Fast Relevant Simulation in Finance

Kazuhiro Iwasawa

Problem Formulation

VAR: Definition Given a portfolio value V, and an initial portfolio value v_o. For a given confidnce p, VAR is defined as a real number $v_p - v_o$ satisfying:

$$P(V \leq v_p) = p$$

Problem: Regular MC loses accuracy for rare event simulation:($p \approx 0.01$)

Objective: Compute VAR Probability Accurately.

Solution: **Importance Sampling**
$\int_A g(x)f(x)dx = \int_A g(x)\frac{f(x)}{\pi(x)}\pi(x)dx$ with smaller variance. $f(x)$= Original Density, $\pi(x)$ = Importance Twisted Density.

Importance Sampling

Most likely points in the VAR region

$$\max_{V \leq v} f(x)$$

where

$$f(x) = \frac{1}{(2\pi)^{d/2}|\Sigma_x|^{1/2}} e^{-\frac{1}{2}x^T \cdot \Sigma_x^{-1} \cdot x}$$

Basis of Large Deviation Theory

- Analytic Approximation (GHS, 2000)

- Non-Linear Optimization (Glass, 1999)

Rare events happen in predictable ways.

Solutions Proposed (Motivated by the Large Deviation Theory):

(Method I) Importance Sampling based on non-linear optimization (following Glass(1999)) with better Importance Sampling Twist Density π.

Target: Portfolios with explicitly known density.

(Method II) Importance sampling based on stochastic optimization combined with the Metropolis Hasting $M(RT^2)$ algorithm with a clever choice of π_θ.

Target: High dimensional portfolios without explicitly known density, or containing path-dependent options.

Method I

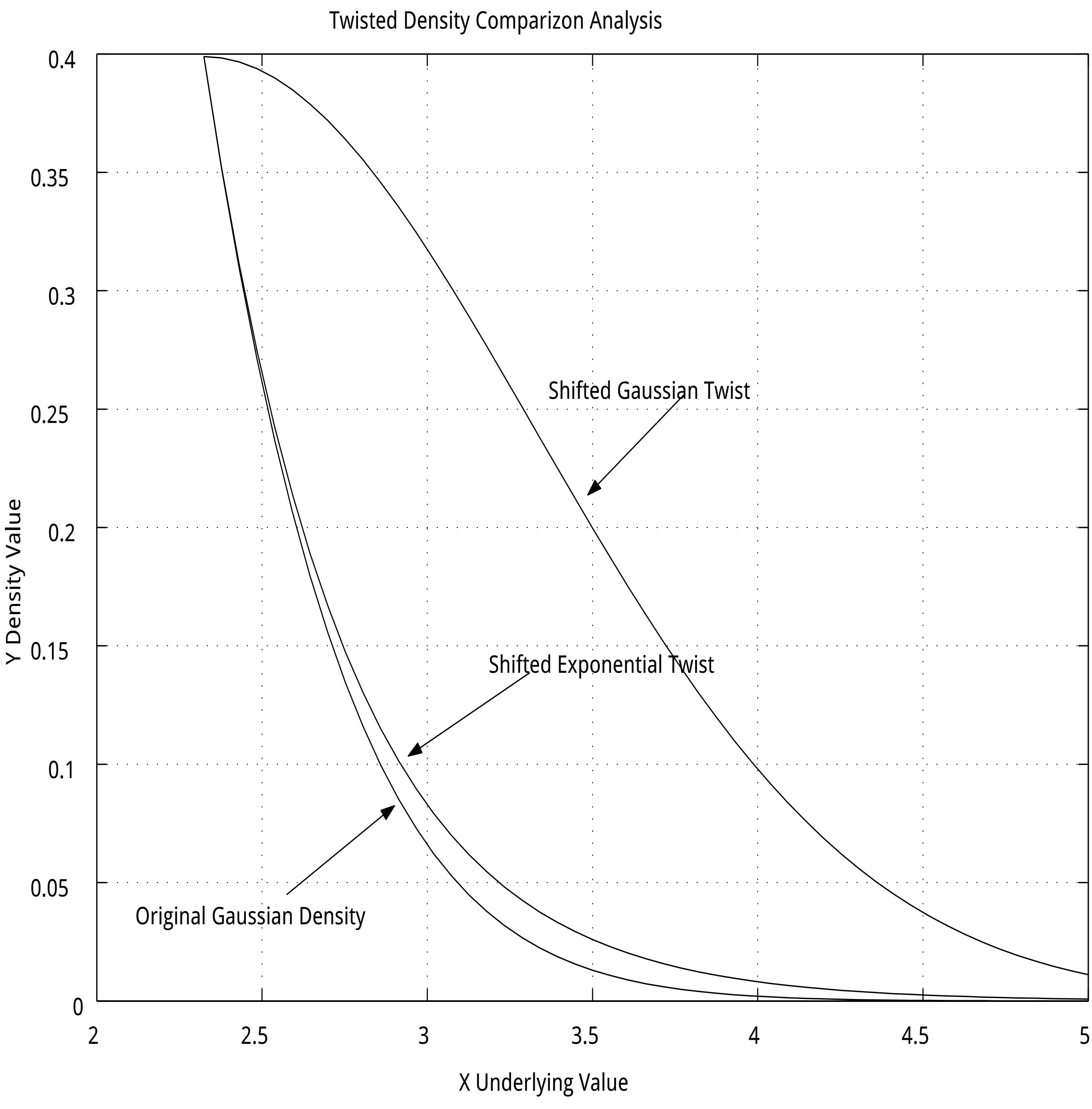
f: Multivariate Correlated Gaussian
(Price: $S_i = C_1 e^{C_2 x_i}$)

β maximizes f ober $B_v \equiv (V \leq v)$.

- $\pi \sim f$ shifted to mean β

- $\pi \sim$ simple exponential normal to ∂B_v

Sampling method: Independent samples from π.


Twisted Density Comparizon Analysis
Y Density Value
X Underlying Value
Shifted Gaussian Twist
Shifted Exponential Twist
Original Gaussian Density
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
2
2.5
3
3.5
4
4.5
5

Exponential Twist For Importance Sampling

Form exponential twist π_e in the direction x_β (x_1 coordinate) perpendicular to the VAR surface $V = v$ at each local minima β, and gaussian density π_g in other directions (x_k coordinates, $k = 2, 3, ..., d$), $\mu = |x_\beta|$.

