A Bayesian Multivariate Functional Dynamic Linear Model

David Rupppert
Department of Statistical Science and School of OR&IE
Cornell University

June 29, 2015

Joint work with:

Daniel Kowal

David Matteson

Outline

- Multiple functional time series.
- Two examples.
  - Yield curves in four economies
  - Local field potentials measured in rats' brains
- Bayesian dynamic factor model.
  - Constrained spline modeling of factor loading curves
  - Markov switching model for the effects of changes in US yield curves on the yield curves of the other economies
  - Stochastic volatility
  - Time-frequency analysis of the local field potentials.

In my experience, a Bayesian analysis is the best approach to a complex problem with many latent variables.

- I am not a Bayesian, but I use Bayesian methodology.

– Tom Loredo, astrostatistician

Assume that we observe multiple time series of functional data:

$$Y_t^{(c)}(\tau).$$

(a) For each c and t, $Y_t^{(c)}(\tau)$ is a *function* of $\tau \in \mathcal{T}$;

(b) For each c and τ, $Y_t^{(c)}(\tau)$ is a *time series* for $t = 1, \ldots, T$; and

(c) For each t and τ, $Y_t^{(c)}(\tau)$ is a *multivariate* observation with outcomes $c = 1, \ldots, C$.

- $\mathcal{T} \subseteq \mathbb{R}^d$ is a compact set.
- $d = 1$ and $\mathcal{T} = [0, 1]$ are typical.

First example:

$Y_t^{(c)}(\tau)$ is the yield curve (function of maturity τ) in tth week and cth economy.

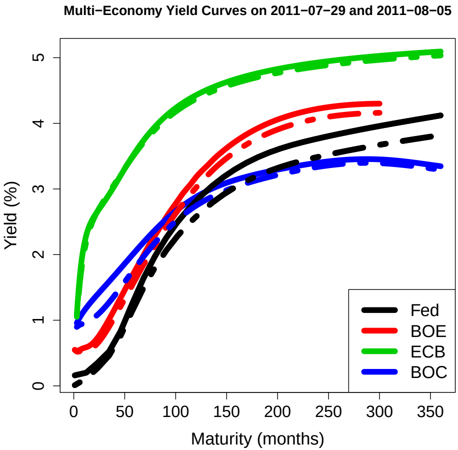
- 1 unit of currency invested in week t will grow to $\exp\{(\tau/12)\, Y_t^{(c)}(\tau)\}$ units in τ months.
- Therefore, $Y_t^{(c)}(\tau)$ is the average yearly interest rate that is earned on a zero-coupon bond maturing in τ months and purchased during the tth week.

We have yield curves from

1. Federal Reserve (Fed)
2. Bank of England (BOE)
3. European Central Bank (ECB)
4. Bank of Canada (BOC)

So $C = 4$.

Yield Curves on Two Dates



Solid curves: 2011-07-29. Dashed curves: 2011-08-05

Example: time-frequency analysis

- Suppose we observe C times series.
- Bin the time dimension into T bins.
- Within each bin, and for each time series, compute the Fourier transform to obtain a function of frequency τ.
  - t is the bin index
  - $Y_t^{(c)}(\tau)$ is the smoothed Fourier transform (spectrum) of the cth series in the tth bin.

Our example: local field potential (LFP) data collected on rats at two brain locations.

- A pair of electrodes is placed in each brain region.
- In feature binding, a rat's brain must amalgamate distinct sensory information into a single neural representation.
- There is interest in the synchronization between the regions, in particular, whether
  - synchronization is increased during feature binding.
- $c = 1$ or 2 indicates brain region.
- t is the bin index.
- τ is frequency.

Define

$$\mathbf{Y}_t(\tau) = \left[Y_t^{(1)}(\tau),\, Y_t^{(2)}(\tau), \ldots,\, Y_t^{(C)}(\tau)\right]'.$$

The multivariate functional dynamic linear model (MFDLM) is:

$$\begin{cases} \mathbf{Y}_t(\tau) = \mathbf{F}(\tau)\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t(\tau), & \left[\boldsymbol{\epsilon}_t(\tau)\middle|\mathbf{E}_t\right] \overset{indep}{\sim} N\left(\mathbf{0}, \mathbf{E}_t\right) \\ \boldsymbol{\beta}_t = \mathbf{X}_t\boldsymbol{\theta}_t + \boldsymbol{\nu}_t, & \left[\boldsymbol{\nu}_t\middle|\mathbf{V}_t\right] \overset{indep}{\sim} N(\mathbf{0}, \mathbf{V}_t) \\ \boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, & \left[\boldsymbol{\omega}_t\middle|\mathbf{W}_t\right] \overset{indep}{\sim} N(\mathbf{0}, \mathbf{W}_t) \\ \int_{\mathcal{T}} \mathbf{F}'(\tau)\mathbf{F}(\tau)\ d\tau = \mathbf{I}_{KC\times KC} & \text{for identifiability.} \end{cases}$$

We will look more closely at each of the four equations in this model.

