# Credit Risk: Simple Closed Form Approximate Maximum Likelihood Estimator, Large deviations and fast simulation

Sandeep Juneja

Tata Institute of Fundamental Research, India

joint work with Anand Deo (TIFR)

Work initially conducted at CAFRAL, Reserve Bank of India

IGIDR, September 2 2017

# The credit risk problems considered

- We want to estimate the conditional default probability of any firm as a function of given global and company specific information.
- This over short time periods - a month or a quarter, as well as longer time horizons.
- Analogous to predicting a person's health (mortality) as a function of his blood pressure, sugar, cholesterol, pollution, income, taxes (un)paid, etc.
- We model in discrete time and assume conditional probabilities have a popular default intensity type or logit type form

# The credit risk problems considered

- We want to estimate the conditional default probability of any firm as a function of given global and company specific information.
- This over short time periods - a month or a quarter, as well as longer time horizons.
- Analogous to predicting a person's health (mortality) as a function of his blood pressure, sugar, cholesterol, pollution, income, taxes (un)paid, etc.
- We model in discrete time and assume conditional probabilities have a popular default intensity type or logit type form

# The credit risk problems considered

- We want to estimate the conditional default probability of any firm as a function of given global and company specific information.
- This over short time periods - a month or a quarter, as well as longer time horizons.
- Analogous to predicting a person's health (mortality) as a function of his blood pressure, sugar, cholesterol, pollution, income, taxes (un)paid, etc.
- We model in discrete time and assume conditional probabilities have a popular default intensity type or logit type form

# The credit risk problems considered

- We want to estimate the conditional default probability of any firm as a function of given global and company specific information.
- This over short time periods - a month or a quarter, as well as longer time horizons.
- Analogous to predicting a person's health (mortality) as a function of his blood pressure, sugar, cholesterol, pollution, income, taxes (un)paid, etc.
- We model in discrete time and assume conditional probabilities have a popular default intensity type or logit type form

# Big picture - contributions

- The literature posits a parametric form for conditional default probabilities. Solves for parameters by maximising the likelihood function.
- Computationally intensive, solution has a black box flavour - drivers of the parameters not clear.
- We observe, in some popular settings, that since these probabilities are small, and co-variates can be transformed to be Gaussian, the MLE has a simple closed form approximation

# Big picture - contributions

- The literature posits a parametric form for conditional default probabilities. Solves for parameters by maximising the likelihood function.

- Computationally intensive, solution has a black box flavour - drivers of the parameters not clear.

- We observe, in some popular settings, that since these probabilities are small, and co-variates can be transformed to be Gaussian, the MLE has a simple closed form approximation

# Big picture - contributions

- The literature posits a parametric form for conditional default probabilities. Solves for parameters by maximising the likelihood function.
- Computationally intensive, solution has a black box flavour - drivers of the parameters not clear.
- We observe, in some popular settings, that since these probabilities are small, and co-variates can be transformed to be Gaussian, the MLE has a simple closed form approximation

- These are almost as good as MLE when the model is correctly specified - Performance slightly worsens for large number of firms (5,000 plus), large default probabilities (5%)

- Equally good or equally bad for mis-specified models, including on empirical data.

- We characterize the performance of the proposed approximate MLE as well as MLE in an asymptotic regime - probabilities decrease to zero, number of firms and number of time periods increase to infinity

- Some numerical/empirical support

- These are almost as good as MLE when the model is correctly specified - Performance slightly worsens for large number of firms (5,000 plus), large default probabilities (5%)
- Equally good  or equally bad for mis-specified models, including on empirical data.
- We characterize the performance of the proposed approximate MLE as well as MLE in an asymptotic regime - probabilities decrease to zero, number of firms and number of time periods increase to infinity
- Some numerical/empirical support

- These are almost as good as MLE when the model is correctly specified - Performance slightly worsens for large number of firms (5,000 plus), large default probabilities (5%)
- Equally good or equally bad for mis-specified models, including on empirical data.
- We characterize the performance of the proposed approximate MLE as well as MLE in an asymptotic regime - probabilities decrease to zero, number of firms and number of time periods increase to infinity
- Some numerical/empirical support

- These are almost as good as MLE when the model is correctly specified - Performance slightly worsens for large number of firms (5,000 plus), large default probabilities (5%)
- Equally good or equally bad for mis-specified models, including on empirical data.
- We characterize the performance of the proposed approximate MLE as well as MLE in an asymptotic regime - probabilities decrease to zero, number of firms and number of time periods increase to infinity
- Some numerical/empirical support

# The credit risk problems maybe not considered ...

- We analyze the portfolio credit risk problem and develop an asymptotic regime where
  - we conduct large deviations tail analysis of large losses
  - Develop fast simulation techniques for computing tail risk measures
- Both in calibration as well as in portfolio analysis, we include contagion effects in the model in a meaningful way.

# The credit risk problems maybe not considered ...

- We analyze the portfolio credit risk problem and develop an asymptotic regime where
  - we conduct large deviations tail analysis of large losses
  - Develop fast simulation techniques for computing tail risk measures
- Both in calibration as well as in portfolio analysis, we include contagion effects in the model in a meaningful way.

# The credit risk problems maybe not considered ...

- We analyze the portfolio credit risk problem and develop an asymptotic regime where
  - we conduct large deviations tail analysis of large losses
  - Develop fast simulation techniques for computing tail risk measures
- Both in calibration as well as in portfolio analysis, we include contagion effects in the model in a meaningful way.

# The credit risk problems maybe not considered ...

- We analyze the portfolio credit risk problem and develop an asymptotic regime where
  - we conduct large deviations tail analysis of large losses
  - Develop fast simulation techniques for computing tail risk measures
- Both in calibration as well as in portfolio analysis, we include contagion effects in the model in a meaningful way.

# Review and models considered

# Popular reduced form intensity based models

- For each firm $i$, default explaining covariates such as prevailing interest rates, GDP, distance to default, cash over total assets, modelled as a continuous Markov process, example,

$$dX_{i,t} = A_i(\theta_i - X_{i,t})dt + \Sigma_i dW_{i,t}$$

for $0 \leq t \leq T$.

- Firm $i$ has a doubly stochastic default intensity process

$$\lambda_i(t) = \Lambda_i(\beta, X_{i,t})$$

where $\beta$ is the set of parameters to be estimated.

- **Conditioned on the covariates, default is an arrival from a non-homogeneous Poisson process.**

# Popular reduced form intensity based models

- For each firm $i$, default explaining covariates such as prevailing interest rates, GDP, distance to default, cash over total assets, modelled as a continuous Markov process, example,

$$dX_{i,t} = A_i(\theta_i - X_{i,t})dt + \Sigma_i dW_{i,t}$$

for $0 \leq t \leq T$.

- Firm $i$ has a doubly stochastic default intensity process

$$\lambda_i(t) = \Lambda_i(\beta, X_{i,t})$$

where $\beta$ is the set of parameters to be estimated.

- Conditioned on the covariates, default is an arrival from a non-homogeneous Poisson process.

# Popular reduced form intensity based models

- For each firm $i$, default explaining covariates such as prevailing interest rates, GDP, distance to default, cash over total assets, modelled as a continuous Markov process, example,

$$dX_{i,t} = A_i(\theta_i - X_{i,t})dt + \Sigma_i dW_{i,t}$$

for $0 \leq t \leq T$.

- Firm $i$ has a doubly stochastic default intensity process

$$\lambda_i(t) = \Lambda_i(\beta, X_{i,t})$$

where $\beta$ is the set of parameters to be estimated.