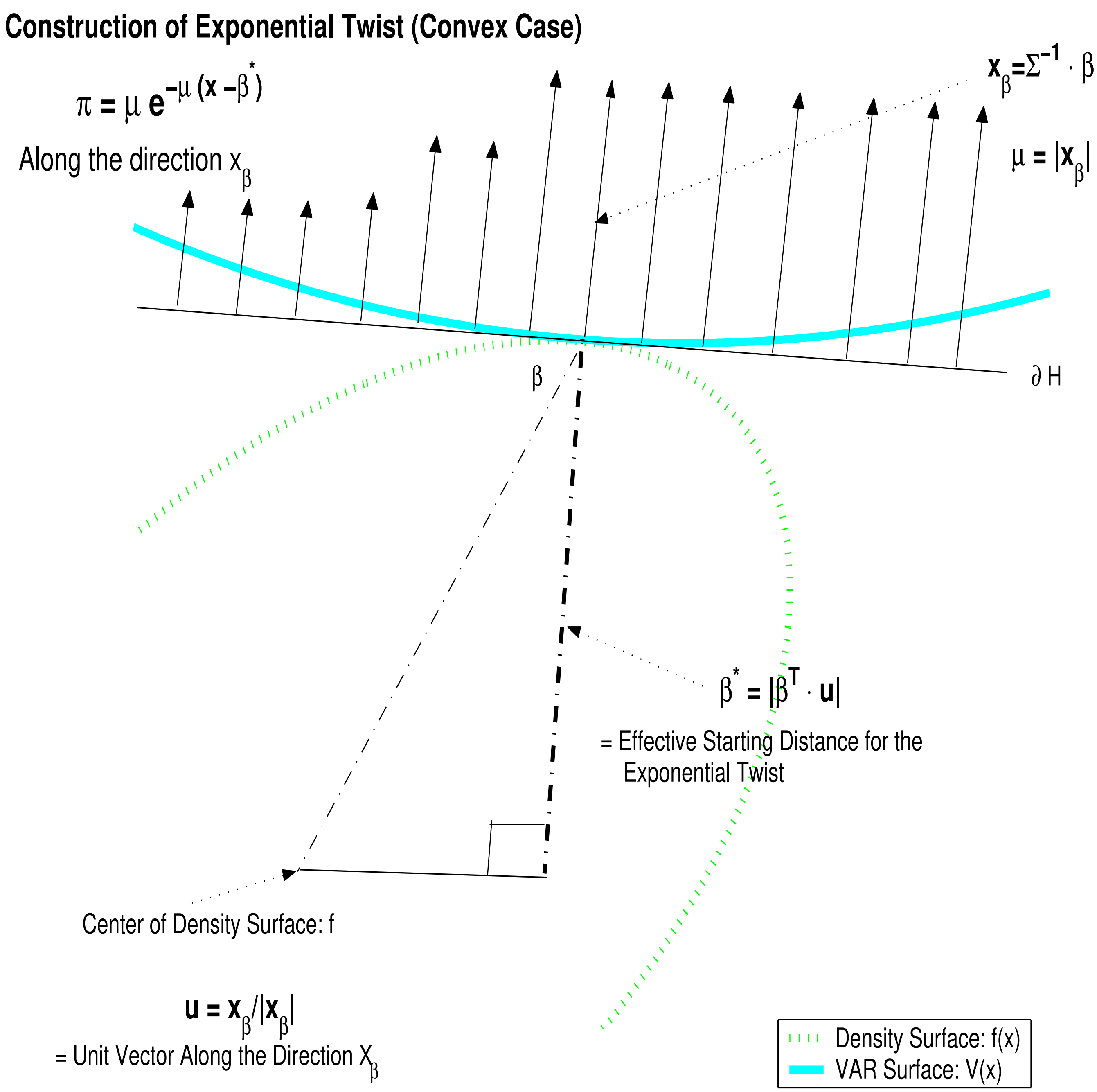
$$\pi_e(x_1) = \mu e^{-\mu(x_1 - \beta^*)}, \;\; x_1 \geq \beta^*$$

$$|\beta^*| = \begin{cases} |\beta^T \cdot u_1| & \text{if } B_v \in H \text{ (i.e.,} B_v \text{ is convex)} \\ |y^T \cdot u_1| & \text{otherwise (i.e.,} B_v \text{ is concave).} \end{cases}$$

