

Predictive Analytics Methods Applied to Insurance

Elucidor, LLC

Forsides

Paris, November 5th 2015

Marc Raymond, IA; Forsides
marc.raymond@forsides.fr
+33-6-42-18-94-08

Néfissa Sator, IA, CERA, MAAA; Forsides
nefissa.sator@forsides.com
+1-646-934-2061

Howard Zail, FSA, MAAA; Elucidor LLC
hzail@elucidor.com
+1-212-532-6434

Introduction

Néfissa Sator
Forsides, SVP USA

Context

- Meeting in NYC in 2015
- Collaboration on the presentation of Predictive Analytics techniques to the Actuarial Association of Greater NY (ASNY)
 - Applied to Long Term Care (longevity and Interest rate risks)
 - For the ORSA framework

Introduction to Machine Learning techniques to improve GLM capabilities

- Experience feedback with an auto-insurance portfolio

Introduction to Predictive Analytics techniques applied to Insurance to improve

- Pricing, Underwriting, Reinsurance program, Capital and risk management

Introduction to Model Validation techniques to comfort new modelling methods

Practical implementation and concrete examples

- Group life pricing and Capital management
- Mortality improvement

Machine Learning Techniques

Marc Raymond

Forsides, Deputy Managing Director

Un modèle bien connu : le GLM

- On suppose disposer d'un échantillon (X_i, Y_i) , où les variables X_i sont des informations exogènes sur l'assuré ou sur le bien assuré, et où Y_i est la variable d'intérêt, qui sera
 - une variable de comptage, à valeurs dans N , par exemple le nombre d'accidents de l'assuré i l'an passé
 - une variable positive, à valeurs dans R^+ , par exemple le coût du sinistre i , ou bien la durée entre la survenance et la déclaration du sinistre, ou encore les règlements à venir dans un triangle de paiement.

➔ Objectif : estimer $E[Y|X]$ sous la forme $g(E[Y|X]) = a_0 \times a_1 \times a_2 \times \dots \times a_n$

Avantages

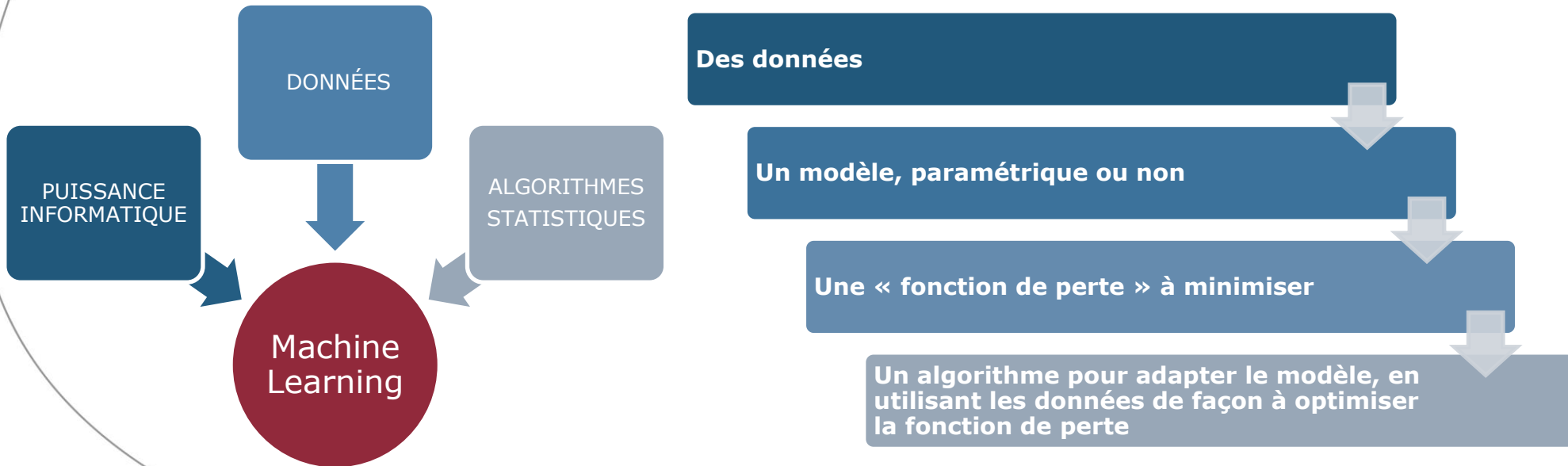
- Performant pour décrire l'influence d'un nombre réduit de caractéristiques sur une variable à expliquer
- Modèle le plus connu et utilisé par tous : facilité de communication et de compréhension
- Traçabilité

Inconvénients

- Modèle limité lorsque le nombre de variables devient important
- Hypothèses fortes sur la distribution des données
- Inadéquation possible du modèle avec la variable à expliquer
- Capacité limitée de croisement de variables

Les algorithmes de machine learning

- L'augmentation de la puissance de calcul a permis le développement de ces nouvelles méthodes d'analyse de données dites de Machine Learning, utilisant une approche algorithmique et issues le plus souvent de travaux d'informaticiens que de statisticiens. Ces méthodes permettent de capter les relations non linéaires dans les données et sont robustes lorsque le nombre de variables augmente



Apprentissage supervisé : prédiction et classement

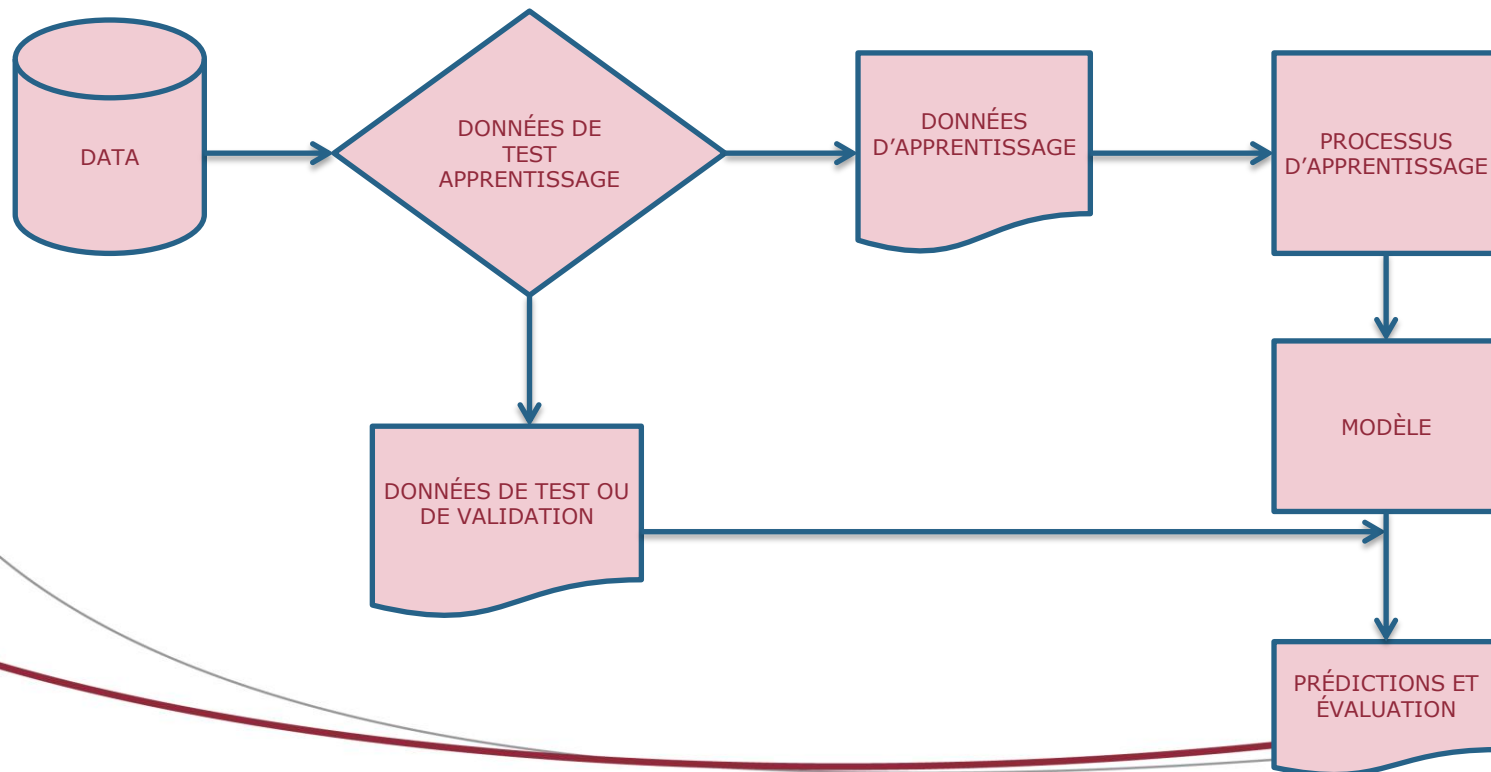
Modèle	Tâche
Arbres - CART	Classement + prédiction
Réseaux de neurones	Classement + prédiction
SVM	Classement + prédiction
K-plus proches voisins	Classement
Bagging et Boosting	Classement + prédiction

Apprentissage non supervisé : classification

Modèles	Tâches
Règles d'association	Classification
K-means	Classification
Réseaux de Kohonen	Classification

