Time Series Modeling with Unobserved Components

Rajesh Selukar

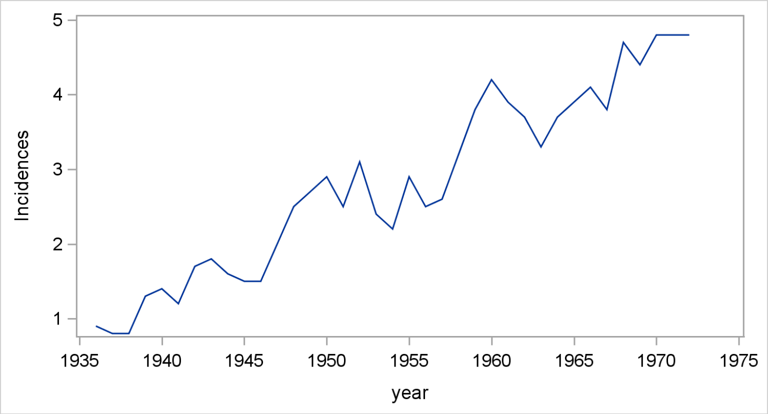
SAS Institute Inc., Cary, NC
rajesh.selukar@sas.com

Unobserved Components Model

- Response Time Series = Superposition of components such as Trend, Seasons, Cycles, and Regression effects
- Each component in the model captures some important feature of the series dynamics.
- Components in the model have their own probabilistic models.
- The probabilistic component models include meaningful deterministic patterns as special cases.

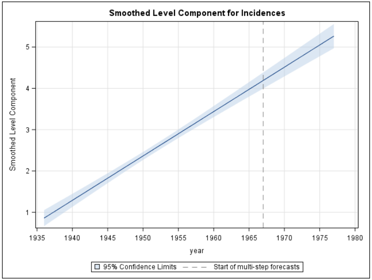
Melanoma Incidences in Connecticut

The age-adjusted numbers of melanoma incidences per 100,000 for the years of 1936 to 1972 (from Connecticut Tumor Registry):

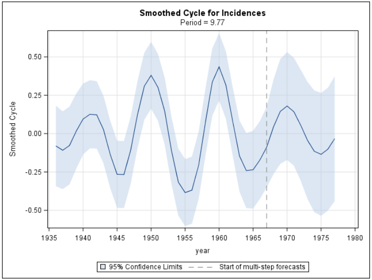


Melanoma Incidences = Trend + Cycle + Irregular

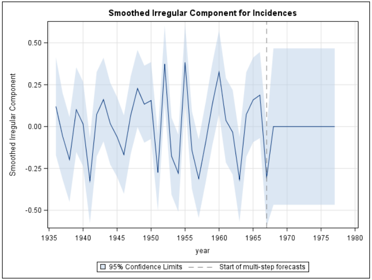
Estimated trend $\mu_t = 0.75 + 0.11t$



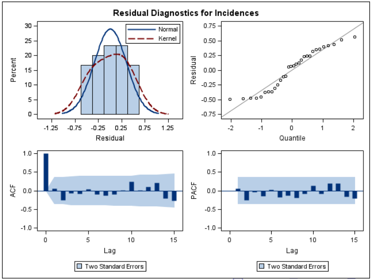
Estimated cycle component



Estimated irregular component

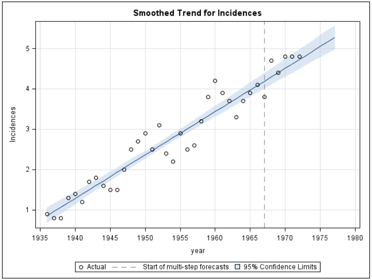


Residual diagnostics

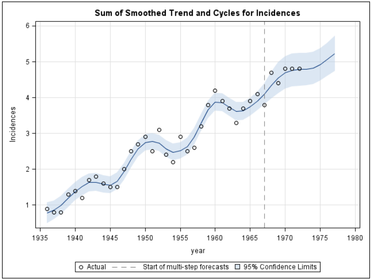


How Did the Components Add-Up

The estimated linear trend

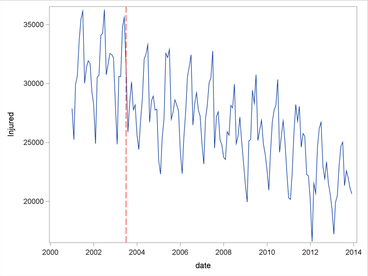


Trend + Cycle

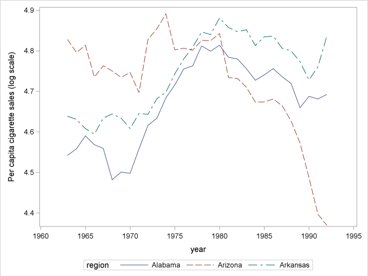


More Examples of Time Series

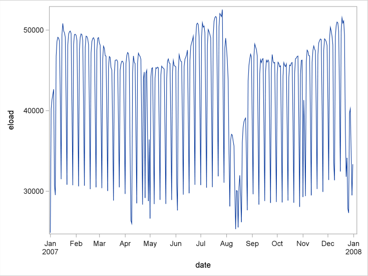
Monthly traffic injuries in Italy



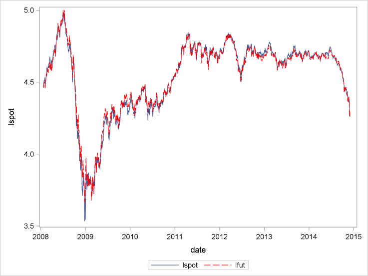
Region-wise yearly per-capita cigarette sales



Hourly electricity load at 10 am during 2007-2008 in Italy



Daily spot price (lspot) and future contract (lfut) for Brent crude



Unobserved Components Model

$$\mathbf{Y_t} = \mathbf{X_t}\beta + \boldsymbol{\mu}_\mathbf{t} + \boldsymbol{\psi}_\mathbf{t} + \ldots + \boldsymbol{\epsilon}_\mathbf{t}$$

- Univariate or multivariate response at time t: $\mathbf{Y_t}$
- Effect of regression variables: $\mathbf{X_t}\beta$
- Time varying mean (level/trend): $\boldsymbol{\mu}_t$
- Periodic/Seasonal component: $\boldsymbol{\psi}_t$
- Noise component: $\boldsymbol{\epsilon}_t$
- A component could be turned on/off, or scaled, based on an external input, e.g., the trend could be scaled as $b_t\ \boldsymbol{\mu}_t$
- All of these components need not be present in a UCM
- Many more types of components are often needed/used

Local Linear Trend (LLT)

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_{t-1} + \boldsymbol{\nu}_t \qquad \text{Level equation}$$
$$\boldsymbol{\eta}_t = \boldsymbol{\eta}_{t-1} + \boldsymbol{\xi}_t \qquad \text{Slope equation}$$

- $\boldsymbol{\nu}_t \sim N(0, \Sigma_\nu)$ are i.i.d. disturbances in the level equation
- $\boldsymbol{\xi}_t \sim N(0, \Sigma_\xi)$ are i.i.d. disturbances in the slope equation
- The initial level $\boldsymbol{\mu}_1$ and the initial slope $\boldsymbol{\eta}_1$ are (usually) unknown vectors

LLT in a vector recursion form:

$$\begin{bmatrix} \boldsymbol{\mu}_t \\ \boldsymbol{\eta}_t \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_{t-1} \\ \boldsymbol{\eta}_{t-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\nu}_t \\ \boldsymbol{\xi}_t \end{bmatrix}$$

Even the simple LLT + Noise model, $\mathbf{Y_t} = \boldsymbol{\mu}_\mathbf{t} + \boldsymbol{\epsilon}_\mathbf{t}$, turns out to be a very versatile model.

