Forecasting demand count series in retail

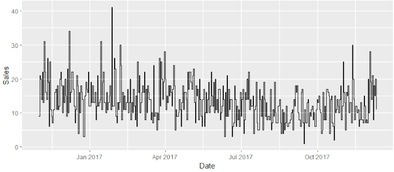
Bruno Flores

Instituto de Ciencias Matemáticas (ICMAT-CSIC)

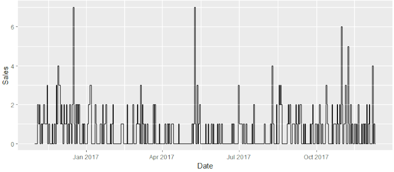
ENBIS-21 Online Conference

13-15 Sep 2021

Count time series , $y_t \in \mathbb{Z}^{\geq 0}$



High demand product.



Intermittent demand.

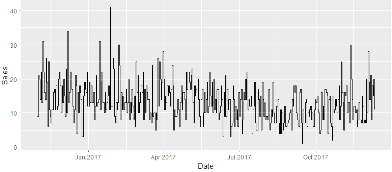
Large retail company

- Thousands of stores
- 4 countries (Europe and the Americas)
- Thousands of references in each store

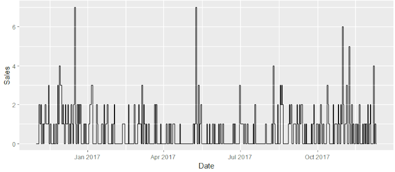
Effects

- Trends (New products, vegan food,...)
- Seasonality (Weekly, yearly)
- Overdispersion
- Price
- Promotion ("Buy one, get one free")
- Substitute goods (Price, stock)

Count time series , $y_t \in \mathbb{Z}^{\geq 0}$



High demand product.



Intermittent demand.

Large retail company

- Thousands of stores
- 4 countries (Europe and the Americas)
- Thousands of references in each store

Effects

- Trends (New products, vegan food,...)
- Seasonality (Weekly, yearly)
- Overdispersion
- Price
- Promotion ("Buy one, get one free")
- Substitute goods (Price, stock)

Objective

Family of models for integration into a DSS:

- **Automatic**. Keep expert intervention to a minimum.
- **Scalable**. Large quantities of data from series and external information
- **Flexible**. High/low demand series, equi/over-dispersion

Approaches

- **Poisson** (Brown, 1959). Problem with **over-dispersion**.
- **Negative binomial** distribution or Poisson but introducing **further randomness** (Snyder et al., 2008).
- **Forecast separately** zero and non-zero values (Croston, 1972) and ideas like **hurdle shifted** or **zero inflated**.
- **Observation-driven** models like INAR, GLARMA or ACP. Less flexible than parameter-driven (Snyder et al., 2008).
- **Bayesian state-space** models (West & Harrison, 1997). With/without use of natural conjugate distributions.

Objective

Family of models for integration into a DSS:

- **Automatic**. Keep expert intervention to a minimum.
- **Scalable**. Large quantities of data from series and external information
- **Flexible**. High/low demand series, equi/over-dispersion

Approaches

- **Poisson** (Brown, 1959). Problem with **over-dispersion**.
- **Negative binomial** distribution or Poisson but introducing **further randomness** (Snyder et al., 2008).
- **Forecast separately** zero and non-zero values (Croston, 1972) and ideas like **hurdle shifted** or **zero inflated**.
- **Observation-driven** models like INAR, GLARMA or ACP. Less flexible than parameter-driven (Snyder et al., 2008).
- **Bayesian state-space** models (West & Harrison, 1997). With/without use of natural conjugate distributions.

Dynamic Generalized Linear Models (DGLM)

Exponential family:

$$p(Y_t|\eta_t, V_t) = exp\{V_t^{-1}[y_t(Y_t)\eta_t - a(\eta_t)]\}b(Y_t, V_t)$$

The **generalized** DLM (West et al., 1985) is defined by

Observation model: $p(Y_t|\eta_t)$ and $g(\eta_t) = \lambda_t = F_t'\theta_t$,

State equation: $\theta_t = G_t\theta_{t-1} + \omega_t$ with $\omega_t \sim (0, W_t)$,

Initial information: $\theta_0 \sim (m_0, C_0)$.

