

Supplement to “A Bayesian Multivariate Functional Dynamic Linear Model”

Daniel R. Kowal, David S. Matteson, and David Ruppert

July 17, 2015

Abstract

This supplement contains the initialization procedure and MCMC sampling algorithm for the proposed model, provides MCMC diagnostics for the applications, presents additional details and extensions of the common trend model of Section 4.1.1, and shows additional figures relevant to the applications.

Recall the proposed *Multivariate Functional Dynamic Linear Model* (MFDLM):

$$\begin{cases} \mathbf{Y}_t(\tau) = \mathbf{F}(\tau)\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t(\tau), & [\boldsymbol{\epsilon}_t(\tau)|\mathbf{E}_t] \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{E}_t) \\ \boldsymbol{\beta}_t = \mathbf{X}_t\boldsymbol{\theta}_t + \boldsymbol{\nu}_t, & [\boldsymbol{\nu}_t|\mathbf{V}_t] \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{V}_t) \\ \boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, & [\boldsymbol{\omega}_t|\mathbf{W}_t] \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{W}_t) \end{cases} \quad (1)$$

for which all of the variables have been defined in the main paper.

To sample from the joint posterior distribution, we use a Gibbs sampler. Because the Gibbs sampler allows blocks of parameters to be conditioned on all other blocks of parameters, it is a convenient approach for our model. First, hierarchical dynamic linear model (DLM) algorithms typically require that $\boldsymbol{\beta}_t$ and $\boldsymbol{\theta}_t$ be the only unknown components, which we can accommodate by conditioning appropriately. Second, our orthonormality approach for $f_k^{(c)}$ fits nicely within a Gibbs sampler, and we can adapt the algorithms described in Wand and Ormerod (2008). And third, the hierarchical structure of our model imposes natu-

ral conditional independence assumptions, which allows us to easily partition the parameters into appropriate blocks.

A Initialization

To initialize the factors $\boldsymbol{\beta}_k^{(c)} = (\beta_{k,1}^{(c)}, \dots, \beta_{k,T}^{(c)})'$ and the factor loading curves (FLCs) $f_k^{(c)}$ for $k = 1, \dots, K$ and $c = 1, \dots, C$, we compute the singular value decomposition (SVD) of the data matrix $\mathbf{Y}^{(c)} = \mathbf{U}^{(c)} \boldsymbol{\Sigma}^{(c)} \mathbf{V}^{(c)'}'$ for $c = 1, \dots, C$. Note that to obtain a data *matrix* $\mathbf{Y}^{(c)}$, with rows corresponding to times t and columns to observations points τ , we need to estimate $Y_t^{(c)}(\tau)$ for any unobserved τ at each time t , which may be computed quickly using splines. However, these estimated data values are *only* used for the initialization step. Letting $\mathbf{U}_{1:K}^{(c)}$ be the first K columns of $\mathbf{U}^{(c)}$, $\boldsymbol{\Sigma}_{1:K}^{(c)}$ be the upper left $K \times K$ submatrix of $\boldsymbol{\Sigma}^{(c)}$, and $\mathbf{V}_{1:K}^{(c)}$ be the first K columns of $\mathbf{V}^{(c)}$, we initialize the factors $(\boldsymbol{\beta}_1^{(c)}, \dots, \boldsymbol{\beta}_K^{(c)}) = \mathbf{U}_{1:K}^{(c)} \boldsymbol{\Sigma}_{1:K}^{(c)}$ and the FLCs $(\mathbf{f}_1^{(c)}, \dots, \mathbf{f}_K^{(c)}) = \mathbf{V}_{1:K}^{(c)}$, where $\mathbf{f}_k^{(c)}$ is the vector of FLC k evaluated at all observation points $\cup_t \mathcal{T}_t^{(c)}$ for outcome c . The $\mathbf{f}_k^{(c)}$ are orthonormal in the sense that $\mathbf{f}_k^{(c)'} \mathbf{f}_j^{(c)} = \mathbf{1}(k = j)$, but they are not smooth. This approach is similar to the initializations in Matteson et al. (2011) and Hays et al. (2012).

Given the factors $\boldsymbol{\beta}_k^{(c)}$ and the FLCs $\mathbf{f}_k^{(c)}$, we can estimate each $\sigma_{(c)}^2$ (or more generally, \mathbf{E}_t) using conditional maximum likelihood, with the likelihood from the observation level of model (1). Similarly, we can estimate each $\lambda_{k,(c)}$ conditional on $\mathbf{f}_k^{(c)}$ by maximizing the likelihood $\mathbf{d}_k^{(c)} \sim N(\mathbf{0}, \mathbf{D}_k^{(c)})$ with respect to $\lambda_{k,(c)}$, where $\mathbf{D}_k^{(c)} = \text{diag}(10^8, 10^8, \lambda_{k,(c)}^{-1}, \dots, \lambda_{k,(c)}^{-1})$. Then, given $\lambda_{k,(c)}$, $\sigma_{(c)}^2$, $\boldsymbol{\beta}_k^{(c)}$, and $\mathbf{f}_k^{(c)}$, we can estimate each $\mathbf{d}_k^{(c)}$ by normalizing the full conditional posterior expectation given in the main paper; i.e., solving the relevant quadratic program and then normalizing the solution. Initializations for the remaining levels proceed similarly as conditional MLEs, but depend on the form chosen for \mathbf{X}_t , \mathbf{V}_t , \mathbf{G}_t , and \mathbf{W}_t . In our applications, this conditional MLE approach produces reasonable starting values for all variables.

A.1 Common Factor Loading Curves

If we wish to implement the common FLCs model $f_k^{(c)} = f_k$ for all k, c , then we instead compute the SVD of the stacked data matrices $(\mathbf{Y}^{(1)'}', \dots, \mathbf{Y}^{(C)'}')' = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}'$, where now the data matrices $\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(C)}$ are imputed using splines for all observation points for all outcomes, $\cup_{t,c}\mathcal{T}_t^{(c)}$, and therefore have the same number of columns. Alternatively, we may improve computational efficiency by choosing a small yet representative subset of observation points $\mathcal{T}^* \subset \cup_{t,c}\mathcal{T}_t^{(c)}$ and then estimating each data matrix $\mathbf{Y}^{(c)}$ for all $\tau \in \mathcal{T}^*$. Let $\mathbf{U}_{1:K}^{(c)}$ be the first K columns of $\mathbf{U}^{(c)}$, where $\mathbf{U}^{(c)}, c = 1, \dots, C$, corresponds to the outcome-specific blocks of $\mathbf{U} = (\mathbf{U}^{(1)'}, \dots, \mathbf{U}^{(C)'})'$. Then, similar to before, we set $(\boldsymbol{\beta}_1^{(c)}, \dots, \boldsymbol{\beta}_K^{(c)}) = \mathbf{U}_{1:K}^{(c)}\boldsymbol{\Sigma}_{1:K}$ for $c = 1, \dots, C$, and $(\mathbf{f}_1, \dots, \mathbf{f}_K) = \mathbf{V}_{1:K}$, where $\boldsymbol{\Sigma}_{1:K}$ is the upper left $K \times K$ submatrix of $\boldsymbol{\Sigma}$ and $\mathbf{V}_{1:K}$ is the first K columns of \mathbf{V} . Again, the \mathbf{f}_k are unsmoothed with $\mathbf{f}_k'\mathbf{f}_j = \mathbf{1}(k = j)$, but now the initialized FLCs are common for $c = 1, \dots, C$. Initialization of the remaining parameters proceeds as before, but now with $\lambda_{k,(c)} = \lambda_k$ and $\mathbf{d}_k^{(c)} = \mathbf{d}_k$, which can be obtained by maximizing the relevant conditional likelihoods under the common FLCs model.

A.2 Computing a range for K

The initialization procedure requires the SVD of the data matrix. If we first center the columns of the data matrix, then the squared components of the diagonal matrix $\boldsymbol{\Sigma}^{(c)}$ (or $\boldsymbol{\Sigma}$) indicate the variance explained by each factor. Therefore, we can estimate the proportion of total variance in the data explained by each factor, without the need to run an MCMC sampler. Using this information, we can either select K based on the minimum number of factors needed to explain a prespecified proportion of total variance explained, such as 95%, or select a range for K based on an interval of proportion of total variance explained, such as (80%, 99%). In the latter case, we can then select K by comparing the marginal likelihood or DIC for each K in this range. Note that in both cases, it may be appropriate to increase the selected value(s) of K by one to account for the initial centering of the data matrix.