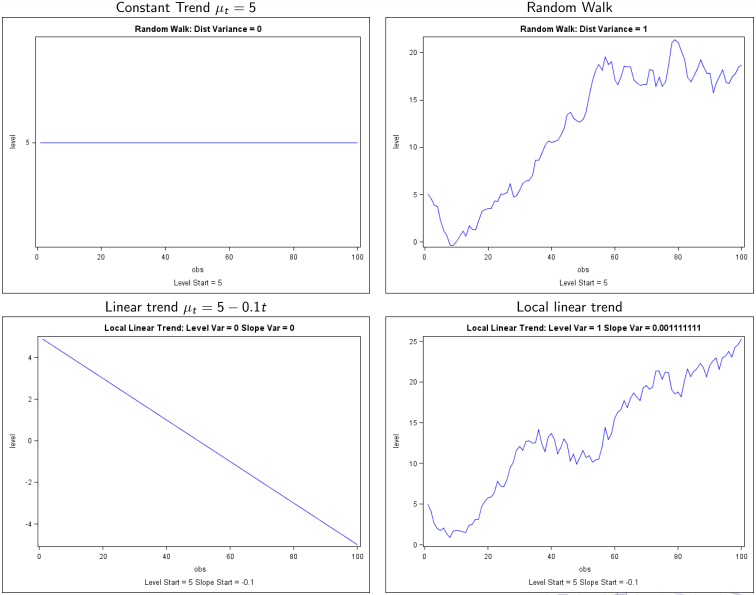
Some Special Cases of LLT

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_{t-1} + \boldsymbol{\nu}_t \qquad \text{Level equation}$$

$$\boldsymbol{\eta}_t = \boldsymbol{\eta}_{t-1} + \boldsymbol{\xi}_t \qquad \text{Slope equation}$$

- Random walk $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\nu}_t$ (initial slope $\boldsymbol{\eta}_1$ and the slope disturbance covariance Σ_ξ are zero).
- Random walk with drift $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_1 + \boldsymbol{\nu}_t$ (the slope disturbance covariance Σ_ξ is zero).
- Integrated random walk $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_{t-1}; \;\; \boldsymbol{\eta}_t = \boldsymbol{\eta}_{t-1} + \boldsymbol{\xi}_t$ (the level disturbance covariance Σ_ν is zero).
- Deterministic time trend $\boldsymbol{\mu}_t = \boldsymbol{\mu}_1 + t\,\boldsymbol{\eta}_1$ (the level disturbance covariance Σ_ν and the slope disturbance covariance Σ_ξ are zero).
- Time invariant mean $\boldsymbol{\mu}_t = \boldsymbol{\mu}_1$ ($\boldsymbol{\eta}_1$, Σ_ξ, and Σ_ν are all zero).

Simulated Univariate Trends



A Recursive Formula For Cycle

For $t = 1, 2, \ldots$, and $0 < \omega < \pi$,

$$\psi_t = a\ \cos(\omega t) + b\ \sin(\omega t)$$

is a cycle with period $2\pi/\omega$, amplitude $\sqrt{a^2 + b^2}$, and phase $\arctan(b/a)$. That is

$$\psi_t = \gamma \cos(\omega t - \phi), \quad \gamma = \sqrt{a^2 + b^2}, \;\; \phi = \arctan(b/a)$$

You can verify that ψ_t satisfies the recursion

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \begin{bmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix}$$

when $\psi_0 = a$ and $\psi_0^* = b$. Moreover, $\psi_t^2 + \psi_t^{*2} = a^2 + b^2, \;\; \forall t$.

A Recursive Formula For Stochastic Cycle

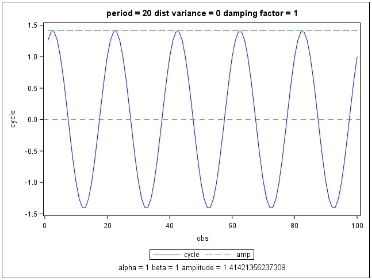
A stochastic generalization of the cycle can be obtained by adding random noise to the cycle recursion and by introducing a damping factor, ρ, for additional modeling flexibility

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \nu_t \\ \nu_t^* \end{bmatrix}$$

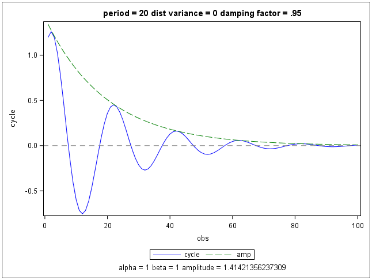
where $0 \leq \rho \leq 1$, and the disturbances ν_t and ν_t^* are independent $N(0, \sigma_\nu^2)$ variables.
The resulting random sequence ψ_t is pseudo-cyclical with time-varying amplitude, phase, and frequency (period).

Simulated Cycles

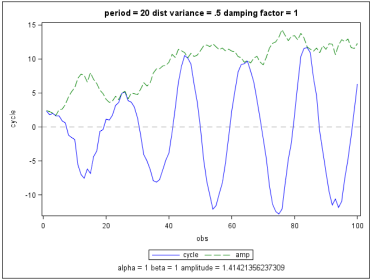
Undamped deterministic cycle



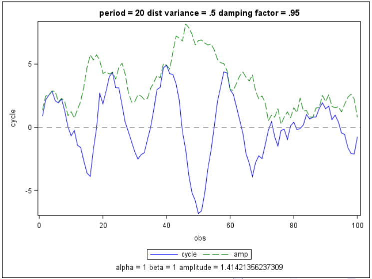
Damped deterministic cycle



Undamped stochastic cycle



Damped stochastic cycle



Stochastic Cycle: Review

$$\boldsymbol{\psi}_t = \rho^t \mathbf{R}_\omega^t \boldsymbol{\psi}_0 + \sum_{j=0}^{t} \rho^{t-j} \mathbf{R}_\omega^{t-j} \boldsymbol{\nu}_j, \quad \mathbf{R}_\omega = \begin{bmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{bmatrix}$$

- If $\rho < 1$, the effect of initial condition and the shocks in the distant past becomes negligible. ψ_t has a stationary distribution with mean zero and variance $\sigma_\nu^2/(1-\rho^2)$.
- If $\rho = 1$, the effect of shocks persists and ψ_t is non-stationary.
- Cycles are very useful as building blocks for constructing more complex periodic patterns. Periodic patterns of almost any complexity can be created by superimposing cycles of different periods and amplitudes. In particular, the seasonal patterns, which are general periodic patterns with integer periods, can be constructed as sums of cycles.

Modeling Seasons

- The seasonal fluctuations are a common source of variation in the time-series data
- The seasonal effects are regarded as corrections to the general trend of the series due to seasonal variations, and these effects sum to zero when summed over the full season cycle
- Therefore, a (deterministic) seasonal component γ_t is a periodic pattern of an integer period s such that the sum

$$\sum_{i=0}^{s-1} \gamma_{t-i} = 0, \quad \forall t$$

Two Representations of Seasonal Pattern (Period = s)

- As a list of s numbers that sum to zero
- As a sum of $[s/2]$ deterministic, undamped cycles, called harmonics, of periods s, $s/2$, $s/3$, ...
  - Here $[s/2] = s/2$ if s is even and $[s/2] = (s-1)/2$ if s is odd.
  - Example: For s = 12, the seasonal pattern can always be written as a sum of six cycles with periods 12, 6, 4, 3, 2.4, and 2.

