Multilevel Monte Carlo Simulation

Mike Giles

mike.giles@maths.ox.ac.uk

Oxford University Mathematical Institute

Oxford-Man Institute of Quantitative Finance

Workshop on Computational Finance

Kyoto, August 10 – 12, 2009

Generic Problem

Stochastic differential equation with general drift and volatility terms:

$$\mathrm{d}S(t) = a(S,t)\,\mathrm{d}t + b(S,t)\,\mathrm{d}W(t)$$

For simple European options, we want to estimate the expected value of an option dependent on the terminal state

$$P = f(S(T))$$

with a uniform Lipschitz bound,

$$|f(U) - f(V)| \ \leq c\ \|U - V\|\,, \quad \forall\ U,V.$$

Standard MC Approach

Euler discretisation with timestep h:

$$\widehat{S}_{n+1} = \widehat{S}_n + a(\widehat{S}_n, t_n)\, h + b(\widehat{S}_n, t_n)\, \Delta W_n$$

Simplest estimator for expected payoff is an average of N independent path simulations:

$$\widehat{Y} = N^{-1} \sum_{i=1}^{N} f(\widehat{S}^{(i)}_{T/h})$$

- weak convergence – $O(h)$ error in expected payoff
- strong convergence – $O(h^{1/2})$ error in individual paths

Standard MC Approach

Mean Square Error is $O\left(N^{-1}+h^2\right)$

- first term comes from variance of estimator
- second term comes from bias due to weak convergence

To make this $O(\varepsilon^2)$ requires

$$N = O(\varepsilon^{-2}), \quad h = O(\varepsilon) \quad \Longrightarrow \quad \text{cost} = O(N\,h^{-1}) = O(\varepsilon^{-3})$$

Aim is to improve this cost to $O\left(\varepsilon^{-2}(\log\varepsilon)^2\right)$, by combining simulations with different numbers of timesteps – same accuracy as finest calculations, but at a much lower computational cost.

Other work

- Many variance reduction techniques to greatly reduce the cost, but without changing the order
- Richardson extrapolation improves the weak convergence and hence the order
- Multilevel method is a generalisation of two-level control variate method of Kebaier (2005), and similar to ideas of Speight (2009)
- Also related to multilevel parametric integration by Heinrich (2001)

Multilevel MC Approach

Consider multiple sets of simulations with different timesteps $h_l = 2^{-l}\,T,\ l = 0, 1, \ldots, L$, and payoff $\widehat{P}_l$

$$\mathbb{E}[\widehat{P}_L] = \mathbb{E}[\widehat{P}_0] + \sum_{l=1}^{L} \mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}]$$

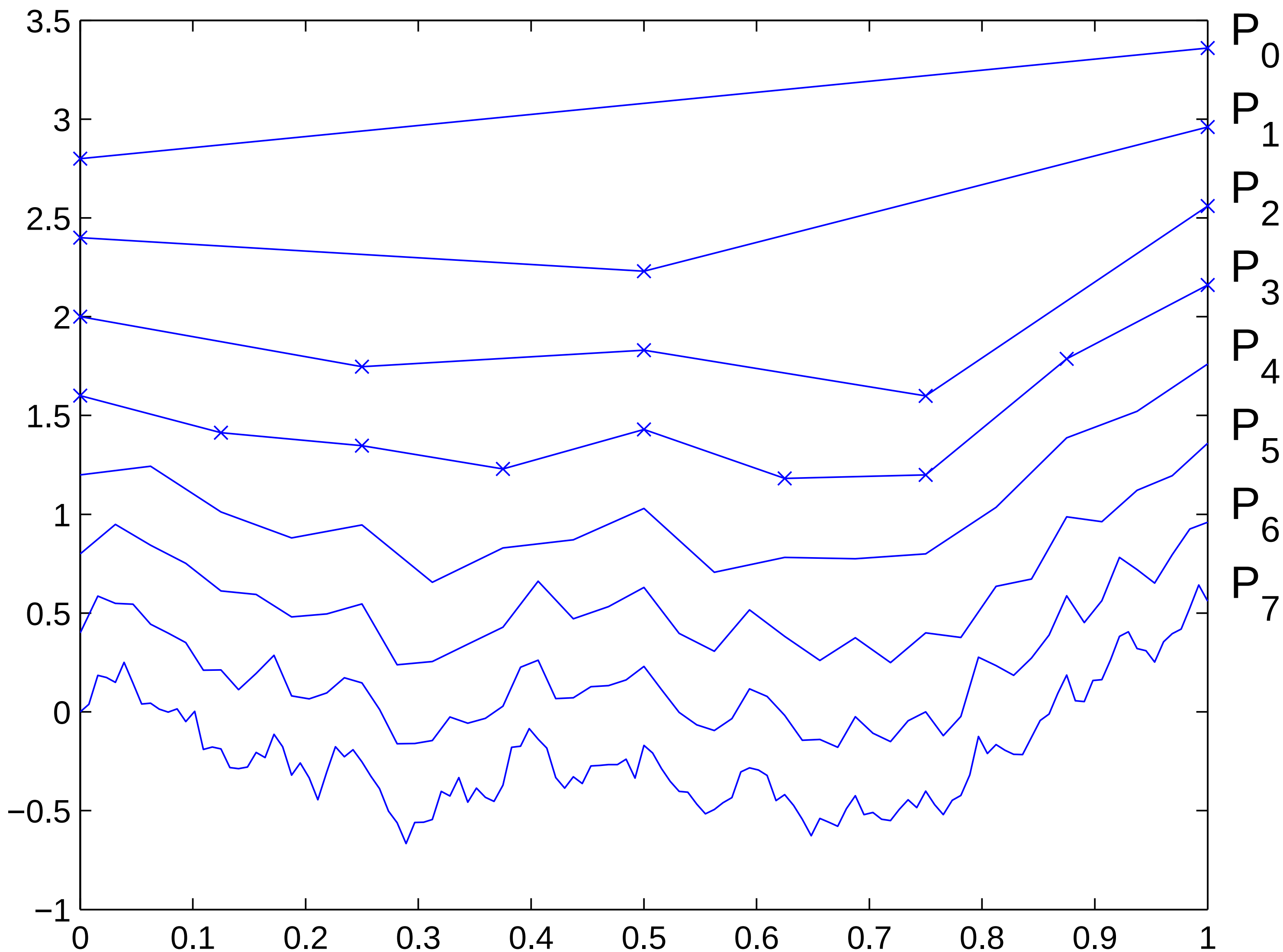
Expected value is same – aim is to reduce variance of estimator for a fixed computational cost.

Key point: approximate $\mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}]$ using N_l simulations with $\widehat{P}_l$ and $\widehat{P}_{l-1}$ obtained using <u>same</u> Brownian path.

$$\widehat{Y}_l = N_l^{-1} \sum_{i=1}^{N_l} \left(\widehat{P}_l^{(i)} - \widehat{P}_{l-1}^{(i)} \right)$$

Multilevel MC Approach

Discrete Brownian path at different levels



Multilevel MC Approach

- each level adds more detail to Brownian path and reduces the error in the numerical integration
- $\mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}]$ reflects impact of that extra detail on the payoff
- different timescales handled by different levels – similar to different wavelengths being handled by different grids in multigrid solvers for iterative solution of PDEs

Multilevel MC Approach

Using independent paths for each level, the variance of the combined estimator is

$$\mathbb{V}\left[\sum_{l=0}^{L} \widehat{Y}_l\right] = \sum_{l=0}^{L} N_l^{-1} V_l, \qquad V_l \equiv \mathbb{V}[\widehat{P}_l - \widehat{P}_{l-1}],$$

and the computational cost is proportional to $\displaystyle\sum_{l=0}^{L} N_l\, h_l^{-1}$.

Hence, the variance is minimised for a fixed computational cost by choosing N_l to be proportional to $\sqrt{V_l\, h_l}$.

The constant of proportionality can be chosen so that the combined variance is $O(\varepsilon^2)$.

Multilevel MC Approach

For the Euler discretisation and the Lipschitz payoff function

$$\mathbb{V}[\widehat{P}_l - P] = O(h_l) \quad \Longrightarrow \quad \mathbb{V}[\widehat{P}_l - \widehat{P}_{l-1}] = O(h_l)$$

and the optimal N_l is asymptotically proportional to h_l.

To make the combined variance $O(\varepsilon^2)$ requires

$$N_l = O(\varepsilon^{-2} L\, h_l).$$

To make the bias $O(\varepsilon)$ requires

$$L = \log_2 \varepsilon^{-1} + O(1) \quad \Longrightarrow \quad h_L = O(\varepsilon).$$

Hence, we obtain an $O(\varepsilon^2)$ MSE for a computational cost which is $O(\varepsilon^{-2} L^2) = O(\varepsilon^{-2} (\log \varepsilon)^2)$.

Results

Geometric Brownian motion:

$$\mathrm{d}S = r\,S\,\mathrm{d}t + \sigma\,S\,\mathrm{d}W, \qquad 0 < t < T,$$

$T\!=\!1,\ \ S(0)\!=\!100,\ \ r\!=\!0.05,\ \ \sigma\!=\!0.2$

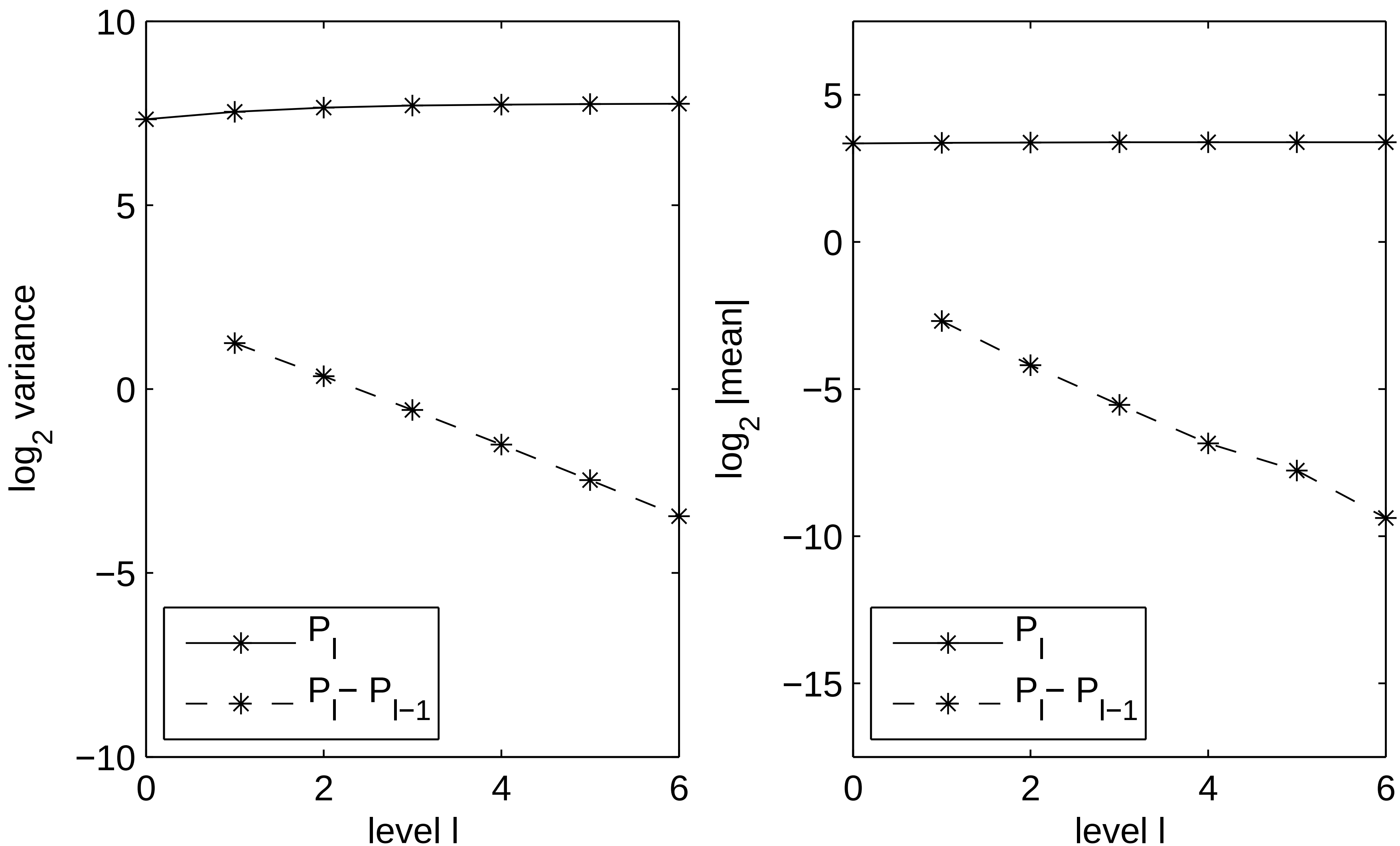
European call option with discounted payoff

$$\exp(-rT)\ \max(S(T)\!-\!K, 0)$$

with strike $K\!=\!100$.

