

Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 901

August 2007

## Real-Time Measurement of Business Conditions

S. Borağan Aruoba  
Francis X. Diebold  
Chiara Scotti

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to International Finance Discussion Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors. Recent IFDPs are available on the Web at [www.federalreserve.gov/pubs/ifdp/](http://www.federalreserve.gov/pubs/ifdp/). This paper can be downloaded without charge from Social Science Research Network electronic library at <http://www.ssrn.com/>.

# Real-Time Measurement of Business Conditions\*

S. Borağan Aruoba<sup>†</sup>

Francis X. Diebold

Chiara Scotti

University of Maryland

University of Pennsylvania and NBER

Federal Reserve Board

July 24, 2007

## Abstract

We construct a framework for measuring economic activity in real time (e.g., minute-by-minute), using a variety of stock and flow data observed at mixed frequencies. Specifically, we propose a dynamic factor model that permits exact filtering, and we explore the efficacy of our methods both in a simulation study and in a detailed empirical example.

*Key Words: Business cycle, Expansion, Recession, State space model, Macroeconomic forecasting, Dynamic factor model*

*JEL Codes: E32, E37, C01, C22*

---

\*For helpful discussion we thank seminar and conference participants at the Board of Governors of the Federal Reserve System, the Federal Reserve Bank of Philadelphia, SCE Cyprus, and American University. We are especially grateful to Carlos Capistran, Martin Evans, Jon Faust, Eric Ghysels, Sharon Kozicki, Alexi Onatski, Frank Schorfheide and Jonathan Wright. We thank the National Science Foundation for research support. The usual disclaimer applies. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

<sup>†</sup>Corresponding author. Department of Economics, University of Maryland, College Park, MD 20742. aruoba@econ.umd.edu

# 1 Introduction

Aggregate business conditions are of central importance in the business, finance, and policy communities, worldwide, and huge resources are devoted to assessment of the continuously-evolving state of the real economy. Literally thousands of newspapers, newsletters, television shows, and blogs, not to mention armies of employees in manufacturing and service industries, including the financial services industries, central banks, government and non-government organizations, grapple daily with the real-time measurement and forecasting of evolving business conditions.

Against this background, we propose and illustrate a framework for real-time business conditions assessment in a systematic, replicable, and statistically optimal manner. Our framework has four key ingredients.

*Ingredient 1.* We work with a dynamic factor model, treating business conditions as an unobserved variable, related to observed indicators. The appeal of latency of business conditions comes from its close coherence with economic theory, which emphasizes that the business cycle is not about any single variable, whether GDP, industrial production, sales, employment, or anything else. Rather, the business cycle is about the dynamics and interactions (“co-movements”) of many variables, as forcefully argued by Lucas (1977) among many others.

Treating business conditions as latent is also a venerable tradition in empirical business cycle analysis, ranging from the earliest work to the most recent, and from the statistically informal to the statistically formal. On the informal side, latency of business conditions is central to many approaches, from the classic early work of Burns and Mitchell (1946) to the recent workings of the NBER business cycle dating committee, as described for example by Hall *et al.* (2003). On the formal side, latency of business conditions is central to the popular dynamic factor framework, whether from the “small data” perspective of Geweke (1977), Sargent and Sims (1977), Stock and Watson (1989, 1991), and Diebold and Rudebusch (1996), or the more recent “large data” perspective of Stock and Watson (2002) and Forni, Hallin, Lippi and Reichlin (2000).<sup>1</sup>

*Ingredient 2.* We explicitly incorporate business conditions indicators measured at different frequencies. Important business conditions indicators do in fact arrive at a variety of frequencies, including quarterly (e.g., GDP), monthly (e.g., industrial production), weekly (e.g., employment), and continuously (e.g., asset prices), and we want to be able to incorpo-

---

<sup>1</sup>For definition and discussion of small-data vs. large-data dynamic factor modeling, see Diebold (2003).

rate all of them, to provide continuously-updated assessments in real time.

*Ingredient 3. We explicitly incorporate a continuously-evolving indicator.* Given that our goal is to track the evolution of real activity in real time, it is crucial to incorporate (or at least not exclude from the outset) the real-time information flow associated with continuously-evolving indicators, such as the yield curve. For practical purposes, in this paper we equate “continuously-evolving” with “daily,” but intra-day information could be used as well.

*Ingredient 4. We extract and forecast latent business conditions using linear yet statistically optimal procedures, which involve no approximations.* The appeal of exact as opposed to approximate procedures is obvious, but achieving exact optimality is not trivial and has proved elusive in the literature, due to complications arising from temporal aggregation of stocks vs. flows in systems with mixed-frequency data.

Related to our concerns and framework is a small but nevertheless significant literature, including Stock and Watson (1989, 1991), Mariano and Murasawa (2003), Evans (2005) and Proietti and Moauro (2006). Our contribution, however, differs from the others, as follows.<sup>2</sup>

Stock and Watson (1989, 1991) work in a dynamic factor framework with exact linear filtering, but they don’t consider data at different frequencies or at high frequencies. We include data at different and high frequencies, while still achieving exact linear filtering. This turns out to be a non-trivial task, requiring an original modeling approach.

Mariano and Murasawa (2003) work in a dynamic factor framework and consider data at different frequencies, but not high frequencies, and their filtering algorithm is not exact. In particular, they invoke an approximation essentially equivalent to assuming that the log of a sum equals the sum of the logs.

Evans (2005) does not use a dynamic factor framework and does not use high-frequency data. Instead, he equates business conditions with GDP growth, and he uses state space methods to estimate daily GDP growth using data on preliminary, advanced and final releases of GDP, as well as a variety of other macroeconomic variables.

Proietti and Moauro (2006) work in the Mariano-Murasawa framework and are able to avoid the Mariano-Murasawa approximation, but only at the cost of moving to a non-linear model, resulting in a filtering scheme that is more tedious than the Kalman filter and that involves approximations of its own.

We proceed as follows. In section 2 we provide a detailed statement of our methodological framework, covering the state space formulation with missing data, optimal filtering and

---

<sup>2</sup>Other related and noteworthy contributions include Shen (1996), Abeyasinghe (2000), Altissimo *et al.* (2002), Liu and Hall (2001), McGuckin, Ozyildirim and Zarnowitz (2003), and Ghysels, Santa Clara and Valkanov (2004).

smoothing, and estimation. In section 3 we report the results of a small simulation exercise, which lets us illustrate our methods and assess their efficacy in a controlled environment. In section 4 we report the results of a four-indicator empirical analysis, using quarterly GDP, monthly employment, weekly initial claims, and the daily yield curve term premium. In section 5 we conclude and offer directions for future research.

## 2 Methodology

Here we propose a state space macroeconomic model with an ultra-high base observational frequency, treating specification, estimation, state extraction and state prediction. Our framework facilitates exactly optimal filtering and forecasting, which we achieve throughout.

### 2.1 Missing Observations and Temporal Aggregation

We assume that the state of the economy evolves at a very high frequency; without loss of generality, call it “daily.”<sup>3</sup> Similarly, we assume that all economic and financial variables evolve daily, although many are not *observed* daily. For example, an end-of-year wealth variable is observed each December 31, and is “missing” for every other day of the year.

Let  $y_t^i$  denote a daily economic or financial variable, and let  $\tilde{y}_t^i$  denote the same variable observed at a lower frequency (without loss of generality, call it the “tilde” frequency). The relationship between  $\tilde{y}_t^i$  and  $y_t^i$  depends on whether  $y_t^i$  is a stock or flow variable. In the case of a stock variable, which by definition is a point-in-time snapshot, we have:

$$\tilde{y}_t^i = \begin{cases} y_t^i & \text{if } y_t^i \text{ is observed} \\ NA & \text{otherwise,} \end{cases}$$

where  $NA$  denotes missing data. In the case of a flow variable, the lower-frequency observations of which are functions of current and past daily observations, we have

$$\tilde{y}_t^i = \begin{cases} f(y_t^i, y_{t-1}^i, \dots, y_{t-D_i}^i) & \text{if } y_t^i \text{ is observed} \\ NA & \text{otherwise,} \end{cases}$$

where  $D_i$  denotes the relevant number of days for the temporal aggregation. For ease of exposition we assume for now that  $D_i$  is fixed, but in our subsequent implementation and

---

<sup>3</sup>In our subsequent empirical work, we will indeed use a daily base observational frequency, but much higher (intra-day) frequencies could be used if desired.

empirical work we allow for time-varying  $D_i$ , which allows us to accommodate, for example, the fact that some months have 28 days, some have 29, some have 30, and some have 31.

Satisfactory treatment of temporal aggregation remains elusive in the literature. Most authors work in logarithms and are effectively forced into the unappealing “approximation” that the log of a sum equals the sum of the logs. Mariano and Murasawa (2003), for example, assume that quarterly GDP is the geometric average of the intra-quarter monthly GDPs.

In contrast, our framework permits exact aggregation. We work in levels, so that flow variables aggregate linearly and exactly. Specifically, we model the levels of all observed variables as stationary deviations from polynomial trends of arbitrary order. The result is a linear state space system for which the standard Kalman filter is optimal, as we now describe in detail.

## 2.2 The Model

We assume that underlying business conditions  $x_t$  evolve daily with  $AR(p)$  dynamics,

$$x_t = \rho_1 x_{t-1} + \dots + \rho_p x_{t-p} + v_t, \quad (1)$$

where  $v_t$  is a white noise innovation with unit variance.<sup>4</sup> We are interested in tracking and forecasting real activity, so we use a single-factor model; that is,  $x_t$  is a scalar, as for example in Stock and Watson (1989). Additional factors could of course be introduced to track, for example, nominal wage/price developments.

We assume that all economic variables  $y_t^i$  evolve daily, although they are not necessarily observed daily. We assume that  $y_t^i$  depends linearly on  $x_t$  and possibly also various exogenous variables and/or lags of  $y_t^i$ :

$$y_t^i = c_i + \beta_i x_t + \delta_{i1} w_t^1 + \dots + \delta_{ik} w_t^k + \gamma_{i1} y_{t-D_i}^i + \dots + \gamma_{in} y_{t-nD_i}^i + u_t^i, \quad (2)$$

where the  $w_t$  are exogenous variables, we include  $n$  lags of the dependent variable, and the  $u_t^i$  are contemporaneously and serially uncorrelated innovations. Notice that we introduce lags of the dependent variable  $y_t^i$  in multiples of  $D_i$ , because the persistence in  $y_t^i$  is actually linked to the lower (tilde) observational frequency of  $\tilde{y}_t^i$ . Persistence modeled only in the higher daily frequency would be inadequate, as it would decay too quickly. We use (2) as the measurement equation for all stock variables.

---

<sup>4</sup>As is well-known, identification of factor models requires normalization either on a factor loading or on the factor variance, and we choose to normalize the factor variance to unity.

Temporal aggregation in our framework is very simple: flow variables observed at a tilde frequency lower than daily are the sums of the corresponding daily variables,

$$\tilde{y}_t^i = \begin{cases} \sum_{j=0}^{D_i-1} y_{t-j}^i & \text{if } y_t^i \text{ is observed} \\ NA & \text{otherwise.} \end{cases}$$

The relationship between an observed flow variable and the factor then follows from (2),

$$\tilde{y}_t^i = \begin{cases} \sum_{j=0}^{D_i-1} c_i + \beta_i \sum_{j=0}^{D_i-1} x_{t-j}^i + \delta_{i1} \sum_{j=0}^{D_i-1} w_{t-j}^1 + \dots + \delta_{ik} \sum_{j=0}^{D_i-1} w_{t-j}^k \\ \quad + \gamma_{i1} \sum_{j=0}^{D_i-1} y_{t-D_i-j}^i + \dots + \gamma_{in} \sum_{j=0}^{D_i-1} y_{t-nD_i-j}^i + u_t^{*i} & \text{if } y_t^i \text{ is observed} \\ NA & \text{otherwise,} \end{cases} \quad (3)$$

where  $\sum_{j=0}^{D_i-1} y_{t-D_i-j}^i$  is by definition the observed flow variable one period ago ( $\tilde{y}_{t-D_i}^i$ ), and  $u_t^{*i}$  is the sum of the  $u_t^i$  over the tilde period. Note that although  $u_t^{*i}$  follows a serially correlated moving average process of order  $D_i - 1$  at the daily frequency, it nevertheless remains white noise when observed at the tilde frequency, due to the cutoff in the autocorrelation function of an  $MA(D_i - 1)$  process at displacement  $D_i - 1$ . Hence we will appropriately treat  $u_t^{*i}$  as white noise in what follows, and we have  $var(u_t^{*i}) = Dvar(u_t^i)$ .

