Aiding Television Media Planning Through Bayesian Inference and Forecasting

Matthew Tiger

Towson University

May 2018

Background - TV Advertising Buying and Selling

- The TV advertising landscape consists of TV sellers, who have airtime available for sale to be used to air advertisements, and TV buyers, who purchase airtime from TV sellers in order to air their desired advertisements.

Background - TV Advertising Buying and Selling

- The TV advertising landscape consists of TV sellers, who have airtime available for sale to be used to air advertisements, and TV buyers, who purchase airtime from TV sellers in order to air their desired advertisements.
- TV sellers form *media plans* based on the goals of TV buyers.

Background - TV Advertising Buying and Selling

- The TV advertising landscape consists of TV sellers, who have airtime available for sale to be used to air advertisements, and TV buyers, who purchase airtime from TV sellers in order to air their desired advertisements.
- TV sellers form *media plans* based on the goals of TV buyers.
- With each media plan, there is a certain number of guaranteed *impressions* based off a declared buy-demographic.

Background - TV Advertising Buying and Selling

- The TV advertising landscape consists of TV sellers, who have airtime available for sale to be used to air advertisements, and TV buyers, who purchase airtime from TV sellers in order to air their desired advertisements.
- TV sellers form *media plans* based on the goals of TV buyers.
- With each media plan, there is a certain number of guaranteed *impressions* based off a declared buy-demographic.

Background - TV Advertising Buying and Selling

- To complicate the matter, the TV buyer may be interested in a sub-target audience and would want their advertising message to reach those additional people as well.

Background - TV Advertising Buying and Selling

- To complicate the matter, the TV buyer may be interested in a sub-target audience and would want their advertising message to reach those additional people as well.
- The TV seller must then provide forecasts for both targets and categorize the content according to what will efficiently target the sub-target audience. Inaccurate forecasts lead to either upset TV buyers or upset TV sellers.

Background - TV Advertising Buying and Selling

- To complicate the matter, the TV buyer may be interested in a sub-target audience and would want their advertising message to reach those additional people as well.
- The TV seller must then provide forecasts for both targets and categorize the content according to what will efficiently target the sub-target audience. Inaccurate forecasts lead to either upset TV buyers or upset TV sellers.
- The forecasts provided by the TV seller are based off audience measurement data. This data is based off of sampled data and could be noisy.

Problem

How can we forecast impressions of this sub-target given the forecasted impressions of the buy-demographic using noisy data?

Problem

How can we forecast impressions of this sub-target given the forecasted impressions of the buy-demographic using noisy data?
Answer: Bayesian Inference!

Motivating Example: Baseball

- Baseball players are judged by their batting average (percentage of hits) but this metric is not informative when the player has few at-bats.

Motivating Example: Baseball

- Baseball players are judged by their batting average (percentage of hits) but this metric is not informative when the player has few at-bats.
- With more information about the league and past historical performances, we are able to come up with a better estimate through Bayesian inference.

Motivating Example: Baseball

- Baseball players are judged by their batting average (percentage of hits) but this metric is not informative when the player has few at-bats.
- With more information about the league and past historical performances, we are able to come up with a better estimate through Bayesian inference.
- Note that these estimates take into account the observed number of at-bats for each player which places larger evidence of skill, or lack thereof, on players with larger numbers of at-bats.

Bayesian Inference

- Bayesian inference is the statistical practice that allows one to provide probability statements that express one's uncertainty on the occurrence of events and then learn about the parameters that determine those probabilities.

Bayesian Inference

- Bayesian inference is the statistical practice that allows one to provide probability statements that express one's uncertainty on the occurrence of events and then learn about the parameters that determine those probabilities.
- A model for the data y conditional on some unknown parameter θ is assigned as the *likelihood* denoted by $p(y|\theta)$.

Bayesian Inference

- Bayesian inference is the statistical practice that allows one to provide probability statements that express one's uncertainty on the occurrence of events and then learn about the parameters that determine those probabilities.
- A model for the data y conditional on some unknown parameter θ is assigned as the *likelihood* denoted by $p(y|\theta)$.
- Based on prior knowledge some probability distribution is given to $p(\theta)$, i.e. the prior distribution for the parameter θ.

Bayesian Inference

- Bayesian inference is the statistical practice that allows one to provide probability statements that express one's uncertainty on the occurrence of events and then learn about the parameters that determine those probabilities.
- A model for the data y conditional on some unknown parameter θ is assigned as the *likelihood* denoted by $p(y|\theta)$.
- Based on prior knowledge some probability distribution is given to $p(\theta)$, i.e. the prior distribution for the parameter θ.
- Through the definition of conditional probability, we have that:

$$p(\theta|y) \propto p(y|\theta)p(\theta). \tag{1}$$

Datasets

- The data used to generate the forecasts come from two main sources TV suppliers and audience measurement companies.

