

Mixed-frequency large-scale factor models

Mirco Rubin

Università della Svizzera Italiana &
swiss:finance:institute

8th ECB Workshop on Forecasting Techniques,
Frankfurt am Main, Germany



June, 14th 2014

Joint work with E. Andreou, P. Gagliardini and E. Ghysels

Latent factors models have been proposed for:

- **High Frequency** (HF, m) datasets: *Stock and Watson (SW 1988, 2002), CFNAI, ...*
- **Low Frequency** (LF, q/a) datasets: *SW (2012), Forni and Reichlin (1998), ...*
- **Mixed Frequency** (MF) datasets, especially for nowcasting:
 - Estimating a factor from MF observables \approx estimating an HF factor with missing observations \Rightarrow state space models
 - **small cross-sectional dimension** N : *Mariano and Murasawa (2003), Arouba et al. (2009), ...*
 - **large** N_H : *FACTOR-MIDAS Marcellino and Schumacher (2010), NEW EUROCOIN Altissimo et al. (2010)*
 - potentially **large** N_H, N_L : *SW (2002), Doz et al. (2011).*

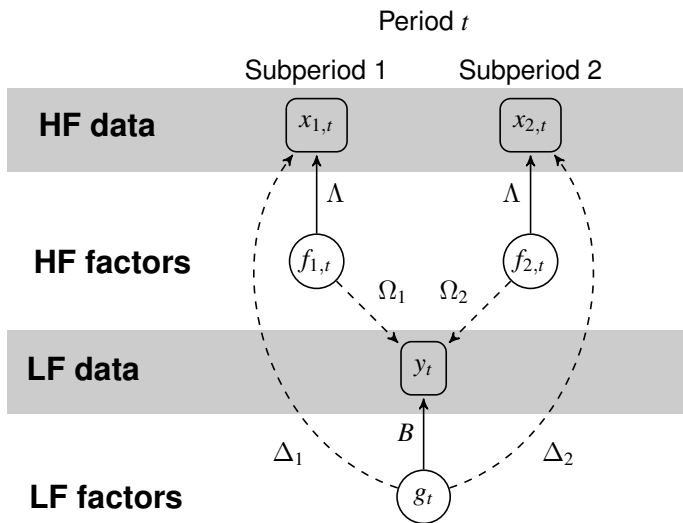
MAIN CONTRIBUTION OF THIS PAPER:

- Large scale factor model allowing for MF observables as well as **HF and LF latent factors.**

- We introduce a **large scale mixed frequency factor model** allowing for:
 - (i) **large panels of both high-frequency (HF) and low-frequency (LF) observed variables** ;
 - (ii) finite number of both **latent HF and LF factors**, i.e. we allow for the presence of both “**fast**” and “**slow**” **evolving factors**.
- We propose an **iterative procedure for the estimation of the latent factors**, based on alternating principal component analysis (PCA) on the HF and LF panels of observations.
- We rely on the recent work on **mixed frequency VAR models** in *Ghysels (2012)* for the modeling and estimation of the factors dynamic [see also *Anderson et al. (2012)* and *McCracken et al. (2013)*].

Introduction: the model

Scheme of the model with **two HF subperiods**:

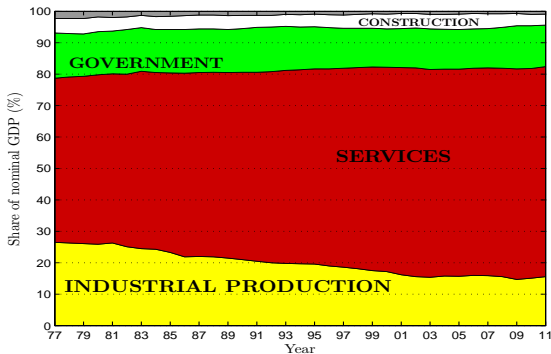


- Business cycle fluctuations are often viewed as a single factor driving a large cross-section of macroeconomic time series, see SW (1989).
- *Foerster, Sarte and Watson (FSW 2011)* find that nearly all variability of US **Industrial Production** (IP) index is due to a common factor to all 117 sectoral IP indexes (**quart. data: HF**).

