Multilevel Monte Carlo for VaR

Mike Giles, Abdul-Lateef Haji-Ali

Mathematical Institute, University of Oxford

Seminar at Bank of America Merrill Lynch

London, March 19, 2018

Outline

- nested expectation for loss probability, VaR and CVaR
- multilevel Monte Carlo (MLMC) for nested expectations
- prior research on loss probability estimation
  - Gordy & Juneja (2010)
  - Broadie, Du & Moallemi (2011)
- MLMC + uniform inner sampling
- MLMC + adaptive inner sampling
- results for a model test problem
- extensions for real portfolios
  - portfolio sub-sampling
  - SDE approximation

Loss probability, VaR and CVaR

Given some risk-neutral expected loss L conditional on some underlying risk factors Y, the probability of a loss exceeding L_η is

$$\eta = \mathbb{P}[L > L_\eta]$$

VaR is then defined implicitly by specifying η and computing L_η, and CVaR (Expected Shortfall) is defined as

$$\mathbb{E}\left[L \mid L > L_\eta\right].$$

The important point is that the loss is a conditional expectation

$$L - L_\eta \equiv \mathbb{E}[X|Y]$$

so the loss probability is a nested expectation

$$\mathbb{P}[L > L_\eta] = \mathbb{E}\left[H\left(\mathbb{E}[X|Y]\right)\right]$$

where $H(\cdot)$ is the Heaviside step function.

Two-level Monte Carlo

If we want to estimate $\mathbb{E}[P_1]$ but it is much cheaper to simulate $P_0 \approx P_1$, then since

$$\mathbb{E}[P_1] = \mathbb{E}[P_0] + \mathbb{E}[P_1 - P_0]$$

we can use the estimator

$$N_0^{-1} \sum_{n=1}^{N_0} P_0^{(0,n)} \;+\; N_1^{-1} \sum_{n=1}^{N_1} \left(P_1^{(1,n)} - P_0^{(1,n)} \right)$$

Benefit: if $P_1 - P_0$ is small, its variance will be small, so won't need many samples to accurately estimate $\mathbb{E}[P_1 - P_0]$, so cost will be reduced greatly.

Multilevel Monte Carlo

MLMC is based on the telescoping sum

$$\mathbb{E}[P_L] \;=\; \mathbb{E}[P_0] + \sum_{\ell=1}^{L} \mathbb{E}[P_\ell - P_{\ell-1}] \;\equiv\; \sum_{\ell=0}^{L} \mathbb{E}[\Delta P_\ell]$$

where P_ℓ represents an approximation of some output P on level ℓ, and $\Delta P_\ell \equiv P_\ell - P_{\ell-1}$ with $P_{-1} \equiv 0$.

If the weak convergence is

$$\mathbb{E}[P_\ell - P] = O(2^{-\alpha\,\ell}),$$

and Z_ℓ is an unbiased estimator for $\mathbb{E}[P_\ell - P_{\ell-1}]$, with variance

$$\mathbb{V}[Z_\ell] = O(2^{-\beta\,\ell}),$$

and expected cost

$$\mathbb{E}[C_\ell] = O(2^{\gamma\,\ell}), \quad \ldots$$

Multilevel Monte Carlo

... then the finest level L and the number of samples N_ℓ on each level can be chosen to achieve an RMS error of ε at an expected cost

$$C = \begin{cases} O\left(\varepsilon^{-2}\right), & \beta > \gamma, \\ O\left(\varepsilon^{-2}(\log \varepsilon)^2\right), & \beta = \gamma, \\ O\left(\varepsilon^{-2-(\gamma-\beta)/\alpha}\right), & 0 < \beta < \gamma. \end{cases}$$

We always try to get $\beta > \gamma$, so the main cost comes from the coarsest levels – use of QMC can then give substantial additional benefits.

Original research in 2006 used 2^ℓ timesteps to approximate an SDE on level ℓ. Since then it has been used in a variety of applications, including SPDEs and stochastic reaction networks.

Multilevel Monte Carlo

Applying it to nested simulation, suppose we want to estimate

$$\mathbb{E}\left[f\left(\mathbb{E}[X|Y]\right) \right]$$

for some general function f, and on level ℓ we use $M_\ell = 2^\ell$ inner samples to estimate $\mathbb{E}[X|Y]$ so

$$P_\ell \equiv f\left(\overline{X}_\ell\right), \quad \overline{X}_\ell = M_\ell^{-1} \sum_{m=1}^{M_\ell} X^{(m)}$$

with independent samples $X^{(m)}$ conditioned on Y.

