RegComponent Modeling in the SeasCen Software Platform

William R. Bell, Lijing Sun, and Demetra Lytras
U.S. Census Bureau

William.R.Bell@census.gov

Disclaimer: The views expressed in this presentation are those of the authors and not those of the U.S. Census Bureau.

References

[1] Bell, William R. (2004), "On RegComponent Time Series Models and Their Applications," in *State Space and Unobserved Component Models: Theory and Applications*, eds. Andrew C. Harvey, Siem Jan Koopman, and Neil Shephard, Cambridge, UK: Cambridge University Press, 248–283.

[2] Bell, William R. (2011), "REGCMPNT – A Fortran Program for Regression Models with ARIMA Component Errors," *Journal of Statistical Software*, Vol. **41**, Issue 7, May 2011, available online at http://www.jstatsoft.org/v41/i07.

RegARIMA model:

$$y_t = x_t'\beta + z_t \qquad z_t \sim \text{ ARIMA model}$$

RegComponent model:

$$y_t = x_t'\beta + \sum_{i=1}^{m} h_{it} z_{it} \qquad z_{it} \sim \text{ ARIMA model} \qquad z_{it} \perp z_{jt} \quad i \neq j$$

h_{it} for $i = 1, \ldots, m$ are series of known constants called "scale factors" (often $h_{it} \equiv 1$)

z_{it} for $i = 1, \ldots, m$ are independent unobserved component series following ARIMA models:

$$\phi_i(B)\delta_i(B)z_{it} = \theta_i(B)\epsilon_{it}$$

RegComponent model calculations:

Calculations done via Kalman filter/smoother with suitable initialization.

- KF initialization of Bell and Hillmer (1991) — produces "transformation approach" estimates of Ansley and Kohn (1985)
- IGLS approach to likelihood maximization
- fixed point smoother of reduced dimension

RegComponent model – special cases

$$y_t = x_t'\beta + \sum_{i=1}^{m} h_{it} z_{it}$$

- RegARIMA model: $m = 1$, $h_{1t} \equiv 1$

- Basic Structural Model (Harvey 1989) with trading-day regression variables: $m = 3$, $h_{it} \equiv 1$, $z_{1t} = S_t$, $z_{2t} = T_t$, $z_{3t} = I_t$

$$y_t = \left(S_t + \sum_{i=1}^{6} \beta_i D_{it}\right) + T_t + I_t$$

where

$$\begin{aligned} (1 + B + \cdots + B^{11})S_t &= \epsilon_{1t} & \epsilon_{1t} \sim N(0, \sigma_1^2) \\ (1 - B)^2 T_t &= (1 - \eta B)\epsilon_{2t} & \epsilon_{2t} \sim N(0, \sigma_2^2) \\ I_t &= \epsilon_{3t} & \epsilon_{3t} \sim N(0, \sigma_3^2) \end{aligned}$$

- Structural model with trigonometric seasonal (Harvey 1989):

$$
\begin{aligned}
S_t &= \sum_{i=1}^{6} S_{it} \\
(1-B)^2 T_t &= (1-\eta B)\epsilon_{7t} \\
I_t &= \epsilon_{8t}
\end{aligned}
$$

with

$$
\begin{aligned}
(1 - 1.732B + B^2)S_{1t} &= (1 - .577B)\epsilon_{1t} \\
(1 - B + B^2)S_{2t} &= (1 - .268B)\epsilon_{2t} \\
(1 + B^2)S_{3t} &= \epsilon_{3t} \\
(1 + B + B^2)S_{4t} &= (1 + .268B)\epsilon_{4t} \\
(1 + 1.732B + B^2)S_{5t} &= (1 + .577B)\epsilon_{5t} \\
(1 + B)S_{6t} &= \epsilon_{6t}
\end{aligned}
$$

Trigonometric seasonal can equivalently be written as (Hannan 1970)

$$S_t = \sum_{k=1}^{6} \left[\cos\left(\frac{2\pi kt}{12}\right) \alpha_{kt} + \sin\left(\frac{2\pi kt}{12}\right) \beta_{kt} \right]$$

where

$$(1-B)\alpha_{kt} \;=\; \epsilon_{k1t} \sim i.i.d.\ N(0, \sigma_k^2) \qquad k = 1, \ldots, 6$$

$$(1-B)\beta_{kt} \;=\; \epsilon_{k2t} \sim i.i.d.\ N(0, \sigma_k^2) \qquad k = 1, \ldots, 5$$

Ex: Model a time series with sampling error

(U.S. teenage unemployment 1/72 – 12/83)

Component model from Bell and Hillmer (1987)

$$y_t = Y_t + e_t$$

$$\begin{aligned} (1 - B)(1 - B^{12})Y_t &= (1 - .27B)(1 - .68B^{12})b_t \\ \sigma_b^2 &= 4,293 \end{aligned}$$

$$e_t = h_t \tilde{e}_t \qquad h_t^2 = \text{var}(e_t) = a y_t + b y_t^2$$

$$(1 - .6B)\tilde{e}_t = (1 - .3B)c_t \quad \sigma_c^2 = .8767 \Rightarrow \text{var}(\tilde{e}_t) = 1$$

Ex: Model a series with time-varying TD coefficients

(U.S. retail sales of department stores, 1/67 - 12/93 ($n = 324$))

- Series has trading-day effects – sales higher on Friday and Saturday than other days, which affects monthly sales figures.

- But, has the pattern changed over time? Let TD coefficients follow random walk models.