The first equation is:

$$\underbrace{\mathbf{Y}_t(\tau)}_{C\times 1} = \underbrace{\mathbf{F}(\tau)}_{C\times KC} \underbrace{\boldsymbol{\beta}_t}_{KC\times 1} + \boldsymbol{\epsilon}_t(\tau), \qquad \left[\boldsymbol{\epsilon}_t(\tau) \middle| \mathbf{E}_t\right] \overset{indep}{\sim} N\left(\mathbf{0}, \mathbf{E}_t\right)$$

- The KC columns of $\mathbf{F}(\tau)$ are factor loading curves, K for each outcome.
- $\boldsymbol{\beta}_t$ contains KC factors, K for each outcome.

$$\mathbf{F}(\tau) = \left\{\begin{matrix} \underbrace{\mathbf{F}^{(1)}(\tau)}_{1\times K} & 0 & \cdots & 0 \\ 0 & \mathbf{F}^{(2)}(\tau) & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \mathbf{F}^{(C)}(\tau) \end{matrix}\right\} \text{ and } \boldsymbol{\beta}_t = \begin{pmatrix} \underbrace{\boldsymbol{\beta}_t^{(1)}}_{K\times 1} \\ \vdots \\ \boldsymbol{\beta}_t^{(C)} \end{pmatrix}$$

From the previous frame:

$$\underbrace{\mathbf{Y}_t(\tau)}_{C\times 1} = \underbrace{\mathbf{F}(\tau)}_{C\times KC} \underbrace{\boldsymbol{\beta}_t}_{KC\times 1} + \boldsymbol{\epsilon}_t(\tau), \qquad \left[\boldsymbol{\epsilon}_t(\tau) \middle| \mathbf{E}_t\right] \overset{indep}{\sim} N\left(\mathbf{0}, \mathbf{E}_t\right)$$

The covariance matrix $\mathbf{E}_t$ could be

- constant (independent of t), or
- might follow a stochastic volatility model.

The second equation is:

$$\underbrace{\boldsymbol{\beta}_t}_{KC\times 1} = \underbrace{\mathbf{X}_t}_{KC\times p} \underbrace{\boldsymbol{\theta}_t}_{p\times 1} + \boldsymbol{\nu}_t, \qquad [\boldsymbol{\nu}_t|\mathbf{V}_t] \overset{indep}{\sim} N(\mathbf{0}, \mathbf{V}_t)$$

- $\mathbf{X}_t$ is a matrix of known covariates that might affect $\boldsymbol{\beta}_t$.
- $\boldsymbol{\theta}_t$ contains the p regression coefficients.
- Again, the covariance matrix $\mathbf{V}_t$ could be constant.

Dynamic Model for Regression Coefficients

The third equation is what makes the model dynamic and is:

$$\underbrace{\boldsymbol{\theta}_t}_{p\times 1} = \underbrace{\mathbf{G}_t}_{p\times p}\underbrace{\boldsymbol{\theta}_{t-1}}_{p\times 1} + \boldsymbol{\omega}_t, \qquad [\boldsymbol{\omega}_t|\mathbf{W}_t] \overset{indep}{\sim} N(\mathbf{0}, \mathbf{W}_t).$$

$\mathbf{G}_t$ is an evolution matrix.

- $\mathbf{G}_t$ might be independent of t or could follow its own dynamic model.
- If $\mathbf{G}_t \equiv \mathbf{G}$, then θ_t is a VAR(1) process.

The fourth equation is:

$$\int_{\mathcal{T}} \mathbf{F}'(\tau)\mathbf{F}(\tau)\ d\tau = \mathbf{I}_{KC \times KC} \qquad \text{for identifiability.}$$

This is a typical constraint used in factor models.

- For example, in PCA.

