Real-Time Measurement of Business Conditions

S. Borağan Aruoba
University of Maryland

Francis X. Diebold
University of Pennsylvania and NBER

Chiara Scotti
Federal Reserve Board

October 5, 2007

3rd Annual Workshop on Macroeconomic Forecasting, Analysis and Policy with Data Revisions
Montreal, CANADA
Introduction

Goal of this paper

We propose and illustrate a framework for real-time business conditions assessment in a systematic, replicable, and statistically optimal manner.
(1) Dynamic Factor Model

- Business conditions is an unobserved variable, related to observed indicators.

- Business cycle is about “co-movements” of many variables.

- Extract the common factor.
(1) Dynamic Factor Model

- Business conditions is an unobserved variable, related to observed indicators.
- Business cycle is about “co-movements” of many variables.
- Extract the common factor.

(2) Indicators at different frequencies

- Important indicators have different frequencies: Quarterly (e.g. GDP), Monthly (e.g. Industrial Production), Weekly (e.g. Initial Jobless Claims), and continuously (asset prices).
- We propose a framework that can in principle combine any frequency.
(3) Continuously-evolving indicator

- Real activity evolves continuously

- Our framework allows us to incorporate high-frequency data to extract a “continuous” measure of business conditions.
(3) Continuously-evolving indicator

- Real activity evolves continuously

- Our framework allows us to incorporate high-frequency data to extract a “continuous” measure of business conditions.

(4) Linear, statistically optimal procedure with no approximations

- Temporal aggregation requires an assumption about linking high frequency (latent) indicators to lower-frequency observations.

- Our framework uses an “exact” aggregation.
• **Dynamic Factor Models** : Stock and Watson (1989, 1991)
  One frequency, non-stationary factor

• **Mixed-Frequency Data** : Mariano and Murasawa (2003) and Proietti and Moauro (2006)
  Dynamic factor models, highest frequency monthly, approximations.

• Evans (2005)
  Not a factor model, daily GDP growth, no high-frequency observations, approximation.
Methodology - Missing Observations and Temporal Aggregation

- State of the economy evolves at a high frequency, e.g. daily.

- All economic and financial variables evolve daily, many are not observed daily.
• State of the economy evolves at a high frequency, e.g. daily.

• All economic and financial variables evolve daily, many are not observed daily.

• $y^i_t$: Daily economic or financial variable.

• $\tilde{y}^i_t$: Observed at the (lower) tilde frequency.
Methodology - Missing Observations and Temporal Aggregation

- State of the economy evolves at a high frequency, e.g. daily.
- All economic and financial variables evolve daily, many are not observed daily.
- $y_t^i$: Daily economic or financial variable.
- $\tilde{y}_t^i$: Observed at the (lower) tilde frequency.
- For a stock variable

$$\tilde{y}_t^i = \begin{cases} 
y_t^i & \text{if } y_t^i \text{ is observed} \\
NA & \text{otherwise,} 
\end{cases}$$
• State of the economy evolves at a high frequency, e.g. daily.

• All economic and financial variables evolve daily, many are not observed daily.

• \( y_t^i \) : Daily economic or financial variable.

• \( \tilde{y}_t^i \) : Observed at the (lower) tilde frequency.

• For a stock variable

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

• For a flow variable

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

\( D_i \) : number of days relevant in the aggregation.
Temporal aggregation typically involves approximations.

- Mariano and Murasawa (2003): Quarterly GDP is geometric average of intra-quarter monthly GDPs. \((\text{log of a sum is sum of the log})\)
Temporal aggregation typically involves approximations.

- Mariano and Murasawa (2003) : Quarterly GDP is geometric average of intra-quarter monthly GDPs. (log of a sum is sum of the log)

- Evans (2005) : Quarterly GDP growth rate is the sum of intra-quarter daily growth rates.
Temporal aggregation typically involves approximations.

- Mariano and Murasawa (2003): Quarterly GDP is geometric average of intra-quarter monthly GDPs. (log of a sum is sum of the log)

- Evans (2005): Quarterly GDP growth rate is the sum of intra-quarter daily growth rates.

Temporal aggregation typically involves approximations.

- Mariano and Murasawa (2003) : Quarterly GDP is geometric average of intra-quarter monthly GDPs. (log of a sum is sum of the log)

- Evans (2005) : Quarterly GDP growth rate is the sum of intra-quarter daily growth rates.