B Sampling the MFDLM

For greater generality, we present our sampling algorithm for non-common FLCs; i.e., we retain dependence on c for $\mathbf{d}_k^{(c)}$ and $\lambda_{k,(c)}$. When applicable, we discuss the necessary modifications for the common FLCs model.

The algorithm proceeds in four main blocks:

1. Sample the smoothing parameters $\lambda_{k,(c)}$ and the basis coefficients $\mathbf{d}_k^{(c)}$ for the FLCs.

Using uniform priors on the standard deviations $\lambda_{k,(c)}^{-1/2}$ and enforcing the ordering constraints $\lambda_{1,(c)} > \lambda_{2,(c)} > \dots > \lambda_{K,(c)}$, the conditional priors are $\lambda_{k,(c)}^{-1/2} \sim \text{Uniform}(\ell_{k,(c)}, u_{k,(c)})$, where $\ell_{1,(c)} = 0$, $\ell_{k,(c)} = \lambda_{k-1,(c)}^{-1/2}$ for $k = 2, \dots, K$, $u_{k,(c)} = \lambda_{k+1,(c)}^{-1/2}$ for $k = 1, \dots, K-1$, and $u_{K,(c)} = 10^4$. For $k = 1, \dots, K, c = 1, \dots, C$, the full conditional distribution for $\lambda_{k,(c)}$ is $\text{Gamma}\left(\frac{1}{2}(M+1), \frac{1}{2} \sum_{j=3}^{M+4} d_{k,(c),j}^2\right)$ truncated to the interval $(u_{k,(c)}^{-2}, \ell_{k,(c)}^{-2})$, where M is the number of interior knots, $d_{k,(c),j}$ are the components of $\mathbf{d}_k^{(c)}$, and $\ell_{1,(c)}^{-2} = \infty$. For the common FLCs model, we simply replace $\mathbf{d}_k^{(c)}$ with \mathbf{d}_k to obtain the full conditional posterior for λ_k . To reduce dependence of the ordering of $\lambda_{k,(c)}$ on the initialization procedure of Section A—which fixes the ordering without accounting for the smoothness of the FLCs $f_k^{(c)}$ —we run the first 10 MCMC iterations without enforcing the ordering constraints, so $\ell_{k,(c)} = 0$ and $u_{k,(c)} = 10^4$ for $k = 1, \dots, K$. At the end of this brief trial run, we reorder $\lambda_{k,(c)}, f_k^{(c)}$, and $\beta_{k,t}^{(c)}$ to reflect the ordering constraint; we may reorder the other parameters as well, but typically this is not necessary. We can sample $\lambda_{k,(c)}$ from the truncated Gamma distribution using the following procedure:

- (a) Sample $U \sim \text{Uniform}(a, b)$, where $a = F_G(u_{k,(c)}^{-2})$ and $b = F_G(\ell_{k,(c)}^{-2})$, with $F_G(\cdot)$ the distribution function of the full conditional Gamma distribution given above;
- (b) Set $\lambda_{k,(c)} = F_G^{-1}(U)$.

After sampling the $\lambda_{k,(c)}$, we sample and then normalize the $\mathbf{d}_k^{(c)}$ with a modified version

of the efficient Cholesky decomposition approach of Wand and Ormerod (2008):

- (a) Compute the (lower triangular) Cholesky decomposition $\mathbf{B}_k^{-1} = \bar{\mathbf{B}}_L \bar{\mathbf{B}}_L'$;
- (b) Use forward substitution to obtain $\bar{\mathbf{b}}$ as the solution to $\bar{\mathbf{B}}_L \bar{\mathbf{b}} = \mathbf{b}_k$, then use backward substitution to obtain \mathbf{d}_k^U as the solution to $\bar{\mathbf{B}}_L' \mathbf{d}_k^U = \bar{\mathbf{b}} + \bar{\mathbf{z}}$, where $\bar{\mathbf{z}} \sim N(\mathbf{0}, \mathbf{I}_{(M+4) \times (M+4)})$;
- (c) Use forward substitution to obtain $\bar{\mathbf{L}}$ as the solution to $\bar{\mathbf{B}}_L \bar{\mathbf{L}} = \mathbf{L}_{[-k]}$, then use backward substitution to obtain $\tilde{\mathbf{L}}$ as the solution to $\bar{\mathbf{B}}_L' \tilde{\mathbf{L}} = \bar{\mathbf{L}}$;
- (d) Set $\mathbf{d}_k^* = \mathbf{d}_k^U - \tilde{\mathbf{L}} (\mathbf{L}_{[-k]}' \tilde{\mathbf{L}})^{-1} \mathbf{L}_{[-k]}' \mathbf{d}_k^U$;
- (e) Retain the vector $\mathbf{d}_k^{(c)} = \mathbf{d}_k^* / \sqrt{\mathbf{d}_k^{*'} \mathbf{J}_\phi \mathbf{d}_k^*}$ and set $\boldsymbol{\beta}_k^{(c)} = \sqrt{\mathbf{d}_k^{*'} \mathbf{J}_\phi \mathbf{d}_k^*} \boldsymbol{\beta}_k^{(c)}$.

The definitions of \mathbf{B}_k and \mathbf{b}_k depend on whether or not we use the common FLCs model with $f_k^{(c)} = f_k$ (see Section 3 of the paper). The sample $\mathbf{d}_k^U \sim N(\mathbf{B}_k \mathbf{b}_k, \mathbf{B}_k)$ in (b) is unconstrained, while steps (c) and (d) incorporate the linear orthogonality constraints: the random variable $\mathbf{d}_k^* = \mathbf{d}_k^U - \mathbf{B}_k \mathbf{L}_{[-k]} (\mathbf{L}_{[-k]}' \mathbf{B}_k \mathbf{L}_{[-k]})^{-1} \mathbf{L}_{[-k]}' \mathbf{d}_k^U$ follows the correct distribution $N(\tilde{\mathbf{B}}_k, \mathbf{b}_k, \tilde{\mathbf{B}}_k)$, which conditions on the linear orthogonality constraints $\mathbf{d}_k' \mathbf{L}_{[-k]} = \mathbf{0}$. Steps (c) and (d) compute this random variable efficiently (see Gelfand et al., 2010, Chapter 12 for more details). The scaling of $\mathbf{d}_k^{(c)}$ and $\boldsymbol{\beta}_k^{(c)}$ in (d) enforces the unit-norm constraint on $f_k^{(c)}$ yet ensures that $f_k^{(c)}(\tau) \boldsymbol{\beta}_k^{(c)}$ —which appears in the posterior distribution of $\mathbf{d}_j^{(c)}$ for all $j \neq k$ —is unaffected by the normalization. To encourage better mixing, we randomly select the order of $k = 1, \dots, K$ in which to sample $\lambda_{k,(c)}$ and $\mathbf{d}_k^{(c)}, c = 1, \dots, C$.

2. Sample the factors $\boldsymbol{\beta}_t$ (and $\boldsymbol{\theta}_t$, if present) conditional on all other parameters in (1) using the state space sampler of Durbin and Koopman (2002); Koopman and Durbin (2003, 2000), which is optimized when \mathbf{E}_t is diagonal. For general hierarchical models, we may modify the hierarchical DLM algorithms of Gamerman and Migon (1993).

For the prior distributions, we only need to specify the distribution of $\boldsymbol{\beta}_0$ (and $\boldsymbol{\theta}_0$);

the remaining distributions are computed recursively using \mathbf{F} , \mathbf{X}_t , \mathbf{G}_t and the error variances. For simplicity, we let $\beta_{k,0}^{(c)} \stackrel{iid}{\sim} N(0, 10^4)$, which is a common choice for DLMs.