- **Conditioned on the covariates, default is an arrival from a non-homogeneous Poisson process.**

# Conditional probability under discretization

$$P(\text{ no default by } t+1 | \text{ no default by time } t) = E_t e^{-\int_t^{t+1} \lambda_i(s) ds}$$

- If we assume that over time period $[t, t+1)$

$$\lambda_i(s) = \exp(\beta^\mathsf{T} X_{i,t} - \alpha)$$

then

$$P(\text{ default by } t+1 | \text{ no default by time } t) = 1 - \exp(-\exp(\beta^\mathsf{T} X_{i,t} - \alpha))$$

# Conditional probability under discretization

$$P(\text{ no default by } t+1| \text{ no default by time } t) = E_t e^{-\int_t^{t+1} \lambda_i(s) ds}$$

- If we assume that over time period $[t, t+1)$

$$\lambda_i(s) = \exp(\beta^{\mathsf{T}} X_{i,t} - \alpha)$$

then

$$P(\text{ default by } t+1| \text{ no default by time } t) = 1 - \exp(-\exp(\beta^{\mathsf{T}} X_{i,t} - \alpha))$$

# Discrete Logit models

- Covariates affecting Firm $i$ follow a stationary process $\{X_{i,t}\}$
- Conditional default probability at period $t$ to default in $[t, t+1)$

$$p_t(X_{i,t}) = \frac{\exp(\beta^T X_{i,t} - \alpha)}{1 + \exp(\beta^T X_{i,t} - \alpha)}$$

# Discrete Logit models

- Covariates affecting Firm $i$ follow a stationary process $\{X_{i,t}\}$

- Conditional default probability at period $t$ to default in $[t, t+1)$

$$p_t(X_{i,t}) = \frac{\exp(\beta^T X_{i,t} - \alpha)}{1 + \exp(\beta^T X_{i,t} - \alpha)}$$

# Model and Maximum likelihood estimation

# We consider a discrete time model

- Covariates - Autoregressive process

$$X_{i,t+1} = \mathbf{A}X_{i,t} + \tilde{\mathbf{E}}_{i,t+1}$$

- Conditional default probability at period $t$ to default within $[t, t+1)$

$$p(X_{i,t}) = 1 - \exp\left(-\exp(\beta^T X_{i,t} - \alpha)\right)$$

or

$$p(X_{i,t}) = \frac{\exp(\beta^T X_{i,t} - \alpha)}{1 + \exp(\beta^T X_{i,t} - \alpha)},$$

- Parameters $\beta$, $\alpha$ need to be estimated from data. Duffie et al 2006, Duan et al 2007, Chava and Jarrow 2004, Duan, Sun Wang 2012, Shumway 2002.

# We consider a discrete time model

- Covariates - Autoregressive process

$$X_{i,t+1} = \mathbf{A} X_{i,t} + \tilde{\mathbf{E}}_{i,t+1}$$

- Conditional default probability at period $t$ to default within $[t, t+1)$

$$p(X_{i,t}) = 1 - \exp\left(-\exp(\beta^T X_{i,t} - \alpha)\right)$$

or

$$p(X_{i,t}) = \frac{\exp(\beta^T X_{i,t} - \alpha)}{1 + \exp(\beta^T X_{i,t} - \alpha)},$$

- Parameters $\beta$, $\alpha$ need to be estimated from data. Duffie et al 2006, Duan et al 2007, Chava and Jarrow 2004, Duan, Sun Wang 2012, Shumway 2002.

# We consider a discrete time model

- Covariates - Autoregressive process

$$X_{i,t+1} = \mathbf{A} X_{i,t} + \tilde{\mathbf{E}}_{i,t+1}$$

- Conditional default probability at period $t$ to default within $[t, t+1)$

$$p(X_{i,t}) = 1 - \exp\left(-\exp(\beta^T X_{i,t} - \alpha)\right)$$

or

$$p(X_{i,t}) = \frac{\exp(\beta^T X_{i,t} - \alpha)}{1 + \exp(\beta^T X_{i,t} - \alpha)},$$

- Parameters $\beta$, $\alpha$ need to be estimated from data. Duffie et al 2006, Duan et al 2007, Chava and Jarrow 2004, Duan, Sun Wang 2012, Shumway 2002.

# We consider a discrete time model

- Covariates - Autoregressive process

$$X_{i,t+1} = \mathbf{A}X_{i,t} + \tilde{\mathbf{E}}_{i,t+1}$$

- Conditional default probability at period $t$ to default within $[t, t+1)$

$$p(X_{i,t}) = 1 - \exp\left(-\exp(\beta^T X_{i,t} - \alpha)\right)$$

or

$$p(X_{i,t}) = \frac{\exp(\beta^T X_{i,t} - \alpha)}{1 + \exp(\beta^T X_{i,t} - \alpha)},$$

- Parameters $\beta$, $\alpha$ need to be estimated from data. Duffie et al 2006, Duan et al 2007, Chava and Jarrow 2004, Duan, Sun Wang 2012, Shumway 2002.

# Maximum likelihood method to estimate parameters

- Default data

$$(x_{i,t}, d_{i,t}) \text{ for } t = 1, , 2, \ldots, T, i = 1, \ldots, m,$$

where $d_{i,t} = 1$ if company $i$ defaults in $[t, t+1)$ and zero otherwise.

- Likelihood $\mathcal{L}$ of seeing the data

$$\mathcal{L} = \prod_{i,t} p(x_{i,t})^{d_{i,t}}(1 - p(x_{i,t}))^{1-d_{i,t}}$$

- This is optimized numerically to find $\beta$ and $\alpha$.
- Computationally intensive; black box.

# Maximum likelihood method to estimate parameters

- Default data

$$(x_{i,t}, d_{i,t}) \text{ for } t = 1, , 2, \ldots, T, i = 1, \ldots, m,$$

where $d_{i,t} = 1$ if company $i$ defaults in $[t, t+1)$ and zero otherwise.

- Likelihood $\mathcal{L}$ of seeing the data

$$\mathcal{L} = \prod_{i,t} p(x_{i,t})^{d_{i,t}}(1 - p(x_{i,t}))^{1-d_{i,t}}$$

- This is optimized numerically to find $\beta$ and $\alpha$.
- Computationally intensive; black box.

# Maximum likelihood method to estimate parameters

- Default data

$$(x_{i,t}, d_{i,t}) \text{ for } t = 1, , 2, \ldots, T, i = 1, \ldots, m,$$

where $d_{i,t} = 1$ if company $i$ defaults in $[t, t+1)$ and zero otherwise.

- Likelihood $\mathcal{L}$ of seeing the data

$$\mathcal{L} = \prod_{i,t} p(x_{i,t})^{d_{i,t}} (1 - p(x_{i,t}))^{1-d_{i,t}}$$

- This is optimized numerically to find $\beta$ and $\alpha$.
- Computationally intensive; black box.

# MLE: Default intensity model

- $$p(x_{i,t}) = 1 - \exp(-e^{\beta^\intercal x_{i,t} - \alpha}).$$

- Setting the partial derivatives w.r.t. $(\beta, \alpha)$ to zero,

$$\sum_{i,t} \frac{x_{i,t} e^{\beta^\intercal x_{i,t} - \alpha}}{1 - \exp(-e^{\beta^\intercal x_{i,t} - \alpha})} d_{i,t} = \sum_{i,t} x_{i,t} e^{\beta^\intercal x_{i,t} - \alpha}$$

  and

$$\sum_{i,t} \frac{e^{\beta^\intercal x_{i,t} - \alpha}}{1 - \exp(-e^{\beta^\intercal x_{i,t} - \alpha})} d_{i,t} = \sum_{i,t} e^{\beta^\intercal x_{i,t} - \alpha}.$$

# MLE: Logit model

- $$p(x_{i,t}) = \frac{\exp(\beta^\intercal x_{i,t} - \alpha)}{1 + \exp(\beta^\intercal x_{i,t} - \alpha)}.$$

Setting the partial derivatives to zero:

$$\sum_{i,t} x_{i,t} d_{i,t} = \sum_{i,t} x_{i,t} \frac{\exp(\beta^\intercal x_{i,t} - \alpha)}{1 + \exp(\beta^\intercal x_{i,t} - \alpha)}.$$

and

$$\sum_{i,t} d_{i,t} = \sum_{i,t} \frac{\exp(\beta^\intercal x_{i,t} - \alpha)}{1 + \exp(\beta^\intercal x_{i,t} - \alpha)}.$$