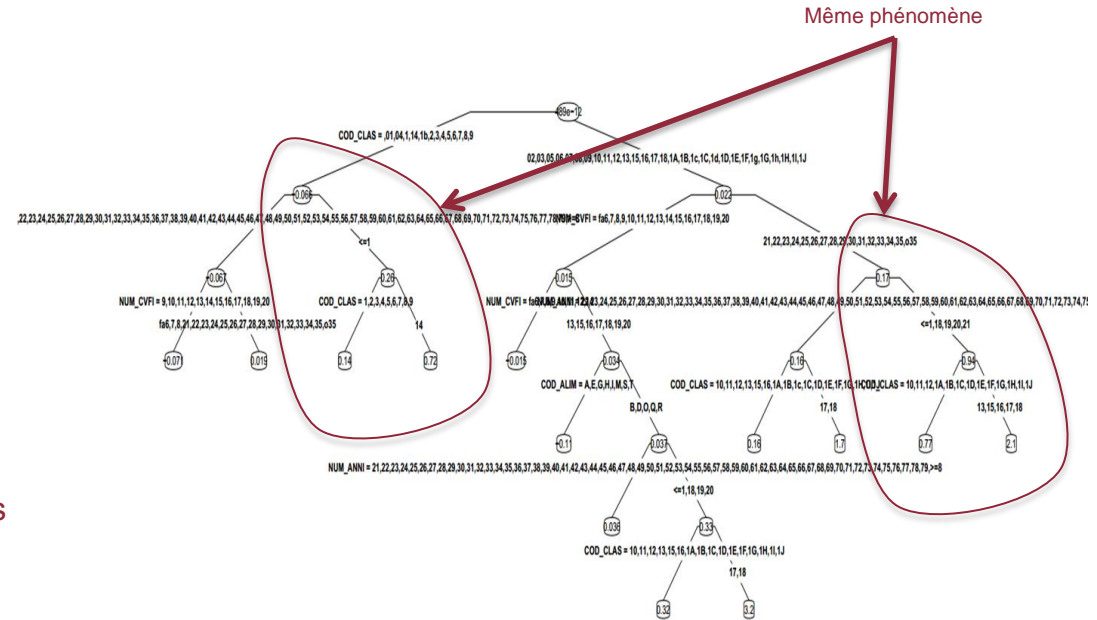
Pour atteindre l'objectif visé dans un projet de data science, il s'agit d'être capable de mesurer la qualité du modèle lors de la phase d'apprentissage (calibrage du modèle) et s'assurer également que le modèle fonctionnera aussi bien dans l'environnement de production que sur les données d'apprentissage

- Evaluation du modèle
- Validation du modèle



De manière générale, on minimise une certaine fonction de perte (de mesure de l'homogénéité ou d'impureté des régions représentées par les nœuds fils obtenus à chaque division)

- Ainsi, en régression (si la variable y à prédire est quantitative), on cherche de façon naturelle à minimiser la variance des nœuds fils (i.e. la variance intra-groupe). Un nœud sera pur si sa variance est nulle (toutes les observations sont « égales ») et impure si la variance des observations est élevée. Plus précisément, pour un arbre de régression, on minimise l'impureté des nœuds fils $I(j, s)$ calculée précédemment,

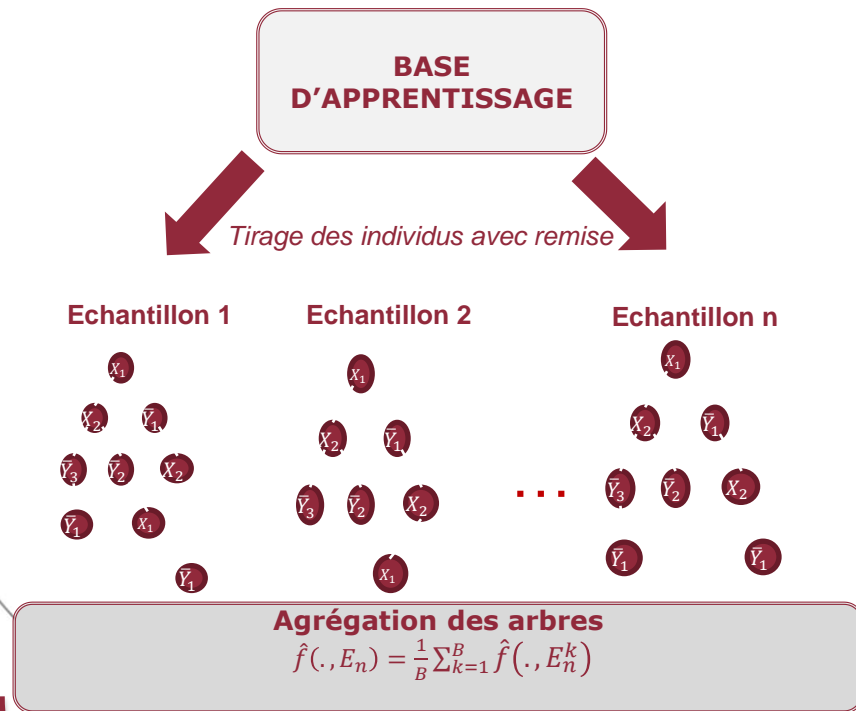


Points de difficultés

- Un seul arbre n'est pas robuste
- Un seul arbre est instable.
- Les mêmes phénomènes peuvent se trouver à plusieurs niveaux de l'arbre
- Les résultats sont très liés au seuil de « pureté » retenu

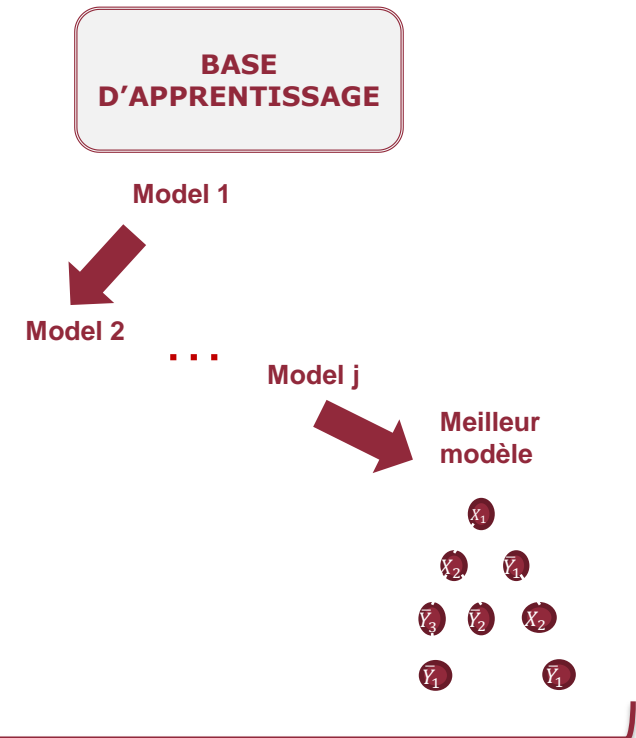
Random Forest

- Répétition d'arbres de décisions avec :
 - sélection aléatoire de variables
 - bootstrap sur un échantillon d'individus



Gradient Boosting Model

Le principe de base est de construire une séquence de modèles de sorte que chaque étape, chaque modèle ajouté à la combinaison, apparaisse comme un pas vers une meilleure solution.



Un bon pouvoir prédictif mais une grande difficulté à interpréter les résultats .

Retour d'expérience

Un cas concret en assurance automobile

Marc Raymond

Forsides, Deputy Managing Director

De la tarification classique ...

Tarification avec un modèle Linéaire Généralisé

- régression
- récupération des coefficients
- calcul des résidus : écart entre la prédiction et la charge sinistre
- tarification



... à la tarification avancée

Pour faciliter la mise en œuvre, il a été décidé de conserver la tarification réalisée avec le GLM mais....

....de l'enrichir avec l'utilisation d'algorithmes de machine learning :

- amélioration de cette tarification (prise en compte de l'ensemble des variables et amélioration du pouvoir prédictif)
- intégration de nouveaux critères dans le modèle linéaire généralisé

De la tarification classique ...

Tarification avec un modèle Linéaire Généralisé

- Régression linéaire
- récupération des coefficients
- calcul des résidus : écart entre la prédiction et la charge sinistre
- tarification

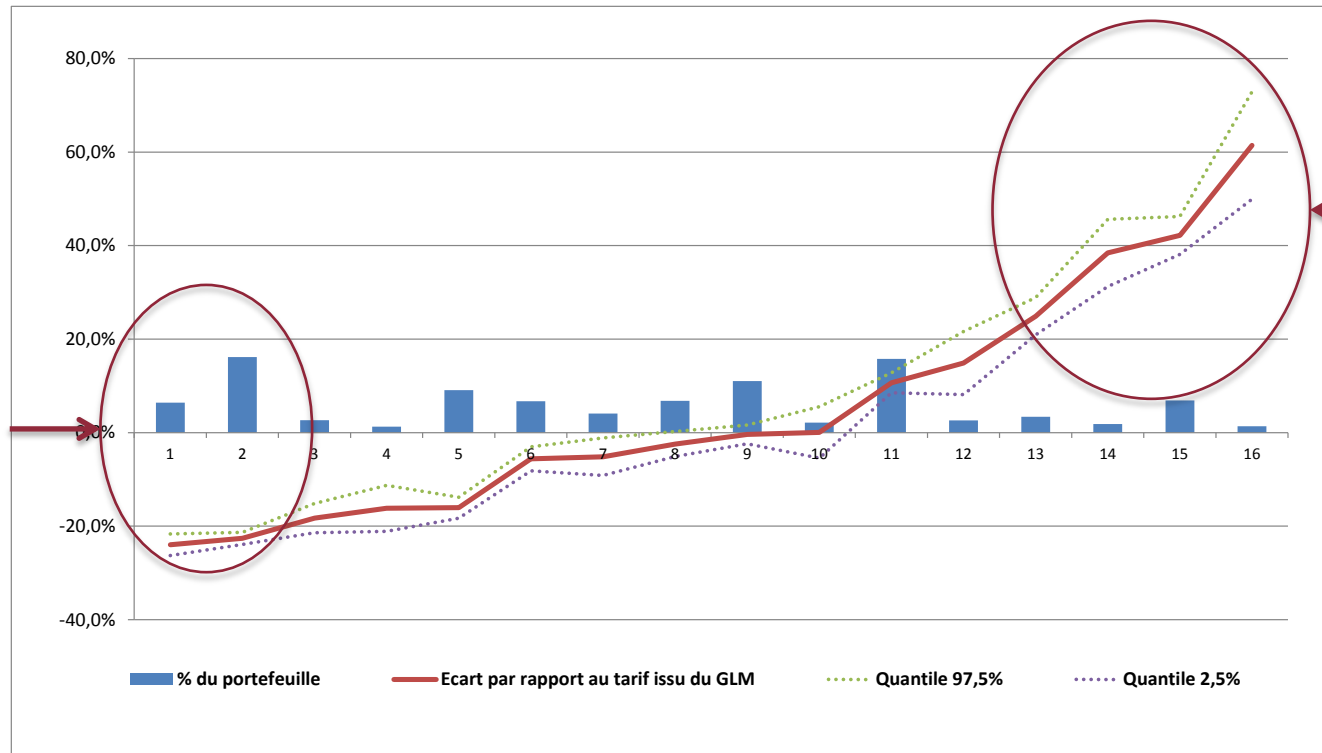
... à la tarification avancée

Les étapes :