We model the factor loading curves with splines:

$$f_k(\tau) = \boldsymbol{\phi}'(\tau)\boldsymbol{d}_k$$

- $f_k(\tau)$ is the loading curve for the kth factor and for all outcomes.
    - for simplicity, in our examples we assume the same loading curves across all outcomes
- $\boldsymbol{\phi}(\tau)$ is a vector of spline basis functions.
- $\boldsymbol{d}_k$ is a vector of spline coefficients for the kth loading curve.

We use a penalty

$$\mathcal{P}(\boldsymbol{d}_k) = \boldsymbol{d}_k' \boldsymbol{\Omega}_\phi \boldsymbol{d}_k = \int_{\tau \in \mathcal{T}} \left[\ddot{f}_k(\tau) \right]^2 \ d\tau \tag{1}$$

In our Bayesian approach, the penalty is implemented by an O'Sullivan spline basis (Wand and Ormerod, 2008) .

- The first two basis functions are 1 (constant) and x.
- The remaining basis functions are orthogonal to the first two. (Nonlinear)
- Penalty (1) is equal to the sum of squared coefficients of the nonlinear basis functions.
- The prior on $\boldsymbol{d}_k$ is $N(0, \mathrm{diag}(10^6, 10^6, \lambda_k^{-1}, \ldots, \lambda_k^{-1}))$.

Selecting a Penalty Parameter

Three common methods for selecting tuning parameter:

1. CV, GCV, AIC, or Mallow C_p.
2. Mixed model/REML (empirical Bayes).
3. Fully Bayes.

- There was been considerable research comparing 1. and 2. starting with Wahba (1985).
  - See Krivobokov (2013) for a review and recent results.
- The tuning parameter selected by 1. is more variable and more likely to undersmooth compared to 2. or 3.
- Approach 3. accounts for uncertainty in the smoothing parameter when making inference about other parameters.

Fitting the MFDLM

We used a rather standard approach to a Bayesian analysis:

- Noninformative priors mostly, but
  - hierarchical priors for the splines
- Gibbs sampling, with
  - Metropolis steps where needed.

There is an extensive literature on yield curve modeling:

- Nelson-Siegel (1987): three-parameter parametric model
- Svensson (1994): four-parameter extension of Nelson-Siegel model
- Hays et al. (2012): functional dynamic model—for single economies

These will be discussed in more detail later.

Yield Curves: Advantages of MFDLM

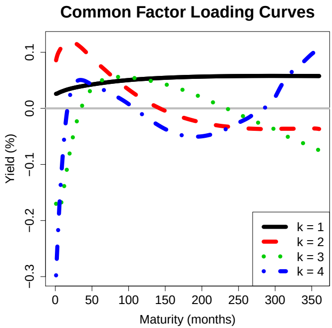
With the multivariate functional dynamic linear model it is easy to:

- handle multiple economies, especially interactions between them,
- add a hidden Markov model for regime-switching,
- add covariates, e.g., indicators of changes in government policies,
- allow conditional heteroscedasticity, e.g., with stochastic volatility models.

We assume common functional loading curves, $f_1, \ldots, f_K$, so that the conditional expectation of the yield curve for economy c, conditional on past information, is

$$\mu_t^{(c)}(\tau) \equiv \sum_{k=1}^{K} \beta_{k,t}^{(c)} f_k(\tau)$$

Estimates of Common Loading Curves



- $k = 1$: parallel shifts
- $k = 1$: changes in slope
- $k = 3$: changes in convexity (curvature)

Recall the general form of our model:

$$\mathbf{Y}_t(\tau) = \mathbf{F}(\tau)\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t(\tau), \quad \left[\boldsymbol{\epsilon}_t(\tau)\middle|\mathbf{E}_t\right] \overset{indep}{\sim} N\left(\mathbf{0}, \mathbf{E}_t\right)$$

$$\boldsymbol{\beta}_t = \mathbf{X}_t\boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \quad \left[\boldsymbol{\nu}_t\middle|\mathbf{V}_t\right] \overset{indep}{\sim} N(\mathbf{0}, \mathbf{V}_t)$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \left[\boldsymbol{\omega}_t\middle|\mathbf{W}_t\right] \overset{indep}{\sim} N(\mathbf{0}, \mathbf{W}_t)$$

$$\int_{\mathcal{T}} \mathbf{F}'(\tau)\mathbf{F}(\tau)\ d\tau = \mathbf{I}_{KC \times KC} \qquad \text{for identifiability.}$$

We will eliminate $\boldsymbol{\theta}_t$ by assuming that $\mathbf{X}_t$ is the identity matrix and $\mathbf{V}_t = \mathbf{0}$ so that $\boldsymbol{\beta}_t = \boldsymbol{\theta}_t$.