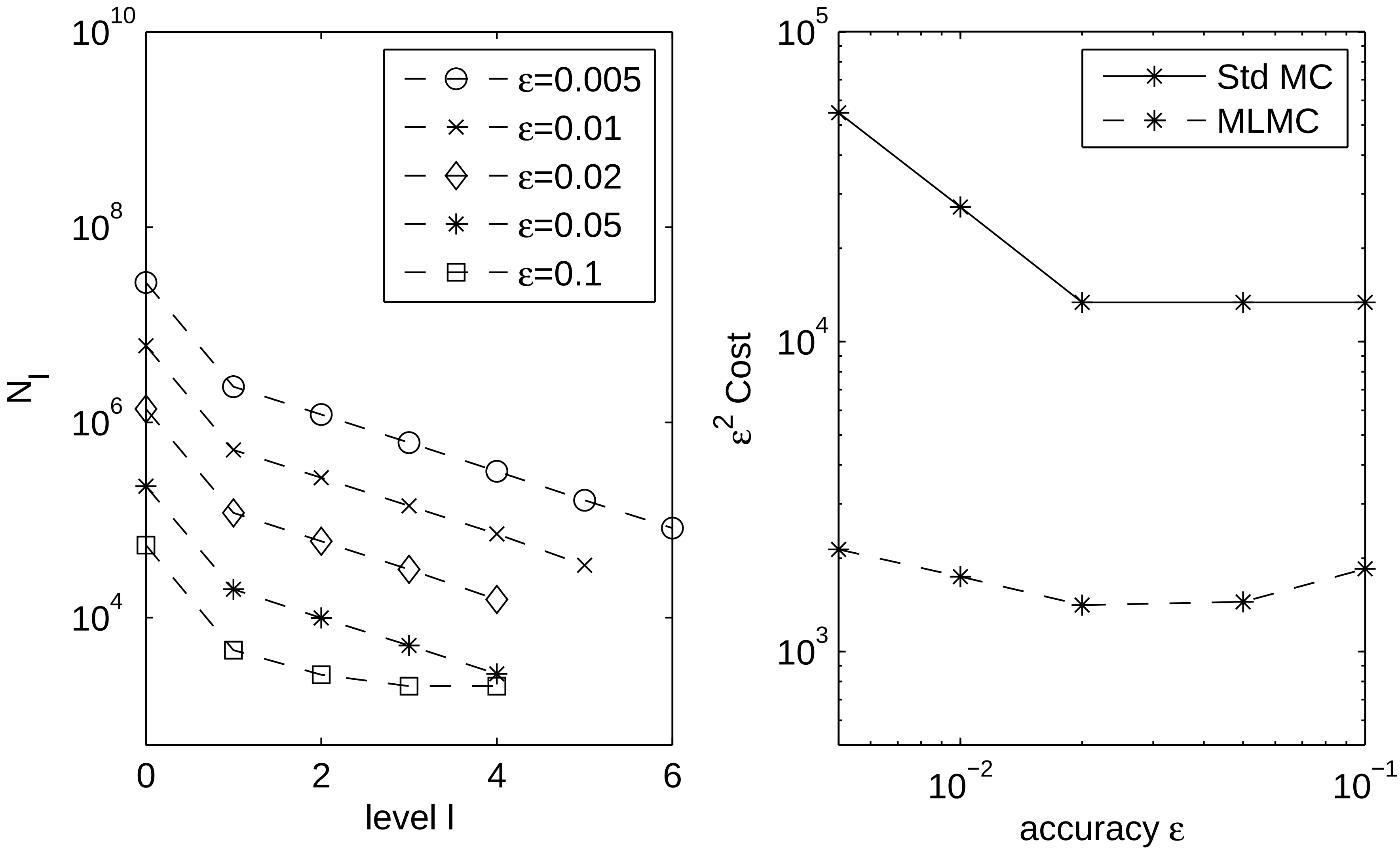
MLMC Results

GBM: European call, $\exp(-rT) \max(S(T)-K, 0)$



MLMC Results

GBM: European call, $\exp(-rT)\ \max(S(T)-K,0)$



MLMC Approach

So far, have kept things very simple:

- European option
- Euler discretisation
- single underlying in example

We now generalise it:

- arbitrary path-dependent options
- arbitrary discretisation
- assume certain properties for weak convergence and variance of multilevel correction
- obtain order of cost to achieve r.m.s. accuracy ε

MLMC Approach

Theorem: *Let P be a functional of the solution of a stochastic o.d.e., and $\widehat{P}_l$ the discrete approximation using a timestep $h_l = 2^{-l}\,T$.*

If *there exist independent estimators $\widehat{Y}_l$ based on N_l Monte Carlo samples, with computational complexity (cost) C_l, and positive constants $\alpha \geq \frac{1}{2}, \beta, c_1, c_2, c_3$ such that*

i) $\left| \mathbb{E}[\widehat{P}_l - P] \right| \leq c_1\, h_l^{\alpha}$

ii) $\mathbb{E}[\widehat{Y}_l] = \begin{cases} \mathbb{E}[\widehat{P}_0], & l=0 \\ \mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}], & l>0 \end{cases}$

iii) $\mathbb{V}[\widehat{Y}_l] \leq c_2\, N_l^{-1} h_l^{\beta}$

iv) $C_l \leq c_3\, N_l\, h_l^{-1}$

Multilevel MC Approach

then *there exists a positive constant* c_4 *such that for any* $\varepsilon < e^{-1}$ *there are values* L *and* N_l *for which the multilevel estimator*

$$\widehat{Y} = \sum_{l=0}^{L} \widehat{Y}_l,$$

has Mean Square Error $MSE \equiv \mathbb{E}\left[\left(\widehat{Y} - \mathbb{E}[P]\right)^2\right] < \varepsilon^2$

with a computational complexity C *with bound*

$$C \leq \begin{cases} c_4\, \varepsilon^{-2}, & \beta > 1, \\ c_4\, \varepsilon^{-2} (\log \varepsilon)^2, & \beta = 1, \\ c_4\, \varepsilon^{-2-(1-\beta)/\alpha}, & 0 < \beta < 1. \end{cases}$$

Milstein Scheme

The theorem suggests use of Milstein approximation
– better strong convergence, same weak convergence

Generic scalar SDE:

$$\mathrm{d}S(t) = a(S,t)\ \mathrm{d}t + b(S,t)\ \mathrm{d}W(t), \quad 0\!<\!t\!<\!T.$$

Milstein scheme:

$$\widehat{S}_{n+1} = \widehat{S}_n + a\,h + b\,\Delta W_n + \tfrac{1}{2}\,b'\,b\left((\Delta W_n)^2 - h\right).$$

Milstein Scheme

In scalar case:

- $O(h)$ strong convergence
- $O(\varepsilon^{-2})$ complexity for Lipschitz payoffs – trivial
- $O(\varepsilon^{-2})$ complexity for more complex cases using carefully constructed estimators based on Brownian interpolation or extrapolation
  - digital, with discontinuous payoff
  - Asian, based on average
  - lookback and barrier, based on min/max
- This extends naturally to basket options based on a weighted average of assets linked only through the correlation in the driving Brownian motion

Milstein Scheme

Brownian interpolation: within each timestep, model the behaviour as simple Brownian motion conditional on the two end-points

$$\begin{aligned}\widehat{S}(t) &= \widehat{S}_n + \lambda(t)(\widehat{S}_{n+1} - \widehat{S}_n) \\ &\quad + b_n \Big(W(t) - W_n - \lambda(t)(W_{n+1} - W_n) \Big),\end{aligned}$$

where

$$\lambda(t) = \frac{t - t_n}{t_{n+1} - t_n}$$

There then exist analytic results for the distribution of the min/max/average over each timestep, and probability of crossing a barrier.

Milstein Scheme

Brownian extrapolation for final timestep:

$$\widehat{S}_N = \widehat{S}_{N-1} + a_{N-1} h + b_{N-1} \Delta W_N$$

Considering all possible ΔW_N gives Gaussian distribution, for which a digital option has a known conditional expectation – example in Glasserman's book of payoff smoothing to allow pathwise calculation of Greeks.

This payoff smoothing can be extended to general multivariate cases (not just baskets) through a “vibrato” Monte Carlo technique which is suitable for both efficient multilevel analysis and the computation of Greeks

Results

Basket of 5 underlying assets, each GBM with

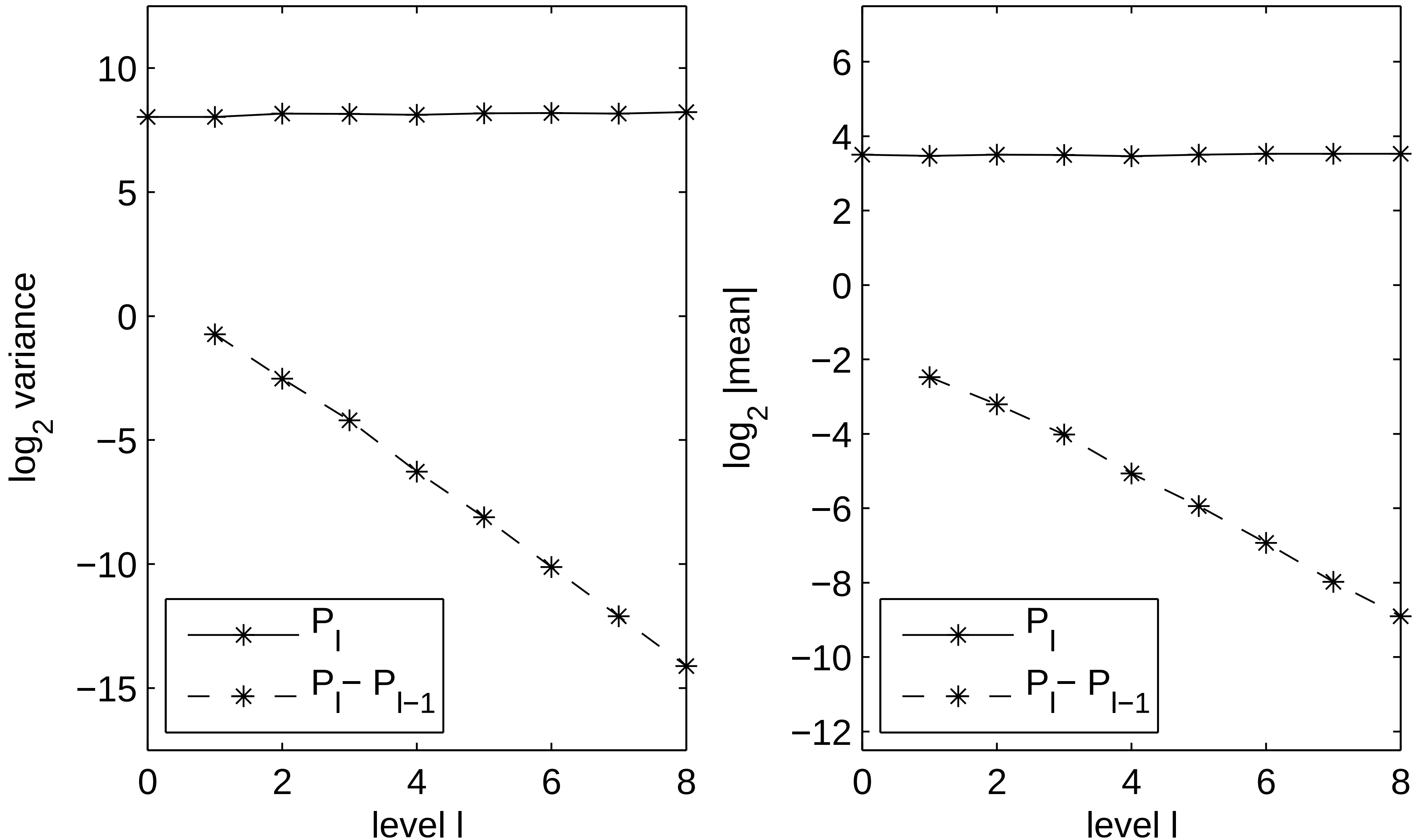
$$r = 0.05, \;\; T = 1, \;\; S_i(0) = 100, \;\; \sigma = (0.2,\, 0.25,\, 0.3,\, 0.35,\, 0.4),$$

and correlation $\rho = 0.25$ between each of the driving Brownian motions.