- Etape 0 : Etude de la significativité des variables et de leur pouvoir prédictif
- Etape 1 : CART Optimisé
Première approche de classification de type CART et choix des paramètres optimaux
- Etape 2 : Random Forest décrypté
- Etape 3 : Identification et explication des segments les plus prédictifs des résidus
- Etape 4 : Intégration de nouveaux critères retenus dans le GLM

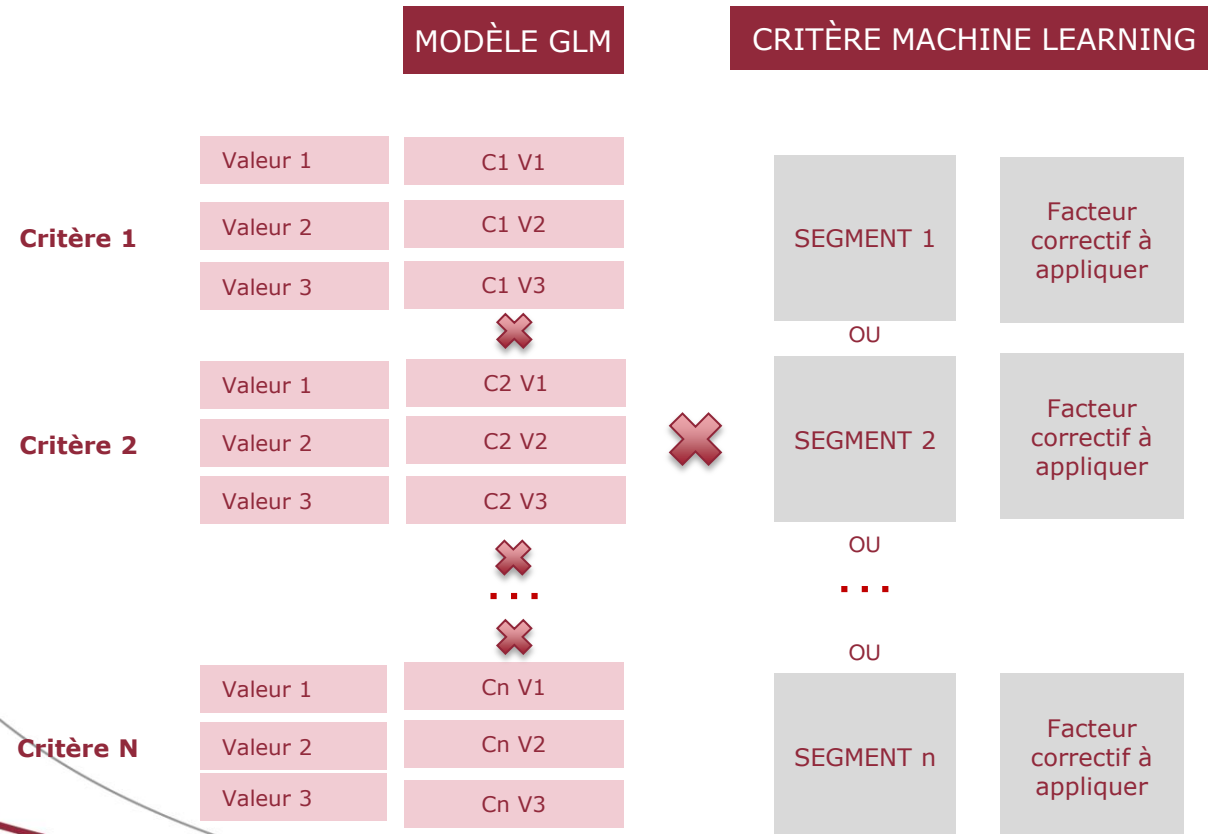
La réintégration des segments identifiés par le « machine learning » dans le GLM donne les résultats suivants :

2 segments représentant 22,5 % des clients et pour lesquels le GLM sous-estime le coût du risque en moyenne de 23 %



4 segments représentant 13,5 % des clients et pour lesquels le GLM sous-estime le coût du risque en moyenne de 39 %

Pour les segments les plus pertinents intégration dans la nouvelle tarification



Exemples de nouveaux segments

Age > 55 ans
 Véhicule > 8 ans
 Véhicule peu puissant
 BM < 0,6

Age 30-50 ans
 Berline
 Véhicule Diesel
 Zone rurale

Age < 35 ans
 Pas de sinistre sur les 36 derniers mois
 Véhicule haut de gamme

Les enjeux d'une démarche « Big data »

- Un secteur en pleine mutation sous les effets couplés d'une massification des données et des innovations technologiques (capacité de calcul et de stockage, algorithmes prédictifs,...) - *“There was **more data produced in 2011 alone than in all of the rest of human history combined back to the invention of the alphabet**”* (Robert Kirkpatrick, Director of UN Global Pulse)
- De nombreuses questions se posent :
 - Un modèle de rupture à concevoir ?
 - Une redéfinition des business model actuel ?
 - Le risque d'entrée de nouveaux acteurs ?
- Mais une réalité déjà perceptible :
 - Des réalisations concrètes encore limitées mais un avantage concurrentiel important aux premiers acteurs intégrant ces nouvelles technologies

Retour d'expérience - les points clés d'une démarche réussie

- Traiter les données pour avoir une base de travail sans biais et pertinente
- Posséder une bonne compréhension de ce qu'on fait et des outils utilisés
- Éviter le côté « boîte noire » inhérent aux algorithmes de machine learning et concevoir une approche pragmatique et intelligible par les interlocuteurs : lien métier / technicité
- Ne pas « renier » le passé mais l'améliorer : arbitrage coût / efficacité du modèle

Predictive Analytics Techniques

Howard Zail

Elucidor LLC, Partner

Agenda

- Motivation
- Example 1: Group Life Pricing and Risk Management
 - Bayesian approach
 - Probabilistic Programming
 - VAR / Risk adjusted returns
- Example 2: State Space Modeling
- Identifying and Selecting Risk Drivers
- Implementation

I: Motivation

The Value to Insurers of the Bayesian Approach to Predictive Analytics

Current State of Insurers using Bayesian Predictive Analytics



Most (if not all) can significantly benefit from adopting Bayesian methods

Few insurers use the methods

What We Will Show

- The Bayesian paradigm provides a natural way for analyzing both ***expectation*** and ***volatility*** (risk)
- Incorporates both ***non-systematic*** (noise) and ***systematic*** (model or parameter) risk automatically
- Highly flexible and capable of managing significant ***complexity***
- Probabilistic programming (e.g Stan, JAGS, home-grown MCMC) provides a means for ***practical implementation*** of a wide range of analyses
- Incorporate ***professional opinion*** in a cohesive manner

II: Example

Group Life Insurance Pricing and Capital Management

The Challenge

- A group life insurer has the following experience for the prior year:
 - 10,000 (n_1) covered insured in prior year
 - 10 (x_1) claims with fixed \$5,000 per claim (.1% rate)
 - Assume age, sex profile remains fixed
- For the coming year, insurer expects to cover 20,000 (n_2) insureds. Relying solely on the data:
 - How many claims do we anticipate (x_2)?
 - What premium should we charge?
 - How much capital should we hold?