# Key insight

- Re-express

$$\frac{1}{\#\ i,t}\sum_{i,t} d_{i,t} = E\left(\exp(\beta^{\mathsf{T}} X_{i,t} - \alpha)\right)$$

$$+\left(\frac{1}{\#\ i,t}\sum_{i,t}\frac{\exp(\beta^{\mathsf{T}} x_{i,t} - \alpha)}{1+\exp(\beta^{\mathsf{T}} x_{i,t} - \alpha)} - E\left(\exp(\beta^{\mathsf{T}} X_{i,t} - \alpha)\right)\right).$$

and

$$\frac{1}{\#\ i,t}\sum_{i,t} x_{i,t} d_{i,t} = EX_{i,t}\left(\exp(\beta^{\mathsf{T}} X_{i,t} - \alpha)\right)$$

$$+\left(\frac{1}{\#\ i,t}\sum_{i,t} x_{i,t}\frac{\exp(\beta^{\mathsf{T}} x_{i,t} - \alpha)}{1+\exp(\beta^{\mathsf{T}} x_{i,t} - \alpha)} - EX_{i,t}\left(\exp(\beta^{\mathsf{T}} X_{i,t} - \alpha)\right)\right).$$

# Key insight

- Re-express

$$\frac{1}{\#\ i,t}\sum_{i,t} d_{i,t} = E\left(\exp(\beta^\intercal X_{i,t}-\alpha)\right)$$

$$+\left(\frac{1}{\#\ i,t}\sum_{i,t}\frac{\exp(\beta^\intercal x_{i,t}-\alpha)}{1+\exp(\beta^\intercal x_{i,t}-\alpha)} - E\left(\exp(\beta^\intercal X_{i,t}-\alpha)\right)\right).$$

and

$$\frac{1}{\#\ i,t}\sum_{i,t} x_{i,t}d_{i,t} = EX_{i,t}\left(\exp(\beta^\intercal X_{i,t}-\alpha)\right)$$

$$+\left(\frac{1}{\#\ i,t}\sum_{i,t}x_{i,t}\frac{\exp(\beta^\intercal x_{i,t}-\alpha)}{1+\exp(\beta^\intercal x_{i,t}-\alpha)} - EX_{i,t}\left(\exp(\beta^\intercal X_{i,t}-\alpha)\right)\right).$$

- When $X$ is Gaussian:

$$E\left[\exp(\beta^{\mathsf{T}}X)\right] = \exp(\frac{1}{2}\beta^{\mathsf{T}}\Sigma\beta) \text{ and } E\left[X\exp(\beta^{\mathsf{T}}X)\right] = \Sigma\beta\exp(\frac{1}{2}\beta^{\mathsf{T}}\Sigma\beta).$$

- This suggests that we set the estimator to dominant term

$$\hat{\beta} = \Sigma_i^{-1}\left(\frac{\sum_{i,t} x_{i,t} d_{i,t}}{\sum_{i,t} d_{i,t}}\right).$$

$\hat{\alpha}$ is chosen to match the observed default probability.

- When $X$ is Gaussian:

$$E\left[\exp(\beta^\mathsf{T} X)\right] = \exp(\frac{1}{2}\beta^\mathsf{T}\Sigma\beta) \text{ and } E\left[X\exp(\beta^\mathsf{T} X)\right] = \Sigma\beta\exp(\frac{1}{2}\beta^\mathsf{T}\Sigma\beta).$$

- This suggests that we set the estimator to dominant term

$$\hat{\beta} = \Sigma_i^{-1}\left(\frac{\sum_{i,t} x_{i,t} d_{i,t}}{\sum_{i,t} d_{i,t}}\right).$$

$\hat{\alpha}$ is chosen to match the observed default probability.

# How good is the estimator

# Analysis of estimator quality in asymptotic regime

- We model conditional default probabilities as

$$p_\gamma(X_{i,t}) = \exp(\beta^T X_{i,t} - \alpha(\gamma))(1 + o_\gamma(1))$$

where $\alpha(\gamma)$ is of order $\log(1/\gamma)$, $\{X_t\}$ is a vector autoregressive process.

- Conditional probabilities of order $\gamma$ ( $\approx 10^{-3}$)
- Number of companies is of order $\frac{1}{\gamma^\delta}$ for $\delta > 0$
- Number of time periods of provided data is $\frac{1}{\gamma^\zeta}$ for $\zeta \in (0, 1)$.

# Analysis of estimator quality in asymptotic regime

- We model conditional default probabilities as

$$p_\gamma(X_{i,t}) = \exp(\beta^T X_{i,t} - \alpha(\gamma))(1 + o_\gamma(1))$$

where $\alpha(\gamma)$ is of order $\log(1/\gamma)$, $\{X_t\}$ is a vector autoregressive process.

- Conditional probabilities of order $\gamma$ ( $\approx 10^{-3}$)
- Number of companies is of order $\frac{1}{\gamma^\delta}$ for $\delta > 0$
- Number of time periods of provided data is $\frac{1}{\gamma^\zeta}$ for $\zeta \in (0, 1)$.

# Analysis of estimator quality in asymptotic regime

- We model conditional default probabilities as

$$p_\gamma(X_{i,t}) = \exp(\beta^T X_{i,t} - \alpha(\gamma))(1 + o_\gamma(1))$$

where $\alpha(\gamma)$ is of order $\log(1/\gamma)$, $\{X_t\}$ is a vector autoregressive process.

- Conditional probabilities of order $\gamma$ ( $\approx 10^{-3}$)
- Number of companies is of order $\frac{1}{\gamma^\delta}$ for $\delta > 0$
- Number of time periods of provided data is $\frac{1}{\gamma^\zeta}$ for $\zeta \in (0, 1)$.

# Analysis of estimator quality in asymptotic regime

- We model conditional default probabilities as

$$p_\gamma(X_{i,t}) = \exp(\beta^T X_{i,t} - \alpha(\gamma))(1 + o_\gamma(1))$$

where $\alpha(\gamma)$ is of order $\log(1/\gamma)$, $\{X_t\}$ is a vector autoregressive process.

- Conditional probabilities of order $\gamma$ ( $\approx 10^{-3}$)
- Number of companies is of order $\frac{1}{\gamma^\delta}$ for $\delta > 0$
- Number of time periods of provided data is $\frac{1}{\gamma^\zeta}$ for $\zeta \in (0, 1)$.

- Our estimator in this regime

$$\hat{\beta} = \Sigma_{X,X}^{-1} \left( \frac{\sum_{i,t} X_{i,t} D_{i,t}}{\sum_{i,t} D_{i,t}} \right)$$

.

- Recall that $\{X_t\}$ captures the underlying covariates. $D_{i,t}$ are default indicators.
- This converges to $\beta$ as $\gamma \to 0$.

- Our estimator in this regime

$$\hat{\beta} = \Sigma_{X,X}^{-1} \left( \frac{\sum_{i,t} X_{i,t} D_{i,t}}{\sum_{i,t} D_{i,t}} \right)$$

.

- Recall that $\{X_t\}$ captures the underlying covariates. $D_{i,t}$ are default indicators.

- This converges to $\beta$ as $\gamma \to 0$.

- Our estimator in this regime

$$\hat{\beta} = \Sigma_{X,X}^{-1} \left( \frac{\sum_{i,t} X_{i,t} D_{i,t}}{\sum_{i,t} D_{i,t}} \right)$$

.

- Recall that $\{X_t\}$ captures the underlying covariates. $D_{i,t}$ are default indicators.

- This converges to $\beta$ as $\gamma \to 0$.