All options are based on arithmetic average $\overline{S}$ of 5 assets, with strike $K = 100$ (and barrier $B = 85$).

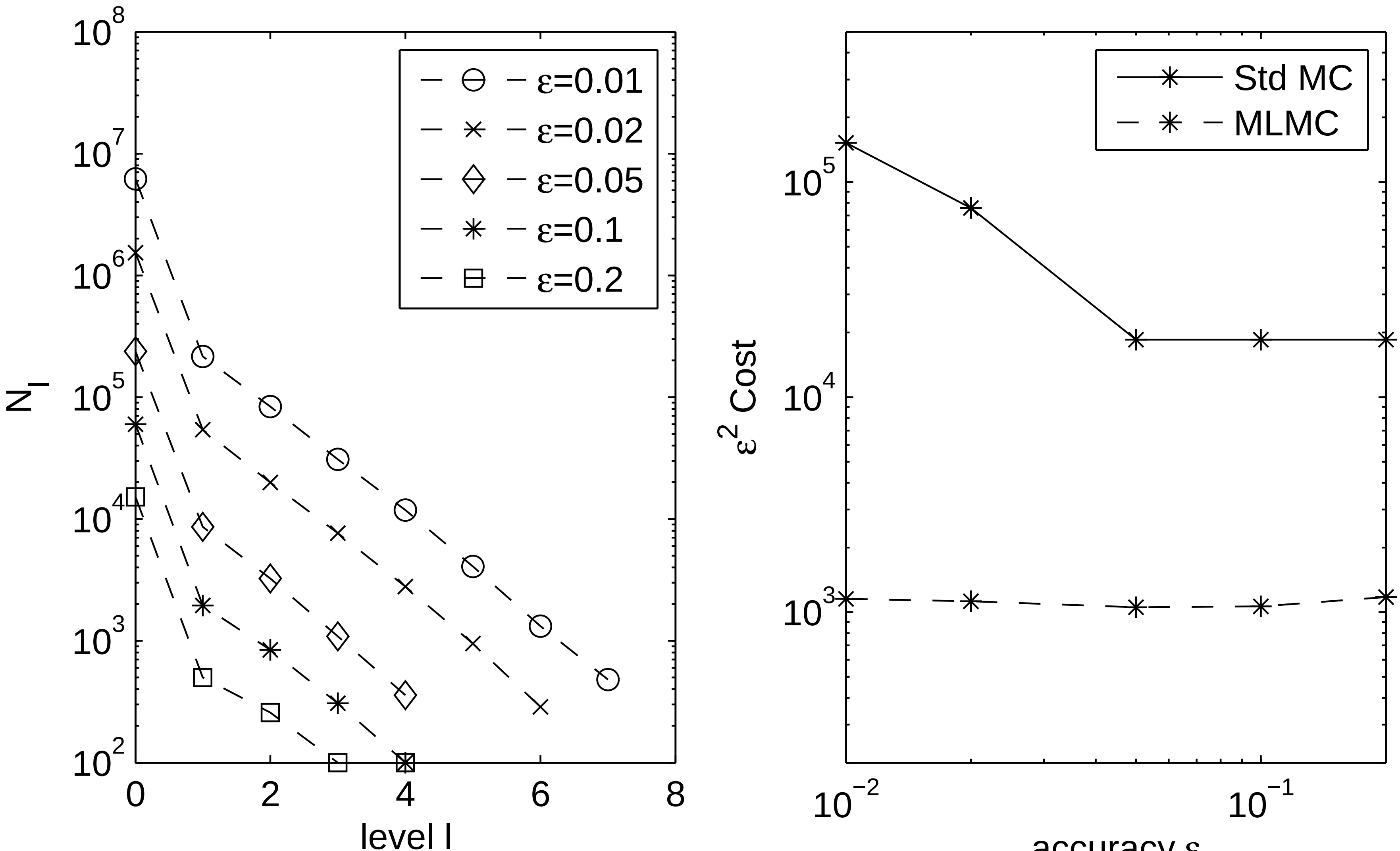
MLMC Results

European call, $\exp(-rT)\ \max(\overline{S}(T) - K, 0)$



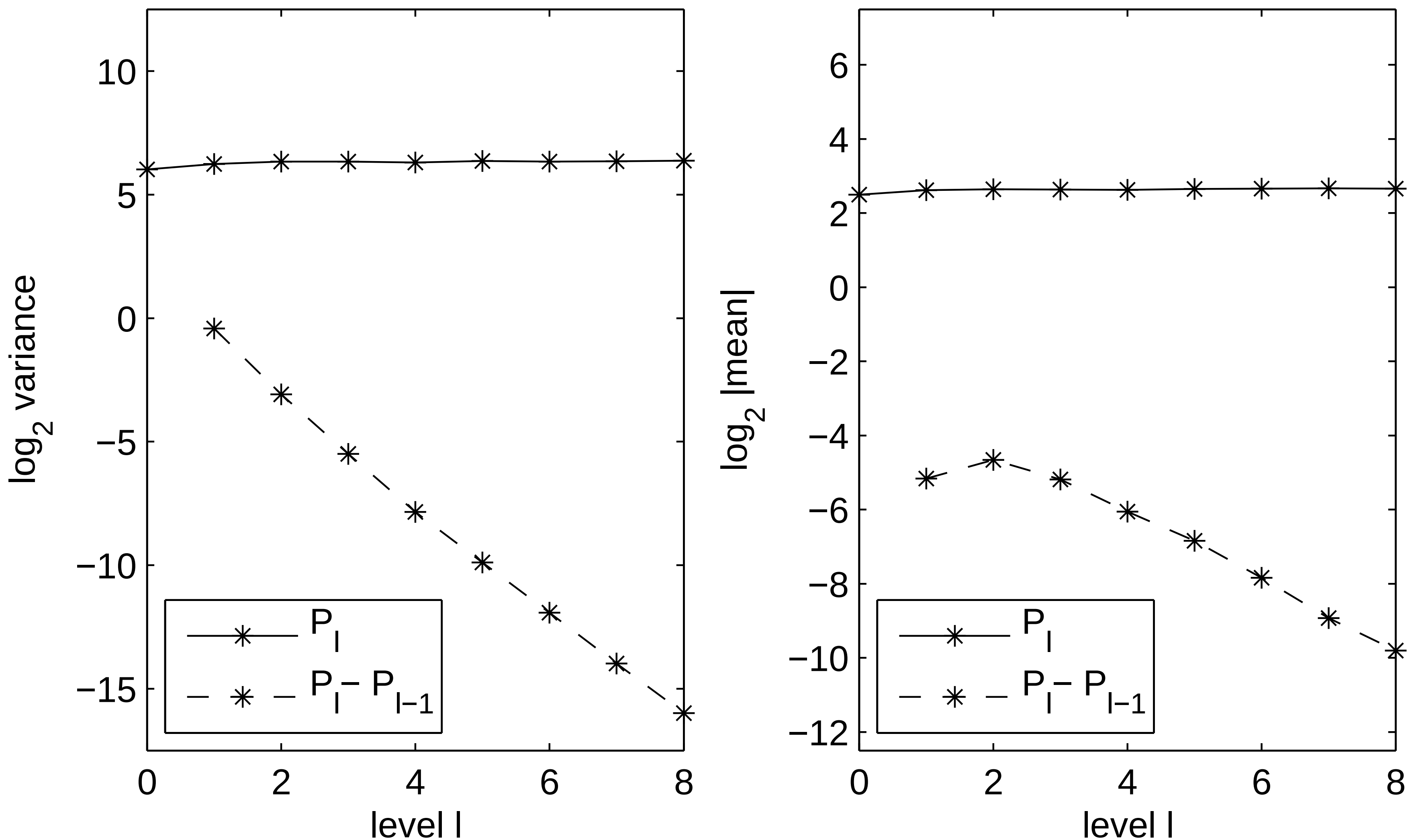
MLMC Results

European call, $\exp(-rT)\ \max(\overline{S}(T)-K,0)$



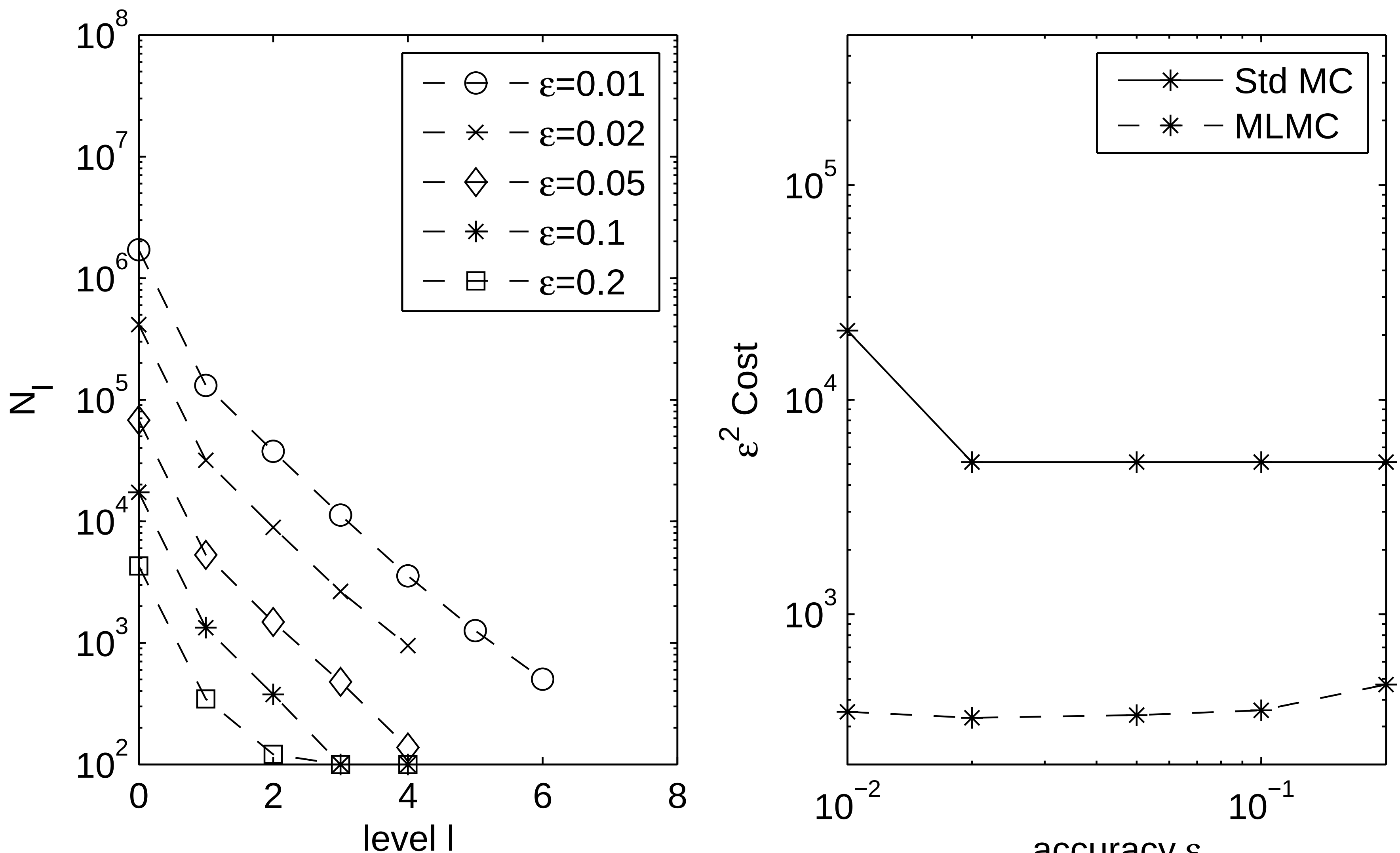
MLMC Results

Asian option, $\exp(-rT)\,\max(T^{-1}\int_0^T \overline{S}(t)\,\mathrm{d}t - K, 0)$