Datasets

- The data used to generate the forecasts come from two main sources TV suppliers and audience measurement companies.
- TV suppliers provide the content that is set to air and the forecasted buy-demographic impressions

Datasets

- The data used to generate the forecasts come from two main sources TV suppliers and audience measurement companies.
- TV suppliers provide the content that is set to air and the forecasted buy-demographic impressions
- Audience measurement companies provide the measured historical impressions.

Datasets

- The data used to generate the forecasts come from two main sources TV suppliers and audience measurement companies.
- TV suppliers provide the content that is set to air and the forecasted buy-demographic impressions
- Audience measurement companies provide the measured historical impressions.
- We will now discuss these datasets in more detail.

Programming Schedule

See below for sample records from the programming schedule provided by TV suppliers:

network	selling title	selling title name	content	content name	start datetime	end datetime
BCST	100	Adult Cartoon 8PM	10	Adult Cartoon	2017-04-02 20:00:00	2017-04-02 20:30:00
BCST	101	Adult Cartoon 8:30PM	10	Adult Cartoon	2017-04-02 20:30:00	2017-04-02 21:00:00

Forecasted Impressions

Sample records from the forecasted impressions data provided by TV suppliers:

selling title	broadcast week	demographic	impressions per unit
100	2017-03-27 06:00:00	F45-49	150000
100	2017-03-27 06:00:00	P18-49	1500000
101	2017-03-27 06:00:00	F45-49	120000
101	2017-03-27 06:00:00	P18-49	1000000

Audience Measurement - Programs

Sample records from Audience Measurement Programs Data:

airing	network	program	telecast	program name	start datetime	end datetime	genre	is first run	is live
35	BCST	1000	301	Adult Cartoon	2017-04-02 20:02:00	2017-04-02 20:30:00	Animation	1	0
36	BCST	1000	302	Adult Cartoon	2017-04-02 20:30:00	2017-04-02 21:00:00	Animation	1	0

Audience Measurement - Viewing

Sample records from Audience Measurement Viewing Data:

program	telecast	respondent	minute	comm secs	weight	age	gender	total comm secs
1000	301	2	3	60	2050	48	F	120
1000	301	2	4	45	2050	48	F	120
1000	301	2	15	60	2050	48	F	120
1000	301	2	16	30	2050	48	F	120
1000	302	2	22	15	2050	48	F	100
1000	302	2	23	60	2050	48	F	100

- The impressions associated to a telecast is given by the Average Commercial Minute (ACM).

Audience Measurement - Measurements

- The impressions associated to a telecast is given by the Average Commercial Minute (ACM).
- This is the weighted sum of the commercial viewing of a target over the total number of commercial seconds.

Audience Measurement - Measurements

- The impressions associated to a telecast is given by the Average Commercial Minute (ACM).
- This is the weighted sum of the commercial viewing of a target over the total number of commercial seconds.
- Define m_i^A, the ACM of target A for airing i as follows:

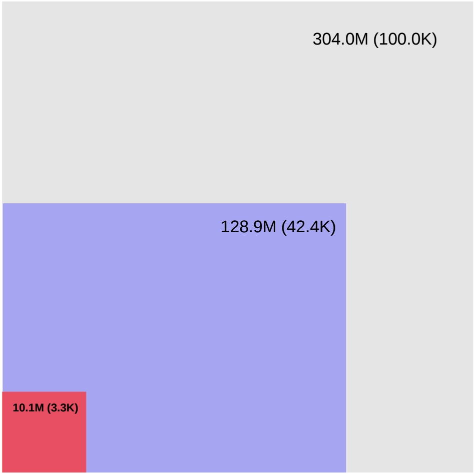
$$m_i^A = \left\lceil \frac{\sum_{k=1}^{n} w_k \sum_{j=1}^{t_i} s_{ij} p_{ijk} \mathbf{1}_A(p_k)}{\sum_{j=1}^{t_i} s_{ij}} \right\rceil . \tag{2}$$

Training Data

- We limit the data from the above data sets to three networks labeled BCST, ETMT, and SPTS during broadcast years 2016 - 2017.
- We consider the in-target audience, target A, to be Females aged 45 - 49 and the buy-demographic audience, target B, to be Persons aged 18-49.
- We combine the information from the above data sets to form the training data:

network	selling title	content	start datetime	end datetime	program	telecast	ACM A	ACM B
BCST	100	10	2017-04-02 20:00:00	2017-04-02 20:30:00	1000	301	110560	1203560
BCST	101	10	2017-04-02 20:30:00	2017-04-02 21:00:00	1000	302	210560	1501000

Audience Size Comparison



Units of Observation

- The units of observation for this model are individual content airings in the media schedule.

Units of Observation

- The units of observation for this model are individual content airings in the media schedule.
- The outcome variable that is measured for each unit i is m_i^A, the ACM of target A, which we denote as y_i for notational convenience.