The cost for each outer sample is clearly proportional to M_ℓ so $\gamma = 1$.

Multilevel Monte Carlo

A particularly good estimator Z_ℓ for $\mathbb{E}[P_\ell - P_{\ell-1}]$ is

$$Z_\ell = f\left(\overline{X}_\ell\right) - \tfrac{1}{2}\left(f\left(\overline{X}^{(a)}_{\ell-1}\right) + f\left(\overline{X}^{(b)}_{\ell-1}\right)\right)$$

where $\overline{X}^{(a)}_{\ell-1}$, $\overline{X}^{(b)}_{\ell-1}$ each use independent sets of $M_{\ell-1}$ inner samples, while $\overline{X}_\ell$ uses the combined set of M_ℓ samples.

If f is linear, $Z_\ell = 0$, while if f has a bounded second derivative then

$$\overline{X}^{(a/b)}_{\ell-1} - \mathbb{E}[X|Y] = O(2^{-\ell/2}) \quad \Longrightarrow \quad Z_\ell = O(2^{-\ell})$$

so $V_\ell \equiv \mathbb{V}[Z_\ell] = O(2^{-2\ell})$ and hence $\beta = 2$. (Giles, 2015)

Bujok, Hambly, Reisinger (2015) previously analysed a credit derivative with $f(x) \equiv \max(0, x)$, and proved $V_\ell = O(2^{-3\ell/2})$, so $\beta = 3/2$.

In both cases we get $\beta > \gamma$, and hence the optimal $O(\varepsilon^{-2})$ complexity.

However, with VaR we are using the discontinuous Heaviside function.

Prior research on VaR

Gordy & Juneja (2010) considered

$$\mathbb{P}\left[\mathbb{E}[X|Y] > 0\right] \;\equiv\; \mathbb{E}\left[H\left(\mathbb{E}[X|Y]\right)\right]$$

using a uniform sampling approach with N outer samples for Y, and M inner samples to estimate $\mathbb{E}[X|Y]$.

The variance for the inner estimator is $O(M^{-1})$, and they prove this produces a bias in the outer estimate of the same order.

Hence, for ε RMS accuracy require

- $M = O(\varepsilon^{-1})$
- $N = O(\varepsilon^{-2})$

and so the total complexity is $O(M\,N) = O(\varepsilon^{-3})$.

Prior research on VaR

Broadie, Du & Moallemi (2011) improved on Gordy & Juneja by noting that few samples are needed for $H(\mathbb{E}[X \mid Y])$ when $|\,\mathbb{E}[X \mid Y]\,|$ is large.

Their adaptive sampling algorithm used something like

$$M = \min(\varepsilon^{-1}, \varepsilon^{-1/2}\sigma/d)$$

where

$$\sigma^2 \equiv \mathbb{V}[X \mid Y], \qquad d = |\mathbb{E}[X \mid Y]|$$

The cross-over is at $d = O(\varepsilon^{1/2})$ so the average number of inner samples is

$$\overline{M} = O(\varepsilon^{-1/2}),$$

reducing the overall complexity to $O(\overline{M}\, N) = O(\varepsilon^{-5/2})$.

This is better, but still not the $O(\varepsilon^{-2})$ that we aim for.

MLMC + uniform sampling

This is essentially the same MLMC estimator as before except that for the numerical analysis we simplify it to

$$Z_\ell = H\left(\overline{X}_\ell\right) - H\left(\overline{X}_{\ell-1}\right)$$

with completely different inner samples for $\overline{X}_\ell$ and $\overline{X}_{\ell-1}$.

Heuristic analysis:

- with $M_\ell = M_0\, 2^\ell$ we get $\overline{X}_\ell - \mathbb{E}[X|Y] = O(2^{-\ell/2})$
  $\Longrightarrow\ Z_\ell = O(1)$ for only $O(2^{-\ell/2})$ fraction of outer samples
  $\Longrightarrow\ V_\ell \equiv \mathbb{V}[Z_\ell] = O(2^{-\ell/2})$
- this gives $\alpha\!=\!1, \beta\!=\!1/2, \gamma\!=\!1$ so complexity is $O(\varepsilon^{-5/2})$

