Forecasting Call Center Arrivals: A Comparative Study

Rouba Ibrahim, University of Montreal

Joint work with:
Pierre L'Ecuyer, University of Montreal

Overview

Existing Forecasts at Company

- Predictions for daily totals only
- Lead times:
  - "Scheduling forecast" (made 2-3 weeks in advance)
  - "Last intraday forecast" (updated about one day in advance)

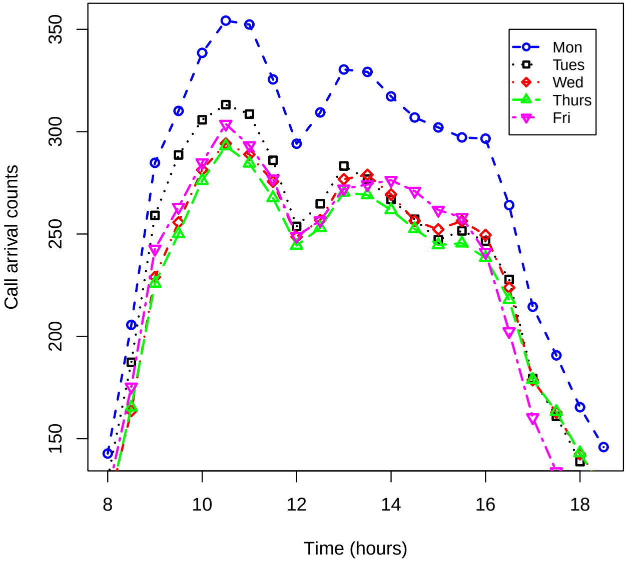
Project

- Develop interval (30 min) predictions
- Dependence structures (interday, intraday)
- Lead times: from weeks to hours

Brief Description of Data

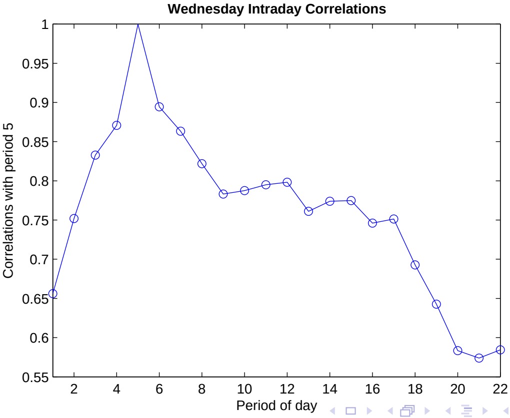
- Call type: Type A
- Hundreds of thousands of calls per month
- Arrival counts per period (30 mins)
- Data collected over $D = 329$ days (Oct. 2009 - Nov. 2010)
- Different arrival pattern on Saturdays $\Rightarrow$ Focus on weekdays

Intraday Seasonality



Interday and Intraday Correlations

	Mon	Tues.	Wed.	Thurs.	Fri.
Mon.	1.0	0.48	0.35	0.35	0.34
Tues.		1.0	0.68	0.62	0.62
Wed.			1.0	0.72	0.67
Thurs.				1.0	0.80
Fri.					1.0



Data Transformation

- Day index: $d \in \{1, 2, 3, .., D\}$
- Half-hour interval index: $p \in \{1, 2, 3, ..., P\}$
- N_{dp}: Number of arrivals in *pth* interval of day d

Basic assumption:

$$N_{dp} \sim \text{Poisson}(\lambda_{dp})$$

Variance-Stabilizing Transformation:

$$y_{dp} = \sqrt{N_{dp} + 1/4}$$

Then, for λ_{dp} large:

$$y_{dp} \approx Nor(\sqrt{\lambda_{dp}}, 1/4)$$

Brown, Zhang, and Zhao (2001).

Gaussian Linear Mixed Models (LMM)

$$y = X\beta + Z\gamma + \epsilon$$

- $y = (y_{11}, y_{12}, ..., y_{1P}, ..., y_{D1}, y_{D2}, ..., y_{DP})'$
- X: $(DP \times r)$-design matrix for *fixed effects*
- $\beta = (\beta_1, ..., \beta_r)'$: r-vector of fixed effect coefficients
- Z: $(DP \times s)$-design matrix for *random effects*
- $\gamma = (\gamma_1, ..., \gamma_s)'$: s-vector of random effects
- ϵ: DP-vector of random residual effects

That is,

$$y_{dp} = \sum_{i=1}^{r} x_{dp,i}\beta_i + \sum_{j=1}^{s} z_{dp,j}\gamma_j + \epsilon_{dp}$$

where $x_{dp,i} \in \{0, 1\}$ and $z_{dp,j} \in \{0, 1\}$.

Aldor-Noiman, Feigin, and Mandelbaum (2009).