A Bayesian Approach

- Step 0: Start with Bayesian Formula:

$$\Pr(\theta | x_1) = \frac{\Pr(x_1 | \theta) \cdot \Pr(\theta)}{\Pr(x_1)}$$

$$\propto \{\Pr(x_1 | \theta)\} \cdot \{\Pr(\theta)\}$$

$$\text{Posterior} \propto \{\text{likelihood}\} \cdot \{\text{prior}\}$$

- Step 1: Define non-informative **prior for θ** , the risk of a claim

$$\theta \sim \text{Uniform}(0,1)$$

- Step 2: Define **likelihood**

$$x_1 \sim \text{Binomial}(n = 10000, p = \theta)$$

- Step 3: **Simulate** posterior θ_i for $i = 1, \dots, 4000$
- Step 4: Use **predictive distribution** $\Pr(x_2 | \theta, x_1)$ to calculate
 - VAR, Premium, Capital

Probabilistic Programming: Stan

```
model{
  theta ~ uniform(0,1);      // this defines the prior
  x1 ~ binomial(n1, theta);  // this defines the likelihood
}
generated quantities {
  int x2; // predicted number of claims in new year
  x2 <- binomial_rng(n2,theta); //generates random samples
}
```

Output Generated: 4000 simulated **pairs** of:

- theta
- x2

Solve for Premiums, Capital

- Target:

- Economic Capital at 99.5% VAR level
- Return on Equity: 10%

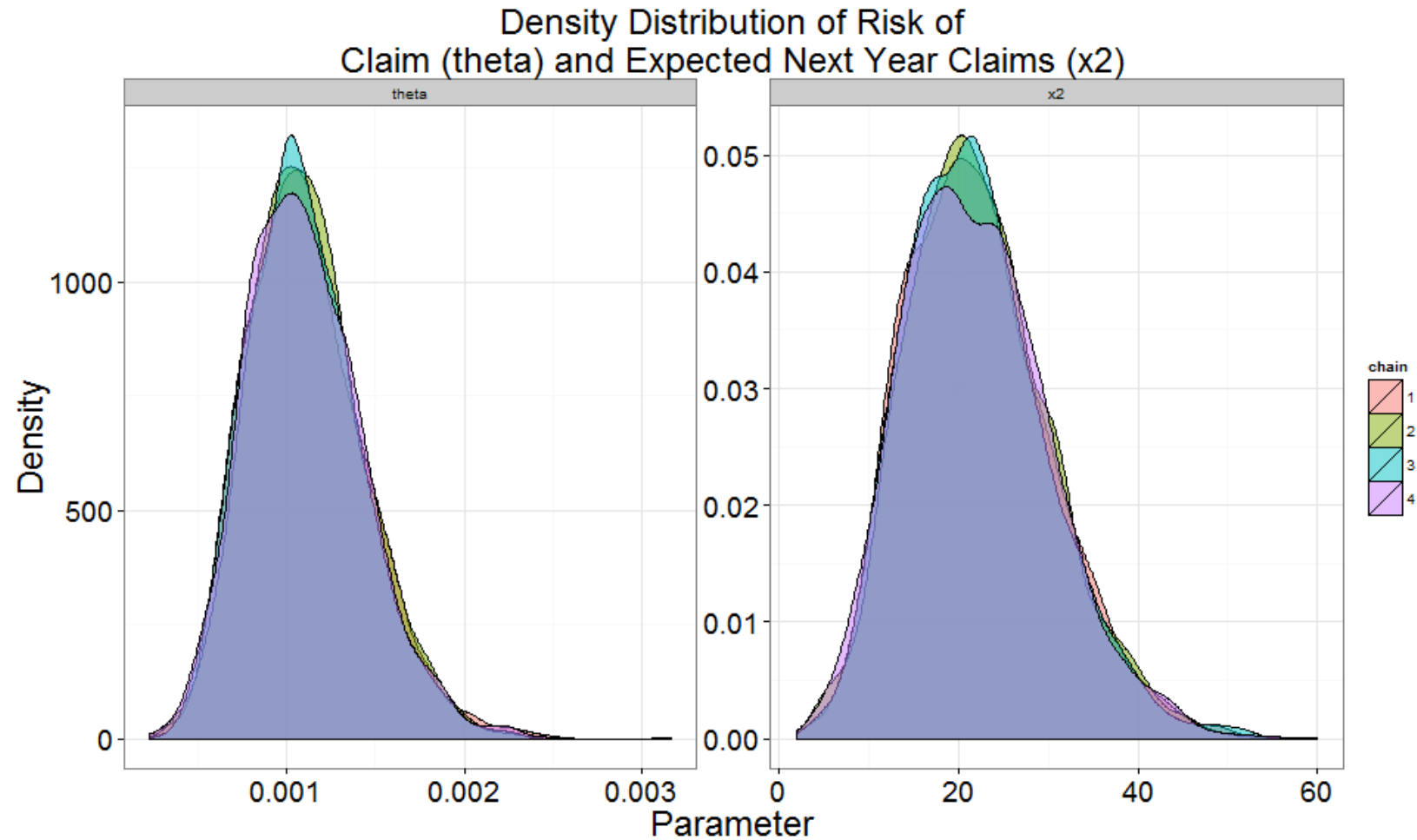
- Solution:

1. $0.995 = \Pr(5,000 \cdot x_2 \leq \text{Claims}_{0.995})$ **VAR Criterion**

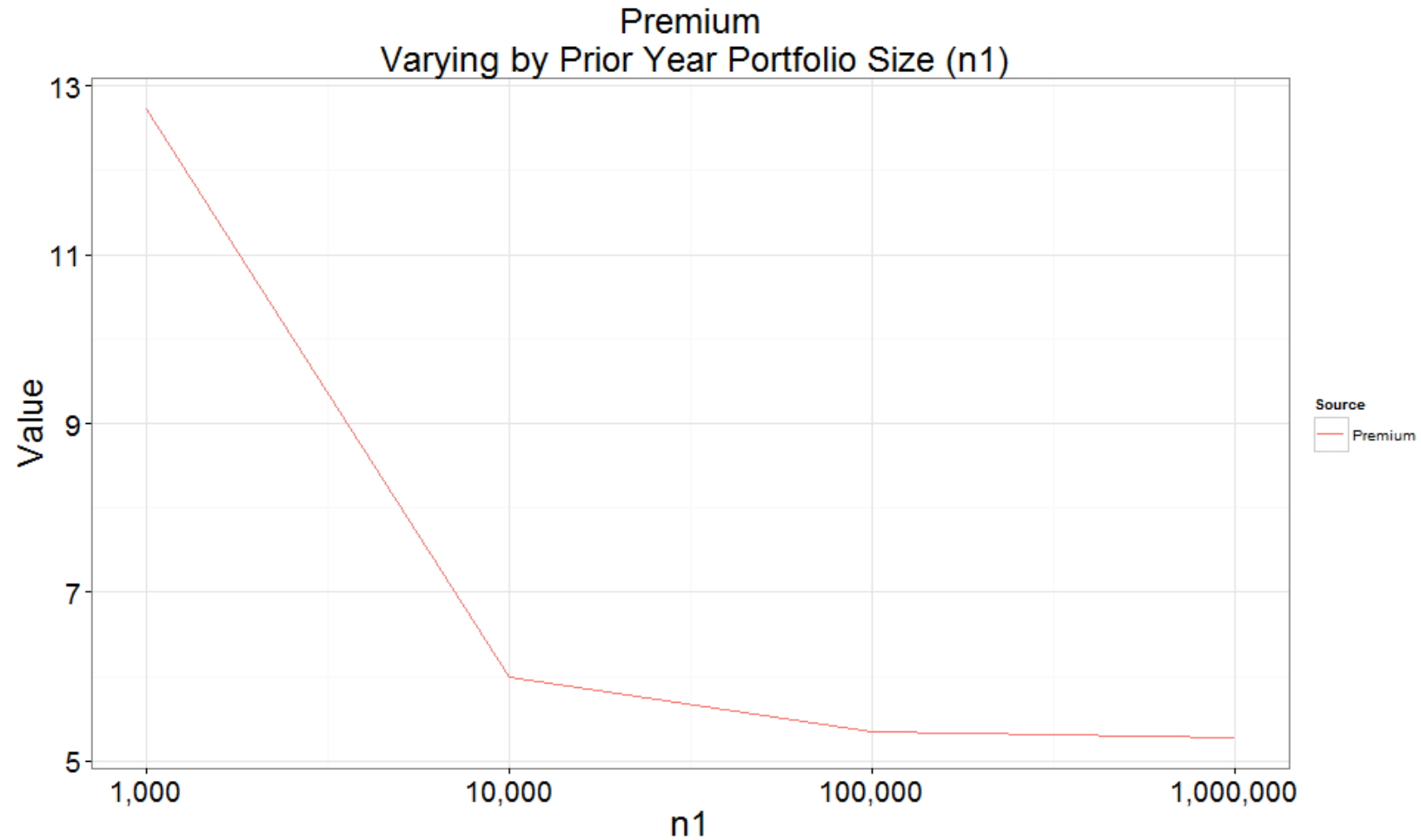
2. $\text{Capital} = \text{Claims}_{0.995} - 20,000 \cdot \text{Premium}$

3. $\frac{20,000(\text{Premium}) - 5,000 \cdot E[\# \text{Claim}]}{\text{Capital}} = 0.1$ **ROE Criterion**