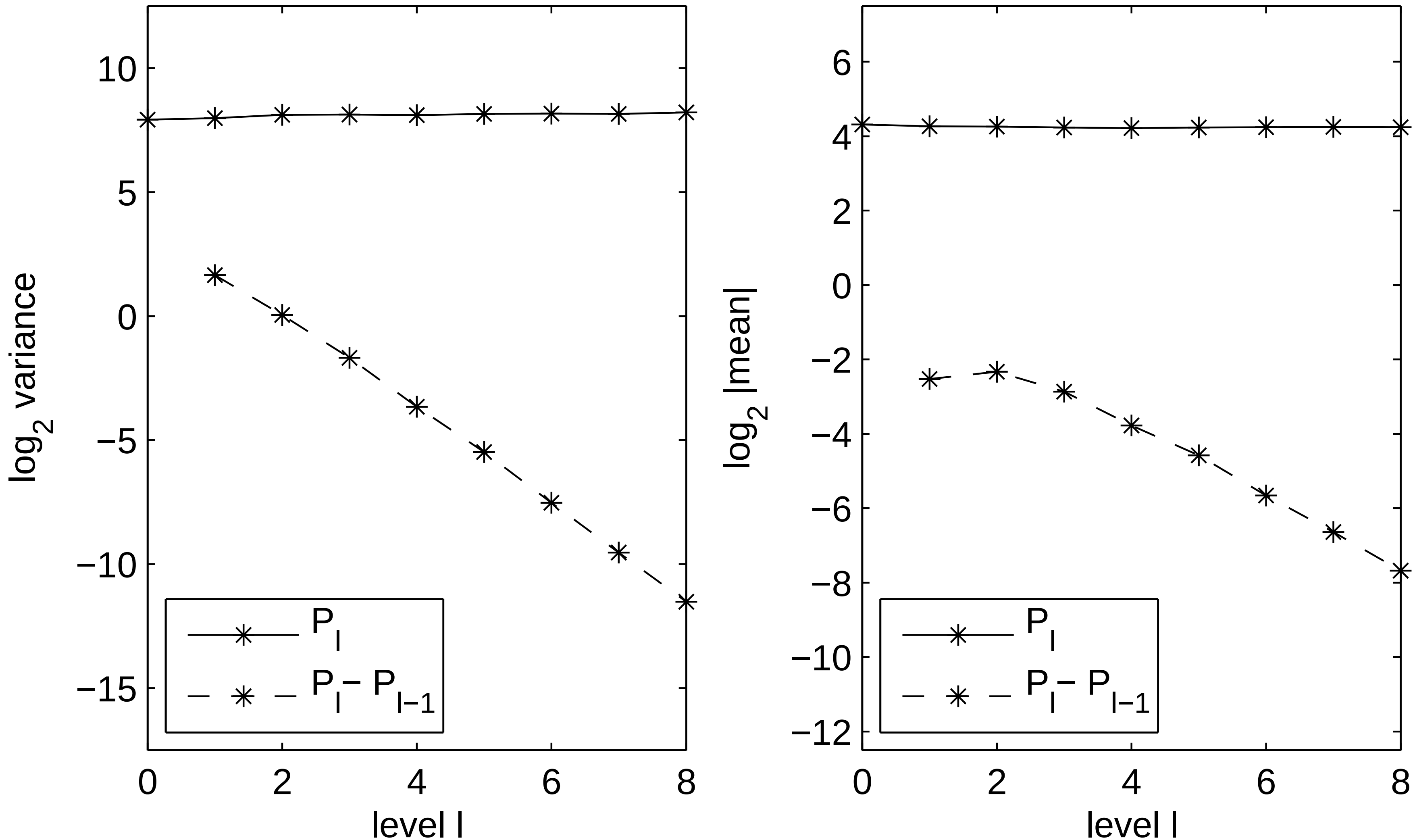
MLMC Results

Asian option, $\exp(-rT)\,\max(T^{-1}\int_0^T \overline{S}(t)\,\mathrm{d}t - K, 0)$



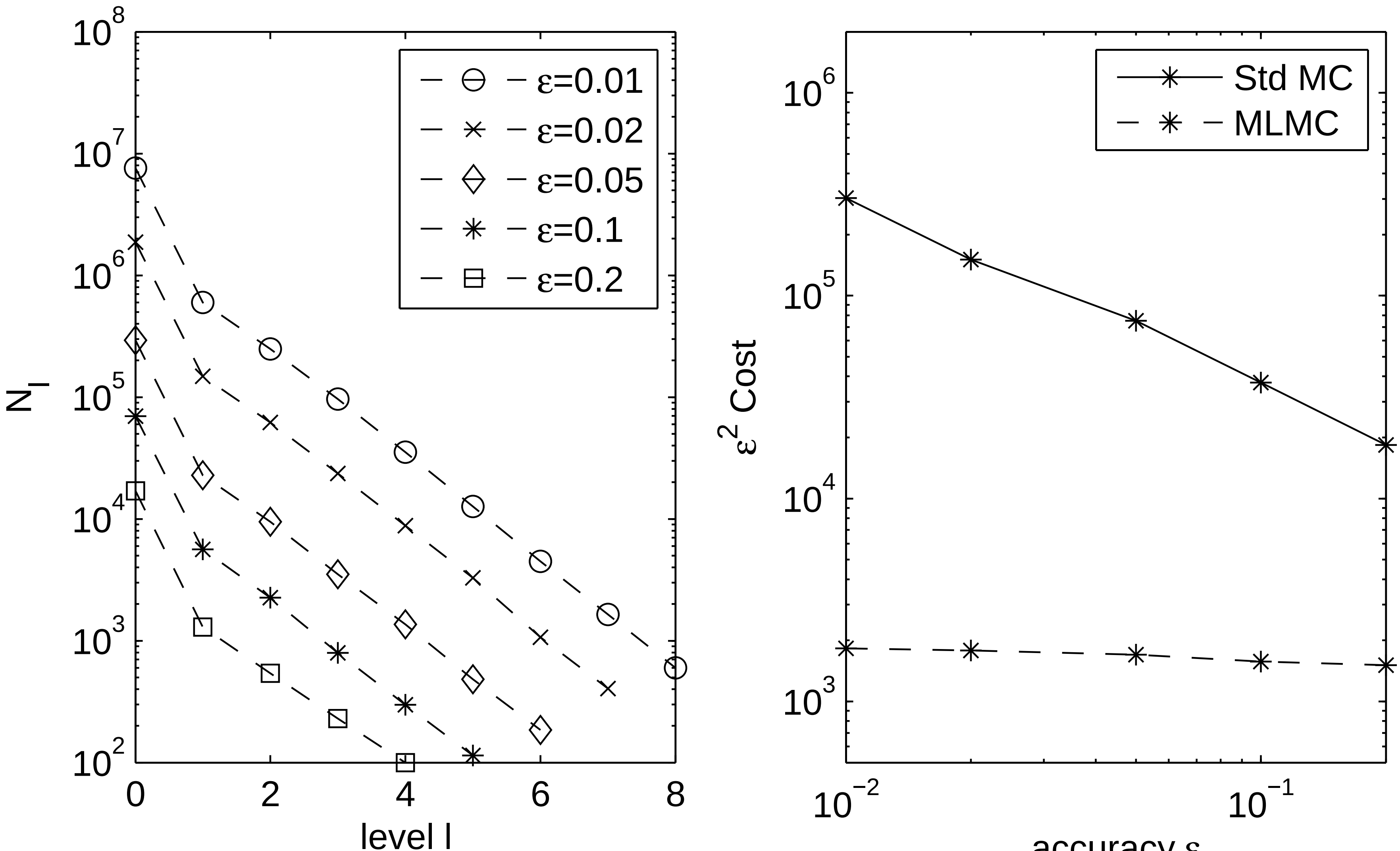
MLMC Results

Lookback option, $\exp(-rT)\,(\overline{S}(T) - \min_{0<t<T} \overline{S}(t))$



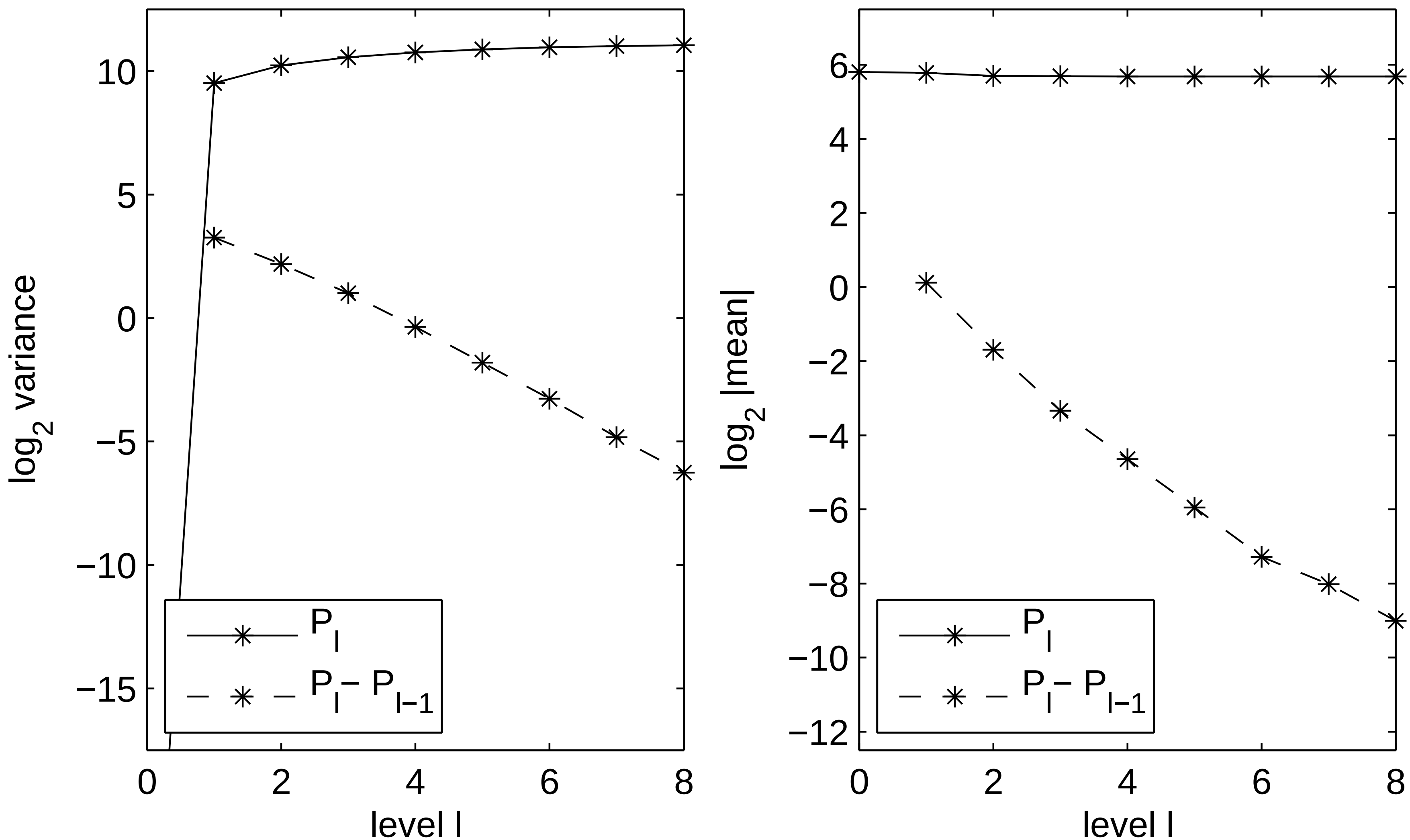
MLMC Results

Lookback option, $\exp(-rT)\left(\overline{S}(T) - \min_{0<t<T} \overline{S}(t)\right)$