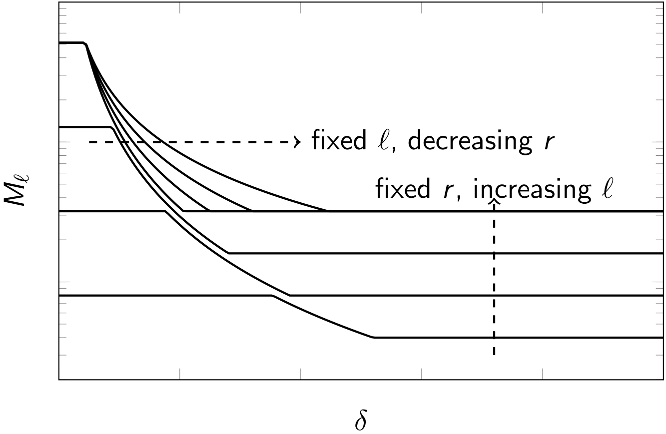
This can be made rigorous provided $\mathbb{E}[X^2] < \infty$ and $\delta \equiv d/\sigma$ has a bounded density near 0. The proof uses Chebyshev's inequality $\mathbb{P}[\,|\overline{X}_\ell - \mathbb{E}[X|Y]| > d\,] \ < \ (\sigma^2/M)/d^2 \ = \ M^{-1}\delta^{-2}$

MLMC + adaptive sampling

Instead of using $M_\ell = M_0\, 2^\ell$ inner samples, we instead want to use

$$M_\ell = M_0\, 4^\ell \max\left(2^{-\ell}, \min(1, (C^{-1} M_0^{1/2} 2^\ell \delta)^{-r})\right), \quad 1 < r < 2,$$

with a minimum of $M_0\, 2^\ell$ and maximum of $M_0\, 4^\ell$.



MLMC + adaptive sampling

Heuristic analysis:

- $r > 1$ ensures the intermediate region is small enough that the average number of inner samples remains $O(2^\ell)$
- $r < 2$ ensures high probability of correct value for $H(\overline{X})$ in intermediate region
- $M_0\, 4^\ell$ in core region ensures $O(2^{-\ell})$ error in computed means, so $Z_\ell = O(1)$ for $O(2^{-\ell})$ fraction of outer samples
- hence, $V_\ell = \mathbb{V}[Z_\ell] = O(2^{-\ell})$ so now $\beta = 1$ and the overall complexity is $O(\varepsilon^{-2} |\log \varepsilon|^2)$

The heuristic analysis is fundamentally correct, but the rigorous analysis for the real algorithm took another year, and the upper bound on r had to be tightened.

Numerical analysis

Big problem: in practice we don't know $\delta = d/\sigma$, so the real adaptive algorithm has to use MC estimates, like Broadie, Du & Moallemi (2011).

Solution: for a given outer sample Y keep doubling the number of inner samples until M_ℓ is big enough based on current estimate $\widehat{\delta} = \widehat{d}/\widehat{\sigma}$, or it reaches the maximum.

Concerns:

- if we use too many samples, the cost may be bigger than we want
- if we use too few samples, the variance may be bigger than we want

The main thrust of the analysis is to prove that the probability of ending up with the "wrong" number of inner samples decays very rapidly as you move away from the "right" number.

Key building block: I

Lemma

Let $\overline{Z}_N$ be an average of N i.i.d. samples of a random variable Z with zero mean and finite q^{th} moment, for $q > 2$. Then there exists a constant C_q depending only on q such that

$$\mathbb{E}[\,|\overline{Z}_N|^q] \leq C_q\, N^{-q/2}\, \mathbb{E}[\,|Z|^q],$$

and for any $z > 0$

$$\mathbb{P}\left[\,|\overline{Z}_N| > z\,\right] \leq \min\left(1, C_q z^{-q} N^{-q/2}\, \mathbb{E}[\,|Z|^q]\right).$$

Proof.

Use Burkholder-Davis-Gundy inequality for the first result, followed by Markov inequality for the second. □

Key building block: II

Corollary

Under the same conditions, if $\sigma^2 = \mathbb{V}[Z]$ *and*

$$\widehat{\sigma}_N^2 = \frac{1}{N}\sum_{n=1}^{N}\left(Z^{(n)} - \overline{Z}_N\right)^2 = \sigma^2 + \frac{1}{N}\sum_{n=1}^{N}\left((Z^{(n)})^2 - \sigma^2\right) \ - \ (\overline{Z}_N)^2,$$

then for any $c_1\!>\!0$ *there exists a* $c_2\!>\!0$ *such that*

$$\mathbb{P}\left[\, |\widehat{\sigma}_N^2 \!-\! \sigma^2| \!>\! c_1\, \sigma^2 \,\right] \leq \min\left(1, c_2 N^{-q/4}\, \mathbb{E}[\,|Z|^q]/\sigma^q\right)$$

Proof.