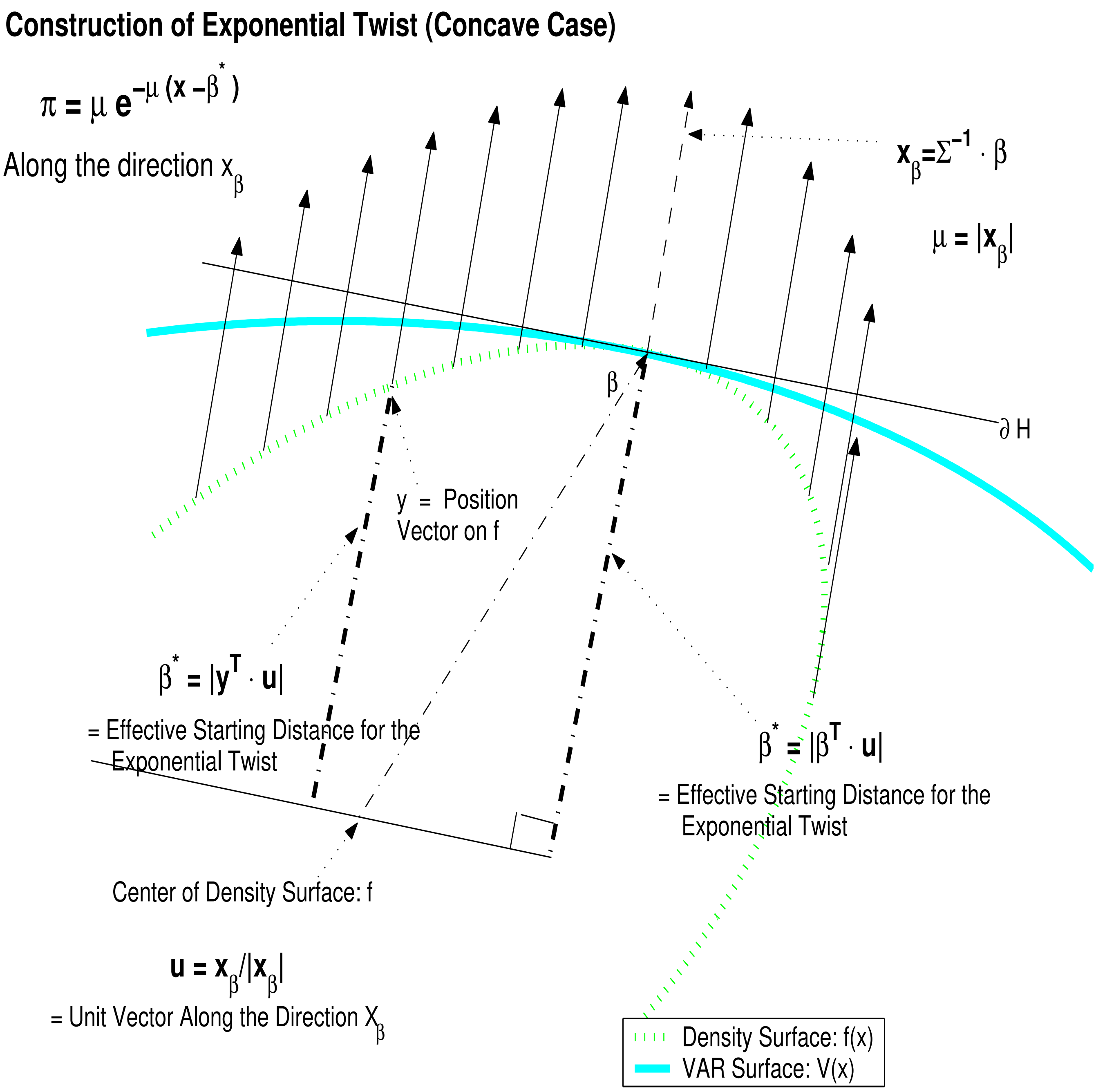
where y is the point on the ellipse: $y^T \cdot \Sigma_x^{-1} \cdot y^T = \beta^T \cdot \beta$.

$$\pi_g(x_k) = \frac{1}{\sqrt{2\pi}\lambda_k} e^{\frac{-x_k^2}{2\lambda_{Min}^2}} \;\; k = 2, 3..., d$$

$$\pi(x) = \prod_{k=2}^{d} \pi_g(x_k) \, \pi_e(x_1)$$



Construction of Exponential Twist (Convex Case)
$\pi = \mu e^{-\mu (x - \beta^*)}$
Along the direction x_β
$x_\beta = \Sigma^{-1} \cdot \beta$
$\mu = |x_\beta|$
β
∂H
$\beta^* = |\beta^T \cdot u|$
= Effective Starting Distance for the Exponential Twist
Center of Density Surface: f
$u = x_\beta / |x_\beta|$
= Unit Vector Along the Direction X_β
Density Surface: f(x)
VAR Surface: V(x)


Construction of Exponential Twist (Concave Case)
$\pi = \mu\, e^{-\mu\,(x-\beta^*)}$
Along the direction x_β
$x_\beta = \Sigma^{-1} \cdot \beta$
$\mu = |x_\beta|$
∂ H
β
y = Position
Vector on f
$\beta^* = |y^T \cdot u|$
= Effective Starting Distance for the
Exponential Twist
$\beta^* = |\beta^T \cdot u|$
= Effective Starting Distance for the
Exponential Twist
Center of Density Surface: f
$u = x_\beta / |x_\beta|$
= Unit Vector Along the Direction X_β
Density Surface: f(x)
VAR Surface: V(x)

Twist for Multiple Local Minima

For β_i with twist density $\pi_{\beta_i}(x)$, $i = 1, 2, ...m$, form the following twist:

$$\pi(x) = \sum_{i=1}^{m} \alpha_i \pi_{\beta_i}(x)$$

where $x \in R^d$, and α_i is a weighting factor such that $\sum_{i=1}^{m} \alpha_i = 1$.

Weight factor selected according to the relative importance of each local minima:

$$\alpha_i = \frac{f(\beta_i)}{\sum_{k=1}^{m} f(\beta_k)}$$

The Laplace Principle (One Dimension Case: Similar for more General Cases: Appendix C)

$$
\begin{aligned}
P(B_v) &= e^{-\frac{v^2}{2}} \frac{1}{\sqrt{2\pi} v} [1 - \frac{1}{v^2} + \frac{3}{v^4}] + o(\frac{1}{v^7}) \\
M_g^2(B_v) &= e^{-v^2} \frac{1}{\sqrt{2\pi} 2v} [1 - \frac{1}{(2v)^2} + \frac{3}{(2v)^4}] + o(\frac{1}{v^7}) \\
M_e^2(B_v) &= e^{-v^2} \frac{1}{2\pi v^2} [1 - \frac{2}{v^2} + \frac{12}{v^4}] + o(\frac{1}{v^7})
\end{aligned}
$$

where $M_g^2 =$ the second moment under the gaussian twist, and $M_e^2 =$ the second moment under the exponential twist.

Based on the Larege Deviation Theory, both Gaussian Twist and Exponential Twist are Aymptotically Efficient. (i.e., $M_g^2(B_v) \sim e^{-v^2}$, and $M_e^2(B_v) \sim e^{-v^2}$ for large v).