Units of Observation

- The units of observation for this model are individual content airings in the media schedule.
- The outcome variable that is measured for each unit i is m_i^A, the ACM of target A, which we denote as y_i for notational convenience.
- Similarly, we denote m_i^B, the ACM of target B for unit i by n_i.

Covariates

- *Time-based* covariates and **Program-based** covariates

Covariates

- *Time-based* covariates and **Program-based** covariates
- Derived from Media Schedule
  - *Broadcast Month*
  - *Day of Week*
  - *Stratified Hour*
  - **Content**
  - **Lead-in Content**

Covariates

- *Time-based* covariates and **Program-based** covariates
- Derived from Media Schedule
  - *Broadcast Month*
  - *Day of Week*
  - *Stratified Hour*
  - **Content**
  - **Lead-in Content**
- Derived from Audience Measurement Data
  - **Genre**
  - **Live-program**
  - **First-run**

Assumptions

- We assume that the response variables y_i are *exchangeable* given the parameters of the model and the covariates of the unit of observation.

Assumptions

- We assume that the response variables y_i are *exchangeable* given the parameters of the model and the covariates of the unit of observation.
- A sequence of random variable is exchangeable if the "joint probability density $p(y_1, \ldots, y_k)$ is invariant to permutations of the indexes."

Assumptions

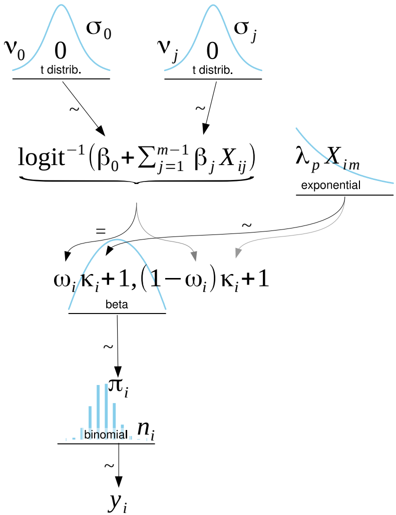
- We assume that the response variables y_i are *exchangeable* given the parameters of the model and the covariates of the unit of observation.
- A sequence of random variable is exchangeable if the "joint probability density $p(y_1, \ldots, y_k)$ is invariant to permutations of the indexes."
- This allows us to model the data as independently and identically distributed given the covariates and unknown parameters.

Model Description

Define model $\mathcal{M}$ to be

$$
\begin{aligned}
y_i | X_i, n_i, \pi_i, \omega_i, \kappa_i &\sim \text{Bin}(n_i, \pi_i) \\
\pi_i | \omega_i, \kappa_i &\sim \text{Beta}\left(\omega_i \kappa_i + 1, (1 - \omega_i)\kappa_i + 1\right) \\
\omega_i &= \text{logit}^{-1}\left(\beta_0 + \sum_{j=1}^{m-1} \beta_j X_{ij}\right), \quad \beta_j \sim t_4(0, \sigma_j^2) \\
&\text{for } 0 \le j \le m \\
\kappa_i | X_{im} &\sim \text{Exp}(\lambda_p X_{im}), \quad \text{for } p = 0, 1,
\end{aligned}
$$

where $\text{logit}^{-1}(\alpha) = \frac{\exp \alpha}{1 + \exp \alpha}$.


ν_0 0 σ_0
t distrib.
ν_j 0 σ_j
t distrib.
~
~
$\text{logit}^{-1}(\beta_0 + \sum_{j=1}^{m-1} \beta_j X_{ij})$
$\lambda_p X_{im}$
exponential
=
~
$\omega_i \kappa_i + 1, (1 - \omega_i) \kappa_i + 1$
beta
~
π_i
binomial n_i
~
y_i

Prior Distribution Choice

- Prior distributions for coefficients β_j and concentration parameter κ_i are chosen to be *weakly informatative*.

Prior Distribution Choice

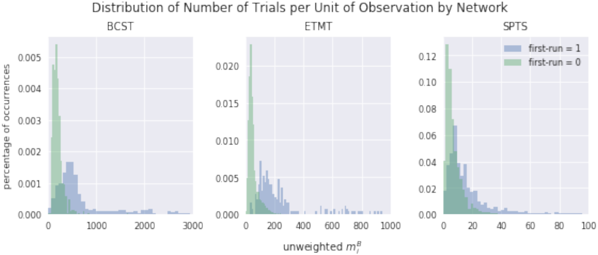
- Prior distributions for coefficients β_j and concentration parameter κ_i are chosen to be *weakly informatative*.
- For the coefficients, this means that $\mu_j = 0, \nu_j = 4$ for all j and that $\sigma_j = 2.5$ if $1 \leq j \leq m$ otherwise $\sigma_0 = 5$.