Again Burkholder-Davis-Gundy and Markov inequalities. □

The Lemma bounds the probability of large errors in $\widehat{d}$, and the Corollary bounds the probability of large errors in $\widehat{\sigma}$.

Final theorem

Theorem

Provided

1. *the random variable* $\delta = d/\sigma$ *has bounded density near 0*
2. *there exists* $q > 2$ *such that*

$$\kappa_q = \sup_y \left\{ \sigma^{-q}\, \mathbb{E}\left[\, |X - \mathbb{E}[X|Y]|^q \,\right] | \, Y = y \right\} < \infty$$

3. *the exponent* r *in the adaptive algorithm satisfies the condition*

$$1 < r < 2 - \left(\sqrt{4q+1} - 1\right) \Big/ q$$

then the adaptive algorithm has expected computational cost $O(2^\ell)$ *per outer sample, and* $V_\ell \equiv \mathbb{V}[Z_\ell] = O(2^{-\ell})$.

Model problem

Taking $\tau \ll 1$ to be a short risk horizon, we consider a loss conditional on $Y \sim N(0,1)$ defined by

$$L \;=\; \mathbb{E}\left[-(\tau^{1/2}\widetilde{Y}_1 + (1-\tau)^{1/2}\widetilde{Y}_2)^2\right] - \mathbb{E}\left[-(\tau^{1/2}Y + (1-\tau)^{1/2}\widetilde{Y}_2)^2 \mid Y\right]$$

where $\widetilde{Y}_1$ and $\widetilde{Y}_2$ are also independent $N(0,1)$ r.v.'s.

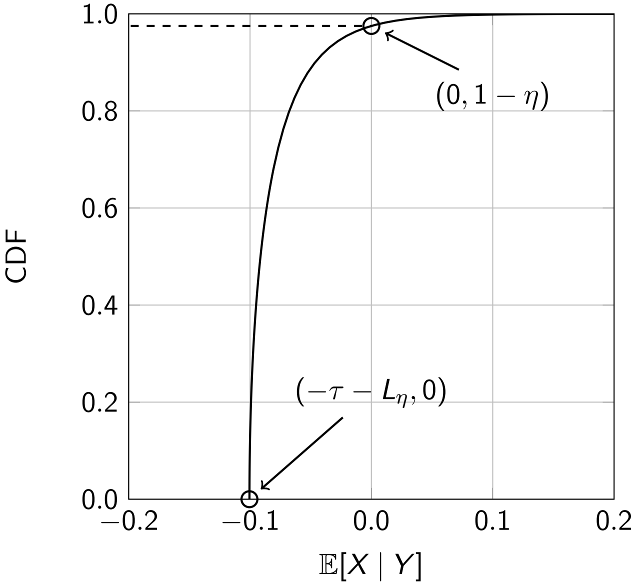
This is intended to model a delta-hedged portfolio with negative Gamma.

A good low-variance definition for X is

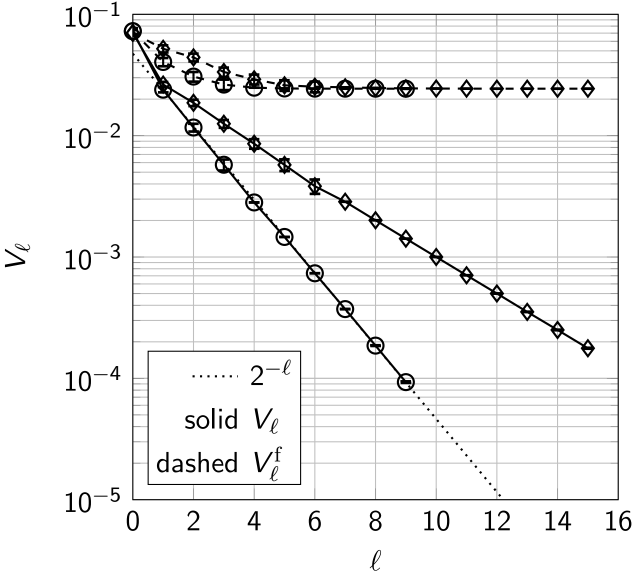
$$\begin{aligned} X &= (\tau^{1/2}Y + (1-\tau)^{1/2}\widetilde{Y}_2)^2 \\ &- \tfrac{1}{2}\left((\tau^{1/2}\widetilde{Y}_1 + (1-\tau)^{1/2}\widetilde{Y}_2)^2 + (-\tau^{1/2}\widetilde{Y}_1 + (1-\tau)^{1/2}\widetilde{Y}_2)^2\right) - L_\eta \end{aligned}$$

The figures include both non-adaptive results, and adaptive using $r = 1.5$.