# Mean square error analysis: Proposed estimator

- Theorem: The mean square error

$$||\hat{\beta} - \beta||^2 = \Theta(\gamma^{\delta+\zeta-1}) + \Theta(\gamma^{\zeta}).$$

- $\delta + \zeta < 1$: No defaults asymptotically
- $\delta < 1$: Increasing $\delta$ helps the estimator. *More firms in dataset improve the estimator*
- $\delta > 1$: increasing $\delta$ no longer matters. *Insensitive to increase in the number of firms*

# Mean square error analysis: Proposed estimator

- Theorem: The mean square error

$$||\hat{\beta} - \beta||^2 = \Theta(\gamma^{\delta+\zeta-1}) + \Theta(\gamma^{\zeta}).$$

- $\delta + \zeta < 1$: No defaults asymptotically
- $\delta < 1$: Increasing $\delta$ helps the estimator. *More firms in dataset improve the estimator*
- $\delta > 1$: increasing $\delta$ no longer matters. *Insensitive to increase in the number of firms*

# Mean square error analysis: Proposed estimator

- Theorem: The mean square error

$$||\hat{\beta} - \beta||^2 = \Theta(\gamma^{\delta+\zeta-1}) + \Theta(\gamma^{\zeta}).$$

- $\delta + \zeta < 1$: No defaults asymptotically
- $\delta < 1$: Increasing $\delta$ helps the estimator. *More firms in dataset improve the estimator*
- $\delta > 1$: increasing $\delta$ no longer matters. *Insensitive to increase in the number of firms*

# Mean square error analysis: Proposed estimator

- Theorem: The mean square error
$$||\hat{\beta} - \beta||^2 = \Theta(\gamma^{\delta+\zeta-1}) + \Theta(\gamma^{\zeta}).$$

- $\delta + \zeta < 1$: No defaults asymptotically
- $\delta < 1$: Increasing $\delta$ helps the estimator. *More firms in dataset improve the estimator*
- $\delta > 1$: increasing $\delta$ no longer matters. *Insensitive to increase in the number of firms*

# Mean square error analysis for MLE

- Theorem: The mean square error equals

$$\Theta(\gamma^{\delta+\zeta-1})$$

- Both the estimators can be shown to satisfy a central limit theorem. Maybe useful for constructing confidence intervals

# Mean square error analysis for MLE

- Theorem: The mean square error equals

$$\Theta(\gamma^{\delta+\zeta-1})$$

- Both the estimators can be shown to satisfy a central limit theorem. Maybe useful for constructing confidence intervals

# Model misspecification: An illustration

- Model generating defaults has two Gaussian factors common to all firms:

$$\frac{\exp(\beta_1 Y_{1,t} + \beta_2 Y_{2,t} - \alpha(\gamma))}{1 + \exp(\beta_1 Y_{1,t} + \beta_2 Y_{2,t} - \alpha(\gamma))},$$

  $Y_{1,t}$ and $Y_{2,t}$ are assumed to have zero mean, variance 1 and correlation $\rho$

- Only the first factor with time series $(Y_{1,t} : 1 \leq t \leq T(\gamma))$ is assumed to be relevant by modeller.

- Both estimators asymptotically converge to

$$\hat{\beta}_1 = \beta_1 + \rho\beta_2$$

  and

$$\hat{\alpha} = \alpha(\gamma) - \frac{\beta_2^2(1-\rho^2)}{2}.$$

# Model misspecification: An illustration

- Model generating defaults has two Gaussian factors common to all firms:

$$\frac{\exp(\beta_1 Y_{1,t} + \beta_2 Y_{2,t} - \alpha(\gamma))}{1 + \exp(\beta_1 Y_{1,t} + \beta_2 Y_{2,t} - \alpha(\gamma))},$$

  $Y_{1,t}$ and $Y_{2,t}$ are assumed to have zero mean, variance 1 and correlation $\rho$

- Only the first factor with time series $(Y_{1,t} : 1 \leq t \leq T(\gamma))$ is assumed to be relevant by modeller.

- Both estimators asymptotically converge to

$$\hat{\beta}_1 = \beta_1 + \rho\beta_2$$

  and

$$\hat{\alpha} = \alpha(\gamma) - \frac{\beta_2^2(1-\rho^2)}{2}.$$

# Model misspecification: An illustration

- Model generating defaults has two Gaussian factors common to all firms:

$$\frac{\exp(\beta_1 Y_{1,t} + \beta_2 Y_{2,t} - \alpha(\gamma))}{1 + \exp(\beta_1 Y_{1,t} + \beta_2 Y_{2,t} - \alpha(\gamma))},$$

  $Y_{1,t}$ and $Y_{2,t}$ are assumed to have zero mean, variance 1 and correlation $\rho$

- Only the first factor with time series $(Y_{1,t} : 1 \leq t \leq T(\gamma))$ is assumed to be relevant by modeller.

- Both estimators asymptotically converge to

$$\hat{\beta}_1 = \beta_1 + \rho\beta_2$$

  and

$$\hat{\alpha} = \alpha(\gamma) - \frac{\beta_2^2(1-\rho^2)}{2}.$$

# Simulation Experiments

# Comparison of RMSE for default probability 1% per annum, model correctly specified

| Time in months | No. of firms | RMSE($\beta_{prop}$) | RMSE($\beta_{ML}$) |
|---|---|---|---|
| 200 | 1000 | 0.1280 | 0.1248 |
| 200 | 3000 | 0.0787 | 0.0707 |
| 200 | 5000 | 0.0685 | 0.0574 |
| 200 | 7000 | 0.0574 | 0.0435 |
| 200 | 10000 | 0.0547 | 0.0374 |
| 100 | 2000 | 0.1232 | 0.1157 |
| 300 | 2000 | 0.0774 | 0.0714 |
| 500 | 2000 | 0.0608 | 0.0547 |
| 700 | 2000 | 0.0565 | 0.0509 |

True Parameters: $(\alpha = 7.5, \beta_1 = -0.2, \beta_2 = 0.5, \beta_3 = 0.5)$. RMSE of the proposed estimator is only slightly larger than that of MLE except when the no. of companies is large.

# Comparison of RMSE for default probability 1% per annum, missing covariate with small and large coefficient

| $\beta_3$ | No. of firms | RMSE($\beta_{prop}$) | RMSE($\beta_{ML}$) |
|---|---|---|---|
| 0.5 | 1000 | 0.1403 | 0.1392 |
| 0.5 | 3000 | 0.0871 | 0.0842 |
| 0.5 | 5000 | 0.0741 | 0.0721 |
| 0.5 | 7000 | 0.0754 | 0.0707 |
| 2 | 1000 | 0.3109 | 0.3231 |
| 2 | 3000 | 0.2958 | 0.3041 |
| 2 | 5000 | 0.3046 | 0.3135 |
| 2 | 7000 | 0.3014 | 0.3072 |

True Parameters: ($\alpha = 7.5, \beta_1 = -0.2, \beta_2 = 0.5$), $\beta_3$. Time period 200. Both the proposed estimator and MLE estimate ($\alpha, \beta_1, \beta_2$) only. The RMSE of the two methods is nearly identical. It worsens as value of $\beta_3$ increases.