Prior Distribution Choice

- Prior distributions for coefficients β_j and concentration parameter κ_i are chosen to be *weakly informatative*.
- For the coefficients, this means that $\mu_j = 0, \nu_j = 4$ for all j and that $\sigma_j = 2.5$ if $1 \leq j \leq m$ otherwise $\sigma_0 = 5$.
- For the concentration parameter, this means that $\lambda_p = 10^{-4}$.

Prior Distribution Choice

- Prior distributions for coefficients β_j and concentration parameter κ_i are chosen to be *weakly informatative*.
- For the coefficients, this means that $\mu_j = 0, \nu_j = 4$ for all j and that $\sigma_j = 2.5$ if $1 \leq j \leq m$ otherwise $\sigma_0 = 5$.
- For the concentration parameter, this means that $\lambda_p = 10^{-4}$.



Computation

- Inference was computed using pymc3, a probabilistic programming language and library for Python.

Computation

- Inference was computed using pymc3, a probabilistic programming language and library for Python.
- The library is powered by the No U-Turn Sampler (NUTS) which is a variant of Hamiltonian Monte Carlo (HMC).

Computation

- Inference was computed using pymc3, a probabilistic programming language and library for Python.
- The library is powered by the No U-Turn Sampler (NUTS) which is a variant of Hamiltonian Monte Carlo (HMC).
- Parameters used for sampling:
  - `target_accept`: 0.95
  - tuned samples: 3000
  - drawn samples: 500
  - number of chains: 4

Convergence

- Approximate convergence to posterior distribution is measured through the *Gelman-Rubin* statistic, denoted by $\hat{R}$.

Convergence

- Approximate convergence to posterior distribution is measured through the *Gelman-Rubin* statistic, denoted by $\hat{R}$.
- Another convergence check is the number of effective samples produced by the simulation, denoted by $\hat{n_{\text{eff}}}$.

Convergence

- Approximate convergence to posterior distribution is measured through the *Gelman-Rubin* statistic, denoted by $\hat{R}$.
- Another convergence check is the number of effective samples produced by the simulation, denoted by $\hat{n_{\text{eff}}}$.
- If $\hat{R}$ is close to 1 then we may assume we have approximate convergence. Further it is recommended that $\hat{n_{\text{eff}}} \geq 10M$ where M is the number of sampled Markov chains for all model parameters.

Convergence

- Approximate convergence to posterior distribution is measured through the *Gelman-Rubin* statistic, denoted by $\hat{R}$.
- Another convergence check is the number of effective samples produced by the simulation, denoted by $\hat{n_{\text{eff}}}$.
- If $\hat{R}$ is close to 1 then we may assume we have approximate convergence. Further it is recommended that $\hat{n_{\text{eff}}} \geq 10M$ where M is the number of sampled Markov chains for all model parameters.
- For each network model, we have that $0.99 \leq \hat{R} \leq 1.01$ and $\hat{n_{\text{eff}}} > 400$ for all model parameters.

Posterior Predictive Checks

- "If the model fits, then replicated data under the model should look similar to observed data."

Posterior Predictive Checks

- "If the model fits, then replicated data under the model should look similar to observed data."
- Generating data using the posterior density and checking some aspect of the generated data set is called a *posterior predictive check*.

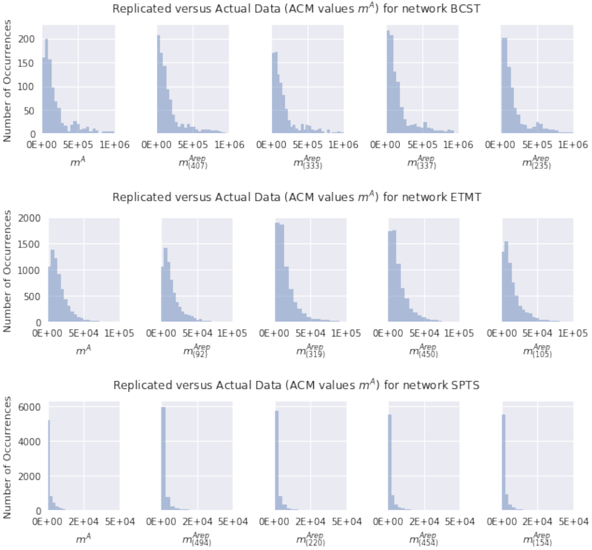
Posterior Predictive Checks

- "If the model fits, then replicated data under the model should look similar to observed data."
- Generating data using the posterior density and checking some aspect of the generated data set is called a *posterior predictive check*.