The exogenous variables  $w_t$  are the key to handling trend. In particular, in the important special case where the  $w_t$  are simply deterministic polynomial trend terms ( $w_{t-j}^1 = t - j$ ,  $w_{t-j}^2 = (t - j)^2$  and so on) we have that

$$\sum_{j=0}^{D_i-1} \left[ c_i + \delta_{i1} (t - j) + \dots + \delta_{ik} (t - j)^k \right] \equiv c_i^* + \delta_{i1}^* t + \dots + \delta_{ik}^* t^k, \quad (4)$$

which yields

$$\tilde{y}_t^i = \begin{cases} c_i^* + \beta_i \sum_{j=0}^{D_i-1} x_{t-j}^i + \delta_{i1}^* t + \dots + \delta_{ik}^* t^k + \gamma_{i1} \tilde{y}_{t-D_i}^i + \dots + \gamma_{in} \tilde{y}_{t-nD_i}^i + u_t^{*i} & \text{if } y_t^i \text{ is observed} \\ NA & \text{otherwise.} \end{cases} \quad (5)$$

We use (5) as the measurement equation for all flow variables. In the appendix we derive the mapping between  $(c, \delta_1, \delta_2, \delta_3)$  and  $(c^*, \delta_1^*, \delta_2^*, \delta_3^*)$  for cubic trends, which we use throughout this paper.<sup>5</sup>

This completes the specification of our model, which has a natural state space form, to which we now turn.

## 2.3 State Space Representation, Filtering and Smoothing

Assembling the discussion thus far, the state space representation of our model is

$$\begin{aligned} y_t &= Z_t \alpha_t + \Gamma_t w_t + \varepsilon_t \\ \alpha_{t+1} &= T \alpha_t + R \eta_t \\ \varepsilon_t &\sim (0, H_t), \eta_t \sim (0, Q), \end{aligned} \tag{6}$$

$t = 1, \dots, \mathcal{T}$ , where  $\mathcal{T}$  denotes the last time-series observation,  $y_t$  is an  $N \times 1$  vector of observed variables,  $\alpha_t$  is an  $m \times 1$  vector of state variables,  $w_t$  is a  $e \times 1$  vector of exogenous variables, and  $\varepsilon_t$  and  $\eta_t$  are vectors of measurement and transition shocks which will collectively contain  $v_t$  and  $u_t^i$ . The vector  $w_t$  includes an entry of unity for the constant,  $k$  trend terms and  $N \times n$  lagged dependent variables,  $n$  for each of the  $N$  elements of the  $y_t$  vector. The exact structure of these vectors will vary across the different setups we consider below. The observed data vector  $y_t$  will have many missing values, reflecting those variables observed at a frequency lower than daily, as well as missing daily data due to holidays. At a minimum, the state vector  $\alpha_t$  will include  $p$  lags of  $x_t$ , as implied by (1). Moreover, because the presence of flow variables requires a state vector containing all lags of  $x_t$  inside the aggregation period, in practice the dimension of  $\alpha_t$  will be much greater than  $p$ . The system parameter matrices  $T, R$  and  $Q$ , are constant, while  $Z_t, \Gamma_t$  and  $H_t$  are not, because of the variation in the number of days in a quarter or month ( $D_i$  for each  $i$ ).<sup>6</sup>

With the model cast in state space form, we can immediately apply the Kalman filter and smoother. We first present the algorithm assuming no missing data values, and then we incorporate missing data. For given parameters, we initialize the Kalman filter using

---

<sup>5</sup>For numerical stability we use  $t/1000, (t/1000)^2$  and  $(t/1000)^3$  instead of simply  $t, t^2$  and  $t^3$  in specifying our cubic trend polynomial. This is simply a normalization and does not affect the other parameters of interest or the log-likelihood. We impose it because in our subsequent empirical work we have over 16,000 daily observations, in which case  $t^3$  can be very large, which might create numerical problems.

<sup>6</sup>Time-varying system matrices pose no problem for the Kalman filter.



$\alpha_1 \sim N(a_1, P_1)$  where  $a_1 = 0_{m \times 1}$  and  $P_1$  solves

$$(I - T \otimes T) \text{vec}(P_1) = \text{vec}(RQR'). \quad (7)$$

Given  $a_1$  and  $P_1$ , for  $t = 1, \dots, \mathcal{T}$ , we use the contemporaneous Kalman filtering equations, which incorporate the computation of the state vector estimate and its associated covariance matrix, denoted by  $a_{t|t}$  and  $P_{t|t}$ .<sup>7</sup> Denote  $\{y_1, \dots, y_t\}$  by  $\mathcal{Y}_t$  for  $t = 1, \dots, \mathcal{T}$ . Then, given  $a_t \equiv E(\alpha_t | \mathcal{Y}_{t-1})$  and  $P_t = \text{var}(\alpha_t | \mathcal{Y}_{t-1})$ , the prediction equations that produce  $a_{t+1}$  and  $P_{t+1}$  are

$$a_{t|t} = a_t + P_t Z_t' F_t^{-1} v_t \quad (8)$$

$$P_{t|t} = P_t - P_t Z_t' F_t^{-1} Z_t P_t' \quad (9)$$

$$a_{t+1} = T a_{t|t} \quad (10)$$

$$P_{t+1} = T P_{t|t} T' + RQR', \quad (11)$$

where

$$v_t = y_t - Z_t a_t - \Gamma_t w_t \quad (12)$$

$$F_t = Z_t P_t Z_t' + H_t. \quad (13)$$

The Kalman smoother computes the conditional expectation of the state vector and its covariance matrix using all the information in the data set, which we denote by  $\hat{\alpha}_t \equiv E(\alpha_t | \mathcal{Y}_{\mathcal{T}})$  and  $V_t \equiv \text{var}(\alpha_t | \mathcal{Y}_{\mathcal{T}})$  for  $t = 1, \dots, \mathcal{T}$ . The Kalman smoother recursions start from  $t = \mathcal{T}$  and work backward. The vector  $r_t$  is a weighted average of the innovations  $v_t$  that happen after period  $t$  with the variance matrix  $N_t$ . We initialize the smoother with

---

<sup>7</sup>We find that using this version of the filter improves the efficiency of the algorithm. See Durbin and Koopman (2001) for details.

$r_{\mathcal{T}} = 0_{m \times 1}$  and  $N_{\mathcal{T}} = 0_{m \times m}$ , and for  $t = 1, \dots, \mathcal{T}-1$  we use

$$K_t = TPZ'_tF_t^{-1} \quad (14)$$

$$L_t = T - K_tZ_t \quad (15)$$

$$r_{t-1} = Z'_tF_t^{-1}v_t + L'_tr_t \quad (16)$$

$$N_{t-1} = Z'_tF_t^{-1}Z_t + L'_tN_tL_t \quad (17)$$

$$\hat{\alpha}_t = a_t + P_tr_{t-1} \quad (18)$$

$$V_t = P_t - P_tN_{t-1}P_t, \quad (19)$$

where we store the matrices  $\{F_t, v_t, a_t, P_t\}_{t=1}^{\mathcal{T}}$  from one run of the Kalman filter. We use the appropriate element of the  $\hat{\alpha}_t$  vector as the extracted factor and the corresponding diagonal element of  $V_t$  as its standard error to compute confidence bands.

We now describe how we handle missing observations (that is, how the Kalman filter optimally handles missing observations). If all elements of the vector  $y_t$  are missing for period  $t$ , we skip updating and the recursion becomes

$$a_{t+1} = Ta_t \quad (20)$$

$$P_{t+1} = TP_tT' + RQR. \quad (21)$$

If some (but not all) elements of  $y_t$  are missing, we replace the observation equation with

$$y_t^* = Z_t^*\alpha_t + \Gamma_t w_t + \varepsilon_t^* \quad (22)$$

$$\varepsilon_t^* \sim N(0, H_t^*), \quad (23)$$

where  $y_t^*$  are the elements of the  $y_t$  vector that are observed.<sup>8</sup> The two vectors are linked by  $y_t^* = W_t y_t$ , where  $W_t$  is a matrix that carries the appropriate rows or  $I_{N \times N}$ ,  $Z_t^* = W_t Z_t$ ,  $\varepsilon_t^* = W_t \varepsilon_t$  and  $H_t^* = W_t H_t W_t'$ . The Kalman filter and smoother work exactly as described above replacing  $y_t$ ,  $Z_t$  and  $H$  with  $y_t^*$ ,  $Z_t^*$  and  $H_t^*$  for period  $t$ .

## 2.4 A Costless Generalization

Note that in the model as specified thus far, current observables depend only on the current state vector. More generally, we might want to allow current observables to depend on both

---

<sup>8</sup>By construction, whenever there is an observation for a particular element of  $y_t$ , there is a corresponding element of  $w_t$ .

the present state vector and various of its leads or lags. This turns out to introduce significant complications in situations such as ours that involve temporal aggregation. Hence we will not pursue it, with one notable exception: For daily variables we can easily accommodate dynamic interaction between observables and the state because there is no temporal aggregation.<sup>a</sup>

In our subsequent empirical work, for example, we allow daily observables to depend on a distributed lag of the state. To promote parsimony, we use a polynomial distributed lag (PDL) specification. Specifically, the measurement equation for a daily variable is

$$y_t^i = c_i + \beta_i^0 x_t + \beta_i^1 x_{t-1} + \dots + \beta_i^{\bar{D}} x_{t-\bar{D}} + \delta_{i1} w_t^1 + \dots + \delta_{ik} w_t^k + \gamma_{i1} y_{t-1}^i + \dots + \gamma_{in} y_{t-n}^i + \varepsilon_t^i, \quad (24)$$

where the elements of  $\{\beta_i^j\}_{j=0}^{\bar{D}}$  follow a low-ordered polynomial given by

$$\beta_i^j = \xi_i^1 + \xi_i^2(j - \bar{c}) + \xi_i^3(j - \bar{c})^2 + \dots + \xi_i^{\bar{D}+1}(j - \bar{c})^{\bar{D}}, \quad (25)$$

where  $\bar{c}$  is equal to  $k/2$  if  $k$  is even and to  $(k - 1)/2$  if  $k$  is odd.<sup>9</sup> We shall use a third-order polynomial.

## 2.5 Estimation

As is well-known, the Kalman filter supplies all of the ingredients needed for evaluating the Gaussian pseudo log likelihood function via the prediction error decomposition,

$$\log L = -\frac{1}{2} \sum_{t=1}^T [N \log 2\pi + (\log |F_t| + v_t' F_t^{-1} v_t)] \quad (26)$$

In calculating the log likelihood, if all elements of  $y_t$  are missing, the contribution of period  $t$  to the likelihood is zero. When some elements of  $y_t$  are observed, the contribution of period  $t$  is  $[N^* \log 2\pi + (\log |F_t^*| + v_t^{*'} F_t^{*-1} v_t^*)]$  where  $N^*$  is the number of observed variables and the other matrices and vectors are obtained using the Kalman filter recursions on the modified system with  $y_t^*$ .

Armed with the ability to evaluate the log likelihood for any given set of parameters, we proceed with estimation using standard methods. In particular, we use a quasi-Newton optimization routine with BFGS updating of the inverse Hessian.

---

<sup>9</sup>Because we assume that daily frequency is the highest available, we can treat flow and stock variables identically when they are observed daily.