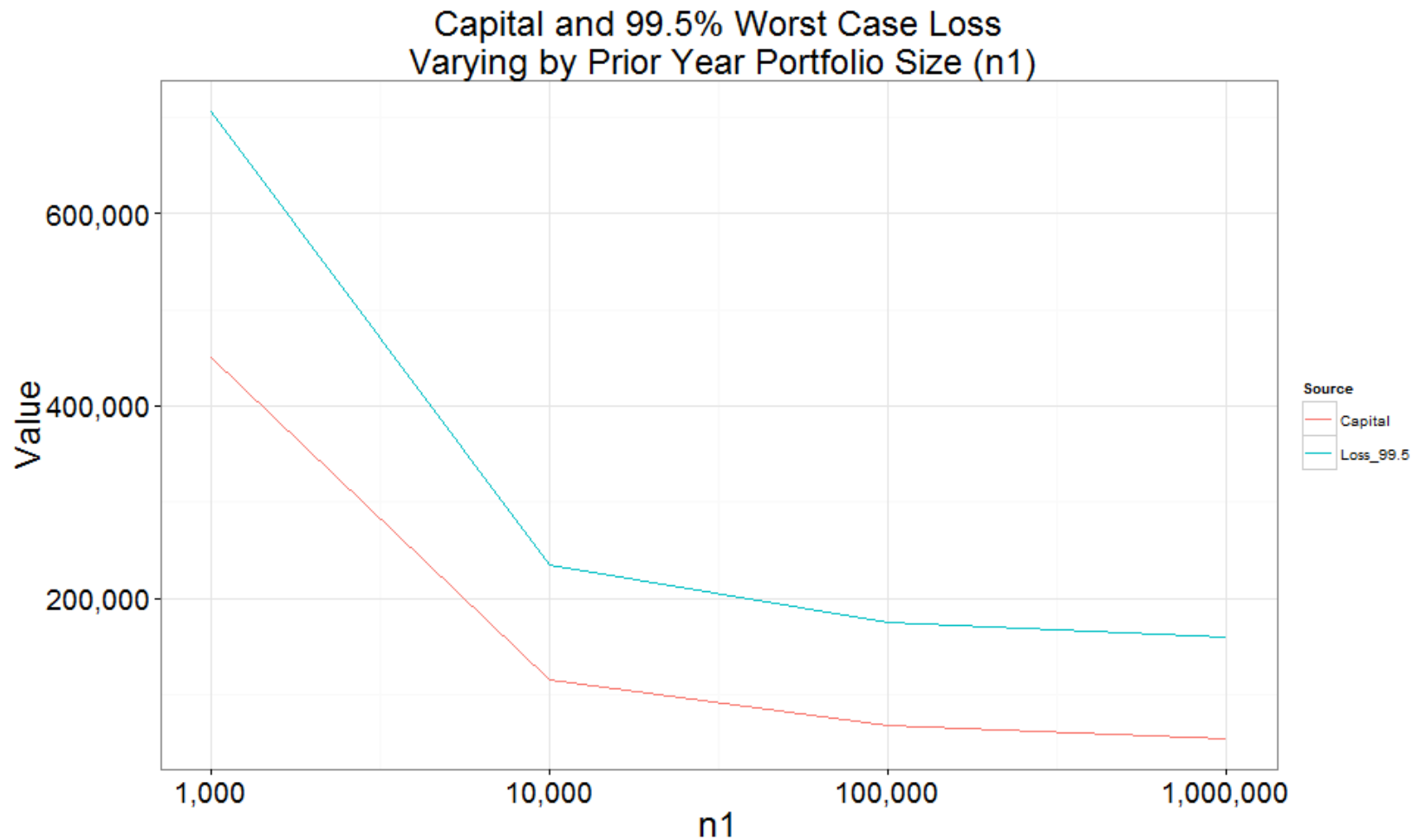
MCMC Results



Premium Sensitivity to Size of Prior Year Portfolio (n_1)



Capital and Loss Sensitivity to Size of Prior Year Portfolio (n_1)



A Word about the “Prior”

- Our example used an artificially simple prior:
 - $\theta \sim \text{Uniform}(0,1)$
- In practice we will have prior information about the claims rate, e.g.
 - $\theta \sim \text{Normal}(\mu = 0.0015, \sigma = 0.0004), \dots, 0 \leq \theta \leq 1$; or
 - $\theta \sim \text{Beta}(a = 15, b = 9985)$ (the mean of this is 0.0015)
- Prior is derived from a combination of prior analysis and professional opinion
- Good Priors (Good Management): Improve Predictions
- Bad Priors (Bad Management): Degrade Predictions

What Have We Achieved?

- Pricing regime that takes into account:
 - Systematic risk
 - Non-systematic risk
 - Professional judgment
 - All available data
- Capital management program that reflects true economic risk
 - Compare to standard formula under Solvency II
- Financial management that reflects our emerging experience

Probabilistic Programming

- Stan member of MCMC languages that include
 - BUGS, JAGS
- Uses Hamiltonian Monte Carlo to conduct inference
 - Can be substantially faster than Gibbs
 - Handles correlated and hierarchical parameters much more efficiently
 - Very active community of users
 - Contains range of advanced tools and functions
- Output is MCMC chains of parameters

Features of Stan

- Called from R, Python, Matlab, Julia, cmd
- Can do both MCMC and Optimization
 - Optimization benefits from internal differentiation engine
 - Works with restricted parameters (e.g. truncated parameters, simplexes)
- Open source, compiled, lots of documentation, worked examples, easy jump from BUGS to Stan, in-built parallelization, diagnostic tools

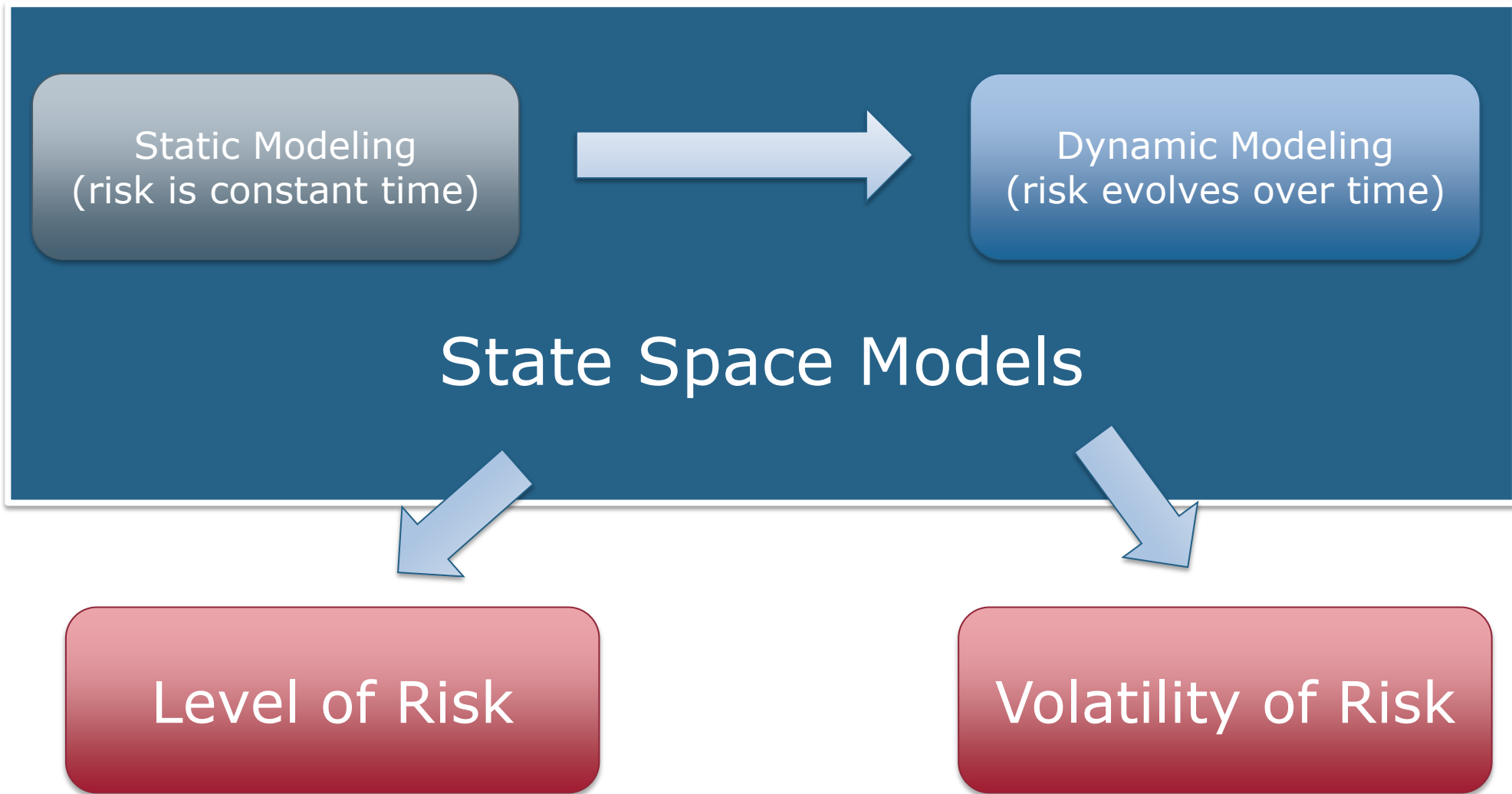
Limitations of Stan

- MCMC is still substantially slower than other predictive analytics that use optimization
- This is mitigated by:
 - Development of models can be a lot easier, faster
 - Stan’s “optimization” option can produce equivalent results to GLM, GAM, Lasso etc.
- Parameters must be continuous:
 - Regime change models cannot (easily) be done
- Not a great solution for massive datasets

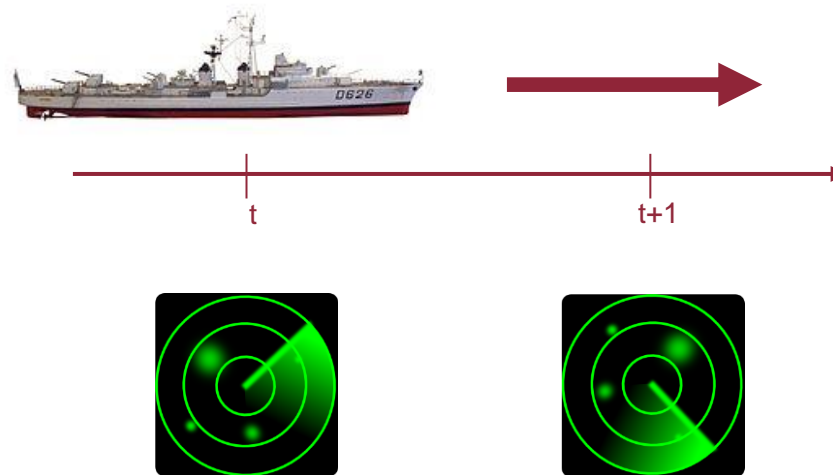
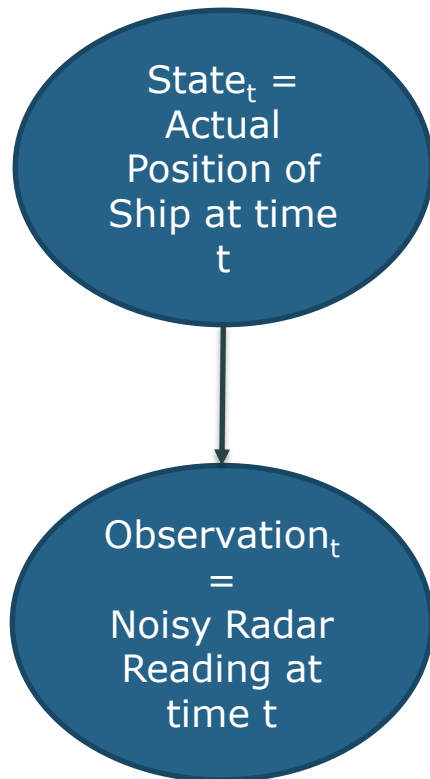
III: Example