The Exponential Twist is More Efficient

The Variances are compued as:

$$
\begin{aligned}
M_g^2 - p^2 &= \frac{e^{-v^2}}{\sqrt{2\pi}v}[\frac{1}{2} - \frac{1}{\sqrt{2\pi}v} + o(\frac{1}{v^2})] \\
M_e^2 - p^2 &= \frac{5e^{-v^2}}{2\pi v^6}[1 + o(\frac{1}{v^8})]
\end{aligned}
$$

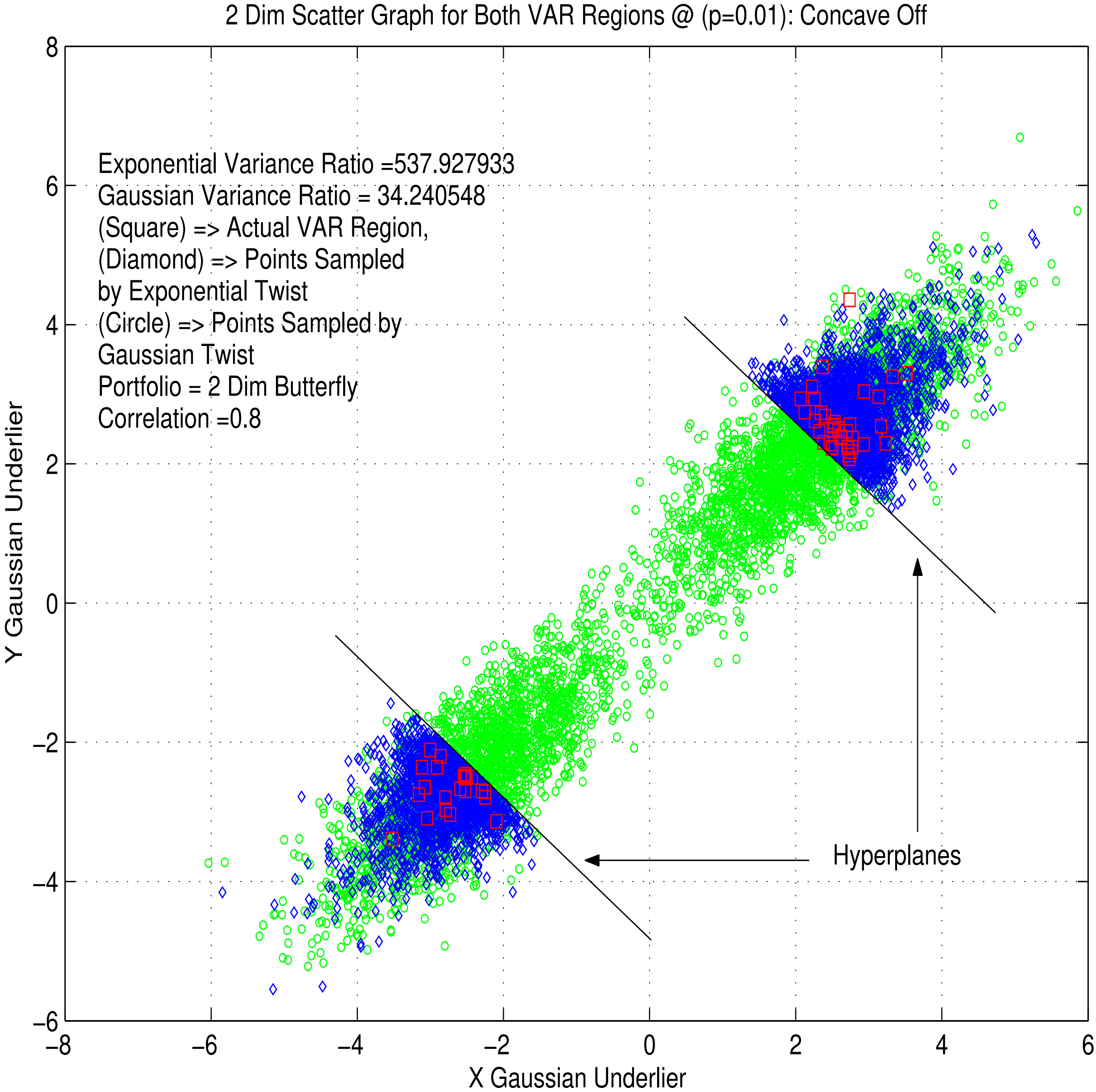
Many Leading Terms Cancel for $M_e^2 - p^2$.

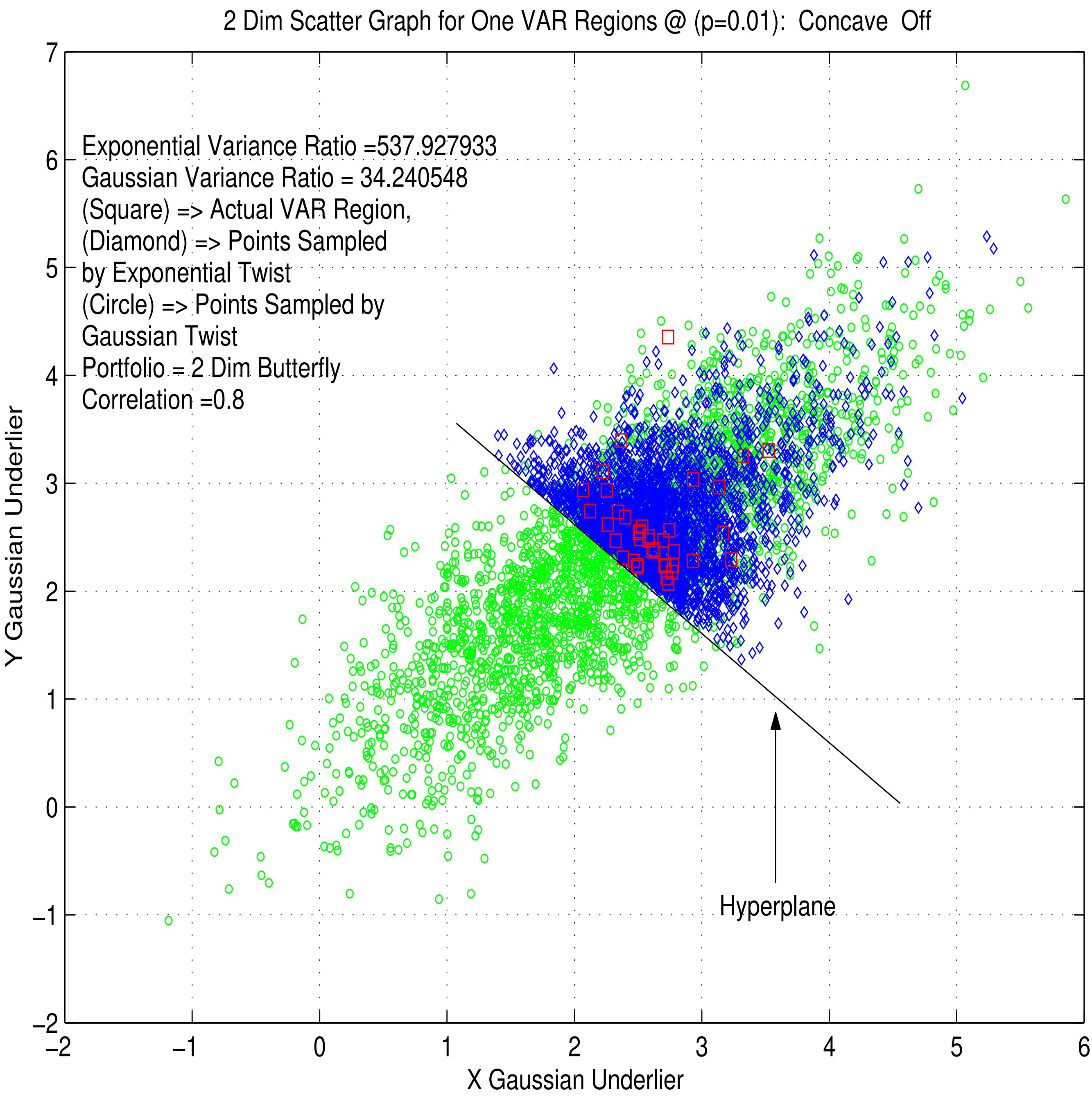
$$
\frac{M_g^2 - p^2}{M_e^2 - p^2} = \frac{\sqrt{2\pi}v^5}{5}[\frac{1}{2} - \frac{1}{\sqrt{2\pi}v} + o(\frac{1}{v^2})]
$$