United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

Model (Bell 2004): Easter effects plus random walk trading-day coefficients plus airline model residuals

$$\log(y_t/m_t) = .032E_{10,t} + \sum_{i=1}^{6} \beta_{it} D_{it} + z_t$$

$$\begin{aligned} (1-B)(1-B^{12})z_t &= (1-.49B)(1-.52B^{12})a_t \\ \hat{\sigma}_a &= .020 \end{aligned}$$

$$\begin{aligned} (1-B)\beta_{it} &= \epsilon_{it} \quad \epsilon_{it} \sim i.i.d.\ N(0, \sigma_i^2) \\ \beta_{7t} &= -(\beta_{1t} + \cdots + \beta_{6t}) \end{aligned}$$

Estimates of component std deviations and of fixed TD coefficients (expressed in percent given log tranformation)

i (day)	1 (Mon)	2 (Tues)	3 (Wed)	4 (Thurs)	5 (Fri)	6 (Sat)
$\hat{\sigma}_i$	.042	.038	0	0	.046	.025
$n^{.5}\hat{\sigma}_i$	.76	.68	0	0	.83	.45
fixed $\hat{\beta}_i$	$-.44$	.22	$-.73$	.68	.79	.99

fixed $\hat{\beta}_7 = -1.50$

Fig. 2. U.S. Retail Sales of Department Stores, 1/67 - 12/93
original series and estimates of time-varying trading-day coefficients

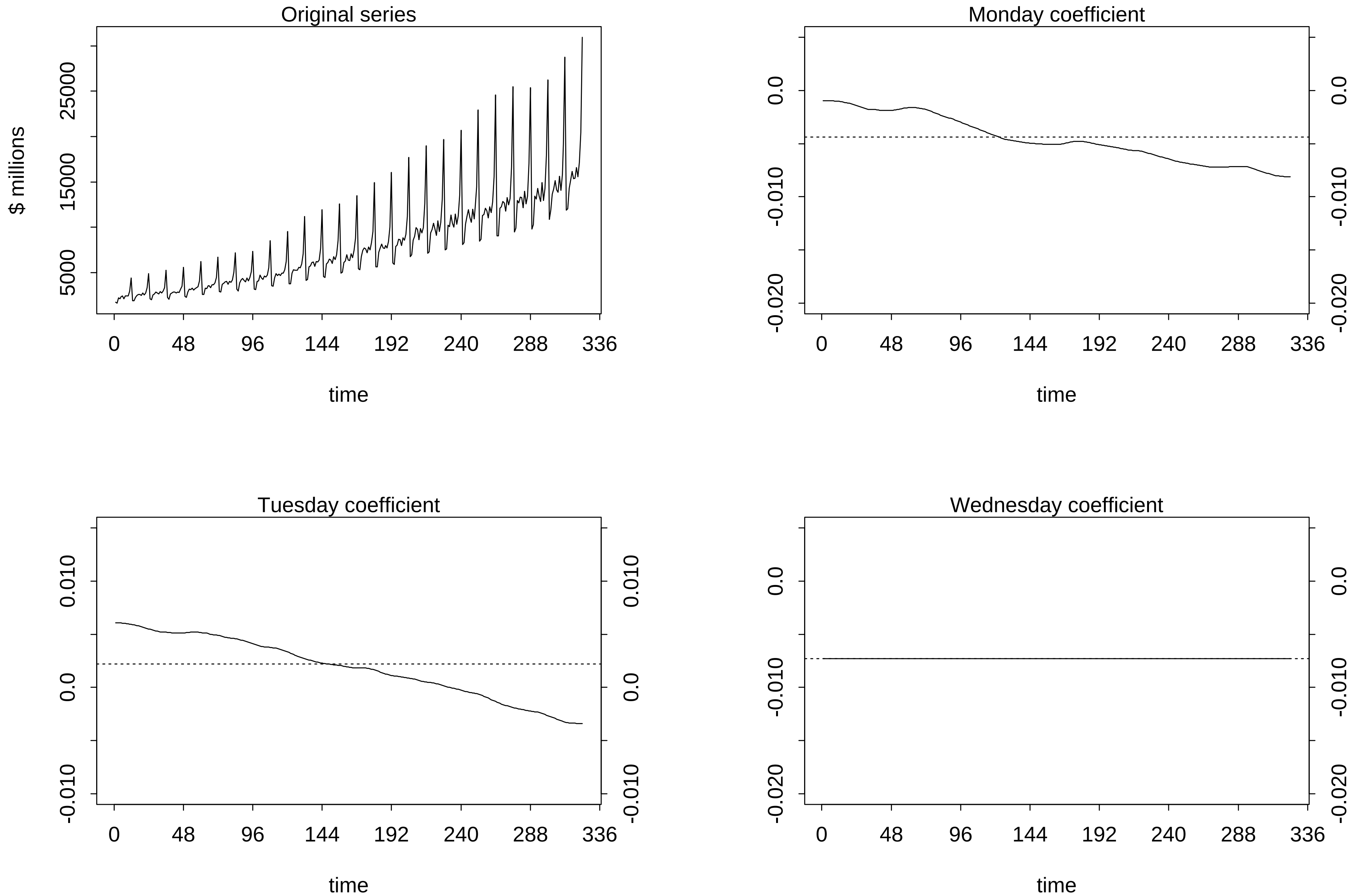

Fig. 2.(cont.) U.S. Retail Sales of Department Stores, 1/67 - 12/93
original series and estimates of time-varying trading-day coefficients