State Space Modeling

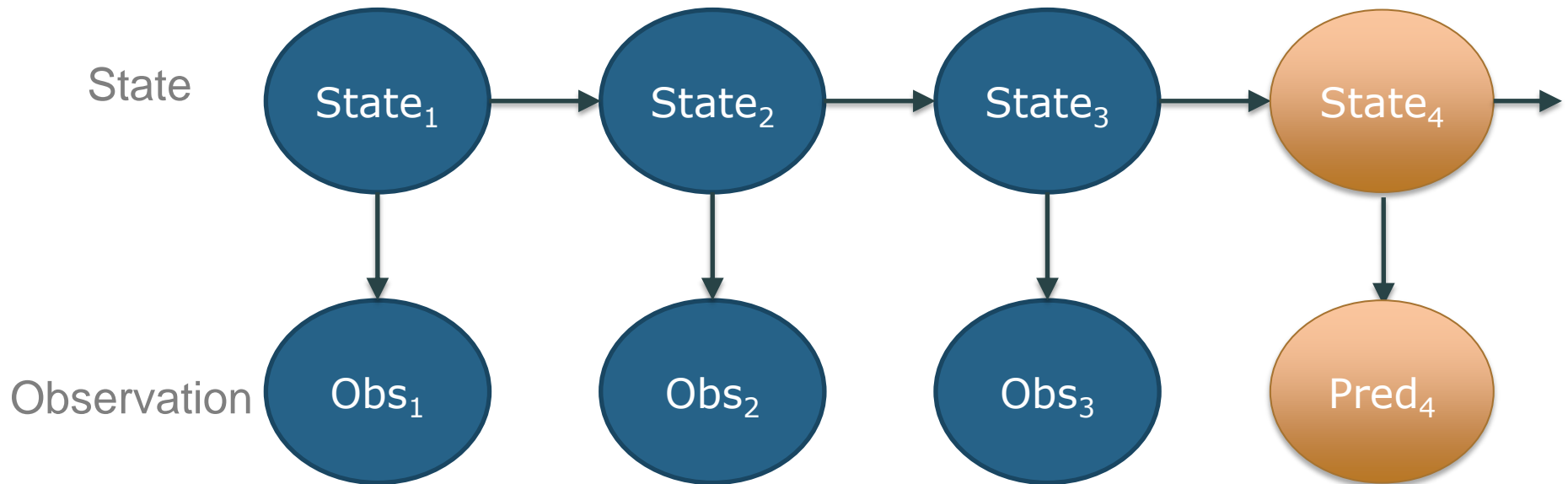
State Space Modeling



A Battle Ships Game Analogy



A Bayesian Network Representation



Solving the Model

- The techniques used to solve state space models are often not trivial
- But, there is extensive literature to help
- Our preferred method:
 - **Full Bayesian Model, solved using Markov chain Monte Carlo techniques**
- Advantages:
 - Provides a complete solution
 - Method is adaptable across many projects
- Disadvantages:
 - Requires lots of computing power (and potentially time)

A Longevity Risk Analysis Example

Model:

q_x = mortality rates from a given table

q_x^t = mortality rates at time t adjusted for improvement

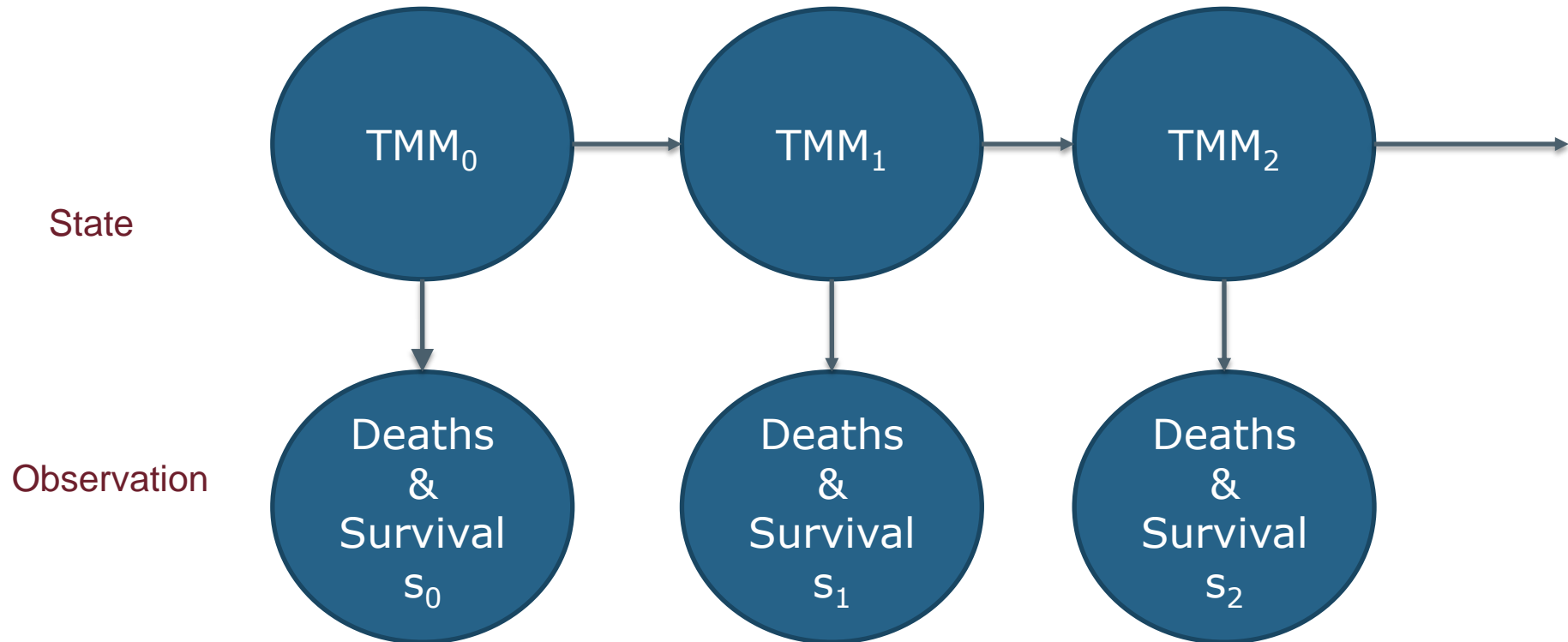
TMM_t = Total Mortality Multiple at time t

$q_x^t = TMM_t * q_x$

Questions:

- What are TMM_t for $t= 0$ to T ?
- What is the confidence interval round each TMM_t ?
- How are TMM_t expected to evolve in the future?

Longevity Risk State Space Representation



The Full Bayesian Longevity Model

State $\longrightarrow TMM_t = TMM_{t-1} + \beta + \eta_t$
Observation $\longrightarrow death(insured)_t \sim Bernoulli(TMM_t \cdot q_{[x]+t})$
 $\eta_t \sim Normal(mean = 0, sd = \sigma_\eta)$

Prior Assumptions:

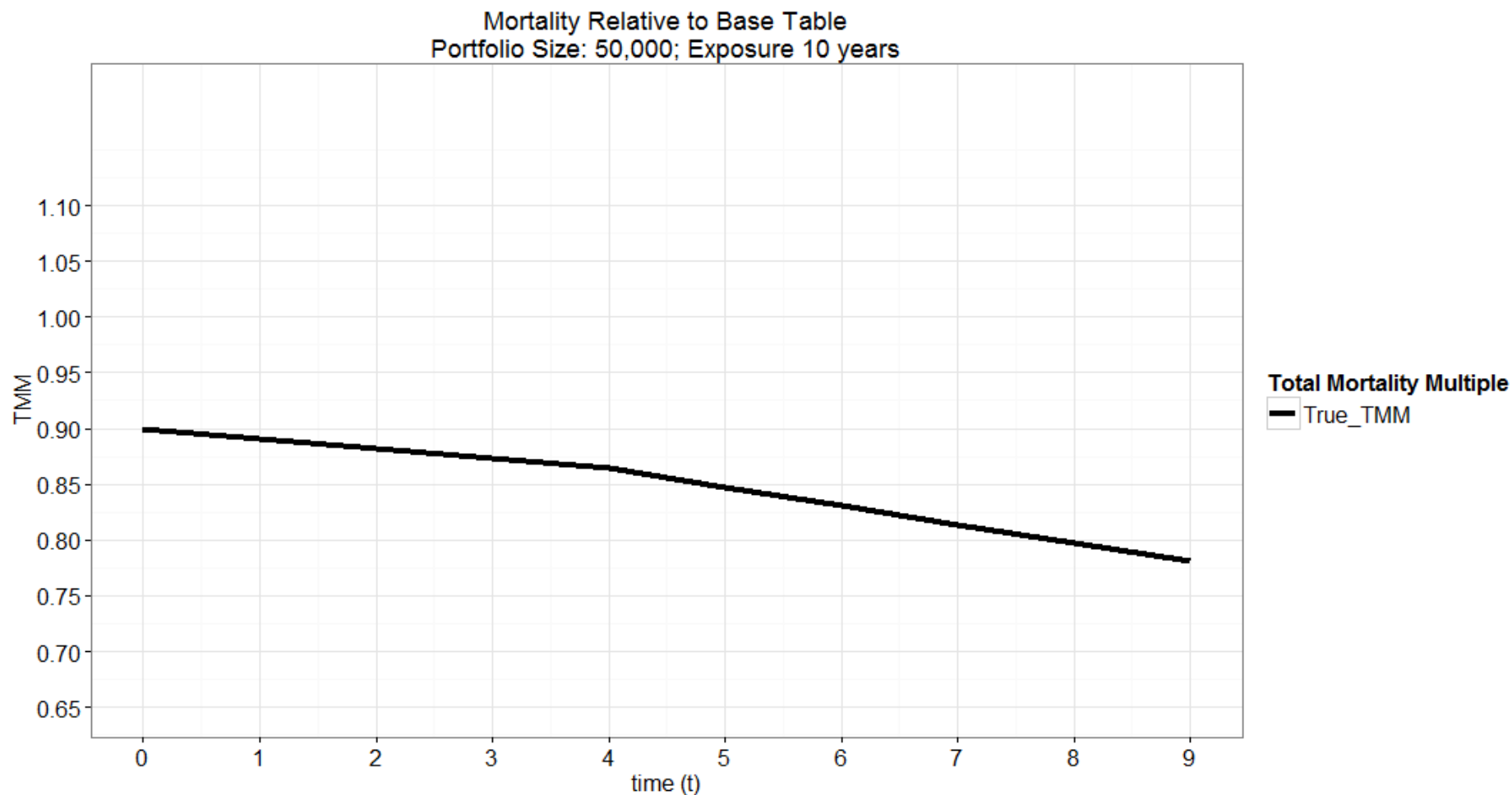
$\beta \sim Normal(mean = \mu_\beta, sd = \sigma_\beta) \dots trend\ coefficient$
 $TMM_0 \sim Normal(mean = \mu_{TMM}, sd = \sigma_{TMM})$
 $\sigma_\eta \sim Normal(0, .1) \ \& \ \sigma_\eta > 0$

* The above model is somewhat of a simplification to the one we use in practice but is useful for pedagogical purposes

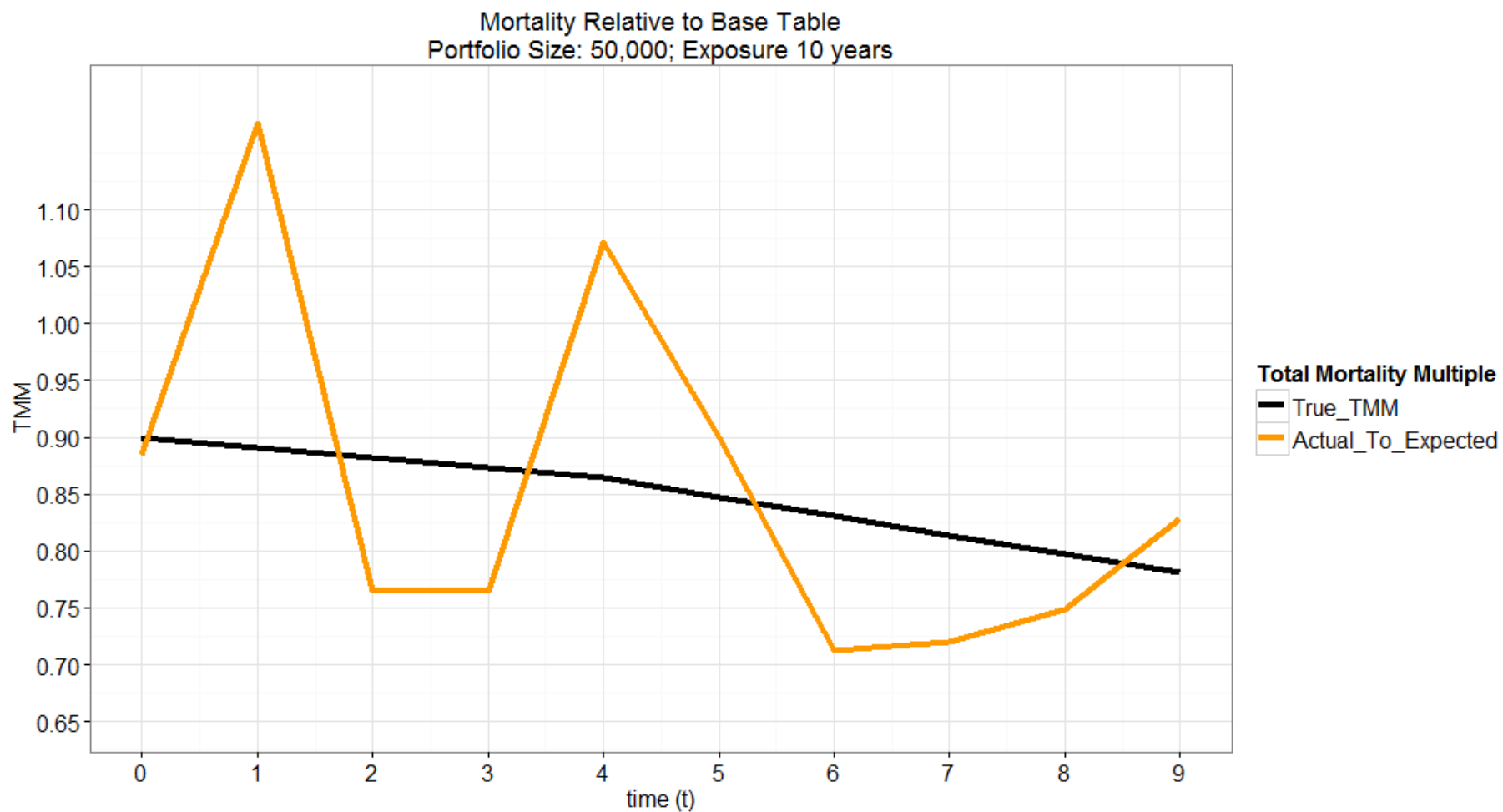
Hypothetical Example

- Background:
 - 50,000 Insured lives, male and female, aged 40 to 60
 - Followed for 10 years
- True (but unknown) Mortality:
 - 90% of 2008 VBT at issue (TMM0 = 90%)
 - Mortality Improvement = 1% p.a. for 4 years, 2% p.a. thereafter
- Approach:
 - Simulate mortality based on true mortality
 - Use State Space Model on simulated outcomes to “reverse engineer” true mortality