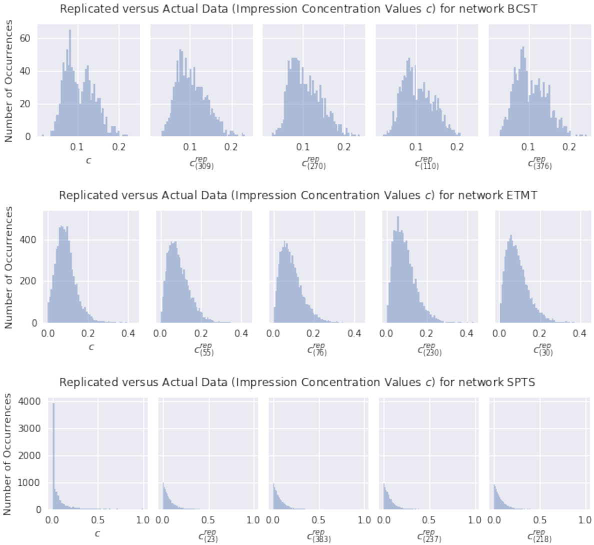
Let y be the observed data and θ be the vector of model parameters. Define y^{rep} to be the replicated data that could have been generated given θ, i.e.

$$p(y^{\text{rep}}|y) = \int p(y^{\text{rep}}|\theta)p(\theta|y)d\theta. \tag{3}$$

Replicated versus Actual Data



Replicated versus Actual Data



Test Statistics

- We can quantify model discrepancies by defining a test quantity $T(y, \theta)$ and then measuring the discrepancy between the observed data and the replicated data.

Test Statistics

- We can quantify model discrepancies by defining a test quantity $T(y, \theta)$ and then measuring the discrepancy between the observed data and the replicated data.
- Formally, we can compute a posterior predictive p-value defined as

$$p_B = \Pr\left(T(y^{\text{rep}}, \theta) \geq T(y, \theta) | y\right).$$

Test Statistics

- We can quantify model discrepancies by defining a test quantity $T(y, \theta)$ and then measuring the discrepancy between the observed data and the replicated data.
- Formally, we can compute a posterior predictive p-value defined as

$$p_B = \Pr\left(T(y^{\text{rep}}, \theta) \geq T(y, \theta) | y\right).$$

- Since we use simulated values of the posterior density, we have that the estimated p-value for S simulations is given by:

$$\hat{p_B} = \frac{1}{S} \sum_{i=1}^{S} [T(y_{(i)}^{\text{rep}}, \theta_{(i)}) \geq T(y, \theta_{(i)})]. \tag{4}$$

Test Statistics - Definition

We define the following test quantities to use in evaluating the fit of model $\mathcal{M}$:

- $T_1(y, \theta) := \min(y)$,

Test Statistics - Definition

We define the following test quantities to use in evaluating the fit of model $\mathcal{M}$:

- $T_1(y, \theta) := \min(y)$,
- $T_2(y, \theta) := \overline{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$,

Test Statistics - Definition

We define the following test quantities to use in evaluating the fit of model $\mathcal{M}$:

- $T_1(y, \theta) := \min(y)$,
- $T_2(y, \theta) := \overline{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$,
- $T_3(y, \theta) := \max(y)$,

Test Statistics - Definition

We define the following test quantities to use in evaluating the fit of model $\mathcal{M}$:

- $T_1(y, \theta) := \min(y)$,
- $T_2(y, \theta) := \overline{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$,
- $T_3(y, \theta) := \max(y)$,
- $T_4(y, \theta) := \text{std}(y) = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \overline{y})^2}{N-1}}$.

Test Statistics - Evaluation - BCST network

Test quantity	$T(y,\theta)$	95% int. for $T(y^{\text{rep}},\theta)$	p_B
$T_1(y,\theta)$ (min)	3701	[6245, 14270]	0.99
$T_2(y,\theta)$ (mean)	227457.84	[2266852.49, 236367.09]	0.95
$T_3(y,\theta)$ (max)	4311038	[3443885, 4989241]	0.34
$T_4(y,\theta)$ (std)	334052.86	[325128.37, 364859.10]	0.90

Test Statistics - Evaluation - ETMT network

Test quantity	$T(y, \theta)$	95% int. for $T(y^{\text{rep}}, \theta)$	p_B
$T_1(y, \theta)$ (min)	0	[9, 182]	1.0
$T_2(y, \theta)$ (mean)	16357.80	[16705.39, 17489.11]	1.0
$T_3(y, \theta)$ (max)	452762	[307901, 760822]	0.78
$T_4(y, \theta)$ (std)	17686.89	[20021.24, 23205.09]	1.0

Test Statistics - Evaluation - SPTS network

Test quantity	$T(y,\theta)$	95% int. for $T(y^{\text{rep}},\theta)$	p_B
$T_1(y,\theta)$ (min)	0	[0, 0]	1.0
$T_2(y,\theta)$ (mean)	3972.45	[3714.91, 4559.66]	0.73
$T_3(y,\theta)$ (max)	526816	[607186, 2239365]	0.99
$T_4(y,\theta)$ (std)	22300.18	[20012.59, 39808.44]	0.88

Residual Analysis

- For a model with unknown parameters θ and predictors x_i, the *predicted* value is $\mathrm{E}(y_i|x_i,\theta)$ and the *residual* is $r_i = y_i - \mathrm{E}(y_i|x_i,\theta)$.