Stochastic Seasonal: Dummy Type

$$\sum_{i=0}^{s-1} \gamma_{t-i} = \nu_t, \quad \nu_t \sim N(0, \sigma_\nu^2)$$

- The periodic pattern sums to zero in the mean.
- The disturbance variance controls the variation in the seasons. If it is zero the model reduces to a deterministic seasonal, equivalent to having (s -1) dummy regressors.

Stochastic Seasonal: Trigonometric Type

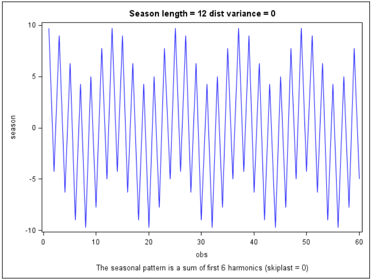
$$\gamma_t = \sum_{j=1}^{[s/2]} \psi_{j,t}$$

where the stochastic cycles $\psi_{j,t}$ have periods $p_j = s/j$.

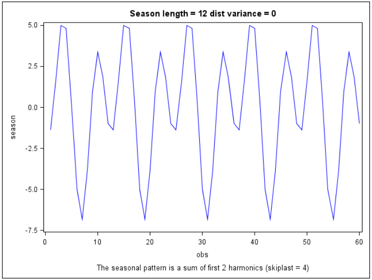
- Here, all the cycles are un-damped, and usually have a common disturbance variance σ_ν^2.
- You can create custom seasonal patterns by dropping some of the harmonics and by judiciously choosing their disturbance variances.
- If all the disturbance variances are zero, the pattern reduces to a deterministic seasonal.

Simulated Seasons with Period = 12

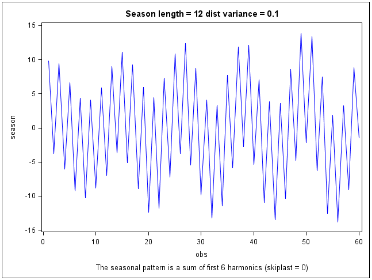
Deterministic Season with All Harmonics



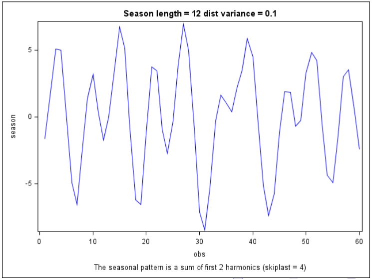
Deterministic Season with First Two Harmonics



Stochastic Season with All Harmonics



Stochastic Season with First Two Harmonics



UCMs and SSMs

- All the unobserved component models (UCMs) discussed in this workshop can also be formulated as (linear) state space models (SSMs).
- An SSM is a dynamic version of the linear regression model where the regression vector evolves with time in a Markovian fashion.
- The SSM formulation of a UCM enables the use of the famous Kalman filter/smoother (KFS) algorithm for UCM based data analysis.

State Space Model and Notation

$$
\begin{aligned}
\mathbf{Y}_t &= \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{X}_t \beta + \boldsymbol{\epsilon}_t && \text{Observation equation} \\
\boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{W}_{t+1} \gamma + \zeta_{t+1} && \text{State transition equation} \\
\boldsymbol{\alpha}_1 &= \mathbf{A}_1 \delta + \mathbf{W}_1 \gamma + \zeta_1 && \text{Initial condition}
\end{aligned}
$$

- Response values y and predictor vectors $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ are recorded at $t = 1, 2, \cdots, n$.
- Number of measurements at $t = p$, say. $\mathbf{Y}_t$ and $\mathbf{X}_t$ denote the vector and matrix formed by vertically stacking y values and $\mathbf{x}$ vectors at t. $Dim(\mathbf{Y}_t) = p$, and $Dim(\mathbf{X}_t) = p \times k$. Similarly, $\mathbf{W}_t$ contains regressor values used in the state equation.
- SSM form is not unique; many equivalent alternate forms are possible.

Latent Quantities in the Model

$$\mathbf{Y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\epsilon}_t \qquad \text{Observation equation}$$

$$\boldsymbol{\alpha}_{t+1} = \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{W}_{t+1} \boldsymbol{\gamma} + \boldsymbol{\eta}_{t+1} \qquad \text{State transition equation}$$

$$\boldsymbol{\alpha}_1 = \mathbf{A}_1 \boldsymbol{\delta} + \mathbf{W}_1 \boldsymbol{\gamma} + \boldsymbol{\eta}_1 \qquad \text{Initial condition}$$

Vector	Dim	Description
$\boldsymbol{\alpha}_t$	m	State vectors
$\boldsymbol{\beta}$	k	Regression vector in the observation equation
$\boldsymbol{\gamma}$	g	Regression vector in the state equation
$\boldsymbol{\delta}$	d	Diffuse part of $\boldsymbol{\alpha}_1$
$\boldsymbol{\epsilon}_t$	p	Observation noise (zero-mean, Gaussian)
$\boldsymbol{\eta}_t$	m	State noise (zero-mean, Gaussian)

- Noise/shock/disturbance variables $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ are mutually independent white noise sequences (possibly with time-varying covariances).

Model System Matrices

$$\mathbf{Y}_t = \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{X}_t\beta + \boldsymbol{\epsilon}_t \qquad \text{Observation equation}$$

$$\boldsymbol{\alpha}_{t+1} = \mathbf{T}_t\boldsymbol{\alpha}_t + \mathbf{W}_{t+1}\gamma + \boldsymbol{\eta}_{t+1} \qquad \text{State transition equation}$$

$$\boldsymbol{\alpha}_1 = \mathbf{A}_1\delta + \mathbf{W}_1\gamma + \boldsymbol{\eta}_1 \qquad \text{Initial condition}$$

Matrix	Dim	Description
$\mathbf{Z}_t$	$p \times m$	Design matrix for $\boldsymbol{\alpha}_t$
$\mathbf{T}_t$	$m \times m$	State transition matrix
$\mathbf{A}_1$	$m \times d$	Diffuse condition specifier made up of 0's and 1's
$Cov(\boldsymbol{\epsilon}_t)$	$p \times p$	Often diagonal
$Cov(\boldsymbol{\eta}_t)$	$m \times m$	Often nondiagonal

- Missing elements are not allowed in any system matrix. However, the system matrices can depend on some unknown parameter vector $\boldsymbol{\theta}$ (which must be estimated first for the model to be practically useful).