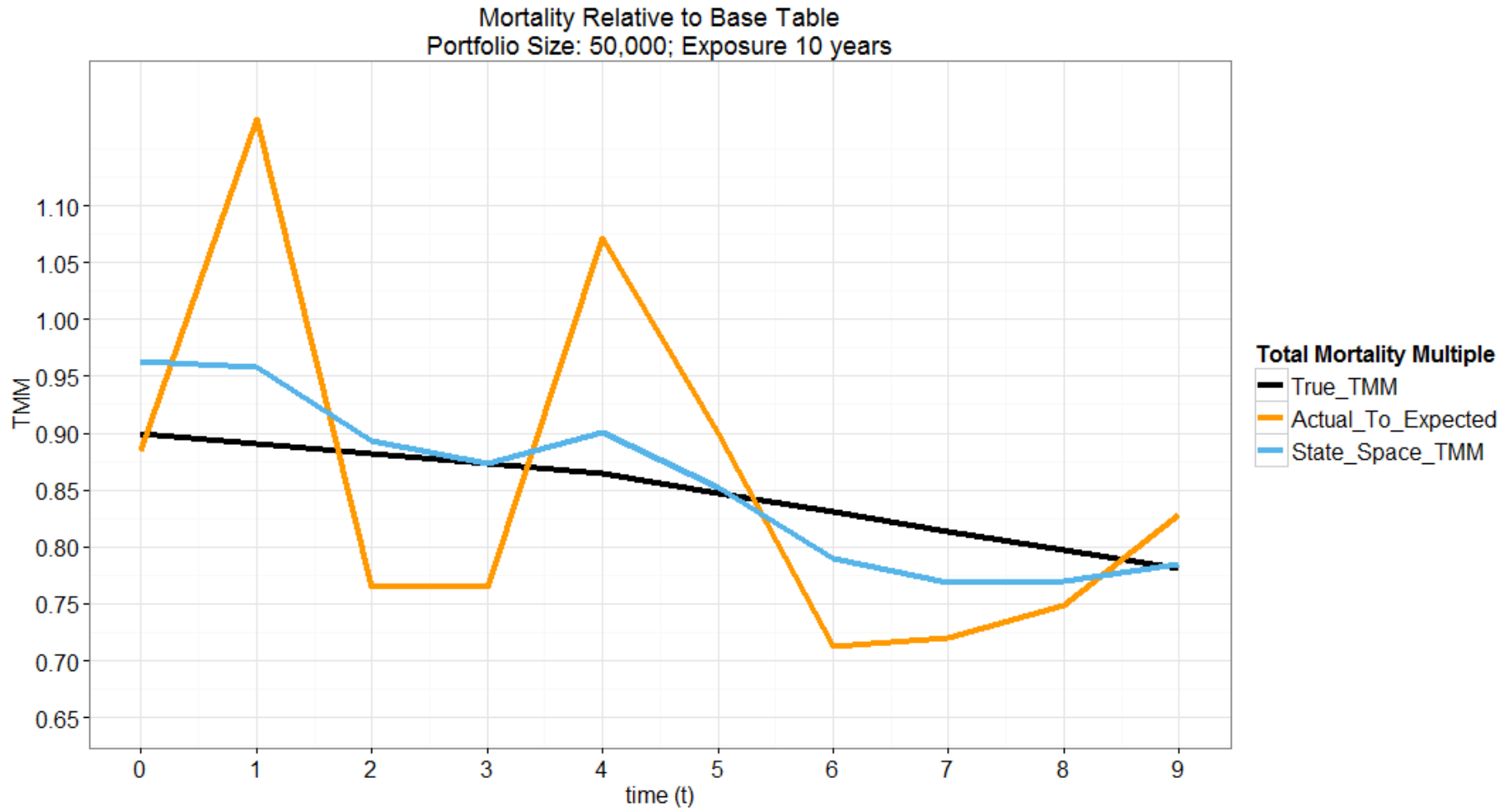
Mortality Assumptions



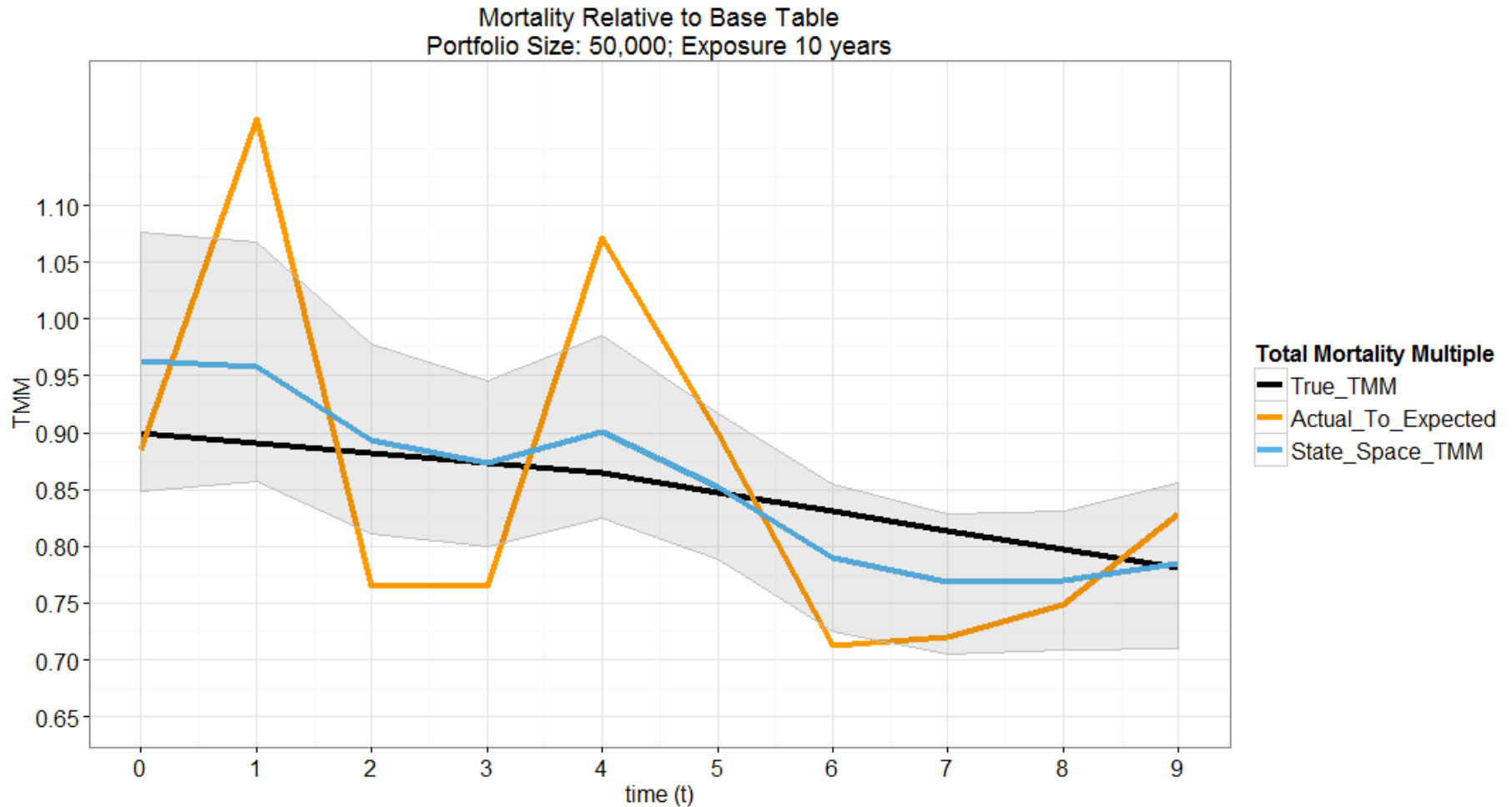
Simulated Mortality



State Space Model



State Space Model with 10% to 90% Confidence Interval



Advantage of the State Space Approach

- We now understand **how mortality has evolved** over the investigation period
 - State space terminology: “smoothed” set of TMM states
- We have a best estimate for **the current level of TMM** rather than an average over the investigation
 - State space terminology: “filtered” TMM state
- We know the **average trend** in mortality improvement, and the uncertainty/volatility of this estimate
- We learn the **credible interval** of our estimates for the current TMM
 - The credible interval shrinks with more data
- We **project the TMM going** forward with a data driven estimate for the volatility (uncertainty) associated with
 - Current TMM value; Improvement Trend
- We do not need to guess as to what **reasonable sensitivity test level** should be

IV: Identifying and Selecting Risk Drivers

Selecting Predictors

- Problem:
 - “Bias”: If too few predictors are selected, the model might not fit the data
 - “Variance”: If too many predictors are included, model will fit data very well, but would have poor predictive power
- Predictor selection is process by which we try to minimize both bias and variance

Possible Methods

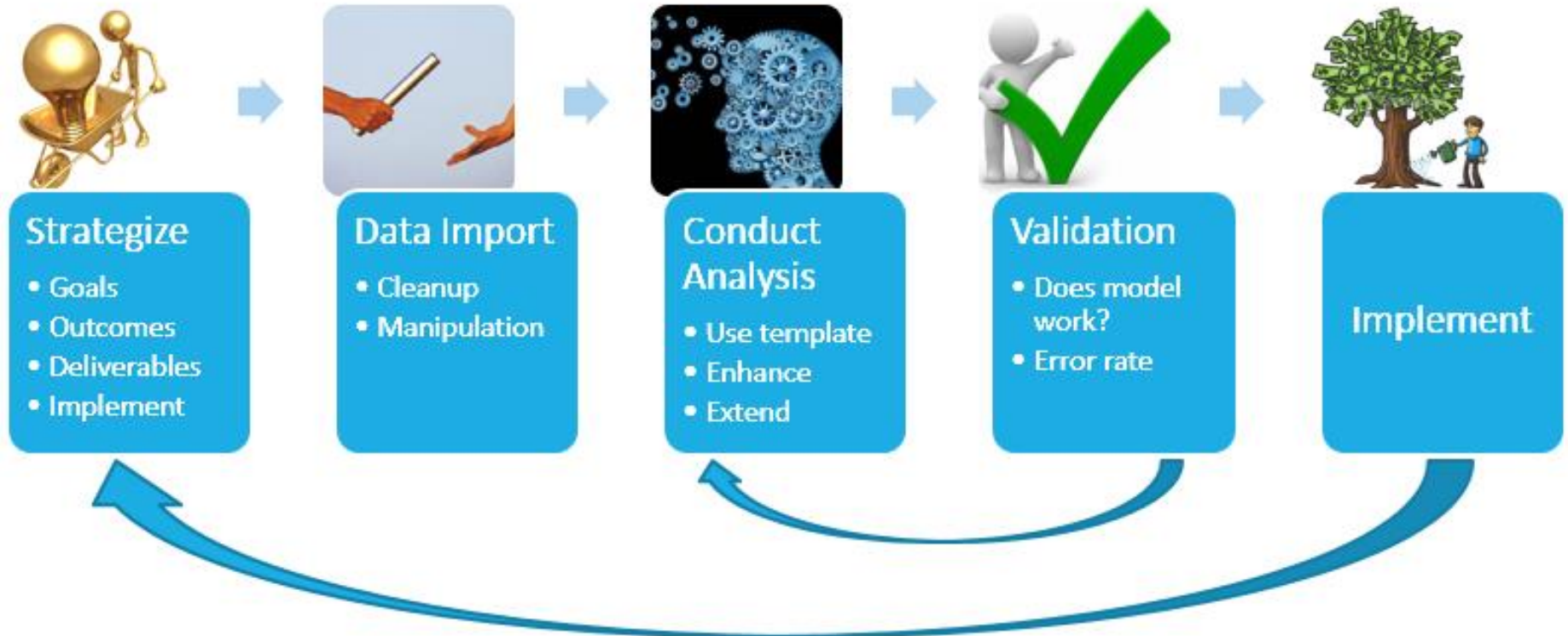
- Modern predictive analytics have come up with a number of possible solutions including:
 - K-Fold Cross Validation
 - Various “criteria” measures
 - AIC, WAIC, DIC
 - Forward and backward stepwise regression
 - Lasso approach
 - Use of strong priors
 - Hierarchical modeling

Relationships Among Death Rates and Risk Drivers

- Linear
- Non linear
 - Manually modeled
 - Splines approach
 - GAM smoothers
 - Gaussian Processes
- Interactive effects
- Bayesian networks dependency structure
- Hierarchical Modeling

V: Implementation Techniques

Predictive Analytics Workflow



Summary

- Bayesian methods directly determine both expected losses and volatility of risk for an insurer
- Bayesian methods scale well with complexity
- Probabilistic programming provides a means for practical implementation of techniques across a wide range of problems
- Methods provide a natural approach to quantify uncertainty for pricing, underwriting and capital management

Thank you for your attention

QUESTIONS ?