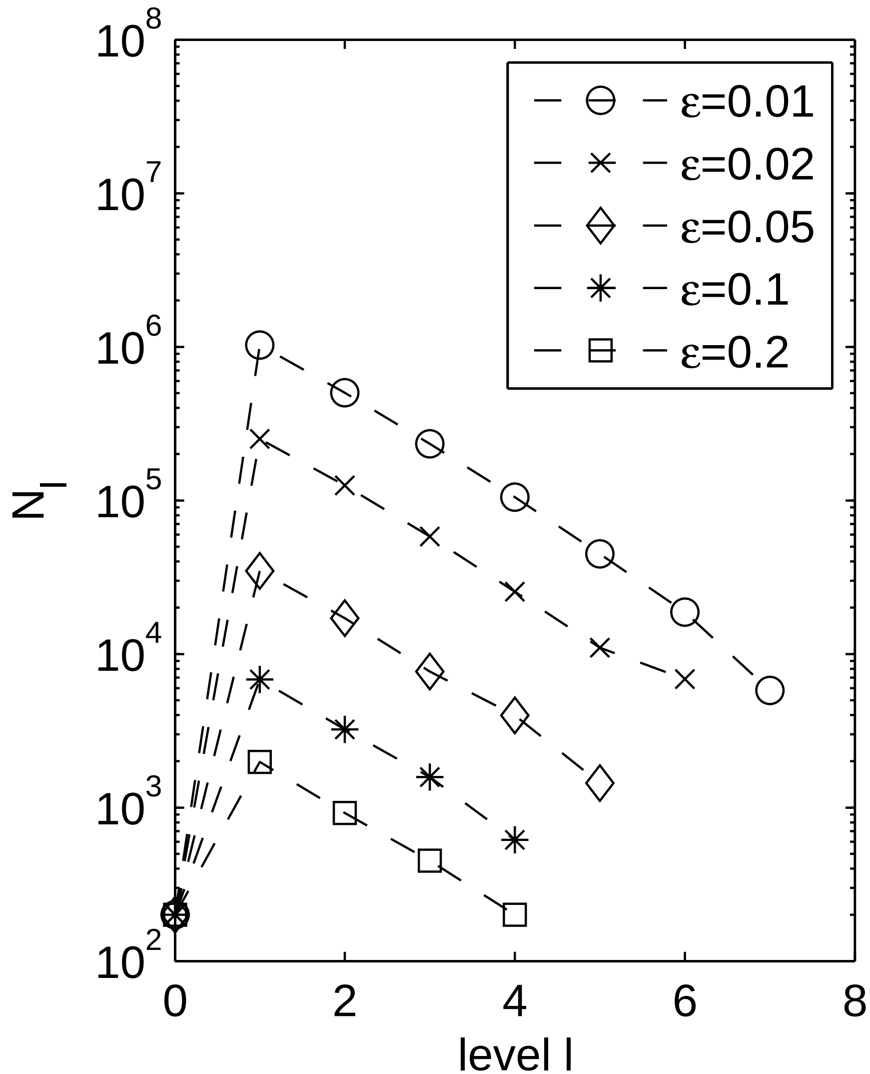
MLMC Results

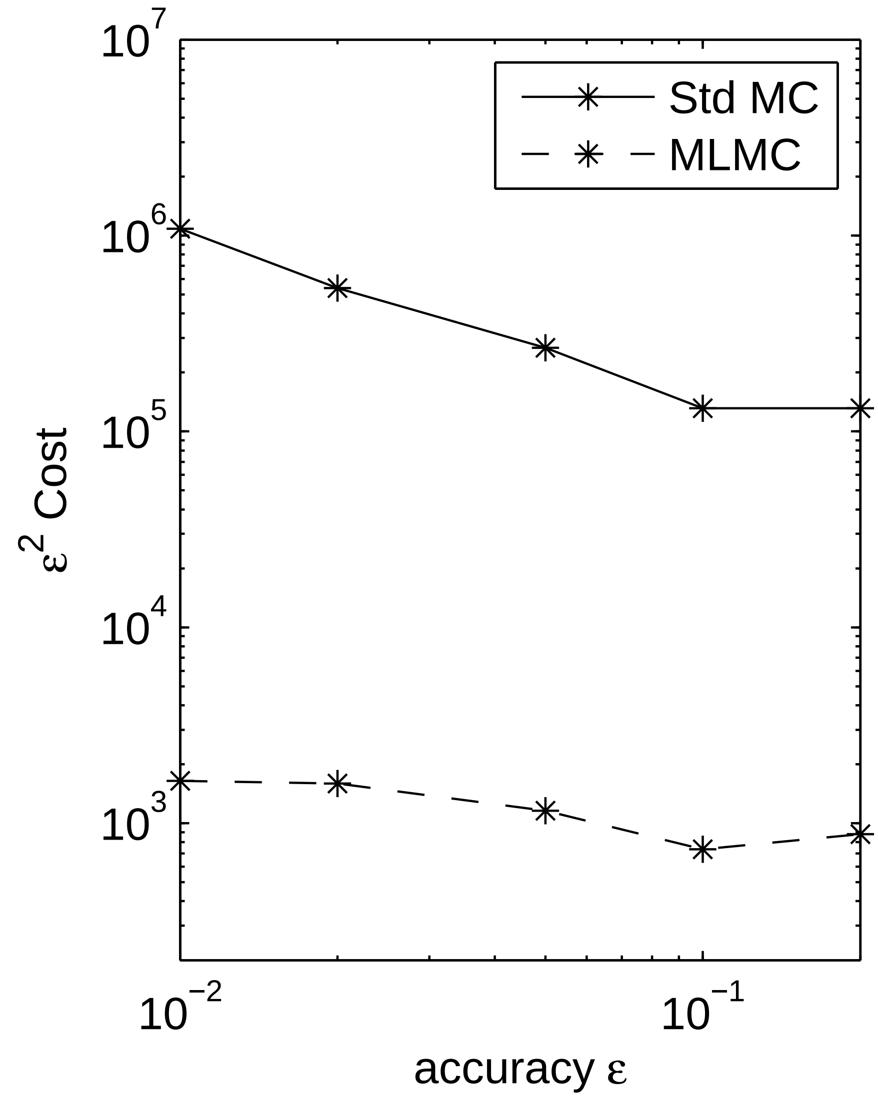
Digital option, $100\ \exp(-rT)\ \mathbf{1}_{\overline{S}(T)>K}$



MLMC Results

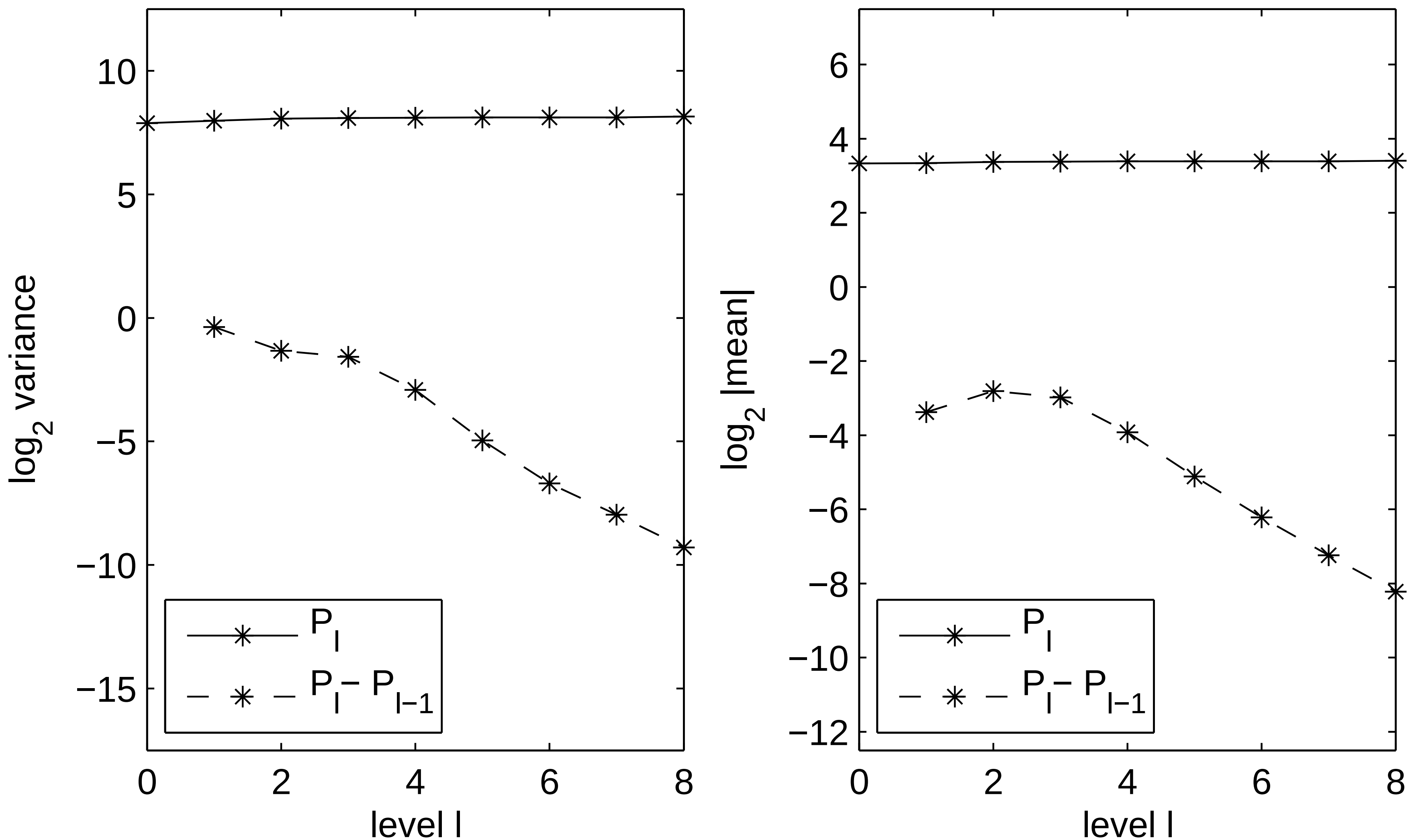
Digital option, $100\ \exp(-rT)\ \mathbf{1}_{\overline{S}(T)>K}$