SSM Form of the Melanoma Incidences UCM

Suppose y denotes the incidences and μ_t, ψ_t, and ϵ_t are the local linear trend, stochastic cycle, and random noise, respectively.

$$y_t = \mu_t + \psi_t + \epsilon_t$$

This model can be expressed as

$$\begin{aligned} y_t &= \mathbf{Z}\boldsymbol{\alpha}_t + \epsilon_t && \text{Observation equation} \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{T}\boldsymbol{\alpha}_t + \boldsymbol{\zeta}_{t+1} && \text{State transition equation} \\ \boldsymbol{\alpha}_1 &= \mathbf{A_1}\boldsymbol{\delta} + \boldsymbol{\zeta}_1 && \text{Initial condition} \end{aligned}$$

where $\boldsymbol{\alpha}_t = [\mu_t\ \eta_t\ \psi_t\ \psi_t^*]$, $\mathbf{Z} = [1\ 0\ 1\ 0]$, $\boldsymbol{\zeta}_t = [\nu_t\ \xi_t\ \nu_t\ \nu_t^*]$, $\boldsymbol{\delta} = [\mu_1\ \eta_1]$, $\mathbf{A_1} = [1\ 0;\ 0\ 1;\ 0\ 0;\ 0\ 0]$, $\boldsymbol{\zeta}_1 \sim N(\mathbf{0},\ [0,\ 0,\ \sigma_\nu^2/(1-\rho^2),\ \sigma_\nu^2/(1-\rho^2)])$, and $\mathbf{T} = [1\ 1\ 0\ 0;\ 0\ 1\ 0\ 0;\ 0\ 0\ \rho\cos(\omega)\ \rho\sin(\omega);\ 0\ 0\ -\rho\sin(\omega)\ \rho\cos(\omega)]$.

Informal Description of the Kalman Filter (KF)

- Assume that the model parameter vector $\boldsymbol{\theta}$ is known.
- KF recursively computes the one-step-ahead predictions of the response values and the latent quantities.
- Let $\mathbf{DATA}_t$ denote all the data up to time t.
- KF recursively computes:

$$\begin{array}{ll} \hat{\mathbf{Y}}_t = \mathrm{E}(\mathbf{Y}_t|\mathbf{DATA}_{t-1}) & \mathbf{F}_t = \mathrm{COV}(\mathrm{E}(\mathbf{Y}_t|\mathbf{DATA}_{t-1})) \\ \hat{\boldsymbol{\alpha}}_t = \mathrm{E}(\boldsymbol{\alpha}_t|\mathbf{DATA}_{t-1}) & \mathbf{P}_t = \mathrm{COV}(\mathrm{E}(\boldsymbol{\alpha}_t|\mathbf{DATA}_{t-1})) \\ \hat{\beta}_t = \mathrm{E}(\beta|\mathbf{DATA}_{t-1}) & \mathbf{G}_t = \mathrm{COV}(\mathrm{E}(\beta|\mathbf{DATA}_{t-1})) \\ \dots & \dots \end{array}$$

- For latent noise vectors, the one-step-ahead predictions are trivial:
  - $\mathrm{E}(\boldsymbol{\epsilon}_t|\mathbf{DATA}_{t-1}) = 0$ and $\mathrm{COV}(\mathrm{E}(\boldsymbol{\epsilon}_t|\mathbf{DATA}_{t-1})) = Cov(\boldsymbol{\epsilon}_t)$
  - $\mathrm{E}(\boldsymbol{\eta}_t|\mathbf{DATA}_{t-1}) = 0$ and $\mathrm{COV}(\mathrm{E}(\boldsymbol{\eta}_t|\mathbf{DATA}_{t-1})) = Cov(\boldsymbol{\eta}_t)$

Kalman Smoother (KS)

- KS computes the smoothed (full-sample) predictions of the missing response values and the latent quantities.
- It is a backward recursive algorithm that uses the one-step-ahead forecasts generated during the KF phase.
- KS computes:

$$
\begin{array}{ll}
\tilde{\mathbf{Y}}_t = \mathrm{E}(\mathbf{Y}_t|\mathbf{DATA}_n) & \tilde{\mathbf{F}}_t = \mathrm{COV}(\mathrm{E}(\mathbf{Y}_t|\mathbf{DATA}_n)), \text{ for missing } \mathbf{Y}_t \\
\tilde{\boldsymbol{\alpha}}_t = \mathrm{E}(\boldsymbol{\alpha}_t|\mathbf{DATA}_n) & \tilde{\mathbf{P}}_t = \mathrm{COV}(\mathrm{E}(\boldsymbol{\alpha}_t|\mathbf{DATA}_n)) \\
\tilde{\boldsymbol{\beta}} = \mathrm{E}(\boldsymbol{\beta}|\mathbf{DATA}_n) & \tilde{\mathbf{G}} = \mathrm{COV}(\mathrm{E}(\boldsymbol{\beta}|\mathbf{DATA}_n)) \\
\cdots \quad \cdots & \cdots \quad \cdots \\
\tilde{\boldsymbol{\eta}}_t = \mathrm{E}(\boldsymbol{\eta}_t|\mathbf{DATA}_n) & \tilde{\mathbf{H}}_t = \mathrm{COV}(\mathrm{E}(\boldsymbol{\eta}_t|\mathbf{DATA}_n))
\end{array}
$$

- KS also yields other useful quantities, such as delete-one cross validation measures and structural break statistics.

UC Modeling: General Steps

Phase 1: Choose a good UCM for the observed data.

1. Propose a tentative UCM.
2. If the specified UCM has unknown parameters, estimate them.
3. Check the model adequacy and complexity (residual analysis, other diagnostics, ...).
4. If the model is inadequate or overly complex, modify it (back to the beginning).

Phase 2: Deploy the chosen UCM

- Use the estimated regression vectors for decision making
- Interpolate/extrapolate response values, latent components, ...
- Obtain a seasonal decomposition of the data sequence
- ...

KFS is the main computational tool for both the phases

KFS for Model Fitting and Diagnostics (Phase 1)

- Start with a proposed UCM, possibly with unknown parameter vector $\boldsymbol{\theta}$.
- KF yields one-step-ahead residuals and the likelihood of the data (at a specific trial value of $\boldsymbol{\theta}$):
    - $\mathbf{R}_t = (\mathbf{Y}_t - \hat{\mathbf{Y}}_t) \sim N(\mathbf{0}, \mathbf{F}_t)$ is an uncorrelated sequence.
    - $-2\text{logL}(\boldsymbol{\theta}, \mathbf{DATA}_n) = \sum_{t=1}^{n} \log(Det(\mathbf{F}_t)) + \mathbf{R}_t^{'}\mathbf{F}_t^{-1}\mathbf{R}_t + \cdots$
- Obtain the ML estimate of $\boldsymbol{\theta}$ by maximizing $\text{logL}(\boldsymbol{\theta}, \mathbf{DATA}_n)$ with respect to $\boldsymbol{\theta}$.
- Check the fitted model for adequacy and compare with other fitted models:
    - Residual analysis, structural break analysis, ...
    - Compare models by using information criteria.
    - KS yields delete-one cross validation measures, which can also be used for model comparison.

KFS for the Series and Component Interpolation/Extrapolation, ... (Phase 2)

Once a suitable model is decided, you can use the KFS again for

- Forecasting and interpolating the response series
- Estimating and forecasting the unobserved components and their linear combinations
- Estimating the sizes and types of structural breaks
- ...

State Space Modeling: Computational Cost

- $n =$ number of distinct time points, $m = \text{Dim}(\boldsymbol{\alpha_t})$
- Cost of single KFS run:
  - Number of multiplications $\sim \; nm^3$
  - Memory requirement of a KF run $\sim \; m^2$
  - Memory requirement of a KS run $\sim \; nm^2$ (output of a full KF run must be stored)
- ML estimation of parameter vector ($\boldsymbol{\theta}$) involves several runs of KFS (KF is used for likelihood computation, and KS is useful for the likelihood gradient computation).
- Computational/memory costs increase rapidly with m (only linearly with n).
- In some situations, the computational efficiency can be improved by exploiting the sparsity of the system matrices.

