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

- [Smoothed Trend for Incidences]
- [Sum of Smoothed Trend and Cycles for Incidences]
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

\[ Y_t = X_t \beta + \mu_t + \psi_t + \ldots + \epsilon_t \]

- Univariate or multivariate response at time \( t \): \( Y_t \)
- Effect of regression variables: \( X_t \beta \)
- Time varying mean (level/trend): \( \mu_t \)
- Periodic/Seasonal component: \( \psi_t \)
- Noise component: \( \epsilon_t \)
- A component could be turned on/off, or scaled, based on an external input, e.g., the trend could be scaled as \( b_t \mu_t \)
- All of these components need not be present in a UCM
- Many more types of components are often needed/used
Local Linear Trend (LLT)

\[
\begin{align*}
\mu_t &= \mu_{t-1} + \eta_{t-1} + \nu_t & \text{Level equation} \\
\eta_t &= \eta_{t-1} + \xi_t & \text{Slope equation}
\end{align*}
\]

- \( \nu_t \sim N(0, \Sigma_\nu) \) are i.i.d. disturbances in the level equation
- \( \xi_t \sim N(0, \Sigma_\xi) \) are i.i.d. disturbances in the slope equation
- The initial level \( \mu_1 \) and the initial slope \( \eta_1 \) are (usually) unknown vectors

LLT in a vector recursion form:

\[
\begin{bmatrix}
\mu_t \\
\eta_t
\end{bmatrix} = 
\begin{bmatrix}
1 & 1 \\
0 & 1
\end{bmatrix} 
\begin{bmatrix}
\mu_{t-1} \\
\eta_{t-1}
\end{bmatrix} + 
\begin{bmatrix}
\nu_t \\
\xi_t
\end{bmatrix}
\]

Even the simple LLT + Noise model, \( Y_t = \mu_t + \epsilon_t \), turns out to be a very versatile model.
Some Special Cases of LLT

\[
\mu_t = \mu_{t-1} + \eta_{t-1} + \nu_t \quad \text{Level equation}
\]
\[
\eta_t = \eta_{t-1} + \xi_t \quad \text{Slope equation}
\]

- Random walk \( \mu_t = \mu_{t-1} + \nu_t \) (initial slope \( \eta_1 \) and the slope disturbance covariance \( \Sigma_\xi \) are zero).
- Random walk with drift \( \mu_t = \mu_{t-1} + \eta_1 + \nu_t \) (the slope disturbance covariance \( \Sigma_\xi \) is zero).
- Integrated random walk \( \mu_t = \mu_{t-1} + \eta_{t-1} \); \( \eta_t = \eta_{t-1} + \xi_t \) (the level disturbance covariance \( \Sigma_\nu \) is zero).
- Deterministic time trend \( \mu_t = \mu_1 + t \eta_1 \) (the level disturbance covariance \( \Sigma_\nu \) and the slope disturbance covariance \( \Sigma_\xi \) are zero).
- Time invariant mean \( \mu_t = \mu_1 \) (\( \eta_1 \), \( \Sigma_\xi \), and \( \Sigma_\nu \) are all zero).
Simulated Univariate Trends

Constant Trend $\mu_t = 5$

Random Walk

Linear trend $\mu_t = 5 - 0.1t$

Local linear trend
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

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

You can verify that $\psi_t$ 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, $\psi^2_t + \psi^{*2}_t = a^2 + b^2$, $\forall t$. 
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, $\rho$, 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 $\nu_t$ and $\nu^*_t$ are independent $N(0, \sigma^2_\nu)$ variables.