When $v = 2.326$, $\frac{M_g^2 - p^2}{M_e^2 - p^2} = 11.21$.

As $v \to \infty$, $\frac{M_g^2 - p^2}{M_e^2 - p^2} \sim Cv^5$.

The Exponential Twist is more efficient by a factor of v^5. This is true even for multiple local minima cases (See Next Graphs).


2 Dim Scatter Graph for Both VAR Regions @ (p=0.01): Concave Off
Exponential Variance Ratio =537.927933
Gaussian Variance Ratio = 34.240548
(Square) => Actual VAR Region,
(Diamond) => Points Sampled
by Exponential Twist
(Circle) => Points Sampled by
Gaussian Twist
Portfolio = 2 Dim Butterfly
Correlation =0.8
Hyperplanes
Y Gaussian Underlier
X Gaussian Underlier
8
6
4
2
0
−2
−4
−6
−8
−6
−4
−2
0
2
4
6


2 Dim Scatter Graph for One VAR Regions @ (p=0.01): Concave Off
Exponential Variance Ratio =537.927933
Gaussian Variance Ratio = 34.240548
(Square) => Actual VAR Region,
(Diamond) => Points Sampled
by Exponential Twist
(Circle) => Points Sampled by
Gaussian Twist
Portfolio = 2 Dim Butterfly
Correlation =0.8
Hyperplane
Y Gaussian Underlier
X Gaussian Underlier
7
6
5
4
3
2
1
0
−1
−2
−2
−1
0
1
2
3
4
5
6

Test Results

(1) 10 Short Calls, 5 Short Puts.
(2) 10 Long Calls, 5 Long Puts.
(3) 10 Long Calls, 5 Short Puts.
(4) 10 Short Calls, Puts Delta Hedged *.
(5) 10 DAO Short Calls, 5 Short Puts.
(6) 10 DAO Short Calls, Puts Delta Hedged *.
(7) 10 DAO Short Calls, CON Puts Delta Hedged *.
(8) 10 DAO Short Calls, DAO Puts Delta Hedged *.
(9) 10 DAO Short Calls, 10 Long DAI Calls .
(* involves more than one local minima)

Table 2.1: Variance Ratio Test: GHS Twist vs Gaussian Twist with Optimization

Port.	GHS Ratio	Gaussian Ratio
(1)	3.20013E+01	3.31234E+01
(2)	3.41234E+01	2.83456E+01
(3)	3.31233E+01	3.12223E+01
(4)*	1.75335E+01	2.12234E+01
(5)	1.04843E+01	1.92334E+01
(6)*	9.18233E+00	1.38283E+01
(7)*	3.12344E-01	1.21314E+01
(8)*	7.34232E+00	1.34212E+01
(9)	3.23423E+01	3.81234E+01

(* involves more than one local minima)
Port. = Portfolio
GHS Ratio = Variance ratio by GHS Method (Appendix B)
Gaussian Ratio = Variance ratio by Gaussian Twist with Optimization.

Table 2.2: Variance Ratio Test: Gaussian Twist vs Exponential Twist

Port.	Gaussian Ratio	Exponential Ratio
(1)	3.31234E+01	1.29874E+02
(2)	2.83456E+01	1.31235E+02
(3)	3.12223E+01	4.35323E+02
(4)*	2.12234E+01	4.65234E+01
(5)	1.92334E+01	4.21433E+01
(6)*	1.38283E+01	4.87683E+01
(7)*	1.21314E+01	3.21314E+01
(8)*	1.34212E+01	3.94342E+01
(9)	3.81234E+01	1.12434E+03

(* involves more than one local minima)
Port. = Portfolio
Gaussian Ratio = Variance ratio by Gaussian Twist with Optimization.
Exponential Ratio = Variance ratio by Exponential Twist with Optimization.

Method II

$f(x)$ can be high dimensional.
V(x) may contain path dependent options.
Explicit optimization may be impractical.

Solution:

$$\pi_\theta(x) \sim f(x) \cdot e^{-\theta \cdot \max(V(x), v)}, \quad \theta > 0$$

(1) Sample π_θ using Metropolis $M(RT)^2$ Markov Chain Monte Carlo (MCMC).

(2) Choose θ adaptively so that $V \leq v$ for the half of the samples (i.e., $E_\theta[sign(V - v)] = 0$).