UCM and SSM procedures in SAS/ETS®

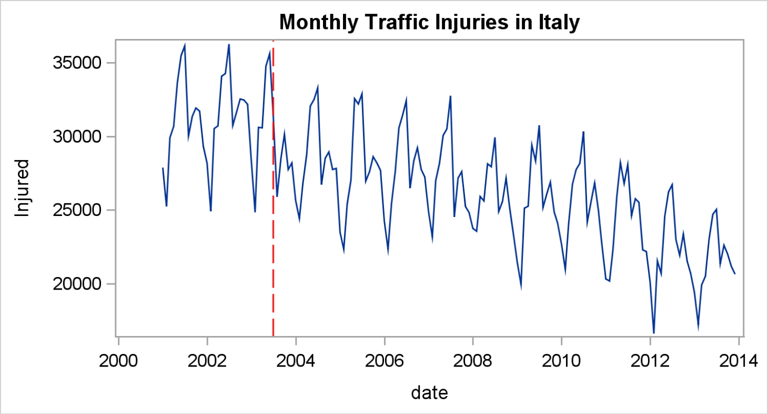
- PROC UCM for modeling univariate response variables
  - rich class of UCMs can be easily specified
  - a variety of diagnostics–tabular and graphical
  - series and component forecasts, and smoothed estimates
- PROC SSM for modeling with general linear SSMs
  - provides a flexible language for specifying very general linear SSMs
  - supports univariate and multivariate time series, panels of univariate and multivariate time series, and longitudinal data
  - keyword support for easy specification of commonly needed univariate and multivariate UCMs
  - ...

Where to Find Additional Info

- Books:
  - Pelagatti, M. M. (2016). *Time Series Modeling with Unobserved Components.* CRC Press.
  - Harvey, A. C. (1989). *Forecasting, Structural Time Series Models, and the Kalman Filter.* Cambridge: Cambridge University Press.
  - Durbin, J., and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods.* 2nd Ed. Oxford: Oxford University Press.
- SAS/ETS® Procedure Documentation:
  - PROC UCM (for univariate UCMs): `http://support.sas.com/documentation/cdl/en/etsug/68148/HTML/default/viewer.htm#etsug_ucm_toc.htm`
  - PROC SSM (for multivariate and other custom UCMs): `http://support.sas.com/documentation/cdl/en/etsug/68148/HTML/default/viewer.htm#etsug_ssm_toc.htm`

Modeling Motor Vehicle Injuries in Italy

- Based on Case Study # 1 from Pelagatti (2016)
- Monthly data on number of injuries due to road accidents
- A new traffic monitoring system introduced in July 2003
- Question: How effective is the new monitoring system?



Injuries = IRW Trend + Seasonal + Irregular

Check for breaks in the level component:

```
proc ucm data=spain.italy;
   id date interval=month;
   level variance=0 noest checkbreak;
   slope;
   season length=12 type=trig;
   irregular;
   model injured;
   estimate plot=panel;
   forecast plot=decomp;
run;
```

Outlier Summary

Obs	date	Break Type	Estimate	Standard Error	Chi-Square	DF	Pr > ChiSq
31	JUL2003	Level	-3856.27331	695.06886	30.78	1	<.0001
30	JUN2003	Level	-2757.10253	695.15074	15.73	1	<.0001

Injuries = IRW + Level Adjustment + Seasonal + Irregular

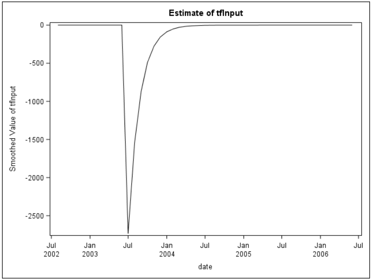
- Point_Jul03 = (date = July 03)
- Shift_Jul03 = (date >= July 03)
- Level Adjustment = β Shift_Jul03 + Transfer Function
- Transfer Function $\lambda_t = \frac{\gamma \text{ Point\_Jul03}}{1-\kappa B}$, where B denotes the lag operator

$$\lambda_t = \kappa\lambda_{t-1} + \gamma \text{ Point\_Jul03}$$

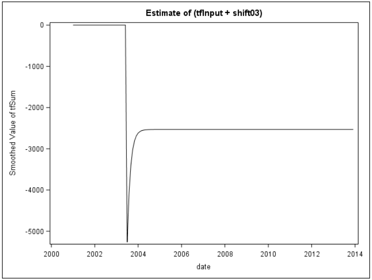
- λ_t is assumed to be zero before Jul 03
- The level adjustment parameters: β, γ, κ
- PROC SSM is used to fit this model

Estimated Components

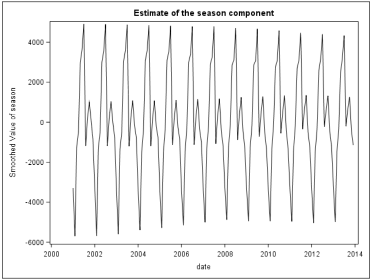
$\lambda_t = 0.564\ \lambda_{t-1} - 2735$ Point_Jul03



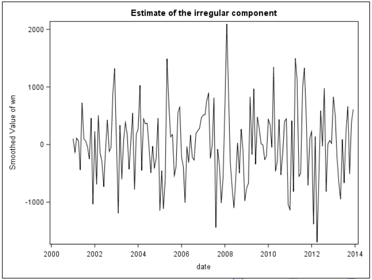
level adjustment $= \lambda_t - 2531$ Shift_Jul03



Seasonal Component

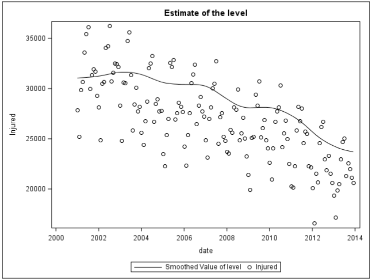


Irregular Component

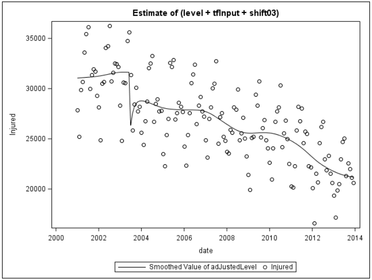


Series Decomposition

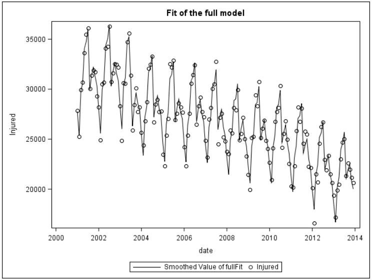
IRW Trend



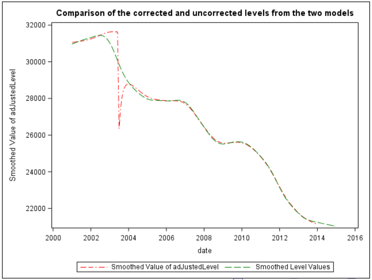
IRW Trend + level adjustment



IRW Trend + level adjustment + Seasonal

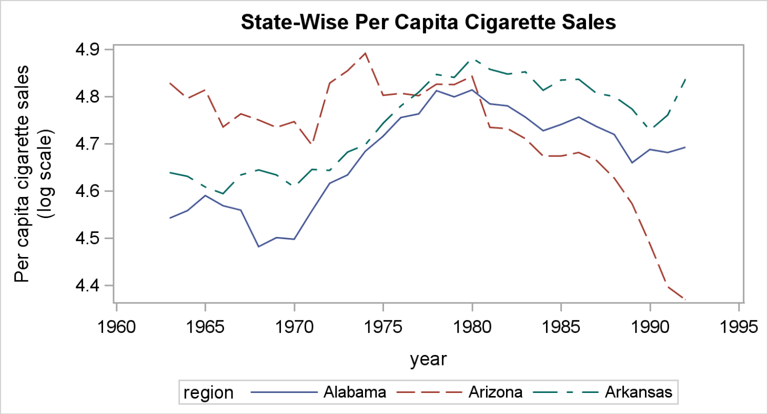


Comparing the levels from the two models



Modeling a Panel of 46 Time Series

- Yearly per capita cigarette sales for 46 states in the USA



Panel of Time Series

Over a span of 30 years (1963 - 1992), a study measured four variables in 46 states in USA:

- The response variable, *lsales*, denotes per capita cigarette sales in the natural log scale
- The regression variables (all in the natural log scale) denote:
  - price per pack of cigarettes (*lprice*)
  - per capita disposable income (*lndi*)
  - minimum price in adjoining regions per pack of cigarettes (*lpimin*)

Question: How do the regression variables *lprice*, *lndi*, and *lpimin* affect the response *lsales*? In particular, is the effect of *lpimin*, called the "boot-legging" effect, significant?

Panel Study: Proposed Model

$$lsales_{i,t} = \boldsymbol{\mu}_{i,t} + lprice\ \boldsymbol{\beta}_1 + lndi\ \boldsymbol{\beta}_2 + lpimin\ \boldsymbol{\beta}_3 + \boldsymbol{\epsilon}_{i,t}$$

- For $1 \leq i \leq 46$, $\boldsymbol{\mu}_{i,t}$ denote the region-specific IRW trends
- As a simplifying assumption, the disturbance variance in the slope equation is taken to be the same for all the 46 regions
- $(lprice\ \boldsymbol{\beta}_1 + lndi\ \boldsymbol{\beta}_2 + lpimin\ \boldsymbol{\beta}_3)$ denotes the contribution of the regression variables
- $\boldsymbol{\epsilon}_{i,t}$ are independent, Gaussian noise values

The regional trends $\boldsymbol{\mu}_{i,t}$ account for the differences between the regions because of unrecorded factors such as demographic changes over time, results of anti-smoking campaigns, and so on.

Panel Study Regression Estimates

Regression Parameter Estimates

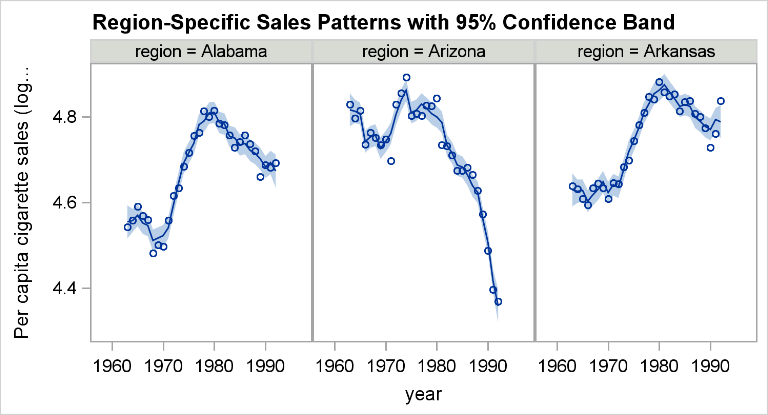
Response Variable	Regression Variable	Estimate	Standard Error	t Value	Pr > |t|
lsales	lprice	-0.3480	0.0232	-15.01	<.0001
lsales	lndi	0.1425	0.0344	4.15	<.0001
lsales	lpimin	0.0619	0.0269	2.30	0.0214

- All three regression variables have statistically significant effects
- The signs of regression coefficients are reasonable:
  - As the cigarette price increases, the sales decrease
  - As the disposable income increases, the sales increase
  - As the prices in the adjoining regions increase, the sales (within the state) increase

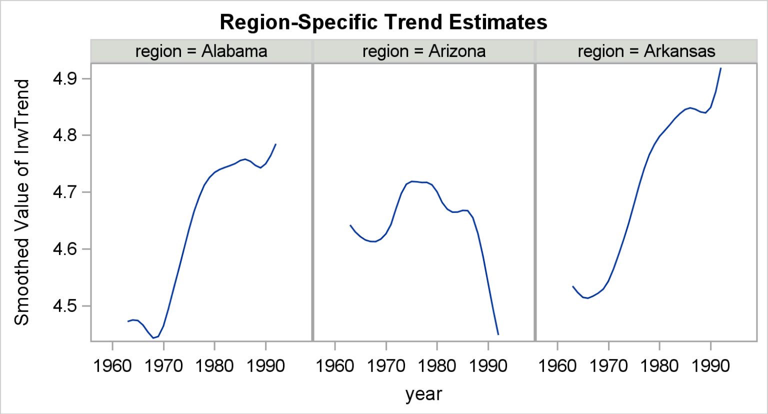
Panel Study: Region-Wise Model Fit

- Trend + Regression Effects =

$$\mu_{i,t} + lprice\ \beta_1 + lndi\ \beta_2 + lpimin\ \beta_3$$



Panel Study: Region-Wise Trend Estimates ($\mu_{i,t}$)

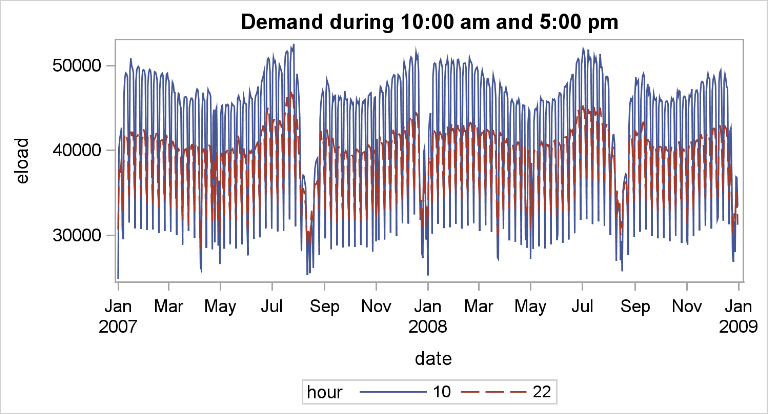


For more information see the "Getting Started" section in the SSM documentation.