The resulting random sequence $\psi_t$ 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

\[ \psi_t = \rho^t R_\omega \psi_0 + \sum_{j=0}^{t} \rho^{t-j} R_\omega^{t-j} \nu_j, \quad R_\omega = \begin{bmatrix} \cos \omega & \sin \omega \\ -\sin \omega & \cos \omega \end{bmatrix} \]

- If \( \rho < 1 \), the effect of initial condition and the shocks in the distant past becomes negligible. \( \psi_t \) has a stationary distribution with mean zero and variance \( \sigma^2_\nu/(1 - \rho^2) \).
- If \( \rho = 1 \), the effect of shocks persists and \( \psi_t \) 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 $\gamma_t$ 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

\[
\sum_{i=0}^{s-1} \gamma_{t-i} = \nu_t, \quad \nu_t \sim N(0, \sigma^2_\nu)
\]

- 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 \( \sigma_{\nu}^2 \).
- 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

Season length = 12 dist variance = 0

The seasonal pattern is a sum of first 8 harmonics (skiplast = 0)

Deterministic Season with First Two Harmonics

Season length = 12 dist variance = 0

The seasonal pattern is a sum of first 2 harmonics (skiplast = 4)

Stochastic Season with All Harmonics

Season length = 12 dist variance = 0.1

The seasonal pattern is a sum of first 8 harmonics (skiplast = 0)

Stochastic Season with First Two Harmonics

Season length = 12 dist variance = 0.1

The seasonal pattern is a sum of first 2 harmonics (skiplast = 4)
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

\[ Y_t = Z_t \alpha_t + X_t \beta + \epsilon_t \quad \text{Observation equation} \]
\[ \alpha_{t+1} = T_t \alpha_t + W_{t+1} \gamma + \zeta_{t+1} \quad \text{State transition equation} \]
\[ \alpha_1 = A_1 \delta + W_1 \gamma + \zeta_1 \quad \text{Initial condition} \]

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

\[ Y_t = Z_t \alpha_t + X_t \beta + \epsilon_t \]
\[ \alpha_{t+1} = T_t \alpha_t + W_{t+1} \gamma + \eta_{t+1} \]
\[ \alpha_1 = A_1 \delta + W_1 \gamma + \eta_1 \]

- **Observation equation**
- **State transition equation**
- **Initial condition**

<table>
<thead>
<tr>
<th>Vector</th>
<th>Dim</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_t )</td>
<td>( m )</td>
<td>State vectors</td>
</tr>
<tr>
<td>( \beta )</td>
<td>( k )</td>
<td>Regression vector in the observation equation</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>( g )</td>
<td>Regression vector in the state equation</td>
</tr>
<tr>
<td>( \delta )</td>
<td>( d )</td>
<td>Diffuse part of ( \alpha_1 )</td>
</tr>
<tr>
<td>( \epsilon_t )</td>
<td>( p )</td>
<td>Observation noise (zero-mean, Gaussian)</td>
</tr>
<tr>
<td>( \eta_t )</td>
<td>( m )</td>
<td>State noise (zero-mean, Gaussian)</td>
</tr>
</tbody>
</table>

Noise/shock/disturbance variables \( \epsilon_t \) and \( \eta_t \) are mutually independent white noise sequences (possibly with time-varying covariances).
Model System Matrices

\[ Y_t = Z_t \alpha_t + X_t \beta + \epsilon_t \quad \text{Observation equation} \]
\[ \alpha_{t+1} = T_t \alpha_t + W_{t+1} \gamma + \eta_{t+1} \quad \text{State transition equation} \]
\[ \alpha_1 = A_1 \delta + W_1 \gamma + \eta_1 \quad \text{Initial condition} \]

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Dim</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Z_t )</td>
<td>( p \times m )</td>
<td>Design matrix for ( \alpha_t )</td>
</tr>
<tr>
<td>( T_t )</td>
<td>( m \times m )</td>
<td>State transition matrix</td>
</tr>
<tr>
<td>( A_1 )</td>
<td>( m \times d )</td>
<td>Diffuse condition specifier made up of 0’s and 1’s</td>
</tr>
<tr>
<td>( \text{Cov}(\epsilon_t) )</td>
<td>( p \times p )</td>
<td>Often diagonal</td>
</tr>
<tr>
<td>( \text{Cov}(\eta_t) )</td>
<td>( m \times m )</td>
<td>Often nondiagonal</td>
</tr>
</tbody>
</table>