We impose several constraints in our estimation. First, to impose stationarity of  $x_t$  we use a result of Barndorff-Nielsen and Schou (1973), who show that under stationarity there is a one-to-one correspondence between the parameters of an  $AR(p)$  process and the first  $p$  partial autocorrelations. Hence we can parameterize the likelihood in terms of the relevant partial autocorrelations, which requires searching only over the unit interval. In our subsequent empirical analysis, we use an  $AR(3)$  process for the factor, which allows for a rich variety of dynamics. Denoting the  $AR(3)$  parameters by  $\rho_i$  and the partial autocorrelations by  $\pi_i$ , the Barndorff-Nielsen-Schou mapping between the two is

$$\rho_1 = \pi_1 - \pi_1\pi_2 - \pi_3\pi_2 \quad (27)$$

$$\rho_2 = \pi_2 - \pi_1\pi_3 + \pi_1\pi_2\pi_3 \quad (28)$$

$$\rho_3 = \pi_3. \quad (29)$$

We then optimize over  $\pi_i \in [-1, 1]$ .<sup>10</sup>

Second, we impose non-negativity of the variance terms in the diagonal elements of  $Q$  and  $H_t$  matrices by estimating natural logarithms of these elements. Similarly, we restrict the factor loadings on some of the variables to have a certain sign (e.g., positive for GDP and negative for initial jobless claims) using the same transformation.

Searching for a global optimum in a parameter space with more than 30 dimensions is a challenging problem. It is not intractable, however, if the iterations are initialized cleverly. To do so, we exploit knowledge gained from a simulation study, to which we now turn.

### 3 A Simulation-Based Example

Here we illustrate our methods in a simulation. This allows us to assess their efficacy in a controlled environment, and to gain insights of relevance to our subsequent fitting of the model to real data. We work with an  $AR(1)$  real activity factor and three observed indicators, which are driven by the factor and linear trend.<sup>11</sup> We generate forty years of daily data, and then we transform them to obtain the data observed by the econometrician. Specifically,  $y_t^1$  is a daily financial variable so we eliminate weekend observations,  $y_t^2$  is a monthly stock variable so we eliminate all observations except the last observation of each month, and  $y_t^3$  is a quarterly flow variable so we eliminate all observations except the last observation of

---

<sup>10</sup>We use a hyperbolic tangent function to search over  $\pi_i$ , because for  $y \in R$ ,  $x = \tanh(y) \in [-1, 1]$ .

<sup>11</sup>For simplicity in the simulation, we do not use higher order trend terms, lagged dependent variables, or polynomial distributed lags.

the quarter, which we set equal to the sum of the intra-quarter daily observations. After obtaining the observed data we estimate the model given in (6), with the system vectors and matrices are defined as:

$$\begin{aligned}
y_t &= \begin{bmatrix} \tilde{y}_t^1 \\ \tilde{y}_t^2 \\ \tilde{y}_t^3 \end{bmatrix}, \quad \alpha_t = \begin{bmatrix} x_t \\ x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-\bar{q}+1} \\ x_{t-\bar{q}} \end{bmatrix}, \quad w_t = \begin{bmatrix} 1 \\ t \end{bmatrix}, \quad \varepsilon_t = \begin{bmatrix} u_t^1 \\ u_t^2 \\ u_t^{*3} \end{bmatrix}, \quad v_t = \eta_t, \quad R = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\
Z_t &= \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 \\ 0 & 0 & \beta_3 \\ \vdots & \vdots & \vdots \\ 0 & 0 & \beta_3 \text{ or } 0 \\ 0 & 0 & \beta_3 \text{ or } 0 \\ 0 & 0 & \beta_3 \text{ or } 0 \end{bmatrix}', \quad \Gamma_t = \begin{bmatrix} c_1 & \delta_1 \\ c_2 & \delta_2 \\ c_{3t}^* & \delta_{3t}^* \end{bmatrix}, \quad T = \begin{bmatrix} \rho_1 & \rho_2 & \rho_3 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \quad (30) \\
\begin{bmatrix} \varepsilon_t \\ v_t \end{bmatrix} &\sim N \left( \begin{bmatrix} 0_{3 \times 1} \\ 0 \end{bmatrix}, \begin{bmatrix} H_t & 0 \\ 0 & Q \end{bmatrix} \right), \quad H_t = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_{3t}^{*2} \end{bmatrix}, \quad Q = 1
\end{aligned}$$

where  $\bar{q}$  is the maximum number of days in a quarter. For convenience, in our notation we treat  $q$ , the number of days in a quarter (the counterpart of  $D_i$  in the previous section), as fixed. In our implementation, however, we make the necessary adjustments to account for the exact number of days in a quarter, which is either 90, 91 or 92, depending on the quarter and whether or not the year is a leap year. All of the relevant matrices and vectors allow for the largest possible value,  $\bar{q}$ , and we adjust the matrices  $Z_t$ ,  $\Gamma_t$  and  $H_t$  every quarter as follows.<sup>12</sup> Each quarter, if  $q < \bar{q}$ , we first set the first  $q$  elements of the third row of  $Z_t$  to  $\beta_3$  and we set the remaining elements to zero. Next, we use  $D = q$  in the formulas derived in the appendix that map our original parameters  $c_3$  and  $\delta_3$  into  $c_{3t}^*$  and  $\delta_{3t}^*$ , and we substitute in  $\Gamma_t$ . Finally, we set the third diagonal element of  $H_t$  to  $\sigma_{3t}^{*2} = q\sigma_3^2$ . All of the adjustments follow from the discussion in the previous section. First, the quarterly flow variable requires

<sup>12</sup>Note that the third rows of  $Z_t$ ,  $\Gamma_t$  and  $H_t$  are only relevant when  $\tilde{y}_t^3$  is observed. For all other days, the contents of the third rows of these matrices do not affect any calculations. When there is an observation for  $\tilde{y}_t^3$  we look at the number of days in that particular quarter,  $q$ , and make the adjustments.

summing the factors over the quarter, and our adjustment of the third row of  $Z_t$  ensures that we sum only the relevant factors. Second, the adjustment of the elements of  $\Gamma_t$  is obvious. Finally, because  $\varepsilon_t^{*3}$  is the sum of  $q$  iid normal innovations each with variance  $\sigma_3^2$ , its variance is  $q\sigma_3^2$ .

We use a multi-step estimation procedure that helps us obtain accurate startup values for certain of the model parameters, after which we estimate all parameters simultaneously. This approach is of course most helpful in higher-dimensional systems than the simple one at hand, but we illustrate it here because we use it in our subsequent (higher-dimensional) empirical work.

Specifically, we first use only the first two variables, estimating the model using naive startup values and experimenting with them to make sure we reach the global maxima. Because we exclude the quarterly flow variable from this model, the system estimated is small and the cost of experimentation is low. Once the model is estimated, we use the Kalman smoother to extract the factor,  $\hat{x}_t$ . Then we obtain startup values for the third equation via least-squares estimation of the auxiliary model,

$$\tilde{y}_t^3 = \sum_{j=0}^{q-1} [a + d(t-j)] + b(\hat{x}_t + \hat{x}_{t-1} + \dots + \hat{x}_{t-q}) + e_t. \quad (31)$$

We then estimate the full model using the estimates of  $a$ ,  $b$ ,  $d$  and  $var(e_t)/\bar{q}$  obtained from this auxiliary regression as startup values for the third equation's parameters, and the originally-estimated parameters of the first and second equations as startup values for those equations' parameters.

To illustrate the performance of our methodology we compare the true and smoothed factor and indicators. To obtain the "smoothed indicators," we run the smoothed factor through equation (2). In Figure 1 we plot the true and smoothed versions of the factor, and the high-frequency (true), observed and smoothed versions of two indicators (the daily financial variable and the monthly stock variable) over a 6-month period in the sample. The first panel shows the very close relationship between the smoothed and the true factor. In fact, over the full sample the correlation between the two is greater than 0.96. In the second panel, the observed and high-frequency (true) indicators are identical except for weekends, and the smoothed signal tries to fill in the missing values in the observed indicator by using the information from other variables. Finally, in the third panel, the observed indicator is represented by dots, which are the end-of-month-values of the true signal. Our smoothed signal tries to fill in the remaining values and performs quite well. Over the full sample,

the correlations between the smoothed and true indicators for both  $Y1$  and  $Y2$  are 0.997. Overall, this example shows that our methodology is well-suited to extract the factor in an environment with missing data and/or time aggregation issues.

## 4 Empirical Analysis

Now we apply our framework to real data. First, we describe the data, and then we discuss our empirical results

### 4.1 Data

Our analysis covers the period from April 1, 1962 through February 20, 2007, which is over 45 years of daily data. Because it is not realistic to assume that economic activity stops over the weekends is not realistic, we use a seven-day week instead of using only business days. We use four variables in our analysis. Below we list these variables and describe how we handle missing data and time-aggregation issues.<sup>13</sup>

- *Yield curve term premium* defined as the difference between the yield of the ten-year and the three-month Treasury yields. This is a daily variable with missing values for weekends and holidays and no time-aggregation issues.
- *Initial claims for unemployment insurance*. This is a weekly flow variable covering the seven-day period from Sunday through Saturday. The value for Saturdays is the sum of the daily values for the previous seven days and other days have missing values.
- *Employees on nonagricultural payrolls*. This is a monthly stock variable, observed on the last day of the month, with missing values for other days.
- *Real GDP*. This is a quarterly flow variable. The value for the last day of the quarter is the sum of the daily values for all the days in the quarter, other days have missing values.

### 4.2 Model

The state variable  $x_t$  follows an  $AR(3)$  process and we also assume  $AR(3)$  structures for the observed variables at their observation frequency. For weekly initial claims, monthly

---

<sup>13</sup>For numerical stability we adjust the units of some of our observed variables. (e.g. we divide employment by 10,000 and initial jobless claims by 1,000)

employment and quarterly GDP, this simply means that the lagged values of these variables are elements of the  $w_t$  vector. We denote these by  $\tilde{y}_{t-iW}^2$ ,  $\tilde{y}_{t-iM}^3$  and  $\tilde{y}_{t-iq}^4$  for  $i = 1, 2, 3$ , where  $W$  denotes the number of days in a week,  $M$  denotes the number of days in a month and  $q$  denotes the number of days in a quarter.<sup>14</sup> For the term premium, on the other hand, we choose to model the autocorrelation structure using an  $AR(3)$  process for the measurement equation innovation,  $u_t^1$ , instead of adding three lags of the term premium in  $w_t$ . We choose to follow this route because of the missing term premium observations due to non-business days. If we used the lagged term premium as an element of  $w_t$  this would yield only two valid observations for each week and it would make the analysis less reliable.<sup>15</sup>

The matrices that define the model are given by

$$y_t = \begin{bmatrix} \tilde{y}_t^1 \\ \tilde{y}_t^2 \\ \tilde{y}_t^3 \\ \tilde{y}_t^4 \end{bmatrix}, \quad \alpha_t = \begin{bmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-\bar{q}-1} \\ x_{t-\bar{q}} \\ u_t^1 \\ u_{t-1}^1 \\ u_{t-2}^1 \end{bmatrix}, \quad w_t = \begin{bmatrix} 1 \\ t \\ t^2 \\ t^3 \\ \tilde{y}_{t-W}^2 \\ \tilde{y}_{t-2W}^2 \\ \tilde{y}_{t-3W}^2 \\ \tilde{y}_{t-M}^3 \\ \tilde{y}_{t-2M}^3 \\ \tilde{y}_{t-3M}^3 \\ \tilde{y}_{t-q}^4 \\ \tilde{y}_{t-2q}^4 \\ \tilde{y}_{t-3q}^4 \end{bmatrix}, \quad \varepsilon_t = \begin{bmatrix} 0 \\ u_t^{*2} \\ u_t^3 \\ u_t^{*4} \end{bmatrix}, \quad v_t = \begin{bmatrix} \eta_t \\ \zeta t \end{bmatrix}, \quad R = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

---

<sup>14</sup>Once again, the notation in the paper assumes  $M$  and  $q$  are constant over time but in the implementation we adjust them according to the number of days in the relevant month or quarter. The number of days in a week is always seven.