Hourly Electricity Load in Italy

- Based on Case Study #3 in Pelagatti (2016)
- Hourly load history available for nine years: 01 Jan 2005 to 31 Dec 2014
- Such data exhibit several different types of seasonal behavior at different time scales:
  - Hour of the day pattern (season length 24 in hours)
  - Hour of the week pattern (season length 168 in hours)
  - Day of the week pattern (season length 7 in days)
  - Day of the year pattern (approx season length 365 in days)
- Load on holidays is usually different

Hourly Load During 10:00 am and 10:00 pm



Model for Electricity Load

Many ways to model these data. A modeling strategy that works reasonably well is as follows:

- Model the load in each hour of the day separately (i.e., 24 separate daily time series). The model for each series:

$$load_t = \mu_t + \mathbf{X}\beta + \gamma_t^7 + \gamma_t^{365} + \epsilon_t$$

where

- μ_t is a random walk trend
- $\mathbf{X}\beta$ is the correction for special days (mainly holidays)
- γ_t^7 is a trigonometric seasonal with season length = 7. Different harmonics use different disturbance variances.
- γ_t^{365} is a trigonometric seasonal with season length = 365. Only the first 16 harmonics used.
- ϵ_t is a Gaussian white noise

PROC UCM Code

```
proc ucm data=load;
   by hour;
   id date interval=day;
   irregular;
   level;
   cycle period=7 rho=1 noest=(period rho);
   cycle period=3.5 rho=1 noest=(period rho);
   cycle period=2.3333 rho=1 noest=(period rho);
   season length=365 type=trig keeph=1 to 16 by 1;
   model eload = dec24 dec25 dec26 jan1 jan6 aug15
        easterSun easterMon easterTue holidays holySat
        easterSat holySun bridgeDay endYear;
   estimate back=14 plot=panel;
   forecast back=14 lead=14 outfor=loadfor1;
run;
```

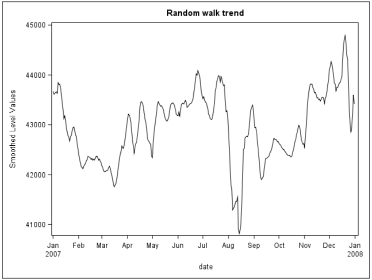
Specifies

$$load_t = \mu_t + \mathbf{X}\beta + \gamma_t^7 + \gamma_t^{365} + \epsilon_t$$

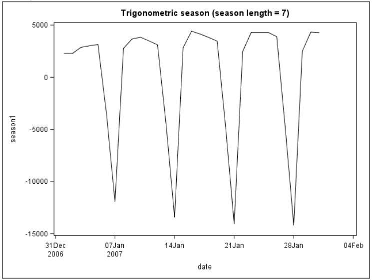
The program takes about 15 minutes to analyze all 24 time series.

Estimated Components for 10:00 am

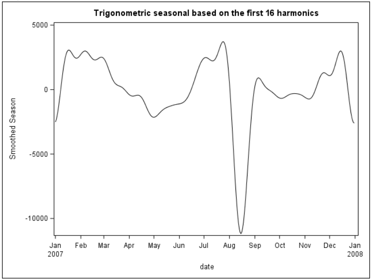
Random walk trend



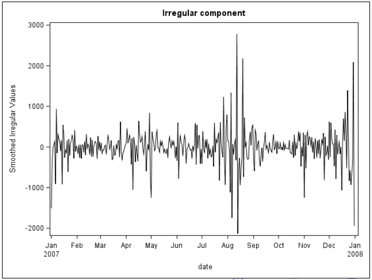
γ_t^7, the trigonometric seasonal with season length $= 7$



γ_t^{365}, the trigonometric seasonal with season length $= 365$

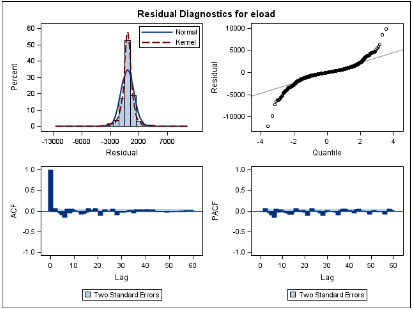


Irregular Component

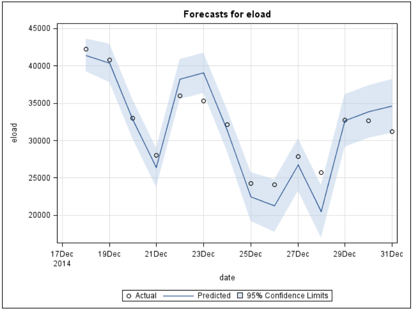


Residual Diagnostics and Forecasts for 10:00 am

Residual Diagnostics



Forecasts for the last 14 days



Adding Temperature Effect in the Load Modeling

- The electricity load, particularly in the residential areas, is quite sensitive to the outside temperature
- The relationship between temp and load is usually nonlinear: the load is higher for lower and upper temp ranges
- If good temp forecasts are available, the earlier model can be improved by adding a nonlinear temp effect, λ_t^{temp}, as follows

$$load_t = \mu_t + \mathbf{X}\beta + \gamma_t^7 + \gamma_t^{365} + \lambda_t^{temp} + \epsilon_t$$

PROC UCM Code with Temp Effect

```
proc ucm data=tempload;
   by hour;
   id date interval=day;
   irregular;
   level;
   season length=7 type=trig;
   season length=365 type=trig keeph=1 to 16 by 1;
   splinereg temp degree=3 nknots=10;
   model eload = dec24 dec25 dec26 jan1 jan6 aug15
        easterSun easterMon easterTue holidays holySat
        easterSat holySun bridgeDay endYear;
   estimate back=14 plot=panel;
   forecast back=14 lead=14 outfor=loadfor1;
run;
```

Specifies

$$load_t = \mu_t + \mathbf{X}\beta + \gamma_t^7 + \gamma_t^{365} + \lambda_t^{temp} + \epsilon_t$$

with λ_t^{temp} as a cubic spline with ten "equally spaced" knots in the observed temperature range

Machine Learning Versus UCMs for Electricity Market Data

Lisi and Pelagatti (2015) analyzed daily electricity load and price data by using UCMs and two popular machine learning techniques (support vector machine regression and random forest regression). General conclusions from their study:

- Loads are very regular and both UCMs and ML models do a good job.
- Prices are more messy and the UCMs do better than their ML counterparts.
- UCMs are more interpret-able and easier to "tune".
- Their presentation is available at `https://www.researchgate.net/publication/301547386_Component_estimation_for_electricity_market_data_deterministic_or_stochastic`

Common Trends in Multivariate RW

Consider an N-dimensional random walk $\boldsymbol{\mu}_t$:

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \sim N(0, \Sigma_\nu)$$

Suppose $\text{rank}(\Sigma_\nu) = k$, $1 \leq k < N$. Then,

$$\begin{aligned} \boldsymbol{\mu}_t &= \boldsymbol{\Theta}_\mu \, \boldsymbol{\mu}_t^\dagger + \boldsymbol{\theta}_\mu \\ \boldsymbol{\mu}_t^\dagger &= \boldsymbol{\mu}_{t-1}^\dagger + \boldsymbol{\nu}_t^\dagger, \quad \boldsymbol{\nu}_t^\dagger \sim N(0, \Sigma_\nu^\dagger) \end{aligned}$$

where $\dim(\boldsymbol{\mu}_t^\dagger) = k$, $\boldsymbol{\Theta}_\mu = \left(\begin{smallmatrix} I_k \\ \boldsymbol{\Theta}_{(N-k)\times k} \end{smallmatrix}\right)$, and $\boldsymbol{\theta}_\mu = \left(\begin{smallmatrix} 0_k \\ \theta_{(N-k)\times 1} \end{smallmatrix}\right)$. That is,

- The N-dimensional random walk $\boldsymbol{\mu}_t$ is driven by a k-dimensional random walk $\boldsymbol{\mu}_t^\dagger$
- $\boldsymbol{\Theta}$ is called the loading matrix
- The elements of $\boldsymbol{\Theta}$, θ, and $\Sigma_\nu^\dagger$ are the new parameters