Residual Analysis

- For a model with unknown parameters θ and predictors x_i, the *predicted* value is $\mathrm{E}(y_i|x_i,\theta)$ and the *residual* is $r_i = y_i - \mathrm{E}(y_i|x_i,\theta)$.
- The *standardized residual* is given by $r_i/\mathrm{std}(y)$.

Residual Analysis

- For a model with unknown parameters θ and predictors x_i, the *predicted* value is $\mathrm{E}(y_i|x_i, \theta)$ and the *residual* is $r_i = y_i - \mathrm{E}(y_i|x_i, \theta)$.
- The *standardized residual* is given by $r_i/\mathrm{std}(y)$.
- Using the simulated posterior density, we can compute $\mathrm{E}(y_i|x_i, \theta)$ to be the mean of the replicated hold-out data itself.

Residual Analysis - Actual versus Replicated

Residual Analysis - Test Statistic Evaluation

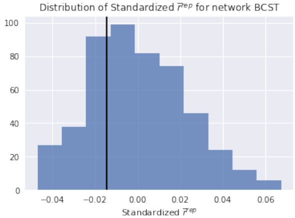
We can measure the residual mis-fit through the following test statistic:

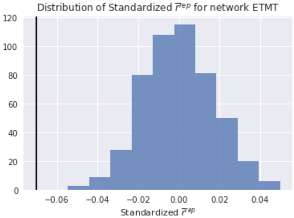
$$T(y, \theta, x) = \frac{\bar{r}}{\text{std}(y)}.$$

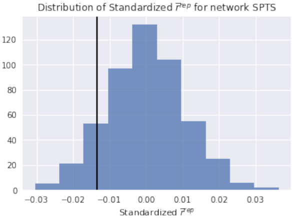
Residual Analysis - Test Statistic Evaluation

We can measure the residual mis-fit through the following test statistic:

$$T(y, \theta, x) = \frac{\bar{r}}{\text{std}(y)}.$$







Industry Standard Model

- The industry standard model used today to forecast impressions is based on the average of observed data.

Industry Standard Model

- The industry standard model used today to forecast impressions is based on the average of observed data.
- Define model $\mathcal{M}_0$ to be as follows:
  1. Using the content and stratified hour, take the average impression concentration c_i of the train set.
  2. Using the stratified hour, take the average impression concentration c_i' of the train set.
  3. The forecasted m_i^A for airing i is then $c_i m_i^B$ if it exists, otherwise it is $c_i' m_i^B$.

Industry Standard Model

- The industry standard model used today to forecast impressions is based on the average of observed data.
- Define model $\mathcal{M}_0$ to be as follows:
    1. Using the content and stratified hour, take the average impression concentration c_i of the train set.
    2. Using the stratified hour, take the average impression concentration c_i' of the train set.
    3. The forecasted m_i^A for airing i is then $c_i m_i^B$ if it exists, otherwise it is $c_i' m_i^B$.
- We wish to compare the performance of model $\mathcal{M}$ to model $\mathcal{M}_0$ and determine if the proposed model outperforms the industry standard model.

Analysis - Error Metrics

- The main error metric we will use to evaluate model performance is the Mean Absolute Error. It is defined as

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i^{\text{act}} - y_i^{\text{pred}}|$$

Analysis - Error Metrics

- The main error metric we will use to evaluate model performance is the Mean Absolute Error. It is defined as

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i^{\text{act}} - y_i^{\text{pred}}|$$

- Additionally, we can use an analogous metric to evaluate the probabilistic forecasts called the Continuous Ranked Probability Score (CRPS).

Analysis - Error Metrics

- The main error metric we will use to evaluate model performance is the Mean Absolute Error. It is defined as

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i^{\text{act}} - y_i^{\text{pred}}|$$

- Additionally, we can use an analogous metric to evaluate the probabilistic forecasts called the Continuous Ranked Probability Score (CRPS).
- Let F be the cumulative distribution function of a random variable X and x be the observed value. Then we have that:

$$\text{CRPS}(F, x) = \int_{-\infty}^{\infty} (F(y) - H(y - x))^2 \, dy \tag{5}$$

Note that this is a generalization of the Mean Absolute Error.

Analysis - Calibration

- We are also interested in the *calibaration* of the forecasted probability distributions.

Analysis - Calibration

- We are also interested in the *calibaration* of the forecasted probability distributions.
- We can assess the calibration of the probabilistic forecasts by calculating Credible Regions (CRs) for each unit of observation and determining the proportion of units that fall within the CRs compared to the number of units.