3. Sample the state evolution matrix \mathbf{G}_t (if unknown). \mathbf{G}_t may have a special form (see Section D of this supplement) or provide a more common time series model such as a VAR. In the latter case, we may choose some structure for $\mathbf{G}_t = \mathbf{G}$, e.g. diagonality to allow dependence between $\beta_{k,t}^{(c)}$ and $\beta_{k,t-1}^{(c)}$, or K blocks of dimension $C \times C$ to allow dependence between $\beta_{k,t}^{(c)}$ and $\beta_{k,t-1}^{(c')}$ for $c, c' = 1, \dots, C$. It is particularly convenient to assume a Gaussian prior for the nonzero entries of \mathbf{G} , which is a conjugate prior for $\text{vec}_0(\mathbf{G})$, where vec_0 stacks the nonzero entries of the matrix (by column) into a vector.
4. Sample each of the remaining error variance parameters separately: \mathbf{E}_t , \mathbf{V}_t , and \mathbf{W}_t . These distributions depend on our assumptions for the model structure, but we typically prefer conjugate priors when available. In both applications, we fix $\mathbf{V}_t = \mathbf{0}_{CK \times CK}$ to remove a level in the hierarchy, and let $\mathbf{E}_t = \text{diag}(\sigma_{(1)}^2, \dots, \sigma_{(C)}^2)$ with $\sigma_{(c)}^{-2} \stackrel{iid}{\sim} \text{Gamma}(0.001, 0.001)$, for which the full conditional posterior distribution is

$$\text{Gamma}\left(0.001 + \frac{1}{2} \sum_{t \in T^{(c)}} |\mathcal{T}_t^{(c)}|, 0.001 + \frac{1}{2} \sum_{t \in T^{(c)}} \sum_{\tau \in \mathcal{T}_t^{(c)}} \left\{ Y_t^{(c)}(\tau) - \sum_{k=1}^K \beta_{k,t}^{(c)} f_k^{(c)}(\tau) \right\}^2\right).$$

In the random walk factor model of (9), we have $\boldsymbol{\beta}_{k,i,s,t} = \boldsymbol{\beta}_{k,i,s,t-1} + \boldsymbol{\omega}_{k,i,s,t}$ with $\boldsymbol{\omega}_{k,i,s,t} \stackrel{\text{indep}}{\sim} N(\mathbf{0}, \mathbf{W}_k)$ for $t = 2, \dots, 15$. Using the Wishart prior $\mathbf{W}_k^{-1} \sim \text{Wishart}((\rho R)^{-1}, \rho)$, the full conditional posterior distribution for the precision is $\mathbf{W}_k^{-1} \sim \text{Wishart}((\rho R + \sum_{i,s,t} \mathbf{w}_{k,i,s,t} \mathbf{w}_{k,i,s,t}')^{-1}, \rho + 4480)$, where $\mathbf{w}_{k,i,s,t} = \boldsymbol{\beta}_{k,i,s,t} - \boldsymbol{\beta}_{k,i,s,t-1}$ is conditional on the factors and 4480 counts the indices (i, s, t) in the summation. We let $R^{-1} = \mathbf{I}_{C \times C}$, which is the expected prior precision, and $\rho = C \geq \text{rank}(R^{-1})$.

For the stochastic volatility model of Section 4.1.2, we use the prior distributions and sampling algorithm given in Kastner and Frühwirth-Schnatter (2014), implemented via the R package `stochvol` (Kastner, 2015). Letting $\sigma_{k,(c),t}^2 = \exp(h_{k,t}^{(c)})$, the model is $h_{k,t}^{(c)} =$

$\xi_{k,0}^{(c)} + \xi_{k,1}^{(c)}(h_{k,t-1}^{(c)} - \xi_{k,0}^{(c)}) + \zeta_{k,t}^{(c)}$, where $\zeta_{k,t}^{(c)} \stackrel{\text{indep}}{\sim} N(0, \sigma_{H,k,(c)}^2)$ for $t = 2, \dots, T$ and $h_{k,1}^{(c)} \sim N(\xi_{k,0}^{(c)}, \sigma_{H,k,(c)}^2 / (1 - (\xi_{k,1}^{(c)})^2))$ with $|\xi_{k,1}^{(c)}| < 1$ for stationarity. The accompanying priors are $\xi_{k,0}^{(c)} \stackrel{\text{indep}}{\sim} N(0, 10^4)$, $(\xi_{k,1}^{(c)} + 1)/2 \stackrel{\text{indep}}{\sim} \text{Beta}(5, 1.5)$, and $\sigma_{H,k,(c)}^2 \stackrel{\text{indep}}{\sim} \text{Gamma}(\frac{1}{2}, \frac{1}{2})$. The hyperparameters for the Beta prior are chosen reflect the high persistence of volatility commonly found in financial data, and the prior for $\sigma_{H,k,(c)}^2$ corresponds to a half-normal distribution. For additional motivation for the stochastic volatility approach over GARCH models, see Danielsson (1998). Note that the sampling algorithm of Kastner and Frühwirth-Schnatter (2014) requires a Metropolis step, and therefore the methods of Chib and Jeliazkov (2001) are more appropriate for marginal likelihood computations.

Recall that we construct a posterior distribution of $\mathbf{d}_k^{(c)}$ without the unit norm constraint, and then normalize the samples from this distribution. As a result, the conditions of Theorem 1 are satisfied and the (unnormalized) full conditional posterior distribution of $\mathbf{d}_k^{(c)}$ is Gaussian, both of which are convenient results. The normalization step 1.(d) is interpretable, corresponding to the projection of a Gaussian distribution onto the unit sphere. Note that rescaling the factors $\boldsymbol{\beta}_k^{(c)}$ in 1.(d) does not affect the remainder of the sampling algorithm (steps 2. - 4.). The rescaled $\boldsymbol{\beta}_k^{(c)}$ are from the previous MCMC iteration, which does not affect the full conditional distributions of step 2. in the current MCMC iteration. The subsequent steps 3., 4., and 1. are then conditional on the newly sampled factors $\boldsymbol{\beta}_k^{(c)}$ from step 2., which have not been rescaled.

C MCMC Diagnostics

To demonstrate convergence and efficiency of the Gibbs sampler, we provide MCMC diagnostics for both applications. We include trace plots for several variables of interest to asses the mixing and convergence of the simulated chains. The trace plots also suggest reasonable lengths of the burn-in, i.e., the initial simulations that are discarded prior to convergence of

the chain. To measure the efficiency of the sampler, we compute the ratio of the effective sample size to the simulation sample size for several variables. We refer to this quantity as the efficiency factor, which is the reciprocal of the simulation inefficiency factor (e.g., Kim et al., 1998). All diagnostics were computed using the R package `coda` (Plummer et al., 2006).

C.1 Multi-Economy Yield Curves

We ran the MCMC sampler for 7,000 iterations and discarded the first 2,000 iterations as a burn-in. Longer chains and dispersed starting values did not produce noticeably different results. The sampler was run in R, and took 181 minutes on a laptop with a 2.40 GHz Intel i7-4700MQ CPU using one core. We are currently developing an R package for the MFIDLM sampler, and expect sizable gains in computational efficiency by coding the algorithms in C.

Tables C.1, C.2, and C.3 contain the efficiency factors for the common FLCs f_k evaluated at several quantiles of τ , the factors $\beta_{k,t}^{(c)}$ at various times t , and the slopes $\gamma_k^{(c)}$ from the common trend model, respectively. The efficiency of both the FLCs and the factors is exceptional. The FLCs are most efficient for the longer maturities, and several of the efficiency factors for the $\beta_{k,t}^{(c)}$ exceed one. The slopes $\gamma_k^{(c)}$ are less efficient, but still at least 11% for all k, c .

	$\tau = 8$	$\tau = 90$	$\tau = 180$	$\tau = 270$
$f_1(\tau)$	0.52	0.72	0.72	0.71
$f_2(\tau)$	0.48	0.72	0.73	0.71
$f_3(\tau)$	0.66	0.96	0.89	0.92
$f_4(\tau)$	0.54	0.77	0.77	0.91
$f_5(\tau)$	0.61	0.72	0.84	0.85
$f_6(\tau)$	0.58	0.94	0.89	0.85

Table C.1: Efficiency factors for the posterior sampling of $f_k(\tau), k = 1, \dots, 6$, for maturities $\tau \in \{8, 90, 180, 270\}$ months, which are the 2nd, 25th, 50th, and 75th quantiles of the observation points, using model (8) for the yield curve application.