OUR CONTRIBUTIONS:

- We study the effects of the common Industrial Production factor on the growth of **the other sectors** of the U.S. economy, measured as **quarterly GDP (or GROSS OUTPUT) growth of non-IP sectors (annual data: LF)**.
- We extract technological shocks in mixed-frequency settings to include LF non-IP output growth.
- We use our MF factor model to extract 1 HF factor and 1 LF factor, and study their effects on the sectors of the US economy.

Introduction: empirical application



- Industrial Production Sectors (yellow): *Manufacturing* (main component) + *Utilities* + *Forestry, etc.* + *Mining*.
- Services (red): *Retail and Wholesale Trade*, *Transportation*, *Finance*, *Insurance*, *Information*, *Arts, Entertainment*, etc.
- **Declining contribution of Industrial Production Sectors to US GDP.**

- **Large scale latent factor models:** *Chamberlain (1983), Chamberlain and Rothschild (1983), SW (1988, 1989, 2002 a, b, 2012), Bai and Ng (2002), Bai (2003), Forni et al. (1998, 2005, 2009, 2012), ...*
- **Mixed frequency factor models (nowcasting, forecasting and ragged-edge data):** *Mariano and Murasawa (2003), Nunes (2005), Breitung and Schumacher (2008), Giannone et al. (2008), Arouba et al. (2009), Altissimo et al. (2010), Marcellino and Schumacher (2010), Banbura and Runstler (2010), Frale and Monteforte (2010), Doz et al. (2011), Frale et al. (2011), Ghysels et al. (2014), ...*
- **Group specific / hierarchical latent factor models:** *Kose et al. (2003), Diebold (2008), Giannone et al. (2008), Wang (2010), Moench and Ng (2011), Moench et al. (2012), Heaton and Solo (2012), Bai and Ando (2013), ...*

- **MF VAR:** *Chiu, Eraker et al. (2011), Ghysels (2012), Anderson et al. (2012), McCracken et al. (2013), ...*
- **MF data surveys:** *Forni and Marcellino (2013), Andreou et al. (2012), Forni, Ghysels and Marcellino (2013),...*
- **Business cycle fluctuations:** *Long and Plosser (1983, 1987), Horvath (1998, 2000), Dupor (1999), FSW (2011), Gabaix (2011), Acemoglu et al. (2012), Carvalho and Gabaix (2013), Atalay (2014),...*

Outline

- 1 Introduction ✓
- 2 The model
- 3 The estimator
- 4 Empirical application
- 5 Technological shocks
- 6 Work in progress and concluding remarks

The **mixed frequency factor model** in stacked form:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \\ y_t \end{bmatrix} = \begin{bmatrix} \Lambda & 0 & \Delta_1 \\ 0 & \Lambda & \Delta_2 \\ \Omega_1 & \Omega_2 & B \end{bmatrix} \begin{bmatrix} f_{1,t} \\ f_{2,t} \\ g_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ u_t \end{bmatrix}, \quad (1)$$

for $t = 1, \dots, T$, where:

- $x_{1,t}, x_{2,t}$: vectors of HF observations ($N_H \times 1$)
- y_t : vector of LF observations ($N_L \times 1$)
- $f_{1,t}, f_{2,t}$: vectors of HF factors ($K_H \times 1$)
- g_t : vector of LF factors ($K_L \times 1$)

- A **structural VAR(1) model for the factor process:**

$$\begin{bmatrix} I_{K_H} & 0 & 0 \\ -R_H & I_{K_H} & 0 \\ 0 & 0 & I_{K_L} \end{bmatrix} \begin{bmatrix} f_{1,t} \\ f_{2,t} \\ g_t \end{bmatrix} = \begin{bmatrix} 0 & R_H & A_1 \\ 0 & 0 & A_2 \\ M_1 & M_2 & R_L \end{bmatrix} \begin{bmatrix} f_{1,t-1} \\ f_{2,t-1} \\ g_{t-1} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \\ w_t \end{bmatrix},$$

where $(v'_{1,t}, v'_{2,t}, w'_t)'$ is a multivariate white noise process with mean 0 and variance-covariance matrix:

$$\Sigma = \begin{bmatrix} \Sigma_H & 0 & \Sigma_{HL,1} \\ & \Sigma_H & \Sigma_{HL,2} \\ & & \Sigma_L \end{bmatrix}.$$

- Allows for coupling between HF and LF factor dynamics.
- We collect the parameters of the VAR process in the vector θ .