$c = 1$ is the Fed. We will investigate how this dominant economy affects the other three economies. Our model is

$$\Delta\beta_{k,t}^{(1)} = \omega_{k,t}^{(1)} \tag{1}$$

$$\Delta\beta_{k,t}^{(c)} = s_{k,t}^{(c)}(\gamma_k^{(c)}\Delta\beta_{k,t}^{(1)}) + \omega_{k,t}^{(c)} \quad c = 2,\ldots,C \tag{2}$$

In equation (1):

- Δ is the differencing operator
- The $\omega_{k,t}^{(c)}$ are AR(r) processes
  - so $\beta_{k,t}^{(1)}$ is an ARIMA(1,1,0) process

From the previous frame:

$$\Delta\beta_{k,t}^{(1)} = \omega_{k,t}^{(1)} \tag{1}$$

$$\Delta\beta_{k,t}^{(c)} = s_{k,t}^{(c)}(\gamma_k^{(c)}\Delta\beta_{k,t}^{(1)}) + \omega_{k,t}^{(c)} \quad c = 2, \dots, C \tag{2}$$

In the equation (2):

- $\left\{s_{k,t}^{(c)} : t = 1, \dots, T\right\}$ is a discrete Markov chain with states $\{0, 1\}$.
- $s_{k,t}^{(c)} = 1$ implies that at time t, the kth factor for the cth economy is affected by the kth factor of the US economy.
- $\gamma_k^{(c)} \in \mathbb{R}$ is the economy-specific slope term for the kth factor.

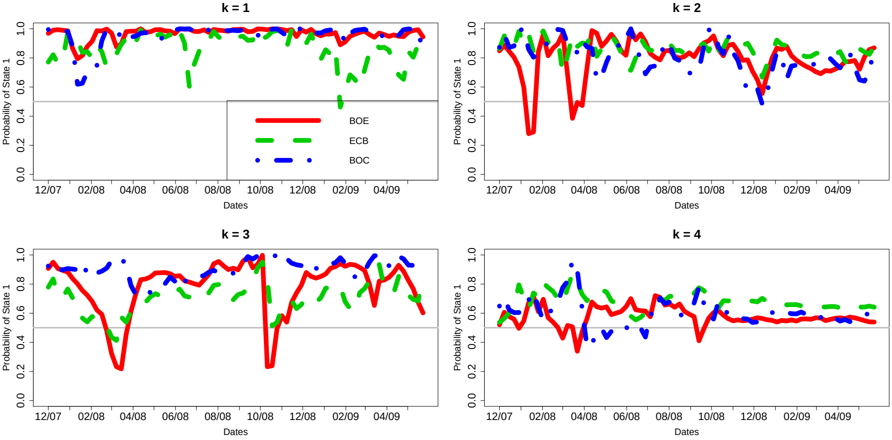
From the previous frame:

$$\Delta\beta_{k,t}^{(1)} = \omega_{k,t}^{(1)} \tag{3}$$

$$\Delta\beta_{k,t}^{(c)} = s_{k,t}^{(c)}(\gamma_k^{(c)}\Delta\beta_{k,t}^{(1)}) + \omega_{k,t}^{(c)} \quad c = 2, \ldots, C \tag{4}$$

- This is the autoregressive regime switching model of Albert and Chib (1993) and McCulloch and Tsay (1993).
- Define $q_{ii',k}^{(c)}$ to be the probability of switching from state i to state i', $i, i' \in \{0, 1\}$.

Posterior estimates of $P(s_{k,t}^{(c)} = 1)$.



We see that changes in the first factor for the US economy are associated with similar changes in the first factor for the other three economies, especially BOE and BOC.

Recall the dynamic model:

$$\Delta\beta_{k,t}^{(c)} = s_{k,t}^{(c)}(\gamma_k^{(c)}\Delta\beta_{k,t}^{(1)}) + \omega_{k,t}^{(c)},\ \ c = 2, \ldots, C.$$

We modeled the noise $\omega_{k,t}^{(c)}$ as AR(r) with stochastic volatility:

$$\omega_{k,t}^{(c)} = \sum_{i=1}^{r} \psi_{k,i}^{(c)}\omega_{k,t-i}^{(c)} + \sigma_{k,(c),t}\, z_{k,t}^{(c)}$$

where, following Kim et al. (1998) and Chib et al. (2002),

- $\log(\sigma_{k,(c),t})$ is an AR(1) process.