<sup>15</sup>Alternatively we could have used  $AR(3)$  measurement errors for all variables. But this persistence in the daily frequency would essentially disappear when we aggregate the variables to the monthly or quarterly frequency.



$$Z_t = \begin{bmatrix} \beta_1^0 & \beta_2 & \beta_3 & \beta_4 \\ \beta_1^1 & \beta_2 & 0 & \beta_4 \\ \vdots & \vdots & \vdots & \vdots \\ \beta_1^6 & \beta_2 & 0 & \beta_4 \\ \beta_1^7 & 0 & 0 & \beta_4 \\ \vdots & \vdots & \vdots & \vdots \\ \beta_1^{\bar{q}-2} & 0 & 0 & \beta_4 \text{ or } 0 \\ \beta_1^{\bar{q}-1} & 0 & 0 & \beta_4 \text{ or } 0 \\ \beta_1^{\bar{q}} & 0 & 0 & \beta_4 \text{ or } 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}', \quad \Gamma_t = \begin{bmatrix} c_1 & c_{2t}^* & c_3 & c_{4t}^* \\ \delta_{11} & \delta_{21t}^* & \delta_{31} & \delta_{41t}^* \\ \delta_{12} & \delta_{22t}^* & \delta_{32} & \delta_{42t}^* \\ \delta_{13} & \delta_{23t}^* & \delta_{33} & \delta_{43t}^* \\ 0 & \gamma_{21} & 0 & 0 \\ 0 & \gamma_{22} & 0 & 0 \\ 0 & \gamma_{23} & 0 & 0 \\ 0 & 0 & \gamma_{31} & 0 \\ 0 & 0 & \gamma_{32} & 0 \\ 0 & 0 & \gamma_{33} & 0 \\ 0 & 0 & 0 & \gamma_{41} \\ 0 & 0 & 0 & \gamma_{42} \\ 0 & 0 & 0 & \gamma_{43} \end{bmatrix}' \quad (32)$$

$$T = \begin{bmatrix} \rho_1 & \rho_2 & \rho_3 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \gamma_{11} & \gamma_{12} & \gamma_{13} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} \varepsilon_t \\ v_t \end{bmatrix} \sim N \left( \begin{bmatrix} 0_{4 \times 1} \\ 0_{2 \times 1} \end{bmatrix}, \begin{bmatrix} H_t & 0 \\ 0 & Q \end{bmatrix} \right), \quad H_t = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \sigma_{2t}^{*2} & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_{4t}^{*2} \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & 0 \\ 0 & \sigma_1^2 \end{bmatrix}$$

where the matrices and vectors correspond to the system in Section 2.2 and we have  $N = 4$ ,  $k = 13$ ,  $m = 95$ ,  $p = 3$  and  $r = 2$ . We use the current and 91 lags of the factor in our state vector because the maximum of days possible in a quarter is 92, which we denote by  $\bar{q}$ .<sup>16</sup> As we did in the simulation example, we use the transformation given in the appendix

<sup>16</sup>If there are  $q$  days in a quarter, on the last day of the quarter, we need the current and the  $q - 1$  lags

to convert the coefficients with “\*” to those without. Also in every quarter, we adjust the number of non-zero elements in the fourth row of the  $Z_t$  matrix to reflect the number of days in that quarter. When estimating this system, we restrict  $\beta_3$  and  $\beta_4$  to be positive and  $\beta_2$  to be negative to reflect our expectation of the relationship between these variables and the common factor.<sup>17</sup>

## 4.3 Results

### 4.3.1 Estimation

It is worth emphasizing the size of this model. We have 16,397 daily observations, 95 state variables and 42 coefficients. Using a fairly efficiently programmed Kalman filter routine in MATLAB, one evaluation of the log-likelihood takes about 25 seconds. As such, one iteration (including the calculation of the Jacobian) takes a minimum of eighteen minutes. Clearly, it is very costly to look over an “irrelevant” part of the parameter space as it may take the estimation routine many hours or days to find the “right” path, if at all. To tackle this problem, we follow the algorithm outlined earlier: We start by a smaller system, one that has only the term premium and employment. Once we estimate this system we get the smoothed factor and estimate the auxiliary regression for real GDP. Using the estimated values from the smaller system and the auxiliary regression as the starting guesses, we estimate the system with real GDP. We repeat this for initial claims.

### 4.3.2 Factor

First we focus on the factor and its properties. In Figure 2 we plot the smoothed factor from the estimation along with 95% confidence bands, with NBER recessions shaded. Because the NBER provides only months of the turning points, we assume recessions start on the first day of the month and end on the last day of the month. We can make a few important observations. First, the smoothed factor declines sharply around the recession start dates announced by the NBER. Although the beginning of recessions and the decline of the smoothed factor do not always coincide, the factor shows the same sharp decline pattern at the start of each of the six recessions in the sample. Second, recoveries do not all have the same pattern. For the recessions in 1974, 1980 and 1982 the recoveries coincide with as sharp

---

of the factor for the measurement equation of GDP.

<sup>17</sup>In our experience with smaller systems, when we do not impose a sign restriction the estimation may yield a factor which is negatively correlated with GDP. Imposing the sign restriction reverses the correlation with virtually no change in the likelihood.

reversals in the factor as the recessions. For the three remaining recessions, as well as the 1961 recession which ends just before our sample starts, the factor is more sluggish during the recoveries, especially so for the 1990 recession as is well-known. We will turn to this in more detail when we zoom in around turning points below. Finally, there seem to be few, if any, “false positives” where our factor shows patterns similar to recessions in a period which is not a recession. Overall, we conclude that our smoothed factor tracks the U.S. business cycle well.

### 4.3.3 Smoothed Indicators, Smoothed Factor, and Turning Points

One of the most powerful aspects of our framework is its ability to produce high frequency measures of indicators that are observed in much lower frequencies. To demonstrate this, and to investigate how our indicators behave around turning points, we compute the smoothed daily signals for our four indicators. To do this, we use the relationship given by (2) for all non-daily variables and by (24) for term premium. Given the presence of lagged terms, we initialize each variable by the mean implied by the steady state of signal equation. To avoid small-sample problems with this initialization, we drop about 4.5 years’ worth of data and use the daily indicators starting from 01/01/1969.<sup>18</sup>

As a first task, we want to verify that the daily smoothed versions of the indicators are consistent with their lower frequency counterparts. For a stock variable such as employment, the daily observation at the end of the observation period (month in this case) will be comparable with the observed value. On the other hand, for a flow variable such as initial claims, we use the appropriate temporal aggregation (adding over the days of the week in this case). In Figure 3 we plot the actual and smoothed term premium. Perhaps not surprisingly, because the term premium is observed on all business days, there is a very close match between the two and the correlation is over 0.99. In Figure 4 we plot the seven-day sum of the daily smoothed initial claims (dark circles) versus the actual weekly initial claims (light circles) and the correlation here is 0.76. It seems that the smoothed version is, by its nature, not as extreme as the actual series but the match between the two is very high. In the first panel of Figure 5 we plot the actual monthly (circles) and the smoothed daily (solid line) employment. Because the non-stationary nature of the variable can be deceiving, we also plot the detrended versions of the two series, where we detrend using cubic polynomials in trend. The correlations are over 0.99 in both cases. In Figure 6 we plot the actual quarterly

---

<sup>18</sup>Because the number of days in a quarter is not fixed, we assume this number is 91 in (2) for GDP and in the temporal aggregation for Figure 6 below. We also start from 01/01/1975 for GDP because it is less frequently observed and hence more time is needed to remove the effects of initialization.

(light circles) and the smoothed quarterly (dark circles) GDP both as level and as deviation from a cubic trend. The match between the two are weaker compared to the other variables with a correlation of 0.31. This is because GDP is very infrequently observed compared to the other variables with only 128 observations. Overall, we conclude that the smoothed daily indicators that we obtain are very reliable.

Next we zoom in around the turning points of the six business cycles that are in our sample, as determined by the NBER.<sup>19</sup> We consider a window that starts twelve months before the month of the peak and ends twelve months after the month of the trough. First, we look at the smoothed daily factor in the six windows, which we plot in Figure 7, where we superimpose NBER Recession dates. Two things that are of interest is how the turning points of the daily factor compare with the official turning points and the behavior of the factor during recessions. We should keep in mind that neither matching the NBER dates is a success, nor not matching them is a failure, at least because there might be some discrepancy due to the lower frequency (monthly) the NBER uses. Generally speaking our factor is decreasing during most of the recessions but there is quite a bit of disagreement about the turning points, especially the peaks. In all recessions except for the 1973-1975 recession, our factor reaches its peak and starts falling at least five months before the official peak of the expansion. For the 1980 recession, the peak of the factor seems to be beyond our twelve-month window. As for the troughs, in all but the last two recessions our factor has already picked up before the trough of the recession even though the difference is at most three months. For the last two recessions, the turning point of the factor seems to exactly match the official end of the recession. In Figure 8 we plot daily smoothed initial claims in the same six windows. We see that initial claims are either relatively flat (i.e. non-decreasing) or increasing even a year before the official start of a recession in all the six episodes. Moreover, the increase in initial claims that is sustained during all the recessions starts to slow down or get reversed either exactly at the trough or shortly thereafter in all six episodes. In Figure 9 we plot daily smoothed employment in the six windows. The cyclical behavior of employment over the business cycle is very clearly visible in the figures. In the first three recessions, employment reaches its peak after the economy falls in to recession while in the latter three recessions the peak of employment coincides with the peak of the business cycle. As for the troughs, the trough of employment coincides with the trough of the business cycle for the first four episodes while it significantly lags the business cycle for

---

<sup>19</sup>These are December 1969-November 1970, November 1973-March 1975, January 1980-July 1980, July 1981-November 1982, July 1990-March 1991 and March 2001-November 2001.

the latter two episodes. This finding reinforces the “jobless recovery” description commonly attributed to the 1990-1991 recession. Moreover it is an indication that the 2001 recession is similar to the 1990-1991 recession and these two episodes are different from the previous recessions in the US. In fact, comparing the lower two panels of Figures 7-9 with the other panels, this observation is very apparent. Finally, in Figure 10 we plot daily smoothed GDP over the last four recessions in our sample. Except for the 1980 recession, GDP starts to decline right about the same time as the official peak date and continues the decline past the official trough. In fact, the trough of GDP for these three recessions are all beyond our twelve-month window. Despite a slowdown during the 1980 recession, GDP does not start to decline until the peak of the 1981-1982 recession.

## 5 Summary and Concluding Remarks

We have constructed a framework for measuring macroeconomic activity in real time, using a variety of stock and flow data observed at mixed frequencies, including ultra-high frequencies. Specifically, we have proposed a dynamic factor model that permits exactly optimal extraction of the latent state of macroeconomic activity, and we have illustrated it both in simulation environments and in a sequence of progressively richer empirical examples. We also provided some examples of the applications of the framework which yield useful insights for understanding comovements of variables over the business cycle.

We look forward to a variety of variations and extensions of our basic theme, including but not limited to:

(1) Incorporation of indicators beyond macroeconomic and financial data. In particular, it will be of interest to attempt inclusion of qualitative information such as headline news.

(2) Construction of a real time composite leading index (CLI). Thus far we have focused only on construction of a composite *coincident index* (CCI), which is the more fundamental problem, because a CLI is simply a forecast of a CCI. Explicit construction of a leading index will nevertheless be of interest.

(3) Allowance for nonlinear regime-switching dynamics. The linear methods used in this paper provide only a partial (linear) statistical distillation of the rich business cycle literature. A more complete approach would incorporate the insight that expansions and contractions may be probabilistically different regimes, separated by the “turning points” corresponding to peaks and troughs, as emphasized for many decades in the business cycle literature and rigorously embodied Hamilton’s (1989) Markov-switching model. Diebold and Rudebusch

(1996) and Kim and Nelson (1998) show that the linear and nonlinear traditions can be naturally joined via dynamic factor modeling with a regime-switching factor. Such an approach could be productively implemented in the present context, particularly if interest centers on turning points, which are intrinsically well-defined only in regime-switching environments.

(4) Comparative assessment of experiences and results from “small data” approaches, such as ours, vs. “big data” approaches. Although much professional attention has recently turned to big data approaches, as for example in Forni, Hallin, Lippi and Reichlin (2000) and Stock and Watson (2002), recent theoretical work by Boivin and Ng (2006) shows that bigger is not necessarily better. The matter is ultimately empirical, requiring detailed comparative assessment. It would be of great interest, for example, to compare results from our approach to those from the Altissimo *et al.*(2002) EuroCOIN approach, for the same economy and time period. Such comparisons are very difficult, of course, because the “true” state of the economy is never known, even *ex post*.