STANDARD IDENT. CONDITIONS FOR LATENT FACTOR MODELS

- orthogonality of factors: $f_{1,t}, f_{2,t} \perp g_t$;
- standardized factors:

$$E \begin{pmatrix} f_{1,t} \\ f_{2,t} \\ g_t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad V \begin{pmatrix} f_{1,t} \\ f_{2,t} \\ g_t \end{pmatrix} = \begin{bmatrix} I_{K_H} & \Phi & 0 \\ & I_{K_H} & 0 \\ & & I_{K_L} \end{bmatrix}. \quad (2)$$

NON-STANDARD IDENTIFICATION CONDITIONS

Ensure that rotation invariance of model (1) - (2) allows only for

- independent rotations among the components of $f_{1,t}$, among those of $f_{2,t}$, and among those of g_t , such that
 - rotations of $f_{1,t}$ and $f_{2,t}$ are the same.
- ⇔ Maintain the interpretation of HF and LF factors and the fact that $f_{1,t}$ and $f_{2,t}$ are consecutive observations of the same process.

The estimators of the factor values (I)

An iterative procedure:

- Let $\hat{G}^{(p-1)} = [\tilde{g}_1, \dots, \tilde{g}_T]'$ be the matrix of estimated LF factors obtained in the previous iteration;

- STEP 1: UPDATE OF THE HF ESTIMATE**

- (i) regress each sub-panel of the HF observations $x_{j,t}$ on \tilde{g}_t to obtain the estimated loadings matrices $\hat{\Delta}_1$ and $\hat{\Delta}_2$, and the residuals:

$$\hat{\xi}_{j,t} = x_{j,t} - \hat{\Delta}_j \tilde{g}_t, \quad j = 1, 2, \quad t = 1, \dots, T;$$

- (ii) collect the residuals in the $(2T \times N_H)$ matrix:

$$\hat{\Xi} = [\hat{\xi}_{1,1}, \hat{\xi}_{2,1}, \dots, \hat{\xi}_{1,T}, \hat{\xi}_{2,T}]';$$

- (iii) the estimated HF factors $\hat{F}^{(p)} = [\hat{f}_{1,1}, \hat{f}_{2,1}, \dots, \hat{f}_{1,T}, \hat{f}_{2,T}]'$ are obtained by PCA:

$$\left(\frac{1}{2N_H T} \hat{\Xi} \hat{\Xi}' \right) \hat{F}^{(p)} = \hat{F}^{(p)} V_F.$$

The estimators of the factor values (II)

● STEP 2: UPDATE OF THE LF ESTIMATE

- (i) regress the LF observations y_t on $\hat{f}_{1,t}$ and $\hat{f}_{2,t}$ to obtain the estimated loadings matrices Ω_1 and Ω_2 , and the residuals:

$$\hat{\psi}_t = y_t - \hat{\Omega}_1 \hat{f}_{1,t} - \hat{\Omega}_2 \hat{f}_{2,t}, \quad t = 1, \dots, T;$$

- (ii) collect the residuals in the $(T \times N_L)$ matrix:

$$\hat{\Psi} = [\hat{\psi}_1, \dots, \hat{\psi}_T]';$$

- (iii) the estimated LF factors $\hat{G}^{(p)} = [\hat{g}_1, \dots, \hat{g}_T]'$ are obtained by PCA:

$$\left(\frac{1}{N_L T} \hat{\Psi} \hat{\Psi}' \right) \hat{G}^{(p)} = \hat{G}^{(p)} V_G.$$

- The procedure is iterated replacing $\hat{G}^{(p-1)}$ with $\hat{G}^{(p)}$ in step 1.
- **INITIALIZATION OF ITERATIVE PROCEDURE:** PCA in step 1 assuming $\hat{G}^{(0)} = 0$, i.e. PCA on HF data.

