I.F. Some Time Series Models (Dynamic Linear Models)

(Not in RWC)

- I will describe so-called dynamic linear models, sometimes called Kalman filters.
- These kinds of models can be used two ways:

1. **As filters:** In (for example) inertial navigation systems, sensors measure (with error) current location, velocity, etc. and combine it with updated estimates of the previous location, etc. to produce a filtered estimate. Purpose: improved current estimate.

2. **As smoothers:** Given a "complete" time series, the model is used to smooth estimates of the underlying state of the measured system at each time point. Purpose: smoothing.

We will talk about use #2 (smoothing)

IF 1 2/18/08 revised 2/18/10
Example: Global mean surface temperature data.

Temperature deviations $Y_t, t = 0, ..., 124$ modeled as:
(Lui, HKC 2010)

$$Y_t = \mu_t + \eta_t, t = 0, ..., 124$$

**OBSERVATION EQUATION**

Local level

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \omega_{1,t}, t = 1, ..., 124$$

**STATE EQUATIONS.**

Local trend

$$\beta_t = \beta_{t-1} + \omega_{2,t}, t = 1, ..., 123$$

$\eta \sim N(0, \sigma_\eta^2 I_{125})$, $\omega_1 \sim N(0, \sigma_1^2 I_{124})$, $\omega_2 \sim N(0, \sigma_2^2 I_{123})$

The state of this system evolves through time according to the state equation.

Observations of the system relate to the state of the system according to the observation equation.

If $\omega_1 = 0_{124}$ and $\omega_2 = 0_{123}$, this model is a simple straight line.

$\sigma_\eta^2, \sigma_1^2$ to allow the level ($\mu_t$) and trend ($\beta_t$) respectively to vary over time.

IF/2 2/18/08 revised 2/18/10
It is easiest to express this in the alternative formulation (constant case)

**OBS. EQN:** \[ y_t = \eta_t \quad + n_t \quad t = 0, \ldots, 124 \]

**STATE EQNS:** \[ 0 = \eta_{t-1} - \eta_t + \beta_{t-1} w_{1,t} \quad t = 1, \ldots, 124 \]

\[ 0 = \beta_{t-1} \eta_{t-1} + w_{2,t} \quad t = 1, \ldots, 123 \]

Assembled in the form of a linear model:

\[
\begin{bmatrix}
Y_t^n \\
O_{123} \\
0_{124}
\end{bmatrix} = \begin{bmatrix}
I_{125} & 0_{125 \times 124} \\
1-1 0 \cdots 0 0 & 1 0 0 \cdots 0 0 \\
0 1-1 \cdots 0 0 & 0 1 0 \cdots 0 0 \\
\vdots & \vdots \\
0 0 0 \cdots 1-1 & 0 0 0 \cdots 1-1 \\
0_{123 \times 125} & 1-1 0 \cdots 0 0 & 0 1-1 \cdots 0 0 & 0 0 0 \cdots 1-1
\end{bmatrix} \begin{bmatrix}
\mu_0 \\
\mu_1 \\
\vdots \\
\mu_{123} \\
\eta_{124}
\end{bmatrix} + \begin{bmatrix}
\Gamma \sim W_1 \\
0 \sim W_2 \\
0 \sim W_3
\end{bmatrix}
\]

(Differs slightly from Cui, HKC (2010))

IF \( b \) 2/18/08 revised 2/18/10
With a bit of cleverness and a re-parameterization, this can be written as a MLM (C, H, K, 2010)

\[ \beta_1 = \beta_0 + \omega_{3,1} \rightarrow \mu_1 = \mu_0 + \omega_{1,1} + \beta_0 \]

\[ \beta_2 = \beta_1 + \omega_{2,2} = \beta_0 + \frac{2}{i} \omega_{2,i} \rightarrow \mu_2 = \mu_1 + \omega_{1,2} + \beta_1 \]

\[ \beta_3 = \beta_2 + \omega_{2,3} = \beta_0 + \frac{3}{i} \omega_{2,i} \rightarrow \mu_3 = \mu_2 + \omega_{1,3} + \beta_2 \]

\[ \mu_t = \mu_0 + t\beta_0 + \sum_{i=1}^{t} \omega_{1,i} + \sum_{i=1}^{t} (t-i)\omega_{2,i} \]