In Figures C.1, C.2, and C.3, we present the trace plots for the FLCs, the factors, and the slopes, respectively. The vertical gray bars indicate the selected burn-in of 2,000 iterations.

	2006-02-10	2007-07-06	2008-12-05	2010-04-30	2011-09-23	2013-02-22
$k = 1, c = 1$	0.83	0.91	1.00	0.83	0.91	1.00
$k = 2, c = 1$	0.96	0.42	0.55	0.96	0.42	0.55
$k = 3, c = 1$	0.98	0.68	1.00	0.98	0.68	1.00
$k = 4, c = 1$	0.91	1.00	1.01	0.91	1.00	1.01
$k = 5, c = 1$	0.72	1.00	1.00	0.72	1.00	1.00
$k = 6, c = 1$	0.41	0.90	1.00	0.41	0.90	1.00
$k = 1, c = 2$	0.95	1.00	1.00	0.95	1.00	1.00
$k = 2, c = 2$	1.10	1.00	1.00	1.10	1.00	1.00
$k = 3, c = 2$	0.82	1.00	1.00	0.82	1.00	1.00
$k = 4, c = 2$	1.04	1.02	1.00	1.04	1.02	1.00
$k = 5, c = 2$	0.95	1.00	1.00	0.95	1.00	1.00
$k = 6, c = 2$	1.00	1.00	1.00	1.00	1.00	1.00
$k = 1, c = 3$	1.00	1.00	1.00	1.00	1.00	1.00
$k = 2, c = 3$	1.00	0.94	0.94	1.00	0.94	0.94
$k = 3, c = 3$	1.00	1.00	1.00	1.00	1.00	1.00
$k = 4, c = 3$	1.00	1.00	1.00	1.00	1.00	1.00
$k = 5, c = 3$	1.00	1.06	1.00	1.00	1.06	1.00
$k = 6, c = 3$	1.00	1.00	0.94	1.00	1.00	0.94
$k = 1, c = 4$	1.00	0.95	0.96	1.00	0.95	0.96
$k = 2, c = 4$	1.00	1.00	1.04	1.00	1.00	1.04
$k = 3, c = 4$	1.00	1.00	1.00	1.00	1.00	1.00
$k = 4, c = 4$	0.92	1.00	1.00	0.92	1.00	1.00
$k = 5, c = 4$	1.00	0.93	1.00	1.00	0.93	1.00
$k = 6, c = 4$	1.00	1.00	1.00	1.00	1.00	1.00

Table C.2: Efficiency factors for the posterior sampling of $\beta_{k,t}^{(c)}$ for various times t , using model (8) for the yield curve application.

Again, the FLCs and the factors demonstrate exceptional MCMC performance. Interestingly, the initializations of the FLCs appear to be farthest from the posterior modes for shorter maturities. The slopes $\gamma_k^{(c)}$ were initialized at zero, yet congregated around the posterior modes rapidly.

C.2 Multivariate Time-Frequency Analysis for Local Field Potential

We ran the MCMC sampler for 7,000 iterations and discarded the first 2,000 iterations as a burn-in. Longer chains and dispersed starting values did not produce noticeably different results. The sampler was run in R, and took 367 minutes on a laptop with a 2.40 GHz Intel i7-4700MQ CPU using one core.

Tables C.4, and C.5 contain the efficiency factors for the sample means $\bar{\mu}_t^{(c)}(\tau)$ and the

	Economy		
	BOE	ECB	BOC
$k = 1$	0.44	0.39	0.15
$k = 2$	0.12	0.11	0.12
$k = 3$	0.40	0.38	0.19
$k = 4$	0.42	0.26	0.19

Table C.3: Efficiency factors for the posterior sampling of $\gamma_k^{(c)}$, using model (8) for the yield curve application.

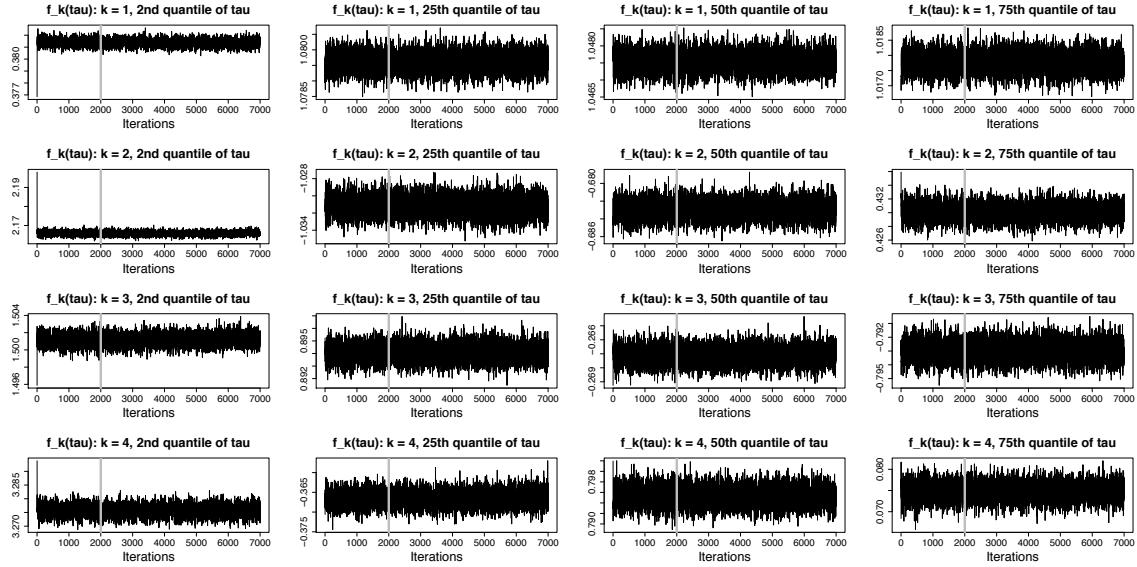


Figure C.1: Trace plots of the posterior samples of $f_k(\tau)$, $k = 1, 2, 3, 4$, for the 2nd, 25th, 50th, and 75th quantiles of the observation points, using model (8) for the yield curve application.

factors $\beta_{k,i,s,t}^{(c)}$ for various rats i , trials s , and time bins t , respectively. For $\bar{\mu}_t^{(c)}(\tau)$, we compute quantiles of the efficiency factors across all c, t, τ : the minimum efficiency factor is 78%, while the overwhelming majority of the efficiency factors are at least one. Since we compute pointwise HPD credible intervals for $\bar{\mu}_t^{(c)}(\tau)$ for all c, t, τ , it is encouraging that the MCMC sampler is extremely efficient for these parameters. As in the previous application, the MCMC efficiency of the factors is exceptional. In Figures C.4 and C.5, we present the trace plots for $\bar{\mu}_t^{(c)}(\tau)$ and $\beta_{k,i,s,t}^{(c)}$. The MCMC performance for both sets of parameters appears to be very good.

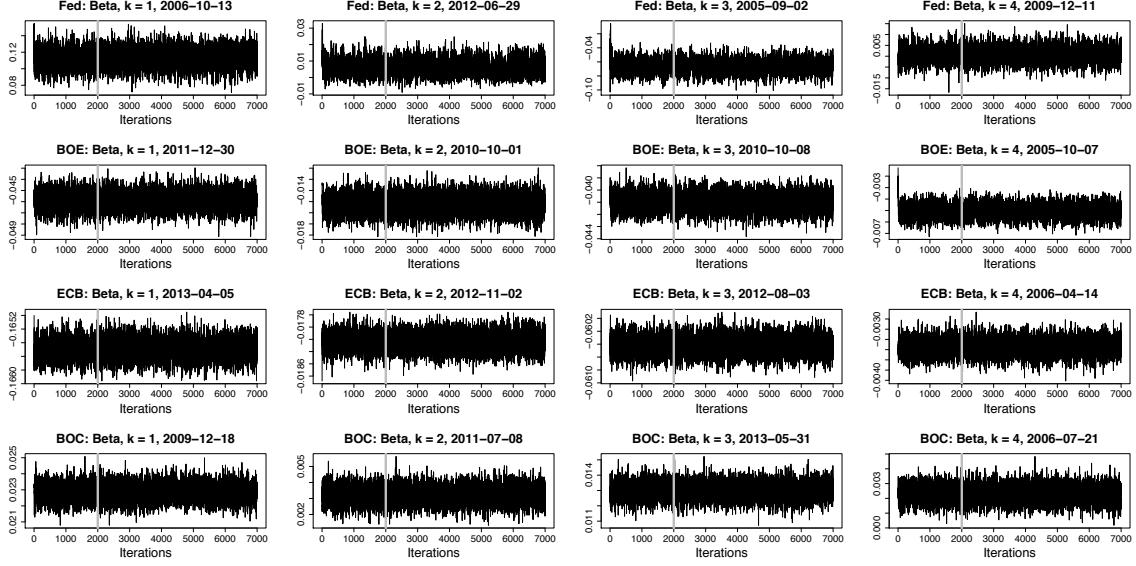


Figure C.2: Trace plots of the posterior samples of $\beta_{k,t}^{(c)}$, $k = 1, 2, 3, 4$, for various times t , using model (8) for the yield curve application. The vertical gray bar indicates the selected burn-in of 2,000 iterations.