The estimator of the factor dynamics parameters

The parameter vector θ in the structural VAR factor dynamics is estimated by:

- deriving the reduced form:

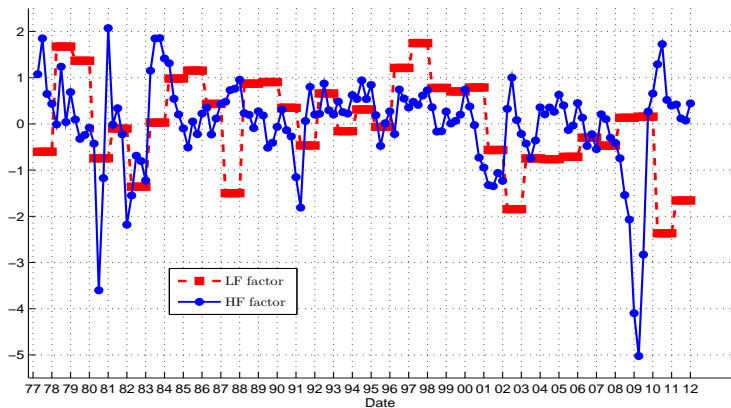
$$\begin{pmatrix} f_{1,t} \\ f_{2,t} \\ g_t \end{pmatrix} = C(\theta) \cdot \begin{pmatrix} f_{1,t-1} \\ f_{2,t-1} \\ g_{t-1} \end{pmatrix} + \zeta_t, \quad \zeta_t \sim (0, \Sigma_\zeta(\theta)),$$

- replacing $f_{1,t}$, $f_{2,t}$ and g_t by the estimates $\hat{f}_{1,t}$, $\hat{f}_{2,t}$ and \hat{g}_t for any date t ,
- performing either unconstrained LS, or constrained ML, as in *Ghysels (2012)*.

- **Asymptotic properties:** Assuming $N_H, N_L, T \rightarrow \infty$ s.t. $N_H, N_L \geq T$ and regularity conditions we show that the estimators of the factor values and parameter θ are consistent, with rate of convergence \sqrt{T} .
- **Monte Carlo analysis:** simulations from a DGP calibrated on the empirical application show good finite sample properties of the factor estimators.

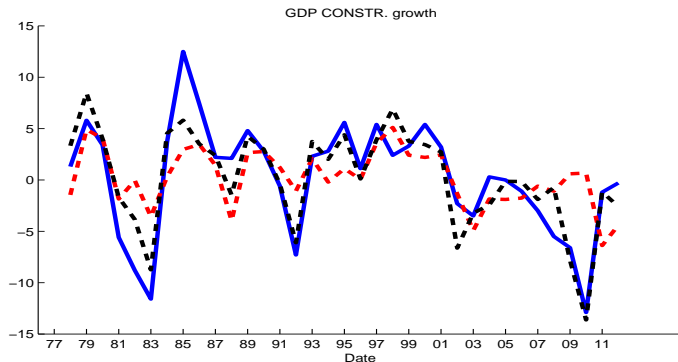
- **Is the U.S. economy driven mainly by IP shocks, or are there other relevant sources of co-movement in the other sectors?**
- **High frequency panel (X): quarterly** observations of growth rates in the sample period **1977.Q1-2011.Q4** for the *117 industrial production indexes* considered by Foester et al. (2011)
 $\Rightarrow N_H = 117$ (source: FED).
- **Low frequency panel (Y): yearly** observations of growth rates in the sample period **1977-2011** for the following 42 GDP sectors: **35 services, Construction, Farms, Forestry-Fishing and related activities, General government (federal), Government enterprises (federal), General government (states and local) and Government enterprises (states and local)**
 $\Rightarrow N_L = 42$ (source: BEA).
- *Information Criteria* from Bai and Ng (2002), computed on X , Y and different subpanels, suggest the presence of **one HF factor and one LF factor** $\Rightarrow K_H = K_L = 1$.