Our framework is linear (Kalman filter is statistically optimal), does not involve an approximation.

**Critical assumption** : Levels of all observed variables are stationary deviations from polynomial trends of arbitrary order.
• Daily business conditions:

\[ x_t = \rho_1 x_{t-1} + \ldots + \rho_p x_{t-p} + v_t \]
Methodology - State Space Formulation

- Daily business conditions:

\[ x_t = \rho_1 x_{t-1} + \ldots + \rho_p x_{t-p} + v_t \]

- Daily economic/financial variables (except those observed daily)

\[ y^i_t = c_i + \beta_i x_t + \delta_{i1} w_1^k + \ldots + \delta_{ik} w_k^t + \gamma_{i1} y^i_{t-D_i} + \ldots + \gamma_{in} y^i_{t-nD_i} + \varepsilon^i_t \]
• Daily business conditions:

\[ x_t = \rho_1 x_{t-1} + \ldots + \rho_p x_{t-p} + v_t \]

• Daily economic/financial variables (except those observed daily)

\[ y_t^i = c_i + \beta_i x_t + \delta_{i1} w^1_t + \ldots + \delta_{ik} w^k_t + \gamma_{i1} y_{t-D_i}^i + \ldots + \gamma_{in} y_{t-nD_i}^i + \varepsilon_t^i \]

• The measurement equation for all non-daily stock variables:

\[ y_t^i = \begin{cases} 
  c_i + \beta_i x_t + \delta_{i1} w^1_t + \ldots + \delta_{ik} w^k_t + \gamma_{i1} y_{t-D_i}^i + \ldots + \gamma_{in} y_{t-nD_i}^i + \varepsilon_t^i & y_t^i \text{ observed} \\
  NA & \text{otherwise.} 
\end{cases} \]
• Exact linear temporal aggregation

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

- Exact linear temporal aggregation

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

- The **measurement equation** for all non-daily **flow** variables:

\[
\tilde{y}_t^i = \begin{cases} 
  \sum_{j=0}^{D_i-1} c_i + \beta_i \sum_{j=0}^{D_i-1} x_{t-j}^i + \delta_{i1} \sum_{j=0}^{D_i-1} w_{t-j}^1 + \ldots + \delta_{ik} \sum_{j=0}^{D_i-1} w_{t-j}^k & y_t^i \text{ observed} \\
  + \gamma_{i1} \sum_{j=0}^{D_i-1} y_{t-D_i-j}^i + \ldots + \gamma_{in} \sum_{j=0}^{D_i-1} y_{t-nD_i-j}^i + \varepsilon_t^i & \text{otherwise,}
\end{cases}
\]
Methodology - State Space Formulation

\[
\tilde{y}_t = \begin{cases} 
D_i-1 & \sum_{j=0}^{D_i-1} c_i + \beta_i \sum_{j=0}^{D_i-1} x_t^{i-j} + \delta_{i1} \sum_{j=0}^{D_i-1} w_{t-j}^1 + \ldots + \delta_{ik} \sum_{j=0}^{D_i-1} w_{t-j}^k \\
+ \gamma_{i1} \sum_{j=0}^{D_i-1} y_{t-D_i-j}^i + \ldots + \gamma_{in} \sum_{j=0}^{D_i-1} y_{t-nD_i-j}^i + \varepsilon_t^i \\
NA & y_t^i \text{ observed} \\
& y_t^i \text{ otherwise,}
\end{cases}
\]

- Observations:
  - \( \sum_{j=0}^{D_i-1} y_{t-nD_i-j}^i = \tilde{y}_{t-nD_i}^i \) by definition (observed flow variable \( n \) periods ago)
  - \( \varepsilon_t^i \) is the sum of the \( \varepsilon_t^i \) over the tilde period.
  - Appropriately treat \( \varepsilon_t^i \) as white noise with \( var(\varepsilon_t^i) = D_i \cdot var(\varepsilon_t^i) \).
Methodology - State Space Formulation

- We use $w_t^j$ to handle trend, using:

\[
\sum_{j=0}^{D_i-1} \left[ c_i + \delta_{i1} (t - j) + ... + \delta_{ik} (t - j)^k \right] \equiv c_i^* + \delta_{i1}^* t + ... + \delta_{ik}^* t^k,
\]
Methodology - State Space Formulation