\[ \text{fixed effects} \quad \text{Random effects} \]
In something closer to the ML formulation:

\[
\begin{pmatrix}
  y_0 \\
  y_1 \\
  y_2 \\
  \vdots \\
  y_T
\end{pmatrix}
= 
\begin{pmatrix}
  1 & 0 \\
  1 & 1 \\
  1 & 2 \\
  \vdots & \vdots \\
  1 & T
\end{pmatrix}
\begin{pmatrix}
  \mu_0 \\
  \beta_0 \\
  \beta_1 \\
  \beta_2 \\
  \vdots \\
  \beta_T
\end{pmatrix}
+ 
\begin{pmatrix}
  1 & 0 & 0 & \cdots & 0 \\
  1 & 1 & 0 & \cdots & 0 \\
  1 & 1 & 1 & \cdots & 0 \\
  \vdots & \vdots & \vdots & \vdots & \vdots \\
  1 & 1 & 1 & \cdots & 1
\end{pmatrix}
\begin{pmatrix}
  w_{1,1} \\
  w_{1,2} \\
  \vdots \\
  w_{1,T}
\end{pmatrix}
+ 
\begin{pmatrix}
  0 & 0 & 0 & \cdots & 0 \\
  0 & 0 & 0 & \cdots & 0 \\
  2 & 1 & 0 & \cdots & 0 \\
  \vdots & \vdots & \vdots & \vdots & \vdots \\
  T-1 & T-2 & T-3 & \cdots & 1
\end{pmatrix}
\begin{pmatrix}
  w_{2,1} \\
  w_{2,2} \\
  \vdots \\
  w_{2,T-1}
\end{pmatrix}
+ 
\begin{pmatrix}
  \varepsilon
\end{pmatrix}
\]

\( \text{cov}(\varepsilon) = R = \sigma^2 I_{T+1} \)

\( \mathbf{u} = \left( \begin{array}{c} \mathbf{w}_1 \\ \mathbf{w}_2 \end{array} \right) \),

\( \text{cov}(\mathbf{u}) = 
\begin{bmatrix}
  \sigma^2 I_T \\
  0 \\
  0 \\
  \sigma^2 I_{T-1}
\end{bmatrix}
\)

Note: Complete collinearity in Z's two sets of columns.
More general DLM:

\[ Y_t = F_t \theta_t + n_t \]

\[ \theta_t = G_t \theta_{t-1} + w_t \]

\[ \sigma_{p \times p} \]

\[ y_t \text{ r-dimensional} \]

\[ F_{r \times p}, \theta_{p \times 1} \]

\[ n_t \sim N_n(0, \Sigma_t) \]

\[ w_t \sim N_p(0, \Sigma_t) \]

This framework allows a huge class of models, including flexible seasonal effects, covariate effects, interaction effects ...

West M, Harrison J "Bayesian Forecasting and Dynamic Models" (2nd ed 1997, Springer) gives an encyclopedic treatment of these models mainly as filters and, by extension, forecasting tools.

Previous model: \[ \theta_t = (n_t) \quad F_t = [1, 0] \]

\[ G_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \]

\[ \sigma_{1 \times 1} \]

r=1, p=2

IF/014 2/18/08 revised 2/8/10
Example: Testing a new method to localize epileptic activity during surgery (M. Laux, D. Hochman, M. Hugend) manuscript
$y_t = \text{intensity of pixel at time } t = 1, \ldots, 650 \text{ (interval 0.28 sec)}$

(535nm)

Stimulation starts at step 75, ends at step 94

**OBSERVATION EQUATION**

$y_t = s_t + h_t + r_t + \epsilon_t$

- $s_t$: smoothed response
- $h_t$: heartbeat
- $r_t$: respiration
- $\epsilon_t$: iid error

(Stimulus detectable?)

Here are 100 time steps from one person's data:

![Graph of time steps](image)

IF/14a 2/1810
Observation equation

\[ y_t = S_t + h_t + r_t + v_t \]

smoothed response beat beat respiration error.