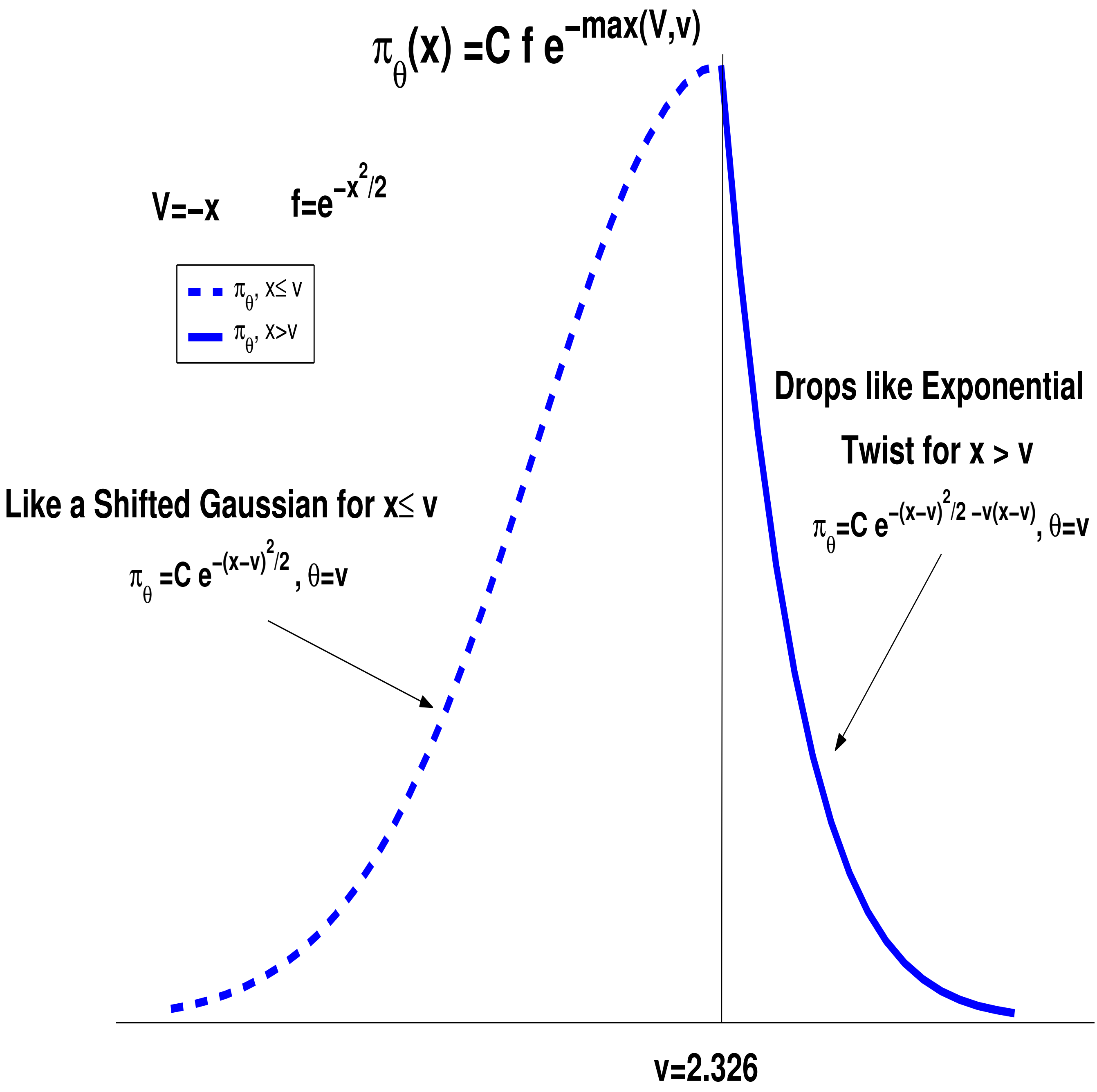
Robbins-Monro algorithm

$$\theta_{n+1} = \theta_n + \frac{1}{n} \cdot (sign(V - v))$$

New Exponential Twist for MCMC

$$\pi_\theta(x) = C(\theta) \cdot f(x) \cdot e^{-\theta \cdot \text{max}(V(x),v)}, \;\; \theta > 0$$

- The twist $f(x) \cdot e^{-\theta V(x)}$ is common.
- $S \sim e^X$ and $V(x) = -S$ gives $\pi \sim e^{-x^2/2+\theta e^x}$, unbounded.
- max$(V(x), v))$ avoids potentially unbounded $\pi_\theta(x)$.


$\pi_\theta(x) = C\, f\, e^{-\max(V,v)}$
$V=-x$
$f=e^{-x^2/2}$
π_θ, $x \le v$
π_θ, $x>v$
Drops like Exponential
Twist for x > v
$\pi_\theta = C\, e^{-(x-v)^2/2 - v(x-v)}$, $\theta=v$
Like a Shifted Gaussian for x≤ v
$\pi_\theta = C\, e^{-(x-v)^2/2}$, $\theta=v$
v=2.326

Metropolis Hasting $M(RT)^2$ Algorithm

(1) **Markov Chain Monte Carlo (MCMC)**

$$P(X_{n+1} \in A | X_n = x) = \int_A T(x|y)dy, \quad A \subset R^d$$

(2) **Rejection Algorithm**

$$\begin{aligned} T(x|y) &= a(x|y)q(y|x) + \\ &\quad \delta_x(y)(1 - \int_A a(x|y)q(y|x)dy) \\ a(x|y) &= 1 \wedge \frac{q(x|y)\pi(y)}{q(y|x)\pi(x)} \end{aligned}$$

$a(x|y) =$ Acceptance Probability of y from x.
$q(y|x) =$ Proposal Density of Sampling y given x.

Metropolis Hasting $M(RT)^2$ Algorithm

Detail Balance

$T(x|y) \cdot \pi(x) = T(y|x) \cdot \pi(y), \quad \forall x, y, \in R^d, x \neq y$

Stationarity

$$\int_A T(y|x)\pi(y)dy = \pi(x) \quad \forall x, y \in A$$

Detail Balance $\rightarrow$ Stationarity

$\int_A T(y|x)\pi(y)dy = \int_A T(x|y)\pi(x)dy$
$= \pi(x) \int_A T(x|y)dy = \pi(x)$

Metropolis Hasting $M(RT)^2$ Algorithm

(1) Given X_n, sample Y_n from some proposal ditribution $q(y|x = X_n)$

(2) Sample ξ from uniform distribution on $[0, 1]$.

(3) Accept Y_n with probability $= \frac{q(X_n|Y_n)\pi(Y_n)}{q(Y_n|X_n)\pi(X_n)}$. That is, if $\xi < \frac{q(X_n|Y_n)\pi(Y_n)}{q(Y_n|X_n)\pi(X_n)}$, then accept the it and let $X_{n+1} = Y_n$.