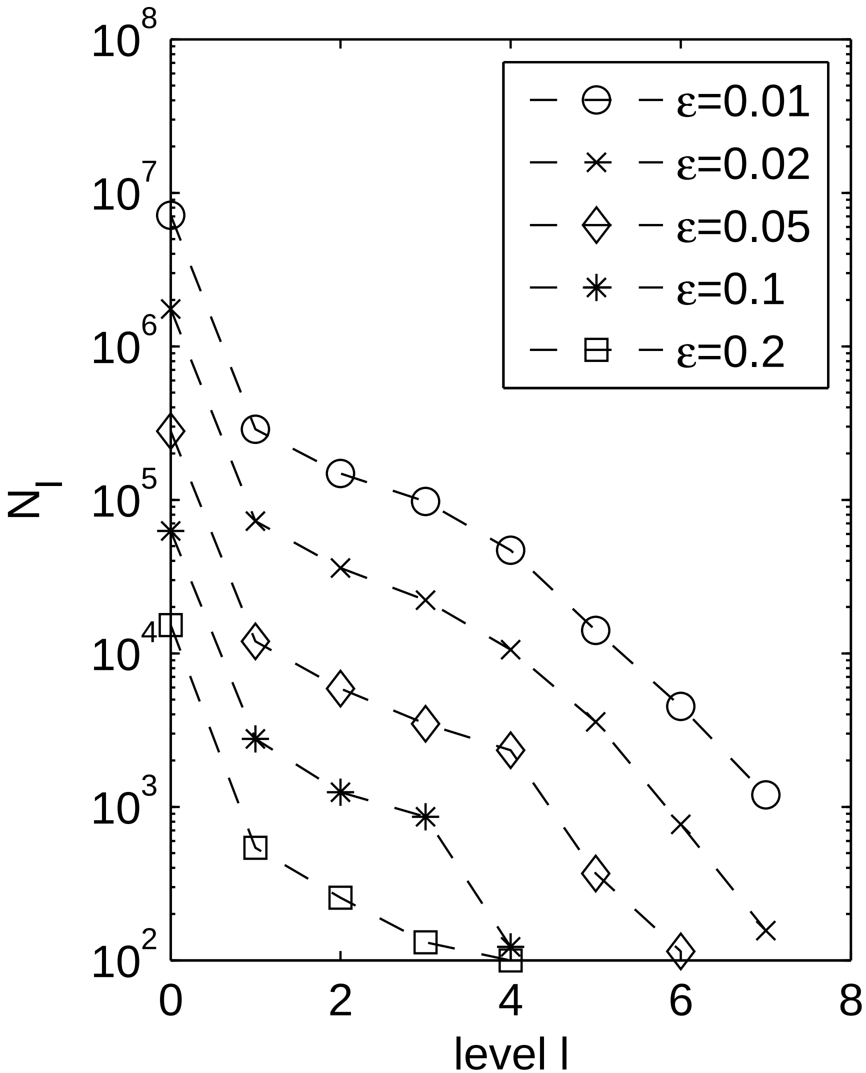
MLMC Results

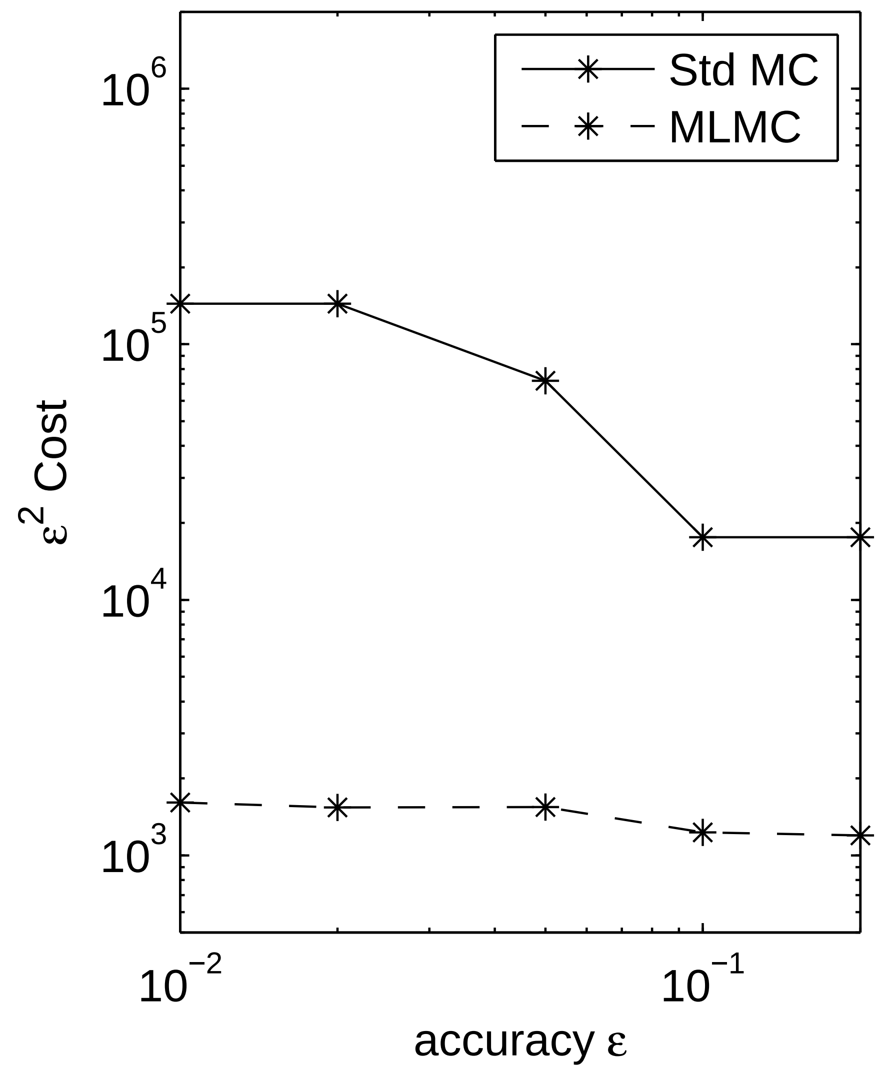
Barrier option, $\exp(-rT)\ \max(\overline{S}(T)-K,0)\ \mathbf{1}_{\min_{0<t<T}\overline{S}(t)>B}$



MLMC Results

Barrier option, $\exp(-rT)\ \max(\overline{S}(T)-K,0)\ \mathbf{1}_{\min_{0<t<T} \overline{S}(t)>B}$





MLMC Numerical Analysis

	Euler		Milstein	
option	numerics	analysis	numerics	analysis
Lipschitz	$O(h)$	$O(h)$	$O(h^2)$	$O(h^2)$
Asian	$O(h)$	$O(h)$	$O(h^2)$	$O(h^2)$
lookback	$O(h)$	$O(h)$	$O(h^2)$	$o(h^{2-\delta})$
barrier	$O(h^{1/2})$	$o(h^{1/2-\delta})$	$O(h^{3/2})$	$o(h^{3/2-\delta})$
digital	$O(h^{1/2})$	$O(h^{1/2}\log h)$	$O(h^{3/2})$	$o(h^{3/2-\delta})$

Table: convergence for V_l as observed numerically and proved analytically for both the Euler and Milstein discretisations. δ can be any strictly positive constant.

MLMC Numerical Analysis

Analysis for Euler discretisation for scalar SDE:

- lookback and barrier: Giles, Higham & Mao (*Finance & Stochastics, 13(3), 2009*)
- digital: Avikainen (*Finance & Stochastics, 13(3), 2009*)

Analysis for Milstein discretisation for scalar SDE:

- Giles, Debrabant & Rößler (TU Darmstadt)
- uses boundedness of all moments to bound the contribution to V_l from “extreme” paths (e.g. for which $\max_n |\Delta W_n| > h^{1/2-\delta}$ for some $\delta\!>\!0$)
- uses asymptotic analysis to bound the contribution from paths which are not “extreme”

Milstein scheme

Milstein scheme for multi-dimensional SDEs generally requires Lévy areas:

$$A_{jk,n} = \int_{t_n}^{t_{n+1}} (W_j(t) - W_j(t_n))\, \mathrm{d}W_k - (W_k(t) - W_k(t_n))\, \mathrm{d}W_j .$$

- $O(h^{1/2})$ strong convergence in general if omitted
- Can still get good convergence for Lipschitz payoffs by using $W^c(t) = \frac{1}{2}(W^{f1}(t) + W^{f2}(t))$ with two fine paths created by antithetic Brownian Bridge construction
- For barrier and digital options, need to simulate Lévy areas – tradeoff between cost and accuracy, optimum may require $O(h^{3/2})$ sub-sampling of Brownian paths, giving $O(h^{3/4})$ strong convergence

Results

Heston stochastic volatility model:

$$\mathrm{d}S = r\, S\, \mathrm{d}t + \sqrt{v}\, S\, \mathrm{d}W_1, \qquad 0 < t < T,$$

$$\mathrm{d}v = \kappa(\theta - v) + \xi \sqrt{v}\, \mathrm{d}W_2, \qquad 0 < t < T,$$

with $T = 1, \;\; S(0) = 100, \;\; r = 0.05, \;\; \theta = 0.04, \;\; \xi = 0.25$
and differing values of κ.

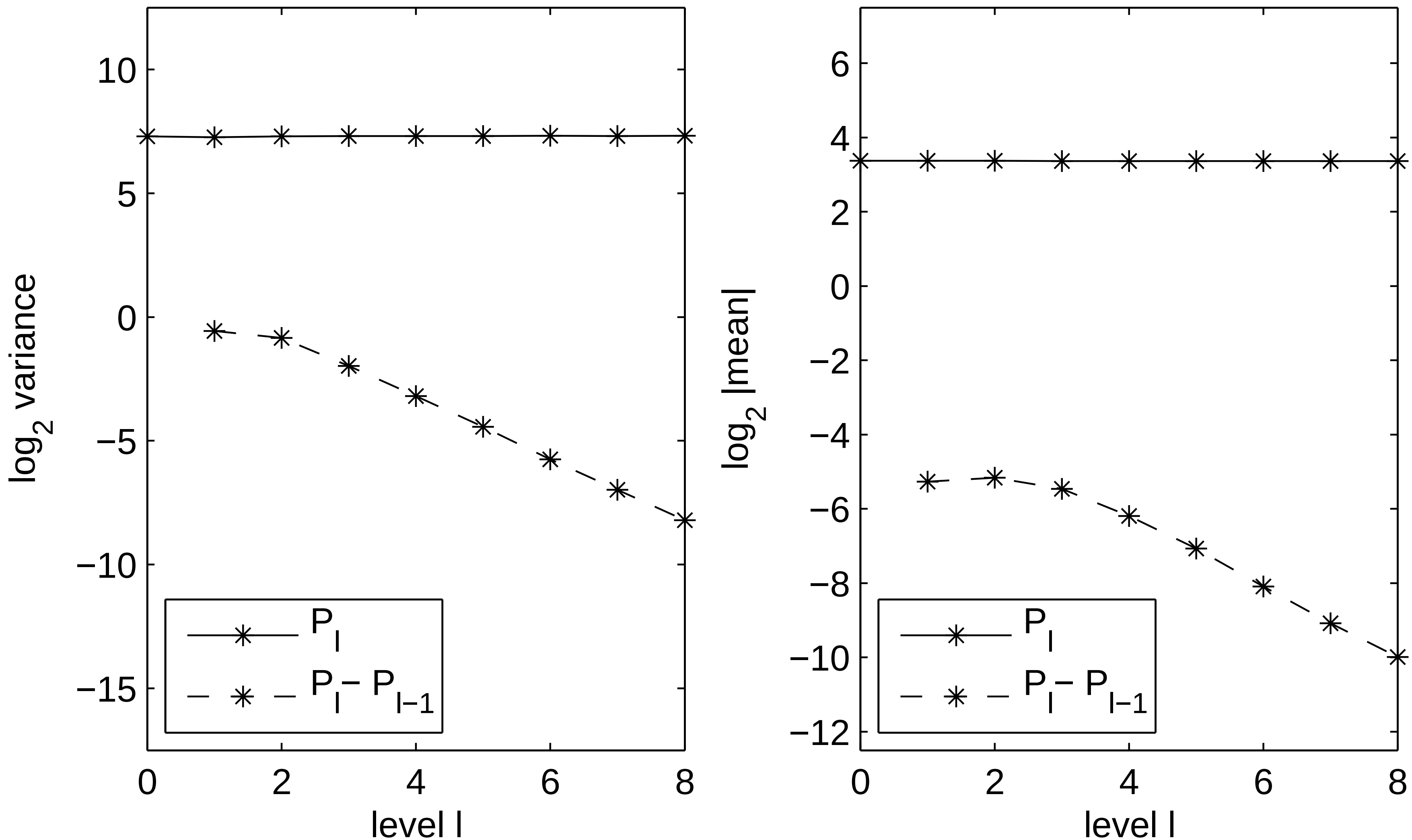
European call option with discounted payoff

$$\exp(-rT)\;\max(S(T) - K, 0)$$

with strike $K = 100$.