Model Selection

Selected Fixed Effects:

- Day of Week
- Period of Day
- Cross terms: Day of Week $\times$ Period of Day

Random Day Effects γ: Interday Dependence

- Daily deviation from fixed weekday effect
- $Var[\gamma] = G$
- Autoregressive AR(1) covariance structure: σ_G^2 and ρ_G

Residuals ϵ: Intraday Dependence

- Period-by-period deviation from observed values
- $Var[\epsilon] = R^* + \sigma^2 I_P$
- R^* has an AR(1) covariance structure: $\sigma_{R^*}^2$ and ρ_{R^*}

Under our model assumptions:

$$\sigma^2 \approx 0.25$$

Other Models

Top-Down Approach:

Forecast for period k of day d:

$$\hat{y}_{dp} = \hat{y}_d \times \hat{p}_{q_d,p} \ ,$$

where

- $q_d = \{1, 2, 3, 4, 5\}$ is type of day d
- $\hat{y}_d$ = daily volume forecast for day d (Bell forecast)
- $\hat{p}_{q_d,p}$ = point estimate of proportion of calls in period p of day type q_d

Benchmark Model 1: Fixed Effects Model

- Same fixed effects as selected LMM
- No random effects
- Independent residuals

Benchmark "Model" 2: Holt Winters

- No model assumptions
- Additive daily seasonality
- No trend

Performance Metrics

Let $\hat{N}_{dp}$ be the predicted value of N_{dp}.

Measures Per Period

- Squared Error: $\text{SE}_{dp} = (\hat{N}_{dp} - N_{dp})^2$
- Relative Error: $\text{RE}_{dp} = 100.\frac{|\hat{N}_{dp} - N_{dp}|}{N_{dp}}$
- $\text{Cover}_{dp} = \mathbf{I}(N_{dp} \in (\text{Lower}_{dp}, \text{Upper}_{dp}))$
- $\text{Width}_{dp} = \text{Upper}_{dp}$ - Lower_{dp}

Predictions

- Forecast lead time: 1 day, 1 week, 2 weeks
- Forecast horizon: 85 days between Aug 19 and Nov 11, 2010
- 1320 predicted values
- Learning period: all previous days
- Roll horizon for each day

Performance of the Models

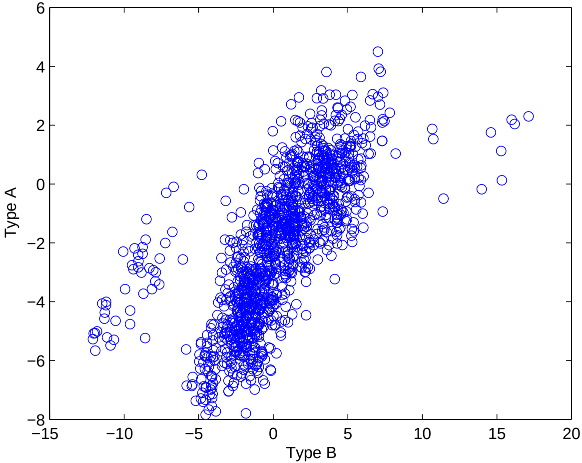
Predictions for a Forecast Lead-Time of 2 weeks

	RMSE	APE	Coverage	Width
LMM	41.7	16.4	0.95	182
Bell	45.7	18.2	–	–
Fixed	40.3	15.3	0.22	22.4
Holt Winters	67.0	29.0	–	–

Predictions for a Forecast Lead-Time of 1 Day

	RMSE	APE	Coverage	Width
LMM	30.4	12.9	0.96	157
Bell	33.9	13.9	–	–
Fixed	35.7	15.1	0.22	21.9
Holt Winters	60.8	26.4	–	–

Correlations Between Type A and Type B Calls



Estimated correlation $= 0.71$.

New Model: Bivariate LMM

$$y_A = X\beta_A + Z\gamma_A + \epsilon_A$$

$$y_B = X\beta_B + Z\gamma_B + \epsilon_B$$

Interday Dependence

- γ_A and γ_B:
  - Daily deviation from fixed weekday effect
  - γ_A is independent of γ_B
  - AR(1) covariance structures: σ_A^2, ρ_A; σ_B^2, ρ_B

Intraday Dependence

- ϵ_A and ϵ_B are correlated with covariance matrix

$$\begin{pmatrix} \sigma_A^2 & \sigma_{A,B} \\ \sigma_{B,A} & \sigma_B^2 \end{pmatrix} \otimes \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^P \\ \rho & 1 & \rho & \dots & \rho^{P-1} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \rho^P & \rho^{P-1} & \rho^{P-2} & \dots & 1 \end{pmatrix}$$

Comparison with LMM: Forecast Lead Time 1 Day

- Forecast horizon: August 19, 2010 - November 11, 2010
- Learning period: 58 days

Predictions for a Forecast Lead Time of 1 Day

	RMSE	APE	Coverage	Width
Biv. LMM	33.4	14.2	0.90	128
LMM	39.0	16.4	0.86	126

Predictions for a Forecast Lead Time of 1/2 Day

	RMSE	APE	Coverage	Width
Biv. LMM	28.5	10.9	0.92	102
LMM	30.1	11.8	0.90	103

$\Rightarrow$ **Obtain better forecasts!**