Mixture of DGLMs

Variant of **DCMM** (Berry & West, 2019):
$\{y_t\}$ original time series, $\{z_t\}$ binary time series with $z_t = \mathbb{1}_{(y_t > 0)}$

$$z_t \sim Ber(\pi_t) \quad \text{and} \quad y_t | z_t = \begin{cases} 0, & \text{if } z_t = 0 \\ 1 + x_t, \quad x_t \sim Neg\text{-}Bin(r_t, p_t) & \text{if } z_t = 1 \end{cases}$$

with linear predictors (λ_t) for the Ber and Neg-Bin DGLMs:

$$logit(\pi_t) = F_t^{0\prime} \theta_t^0 \quad \text{and} \quad log(p_t) = F_t^{+\prime} \theta_t^+$$

and state evolution equations:

$$\theta_t^0 = G_t^0 \theta_{t-1}^0 + \omega_t^0 \quad \text{and} \quad \theta_t^+ = G_t^+ \theta_{t-1}^+ + \omega_t^+$$

Note: We consider r_t fixed for all t so that the Neg-Bin belongs to the exponential family. Calculated with initial obs. using EM algorithm. Zero inflated with $y_t \sim Neg\text{-}Bin(r_t, p_t)$.

Mixture of DGLMs

Variant of **DCMM** (Berry & West, 2019):
$\{\mathrm{y_t}\}$ original time series, $\{\mathrm{z_t}\}$ binary time series with $z_t = \mathbb{1}_{(y_t>0)}$

$$z_t \sim Ber(\pi_t) \quad \text{and} \quad y_t|z_t = \begin{cases} 0, & \text{if } \; z_t = 0 \\ 1 + x_t, \quad x_t \sim Neg\text{-}Bin(r_t, p_t) & \text{if } \; z_t = 1 \end{cases}$$

with linear predictors (λ_t) for the Ber and Neg-Bin DGLMs:

$$logit(\pi_t) = F_t^{0'}\theta_t^0 \quad \text{and} \quad log(p_t) = F_t^{+'}\theta_t^+$$

and state evolution equations:

$$\theta_t^0 = G_t^0\theta_{t-1}^0 + \omega_t^0 \quad \text{and} \quad \theta_t^+ = G_t^+\theta_{t-1}^+ + \omega_t^+$$

Note: We consider r_t fixed for all t so that the Neg-Bin belongs to the exponential family. Calculated with initial obs. using EM algorithm. Zero inflated with $y_t \sim Neg\text{-}Bin(r_t, p_t)$.

Fitting and forecasting procedure

- **One step ahead forecast for state $\theta_t|\mathcal{D}_{t-1} \sim (a_t, R_t)$**

$$a_t = G_t m_{t-1} \qquad R_t = G_t C_{t-1} {G_t}' + W_t$$

- **One step ahead forecast for observations**: $(f_t = {F_t}' a_t,\ q = {F_t}' R_t F_t)$

$$z_t|\mathcal{D}_{t-1} \sim Ber(\alpha_t^0/(\alpha_t^0+\beta_t^0)) \quad \text{and} \quad x_t|\mathcal{D}_{t-1} \sim BNB(\beta_t^+ r_t+1, \alpha_t^+, r_t)$$

with hyper-parameters $\alpha_t^0, \alpha_t^+, \beta_t^0, \beta_t^+$ satisfying

$$f^0 = \gamma(\alpha_t^0) - \gamma(\beta_t^0) \qquad q^0 = \dot{\gamma}(\alpha_t^0) + \dot{\gamma}(\beta_t^0)$$
$$f^+ = \gamma(\alpha_t^+) - \gamma(\alpha_t^+ + \beta_t^+ r_t + 1) \qquad q^+ = \dot{\gamma}(\alpha_t^+) - \dot{\gamma}(\alpha_t^+ + \beta_t^+ r_t + 1)$$

- **Update after observing** y_t, $\theta_t|\mathcal{D}_t \sim (m_t, C_t)$

$$m_t = a_t + R_t F_t(\hat{f}_t - f_t)/q_t \qquad C_t = R_t - R_t F_t {F_t}' R_t (1 - \hat{q}_t/q_t)/q_t$$

with

$$\hat{f}_t^0 = \gamma(\alpha_t^0 + z_t) - \gamma(\beta_t^0 + 1 - z_t) \qquad \hat{f}_t^+ = \gamma(\alpha_t^+ + x_t) - \gamma(\alpha_t^+ + x_t + \beta_t^+ r_t + r_t + 1)$$
$$\hat{q}_t^0 = \dot{\gamma}(\alpha_t^0 + z_t) + \dot{\gamma}(\beta_t^0 + 1 - z_t) \qquad \hat{q}_t^+ = \dot{\gamma}(\alpha_t^+ + x_t) - \dot{\gamma}(\alpha_t^+ + x_t + \beta_t^+ r_t + r_t + 1)$$