Other Factor Models: Jungbacker et al.

Jungbacker et al. (2014) also use splines for the factor loading curves

- They use Wald, Lagrange multiplier, and likelihood ratio tests to reduce the number of knots.
- They work with only the US economy.
- They assume constant conditional volatility.

Other Factor Models: Nelson-Siegel Model

The Nelson-Siegel (1987) model for the yield curve is:

$$Y(\tau) = \theta_0 + (\theta_1 + \theta_2\tau)\exp(-\theta_3\tau).$$

- This model is widely used in the banking industry.
- Empirically, it does not fit as well as spline models, which are also widely used.

Svensson's (1994) model adds $\theta_4\tau\exp(-\theta_5\tau)$ to the Nelson-Siegel model.

- This improves the fit, but still not as much as using splines.

Nelson and Siegel formulated their model as a static one.

- Diebold and Li (2006) reformulated the Nelson-Siegel model in dynamic form.
- λ determines the location of the maximum of the third factor loading curve.
  - Diebold and Li fixed λ in advance so that the location of this maximum was sensible.
- With λ fixed, the Nelson-Siegel model is linear.

Nelson-Siegel Model in Factor Form

The Nelson-Siegel model can be expressed as a dynamic model with three parametric factor loading curves:

$$1 \text{ (constant)}$$

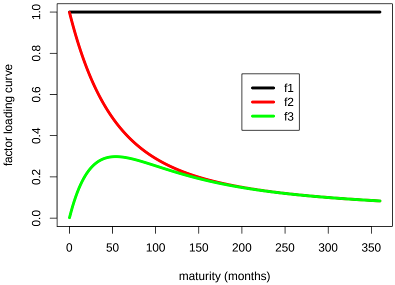
$$\frac{1 - \exp(-\lambda\tau)}{\lambda t}$$

$$\frac{1 - \exp(-\lambda\tau)}{\lambda\tau} - \exp(-\lambda\tau).$$

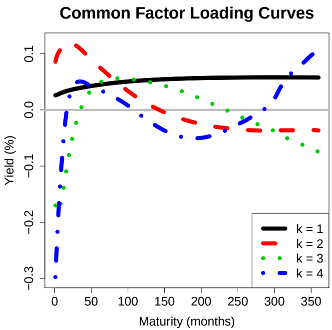
Here $\lambda = \theta_3$ is fixed and the factors are linear combinations of θ_0, θ_1, and θ_2.

Recall: The Nelson-Siegel model is $\theta_0 + (\theta_1 + \theta_2\tau)\exp(-\theta_3\tau)$.

Empircal Versus Nelson-Siegel Factor Loading Curves



Nelson-Siegel with
$\lambda = 0.4/12$



Empirical from the MFDLM

Hays et al. (2012) developed a functional dynamic factor model and applied it to yield curve forecasting.

- We view our work at least somewhat as a generalization of theirs.
- They worked with one economy.
- Their analysis is non-Bayesian.
  - A stochastic EM-algorithm was used.
  - Smoothing parameters were fit by GCV.
  - Stochastic volatility and covariates seem more difficult to add compared to within a Bayesian analysis.
- They found that their model outperformed a dynamic Nelson-Siegel in terms of forecasting accuracy.

Local Field Potentials

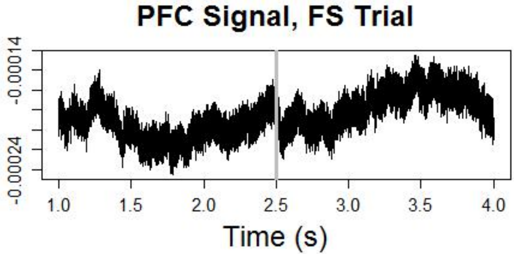
- Neural activity in a region of the brain can be detected by recording from a pair of electrode in that region.
  - The electrodes measure the local field potential (LFP) in the region.
- Usually done with rodents.
  - Performed on humans only rarely and only for medical reasons.
- LFP data are similar to data obtained from EEG where electrodes are placed on the scalp.
  - LPF signals are more localized than EEG signals.