# Empirical Analysis

# Sample Data Characteristics (From Risk Management Institute, NUS)

1. Number of Companies: 2,000
2. Time Periods:251
3. Defaults:168
4. Default Probability per year: 1.12%
5. Number of Variables Available: 8

# Description of Variables

**Macroeconomic Variables**

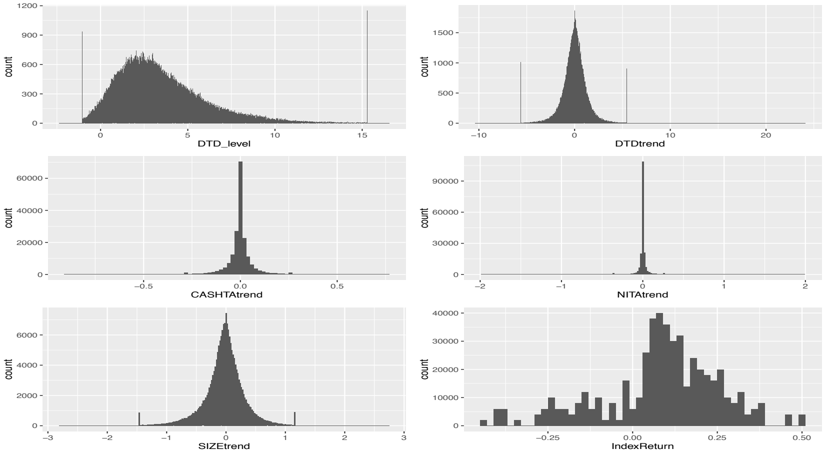
1. IndexReturn: trailing 1-year return on the S&P500 index
2. Treasury rate: 3-month US Treasury bill rate

**Firm-Specific Variables**

1. DTD: firms distance-to-default
2. CASH/TA: ratio of the sum of cash and short-term investments to the total assets
3. NI/TA: ratio of net income to the total assets
4. SIZE: log of the ratio of firms market equity value to the average market equity value of the S&P500 firm
5. M/B: market-to-book asset ratio
6. SIGMA: 1-year idiosyncratic volatility

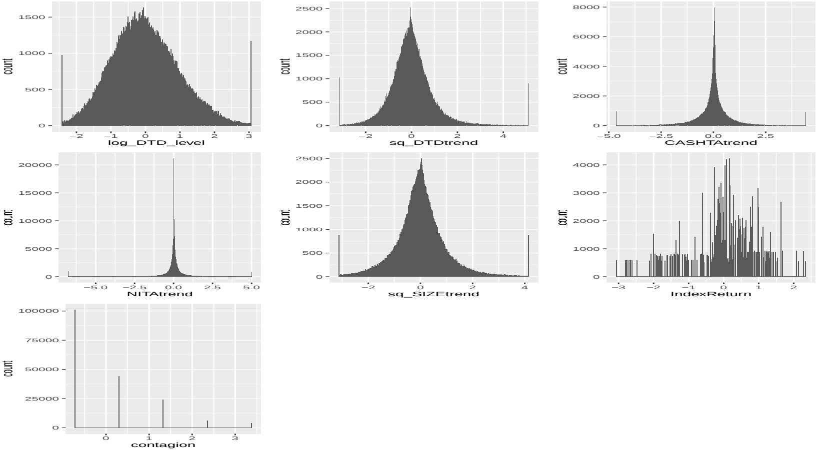
# Assumption of Normality/Boundedness

Figure: Frequency Plots of Variables without Transformation



# Assumption of Normality/Boundedness

Figure: Frequency Plots of Variables after Transformation



# Calibration Betas

Table: Combined Beta Table

| Decile | Our Calibration | Duffie's MLE | Logit |
|---|---|---|---|
| Constant | -9.251 | -6.739 | -9.344 |
| log_DTD_level | -1.330 | -0.425 | -1.837 |
| sq_DTDtrend | -0.199 | 0.320 | -1.267 |
| CASHTAtrend | -0.035 | 0.006 | -0.045 |
| NITAtrend | -0.417 | -0.108 | -0.060 |
| sq_SIZEtrend | -1.477 | -0.615 | -0.565 |
| IndexReturn | -0.342 | -0.089 | -0.218 |

# Calibration Results

Table: Combined Accuracy Table

| Decile | Our Calibration | Duffie's MLE | Logit |
|---|---|---|---|
| 1 | 0.895 | 0.842 | 0.763 |
| 2 | 0.974 | 0.947 | 0.921 |
| 3 | 0.974 | 0.974 | 0.947 |
| 4 | 1 | 0.974 | 0.947 |
| 5 | 1 | 0.974 | 0.947 |
| 6 | 1 | 0.974 | 0.974 |
| 7 | 1 | 0.974 | 1 |
| 8 | 1 | 1 | 1 |
| 9 | 1 | 1 | 1 |

# Calibration Betas with Contagion

Table: Combined Beta Table with Contagion

| Variable | Our Calibration | Duffie's MLE | Logit |
|---|---|---|---|
| Constant | -9.806 | -6.811 | -9.145 |
| log_DTD_level | -1.281 | -0.322 | -1.587 |
| sq_DTDtrend | -0.174 | 0.072 | -1.235 |
| CASHTAtrend | -0.033 | 0.181 | -0.042 |
| NITAtrend | -0.410 | 0.223 | -0.061 |
| sq_SIZEtrend | -1.462 | -0.755 | -0.582 |
| IndexReturn | 0.021 | 0.007 | -0.198 |
| Contagion | 1.117 | 0.194 | 0.046 |

# Calibration Results with Contagion

Table: Combined Accuracy Table with Contagion

| Decile | Our Calibration | Duffie's MLE | Logit |
|---|---|---|---|
| 1 | 0.921 | 0.868 | 0.763 |
| 2 | 0.974 | 0.947 | 0.921 |
| 3 | 0.974 | 0.974 | 0.947 |
| 4 | 1 | 1 | 0.947 |
| 5 | 1 | 1 | 0.974 |
| 6 | 1 | 1 | 0.974 |
| 7 | 1 | 1 | 1 |
| 8 | 1 | 1 | 1 |
| 9 | 1 | 1 | 1 |

Table: Computer Generated Data Coefficient Results

| | Underlying Logit Betas | Our Betas | Duffie Betas |
|---|---|---|---|
| Constant | -7.600 | -7.490 | -7.566 |
| CVar1 | 0.225 | 0.198 | 0.224 |
| CVar2 | 0.549 | 0.536 | 0.561 |
| CVar3 | -1.417 | -1.376 | -1.408 |
| MVar1 | 0.500 | 0.455 | 0.517 |
| MVar2 | 0.700 | 0.687 | 0.664 |

# Portfolio Credit Risk: Tail Analysis

# Our portfolio framework

- Consider a portfolio with $n$ borrowers.

- For each obligor $i$, the conditional default probability in period $[t, t+1)$ has the form

$$p_{i,t} = F(-\alpha_i + \beta^T X_{i,t})$$

where $F$ is a strictly increasing distribution function.

- The covariates follow a vector autoregressive process

$$X_{i,t+1} = \mathbf{A} X_{i,t} + \tilde{\mathbf{E}}_{i,t+1}$$

# Our portfolio framework

- Consider a portfolio with $n$ borrowers.
- For each obligor $i$, the conditional default probability in period $[t, t+1)$ has the form

$$p_{i,t} = F(-\alpha_i + \beta^T X_{i,t})$$

where $F$ is a strictly increasing distribution function.

- The covariates follow a vector autoregressive process

$$X_{i,t+1} = \mathbf{A} X_{i,t} + \tilde{\mathbf{E}}_{i,t+1}$$

# Our portfolio framework

- Consider a portfolio with $n$ borrowers.
- For each obligor $i$, the conditional default probability in period $[t, t+1)$ has the form

$$p_{i,t} = F(-\alpha_i + \beta^T X_{i,t})$$

where $F$ is a strictly increasing distribution function.

- The covariates follow a vector autoregressive process

$$X_{i,t+1} = \mathbf{A} X_{i,t} + \tilde{\mathbf{E}}_{i,t+1}$$

- Let $D_{i,t}$ denote the event that obligor $i$ defaults at time $t$. Then, loss suffered equals $e_i$. This may be random.

- One illustrative performance measure of interest may be

$$P(\sum_{i=1}^{n} e_i I(D_{i,t_1}) \geq n a_{t_1}, \sum_{i=1}^{n} e_i I(D_{i,t_2}) \geq n a_{t_2})$$

That is, large losses observed jointly in two time periods $t_1$ and $t_2$.

- Let $D_{i,t}$ denote the event that obligor $i$ defaults at time $t$. Then, loss suffered equals $e_i$. This may be random.

- One illustrative performance measure of interest may be

$$P(\sum_{i=1}^{n} e_i I(D_{i,t_1}) \geq n a_{t_1}, \sum_{i=1}^{n} e_i I(D_{i,t_2}) \geq n a_{t_2})$$

That is, large losses observed jointly in two time periods $t_1$ and $t_2$.