## References

- [1] Abeyasinghe, T. (2000), "Modeling Variables of Different Frequencies," *International Journal of Forecasting*, 16, 117-119.
- [2] Altissimo, F., Bassanetti, A., Cristadoro, R., Forni, M., Hallin, M., Lippi, M., Reichlin, L. and Veronese, G. (2001), "Eurocoin: A Real Time Coincident Indicator of the Euro Area Business Cycle," CEPR Discussion Paper No. 3108.
- [3] Barndorf-Nielsen, O. and G. Schou (1973), "On the Parametrization of Autoregressive Models by Partial Autocorrelations," *Journal of Multivariate Analysis*, 3, 408-419.
- [4] Boivin, J. and Ng, S. (2006), "Are More Data Always Better for Factor Analysis," *Journal of Econometrics*, 127, 169-194.
- [5] Burns, A.F. and Mitchell, W.C. (1946), *Measuring Business Cycles*, New York, NBER.
- [6] Diebold, F.X. (2003), "Big Data' Dynamic Factor Models for Macroeconomic Measurement and Forecasting" (Discussion of Reichlin and Watson papers), in M. Dewatripont, L.P. Hansen and S. Turnovsky (Eds.), *Advances in Economics and Econometrics, Eighth World Congress of the Econometric Society* . Cambridge: Cambridge University Press, 115 122.
- [7] Diebold, F.X. and Rudebusch, G. (1996), "Measuring Business Cycles: A Modern Perspective," *Review of Economics and Statistics*, 78, 67 77.
- [8] Durbin and Koopman (2001), *Time Series Analysis by State Space Methods*, Oxford University Press.
- [9] Evans, M.D.D. (2005), "Where Are We Now?: Real Time Estimates of the Macro Economy," *The International Journal of Central Banking*, September.
- [10] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000), "The Generalized Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82, 540-554.
- [11] Geweke, J.F. (1977), "The Dynamic Factor Analysis of Economic Timeseries Models," in D. Aigner and A. Goldberger (eds.), *Latent Variables in Socio economic Models*, North Holland, 1977, pp. 365 383.

- [12] Ghysels, E., Santa-Clara, P. and Valkanov, R.(2004), “The MIDAS Touch: Mixed Data Sampling Regression Models,” Manuscript, University of North Carolina.
- [13] Hall, R.E., et al. (2003), “The NBER’s Recession Dating Procedure,” Available at <http://www.nber.org/cycles/recessions.html>
- [14] Hamilton, J.D. (1989), “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle,” *Econometrica*, 57, 357-384.
- [15] Kim, C.-J. and Nelson, C.R. (1998), *State Space Models with Regime Switching: Classical and Gibbs Sampling Approaches with Applications*. Cambridge, Mass.: MIT Press.
- [16] Liu, H. and Hall, S.G. (2001), “Creating High-frequency National Accounts with State-Space Modelling: A Monte Carlo Experiment,” *Journal of Forecasting*, 20, 441-449.
- [17] Lucas, R.E. (1977), “Understanding Business Cycles,” *Carnegie Rochester Conference Series on Public Policy*, 5, 7-29.
- [18] Mariano, R.S. and Murasawa, Y. (2003), “A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series,” *Journal of Applied Econometrics*, 18, 427-443.
- [19] McGuckin, R.H., Ozyildirim, A. and Zarnowitz, V. (2003), “A More Timely and Useful Index of Leading Indicators,” Manuscript, Conference Board.
- [20] Proietti, T. and Moauro, F. (2006), “Dynamic Factor Analysis with Non Linear Temporal Aggregation Constraints,” *Applied Statistics*, 55, 281-300.
- [21] Sargent, T.J. and Sims, C.A. (1977), “Business Cycle Modeling Without Pretending to Have Too Much A Priori Economic Theory,” in C. Sims (ed.), *New Methods in Business Research*. Minneapolis: Federal Reserve Bank of Minneapolis.
- [22] Shen, C.-H. (1996), “Forecasting Macroeconomic Variables Using Data of Different Periodicities,” *International Journal of Forecasting*, 12, 269-282.
- [23] Stock, J.H. and Watson, M.W. (1989), “New Indexes of Coincident and Leading Economic Indicators,” *NBER Macro Annual*, Volume 4. Cambridge, Mass.: MIT Press.
- [24] Stock, J.H. and Watson, M.W. (1991), “A Probability Model of the Coincident Economic Indicators.” In K. Lahiri and G. Moore (eds.), *Leading Economic Indicators: New Approaches and Forecasting Records*. Cambridge: Cambridge University Press, 63-89.



[25] Stock, J.H. and Watson, M.W. (2002), “Macroeconomic Forecasting Using Diffusion Indexes,” *Journal of Business and Economic Statistics*, 20, 147-162.

## Appendix

### A The Mapping for Third-Order Trend Polynomial Coefficients

Here we establish the mapping between two sets of parameters. On the one hand, we have

$$\sum_{j=0}^{D-1} \left[ c + \delta_1 \left( \frac{t-j}{1000} \right) + \delta_2 \left( \frac{t-j}{1000} \right)^2 + \delta_3 \left( \frac{t-j}{1000} \right)^3 \right],$$

and on the other hand we have

$$c^* + \delta_1^* \left( \frac{t}{1000} \right) + \delta_2^* \left( \frac{t}{1000} \right)^2 + \delta_3^* \left( \frac{t}{1000} \right)^3.$$

We want to establish the mapping between  $(c, \delta_1, \delta_2, \delta_3)$  and  $(c^*, \delta_1^*, \delta_2^*, \delta_3^*)$ . We have

$$\begin{aligned} & \sum_{j=0}^{D-1} \left[ c + \delta_1 \left( \frac{t-j}{1000} \right) + \delta_2 \left( \frac{t-j}{1000} \right)^2 + \delta_3 \left( \frac{t-j}{1000} \right)^3 \right] \\ = & \sum_{j=0}^{D-1} c + \delta_1 \sum_{j=0}^{D-1} \left( \frac{t}{1000} - \frac{j}{1000} \right) + \delta_2 \sum_{j=0}^{D-1} \left( \frac{t}{1000} - \frac{j}{1000} \right)^2 + \delta_3 \sum_{j=0}^{D-1} \left( \frac{t}{1000} - \frac{j}{1000} \right)^3 \\ = & Dc + \delta_1 \sum_{j=0}^{D-1} \left( \frac{t}{1000} \right) - \delta_1 \sum_{j=0}^{D-1} \left( \frac{j}{1000} \right) \\ & + \delta_2 \sum_{j=0}^{D-1} \left( \frac{t}{1000} \right)^2 + \delta_2 \sum_{j=0}^{D-1} \left( \frac{j}{1000} \right)^2 - 2\delta_2 \sum_{j=0}^{D-1} \frac{tj}{1000^2} \\ & + \delta_3 \sum_{j=0}^{D-1} \left( \frac{t}{1000} \right)^3 - \delta_3 \sum_{j=0}^{D-1} \left( \frac{j}{1000} \right)^3 - 3\delta_3 \sum_{j=0}^{D-1} \left( \frac{t}{1000} \right)^2 \left( \frac{j}{1000} \right) + 3\delta_3 \sum_{j=0}^{D-1} \left( \frac{t}{1000} \right) \left( \frac{j}{1000} \right)^2 \end{aligned}$$

$$\begin{aligned}
&= Dc - \delta_1 \sum_{j=0}^{D-1} \left( \frac{j}{1000} \right) + \delta_2 \sum_{j=0}^{D-1} \left( \frac{j}{1000} \right)^2 - \delta_3 \sum_{j=0}^{D-1} \left( \frac{j}{1000} \right)^3 \\
&\quad + \frac{t}{1000} \left[ D\delta_1 - 2\delta_2 \sum_{j=0}^{D-1} \frac{j}{1000} + 3\delta_3 \sum_{j=0}^{D-1} \left( \frac{j}{1000} \right)^2 \right] \\
&\quad + \left( \frac{t}{1000} \right)^2 \left[ D\delta_2 - 3\delta_3 \sum_{j=0}^{D-1} \left( \frac{j}{1000} \right) \right] \\
&\quad + \left( \frac{t}{1000} \right)^3 (D\delta_3).
\end{aligned}$$

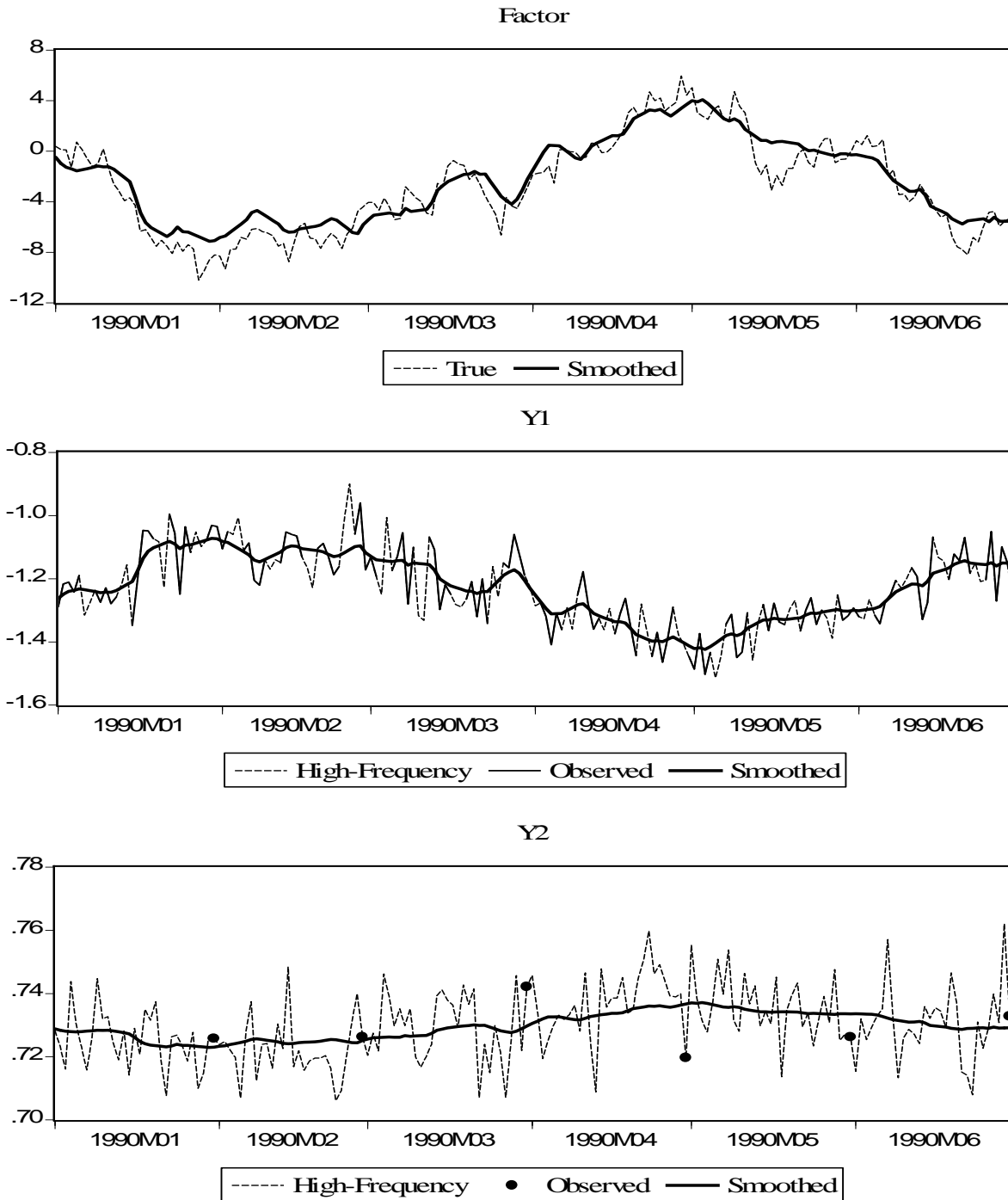
Now, note that

$$\begin{aligned}
\sum_{j=0}^{D-1} j &= \frac{D(D-1)}{2} \\
\sum_{j=0}^{D-1} j^2 &= \frac{D(D-1)[2(D-1)+1]}{6} = \frac{D(D-1)(2D-1)}{6} \\
\sum_{j=0}^{D-1} j^3 &= \left[ \frac{D(D-1)}{2} \right]^2.
\end{aligned}$$