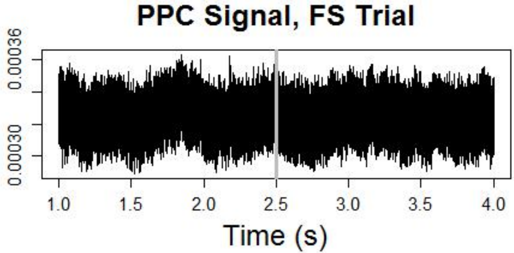
- LFP data are usually analyzed in the frequency domain.
- Our data were collected by PhD student Vladmir Ljubojevic working with Professor Eve De Rosa at U. Toronto.
  - They are now both at Cornell where Vlad is a postdoc.
- Pairs of electrodes where placed in two locations:
  - prefrontal cortex (PFC)
  - posterior parietal cortex (PPC)
- These two regions have been proposed as the neural substrate of attention.
- Sampling was at 4000 Hertz.

- There were two experimental conditions:
  - feature singleton (FS) — the rat processes a single stimulus.
  - feature conjunction (FC) — the rat must integrate two stimuli.
- In the experiments, the rat attempts to select the one of two bowls with a treat.
  - In FS, the correct bowl is indicated by a single smell.
  - In FC, the correct bowl is indicated by two smells.
- FS serves as a baseline.
- We are interested in synchronization of the two brain regions during FC.

- Tasks are repeated for 20 trials per rat for each of FS and FC.
- For each trial, we look at a 3 sec window centered at the time the rat makes a decision.
- The data are binned into 15 time bins.
  - Each is of length 3/8 sec.
  - 50% overlap.

LFP Signals from a Single Trial





- The periodgram (discrete Fourier transform) of the bivariate time series is computed in each bin.
- The periodgrams are smoothed with a modified Daniell kernel to obtain spectra.
- $\tilde{I}_t^{(c)}(\tau)$ is
  - the spectrum of PFC for $c = 1$,
  - the spectrum of PPC for $c = 2$,
  - the cross-spectrum for $c = 3$.
- The squared coherence is

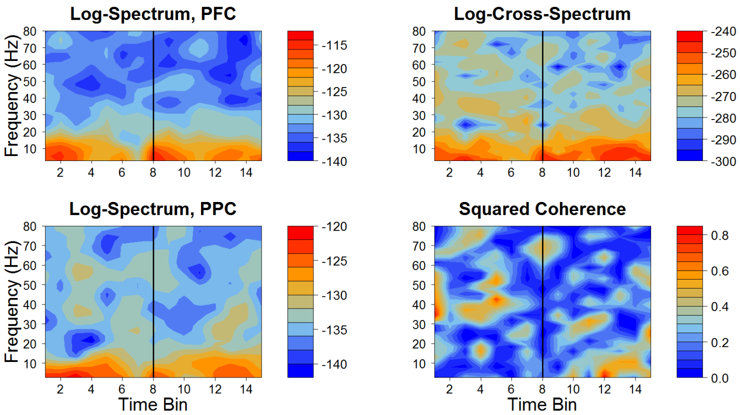
$$\kappa_t^2(\tau) \equiv \frac{|\tilde{I}_t^{(3)}(\tau)|^2}{\tilde{I}_t^{(1)}(\tau)\tilde{I}_t^{(2)}(\tau)}$$

- $Y_t^{(c)}(\tau) = \log\left(\tilde{I}_t^{(c)}(\tau)\right)$ for $c = 1, 2$.
- $0 \leq \kappa_t^2(\tau) \leq 1$
- Let $Y_t^{(3)}(\tau) = \Phi^{-1}(\kappa_t^2(\tau))$
    - The link function Φ^{-1} is the standard normal quantile function.

The indices are:

- $t = 1, \ldots, 15$.
- $c = 1, 2, 3$.
- τ ranges from 0.1 to 80 Hertz.
    - 30 observation points after truncation above 80 Hertz.
    - There were about 750 frequencies, but most were above 80 Hertz and not of interest.