Empirical application: estimated HF and LF factors



- High and low frequency factors in 1977.Q1-2011.Q4, estimated from 117 quarterly IP series of Foester et al. (2011) and 42 GDP (*non-industrial production*) annual series.

Observed and fitted values of changes of GDP-Construction index



- TS of growth rate of GDP Construction index (solid blue).
- TS of fitted value from a regression of the index on the HF factor (red dashed), adjusted R-squared $\bar{R}^2 = 44\%$.
- TS of fitted value from a regression of the index on both the HF and LF factors (black dashed), $\bar{R}^2 = 74\%$.

Empirical application: results (II)

- We regress each LF and HF observed series on the estimated LF factor only, HF factor only, and both the LF and HF factors.
- For each regression we compute \bar{R}^2 .
- Quantiles of the empirical distributions of \bar{R}^2 :

Obs.	Factors	\bar{R}^2 Quantile				
		10%	25%	50%	75%	90%
LF (Y)	LF	-3.0	-2.3	4.4	11.2	22.0
LF (Y)	LF, HF	0.5	8.7	24.5	46.0	56.1
LF (Y)	HF	-4.0	3.9	15.8	33.2	45.2
HF (X)	HF	-0.0	5.3	25.8	38.3	57.0
HF (X)	HF, LF	0.7	5.9	24.8	38.2	57.1
HF (X)	LF	-2.5	-2.0	-1.2	0.4	2.2

Empirical application: results (III)

Change in \bar{R}^2 of the regressions of yearly SECTORAL GDP growth rates on estimated HF and LF factors: $\bar{R}^2(HF + LF) - \bar{R}^2(HF)$

Sector	change in \bar{R}^2	\hat{B}
Social assistance	38.89	0.59
Computer systems design and related services	37.30	0.58
General government (STATES AND LOCAL)	30.67	0.53
Construction	29.82	0.51
Government enterprises (FEDERAL)	24.52	0.47
Rental and leasing services and lessors of intangible assets	23.84	0.47
Wholesale trade	22.71	0.46
Retail trade	19.41	0.42
Management of companies and enterprises	17.10	0.41
Real estate	16.34	0.40
Miscellaneous professional, scientific, and technical services	14.15	0.37
Administrative and support services	13.97	0.36
Legal services	11.25	0.34
Information and data processing services	10.06	-0.34
...		

Empirical application: results (IV)

\bar{R}^2 of the regressions of INDEXES growth rates on estimated HF and LF factors

Sector	(1) $\bar{R}^2(HF)$	(2) $\bar{R}^2(LF)$	(3) $\bar{R}^2(HF + LF)$	(3) - (1) change in \bar{R}^2
HF observations				
Industrial Production	89.46	-0.08	90.03	0.57
LF observations (GDP)				
Aggregate GDP	60.39	20.22	85.48	25.09
Manufacturing	74.20	-0.76	75.89	1.69
Agriculture, forestry, fishing, etc.	-0.61	4.85	4.88	5.49
Construction	44.03	24.88	73.85	29.82
Wholesale trade	30.04	19.06	52.75	22.71
Retail trade	33.05	16.06	52.46	19.41
Transportation and warehousing	54.55	-1.43	54.81	0.26
Information	18.12	-2.54	15.85	-2.26
Finance, insur., real estate, etc.	6.65	21.82	31.69	25.04
Professional and business serv.	47.29	19.17	70.74	23.45
Government	-2.03	12.38	11.98	14.00

Empirical application: results (V)

We estimate by ML the constrained reduced-form VAR(1) on the factor process:

$$\begin{bmatrix} f_{1,t} \\ f_{2,t} \\ f_{3,t} \\ f_{4,t} \\ g_t \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & r_H & a \\ 0 & 0 & 0 & r_H^2 & a(1+r_H) \\ 0 & 0 & 0 & r_H^3 & a(1+r_H+r_H^2) \\ 0 & 0 & 0 & r_H^4 & a(1+r_H+r_H^2+r_H^3) \\ m_1 & m_2 & m_3 & m_4 & r_L \end{bmatrix} \begin{bmatrix} f_{1,t-1} \\ f_{2,t-1} \\ f_{3,t-1} \\ f_{4,t-1} \\ g_{t-1} \end{bmatrix} + \varepsilon_t,$$

where $\varepsilon_t \sim N(0, \Sigma)$.