Hence we obtain

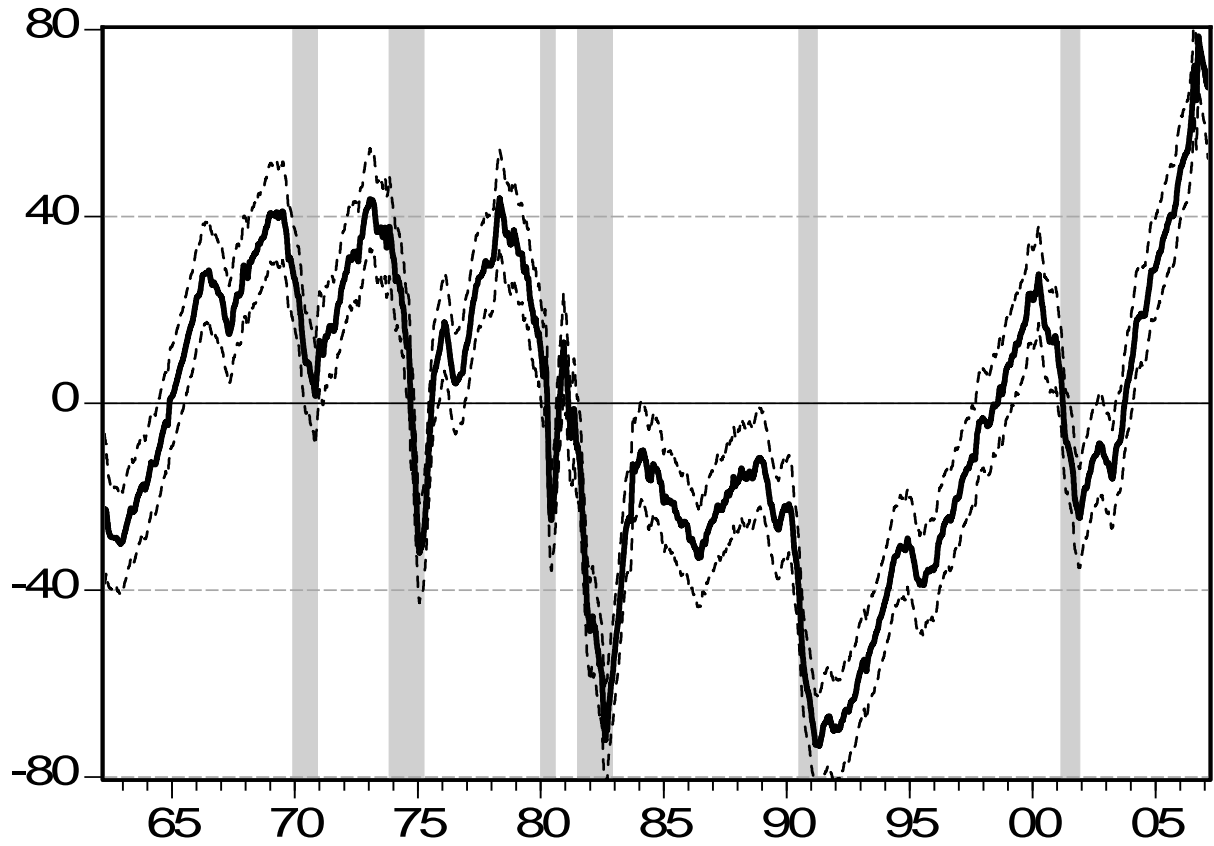
$$\begin{aligned}
c^* &= Dc - \frac{\delta_1 D(D-1)}{2000} + \frac{\delta_2 D(D-1)(2D-1)}{6 \times 10^6} - \frac{\delta_3 [D(D-1)]^2}{4 \times 10^9} \\
\delta_1^* &= D\delta_1 - \frac{\delta_2 D(D-1)}{1000} + \frac{\delta_3 D(D-1)(2D-1)}{2 \times 10^6} \\
\delta_2^* &= D\delta_2 - \frac{3\delta_3 D(D-1)}{2000} \\
\delta_3^* &= D\delta_3.
\end{aligned}$$

**Figure 1 – Simulation: Smoothed Factors and Indicators**



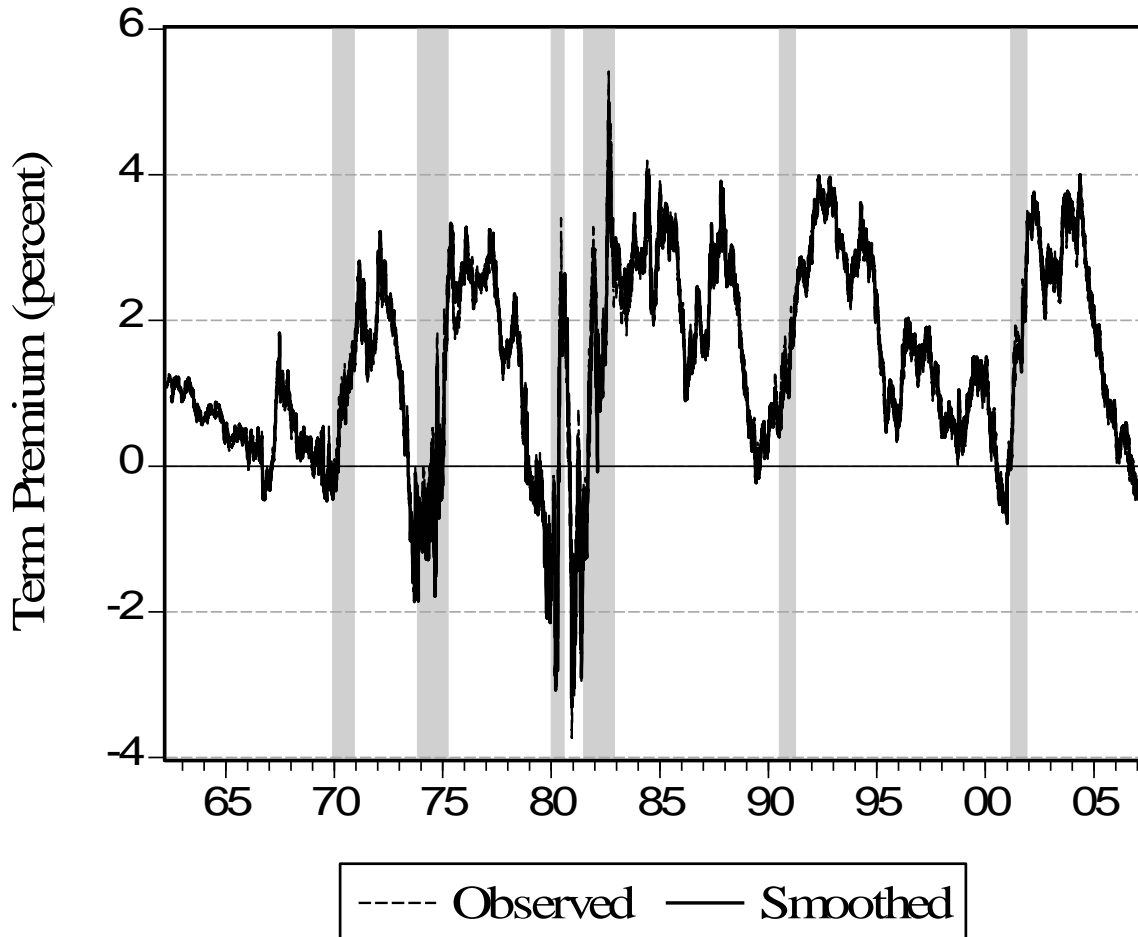
**Notes:** The first panel shows the true and smoothed values of the latent real activity factor. The second panel shows the high-frequency ("daily") values of indicator Y1, the observed values of Y1, and the "smoothed" daily values of Y1 obtained by running the smoothed values of the factor through equation (2) for Y1. The third panel shows the high-frequency ("daily") values of indicator Y2, the observed values of Y2, and the "smoothed" daily values of Y2 obtained by running the smoothed values of the factor through equation (2) for Y2. See text for details.

**Figure 2 – Smoothed U.S. Real Activity Factor**



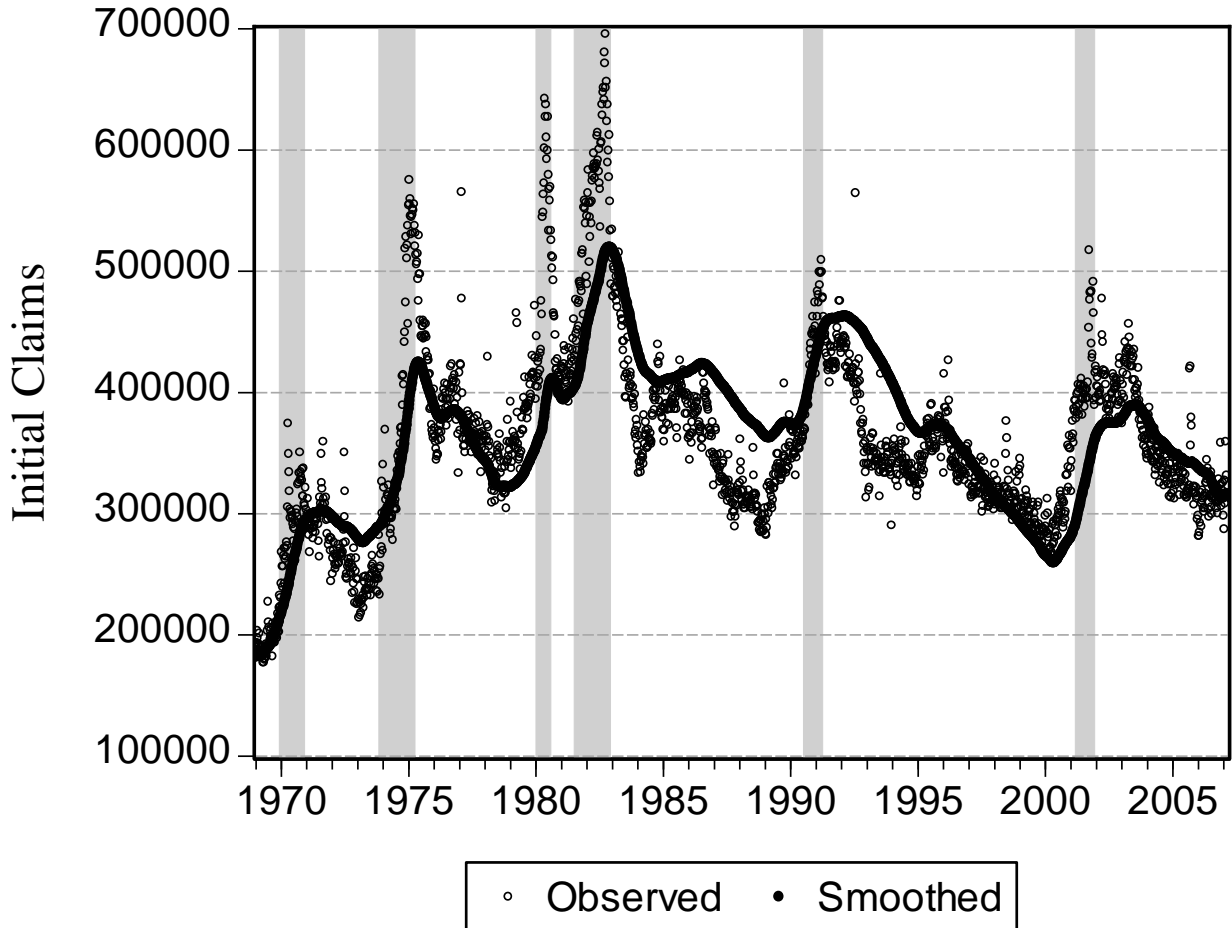
**Notes:** We show the smoothed factor together with ninety-five percent confidence bands (dashed lines). The shaded bars denote NBER recessions. See text for details.