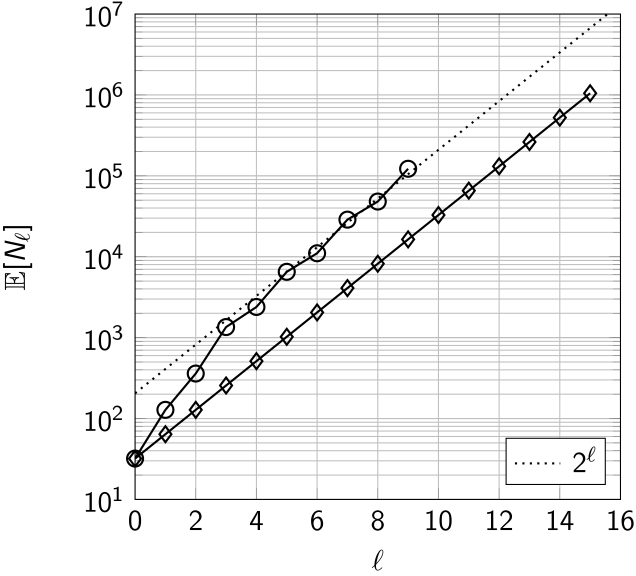
Model problem: $\eta = 0.025$



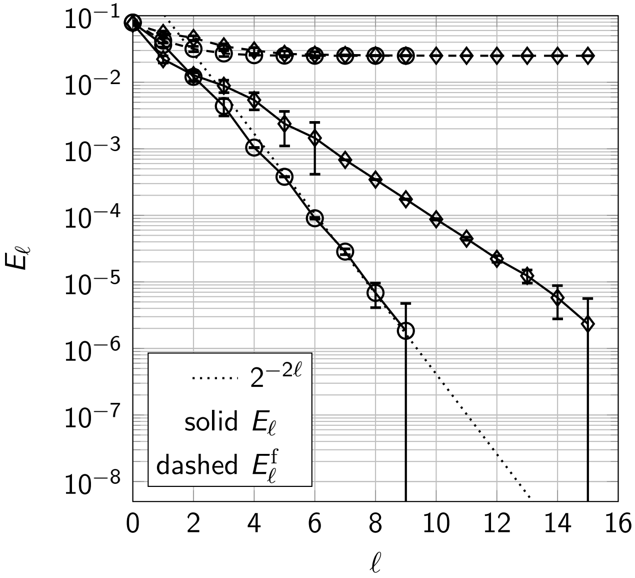
Numerical Results



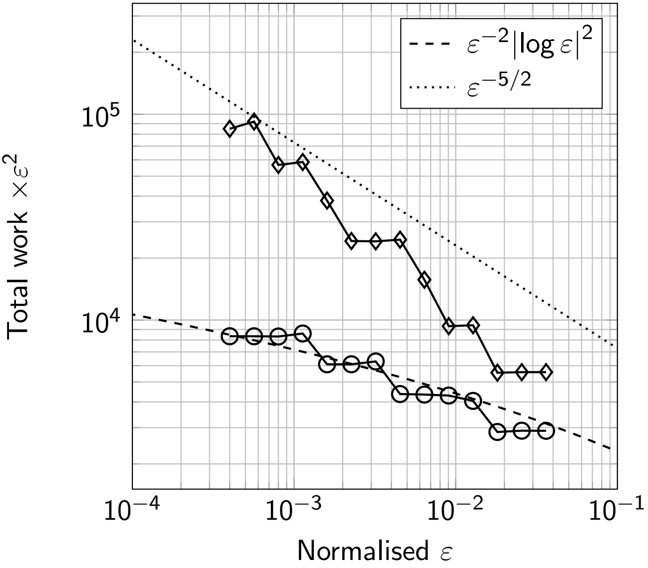
Numerical Results



Numerical Results



Numerical Results



VaR and CVaR

VaR loss L_η is defined implicitly by $\mathbb{P}[L > L_\eta] = \eta$.