- We use $w_t^j$ to handle trend, using:

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

- The **measurement equation** for all non-daily flow variables:

$$
\tilde{y}_t^i = \begin{cases} 
  c_i^* + \beta_i \sum_{j=0}^{D_i-1} x_{t-j}^i + \delta_{i1} t + \ldots + \delta_{ik} t^k + \gamma_{i1} \tilde{y}_{t-D_i}^i + \ldots + \gamma_{in} \tilde{y}_{t-nD_i}^i + \varepsilon_t^i \\
  \text{NA}
\end{cases}
$$
The measurement equation for all daily (flow/stock) variables:

\[ y_t^i = c_i + \beta_i^0 x_t + \beta_i^1 x_{t-1} + \ldots + \beta_i^{\tilde{D}} x_{t-\tilde{D}} + \delta_{i1} w_t^1 + \ldots + \delta_{ik} w_t^k + \gamma_{i1} y_{t-1}^i + \ldots + \gamma_{in} y_{t-n}^i + \epsilon_t^i \]

where the elements of \( \{\beta_i^j\}_{j=0}^{\tilde{D}} \) follow a low-ordered polynomial.
Methodology - Initialization, Filtering and Smoothing

\[ y_t = Z_t \alpha_t + \Gamma_t w_t + \varepsilon_t \]
\[ \alpha_{t+1} = T \alpha_t + R \eta_t \]
\[ \varepsilon_t \sim (0, H_t), \eta_t \sim (0, Q), \]

**Remark:** Matrices \( T, R, \) and \( Q \) are constant. \( Z_t, \Gamma_t \) and \( H_t \) are not because of variation in the number of days in a month/quarter.
\[ y_t = Z_t \alpha_t + \Gamma_t w_t + \varepsilon_t \]
\[ \alpha_{t+1} = T \alpha_t + R \eta_t \]
\[ \varepsilon_t \sim (0, H_t), \eta_t \sim (0, Q), \]

**Remark:** Matrices \( T, R, \) and \( Q \) are constant. \( Z_t, \Gamma_t \) and \( H_t \) are not because of variation in the number of days in a month/quarter.

Standard Kalman filter initialization, filtering and smoothing and estimation via maximum likelihood, adjusted for missing observations.
Methodology - Initialization, Filtering and Smoothing

\[ y_t = Z_t \alpha_t + \Gamma_t w_t + \varepsilon_t \]
\[ \alpha_{t+1} = T_\alpha \alpha_t + R\eta_t \]
\[ \varepsilon_t \sim (0, H_t), \eta_t \sim (0, Q), \]

Remark: Matrices \( T, R, \) and \( Q \) are constant. \( Z_t, \Gamma_t \) and \( H_t \) are not because of variation in the number of days in a month/quarter.

Standard Kalman filter initialization, filtering and smoothing and estimation via maximum likelihood, adjusted for missing observations.

Initialization: We have a stationary system. Use \( \alpha_1 \sim N(a_1, P_1) \) where \( a_1 = 0_{m \times 1} \) and \( P_1 \) solves

\[ (I - T \otimes T) \text{vec}(P_1) = \text{vec}(RQR') \]
Filtering : Use the “contemporaneous” Kalman filter equations:

Without missing observations

\[ v_t = y_t - Z_t a_t - \Gamma_t w_t \]
\[ F_t = Z_t P_t Z'_t + H_t \]
\[ a_{t|t} = a_t + P_t Z'_t F_t^{-1} v_t \]
\[ P_{t|t} = P_t - P_t Z'_t F_t^{-1} Z_t P'_t \]
\[ a_{t+1} = T a_{t|t} \]
\[ P_{t+1} = T P_{t|t} T' + RQR'. \]
**Filtering** : Use the “contemporaneous” Kalman filter equations:

**Without missing observations**

\[
\begin{align*}
v_t &= y_t - Z_t a_t - \Gamma_t w_t \\
F_t &= Z_t P_t Z'_t + H_t \\
a_{t|t} &= a_t + P_t Z'_t F_t^{-1} v_t \\
P_{t|t} &= P_t - P_t Z'_t F_t^{-1} Z_t P_t' \\
a_{t+1} &= T a_{t|t} \\
P_{t+1} &= T P_{t|t} T' + R Q R'.
\end{align*}
\]

**All observations in** \(y_t\) **missing**

\[
\begin{align*}
a_{t+1} &= T a_t \\
P_{t+1} &= T P_t T' + R Q R.
\end{align*}
\]

**Some observations in** \(y_t\) **missing** : Replace measurement equation with

\[
\begin{align*}
y_t^* &= Z_t^* \alpha_t + \Gamma_t w_t + \varepsilon_t^* \\
\varepsilon_t^* &\sim N(0, H_t^*),
\end{align*}
\]
Estimation: Use the prediction error decomposition to get $\log L$. Use a classical approach today.