**Figure 3 – Smoothed Indicators I: Term Premium**



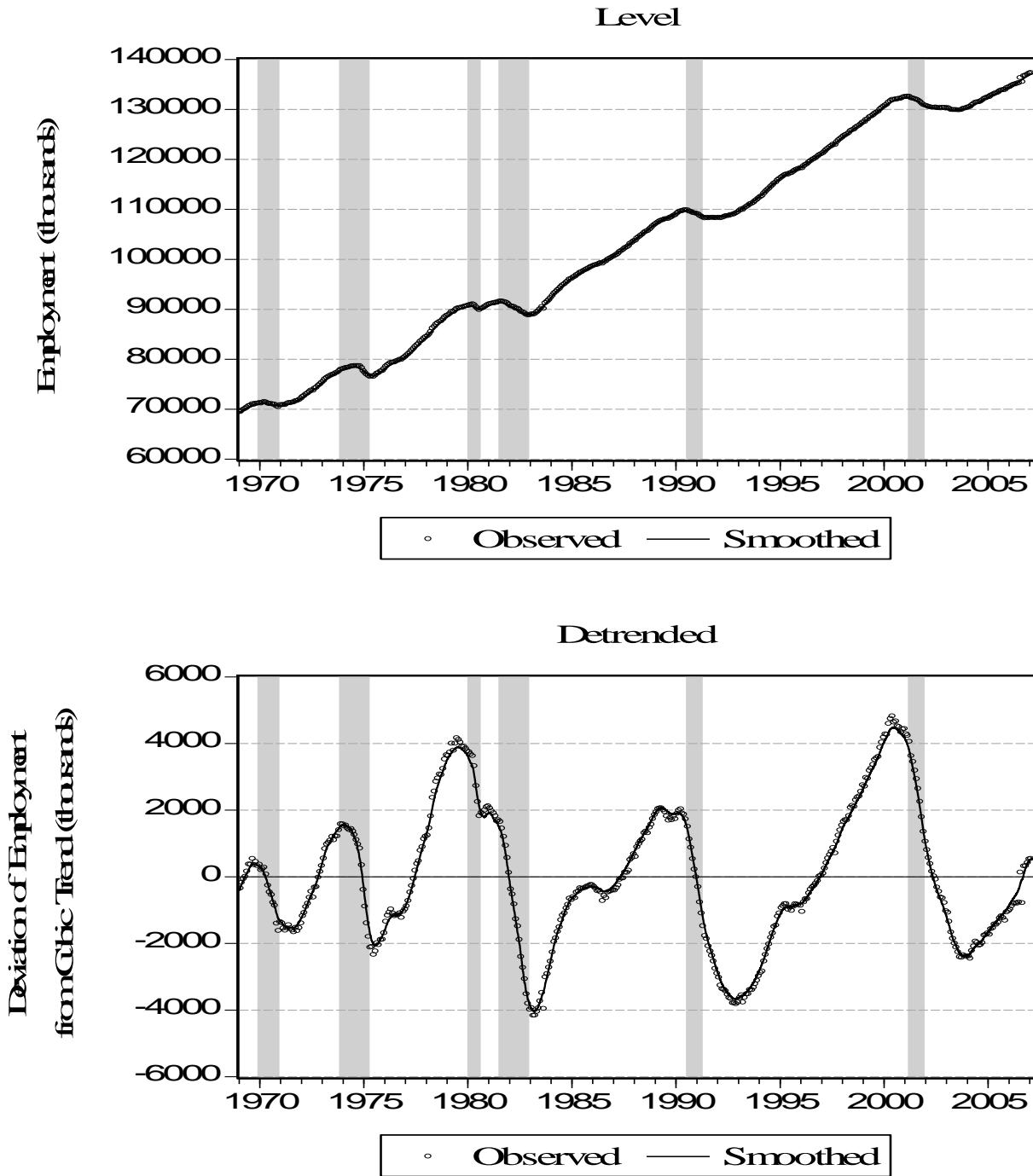
**Notes:** We show the observed and “smoothed” daily term premium. We obtain the smoothed daily term premium by running the smoothed values of the factor through equation (24) for the term premium. See text for details.

**Figure 4 – Smoothed Indicators II: Initial Jobless Claims**



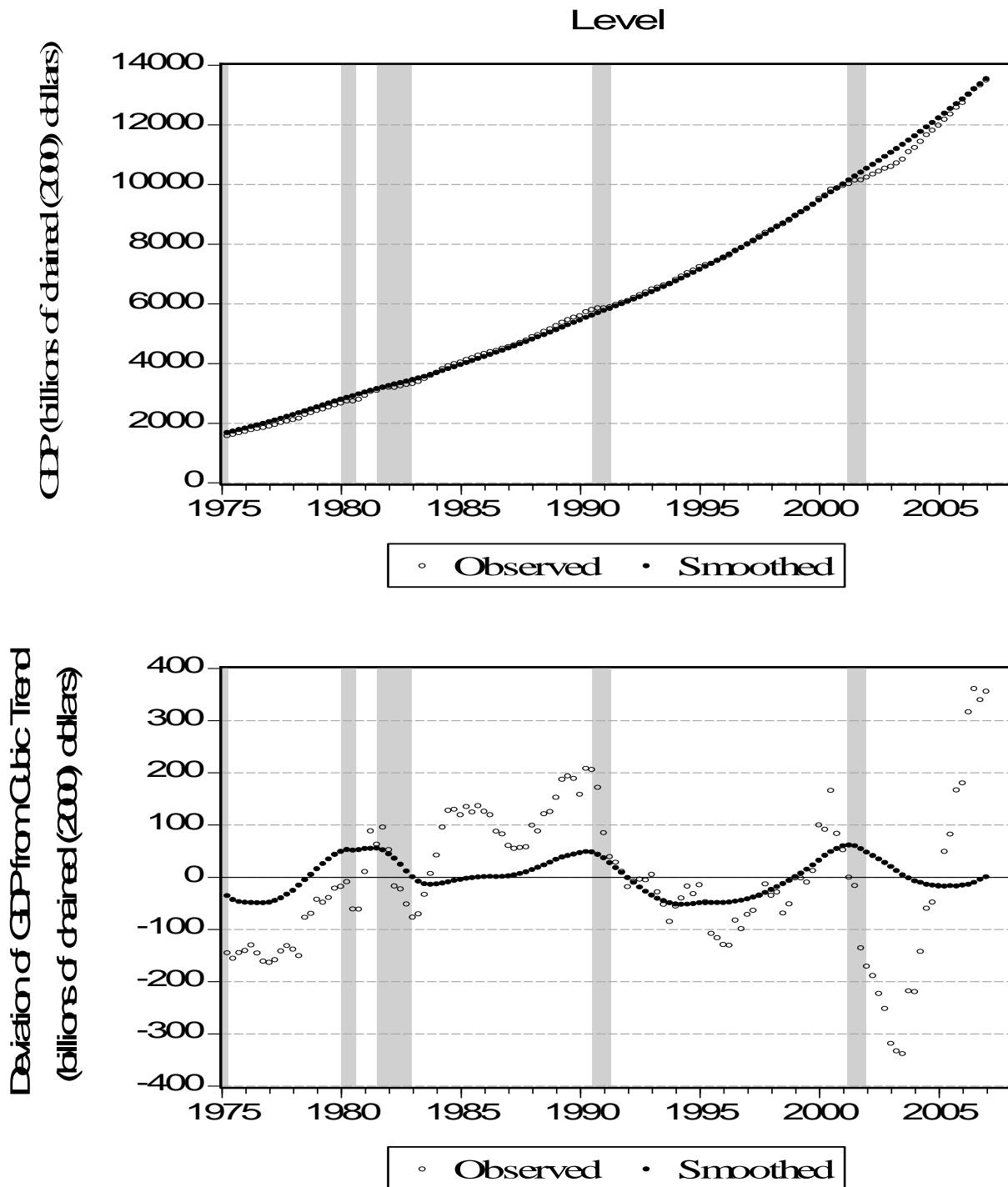
**Notes:** We show the observed and “smoothed” weekly initial jobless claims. We obtain the smoothed daily initial claims by running the smoothed values of the factor through equation (2) for initial claims. We obtain smoothed weekly initial claims (Saturdays) by summing daily smoothed initial claims over the last seven days. See text for details.

**Figure 5 – Smoothed Indicators III: Employment**



**Notes:** We show observed (monthly) and “smoothed” (daily) employment. The top panel shows observed and smoothed values of total employment, and the bottom panel shows observed and smoothed values of detrended employment. We obtain smoothed daily employment by running smoothed values of the factor through equation (2) for employment. See text for details.

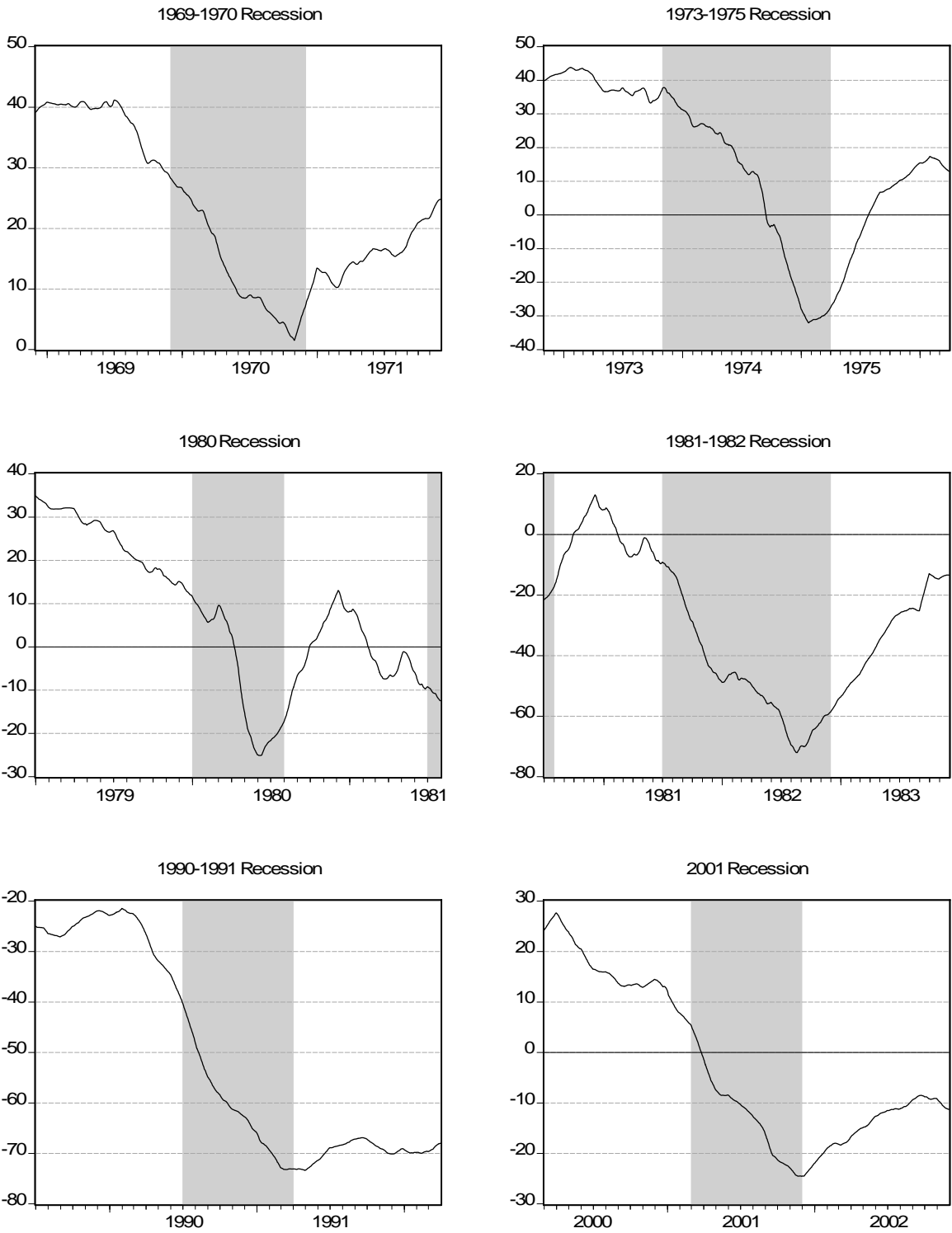
**Figure 6 – Smoothed Indicators IV: GDP**



**Notes:** We show observed (monthly) and “smoothed” quarterly GDP. We obtain the smoothed daily GDP by running the smoothed values of the factor through equation (2) for GDP. We obtain smoothed quarterly GDP by summing daily smoothed GDP over the last 91 days. The top panel shows observed and smoothed values of GDP, and the bottom panel shows observed and smoothed values of detrended GDP. See text for details.