Coefficient	Estimate	Std. error
r_H	0.6542	0.0651
a	-0.0268	0.0665
m_1	0.1677	0.2134
m_2	0.2821	0.3008
m_3	-0.1756	0.4968
m_4	0.2207	0.2233
r_L	0.3643	0.1438
σ_H	0.7498	0.1283
σ_L	0.8163	0.2743
ρ_{HL}	0.0055	0.0962

Empirical application: structural factor analysis (I)

- **Issue:** Sectoral linkages lead to propagation of sector-specific shocks
- **Solution:** Multi-industry real business cycle model FSW(2011):
 - (i) Perfectly competitive industries produce N goods using: **capital, labor and intermediate inputs** \Rightarrow capital and intermediate inputs are input for some industries and output of others.
 - (ii) Productivity of each sector follows a random walk with innovations $\varepsilon_t =$ **industry specific (technological) shocks purged from interlinkages due to materials and capital use.**
 - (iii) Model yields VARMA(1,1) model for **SECTORAL OUTPUT GROWTH** $X_t = [\Delta \ln Y_{1t}, \dots, \Delta \ln Y_{Nt}]$:

$$(I - \Phi L)X_t = (\Pi_0 + \Pi_1 L)\varepsilon_t,$$

where Φ , Π_0 and Π_1 depend on **Input-Output matrices, Capital Flows matrices** and other model parameters.

- **OUR CONTRIBUTION:** we build on FSW (2011) to extract **technological shocks in mixed-frequency settings** in order to **include also non-IP output growth** in the structural model.
- LF panel (Y): **yearly** observations of growth rates of **GROSS OUTPUT GROWTH** for the 38 GDP sectors: *35 services, Construction, Farms, Forestry-Fishing* (**1988-2011**, BEA) $\Rightarrow N_L = 38, T = 24$.
- HF panel (X): **quarterly** observations of growth rates of the *117 IP indexes* (**sample: 1988.Q1-2011.Q4**) $\Rightarrow N_H = 117$.
- **Input-output tables** for the IP and non-IP sectors (1997, BEA).
- **Capital flows tables** for the IP and non-IP sectors (1997, BEA).
- Values of the other parameters of the general equilibrium model are analogous to FSW (2011).

Empirical application: structural factor analysis (III)

- **IDEA:** use the algorithms in King and Watson (2002) to obtain the estimates of the HF technological shocks $\hat{\varepsilon}_t^{X,q}$ at each quarter ($q = 1, 2, 3, 4$):

$$\begin{bmatrix} \hat{\varepsilon}_t^{X,q} \\ \hat{\varepsilon}_t^Y \end{bmatrix} = (\hat{\Pi}_0 + \hat{\Pi}_1 L)^{-1} (I - \hat{\Phi} L) \begin{bmatrix} X_t^q \\ Y_t \end{bmatrix},$$

where X_t^q is the panel of quarterly growth rates of IP, for quarter q only.

- The LF technological shocks $\hat{\varepsilon}_t^Y$ are obtained in the same way using yearly growth rates of IP indexes instead of X_t^q .
- **Estimate a mixed frequency factor model** with 1 HF and 1 LF factors on the panels of mixed frequency technological shocks $\hat{\varepsilon}_t^X$ and $\hat{\varepsilon}_t^Y$.
- Check the explanatory power of these common mixed-frequency factors for aggregate output growth and other aggregate indexes.

Empirical application: structural factor analysis (IV)

\bar{R}^2 of the regressions of INDEXES growth rates on estimated 1 HF and 1 LF factors from technological shocks. HF data: IP indexes, LF data: non-IP real GROSS OUTPUT, sample: 1989-2011.