This can be estimated by a stochastic root-finding algorithm, with the acceptable error ε being steadily reduced during the iteration.

Given an estimate $\widetilde{L}_\eta$, Rockafellar & Uryasev (2000) show that CVaR is

$$\begin{aligned}
\mathbb{E}[\, L \,|\, L > L_\eta] &= L_\eta + \eta^{-1}\mathbb{E}[\max(0, L - L_\eta)] \\
&= \min_x \left\{ x + \eta^{-1}\mathbb{E}[\max(0, L - x)] \right\} \\
&= \widetilde{L}_\eta + \eta^{-1}\mathbb{E}[\max(0, L - \widetilde{L}_\eta)] + O(\widetilde{L}_\eta - L_\eta)^2
\end{aligned}$$

For ε RMS error, first estimate $\widetilde{L}_\eta$ to accuracy $O(\varepsilon^{1/2})$ at cost $o(\varepsilon^{-2})$.

Then estimate $\eta^{-1}\mathbb{E}[\max(0, L - \widetilde{L}_\eta)]$ to accuracy ε using MLMC + uniform sampling. We prove that $V_\ell = O(2^{-3\ell/2})$ so complexity is $O(\varepsilon^{-2})$.

Real portfolios

In a real delta-hedged portfolio with K products, X can be expressed as a sum of delta-hedged products

$$X = \sum_{k=1}^{K} X_k$$

We now have 3 choices:

1. simulate all underlying assets to compute all X_k, and hence X
2. independently simulate underlying assets needed for each X_k, then sum to get X — more costly but lower variance (Gordy & Juneja)
3. replace X by X_k/p_k with k selected randomly with probability p_k so

$$\mathbb{E}[X_k/p_k] \;=\; \sum_{k'=1}^{K} \mathbb{P}[k\!=\!k']\,\mathbb{E}[X_{k'}/p_{k'}] \;=\; \sum_{k'=1}^{K} \mathbb{E}[X_{k'}] \;=\; \mathbb{E}[X]$$

This makes the overall cost independent of the number of products K.

Real portfolios

We have implemented this for a synthetic portfolio with a delta-neutral mix of Black-Scholes put and call options with various strikes and maturities, all short to give negative Gamma.

The nominal values of each were similar so uniform sub-sampling was used. In future work, we'll make p_k dependent on the nominal values or individual Gammas.

We've also incorporated an MLMC treatment of time discretisation of underlying SDEs.

Future work will also address the fact that options are priced in different ways with different costs (closed form, semi-analytic, MC, finite differences)

Ideally, we would like to collaborate with banks on this work, so that our research is relevant to industry challenges / needs.

Conclusions

- VaR/CVaR is a great new application area for MLMC
- so far, banks haven't been very interested in MLMC, perhaps because the savings have been relatively modest in practice
  – with VaR/CVaR I think the savings may be quite large
- two keys to performance:
  - MLMC approach with more inner samples on "finer" levels
  - adaptive number of inner samples (for VaR)
- sub-sampling the portfolio will offer significant additional savings for large portfolios

Webpages:

`http://people.maths.ox.ac.uk/gilesm/mlmc.html`
`http://people.maths.ox.ac.uk/gilesm/slides.html`
`http://people.maths.ox.ac.uk/gilesm/mlmc_community.html`

References

- R.T. Rockafellar & S. Uryasev. "Optimization of conditional value-at-risk." *Journal of Risk*, 2:21-42, 2000
- M.B. Giles. "Multilevel Monte Carlo path simulation." *Operations Research*, 56(3):607-617, 2008.
- M.B. Gordy and S. Juneja. "Nested simulation in portfolio risk measurement." *Management Science*, 56(10):1833-1848, 2010.
- M. Broadie, Y. Du and C.C. Moallemi. "Efficient risk estimation via nested sequential simulation." *Management Science*, 57(6):1172-1194, 2011.
- K. Bujok, B.M. Hambly and C. Reisinger. "Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives." *Meth. Comp. Applied Prob.*, 17:579-604, 2015.
- M.B. Giles. "Multilevel Monte Carlo methods." *Acta Numerica*, 24, 2015.
- W. Gou. "Estimating Value-at-Risk using multilevel Monte Carlo Maximum Entropy method". Oxford University MSc thesis, 2016.
- M.B. Giles and A. Haji-Ali. "Multilevel nested simulation for efficient risk estimation." arXiv:1802.05016, 2018.