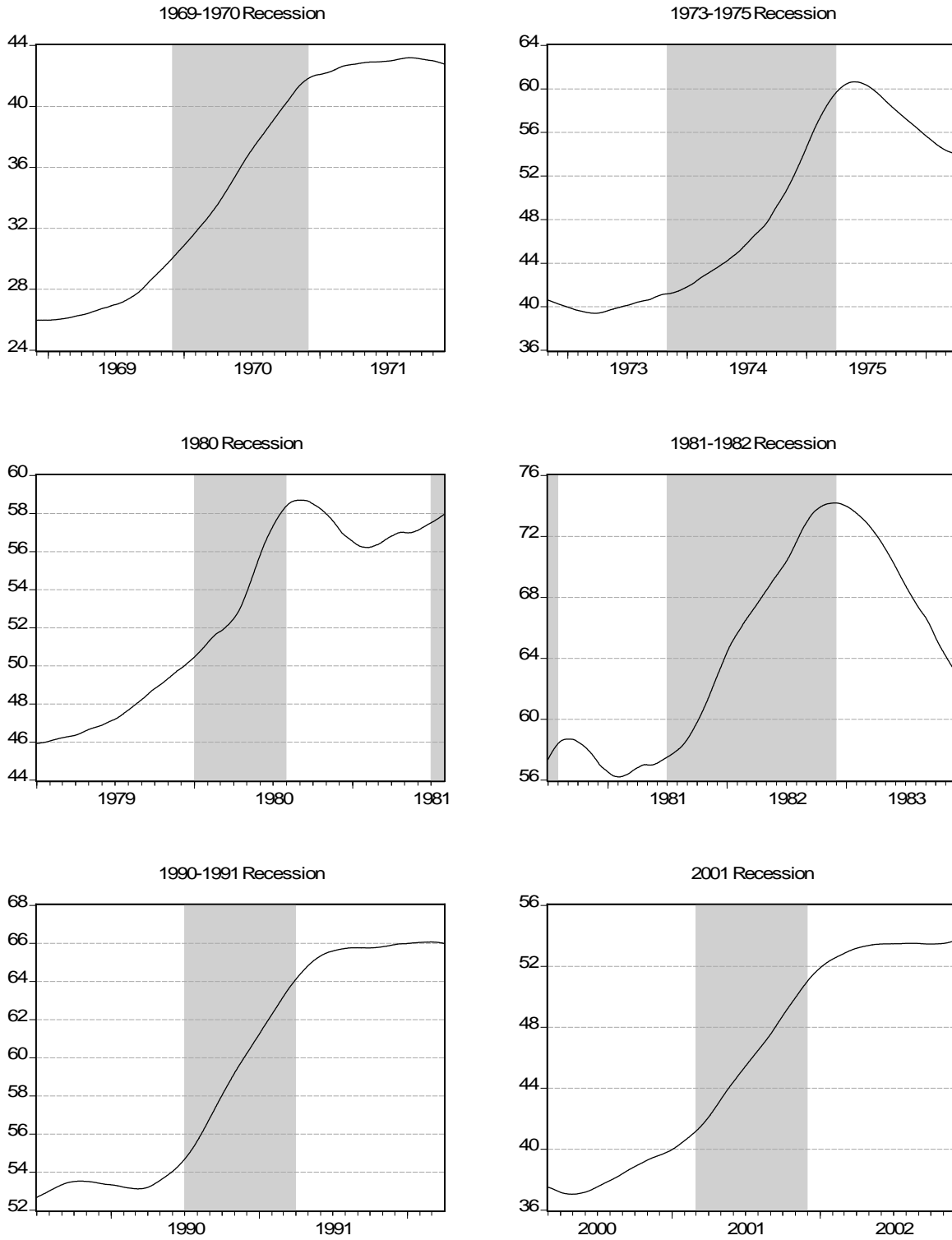


**Figure 7 – Smoothed Factor During Recessions**



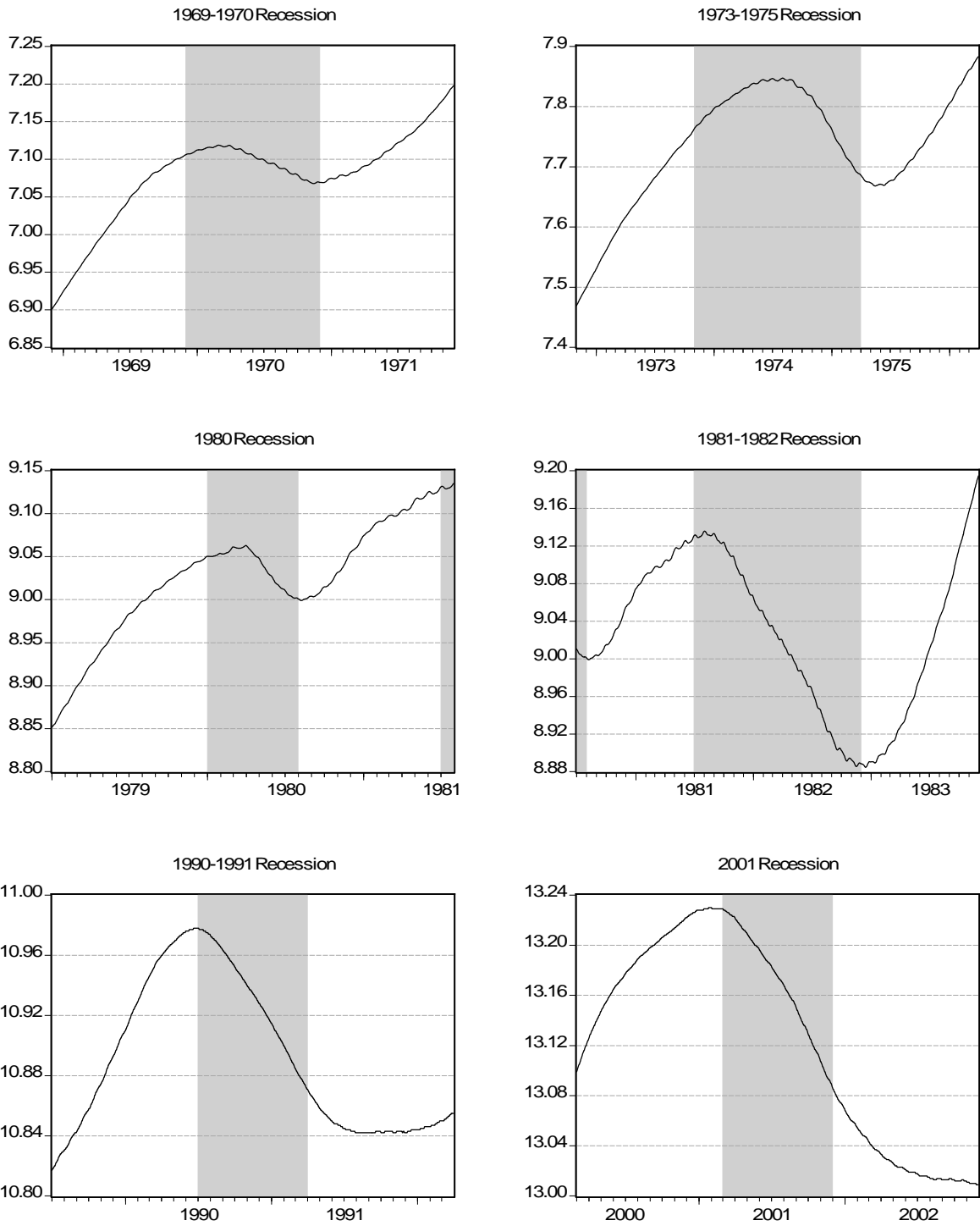
**Notes:** We show the smoothed daily real activity factor during and near the six NBER recessions in our sample. See text for details.

**Figure 8 – Smoothed Initial Claims During Recessions**



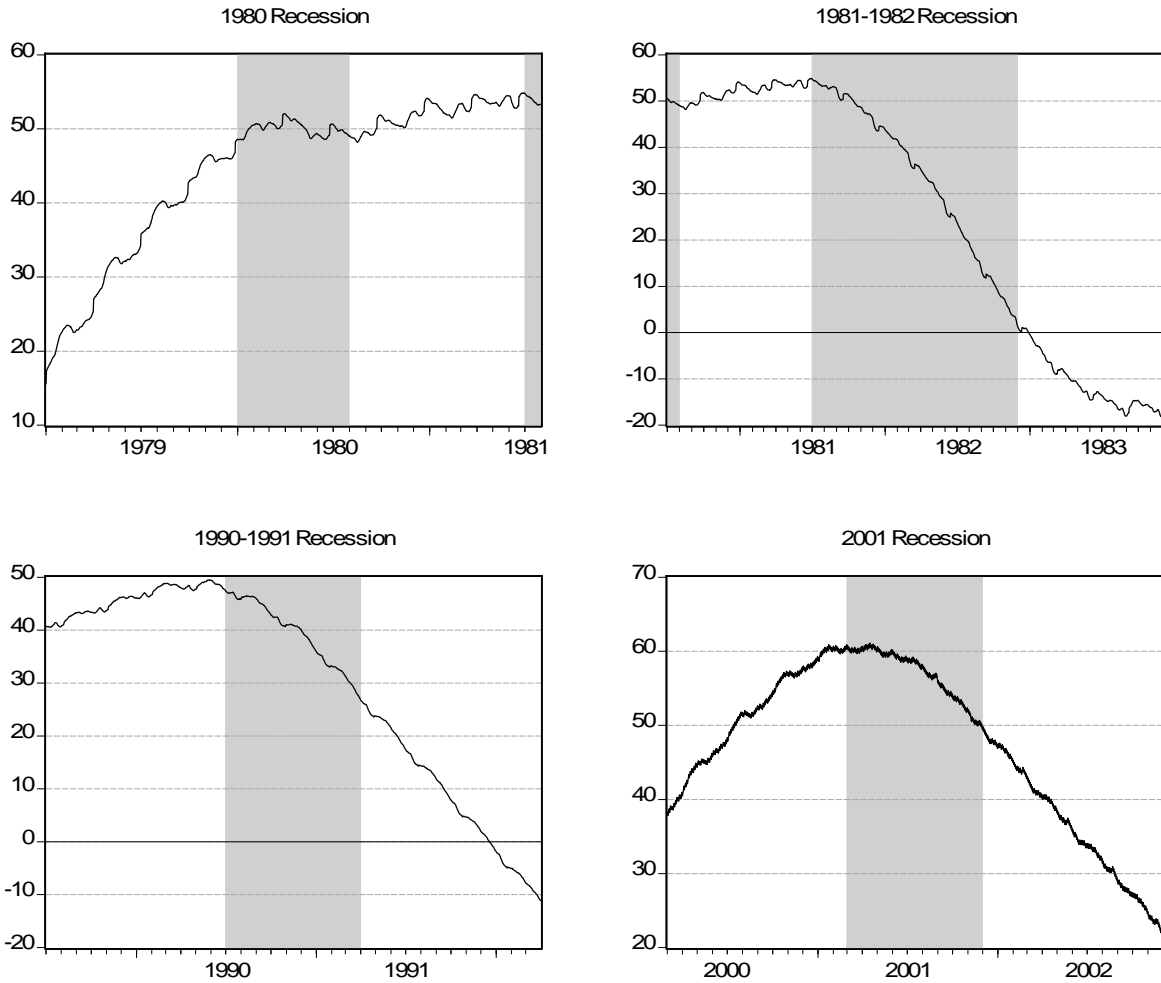
**Notes:** We show the smoothed daily initial claims during and near the six NBER recessions in our sample. See text for details.

**Figure 9 – Smoothed Employment During Recessions**



**Notes:** We show the smoothed daily employment during and near the six NBER recessions in our sample. See text for details.

**Figure 10 – Smoothed GDP During Recessions**



**Notes:** We show the smoothed daily GDP during and near the six NBER recessions in our sample. See text for details.