# Monte Carlo methodology to compute large loss probabilities

- Start with the value $X_{i,0}$ as well as $p_{i,0}$ for each obligor. Check how many default in period $[0, 1)$.
- Increment the factors generating samples of $\tilde{\mathbf{E}}_1$,

$$X_{i,1} = \mathbf{A}X_{i,0} + \tilde{\mathbf{E}}_{i,1}$$

- Generate samples of $Y_{i,1}$ and compute the conditional probabilities

$$p_{i,1} = F(-\alpha_i + \beta^T X_{i,1})$$

- Generate defaults at time 1. Compute the loss amount at this time.
- Carry on till time $t_2$.

# Monte Carlo methodology to compute large loss probabilities

- Start with the value $X_{i,0}$ as well as $p_{i,0}$ for each obligor. Check how many default in period $[0, 1)$.
- Increment the factors generating samples of $\tilde{\mathbf{E}}_1$,

$$X_{i,1} = \mathbf{A}X_{i,0} + \tilde{\mathbf{E}}_{i,1}$$

- Generate samples of $Y_{i,1}$ and compute the conditional probabilities

$$p_{i,1} = F(-\alpha_i + \beta^T X_{i,1})$$

- Generate defaults at time 1. Compute the loss amount at this time.
- Carry on till time $t_2$.

# Monte Carlo methodology to compute large loss probabilities

- Start with the value $X_{i,0}$ as well as $p_{i,0}$ for each obligor. Check how many default in period $[0, 1)$.
- Increment the factors generating samples of $\tilde{\mathbf{E}}_1$,

$$X_{i,1} = \mathbf{A}X_{i,0} + \tilde{\mathbf{E}}_{i,1}$$

- Generate samples of $Y_{i,1}$ and compute the conditional probabilities

$$p_{i,1} = F(-\alpha_i + \beta^T X_{i,1})$$

- Generate defaults at time 1. Compute the loss amount at this time.
- Carry on till time $t_2$.

# Monte Carlo methodology to compute large loss probabilities

- Start with the value $X_{i,0}$ as well as $p_{i,0}$ for each obligor. Check how many default in period $[0, 1)$.
- Increment the factors generating samples of $\tilde{\mathbf{E}}_1$,

$$X_{i,1} = \mathbf{A}X_{i,0} + \tilde{\mathbf{E}}_{i,1}$$

- Generate samples of $Y_{i,1}$ and compute the conditional probabilities

$$p_{i,1} = F(-\alpha_i + \beta^T X_{i,1})$$

- Generate defaults at time 1. Compute the loss amount at this time.
- Carry on till time $t_2$.

# Monte Carlo methodology to compute large loss probabilities

- Start with the value $X_{i,0}$ as well as $p_{i,0}$ for each obligor. Check how many default in period $[0, 1)$.

- Increment the factors generating samples of $\tilde{\mathbf{E}}_1$,

$$X_{i,1} = \mathbf{A}X_{i,0} + \tilde{\mathbf{E}}_{i,1}$$

- Generate samples of $Y_{i,1}$ and compute the conditional probabilities

$$p_{i,1} = F(-\alpha_i + \beta^T X_{i,1})$$

- Generate defaults at time 1. Compute the loss amount at this time.

- Carry on till time $t_2$.

# Monte Carlo methodology ...

- Compute losses at time $t_1$ and at time $t_2$. Get a sample of

$$I(\sum_{i=1}^{n} e_i I(D_{i,t_1}) \geq na_{t_1}, \sum_{i=1}^{n} e_i I(D_{i,t_2}) \geq na_{t_2})$$

- Average of iid samples gives an unbiased probability estimator.
- Problem is intractable to analysis and due to rare events it is computationally prohibitive.

# Monte Carlo methodology ...

- Compute losses at time $t_1$ and at time $t_2$. Get a sample of

$$I(\sum_{i=1}^{n} e_i I(D_{i,t_1}) \geq na_{t_1}, \sum_{i=1}^{n} e_i I(D_{i,t_2}) \geq na_{t_2})$$

- Average of iid samples gives an unbiased probability estimator.
- Problem is intractable to analysis and due to rare events it is computationally prohibitive.

# Monte Carlo methodology ...

- Compute losses at time $t_1$ and at time $t_2$. Get a sample of

$$I(\sum_{i=1}^{n} e_i I(D_{i,t_1}) \geq na_{t_1}, \sum_{i=1}^{n} e_i I(D_{i,t_2}) \geq na_{t_2})$$

- Average of iid samples gives an unbiased probability estimator.

- Problem is intractable to analysis and due to rare events it is computationally prohibitive.

# Some related literature

- Dembo, Deuschel, Duffie (2004) - single period, single factor large deviations.
- Glasserman and Li (2005), Glasserman, Kang, Shahabuddin (07, 08). Single period, Gaussian Copula, large deviations, fast simulation.
- Bassamboo, Juneja, Zeevi (2008) T-Copula single period large deviations, fast simulation.
- Giesecke, Spiliopoulos, R. Sowers, and J. Sirignano (2015). Continuous time model, analysis relatively complex.
- Duan, Sun, Wang (2012). Discrete time multi period model. No large deviations analysis.

# Tail Analysis of Large Losses

## Embedding the portfolio credit risk problem in asymptotic regime

- Consider a portfolio with $n$ obligors. We analyze this portfolio as $n \to \infty$.

- For each obligor $i$, the conditional default probability in period $[t, t+1)$ has the form

$$p_{i,t}^{(n)} = F(-m_n\alpha + \tilde{m}_n\beta^T X_t + \tilde{m}_n\eta^T Y_{i,t})$$

- The common factors again follow a vector autoregressive process independent of $n$,

$$X_{t+1} = \mathbf{A}X_t + \tilde{\mathbf{E}}_{t+1}$$

# Embedding the portfolio credit risk problem in asymptotic regime

- Consider a portfolio with $n$ obligors. We analyze this portfolio as $n \to \infty$.

- For each obligor $i$, the conditional default probability in period $[t, t+1)$ has the form

$$p_{i,t}^{(n)} = F(-m_n\alpha + \tilde{m}_n\beta^T X_t + \tilde{m}_n\eta^T Y_{i,t})$$

- The common factors again follow a vector autoregressive process independent of $n$,

$$X_{t+1} = \mathbf{A}X_t + \tilde{\mathbf{E}}_{t+1}$$

# Embedding the portfolio credit risk problem in asymptotic regime

- Consider a portfolio with $n$ obligors. We analyze this portfolio as $n \to \infty$.
- For each obligor $i$, the conditional default probability in period $[t, t+1)$ has the form

$$p_{i,t}^{(n)} = F(-m_n\alpha + \tilde{m}_n\beta^T X_t + \tilde{m}_n\eta^T Y_{i,t})$$

- The common factors again follow a vector autoregressive process independent of $n$,

$$X_{t+1} = \mathbf{A}X_t + \tilde{\mathbf{E}}_{t+1}$$

# Illustrative large deviations result

**Theorem**

*When, $\frac{m_n}{\tilde{m}_n} \to \infty$, and $\liminf_n \tilde{m}_n > 0$, under mild conditions,*

$$\lim_{n\to\infty} \frac{1}{n} \log P\left(\sum_{i=1}^{n} e_i I(D_{i,t}) \geq na\right) = -q(t),$$

*where $q(t)$ equals*

$$\frac{\alpha_1^2}{\sum_{k=1}^{t} \sum_{p=1}^{d} h_{t-k,p}^2}.$$

*Note that it strictly reduces with $t$.*

- 
- Similar results when $\tilde{m}_n \to 0$.

# Illustrative large deviations result

## Theorem

*When,* $\frac{m_n}{\tilde{m}_n} \to \infty$*, and* $\liminf_n \tilde{m}_n > 0$*, under mild conditions,*

$$\lim_{n\to\infty} \frac{1}{n} \log P\Big(\sum_{i=1}^{n} e_i I(D_{i,t}) \geq na\Big) = -q(t),$$

*where* $q(t)$ *equals*

$$\frac{\alpha_1^2}{\sum_{k=1}^{t} \sum_{p=1}^{d} h_{t-k,p}^2}.$$

*Note that it strictly reduces with* $t$*.*

- 
- Similar results when $\tilde{m}_n \to 0$.