(4) If step (3) is not satisfied, then reject it and let $X_{n+1} = X_n$.

(5) Repeat the above steps (1)-(4) for the rest of sampling.

Metropolis-within-Gibbs Algorithm

This is a multi-dimensional version of Metropolis Hasting $M(RT)^2$ Algorithm. One dimensional Metropolis algorithm is applied one coordinate at a time.

$$\begin{array}{rcl}
(X_1^1, X_1^2, ..., X_1^d) & \rightarrow & (X_2^1, X_1^2, ..., X_1^d) \\
(X_2^1, X_2^2, ..., X_1^d) & \rightarrow & ..... \\
(X_2^1, X_2^2, ..., X_2^d) & \rightarrow & (X_3^1, X_2^2, ..., X_2^d) \\
...... & \rightarrow & ..... \\
(X_n^1, X_n^2, ..., X_{n-1}^d) & \rightarrow & (X_n^1, X_n^2, ..., X_n^d)
\end{array}$$

(1) Sequential Scan Metropolis-within-Gibbs Algorithm

Each Coordinate is sampled sequentially in a fixed order.

(2) Random Scan Metropolis-within-Gibbs Algorithm

Each Coordinate is sampled in random order.

New Algorithm for VAR Probability Computation

Perform Metropolis-within-Gibbs Algorithm

(1) Accept with Probability $= \frac{q(X_n^k|Y_n^k)\pi(Y_n^k)}{q(Y_n^k|X_n^k)\pi(X_n^k)}$
That is if $\xi < \frac{q(X_n^k|Y_n^k)\pi(Y_n^k)}{q(Y_n^k|X_n^k)\pi(X_n^k)}$, then let $X_{n+1} = (X_n^1, ...Y_n^k, ...X_n^d)$.

(2) If step (1) is not satidfied, then reject it and let $X_{n+1} = (X_n^1, ...X_n^k, ...X_n^d)$.

(3) Update θ_{n+1} by $\theta_{n+1} = \theta_n + \frac{1}{n} \cdot sign(V - v)$. Make sure to record all the values of $\{\theta_n\}$, and $\{V_n\}$, and Go to step (1)

Control Normalization Constant by Sampling from Conditional Probability

(1: More Accurate Near the Center) $\frac{N}{5}$ runs of the Regular Monte Carlo to compute $p_1 = P(V < v_1)$. ($v_1 > v$ is chosen to approximately corresponds to about $p_1 = 0.1$ or even $p_1 = 0.5$ heuristically).

(2: More Accurate Near the Tail) $\frac{4N}{5}$ runs of the Metropolis to compute the conditional probability $p_2 = P(V < v|V < v_1)$.
The Final VAR Probability $= \widehat{p}_1 \widehat{p}_2$.
Effective Error $= \varepsilon_1 \varepsilon_2 + \varepsilon_1 \widehat{p}_2 + \varepsilon_2 \widehat{p}_1$ where ε_1 = the error for the fist run, ε_2 = the error for the second run.

(Bayes Rule)

$$\begin{aligned}
\widehat{p}_1 \widehat{p}_2 &\approx P(V < v_1) P(V < v|V < v_1) \\
&= P(V < v_1) \frac{P(\{V < v\} \cap \{V < v_1\})}{P(V < v_1)} \\
&= P(V < v_1) \frac{P(V < v)}{P(V < v_1)} = P(V < v)
\end{aligned}$$

Theoretical Justification

(1: Lyapunov Condition) : Existence of the Lyapunov function $W(\theta) = e^{(\theta-\theta^*)^{2\cdot m}}$

(2: A-Stability) : θ is contained in a compact set $K \subset R^+$.

(3: Geometric Ergodicity) : For a fixed θ, the chain is Geometric Ergodic (Rosenthal(2001))

(4: Continuity of T_θ with respect to θ)

(5: Slowly Decreasing Condition) : Automatically satisfied by $\sum \frac{1}{n} = \infty$, $\sum \frac{1}{n^2} < \infty$

By Delyon (2000), $\theta_n \to \theta^*$ such that $E_{\theta^*}(sign(V-v) = 0$ almost surely.

Application

VAR for Path Depensent Options : Each path corresponding to different times are considered as an independent variable. Brownian Motion: B_{t-s} is independent of $B_s, t > s$.

$$\begin{aligned} \bar{x} &= \{\bar{x}_1, \bar{x}_2, ..., \bar{x}_d\} \\ &= \{\hat{x}_1\sqrt{t_1}, \hat{x}_2\sqrt{t_2 - t_1}, ...\hat{x}_d\sqrt{t_d - t_{d-1}}\} \end{aligned}$$

$S(t_i) = S(0)e^{(r-\frac{\sigma_i^2}{2})t_i} \prod_{k=1}^{i} e^{\sigma_i \bar{x}_k}$.

$$S(t_i) = S(0)e^{(r-\frac{\sigma_i^2}{2})t_i} \prod_{k=1}^{i-1} e^{\sigma_i \bar{x}_k} \cdot e^{\sigma_i \bar{x}_i} = S(t_{i-1})e^{\sigma_i \bar{x}_i}$$

recursively.

Problem Formulation

E.g. Asian Call

$$E[e^{-rT}(\sum_{i=1}^{d} \frac{S(t_i)}{d} - K)^+]$$

Pricing Basket Option on a Portfolio

Application for Pricing: Far-out-of-the money puts or calls on baskets (For Insuarance purpose to hedge against extreme risk).

Basket Call Price:

$$\frac{\sum_{i=1}^{N}(v-V)^{+}e^{\theta max(V,v)}}{\sum_{i=1}^{N}e^{\theta max(V,v)}}$$

under the new Girsanov measure $\pi_{\theta} = fe^{-\theta max(V,v)}$.

Intuition:

$\pi_{\theta} = C \cdot fe^{-\theta max(V,v)} = \widehat{C} \cdot fe^{-\theta max(V-v,0)}$

The limiting distribution tends to be centered around the set $V = v$.

Approach: To treat V as a random number, not its underlying risk factors. The risk factors are the part of the system, but they are all put in the black box approach as this algorithm implicitly find the optimal point based on the behavior of the random number V.