Each piece has its own state equation

\[
\begin{pmatrix}
S_t \\
\text{slope}_t
\end{pmatrix}
= \begin{pmatrix}
1 & 1 \\
0 & 1
\end{pmatrix}
\begin{pmatrix}
S_{t-1} \\
\text{slope}_{t-1}
\end{pmatrix}
+ \begin{pmatrix}
W_{s,t} \\
W_{\text{slope},t}
\end{pmatrix}
\\
W_{s,t} \equiv 0
\\
W_{\text{slope},t} \sim \text{iid } N(0, \sigma_5)
\]

Each cyclic term looks like this:

\[
\begin{pmatrix}
b_t \cos d_t \\
b_t \sin d_t
\end{pmatrix}
= \begin{pmatrix}
\cos \delta & \sin \delta \\
-\sin \delta & \cos \delta
\end{pmatrix}
\begin{pmatrix}
b_{t-1} \cos d_{t-1} \\
b_{t-1} \sin d_{t-1}
\end{pmatrix}
+ \begin{pmatrix}
W_{1,t} \\
W_{2,t}
\end{pmatrix}
\\
W_{1,t} \sim \text{iid } N(0, \sigma_6)
\\
W_{2,t} \sim \text{iid } N(0, \sigma_7)
\]

\[ \delta = \frac{2\pi}{\text{period}} \] . If there's no error, this equation describes a circle of radius \( b_t \equiv b \), where each step is an angle of \( \delta \). With the errors \( b_t \neq b_{t-1} \), \( d_t \neq d_{t-1} + \delta \), and this accommodates functions that aren't sinusoidal and non-constant amplitude/period.

\[ h_t = b_t \cos d_t \]
\[ r_t = b_t \cos d_t \]

IF old 2/8/10
In terms of the notation $y_t = F_t \theta_t + n_t$

$$\theta_t = \theta_{t-1} + \omega_t$$

$$y_t = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} + v_t$$

$$F_t$$

$$h_t = \begin{bmatrix} \text{slope}_t \\ h_t \cos \phi_t \\ b_t \sin \phi_t \\ r_t \cos \phi_t \\ b_t \sin \phi_t \\ b_t \sin \phi_t \end{bmatrix}$$

$$V_t \sim \mathcal{N}(0, \sigma^2)$$

$$G_t = \begin{bmatrix} 1 & 1 & 0 & 1 & \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \end{bmatrix} \\ 0 & 1 & \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} \cos \delta_h & \sin \delta_h \end{bmatrix} & \begin{bmatrix} \cos \delta_h & \sin \delta_h \end{bmatrix} & \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} \cos \delta_r & \sin \delta_r \end{bmatrix} & \begin{bmatrix} \cos \delta_r & \sin \delta_r \end{bmatrix} & \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} \cos \delta_r & \sin \delta_r \end{bmatrix} & \begin{bmatrix} \cos \delta_r & \sin \delta_r \end{bmatrix} & \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} \cos \delta_r & \sin \delta_r \end{bmatrix} & \begin{bmatrix} \cos \delta_r & \sin \delta_r \end{bmatrix} & \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \end{bmatrix} \end{bmatrix}$$

$S_h = \frac{2\pi}{2.78}$

$S_r = \frac{2\pi}{18.75}$

$\omega(\omega_t) = \text{diag}(0, \omega_s, \omega_h, \omega_r, \omega_r)$
Here's the fit, with the pieces of it.
This fits easily into the alternative formulation (Duncan & Horn, JASA 1972 [17]), and into the MLM as follows:

\[ y_0 = F_0 \theta_0 + n_0 \]

\[ \theta_1 = G_1 \theta_0 + w_1 \rightarrow y_1 = F_1 \theta_1 + n_1 = F_1 G_1 \theta_0 + F_1 w_1 + n_1 \]

\[ \theta_2 = G_2 \theta_1 + w_2 \rightarrow y_2 = F_2 \theta_2 + n_2 = F_2 G_2 \theta_0 + F_2 G_2 w_1 + F_2 w_2 + n_2 \]

\[ \theta_3 = \ldots \rightarrow y_3 = F_3 \theta_3 + n_3 = F_3 G_3 \theta_0 + F_3 G_3 w_1 \]

\[ \text{Dot dot dot} \]

So \[ y_t = F_t \theta_t + n_t = F_t G_t \ldots G_1 \theta_0 + F_t G_t \ldots G_t w_t \]

\[ + \sum_{i=1}^{t-1} F_t G_t \ldots G_i w_i + F_t w_t + n_t \]

\[ \text{Final effect} \quad \theta_0 \]

\[ \text{Random effects} (w_t) \]
**Alternative syntaxes for RPMs**

**Main syntax**: Mixed linear models.

**Key idea**: Express a large class of models as mixed linear models.

**Key tools**:  
- Mixed linear model theory, methods, and computing, and ideas adapted from simple linear models.  
- The conventional analysis uses the restricted likelihood, large-sample approximations, and bootstrapping.  
- Bayesian methods rely on MCMC.
**Alternative #1**: Gaussian Markov random fields (Rue & Held 2005).