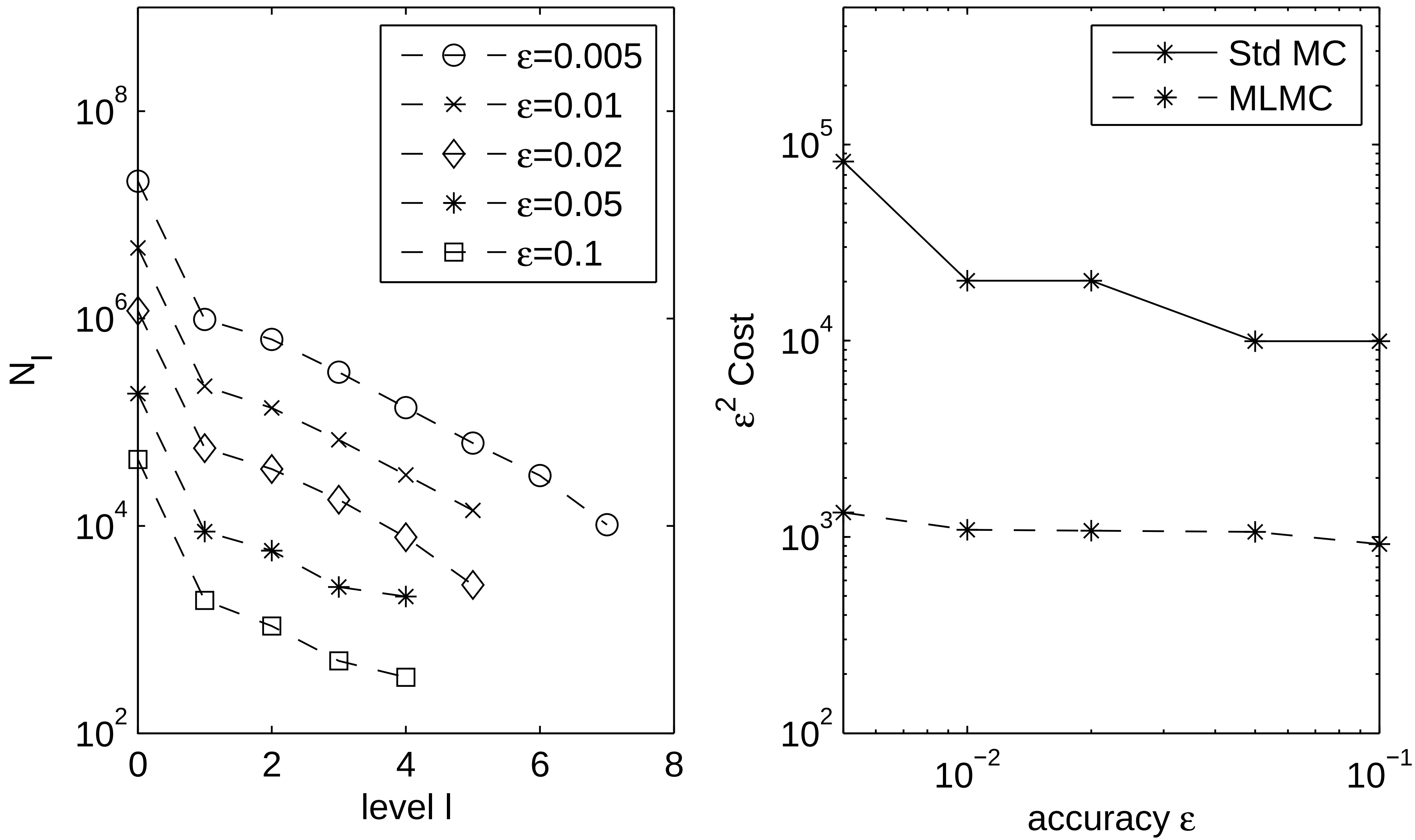
MLMC Results

Heston: European call, $\kappa\theta/\xi^2 = 2/3$



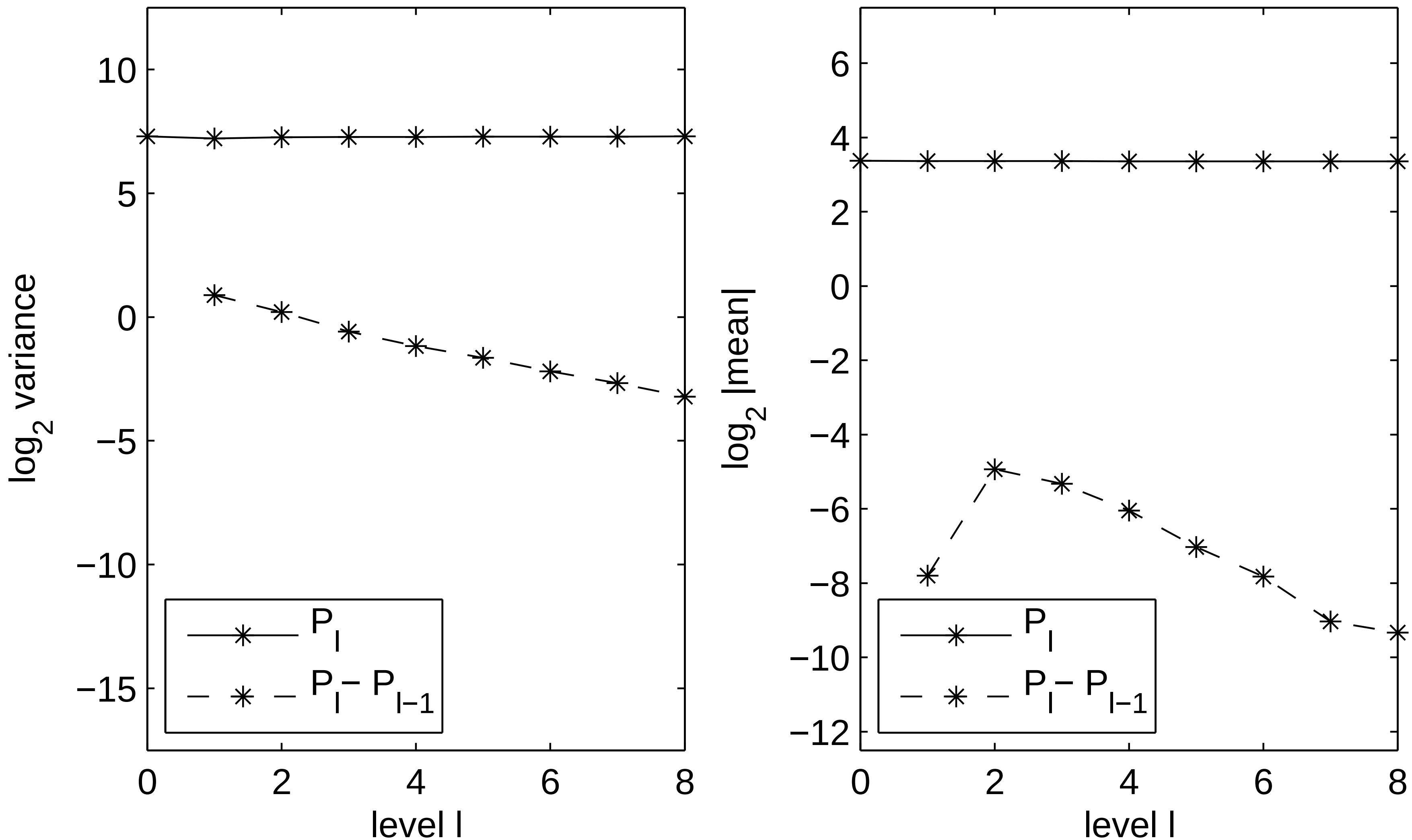
MLMC Results

Heston: European call, $\kappa\theta/\xi^2 = 2/3$



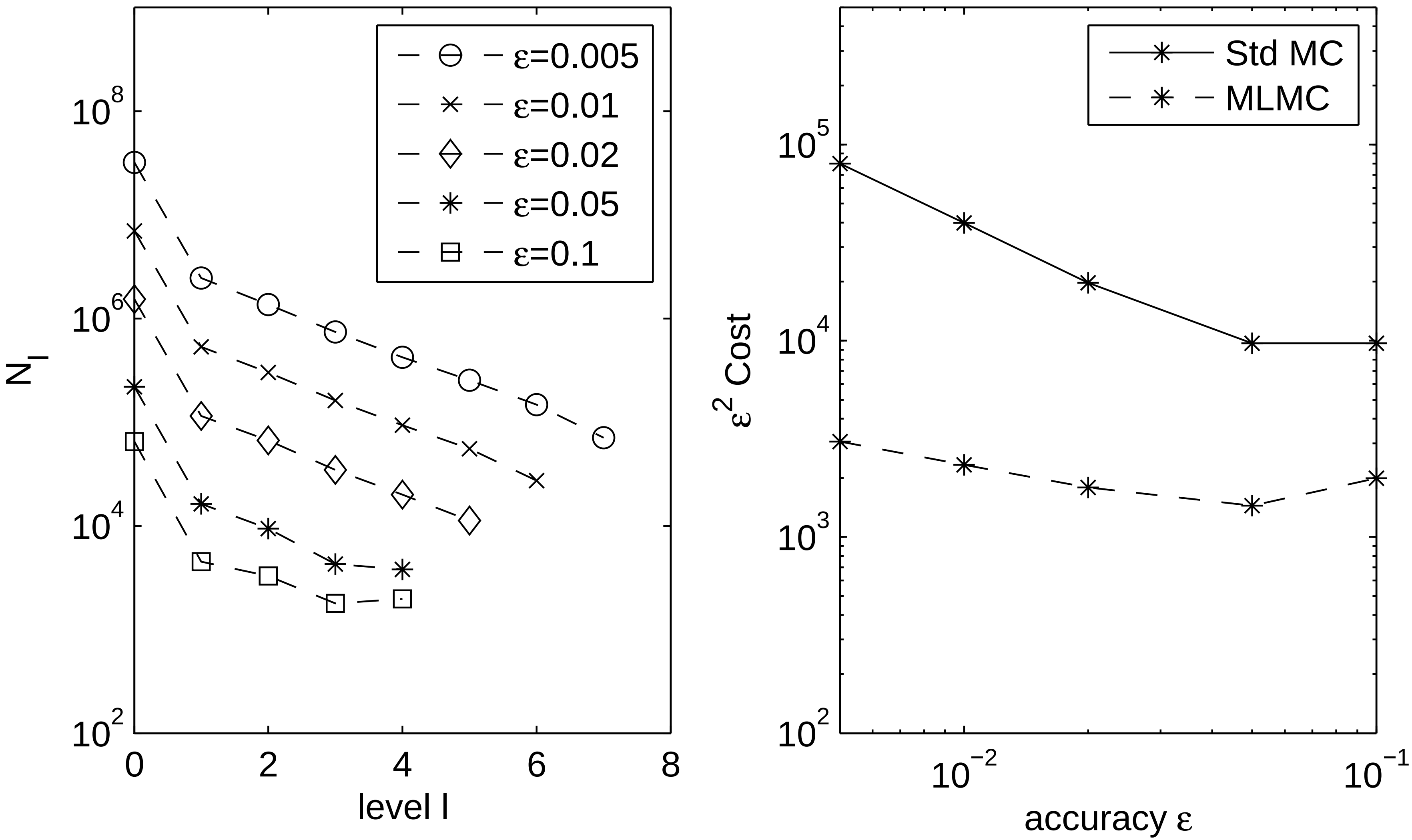
MLMC Results

Heston: European call, $\kappa\theta/\xi^2 = 1/3$



MLMC Results

Heston: European call, $\kappa\theta/\xi^2 = 1/3$



Heston model

How can harder cases be handled better?