Fitting and forecasting procedure

- **One step ahead forecast for state** $\theta_t|\mathcal{D}_{t-1} \sim (a_t, R_t)$

$$a_t = G_t m_{t-1} \qquad R_t = G_t C_{t-1} G_t{}' + W_t$$

- **One step ahead forecast for observations**: $(f_t = F_t{}' a_t,\ q = F_t{}' R_t F_t)$

$z_t|\mathcal{D}_{t-1} \sim Ber(\alpha_t^0/(\alpha_t^0+\beta_t^0))$ and $x_t|\mathcal{D}_{t-1} \sim BNB(\beta_t^+ r_t+1, \alpha_t^+, r_t)$

with hyper-parameters $\alpha_t^0, \alpha_t^+, \beta_t^0, \beta_t^+$ satisfying

$$f^0 = \gamma(\alpha_t^0) - \gamma(\beta_t^0) \qquad q^0 = \dot{\gamma}(\alpha_t^0) + \dot{\gamma}(\beta_t^0)$$
$$f^+ = \gamma(\alpha_t^+) - \gamma(\alpha_t^+ + \beta_t^+ r_t + 1) \qquad q^+ = \dot{\gamma}(\alpha_t^+) - \dot{\gamma}(\alpha_t^+ + \beta_t^+ r_t + 1)$$

- **Update after observing** y_t, $\theta_t|\mathcal{D}_t \sim (m_t, C_t)$

$$m_t = a_t + R_t F_t(\hat{f}_t - f_t)/q_t \qquad C_t = R_t - R_t F_t F_t{}' R_t (1 - \hat{q}_t/q_t)/q_t$$

with

$$\hat{f}_t^0 = \gamma(\alpha_t^0 + z_t) - \gamma(\beta_t^0 + 1 - z_t) \qquad \hat{f}_t^+ = \gamma(\alpha_t^+ + x_t) - \gamma(\alpha_t^+ + x_t + \beta_t^+ r_t + r_t + 1)$$
$$\hat{q}_t^0 = \dot{\gamma}(\alpha_t^0 + z_t) + \dot{\gamma}(\beta_t^0 + 1 - z_t) \qquad \hat{q}_t^+ = \dot{\gamma}(\alpha_t^+ + x_t) - \dot{\gamma}(\alpha_t^+ + x_t + \beta_t^+ r_t + r_t + 1)$$

Fitting and forecasting procedure

- **One step ahead forecast for state** $\theta_t|\mathcal{D}_{t-1} \sim (a_t, R_t)$

$$a_t = G_t m_{t-1} \qquad R_t = G_t C_{t-1} G_t{}' + W_t$$

- **One step ahead forecast for observations**: $(f_t = F_t{}' a_t,\ q = F_t{}' R_t F_t)$

$$z_t|\mathcal{D}_{t-1} \sim Ber(\alpha_t^0/(\alpha_t^0+\beta_t^0)) \quad \text{and} \quad x_t|\mathcal{D}_{t-1} \sim BNB(\beta_t^+ r_t+1, \alpha_t^+, r_t)$$

with hyper-parameters $\alpha_t^0, \alpha_t^+, \beta_t^0, \beta_t^+$ satisfying

$$f^0 = \gamma(\alpha_t^0) - \gamma(\beta_t^0) \qquad q^0 = \dot{\gamma}(\alpha_t^0) + \dot{\gamma}(\beta_t^0)$$
$$f^+ = \gamma(\alpha_t^+) - \gamma(\alpha_t^+ + \beta_t^+ r_t + 1) \qquad q^+ = \dot{\gamma}(\alpha_t^+) - \dot{\gamma}(\alpha_t^+ + \beta_t^+ r_t + 1)$$

