA's", *Survey of Current Business*, 78(3), March, pp. 26-68.

Stock, J. and M. Watson (2002) "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business and Economic Statistics*, 20(2), pp. 147-162,

Urich, T. J., and P. Watchel (1984), "The Effects of Inflation and Money Supple Announcements on Interest Rates", *Journal of Finance,* 39(4), September, pp. 1177-88.