- Missing elements are not allowed in any system matrix.
  However, the system matrices can depend on some unknown parameter vector \( \theta \) (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 $\mu_t$, $\psi_t$, and $\epsilon_t$ 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

$$y_t = Z\alpha_t + \epsilon_t$$ \hspace{1cm} Observation equation

$$\alpha_{t+1} = T\alpha_t + \zeta_{t+1}$$ \hspace{1cm} State transition equation

$$\alpha_1 = A_1\delta + \zeta_1$$ \hspace{1cm} Initial condition

where $\alpha_t = [\mu_t \ \eta_t \ \psi_t \ \psi^*_t]$, $Z = [1 \ 0 \ 1 \ 0]$, $\zeta_t = [\nu_t \ \xi_t \ \nu_t \ \nu^*_t]$, $\delta = [\mu_1 \ \eta_1]$, $A_1 = [1 \ 0; \ 0 \ 1; \ 0 \ 0; \ 0 \ 0]$, $\zeta_1 \sim N(0, [0, 0, \sigma^2_{\nu}/(1 - \rho^2), \sigma^2_{\nu}/(1 - \rho^2)])$, and $T = [1 \ 1 \ 0 \ 0; \ 0 \ 1 \ 0 \ 0; \ 0 \ 0 \ \rho \cos(\omega) \ \rho \sin(\omega); \ 0 \ 0 \ \rho \sin(\omega) \ \rho \cos(\omega)]$. 
Informal Description of the Kalman Filter (KF)

- Assume that the model parameter vector $\theta$ is known.
- KF recursively computes the one-step-ahead predictions of the response values and the latent quantities.
- Let $\text{DATA}_t$ denote all the data up to time $t$.
- KF recursively computes:
  \[
  \hat{Y}_t = E(Y_t|\text{DATA}_{t-1}) \quad F_t = \text{COV}(E(Y_t|\text{DATA}_{t-1})) \\
  \hat{\alpha}_t = E(\alpha_t|\text{DATA}_{t-1}) \quad P_t = \text{COV}(E(\alpha_t|\text{DATA}_{t-1})) \\
  \hat{\beta}_t = E(\beta|\text{DATA}_{t-1}) \quad G_t = \text{COV}(E(\beta|\text{DATA}_{t-1})) \\
  \ldots \quad \ldots
  \]
- For latent noise vectors, the one-step-ahead predictions are trivial:
  - $E(\epsilon_t|\text{DATA}_{t-1}) = 0$ and $\text{COV}(E(\epsilon_t|\text{DATA}_{t-1})) = \text{Cov}(\epsilon_t)$
  - $E(\eta_t|\text{DATA}_{t-1}) = 0$ and $\text{COV}(E(\eta_t|\text{DATA}_{t-1})) = \text{Cov}(\eta_t)$
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:
  \[
  \hat{\bar{Y}}_t = E(Y_t | DATA_n), \quad \hat{\bar{F}}_t = \text{COV}(E(Y_t | DATA_n)), \quad \text{for missing } Y_t
  \]
  \[
  \hat{\bar{\alpha}}_t = E(\alpha_t | DATA_n), \quad \hat{\bar{P}}_t = \text{COV}(E(\alpha_t | DATA_n))
  \]
  \[
  \hat{\bar{\beta}} = E(\beta | DATA_n), \quad \hat{\bar{G}} = \text{COV}(E(\beta | DATA_n))
  \]
  \[
  \ldots \quad \ldots \quad \ldots \quad \ldots
  \]
  \[
  \hat{\bar{\eta}}_t = E(\eta_t | DATA_n), \quad \hat{\bar{H}}_t = \text{COV}(E(\eta_t | DATA_n))
  \]
- 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 $\theta$.
- KF yields one-step-ahead residuals and the likelihood of the data (at a specific trial value of $\theta$):
  - $R_t = (Y_t - \hat{Y}_t) \sim N(0, F_t)$ is an uncorrelated sequence.
  - $-2 \log L(\theta, \text{DATA}_n) = \sum_{t=1}^{n} \log(\text{Det}(F_t)) + R_t' F_t^{-1} R_t + \cdots$
- Obtain the ML estimate of $\theta$ by maximizing $\log L(\theta, \text{DATA}_n)$ with respect to $\theta$.
- 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.
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, $m = \text{Dim}(\alpha_t)$
- 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 ($\theta$) 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:**