- **Update after observing** y_t, $\theta_t|\mathcal{D}_t \sim (m_t, C_t)$

$$m_t = a_t + R_t F_t(\hat{f}_t - f_t)/q_t \qquad C_t = R_t - R_t F_t F_t{}' R_t (1 - \hat{q}_t/q_t)/q_t$$

with

$$\hat{f}_t^0 = \gamma(\alpha_t^0 + z_t) - \gamma(\beta_t^0 + 1 - z_t) \qquad \hat{f}_t^+ = \gamma(\alpha_t^+ + x_t) - \gamma(\alpha_t^+ + x_t + \beta_t^+ r_t + r_t + 1)$$
$$\hat{q}_t^0 = \dot{\gamma}(\alpha_t^0 + z_t) + \dot{\gamma}(\beta_t^0 + 1 - z_t) \qquad \hat{q}_t^+ = \dot{\gamma}(\alpha_t^+ + x_t) - \dot{\gamma}(\alpha_t^+ + x_t + \beta_t^+ r_t + r_t + 1)$$

Forecasts

Resulting mixture distribution for the observations is:

$$p(y_t|D_t, \pi_t) = (1 - \pi_t)\delta_0(y_t) + \pi_t h_t(y_t - 1) \quad (1)$$

with $(\pi_t|D_t) \sim Be(\alpha_t^0, \beta_t^0)$, $\quad h_t(y_t - 1) = BNB(y_t - 1|\beta_t^+ r_t + 1, \alpha_t^+, r_t)$.

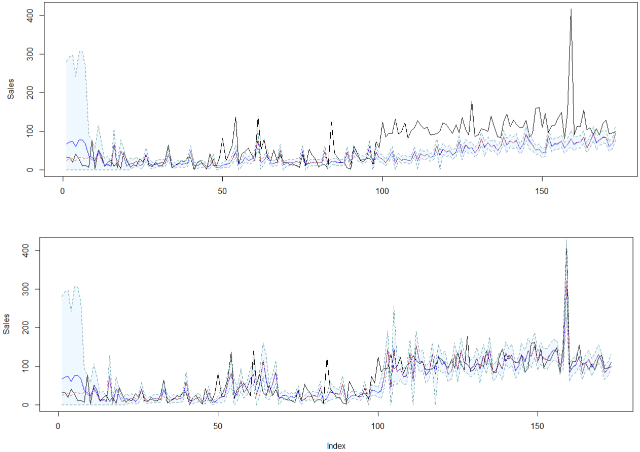
Forecast k steps ahead:

T $\equiv$ "last index of the time series"; N $\equiv$ "number of synthetic paths";
P $\equiv$ "matrix for storing the simulations";
for p *in* $1 : N$ **do**
  for t *in* $T + 1 : T + k$ **do**
    Calculate prior moments for states, $\theta_t|\mathcal{D}_{t-1} \sim (a_t, R_t)$;
    Calculate hyper-parameters $\alpha_t^0, \beta_t^0, \alpha_t^+, \beta_t^+$, use them to draw y_t^* from (1);
    Consider draw as observed value, $y_t = y_t^*$;
    Save the draw, $P_{pt} = y_t$;
    Update state moments with y_t and other relevant information $theta_t|\mathcal{D}_t \sim (m_t, C_t)$;
  end
end

Adaptive discount factors

Difficult specify W_t → use discount factors δ_t ($R_t = G_t C_{t-1} {G_t}' / \delta_t$)

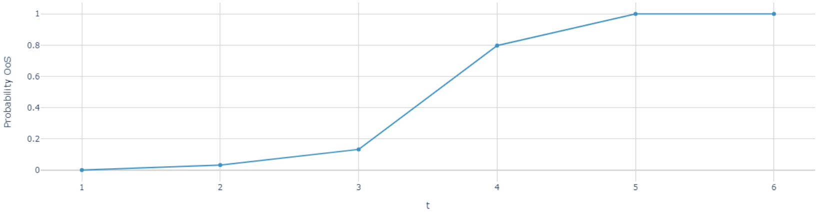
Automatic exception detection and handling routine for δ_t:



Out of Stock (OoS) events

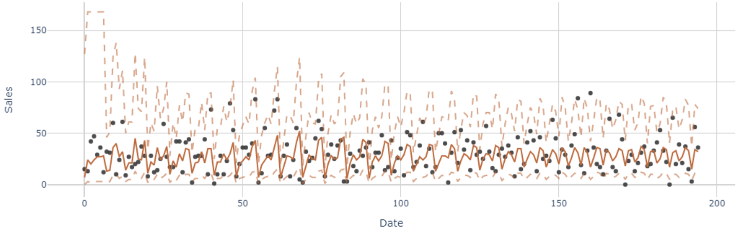
OoS situations entail negative consequences, like economic or image losses.