**Key idea**: Represent components of models and priors as Gaussian Markov random fields (GMRFs), using conditional dependence.

**Key tools:**

- Model the mean structure in a modular fashion, with components being GMRFs or simple effects (fixed effects).
- Bayesian analyses only; non-Bayesian analyses are possible (at least, sometimes)
- "Exact" analyses use MCMC that exploits sparse precision matrices.
- Approximate analyses use integrated nested Laplace approximation (INLA).
- Many models can be represented in this syntax, at worst closely analogous models to MLM models.
Alternative #2: Likelihood inference for models with unobservables (Lee et al 2006).

Key ideas: Extend generalized linear models in several directions, using likelihood-like functions.

Key tools:

• Modular modelling of the observation error distribution (exponential family), the linear predictor, error dispersion, random effects dispersion.

• Can handle other unobservable random variables, e.g., missing data or predictions.

• Analysis: Estimates are maxima of likelihood-like functions; uncertainty is described using curvatures at maxima.
Key idea: Conditional dependence/independence $\iff$ Precision matrix

If $\mathbf{x} = (x_1, \ldots, x_n) \sim \text{Normal mean } \mu$

Precision matrix $Q$, not necessarily invertible

Then $\text{Cov}(x_i, x_j \mid x_{-ij}) = 0 \iff Q_{ij} = 0$

Examples

One way RE model $y_{ij} = \Theta_i + \varepsilon_{ij}, \Theta_i = \mu_i + \delta_i, \varepsilon_{ij}, \delta_i$ independent

$\text{Cov}(y_{ij}, \Theta_{i'} \mid \Theta_i) = 0 \quad \text{if } i \neq i'$

$\text{Cov}(y_{ij}, \mu \mid \Theta_i) = 0$

$\text{Cov}(\Theta_i, \Theta_{i'} \mid \mu) = 0$

ICAR model $y_i = \Theta_i + \varepsilon_i$ for regions $i = 1, \ldots, n$

$\Theta_i \sim \text{CAR}(\tau, Q)$

$\text{Cov}(y_i, \Theta_{i'} \mid \Theta_i) = 0$ if $i \neq i'$

$\text{Cov}(\Theta_i, \Theta_{i'} \mid \Theta_{-ii}) = 0$ if $i$ and $i'$ are not neighbors.

$\text{Cov}(\Theta_i, Q_{ij} \mid \Theta_{-ij}) = -\tau_i$ if $i, j$ neighbors

Recall $\tau_i Q$ has diagonals $\tau_i 2s$

off diagonals $\{-\tau_i \mid (i, j) \text{ neighbors} \}$

$0 \quad \text{if not}$

$\text{Cov}(y, \Theta \mid \mu) = 0$

$\text{Cov}(\Theta, \Theta \mid \mu) = 0$

$\text{Var}(\Theta) = \tau_i 2s$

$\text{Var}(y) = \tau_i 2s$

$I_{6\times6} \quad 2/8/10$
Auto-regressive model \( X_t = \phi X_{t-1} + \epsilon_t \) and \( \epsilon_t \sim \text{N}(0,1) \), \( |\phi| < 1 \)

\( X_t | x_1, \ldots, x_{t-1} \sim \text{N}(\phi x_{t-1}, 1) \); \( X_t | x_{t-1} \) is conditionally independent of \( x_1, \ldots, x_{t-2} \)

If the marginal distribution of \( x_t \) is \( \text{N}(0, \sigma_x^2 = (1-\phi^2)^{-1}) \)

\( X \) is a G-MRF with \( \Omega = \begin{bmatrix} 1-\phi & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & \cdots & 0 & 0 & 0 \\ 0 & -\phi & 1+\phi^2 & -\phi & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1-\phi & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1-\phi \end{bmatrix} \)

This precision matrix is mostly 0's.