Without missing observations

$$\log L = -\frac{1}{2} \sum_{t=1}^{T} [N \log 2\pi + (\log |F_t| + v_t' F_t^{-1} v_t)]$$
Methodology - Initialization, Filtering and Smoothing

**Estimation**: Use the prediction error decomposition to get $\log L$.

Use a classical approach today.

**Without missing observations**

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

**All observations in $y_t$ missing**: Contribution of $t$ to $\log L$ is zero.
Methodology - Initialization, Filtering and Smoothing

**Estimation**: Use the prediction error decomposition to get $log L$.

Use a classical approach today.

**Without missing observations**

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

**All observations in $y_t$ missing**: Contribution of $t$ to $log L$ is zero.

**Some observations in $y_t$ missing**: Contribution of $t$ to $log L$ is

$$
[N^* \log 2\pi + (\log |F_t^*| + v_t^* F_t'^{-1} v_t^*)]
$$
Assumptions/Restrictions

- Variance of $v_t$ is normalized to unity.
- For trend use $t/1000$, $(t/1000)^2$ etc. for numerical stability.
- Use the one-to-one mapping between the parameters of an AR($p$) process and the first $p$ partial autocorrelations shown by Barndorff-Nielsen and Schou (1973) to impose stationarity. For $p = 3$

$$
\begin{align*}
\rho_1 &= \pi_1 - \pi_1 \pi_2 - \pi_3 \pi_2 \\
\rho_2 &= \pi_2 - \pi_1 \pi_3 + \pi_1 \pi_2 \pi_3 \\
\rho_3 &= \pi_3
\end{align*}
$$

- Guarantee non-negativity of variance terms by estimating natural logarithms.
- Loadings on some “obvious” variables are restricted to be positive (e.g. GDP) or negative (e.g. IJC).
- We abstract from real-time data issues (data uncertainty) for now.
An Artificial Data Example

- True daily factor follows an AR(1).
- Three daily variables are linked to this factor and a linear trend.
- Generate 40 years’ worth of data of daily variables.
An Artificial Data Example

- True daily factor follows an AR(1).

- Three daily variables are linked to this factor and a linear trend.

- Generate 40 years’ worth of data of daily variables.

- Transform to observed variables:
  - $y_t^1$ (Daily financial variable): Eliminate weekends.
  - $y_t^2$ (Monthly stock variable): Eliminate all observations except last observation in the month.
  - $y_t^3$ (Quarterly flow variable): All observations except last observation in the quarter are missing and the last observation in the quarter is the sum of daily observation over a quarter.
An Artificial Data Example

\[
y_t = \begin{bmatrix} \tilde{y}_t^1 \\ \tilde{y}_t^2 \\ \tilde{y}_t^3 \end{bmatrix}, \quad \alpha_t = \begin{bmatrix} x_t \\ x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-q+1} \\ x_{t-q} \end{bmatrix}, \quad w_t = \begin{bmatrix} 1 \\ t \end{bmatrix}, \quad \varepsilon_t = \begin{bmatrix} u_t^1 \\ u_t^2 \\ u_t^*3 \end{bmatrix}, \quad v_t = \eta_t
\]

\[
Z = \begin{bmatrix} \beta_1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \beta_2 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \beta_3 & \cdots & \beta_3 \text{ or } 0 & \beta_3 \text{ or } 0 & \beta_3 \text{ or } 0 \\ \end{bmatrix}, \quad \Gamma = \begin{bmatrix} c_1 & \delta_1 \\ c_2 & \delta_2 \\ c_{3t} & \delta_{3t} \end{bmatrix}
\]
An Artificial Data Example

\[ T = \begin{bmatrix}
\rho & 0 & 0 & \cdots & 0 & 0 \\
1 & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & 1 & 0
\end{bmatrix}, \quad R = \begin{bmatrix}
1 \\
0 \\
\vdots \\
0
\end{bmatrix} \]