Experiments

[1] 10 long Asian call with strike = 100, the time grid size equal to 10 equal intervals with final maturity time = 0.05. This results in 10 effective dimensions.
[2] 10 short Asian put with strike = 100. The effective dimension is the same as the above.
[3] 10 long lookback call positions with effective dimension of 10. The gridding was done the same way as the Asian case above.
[4] 10 long lookback call positions with effective dimension of 30. The gridding was done the same way as the Asian case above.
[5] 10 long Americal call option with effective dimension of 5. The exercise times reported by Binomial tree are (0.03, 0.035, 0.04, 0.045, 0.05).
[6] 10 long American call option with effective dimension of 17. The gridding for the binomial tree was increased so that it resulted in

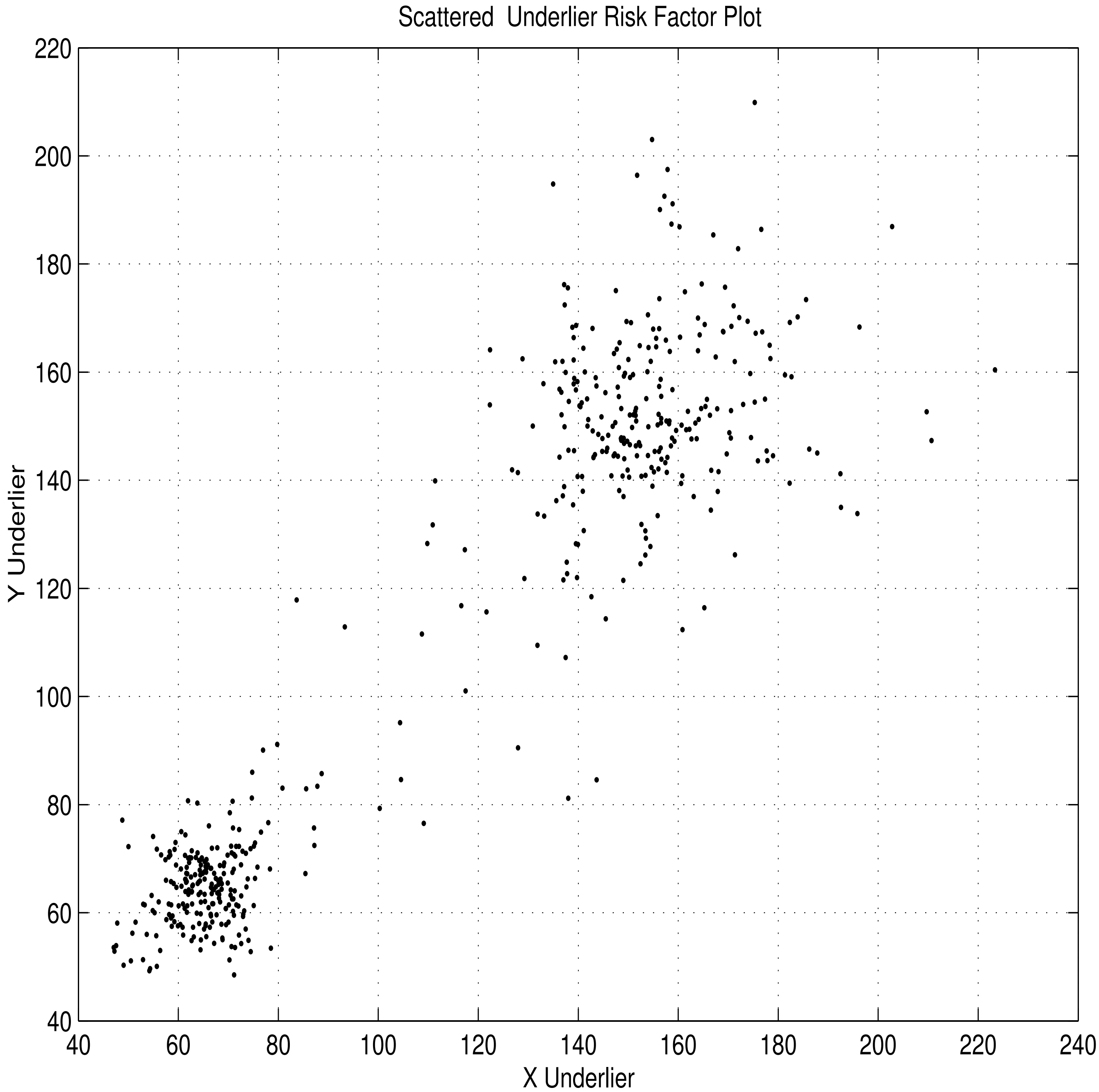
17 instances of exercise times.
[7] Basket option with American option in (6) plus 10 short European calls in 2 dimensions, 5 long European puts in the same dimension as the call, 5 digital call in another 2 dimensions , 6 digital puts in the same dimension as the previous digital call. The blockwide correlation is all set to 0.5. This results in the effective dimension of $17 + 2 + 2 = 21$.
[8] Basket Option with American Option in (6) plus separate butterfly positions in 6 dimensions block correlated with correlation = 0.5: one long call with strike = 90, one long call with strike = 110, and 2 short calls with strike =100.

[MCMC Metropolis-within-Gibbs Algorithm Variance Ratio Test Results]

Probability $\approx$ 1.0E-002;

Portfolio	Variance Ratio	Autocorrelation
1	3.427232E+000	1.544545E+001
2	3.012343E+000	1.685677E+001
3	3.252323E+000	1.532123E+001
4	2.042234E+000	3.114677E+001
5	4.566678E+000	8.123445E+000
6	3.023355E+000	1.912424E+001
7	2.484545E+000	2.342456E+001
8*	2.364311E+000	2.645677E+001

(* is a portfolio with multiple local minima)


Scattered Underlier Risk Factor Plot
Y Underlier
X Underlier
220
200
180
160
140
120
100
80
60
40
40
60
80
100
120
140
160
180
200
220
240

Summary

(I.I.D. Solution) Importance Sampling based on Non-Linear Optimization with Exponential Twist Density π.

Conclusion: Works well for almost all portfolios with explicitly known density as long as all local minima are covered by this twist.

(MCMC Solution) Importance Sampling based on Metropolis MCMC with a clever choice of π_θ where θ is chosen adaptively using the Robbins-Monro algorithm.

Conclusion: Works well for our portfolios.
Issues: Sampling could be trapped in one of the local minima potentially missing many other local minima. The reduction of the autocorrelation time decreases the variance further.