Dynamic linear models

\( y_t = F_t \theta_t + n_t \)

\( \theta_t = C_t \theta_{t-1} + w_t \)

\( \text{cov}(y_{t'}, \theta_t | \theta_t) = 0 \text{ if } t' \neq t \)

\( \text{cov}(\theta_t, \theta_{t'} | \theta_{t'}) = 0 \text{ if } t' < t-1 \)

\( \text{cov}(y_{t'}, \theta_{t'} | \theta_{t-1}, \theta_{t+1}) = 0 \text{ if } t' \neq (t-1, t, t+1) \)
Lots of models fit into this framework, including models with non-normal observables—put "some of" before most classes below.

- Simple REs: - GMRF with diagonal
- Time series: AR models are GMRFs
- State-space models (DLMs or KFs)
- Longitudinal analyses: - see time series
- Graphical models: graphs represent conditional dependence structure
- Semi-parametric regressions/splines: may require some approximating
- Image analysis
- Spatial analysis

This is technically complicated and I don't understand it. I believe it's true GPs can be represented as GMRFs so that the analysis of a given dataset in this representation is identical to the analysis in the original GP formulation, but that other things people do with GPs, e.g., interpolation, are not necessarily the same.
This grows out of generalized L.M.S, which use:
- Error distribution (1-param exponential)
- Linear predictor (link function) mean germfdis
- Analysis by ML, large-sample approximate theory and IRLS for computing.

Lee et al (2006) extends this by:
- Adding random effects to the linear predictor
- Modeling the dispersion parameter of the error dist
  - its own GLM, can include random effects
  - estimated simultaneously
- Includes other "unobservables," e.g. missing data, prediction
- Analysis uses the so-called h-likelihood, so that a model with all possible pieces is analyzed as a series of GLMs.
- Commentators:
  - This includes some new models and unifies some existing models as well.
  - Mainly disagree with Lee et al. about the value of their unified analytic approach.

- Analytic approach:
  - Model syntax
  - Computing method
  - Theory of analysis (controversial)

- Lee et al. on their theory of analysis:
  - Principled, based on probabilistic models & likelihood.
  - "Avoids prior probabilities" => superior to Bayes.
  - Solves all problems in analysis apart from minor technical issues they can solve with 2nd order approximations.
In my view and apparently others' view (see discossants of Lee & Nelder 2009), these claims are ridiculously overstated and cannot be justified. (See esp. Meng's comment.)

Some simple points:

- Their analysis approach is, in fact, an ad hoc patchwork (see quote from L&N 2009 in my back).
- Why do they have to do different things for different unknowns — to avoid Bayes and to avoid known problems.
- Ad hoc is OK if it performs, but this can't perform as well as they claim
  - multiple maxima
  - maxima at boundary values
  - intervals using curvature at maximum rationalized by large sample theory

I have found no mention of any of these problems in LNP (2006) or L&N (2009).
Summing up Part I of the course

- Theory = Syntax for expressing lots of models
  - Tools for understanding analyses of models expressed in that syntax

The right syntax for expressing many models:

- Allows more powerful computing
- Allows theory for many models simultaneously:
  - Linear models (ANOVA, multiple regression)
  - Generalized LMs.
We have the syntax for MLMs.

What do I want in a theory?

The obvious place to start is the things we get from the powerful and beautiful theory of linear models:

1. Find discrepant features of the data (residuals/outsliers)

2. Seek deviations from model assumptions
   - residuals: non-linearity in mean
   - non-constant variance
   - transformations of y

3. Seek data features with large influence on the results (influence diagnostics)

4. Assess evidence for adding predictors (tVar plots)

5. Understand indeterminate results/competition among predictors (collinearity)

I sum 2/20/08
We'll begin by looking at simple extensions of these ideas from linear model theory to MLM.

BUT before we do that ... 

MLMs provide a whole new set of ways to generate mysteries and complications in analyses, and they're much more complicated than linear models.
Therefore, I argue, we need to consider a different *style* for learning about our methods.

We need to use a *scientific* style, alongside the more traditional *mathematical* style.

The next lecture will demonstrate this scientific style on a problem that arises in fitting the "random regressions" model.

The rest of the course will use the math style sometimes and the scientific style other times.