Analysis - Calibration

- We are also interested in the *calibaration* of the forecasted probability distributions.
- We can assess the calibration of the probabilistic forecasts by calculating Credible Regions (CRs) for each unit of observation and determining the proportion of units that fall within the CRs compared to the number of units.
- The forecasts are perfectly calibrated if for a CR of $(1-\alpha)\%$, the proportion of units within the CR is $(1-\alpha)\%$.

Analysis - Units of Observation

- The table below shows the evaluation of error metrics between the models:

network	$\overline{m_i^A}$	$\mathcal{M}_0$ MAE	$\mathcal{M}$ CRPS	$\mathcal{M}$ MAE
BCST	171450.06	23307.81	15247.91	21408.24
ETMT	13034.77	5123.70	3683.04	5226.42
SPTS	3164.84	1844.51	1426.60	1942.76

- The point-forecasts of Model $\mathcal{M}$ perform on-par with, or slightly worse than $\mathcal{M}_0$.

Analysis - Units of Observation

- The table below shows the evaluation of error metrics between the models:

network	$\overline{m_i^A}$	$\mathcal{M}_0$ MAE	$\mathcal{M}$ CRPS	$\mathcal{M}$ MAE
BCST	171450.06	23307.81	15247.91	21408.24
ETMT	13034.77	5123.70	3683.04	5226.42
SPTS	3164.84	1844.51	1426.60	1942.76

- The point-forecasts of Model $\mathcal{M}$ perform on-par with, or slightly worse than $\mathcal{M}_0$.
- However, the probabilistic forecasts of model $\mathcal{M}$ greatly outperform model $\mathcal{M}_0$.

Analysis - Units of Observation

The table below shows the calibration of the probabilistic forecasts at the level of the units of observation.

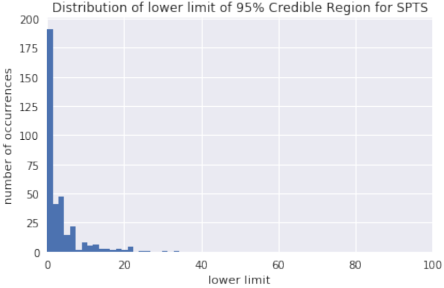
	Credible Region $(1-\alpha)\%$		
network	$\alpha = 0.5$	$\alpha = 0.05$	$\alpha = 0.01$
BCST	0.465	0.922	0.966
ETMT	0.48	0.91	0.95
SPTS	0.416	0.765	0.799

Analysis - Units of Observation - SPTS

- Closer analysis reveals that the right-censored data is causing the miscalibration on the SPTS network.

Analysis - Units of Observation - SPTS

- Closer analysis reveals that the right-censored data is causing the miscalibration on the SPTS network.
- The figure below shows the distribution of lower-limits of 95% CRs in which the outcome was outside the above specified CRs:



- Removing the right-censored data and recalculating the 95% CRs shows that 91.8% of units are within the forecasted CR, consistent with other networks.

Quantiled Media Plans

- We are interested in understanding the performance of aggregations of units of observation since these aggregations form the media plans.

Quantiled Media Plans

- We are interested in understanding the performance of aggregations of units of observation since these aggregations form the media plans.
- To do this, we quantile the units in the hold-out set into 20 bins ordered by the point forecasts of the impression concentration as a proxy for creating media plans. We then compute the total forecasted impressions of the media plan using each model.

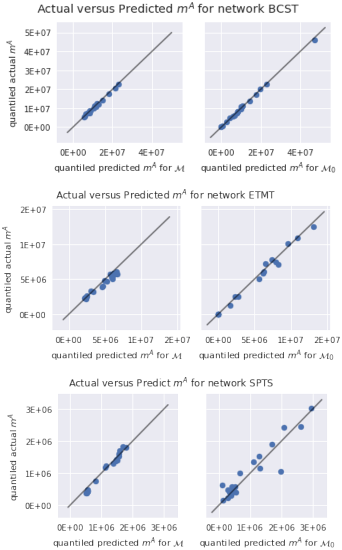
Quantiled Media Plans

- We are interested in understanding the performance of aggregations of units of observation since these aggregations form the media plans.
- To do this, we quantile the units in the hold-out set into 20 bins ordered by the point forecasts of the impression concentration as a proxy for creating media plans. We then compute the total forecasted impressions of the media plan using each model.
- For model $\mathcal{M}_0$, the total forecasted impressions are the sum of the individual point forecasts.

Quantiled Media Plans

- We are interested in understanding the performance of aggregations of units of observation since these aggregations form the media plans.
- To do this, we quantile the units in the hold-out set into 20 bins ordered by the point forecasts of the impression concentration as a proxy for creating media plans. We then compute the total forecasted impressions of the media plan using each model.
- For model $\mathcal{M}_0$, the total forecasted impressions are the sum of the individual point forecasts.
- For model $\mathcal{M}$, under the model assumptions, the units are i.i.d. given the covariates and parameters. We can create the distribution of aggregated impressions through simple sums of the distributions of outcomes at the level of the units of observation.