$$P(stock_t = 0) = P(sales_t > stock_{t-1} + resupply_t)$$

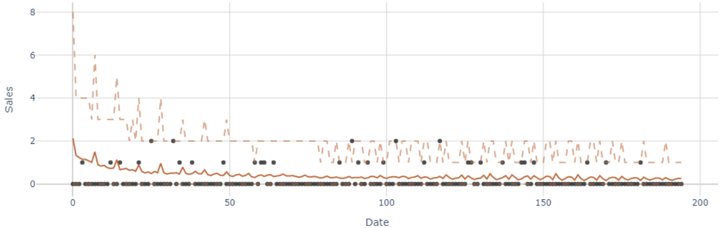


Out of Stock (OoS) probability for beer (SKU '182')

Example



Beer 1 ('SKU 182')



Shaving gel ('SKU 24144')

Multivariate model

Similar to the univariate model:

$y_t = \{\mathrm{y_{it}}\}_{i=1}^{m}$ original ts, $\quad z_t = \{\mathrm{z_{it}}\}_{i=1}^{m}$ binary ts with $z_{it} = \mathbb{1}_{(y_{it}>0)}$

$$
\begin{aligned}
&z_t \sim MBer(p_t) && y_t \sim MNB(r_t, \mu_t)\\
&softmax^{-1}(p_t) = F_t^{0'}{}_t\theta_t^0, && \log(\mu_t) = F_t^{+'}{}_t\theta_t^+,\\
&\theta_t = G_t^0\theta_{t-1}^0 + \omega_t^0 && \theta_t^+ = G_t^+\theta_{t-1}^+ + \omega_t^+
\end{aligned}
$$

with $p_t = (p_{00..00,t},\, p_{00..01,t}, \ldots, p_{11..11,t})$, $\mu_t = (\mu_{1,t}, \ldots, \mu_{m,t})$,
and MBer (Dai et al., 2013) and MNB (Arbous & Kerrich, 1951)

Fitting and forecasting procedure

- **One step ahead forecast for state $\theta_t|\mathcal{D}_{t-1} \sim (a_t, R_t)$**
- **One step ahead forecast for observations**:

$$z_t|\mathcal{D}_{t-1} \sim MBer\Big(\frac{\alpha^0_{00..00,t}}{\beta^0_t}, \ldots, \frac{\alpha^0_{01..11,t}}{\beta^0_t}, \frac{\overbrace{\beta^0_t - \alpha^0_{00..00,t} \cdots - \alpha^0_{01..11,t}}^{b}}{\beta^0_t}\Big)$$

$$h(y_t|\mathcal{D}_{t-1}) = \frac{B(\alpha^+_{1t} + y_{1t}, \ldots, \alpha^+_{mt} + y_{mt}, (\beta^+_t + 1)r_t + 1)}{B(\alpha^+_{1t}, \ldots, \alpha^+_{mt}, \beta^+_t r_t + 1)} \frac{\Gamma(r + \sum_k y_{kt})}{\Gamma(r_t)\prod_k y_{kt}!}.$$

with hyper-parameters satisfying:

$$f^0_t = \begin{pmatrix} \gamma(\alpha^0_{00..00,t}) - \gamma(b) \\ \vdots \\ \gamma(\alpha^0_{01..11,t}) - \gamma(b) \\ 0 \end{pmatrix} \quad Q^0_t = \begin{pmatrix} \dot\gamma(\alpha^0_{00..00,t}) + \dot\gamma(b) & \dot\gamma(b) \cdots & \dot\gamma(b) & 0 \\ \dot\gamma(b) & \ddots & \dot\gamma(b) & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \dot\gamma(b) & \cdot & \dot\gamma(\alpha^0_{01..11,t}) + \dot\gamma(b) & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix}$$

$$f^+_t = \begin{pmatrix} \gamma(\alpha^+_{1t}) - \gamma(\beta^+_t r + 1) + \log(r) \\ \vdots \\ \gamma(\alpha^+_{mt}) - \gamma(\beta^+_t r + 1) + \log(r) \end{pmatrix} \quad Q^+_t = \begin{pmatrix} \dot\gamma(\alpha^+_{1t}) + \dot\gamma(\beta^+ r + 1) & \cdots & \dot\gamma(\beta^+ r + 1) \\ \vdots & \ddots & \vdots \\ \dot\gamma(\beta^+ r + 1) & \cdots & \dot\gamma(\alpha^+_{mt}) + \dot\gamma(\beta^+ r + 1) \end{pmatrix}$$