Time-frequency Plots for FS Trials



We assume

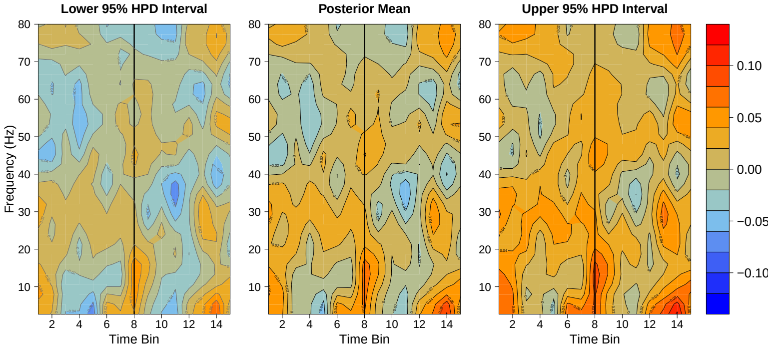
- common factor loading curves, and
- a random walk model for the factors.

$$
\begin{aligned}
Y_{i,s,t}^{(c)}(\tau) &= \sum_{k=1}^{K} \beta_{k,i,s,t}^{(c)} f_k(\tau) + \epsilon_{i,s,t}^{(c)}(\tau), \quad \left[\epsilon_{i,s,t}^{(c)}(\tau) \middle| \sigma_{(c)}^2\right] \overset{indep}{\sim} N(0, \sigma_{(c)}^2) \\
\boldsymbol{\beta}_{k,i,s,t} &= \boldsymbol{\beta}_{k,i,s,t-1} + \boldsymbol{\omega}_{k,i,s,t}, \quad \left[\boldsymbol{\omega}_{k,i,s,t} \middle| \mathbf{W}_k\right] \overset{indep}{\sim} N(\mathbf{0}, \mathbf{W}_k)
\end{aligned}
$$

- $i = 1, \ldots, 8$ is the rat index.
- $s = 1, \ldots, 40$ is the index of the trial within a rat.
- $t = 1, \ldots, 15$ is the index of the time bins.
- $c = 1, 2, 3$ is the outcome index.
- $k = 1, \ldots, K$ is factor index.

- DIC selected $K = 10$ factors.
- $\mu^{(c)}_{i,s,t}(\tau) \equiv \sum_{k=1}^{10} \beta^{(c)}_{k,i,s,t} f_k(\tau)$ is the estimated mean for the ith rat, sth trial within that rat, and tth time bin.
- We focused on the difference in squared coherence between FC and FS.

Posterior and 95% HPD Confidence Intervals for Difference in Squared Coherene



FB is associated with increased coherence in the Theta range (4–8 Hz), Alpha range (8–13 Hz), and Beta range (13–30 Hz).

Albert, J. H. and Chib, S. (1993).
Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts.
Journal of Business & Economic Statistics, 11(1):1–15.

Chib, S., Nardari, F., and Shephard, N. (2002).
Markov chain Monte Carlo methods for stochastic volatility models.
Journal of Econometrics, 108(2):281–316.

Diebold, F. X. and Li, C. (2006).
Forecasting the term structure of government bond yields.
Journal of Econometrics, 130(2):337–364.

Hays, S., Shen, H., and Huang, J. Z. (2012).
Functional dynamic factor models with application to yield curve forecasting.
The Annals of Applied Statistics, 6(3):870–894.

Jungbacker, B., Koopman, S. J., and van der Wel, M. (2013).
Smooth dynamic factor analysis with application to the US term structure of interest rates.
Journal of Applied Econometrics.

Kim, S., Shephard, N., and Chib, S. (1998).
Stochastic volatility: likelihood inference and comparison with ARCH models.
The Review of Economic Studies, 65(3):361–393.

Krivobokova, T. (2013)
Smoothing parameter selection in two frameworks for penalized splines
Journal of the Royal Statistical Society, Series B, 75, 725–741.

McCulloch, R. E. and Tsay, R. S. (1993).
Bayesian inference and prediction for mean and variance shifts in autoregressive time series.
Journal of the American Statistical Association, 88(423):968–978.

References

Nelson, C. R. and Siegel, A. F. (1987).
Parsimonious modeling of yield curves.
Journal of Business, 60(4):473

Svensson, L. E. (1994).
Estimating and interpreting forward interest rates: Sweden 1992-1994.
Technical report, National Bureau of Economic Research.

Wahba, G. (1985)
A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem.
Annals of Statistics, 13, 1378–1402.

Wand, M. P. and Ormerod, J. (2008).
On semiparametric regression with O'Sullivan penalized splines.
Australian and New England Journal of Statistics, 50:179–198.