- could combine multilevel with adaptive time-stepping (Raul Tempone and Anders Szepessy)
- could use Glasserman and Kim's efficient implementation of Broadie and Kaya's exact simulation method
  - multilevel unnecessary for European options, but would give benefits for path-dependent options
  - could also use multilevel to handle a local vol surface

SPDE application

Currently working with Christoph Reisinger on an SPDE application which arises in CDO modelling (Bush, Hambly, Haworth & Reisinger)

$$\mathrm{d}p = -\mu \, \frac{\partial p}{\partial x} \, \mathrm{d}t + \frac{1}{2} \, \frac{\partial^2 p}{\partial x^2} \, \mathrm{d}t + \sqrt{\rho} \, \frac{\partial p}{\partial x} \, \mathrm{d}W$$

with absorbing boundary $p(0,t) = 0$

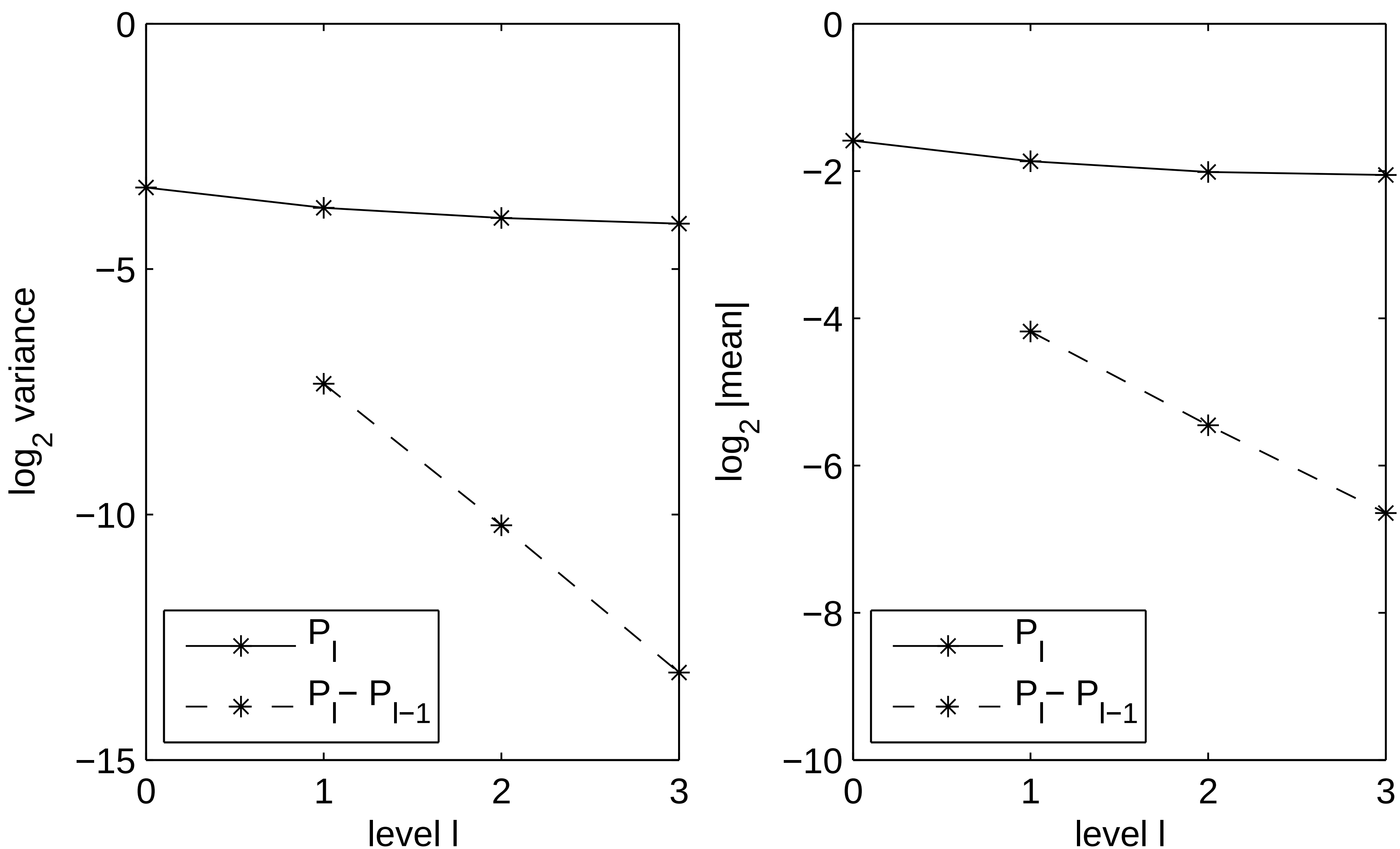
- derived in limit as number of firms $\longrightarrow \infty$
- x is distance to default
- $p(x,t)$ is probability density function
- $\mathrm{d}W$ term corresponds to systemic risk
- $\partial^2 p / \partial x^2$ comes from idiosyncratic risk

SPDE application

- numerical discretisation combines Milstein time-marching with central difference approximations
- coarsest level of approximation uses 1 timestep per quarter, and 10 spatial points
- each finer level uses four times as many timesteps, and twice as many spatial points – ratio is due to numerical stability constraints

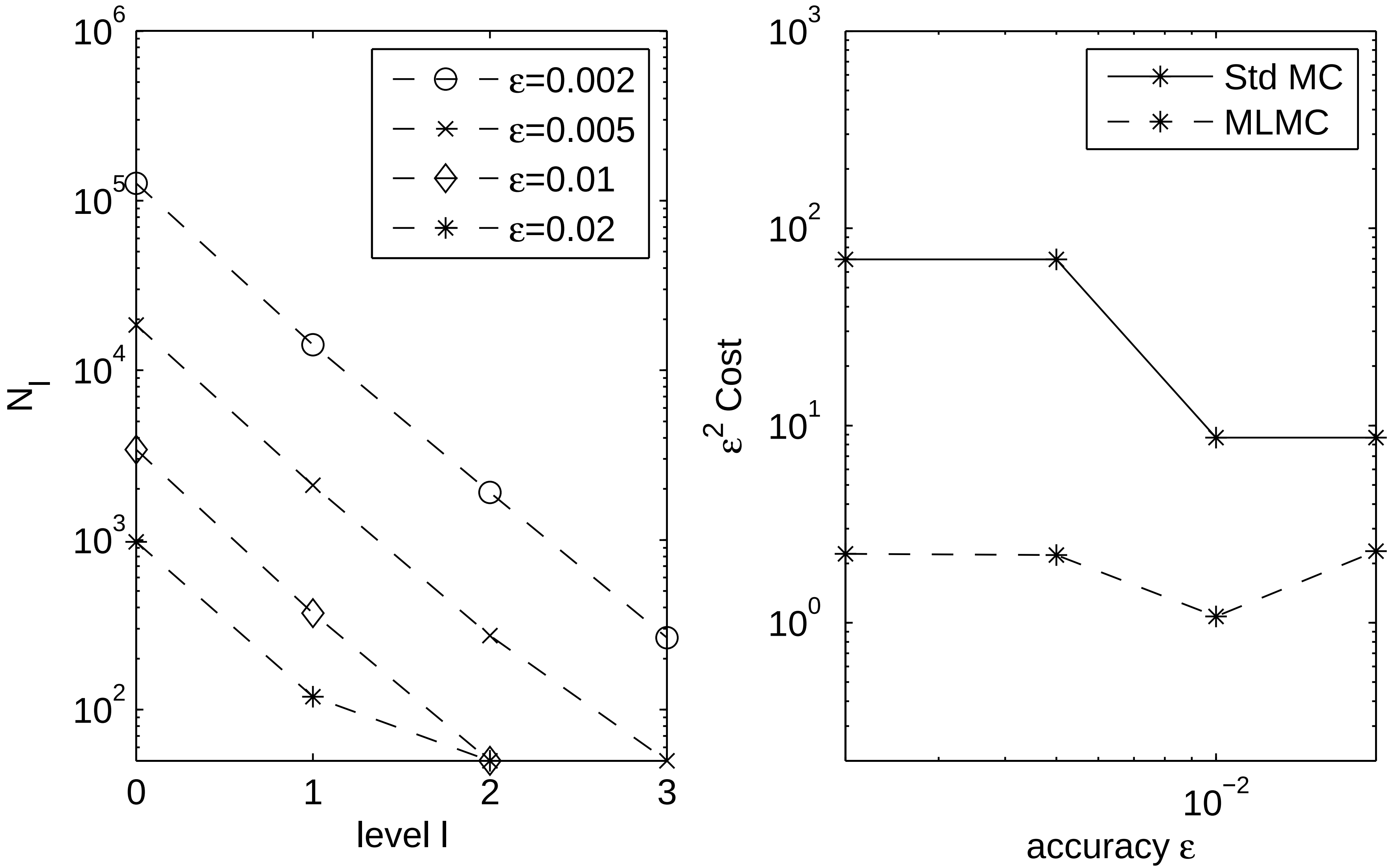
MLMC Results

Fractional loss on equity tranche of a 5-year CDO:



MLMC Results

Fractional loss on equity tranche of a 5-year CDO:



Other work

- Quasi-Monte Carlo:
  - uses deterministic sample “points” to achieve an error which is nearly $O(N^{-1})$ in the best cases
  - little applicable theory due to lack of smoothness, but great results using rank-1 lattice rules developed by Ian Sloan’s group at UNSW
- implementation on GPUs
  - up to 240 cores per GPU, each equivalent to 10-50% of an Intel core for single precision calculations
  - ideally suited for trivially-parallel Monte Carlo applications
  - could use multilevel to correct for difference between single and double precision?

Future work

- “vibrato” technique for digital options:
  - current treatment uses conditional expectation one timestep before maturity, which smooths the payoff
  - the “vibrato” idea generalises this to cases without a known conditional expectation
- Greeks:
  - the multilevel approach should work well, combining pathwise sensitivities with “vibrato” treatment to cope with lack of smoothness
  - can also incorporate the adjoint approach developed with Paul Glasserman – more efficient when many Greeks are wanted for one payoff function

Future work

- variance-gamma, CGMY and other processes:
  - given exact simulation techniques, multilevel benefit is in treating path-dependent options
  - could also handle addition of a local vol surface
- American options – the next big challenge:
  - instead of Longstaff-Schwartz approach, view it as an exercise boundary optimisation problem, and use multilevel optimisation?

Conclusions

Multilevel Monte Carlo method has already achieved

- improved order of complexity
- significant benefits for model problems

but there is still a lot more research to be done, both theoretical and applied.

M.B. Giles, “Multilevel Monte Carlo path simulation”, *Operations Research*, 56(3):607-617, 2008.

M.B. Giles. “Improved multilevel Monte Carlo convergence using the Milstein scheme”, pp. 343-358 in *Monte Carlo and Quasi-Monte Carlo Methods 2006*, Springer, 2007.

Papers are available from:
`www.maths.ox.ac.uk/~gilesm/finance.html`