Min.	25th Quantile	Median	Mean	75th Quantile	Max.
0.7781	1.0000	1.0000	1.0060	1.0000	1.8270

Table C.4: Summary statistics of the efficiency factors for the posterior sampling of $\bar{\mu}_t^{(c)}(\tau)$ across all c, t, τ , using model (9) for the LFP application.

D The Common Trend Hidden Markov Model

Consider the following extension of the common trend model (8) in the main paper:

$$\begin{cases} \beta_{k,t}^{(1)} = \omega_{k,t}^{(1)} \\ \beta_{k,t}^{(c)} = s_{k,t}^{(c)}(\gamma_k^{(c)} \beta_{k,t}^{(1)}) + \omega_{k,t}^{(c)} \quad c = 2, \dots, C \end{cases} \quad (\text{D.1})$$

where $\left\{ s_{k,t}^{(c)} : t = 1, \dots, T \right\}$ is a discrete Markov chain with states $\{0, 1\}$. Model (D.1) reduces to model (8) in the main paper when $s_{k,t}^{(c)} = 1$ for all k, c, t . As with the common trend model, we can use (D.1) to investigate how the factors $\beta_{k,t}^{(c)}$ for each economy $c > 1$ are *directly* related to those of the Fed, $\beta_{k,t}^{(1)}$. Model (D.1) relates each economy $c > 1$ to the Fed using a regression framework, in which we regress $\beta_{k,t}^{(c)}$ on $\beta_{k,t}^{(1)}$ with AR(r) errors, where the (Fed) predictor $\beta_{k,t}^{(1)}$ is present at time t only if $s_{k,t}^{(c)} = 1$. Therefore, the role of the states $s_{k,t}^{(c)}$ is to identify times

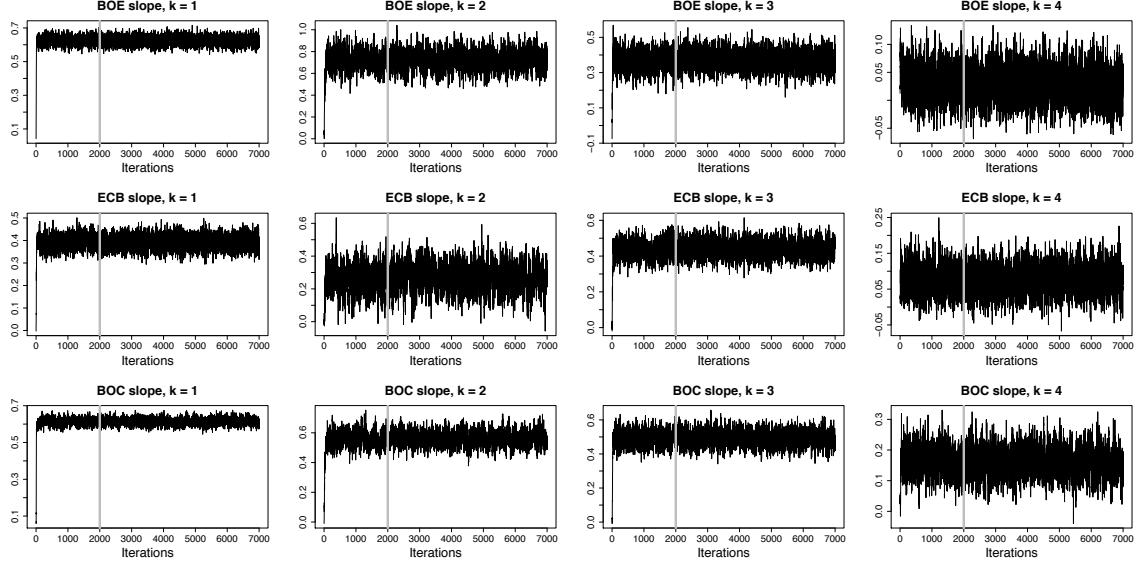


Figure C.3: Trace plots of the posterior samples of $\gamma_k^{(c)}$, $k = 1, 2, 3, 4$, using model (8) for the yield curve application.

t for which $\beta_{k,t}^{(c)}$ is strongly correlated with $\beta_{k,t}^{(1)}$; i.e., the periods for which the week-to-week changes in the features of the yield curves described by f_k are similar for economy c and the Fed. When $s_{k,t}^{(c)} = s_{k,t}^{(c')} = 1$ for $c \neq c'$, we also have dependence between $\beta_{k,t}^{(c)}$ and $\beta_{k,t}^{(c')}$; therefore, in (D.1), the Fed acts as a conduit for *all* contemporaneous dependence between economies.

It is natural for the values of the states $s_{k,t}^{(c)}$ to depend on past values of the states: if $\beta_{k,t}^{(c)}$ is correlated with $\beta_{k,t}^{(1)}$ at time t , then we may perhaps infer something about their relative behavior at time $t + 1$. Following the construction of Albert and Chib (1993), the distribution of $\{s_{k,t}^{(c)} : t = 1, \dots, T\}$, unconditional on the factors $\beta_{k,t}^{(c)}$, is determined by $P(s_{k,t}^{(c)} = 1 | s_{k,t-1}^{(c)} = 0) = q_{01,k}^{(c)}$ and $P(s_{k,t}^{(c)} = 0 | s_{k,t-1}^{(c)} = 1) = q_{10,k}^{(c)}$ with the accompanying Markov property $[s_{k,t}^{(c)} | s_{k,t-1}^{(c)}, s_{k,t-2}^{(c)}, \dots] = [s_{k,t}^{(c)} | s_{k,t-1}^{(c)}]$, where the transition probabilities $q_{01,k}^{(c)}$ and $q_{10,k}^{(c)}$ are unknown. Therefore, (D.1) contains a *hidden Markov model*, where the hidden states $s_{k,t}^{(c)}$ determine whether or not the factors $\beta_{k,t}^{(c)}$ are related to those of the Fed, $\beta_{k,t}^{(1)}$, at time t . As in Albert and Chib (1993), we use conjugate Beta priors for the transition probabilities, and select the hyperparameters so that the bulk of the mass of the prior