Quantiled Media Plans



Quantiled Media Plans - Error Metrics

- The table below shows the error metrics between the two models for the quantiled media plans.

network	$\mathcal{M}_0$ MAE	$\mathcal{M}$ CRPS	$\mathcal{M}$ MAE
BCST	324034.14	277437.29	417797.60
ETMT	297512.52	311316.38	399648.10
SPTS	205096.15	75962.55	90668.35

Quantiled Media Plans - Error Metrics

- The table below shows the error metrics between the two models for the quantiled media plans.

network	$\mathcal{M}_0$ MAE	$\mathcal{M}$ CRPS	$\mathcal{M}$ MAE
BCST	324034.14	277437.29	417797.60
ETMT	297512.52	311316.38	399648.10
SPTS	205096.15	75962.55	90668.35

- The point-forecasts of model $\mathcal{M}$ perform worse on BCST and ETMT, but much better on SPTS compared to $\mathcal{M}_0$.

Quantiled Media Plans - Error Metrics

- The table below shows the error metrics between the two models for the quantiled media plans.

network	$\mathcal{M}_0$ MAE	$\mathcal{M}$ CRPS	$\mathcal{M}$ MAE
BCST	324034.14	277437.29	417797.60
ETMT	297512.52	311316.38	399648.10
SPTS	205096.15	75962.55	90668.35

- The point-forecasts of model $\mathcal{M}$ perform worse on BCST and ETMT, but much better on SPTS compared to $\mathcal{M}_0$.
- The probabilistic forecasts perform on-par or better with the industry standard model.

Quantiled Media Plans - Calibration

- The table below shows the calibration of the probabilistic forecasts at the media-plan-level.

	Credible Region $(1-\alpha)\%$		
network	$\alpha = 0.5$	$\alpha = 0.05$	$\alpha = 0.01$
BCST	0.45	0.95	1.0
ETMT	0.1	0.55	0.75
SPTS	0.4	0.65	0.65

Quantiled Media Plans - Calibration

- The table below shows the calibration of the probabilistic forecasts at the media-plan-level.

	Credible Region $(1-\alpha)\%$		
network	$\alpha = 0.5$	$\alpha = 0.05$	$\alpha = 0.01$
BCST	0.45	0.95	1.0
ETMT	0.1	0.55	0.75
SPTS	0.4	0.65	0.65

- From this table we can see that the model is well-calibrated on the BCST network, but not on the others.

Quantiled Media Plans - Calibration

- The table below shows the calibration of the probabilistic forecasts at the media-plan-level.

	Credible Region $(1-\alpha)\%$		
network	$\alpha = 0.5$	$\alpha = 0.05$	$\alpha = 0.01$
BCST	0.45	0.95	1.0
ETMT	0.1	0.55	0.75
SPTS	0.4	0.65	0.65

- From this table we can see that the model is well-calibrated on the BCST network, but not on the others.
- More analysis is needed to understand which assumptions are violated when aggregating the units of observation on the ETMT and SPTS networks.

Conclusions

- We aimed to create a model that outperformed the Industry Standard model and also provided probabilistic forecasts.

Conclusions

- We aimed to create a model that outperformed the Industry Standard model and also provided probabilistic forecasts.
- The proposed model's point forecasts are on-par with the Industry Standard model at the level of unit of observation and either better or worse for aggregations.

Conclusions

- We aimed to create a model that outperformed the Industry Standard model and also provided probabilistic forecasts.
- The proposed model's point forecasts are on-par with the Industry Standard model at the level of unit of observation and either better or worse for aggregations.
- The proposed model is well-calibrated at the level of units of observation, but not for aggregations of units.

Conclusions

- The data generated under the proposed model for the ETMT network has obvious misfit and this translates into the forecasts.

Conclusions

- The data generated under the proposed model for the ETMT network has obvious misfit and this translates into the forecasts.
- The proposed model greatly outperforms the Industry Standard model on the SPTS network in terms of aggregations of units.

Conclusions

- The data generated under the proposed model for the ETMT network has obvious misfit and this translates into the forecasts.
- The proposed model greatly outperforms the Industry Standard model on the SPTS network in terms of aggregations of units.
- More work is needed to iterate on model and address the misfit and the loss of calibration for aggregations.

Conclusions

- The data generated under the proposed model for the ETMT network has obvious misfit and this translates into the forecasts.
- The proposed model greatly outperforms the Industry Standard model on the SPTS network in terms of aggregations of units.
- More work is needed to iterate on model and address the misfit and the loss of calibration for aggregations.
- The probabilistic forecasts generated are well-calibrated at the unit of observation level and thus can be leveraged to provide media planners with insight as to what the volatility of the media plan might be.