Fitting and forecasting procedure

- **One step ahead forecast for state** $\theta_t|\mathcal{D}_{t-1} \sim (a_t, R_t)$
- **One step ahead forecast for observations**:

$$z_t|\mathcal{D}_{t-1} \sim MBer\left(\frac{\alpha^0_{00..00,t}}{\beta^0_t}, ..., \frac{\alpha^0_{01..11,t}}{\beta^0_t}, \frac{\overbrace{\beta^0_t - \alpha^0_{00..00,t} \cdots - \alpha^0_{01..11,t}}^{b}}{\beta^0_t}\right)$$

$$h(y_t|\mathcal{D}_{t-1}) = \frac{B(\alpha^+_{1t} + y_{1t}, ..., \alpha^+_{mt} + y_{mt}, (\beta^+_t + 1)r_t + 1)}{B(\alpha^+_{1t}, ..., \alpha^+_{mt}, \beta^+_t r_t + 1)} \frac{\Gamma(r + \sum_k y_{kt})}{\Gamma(r_t) \prod_k y_{kt}!}.$$

with hyper-parameters satisfying:

$$f^0_t = \begin{pmatrix} \gamma(\alpha^0_{00..00,t}) - \gamma(b) \\ \vdots \\ \gamma(\alpha^0_{01..11,t}) - \gamma(b) \\ 0 \end{pmatrix} \quad Q^0_t = \begin{pmatrix} \dot\gamma(\alpha^0_{00..00,t}) + \dot\gamma(b) & \dot\gamma(b) \cdots & \dot\gamma(b) & 0 \\ \dot\gamma(b) & \ddots & \dot\gamma(b) & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \dot\gamma(b) & \cdot & \dot\gamma(\alpha^0_{01..11,t}) + \dot\gamma(b) & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix}$$

$$f^+_t = \begin{pmatrix} \gamma(\alpha^+_{1t}) - \gamma(\beta^+_t r + 1) + \log(r) \\ \vdots \\ \gamma(\alpha^+_{mt}) - \gamma(\beta^+_t r + 1) + \log(r) \end{pmatrix} \quad Q^+_t = \begin{pmatrix} \dot\gamma(\alpha^+_{1t}) + \dot\gamma(\beta^+ r + 1) & \cdots & \dot\gamma(\beta^+ r + 1) \\ \vdots & \ddots & \vdots \\ \dot\gamma(\beta^+ r + 1) & \cdots & \dot\gamma(\alpha^+_{mt}) + \dot\gamma(\beta^+ r + 1) \end{pmatrix}$$

Fitting and forecasting procedure

- **Update after observing** y_t, $\theta_t|\mathcal{D}_t \sim (m_t, C_t)$

$$m_t = a_t + R_t F_t Q_t^{-1}(\hat{f}_t - f_t) \qquad C_t = R_t - R_t F_t(Id - Q_t^{-1}\hat{Q}_t)Q_t^{-1}F_t' R_t$$

with moments of the linear predictor $\hat{f}_t, \hat{Q}_t$ updated with the new hyper-parameters $\hat{\alpha}_t = \alpha_t + \phi_t T(x_t)$, $\hat{\beta}_t = \beta_t + \phi_t$.

The resulting forecast distribution is:

$$p(y_{i,t}|D_t, \pi_{i,t}) = (1 - \pi_{i,t})\delta_0(y_{i,t}) + \pi_{i,t} h_{i,t}(y_{i,t} - 1)$$

From this can obtain median to use as a point forecast credible intervals, with percentiles ρ as

$$\begin{cases} 0, & \text{if } (1 - \pi_{i,t}) >= \rho \\ Q_i((\rho - (1 - \pi_{i,t}))/\pi_{i,t}; \alpha_t^+, \beta_t^+, r_t) + 1 & \text{otherwise} \end{cases}$$

Fitting and forecasting procedure

- **Update after observing** $\mathrm{y_t}$, $\theta_t | \mathcal{D}_t \sim (m_t, C_t)$

$$m_t = a_t + R_t F_t Q_t^{-1} (\hat{f}_t - f_t) \qquad C_t = R_t - R_t F_t (Id - Q_t^{-1} \hat{Q}_t) Q_t^{-1} F_t' R_t$$

with moments of the linear predictor $\hat{f}_t$, $\hat{Q}_t$ updated with the new hyper-parameters $\hat{\alpha}_t = \alpha_t + \phi_t T(x_t)$, $\hat{\beta}_t = \beta_t + \phi_t$.