	720	1440	2160	2880	3600	4320
$k = 1, c = 1$	1.00	1.00	1.00	1.00	1.00	1.00
$k = 2, c = 1$	1.00	0.90	1.06	1.03	1.00	1.09
$k = 3, c = 1$	1.00	1.22	1.00	1.00	0.99	1.07
$k = 4, c = 1$	1.00	1.00	1.00	1.00	1.00	1.08
$k = 5, c = 1$	1.00	1.00	1.00	1.05	1.00	1.00
$k = 6, c = 1$	1.00	0.90	1.00	1.11	0.98	1.00
$k = 7, c = 1$	0.93	1.00	1.00	1.00	1.00	1.00
$k = 8, c = 1$	1.00	1.10	1.00	0.94	1.00	1.00
$k = 9, c = 1$	1.00	1.00	1.00	1.00	1.00	1.00
$k = 10, c = 1$	1.00	1.00	1.00	1.00	0.96	1.00
$k = 1, c = 2$	1.00	1.00	1.00	1.00	1.10	1.00
$k = 2, c = 2$	0.94	1.05	1.00	1.00	1.00	1.00
$k = 3, c = 2$	1.00	1.07	1.00	1.00	0.87	0.94
$k = 4, c = 2$	1.13	1.00	1.00	1.01	0.95	0.89
$k = 5, c = 2$	1.00	1.00	1.00	1.00	0.95	1.00
$k = 6, c = 2$	1.00	1.12	1.00	1.05	1.01	1.00
$k = 7, c = 2$	1.00	1.06	1.00	1.00	1.00	1.00
$k = 8, c = 2$	1.00	1.14	1.05	1.00	1.07	1.00
$k = 9, c = 2$	0.88	1.00	0.95	1.00	1.00	1.00
$k = 10, c = 2$	1.00	1.03	1.00	1.00	1.00	1.00
$k = 1, c = 3$	1.00	1.07	1.00	1.00	1.15	0.95
$k = 2, c = 3$	1.00	1.07	1.00	0.95	1.00	0.90
$k = 3, c = 3$	1.00	1.00	1.00	0.95	1.00	1.00
$k = 4, c = 3$	1.00	1.00	0.93	1.06	1.00	1.00
$k = 5, c = 3$	1.00	1.00	1.00	1.00	1.00	1.00
$k = 6, c = 3$	1.00	1.00	0.94	0.97	1.00	1.00
$k = 7, c = 3$	1.00	1.00	0.95	1.00	1.00	1.00
$k = 8, c = 3$	1.00	1.00	1.00	0.86	0.95	1.00
$k = 9, c = 3$	1.15	1.00	1.00	1.00	1.00	1.00
$k = 10, c = 3$	1.04	1.00	1.00	1.00	1.00	1.00

Table C.5: Efficiency factors for the posterior sampling of $\beta_{k,i,s,t}^{(c)}$, using model (9) for the LFP application. The column indexes are the 15th, 30th, 45th, 60th, 75th, and 90th quantiles of 1:4800, which is the concatenated time index across rats $i = 1, \dots, 8$, trials $s = 1, \dots, 40$, and time bins $t = 1, \dots, 15$.

distribution is on $(0, 0.5)$, which reflects the belief that transitions should occur infrequently. Sampling from the posterior distribution of $\left\{ s_{k,t}^{(c)} : t = 1, \dots, T \right\}$ (i.e., conditional on the factors $\beta_{k,t}^{(c)}$) is a straightforward application of Albert and Chib (1993).

D.1 Sampling The Common Trend Hidden Markov Model

While model (D.1) is a useful example of the flexibility of the MFDLM, it is not supported by DIC: the DIC for model (8) is $-2,393,266$, while the DIC for model (D.1) is $-2,393,200$. However, since we can obtain the preferred model (8) from the main paper by setting $s_{k,t}^{(c)} = 1$,

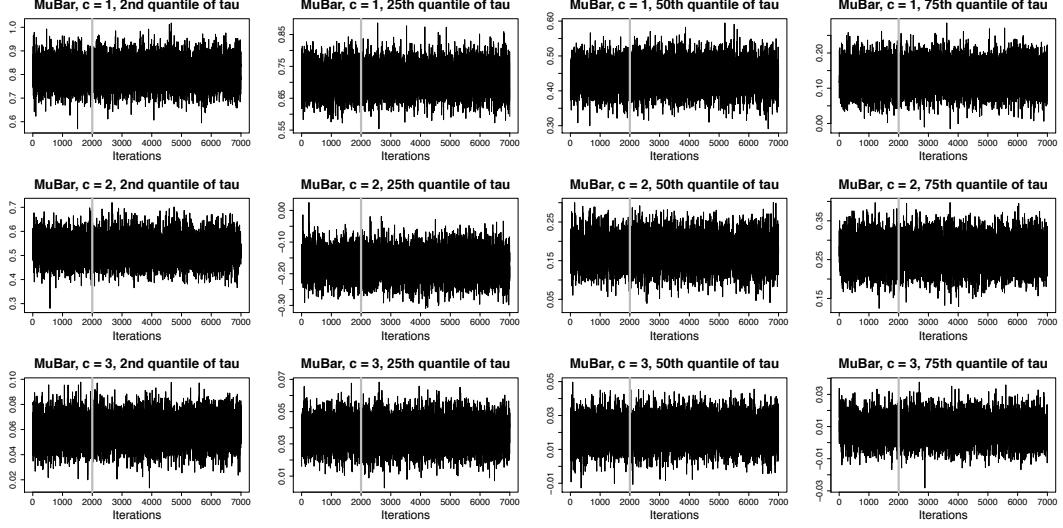


Figure C.4: Trace plots of the posterior samples of $\bar{\mu}_t^{(c)}(\tau)$, for the 2nd, 25th, 50th, and 75th quantiles of the observation points, $c = 1, \dots, C$, and selected time bins, using model (9) for the LFP application. The vertical gray bar indicates the selected burn-in of 2,000 iterations.

we describe the DLM construction for the more general model (D.1). Expressing (D.1) as a DLM allows us to use efficient state space samplers for the factors β_t , as in the algorithm described in Section B.

We can express (D.1) as the $\beta_t = \theta_t$ -level in (1) with $\mathbf{X}_t = \mathbf{I}_{CK \times CK}$ and $\mathbf{V}_t = \mathbf{0}_{CK \times CK}$. Let $\mathbf{L}_{\beta_t} = \mathbf{I}_{CK \times CK} - \mathbf{Q}_t$,

$$\mathbf{Q}_t = \begin{pmatrix} \mathbf{0}_{K \times K} & \mathbf{0}_{K \times K} & \cdots & \mathbf{0}_{K \times K} \\ \mathbf{S}_t^{(2)} \boldsymbol{\gamma}^{(2)} & \mathbf{0}_{K \times K} & \cdots & \mathbf{0}_{K \times K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_t^{(C)} \boldsymbol{\gamma}^{(C)} & \mathbf{0}_{K \times K} & \cdots & \mathbf{0}_{K \times K} \end{pmatrix},$$

where $\mathbf{S}_t^{(c)} = \text{diag}(\{s_{k,t}^{(c)}\}_{k=1}^K)$ and $\boldsymbol{\gamma}^{(c)} = \text{diag}(\{\gamma_k^{(c)}\}_{k=1}^K)$. Note that $\mathbf{L}_{\beta_t}^{-1} = \mathbf{I}_{CK \times CK} + \mathbf{Q}_t$. In vector notation, (D.1) can be written

$$\mathbf{L}_{\beta_t} \beta_t = \Psi \mathbf{L}_{\beta_{t-1}} \beta_{t-1} + \tilde{\omega}_t \quad (\text{D.2})$$



Figure C.5: Trace plots of the posterior samples of $\beta_{k,i,s,t}^{(c)}$ for various (i, s, t) , using model (9) for the LFP application.

where $\Psi = \text{diag}(\{\psi_{k,1}^{(c)}\}_{k,c})$ and $\tilde{\omega}_t$ has elements $\tilde{\omega}_{k,t}^{(c)} = \sigma_{k,(c),t} z_{k,t}^{(c)}$ with $\tilde{\omega}_t \sim N(\mathbf{0}, \tilde{\mathbf{W}}_t)$ and $\tilde{\mathbf{W}}_t = \text{diag}(\{\sigma_{k,(c),t}^2\}_{k,c})$. Inverting \mathbf{L}_{β_t} , the DLM evolution equation is therefore

$$\boldsymbol{\beta}_t = \mathbf{G}_t \boldsymbol{\beta}_{t-1} + \boldsymbol{\omega}_t \quad (\text{D.3})$$

where $\mathbf{G}_t = (\mathbf{I}_{CK \times CK} + \mathbf{Q}_t)\Psi(\mathbf{I}_{CK \times CK} - \mathbf{Q}_{t-1})$ and $\boldsymbol{\omega}_t = (\mathbf{I}_{CK \times CK} + \mathbf{Q}_t)\tilde{\omega}_t \sim N(\mathbf{0}, \mathbf{W}_t)$, with $\mathbf{W}_t = \mathbf{L}_{\beta_t}^{-1} \tilde{\mathbf{W}}_t (\mathbf{L}_{\beta_t}^{-1})'$. Since $\mathbf{Q}_t \Psi \mathbf{Q}_{t-1} = \mathbf{0}_{CK \times CK}$, we have