\[ \begin{bmatrix}
\varepsilon_t \\
v_t
\end{bmatrix} \sim N \left( \begin{bmatrix}
0_{3 \times 1} \\
0
\end{bmatrix}, \begin{bmatrix}
H_t & 0 \\
0 & Q
\end{bmatrix} \right), \quad H_t = \begin{bmatrix}
\sigma_1^2 & 0 & 0 \\
0 & \sigma_2^2 & 0 \\
0 & 0 & \sigma_{3t}^2
\end{bmatrix}, \quad Q = 1 \]
An Artificial Data Example

To mimic our estimation of the full model:

- Estimate the smaller two-variable system with \( y_t^1 \) and \( y_t^2 \) which requires only one lag of \( x_t \) in \( \alpha_t \). A total of 9 parameters and 1 state variable.

- Run the Kalman smoother to get the smooth factor \( \hat{x}_t \).
An Artificial Data Example

To mimic our estimation of the full model:

- Estimate the smaller two-variable system with $y_t^1$ and $y_t^2$ which requires only one lag of $x_t$ in $\alpha_t$. A total of 9 parameters and 1 state variable.

- Run the Kalman smoother to get the smooth factor $\hat{x}_t$.

- Estimate the auxiliary regression

$$\tilde{y}_t^3 = \sum_{j=0}^{q-1} [a + d(t-j)] + b(\hat{x}_t + \hat{x}_{t-1} + \ldots + \hat{x}_{t-q}) + e_t$$
An Artificial Data Example

To mimic our estimation of the full model:

- Estimate the smaller two-variable system with $y_t^1$ and $y_t^2$ which requires only one lag of $x_t$ in $\alpha_t$. A total of 9 parameters and 1 state variable.

- Run the Kalman smoother to get the smooth factor $\hat{x}_t$.

- Estimate the auxiliary regression

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

- Use estimates for $a$, $b$, $d$ and $\text{var}(e_t)/\bar{q}$ as starting values for $c_3$, $\beta_3$, $\delta_3$ and $\sigma_3^2$ along with estimates from the smaller model.

- Estimate the full model. A total of 13 parameters and 92 state variables.
An Artificial Data Example

Factor

True Smoothed

90M01 90M02 90M03 90M04 90M05 90M06

True

Smoothed
An Artificial Data Example

Y2

High-Frequency
Observed
Smoothed

90M01 90M02 90M03 90M04 90M05 90M06

- Line color: Blue for High-Frequency, Green for Smoothed, Red for Observed.
A Four-Variable Model - Data

- April 1, 1962 through February 20, 2007
- 7-day weeks
- Variables:
  - Yield curve term premium defined as the difference between the yield of the 10-year and the 3-month Treasury yields. [TERM] This is a daily variable.
  - Average weekly initial claims for unemployment insurance. [IJC] This is a weekly flow variable covering the 7-day period from Sunday through Saturday. The value for Saturdays is the sum of the daily values for the previous 7-days.
  - Employees on nonagricultural payrolls. [EMP] This is a monthly stock variable, observed on the last day of the month.
  - Real GDP. [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.
A Four-Variable Model - Model

\[ y_t = \begin{bmatrix} \tilde{y}_t^1 \\ \tilde{y}_t^2 \\ \tilde{y}_t^3 \\ \tilde{y}_t^4 \end{bmatrix}, \alpha_t = \begin{bmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-q-1} \\ x_{t-q} \\ u_t^1 \\ u_{t-1}^1 \\ u_{t-2}^1 \end{bmatrix}, \omega_t = \begin{bmatrix} 1 \\ t \\ t^2 \\ t^3 \\ \tilde{y}_{t-W}^2 \\ \tilde{y}_{t-2W}^2 \\ \tilde{y}_{t-3W}^2 \\ \tilde{y}_{t-M}^3 \\ \tilde{y}_{t-2M}^3 \\ \tilde{y}_{t-3M}^3 \\ \tilde{y}_{t-q}^4 \\ \tilde{y}_{t-2q}^4 \\ \tilde{y}_{t-3q}^4 \end{bmatrix}, \varepsilon_t = \begin{bmatrix} 0 \\ u_t^2 \\ u_t^3 \\ u_t^4 \end{bmatrix}, \nu_t = \begin{bmatrix} \eta_t \\ \zeta_t \end{bmatrix}, R = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \]
### A Four-Variable Model - Model