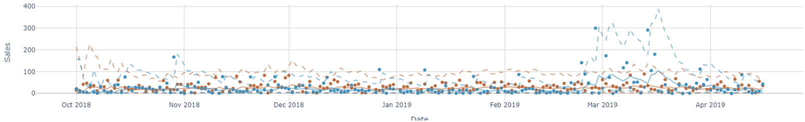
The resulting forecast distribution is:

$$p(y_{i,t} | D_t, \pi_{i,t}) = (1 - \pi_{i,t}) \delta_0(y_{i,t}) + \pi_{i,t} h_{i,t}(y_{i,t} - 1)$$

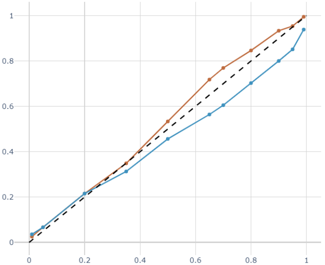
From this can obtain median to use as a point forecast credible intervals, with percentiles ρ as

$$\begin{cases} 0, & \text{if } (1 - \pi_{i,t}) >= \rho \\ Q_i((\rho - (1 - \pi_{i,t}))/\pi_{i,t}; \alpha_t^+, \beta_t^+, r_t) + 1 & \text{otherwise} \end{cases}$$

Fitting and forecasting procedure



Forecasts for Beer 1 and Beer 2 (SKU '182','29352')



Coverage plot of credible intervals for Beer 1 and Beer 2 (SKU '182','29352')

Discussion & Future work

- We provided methodology to forecast product demand taking into account effects encountered in retail
- The family of models introduced satisfies the objectives improve the performance of models commonly used in this application domain.
- Fundamental part of a Decision Support System developed to manage inventory in a large retail company, feeding sales forecasts to other modules.

- In other applications it might be worth considering to account for possible under-dispersion (through the use of CMP distribution) or model separately the number of transactions and the number of products per transaction (Berry et al., 2020) in case that information is available.
- Explore more in depth approaches at other aggregation levels

Discussion & Future work

- We provided methodology to forecast product demand taking into account effects encountered in retail
- The family of models introduced satisfies the objectives improve the performance of models commonly used in this application domain.
- Fundamental part of a Decision Support System developed to manage inventory in a large retail company, feeding sales forecasts to other modules.

- In other applications it might be worth considering to account for possible under-dispersion (through the use of CMP distribution) or model separately the number of transactions and the number of products per transaction (Berry et al., 2020) in case that information is available.
- Explore more in depth approaches at other aggregation levels

References

Arbous, A. G., & Kerrich, J. (1951). Accident statistics and the concept of accident proneness. *Biometrics*, *7*, 340–432.

Berry, L. R., Helman, P., & West, M. (2020). Probabilistic forecasting of heterogeneous consumer transaction-sales time series. *International Journal of Forecasting*, *36*, 552–569.

Berry, L. R., & West, M. (2019). Bayesian forecasting of many count-valued time series. *Journal of Business and Economic Statistics*, *38*, 872–887.

Brown, R. G. (1959). *Statistical forecasting for inventory control*. McGraw-Hill.

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, *23*, 289–303.

Dai, B., Ding, S., & Wahba, G. (2013). Multivariate bernoulli distribution. *Bernoulli Society for Mathematical Statistics and Probability*, *19*, 1465–1483.

Snyder, R. D., Martin, G. M., & Gould, P. (2008). An assessment of alternative state space models for count time series. *Monash*.

West, M., & Harrison, J. (1997). *Bayesian forecasting and dynamic models*. Springer.

West, M., Harrison, J., & Migon, H. (1985). Dynamic generalized models and bayesian forecasting. *Journal of the American Statistical Association*, *80*, 73–97.

Contact Information

Bruno Flores
Predoctoral researcher
Institute of Mathematical Sciences (ICMAT)
Nicolás Cabrera, 13-15, 28049 Madrid
+34 912 999 759
bruno.flores@icmat.es