- **SAS/ETS® Procedure Documentation:**
  - PROC UCM (for univariate UCMs):
  - PROC SSM (for multivariate and other custom UCMs):
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:

```r
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;
```

<table>
<thead>
<tr>
<th>Obs</th>
<th>date</th>
<th>Break</th>
<th>Type</th>
<th>Estimate</th>
<th>Error</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>JUL2003</td>
<td>Level</td>
<td></td>
<td>-3856.27331</td>
<td>695.06886</td>
<td>30.78</td>
<td>1</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>30</td>
<td>JUN2003</td>
<td>Level</td>
<td></td>
<td>-2757.10253</td>
<td>695.15074</td>
<td>15.73</td>
<td>1</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
Injuries = IRW + Level Adjustment + Seasonal + Irregular

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

\[
\lambda_t = \kappa \lambda_{t-1} + \gamma \text{Point_Jul03}
\]

- \( \lambda_t \) is assumed to be zero before Jul 03
- The level adjustment parameters: \( \beta, \gamma, \kappa \)
- PROC SSM is used to fit this model
Estimated Components

\[ \lambda_t = 0.564 \lambda_{t-1} - 2735 \text{ Point-Jul03} \]

level adjustment = \( \lambda_t - 2531 \text{ Shift-Jul03} \)

Seasonal Component

Irregular Component
Series Decomposition

IRW Trend

Estimate of the level

IRW Trend + level adjustment

Estimate of (level + tinput + shift03)

IRW Trend + level adjustment + Seasonal

Fit of the full model

Comparing the levels from the two models

Comparison of the corrected and uncorrected 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, \( l_{sales} \), 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 (\( l_{price} \))
  - per capita disposable income (\( l_{ndi} \))
  - minimum price in adjoining regions per pack of cigarettes (\( l_{pimin} \))

Question: How do the regression variables \( l_{price}, l_{ndi}, \) and \( l_{pimin} \) affect the response \( l_{sales} \)? In particular, is the effect of \( l_{pimin} \), called the ”boot-legging” effect, significant?
Panel Study: Proposed Model

\[ l_{sales_{i,t}} = \mu_{i,t} + l_{price} \beta_1 + l_{Indi} \beta_2 + l_{pimin} \beta_3 + \epsilon_{i,t} \]

- For \( 1 \leq i \leq 46 \), \( \mu_{i,t} \) 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
- \( (l_{price} \beta_1 + l_{Indi} \beta_2 + l_{pimin} \beta_3) \) denotes the contribution of the regression variables
- \( \epsilon_{i,t} \) are independent, Gaussian noise values

The regional trends \( \mu_{i,t} \) 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

| 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 =

$$\mu_{i,t} + \text{lprice } \beta_1 + \text{Indi } \beta_2 + \text{lpimin } \beta_3$$
Panel Study: Region-Wise Trend Estimates ($\mu_{i,t}$)

Region-Specific Trend Estimates

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

Demand during 10:00 am and 5:00 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:

\[
\text{load}_t = \mu_t + X\beta + \gamma^7_t + \gamma^{365}_t + \epsilon_t
\]

where

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

\[ \text{load}_t = \mu_t + \mathbf{X}\beta + \gamma_7^t + \gamma_{365}^t + \epsilon_t \]

The program takes about 15 minutes to analyze all 24 time series.
Estimated Components for 10:00 am

Random walk trend

\( \gamma_t^7 \), the trigonometric seasonal with season length = 7

Irregular Component

\( \gamma_t^{365} \), the trigonometric seasonal with season length = 365
Residual Diagnostics and Forecasts for 10:00 am