\[
Z = \begin{bmatrix}
\beta_1^0 & \beta_1^1 & \cdots & \beta_1^6 & \beta_1^7 & \cdots & \beta_1^{q-1} & \beta_1^q & 1 & 0 & 0 \\
\beta_2 & \beta_2 & \cdots & \beta_2 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
\beta_3 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
\beta_4 & \beta_4 & \cdots & \beta_4 & \beta_4 & \cdots & \beta_4 & \beta_4 & 0 & 0 & 0 \\
\end{bmatrix}
\]

\[
\Gamma = \begin{bmatrix}
c_1 & \delta_{11} & \delta_{12} & \delta_{13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
c_2^* & \delta_{21t}^* & \delta_{22t}^* & \delta_{23t}^* & \gamma_{21} & \gamma_{22} & \gamma_{23} & 0 & 0 & 0 & 0 & 0 \\
c_3 & \delta_{31} & \delta_{32} & \delta_{33} & 0 & 0 & 0 & \gamma_{31} & \gamma_{32} & \gamma_{33} & 0 & 0 \\
c_4^* & \delta_{41t}^* & \delta_{42t}^* & \delta_{43t}^* & 0 & 0 & 0 & 0 & 0 & 0 & \gamma_{41} & \gamma_{42} & \gamma_{43} \\
\end{bmatrix}
\]
A Four-Variable Model - Model

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

\[
\begin{bmatrix}
\varepsilon_t \\
v_t 
\end{bmatrix} \sim N \left( \begin{bmatrix}
0_{4 \times 1} \\
0_{2 \times 1}
\end{bmatrix}, \begin{bmatrix}
H_t & 0 \\
0 & Q
\end{bmatrix} \right), \quad H_t = \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & \sigma_{2t}^2 & 0 & 0 \\
0 & 0 & \sigma_{3t}^2 & 0 \\
0 & 0 & 0 & \sigma_{4t}^2
\end{bmatrix}, \quad Q = \begin{bmatrix}
1 & 0 \\
0 & \sigma_1^2
\end{bmatrix}
\]

Remark: We have 95 state variables, 42 coefficients, 16,397 daily observations.
A Four-Variable Model - Results

Smoothed U.S. Real Activity Factor

[Graph showing trends in smoothed U.S. real activity factor over time from 1965 to 2005. The graph includes two lines: one blue and one red, with shaded areas indicating significant periods.]
A Four-Variable Model - Results

Smoothed Indicators I: Term Premium

![Graph showing observed and smoothed term premiums over time]

- Term Premium (percent)
- Observed vs. Smoothed

Year: 65 to 05
A Four-Variable Model - Results

Smoothed Indicators II: Initial Jobless Claims

- Observed
- Smoothed
A Four-Variable Model - Results

Smoothed Indicators IV: GDP

Level

GDP (billions of chained (2000) dollars)

Deviation of GDP from Cubic Trend (billions of chained (2000) dollars)

Observed
Smoothed

Observed
Smoothed
Smoothened Variables Around Turning Points

The Factor

1969-1970 Recession

1973-1975 Recession

1980 Recession

1981-1982 Recession

1990-1991 Recession

2001 Recession
Smoothed Variables Around Turning Points

Initial Jobless Claims

- 1969-1970 Recession
- 1973-1975 Recession
- 1979 Recession
- 1980 Recession
- 1981-1982 Recession
- 1990-1991 Recession
- 2001 Recession
Smoothed Variables Around Turning Points

Employment

1969-1970 Recession
1973-1975 Recession
1979-1980 Recession
1981-1982 Recession
1990-1991 Recession
2001 Recession
Smoothed Variables Around Turning Points

GDP

1980 Recession

1981-1982 Recession

1990-1991 Recession

2001 Recession
Conclusions and Further Work

• A framework for measuring macroeconomic activity in real time

• Stock/flow variables

• Mixed frequencies

• Exact aggregation

• Optimal extraction of the business conditions as a latent factor

• Extensions:
  – Incorporating qualitative information such as headline news.
  – Constructing a real-time composite leading index.
  – Allow for non-linear regime switching dynamics.
  – Comparative assessment of “small data” and “big data” approaches.