Sector	$\bar{R}^2(HF)$	$\bar{R}^2(LF)$	$\bar{R}^2(HF + LF)$	change in \bar{R}^2
HF observations				
Industrial Production	70.52	9.18	80.60	10.08
LF observations (GROSS OUTPUT)				
GO (all sectors)	48.57	27.61	85.09	36.51
Manufacturing	70.00	16.02	93.50	23.50
Agriculture, forestry, fishing, etc.	-11.65	-3.17	-16.24	-4.58
Construction	19.12	20.95	45.86	26.74
Wholesale trade	78.27	6.95	91.45	13.18
Retail trade	73.40	4.97	83.86	10.46
Transportation and warehousing	68.17	7.51	81.14	12.97
Information	4.57	59.81	78.09	73.51
Finance, insur., real estate, etc.	6.60	13.04	22.88	16.28
Professional and business services	37.74	36.51	84.23	46.49
Educ. serv., health care, etc.	12.33	2.55	16.10	3.76

Empirical application: structural factor analysis (V)

\bar{R}^2 of the regressions of INDEXES growth rates on estimated 1 HF and 1 LF **factors from gross output indexes**. HF data: IP indexes, LF data: non-IP real GROSS OUTPUT, sample: 1988-2011.

Sector	$\bar{R}^2(HF)$	$\bar{R}^2(LF)$	$\bar{R}^2(HF + LF)$	change in \bar{R}^2
HF observations				
Industrial Production	89.28	-0.22	90.01	0.73
LF observations (GROSS OUTPUT)				
GO (all sectors)	70.75	16.73	94.89	24.14
Manufacturing	90.22	-0.38	94.68	4.47
Agriculture, forestry, fishing, etc.	-8.85	0.12	-9.13	-0.28
Construction	28.13	29.35	65.29	37.16
Wholesale trade	85.18	-3.56	85.52	0.34
Retail trade	81.68	-2.75	82.83	1.15
Transportation and warehousing	75.42	3.54	83.81	8.39
Information	25.46	28.87	61.91	36.46
Finance, insur., real estate, etc.	17.72	22.89	46.49	28.78
Professional and business services	45.12	33.85	88.60	43.48
Educ. serv., health care, etc.	3.05	3.28	7.15	4.10

Work in progress (I)

- (i) Selection of the numbers of HF and LF factors.
- (ii) *Main Street vs. Wall Street: a MFFM analysis [AGGR(2014)]*:
 - **Same methodology**, but focus on **macro-finance application**.
 - HF data: **173 monthly US financial series**, mainly equity portfolios returns, for the sample: **1960.1 - 2012.12**.
 - LF data: **212 quarterly US macroeconomic indicators**, mainly those considered by SW (2012), Juardo, Luidvigson and Ng (2012).
 - **Multiple HF (m) financial factors and LF (q) macro factors**
 - **Financial factors (HF)**: $1^{st} \approx$ excess market returns, 2^{nd} to 4^{th} correlate with the Fama and French factors and industry portfolios (Oil, Coal, Utils,...), but feature some significant differences.
 - **Macro factors (LF)**: 1^{st} to 4^{th} correlate with broad real economic activity, Housing Starts and Price indexes.

(ii) *Main Street vs. Wall Street: a MFFM analysis [cont'd]:*

- Monthly financial factors have higher explanatory power than *Fama and French* factors in predicting LF GDP or IP growth rates:
↑ 50% of \bar{R}^2 if added in a predictive regression to LF factors or lagged real output growth.
- Relationship between the HF and LF factors looks significantly different before and after the mid 1980s.

Concluding remarks

- We study large-scale mixed frequency latent factor models with orthogonal “fast” and “slow” factors.
- The unobservable factor space is identifiable disentangling factors at different frequencies.
- We develop consistent estimators of the factor paths (by iterative PCA methods) and of the factor dynamics.
- Application to our panels of growth rates of U.S. industrial and non-industrial sectors shows:
 - (i) 1 LF factor at yearly frequency has explanatory power for non-industrial sectors.
 - (ii) 1 HF factor at quarterly frequency has explanatory power for both industrial and non-industrial production sectors.
 - (iii) Results are still valid after removing comovements from input-output and capital use linkages.