Trivariate RW with Common Trends ($N = 3,\ k = 2$)

Suppose $\boldsymbol{\mu}_t^\dagger$ is a 2-dimensional random walk

$$\boldsymbol{\mu}_t^\dagger = \boldsymbol{\mu}_{t-1}^\dagger + \boldsymbol{\nu}_t^\dagger, \quad \boldsymbol{\nu}_t^\dagger \sim N(0, \Sigma_\nu^\dagger)$$

Then a three dimensional random walk with two common trends has the following form:

$$\begin{aligned} \boldsymbol{\mu}_{1t} &= \boldsymbol{\mu}_{1t}^\dagger \\ \boldsymbol{\mu}_{2t} &= \boldsymbol{\mu}_{2t}^\dagger \\ \boldsymbol{\mu}_{3t} &= \theta_0 + \theta_1\ \boldsymbol{\mu}_{1t}^\dagger + \theta_2\ \boldsymbol{\mu}_{2t}^\dagger \end{aligned}$$

The constants $\theta_0, \theta_1, \theta_2$, and the elements of $\Sigma_\nu^\dagger$ are the model parameters. θ_1 and θ_2 are called factor loadings.

Trivariate RW with Common Trends: An Example

- Example 7.1 from Pelagatti (2016)
- p_t denotes the spot price of Brent crude oil (in the log scale)
- f_t denotes the future price of Brent crude oil (in the log scale)
- r_t denotes the risk free continuously compounded annual interest rate

According to the econometric considerations

$$f_t \sim p_t + \delta r_t$$

where δ denotes the time to delivery (in years). This suggests that the three-dimensional series $\mathbf{y}_t = (p_t \; r_t \; f_t)$ might be driven by a two dimensional mechanism.

Model for $\mathbf{y}_t = (p_t \; r_t \; f_t)$

Suppose $\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\phi}_t$ where $\boldsymbol{\mu}_t$ is a three dimensional random walk with two common trends, and $\boldsymbol{\phi}_t$ is a three dimensional AR(1) process with diagonal coefficient matrix $\mathbf{D}$. In effect,

$$
\begin{aligned}
p_t &= \boldsymbol{\mu}^{\dagger}_{1t} + \boldsymbol{\phi}_{1t} \\
r_t &= \boldsymbol{\mu}^{\dagger}_{2t} + \boldsymbol{\phi}_{2t} \\
f_t &= \theta_0 + \theta_1 \, \boldsymbol{\mu}^{\dagger}_{1t} + \theta_2 \, \boldsymbol{\mu}^{\dagger}_{2t} + \boldsymbol{\phi}_{3t}
\end{aligned}
$$

where

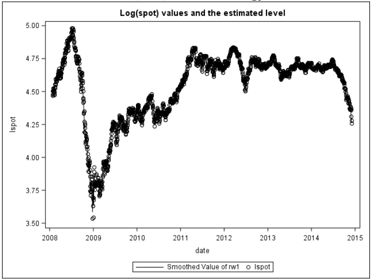
$$
\begin{aligned}
&\boldsymbol{\mu}^{\dagger}_t = \boldsymbol{\mu}^{\dagger}_{t-1} + \boldsymbol{\nu}^{\dagger}_t, \quad \boldsymbol{\nu}^{\dagger}_t \sim N(0, \Sigma^{\dagger}_{\nu}) \\
&\boldsymbol{\phi}_t = \mathbf{D} \, \boldsymbol{\phi}_{t-1} + \boldsymbol{\zeta}_t, \quad \boldsymbol{\zeta}_t \sim N(0, \Sigma_{\zeta}), \quad \text{rank}(\Sigma_{\zeta}) = 2
\end{aligned}
$$

PROC SSM Code

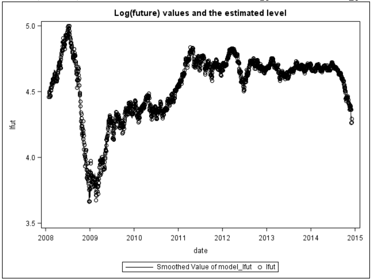
```
proc ssm data=brent; * opt(tech=activeset);
  id date interval=weekday;
  parms load1 load2 / lower=0;
  one = 1.0;
  state rw(2) type=rw cov(g);
  comp rw1 = rw[1];
  comp rw2 = rw[2];
  comp rw3 = (load1 load2)*rw;

  state ar(3) type=VARMA(p(d)=1) cov(rank=2);
  comp ar1 = ar[1];
  comp ar2 = ar[2];
  comp ar3 = ar[3];

  model lspot = rw1 ar1;
  model intrate = rw2 ar2;
  model lfut = one rw3 ar3;
run;
```

Analysis Results

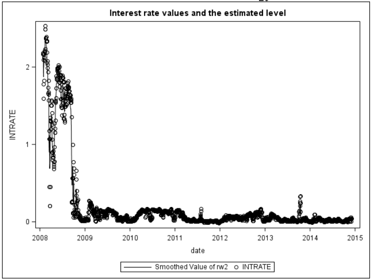
Log(spot) values and $\mu_{1t}^{\dagger}$



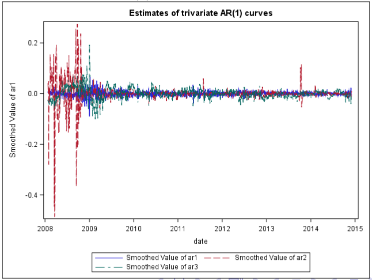
Log(future) values and $0.29 + 0.93676\,\mu_{1t}^{\dagger} + 0.00696\,\mu_{2t}^{\dagger}$



Interest rate values and $\mu_{2t}^{\dagger}$



Three AR(1) processes ϕ_{1t}, ϕ_{2t}, ϕ_{3t}



Summary

- UCMs are structural models:
  - Prior knowledge (or some data exploration) suggests the form of the initial model
  - A variety of models available to capture different types of commonly needed structural patterns such as trend, cycles, etc.
  - The analysis provides the in-sample and out-of-sample estimates of these unobserved structural patterns. Such estimates are important for a variety of purposes: seasonal adjustment, determining the relative sizes of different effects, ...
  - Refinement of the initial model is based on standard statistical techniques: residual diagnostics, information criteria, structural break analysis, etc

Summary Continued ...

- UCMs are structural models (continued):
  - In addition to the interpolation and extrapolation (forecasting) of the response values, the analysis also provides similar estimates for the model components
- UCMs have state space forms
  - Model parameters are estimated by optimizing the likelihood, which is computed by using the Kalman filter
  - interpolation and extrapolation of the response values and the model components is done by using the Kalman filter and smoother

Additional References

- Harvey, A.C. and Trimbur, T. (2003). General model-based filters for extracting cycles and trends in economic time series. The Review of Economics and Statistics 85(2), 244-55.
- Runstler, G. (2004) Modelling phase shifts among stochastic cycles, Econometrics Journal (2004), volume 7, pp. 232-248.
- White papers/workshop slides by Rajesh Selukar:
  - State Space Modeling of Sequence Data https://forecasters.org/wp-content/uploads/gravity_forms/7-621289a708af3e7af65a7cd487aee6eb/2015/07/selukar_rajesh_isf2015.pdf
  - Functional Modeling of Longitudinal Data with the SSM Procedure: http://support.sas.com/resources/papers/proceedings15/SAS1580-2015.pdf