$$\mathbf{G}_t = \begin{pmatrix} \Psi^{(1)} & \mathbf{0}_{K \times K} & \cdots & \mathbf{0}_{K \times K} \\ \gamma^{(2)} \left(\mathbf{S}_t^{(2)} \Psi^{(1)} - \mathbf{S}_{t-1}^{(2)} \Psi^{(2)} \right) & \Psi^{(2)} & \cdots & \mathbf{0}_{K \times K} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma^{(C)} \left(\mathbf{S}_t^{(C)} \Psi^{(1)} - \mathbf{S}_{t-1}^{(C)} \Psi^{(C)} \right) & \mathbf{0}_{K \times K} & \cdots & \Psi^{(C)} \end{pmatrix},$$

where $\Psi^{(c)} = \text{diag}(\{\psi_{k,1}^{(c)}\}_k)$. Similarly, we may compute $\mathbf{W}_t = (\mathbf{I}_{CK \times CK} + \mathbf{Q}_t)\tilde{\mathbf{W}}_t(\mathbf{I}_{CK \times CK} + \mathbf{Q}'_t) = \tilde{\mathbf{W}}_t + \mathbf{Q}_t \tilde{\mathbf{W}}_t + (\mathbf{Q}_t \tilde{\mathbf{W}}_t)' + \mathbf{Q}_t \tilde{\mathbf{W}}_t \mathbf{Q}'_t$. Letting $\sigma_{(c),t}^2 = \text{diag}(\{\sigma_{k,(c),t}^2\}_{k=1}^K)$ so that $\tilde{\mathbf{W}}_t =$

bdiag($\boldsymbol{\sigma}_{(1),t}^2, \dots, \boldsymbol{\sigma}_{(C),t}^2$), we may compute the relevant terms explicitly:

$$\mathbf{Q}_t \tilde{\mathbf{W}}_t = \begin{pmatrix} \mathbf{0}_{K \times K} & \mathbf{0}_{K \times K} & \cdots & \mathbf{0}_{K \times K} \\ \mathbf{S}_t^{(2)} \boldsymbol{\gamma}^{(2)} \boldsymbol{\sigma}_{(1),t}^2 & \mathbf{0}_{K \times K} & \cdots & \mathbf{0}_{K \times K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_t^{(C)} \boldsymbol{\gamma}^{(C)} \boldsymbol{\sigma}_{(1),t}^2 & \mathbf{0}_{K \times K} & \cdots & \mathbf{0}_{K \times K} \end{pmatrix}$$

and

$$\mathbf{Q}_t \tilde{\mathbf{W}}_t \mathbf{Q}'_t = \begin{pmatrix} \mathbf{0}_{K \times K} & \mathbf{0}_{K \times K} & \cdots & \mathbf{0}_{K \times K} \\ \mathbf{0}_{K \times K} & \mathbf{S}_t^{(2)} \boldsymbol{\gamma}^{(2)} \boldsymbol{\sigma}_{(1),t}^2 \mathbf{S}_t^{(2)} \boldsymbol{\gamma}^{(2)} & \cdots & \mathbf{S}_t^{(2)} \boldsymbol{\gamma}^{(2)} \boldsymbol{\sigma}_{(1),t}^2 \mathbf{S}_t^{(C)} \boldsymbol{\gamma}^{(C)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{K \times K} & \mathbf{S}_t^{(C)} \boldsymbol{\gamma}^{(C)} \boldsymbol{\sigma}_{(1),t}^2 \mathbf{S}_t^{(2)} \boldsymbol{\gamma}^{(2)} & \cdots & \mathbf{S}_t^{(C)} \boldsymbol{\gamma}^{(C)} \boldsymbol{\sigma}_{(1),t}^2 \mathbf{S}_t^{(C)} \boldsymbol{\gamma}^{(C)} \end{pmatrix}$$

where again, the component terms are all diagonal, and therefore can be reordered for convenience. Combining terms and simplifying, the error variance matrix is

$$\mathbf{W}_t = \begin{pmatrix} \boldsymbol{\sigma}_{(1),t}^2 & \mathbf{S}_t^{(2)} \boldsymbol{\gamma}^{(2)} \boldsymbol{\sigma}_{(1),t}^2 & \cdots & \mathbf{S}_t^{(C)} \boldsymbol{\gamma}^{(C)} \boldsymbol{\sigma}_{(1),t}^2 \\ \mathbf{S}_t^{(2)} \boldsymbol{\gamma}^{(2)} \boldsymbol{\sigma}_{(1),t}^2 & \boldsymbol{\sigma}_{(2),t}^2 + \mathbf{S}_t^{(2)} (\boldsymbol{\gamma}^{(2)})^2 \boldsymbol{\sigma}_{(1),t}^2 & \cdots & \mathbf{S}_t^{(2)} \mathbf{S}_t^{(C)} \boldsymbol{\gamma}^{(2)} \boldsymbol{\gamma}^{(C)} \boldsymbol{\sigma}_{(1),t}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_t^{(C)} \boldsymbol{\gamma}^{(C)} \boldsymbol{\sigma}_{(1),t}^2 & \mathbf{S}_t^{(2)} \mathbf{S}_t^{(C)} \boldsymbol{\gamma}^{(2)} \boldsymbol{\gamma}^{(C)} \boldsymbol{\sigma}_{(1),t}^2 & \cdots & \boldsymbol{\sigma}_{(C),t}^2 + \mathbf{S}_t^{(C)} (\boldsymbol{\gamma}^{(C)})^2 \boldsymbol{\sigma}_{(1),t}^2 \end{pmatrix}.$$

When $s_{k,t}^{(c)} = 1, c > 1$ the slope parameter $\gamma_k^{(c)}$ may increase or decrease the error variance of the residuals $\tilde{\omega}_{k,t}^{(c)}$ at time t , and determines the contemporaneous covariance between $\tilde{\omega}_{k,t}^{(c)}$ and $\tilde{\omega}_{k,t}^{(1)}$. Similarly, when $s_{k,t}^{(c)} = s_{k,t}^{(c')} = 1$, the product $\gamma_k^{(c)} \gamma_k^{(c')} \boldsymbol{\sigma}_{k,(1),t}^2$ determines the contemporaneous covariance between $\tilde{\omega}_{k,t}^{(c)}$ and $\tilde{\omega}_{k,t}^{(c')}$ at time t .