# Incorporating contagion in tail effects

- Suppose that $\Delta_t$ denotes the random amount of the weighted defaults observed at time $t$.

- For each obligor $i$, the conditional default probability in period $[t, t+1)$ has the form

$$p_{i,t}^{(n)} = F(-\alpha + \tilde{m}_n(\beta + c_1\Delta_t)^T X_t + \tilde{m}_n c_2 \Delta_t Z_t)$$

where $Z_t$ is an independent Gaussian random variable.

- Large $\Delta_t$ increases the sensitivity of default probability to global factors.

- Large $\Delta_t$ impacts the negative sentiment $Z_t$.

# Incorporating contagion in tail effects

- Suppose that $\Delta_t$ denotes the random amount of the weighted defaults observed at time $t$.
- For each obligor $i$, the conditional default probability in period $[t, t+1)$ has the form

$$p_{i,t}^{(n)} = F(-\alpha + \tilde{m}_n(\beta + c_1\Delta_t)^T X_t + \tilde{m}_n c_2 \Delta_t Z_t)$$

where $Z_t$ is an independent Gaussian random variable.

- Large $\Delta_t$ increases the sensitivity of default probability to global factors.
- Large $\Delta_t$ impacts the negative sentiment $Z_t$.

# Incorporating contagion in tail effects

- Suppose that $\Delta_t$ denotes the random amount of the weighted defaults observed at time $t$.
- For each obligor $i$, the conditional default probability in period $[t, t+1)$ has the form

$$p_{i,t}^{(n)} = F(-\alpha + \tilde{m}_n(\beta + c_1\Delta_t)^T X_t + \tilde{m}_n c_2 \Delta_t Z_t)$$

where $Z_t$ is an independent Gaussian random variable.

- Large $\Delta_t$ increases the sensitivity of default probability to global factors.
- Large $\Delta_t$ impacts the negative sentiment $Z_t$.

# Incorporating contagion in tail effects

- Suppose that $\Delta_t$ denotes the random amount of the weighted defaults observed at time $t$.

- For each obligor $i$, the conditional default probability in period $[t, t+1)$ has the form

$$p_{i,t}^{(n)} = F(-\alpha + \tilde{m}_n(\beta + c_1\Delta_t)^T X_t + \tilde{m}_n c_2 \Delta_t Z_t)$$

  where $Z_t$ is an independent Gaussian random variable.

- Large $\Delta_t$ increases the sensitivity of default probability to global factors.

- Large $\Delta_t$ impacts the negative sentiment $Z_t$.

# Incorporating contagion in tail effects

- When $\tilde{m}_n \to 0$ and $n^{1/2}\tilde{m}_n \to \infty$, the above scaling allows both $\beta$ and $c$ to show up in the large deviations tail exponent of large loss probabilities.

- When $\tilde{m}_n = n^{-1/2}$, idiosyncratic defaults also contribute to the large deviations rate.

# Incorporating contagion in tail effects

- When $\tilde{m}_n \to 0$ and $n^{1/2}\tilde{m}_n \to \infty$, the above scaling allows both $\beta$ and $c$ to show up in the large deviations tail exponent of large loss probabilities.
- When $\tilde{m}_n = n^{-1/2}$, idiosyncratic defaults also contribute to the large deviations rate.

# Fast Simulation of Large Losses

# Illustrative rare event simulation problem

- Consider the problem of estimating probability of eighty or more heads in hundred tosses of a fair coin. $(5.58 \times 10^{-10})$.

- Estimator from average of $n$ independent samples

$$\frac{1}{n}\sum_{i=1}^{n} I_i(X_1 + X_2 + \cdots + X_{100} \geq 80).$$

- On average $1.8 \times 10^9$ samples needed to observe a successful sample

- $2.75 \times 10^{12}$ trials needed to get 95% confidence interval of width 5% of the true value.

# Illustrative rare event simulation problem

- Consider the problem of estimating probability of eighty or more heads in hundred tosses of a fair coin. $(5.58 \times 10^{-10})$.

- Estimator from average of $n$ independent samples

$$\frac{1}{n}\sum_{i=1}^{n} I_i(X_1 + X_2 + \cdots + X_{100} \geq 80).$$

- On average $1.8 \times 10^9$ samples needed to observe a successful sample

- $2.75 \times 10^{12}$ trials needed to get 95% confidence interval of width 5% of the true value.

# Illustrative rare event simulation problem

- Consider the problem of estimating probability of eighty or more heads in hundred tosses of a fair coin. $(5.58 \times 10^{-10})$.

- Estimator from average of $n$ independent samples

$$\frac{1}{n}\sum_{i=1}^{n} I_i(X_1 + X_2 + \cdots + X_{100} \geq 80).$$

- On average $1.8 \times 10^9$ samples needed to observe a successful sample

- $2.75 \times 10^{12}$ trials needed to get 95% confidence interval of width 5% of the true value.

# Illustrative rare event simulation problem

- Consider the problem of estimating probability of eighty or more heads in hundred tosses of a fair coin. ($5.58 \times 10^{-10}$).
- Estimator from average of $n$ independent samples

$$\frac{1}{n}\sum_{i=1}^{n} I_i(X_1 + X_2 + \cdots + X_{100} \geq 80).$$

- On average $1.8 \times 10^9$ samples needed to observe a successful sample
- $2.75 \times 10^{12}$ trials needed to get 95% confidence interval of width 5% of the true value.

# Importance sampling to the rescue

- Generate these samples under a new distribution such that $X_i$'s independently equal 1 with probability $p$.

- Unbias the result using the 'likelihood ratio'

$$\frac{1}{n}\sum_{i=1}^{n} I_i(X_1+\cdots+X_{100}\geq 80)\frac{(1/2)^{\sum_{i=1}^{100}X_i}(1/2)^{100-\sum_{i=1}^{100}X_i}}{p^{\sum_{i=1}^{100}X_i}(1-p)^{100-\sum_{i=1}^{100}X_i}}.$$

- When $p=0.8$, 7,932 samples needed for 5% relative accuracy

- When $p=0.99$, $3.69\times 10^{22}$ samples needed for 5% relative accuracy.

# Importance sampling to the rescue

- Generate these samples under a new distribution such that $X_i$'s independently equal 1 with probability $p$.

- Unbias the result using the 'likelihood ratio'

$$\frac{1}{n}\sum_{i=1}^{n} I_i(X_1+\cdots+X_{100}\geq 80)\frac{(1/2)^{\sum_{i=1}^{100}X_i}(1/2)^{100-\sum_{i=1}^{100}X_i}}{p^{\sum_{i=1}^{100}X_i}(1-p)^{100-\sum_{i=1}^{100}X_i}}.$$

- When $p=0.8$, 7,932 samples needed for 5% relative accuracy

- When $p=0.99$, $3.69\times 10^{22}$ samples needed for 5% relative accuracy.

# Importance sampling to the rescue

- Generate these samples under a new distribution such that $X_i$'s independently equal 1 with probability $p$.

- Unbias the result using the 'likelihood ratio'

$$\frac{1}{n}\sum_{i=1}^{n} I_i(X_1+\cdots+X_{100}\geq 80)\frac{(1/2)^{\sum_{i=1}^{100}X_i}(1/2)^{100-\sum_{i=1}^{100}X_i}}{p^{\sum_{i=1}^{100}X_i}(1-p)^{100-\sum_{i=1}^{100}X_i}}.$$

- When $p=0.8$, 7,932 samples needed for 5% relative accuracy

- When $p=0.99$, $3.69\times 10^{22}$ samples needed for 5% relative accuracy.

# Importance sampling to the rescue

- Generate these samples under a new distribution such that $X_i$'s independently equal 1 with probability $p$.