Residual Diagnostics

Residual Diagnostics for eload

Forecasts for the last 14 days

Forecasts for eload

- Actual
- Predicted
- 95% Confidence Limits
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, $\lambda_t^{temp}$, as follows:

$$load_t = \mu_t + X\beta + \gamma_t^7 + \gamma_t^{365} + \lambda_t^{temp} + \epsilon_t$$
PROC UCM Code with Temp Effect

```ucm
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

\[ load_t = \mu_t + X\beta + \gamma_7 t + \gamma_{365} t + \lambda_{temp} t + \epsilon_t \]

with \( \lambda_{temp} t \) as a cubic spline with ten "equally spaced" knots in the observed temperature range.
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 interpretable 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 $\mu_t$:

$$
\mu_t = \mu_{t-1} + \nu_t, \quad \nu_t \sim N(0, \Sigma_\nu)
$$

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

$$
\mu_t = \Theta_\mu \mu_t^\dagger + \theta_\mu \\
\mu_t^\dagger = \mu_{t-1}^\dagger + \nu_t^\dagger, \quad \nu_t^\dagger \sim N(0, \Sigma_\nu^\dagger)
$$

where $\text{dim}(\mu_t^\dagger) = k, \Theta_\mu = (\Theta_{(N-k) \times k}),$ and $\theta_\mu = (\theta_{(N-k) \times 1})$. That is,

- The $N$-dimensional random walk $\mu_t$ is driven by a $k$-dimensional random walk $\mu_t^\dagger$
- $\Theta$ is called the loading matrix
- The elements of $\Theta, \theta,$ and $\Sigma_\nu^\dagger$ are the new parameters
Trivariate RW with Common Trends \((N = 3, \ k = 2)\)

Suppose \(\mu_t^\dagger\) is a 2-dimensional random walk

\[
\mu_t^\dagger = \mu_{t-1}^\dagger + \nu_t^\dagger, \quad \nu_t^\dagger \sim \mathcal{N}(0, \Sigma_{\nu}^\dagger)
\]

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

\[
\mu_{1t} = \mu_{1t}^\dagger \\
\mu_{2t} = \mu_{2t}^\dagger \\
\mu_{3t} = \theta_0 + \theta_1 \mu_{1t}^\dagger + \theta_2 \mu_{2t}^\dagger
\]

The constants \(\theta_0, \theta_1, \theta_2\), and the elements of \(\Sigma_{\nu}^\dagger\) are the model parameters. \(\theta_1\) and \(\theta_2\) 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 $\delta$ 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 $y_t = (p_t \ r_t \ f_t)$

Suppose $y_t = \mu_t + \phi_t$ where $\mu_t$ is a three dimensional random walk with two common trends, and $\phi_t$ is a three dimensional AR(1) process with diagonal coefficient matrix $D$. In effect,

$$p_t = \mu_1^t + \phi_1 t$$
$$r_t = \mu_2^t + \phi_2 t$$
$$f_t = \theta_0 + \theta_1 \mu_1^t + \theta_2 \mu_2^t + \phi_3 t$$

where

$$\mu_t^\dagger = \mu_{t-1}^\dagger + \nu_t^\dagger, \quad \nu_t^\dagger \sim N(0, \Sigma_\nu^\dagger)$$
$$\phi_t = D \phi_{t-1} + \zeta_t, \quad \zeta_t \sim N(0, \Sigma_\zeta), \quad \text{rank}(\Sigma_\zeta) = 2$$
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 $\phi_{1t}$, $\phi_{2t}$, $\phi_{3t}$**

![Graph of Log(spot) values and $\mu_{1t}^\dagger$](image1)

![Graph of Log(future) values](image2)

![Graph of Interest rate values and $\mu_{2t}^\dagger$](image3)

![Graph of Three AR(1) processes](image4)
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

- White papers/workshop slides by Rajesh Selukar:
  - State Space Modeling of Sequence Data