E Additional Figures

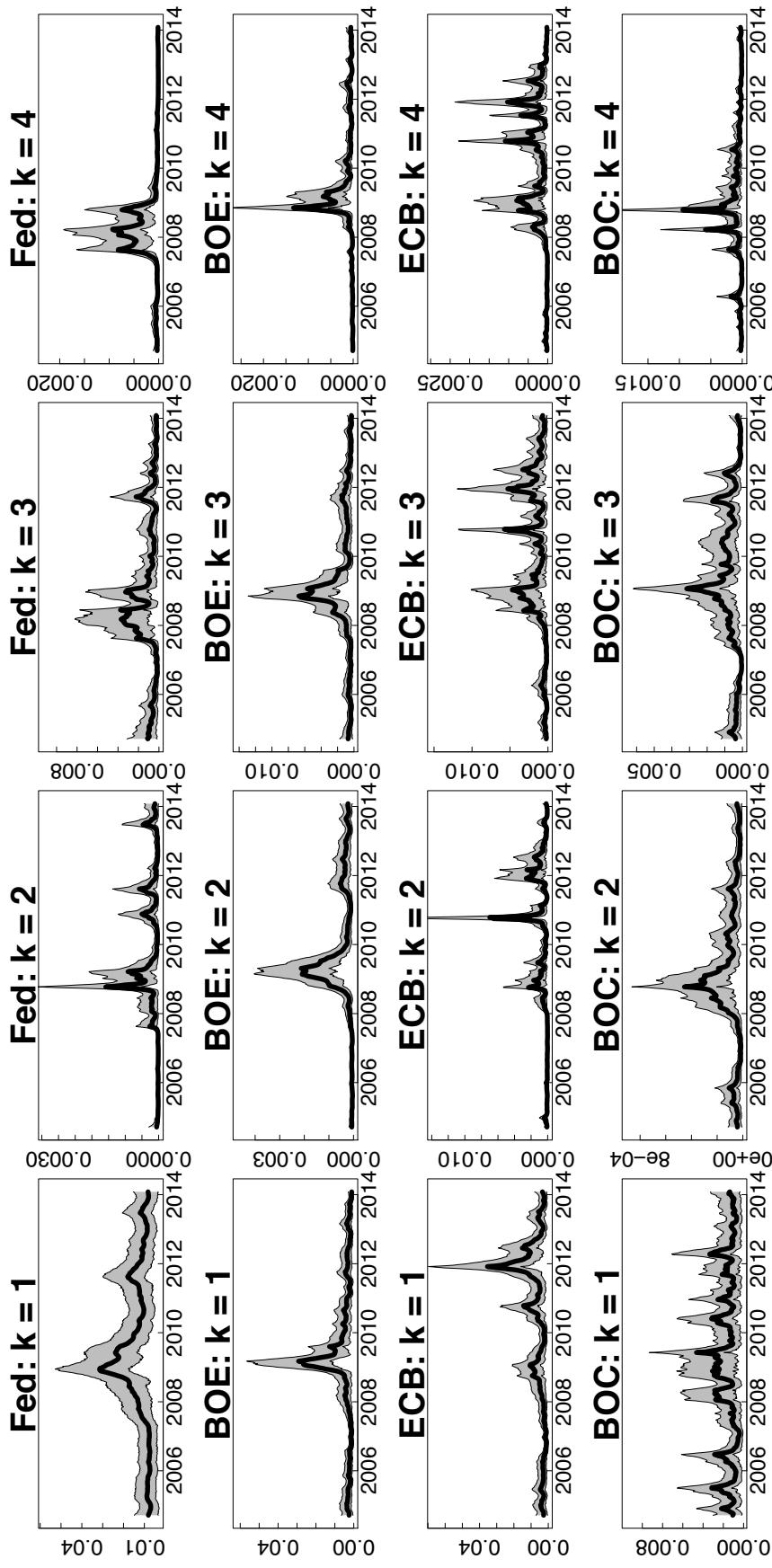


Figure E.1: Posterior means (black line) and 95% HPD intervals (gray shading) of the volatilities $\sigma_{k,(c),t}^2$ from model (8) in the main paper.

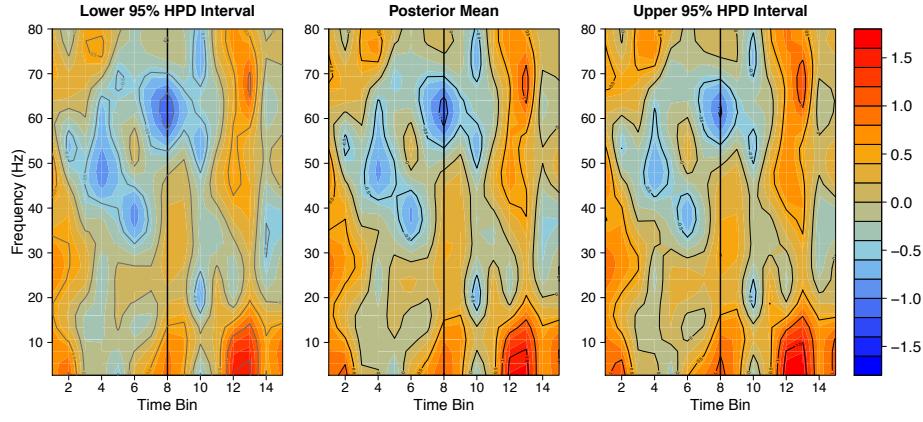


Figure E.2: Pointwise 95% HPD intervals and the posterior mean for $\bar{\mu}_t^{(1)}$, which is the average difference in the PFC log-spectra between the FC and FS trials. The black vertical lines indicate the event time t^* .

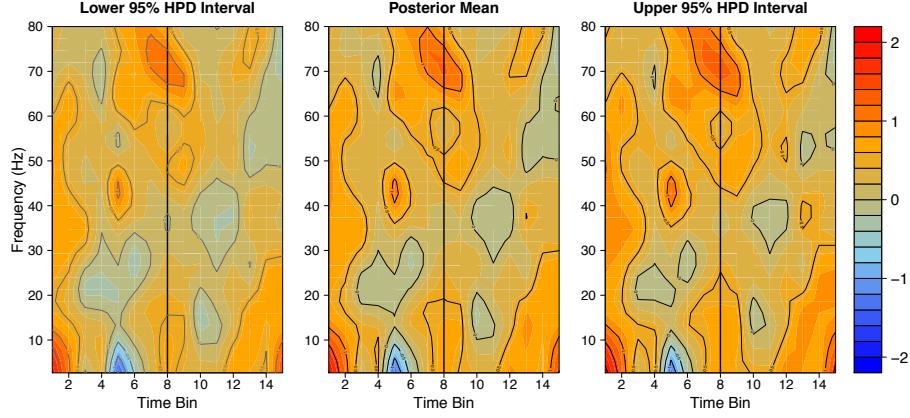


Figure E.3: Pointwise 95% HPD intervals and the posterior mean for $\bar{\mu}_t^{(2)}$, which is the average difference in the PFC log-spectra between the FC and FS trials. The black vertical lines indicate the event time t^* .

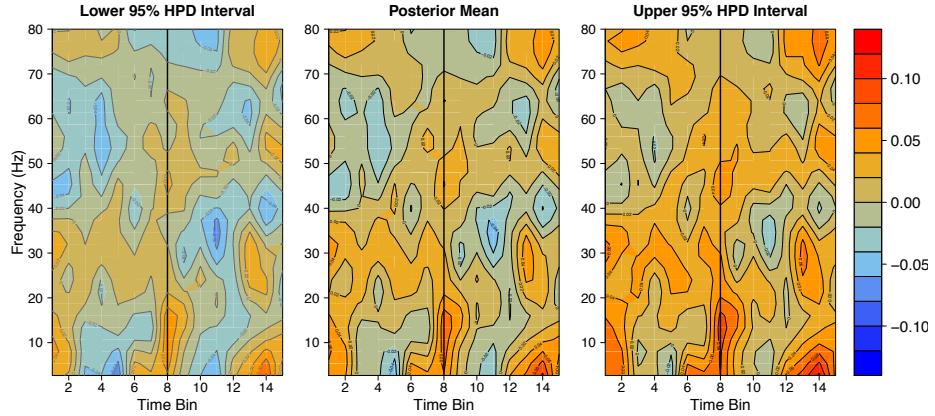


Figure E.4: Pointwise 95% HPD intervals and the posterior mean for $\bar{\mu}_t^{(3)}$, which is the average difference in squared coherence between the FC and FS trials. The black vertical lines indicate the event time t^* .

References

- 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. and Jeliazkov, I. (2001). Marginal likelihood from the Metropolis–Hastings output. *Journal of the American Statistical Association*, 96(453):270–281.
- Daníelsson, J. (1998). Multivariate stochastic volatility models: estimation and a comparison with VGARCH models. *Journal of Empirical Finance*, 5(2):155–173.
- Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–616.
- Gamerman, D. and Migon, H. S. (1993). Dynamic hierarchical models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 629–642.
- Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M. (2010). *Handbook of spatial statistics*. CRC press.
- 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.
- Kastner, G. (2015). stochvol: Efficient Bayesian inference for stochastic volatility (SV) models. *R package version*, 1(0).
- Kastner, G. and Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, 76:408–423.
- 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.

- Koopman, S. J. and Durbin, J. (2000). Fast filtering and smoothing for multivariate state space models. *Journal of Time Series Analysis*, 21(3):281–296.
- Koopman, S. J. and Durbin, J. (2003). Filtering and smoothing of state vector for diffuse state-space models. *Journal of Time Series Analysis*, 24(1):85–98.
- Matteson, D. S., McLean, M. W., Woodard, D. B., and Henderson, S. G. (2011). Forecasting emergency medical service call arrival rates. *The Annals of Applied Statistics*, 5(2B):1379–1406.
- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R news*, 6(1):7–11.
- Wand, M. and Ormerod, J. (2008). On semiparametric regression with O’Sullivan penalized splines. *Australian & New Zealand Journal of Statistics*, 50(2):179–198.