- Unbias the result using the 'likelihood ratio'

$$\frac{1}{n}\sum_{i=1}^{n} I_i(X_1+\cdots+X_{100}\geq 80)\frac{(1/2)^{\sum_{i=1}^{100}X_i}(1/2)^{100-\sum_{i=1}^{100}X_i}}{p^{\sum_{i=1}^{100}X_i}(1-p)^{100-\sum_{i=1}^{100}X_i}}.$$

- When $p=0.8$, 7,932 samples needed for 5% relative accuracy

- When $p=0.99$, $3.69\times 10^{22}$ samples needed for 5% relative accuracy.

# Importance sampling

- Consider estimating the rare event probability $P(A)$.

$$P(A) = EI(A) = \int_{x \in A} f(x)dx = \int_{x \in A} \frac{f(x)}{f^*(x)} f^*(x)dx = E^*[LI(A)]$$

where $L(x) = \frac{f(x)}{f^*(x)}$ is called the likelihood ratio.

- Estimation strategy: Generate independent samples of $L * I(A)$ using $f^*$. Their average is an unbiased and consistent estimator of $P(A)$.

# Importance sampling

- Consider estimating the rare event probability $P(A)$.

$$P(A) = EI(A) = \int_{x \in A} f(x)dx = \int_{x \in A} \frac{f(x)}{f^*(x)} f^*(x)dx = E^*[LI(A)]$$

where $L(x) = \frac{f(x)}{f^*(x)}$ is called the likelihood ratio.

- Estimation strategy: Generate independent samples of $L * I(A)$ using $f^*$. Their average is an unbiased and consistent estimator of $P(A)$.

- Challenge is to find $f^*$ that minimizes the variance or the second moment of the estimator $L * I(A)$.

$$E^* L^2 I(A) = \int_{x \in A} \left( \frac{f(x)}{f^*(x)} \right)^2 f^*(x) dx = \int_{x \in A} \left( \frac{f(x)^2}{f^*(x)} \right) dx$$

- Therefore, whenever $f(x)$ is large, $f^*(x)$ should be large - Should emphasize most likely paths.

- $f^*(x)$ should never be much smaller than $f(x)$ along $A$.

- Challenge is to find $f^*$ that minimizes the variance or the second moment of the estimator $L * I(A)$.

$$E^* L^2 I(A) = \int_{x \in A} \left( \frac{f(x)}{f^*(x)} \right)^2 f^*(x) dx = \int_{x \in A} \left( \frac{f(x)^2}{f^*(x)} \right) dx$$

- Therefore, whenever $f(x)$ is large, $f^*(x)$ should be large - Should emphasize most likely paths.

- $f^*(x)$ should never be much smaller than $f(x)$ along $A$.

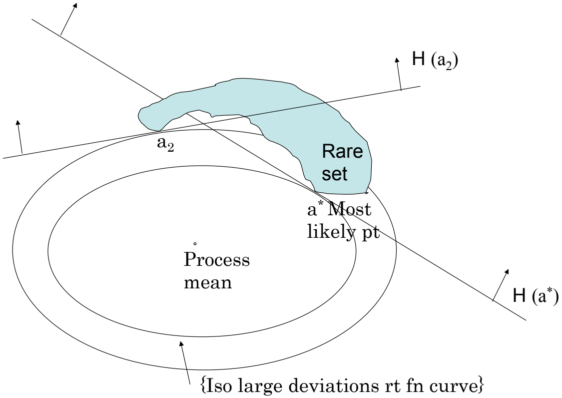
- Challenge is to find $f^*$ that minimizes the variance or the second moment of the estimator $L * I(A)$.

$$E^* L^2 I(A) = \int_{x \in A} \left( \frac{f(x)}{f^*(x)} \right)^2 f^*(x) dx = \int_{x \in A} \left( \frac{f(x)^2}{f^*(x)} \right) dx$$

- Therefore, whenever $f(x)$ is large, $f^*(x)$ should be large - Should emphasize most likely paths.

- $f^*(x)$ should never be much smaller than $f(x)$ along $A$.

# Issues with importance Sampling




- First illustrated by Sadowsky and Bucklew (1991).

# In our problem

- For loss probabilities of order 1 in a 1000, one can expect 100-150 times speed up using an implementable asymptotically optimal importance sampling distribution.

# In our problem

- For loss probabilities of order 1 in a 1000, one can expect 100-150 times speed up using an implementable asymptotically optimal importance sampling distribution.

# Comments and Conclusions

- Using Taylor expansion, same approximations for non-Gaussian light tailed variables, if the corresponding $\beta$ are small.
- Developed closed form expressions for approximations to MLE.
- Conducted asymptotic analysis to prove effectiveness of proposed estimators and empirically verified strong performance relative to existing methods.
- If the underlying model is wrong (the only truth in this talk so far), the exact method and the approximate one are equally bad!
- Developed an asymptotic framework and conducted large deviations methodology for joint distribution of large losses for portfolio credit risk.
- Proposed provably efficient fast simulation techniques for the portfolio model.

# Comments and Conclusions

- Using Taylor expansion, same approximations for non-Gaussian light tailed variables, if the corresponding $\beta$ are small.
- Developed closed form expressions for approximations to MLE.
- Conducted asymptotic analysis to prove effectiveness of proposed estimators and empirically verified strong performance relative to existing methods.
- If the underlying model is wrong (the only truth in this talk so far), the exact method and the approximate one are equally bad!
- Developed an asymptotic framework and conducted large deviations methodology for joint distribution of large losses for portfolio credit risk.
- Proposed provably efficient fast simulation techniques for the portfolio model.

# Comments and Conclusions

- Using Taylor expansion, same approximations for non-Gaussian light tailed variables, if the corresponding $\beta$ are small.
- Developed closed form expressions for approximations to MLE.
- Conducted asymptotic analysis to prove effectiveness of proposed estimators and empirically verified strong performance relative to existing methods.
- If the underlying model is wrong (the only truth in this talk so far), the exact method and the approximate one are equally bad!
- Developed an asymptotic framework and conducted large deviations methodology for joint distribution of large losses for portfolio credit risk.
- Proposed provably efficient fast simulation techniques for the portfolio model.

# Comments and Conclusions

- Using Taylor expansion, same approximations for non-Gaussian light tailed variables, if the corresponding $\beta$ are small.
- Developed closed form expressions for approximations to MLE.
- Conducted asymptotic analysis to prove effectiveness of proposed estimators and empirically verified strong performance relative to existing methods.
- If the underlying model is wrong (the only truth in this talk so far), the exact method and the approximate one are equally bad!
- Developed an asymptotic framework and conducted large deviations methodology for joint distribution of large losses for portfolio credit risk.
- Proposed provably efficient fast simulation techniques for the portfolio model.

# Comments and Conclusions

- Using Taylor expansion, same approximations for non-Gaussian light tailed variables, if the corresponding $\beta$ are small.
- Developed closed form expressions for approximations to MLE.
- Conducted asymptotic analysis to prove effectiveness of proposed estimators and empirically verified strong performance relative to existing methods.
- If the underlying model is wrong (the only truth in this talk so far), the exact method and the approximate one are equally bad!
- Developed an asymptotic framework and conducted large deviations methodology for joint distribution of large losses for portfolio credit risk.
- Proposed provably efficient fast simulation techniques for the portfolio model.

# Comments and Conclusions

- Using Taylor expansion, same approximations for non-Gaussian light tailed variables, if the corresponding $\beta$ are small.
- Developed closed form expressions for approximations to MLE.
- Conducted asymptotic analysis to prove effectiveness of proposed estimators and empirically verified strong performance relative to existing methods.
- If the underlying model is wrong (the only truth in this talk so far), the exact method and the approximate one are equally bad!
- Developed an asymptotic framework and conducted large deviations methodology for joint distribution of large losses for portfolio credit risk.
- Proposed provably efficient fast simulation techniques for the portfolio model.