Static Control variate

- Back to business: killing the variance.
- Let $X,\ X' \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ satisfying

$$\mathbb{E}\,X = \mathbb{E}\,X' = m_X \in \mathbb{R}\setminus\{0\}, \quad \mathrm{Var}(X),\ \mathrm{Var}(X'),\ \mathrm{Var}(X-X') > 0$$

- **Question:** Which random vector (distribution. . .) is more appropriate?
- **(Not so) Naive answer:** *if both X and X' can be simulated with an similar complexity*, the one with the lowest variance. . . *i.e.*

$$X \quad \text{if} \quad \mathrm{Var}(X) < \mathrm{Var}(X'), \quad X' \ \text{ otherwise}$$

- So what ?

Practitioner's viewpoint

Definition

Let $\Xi \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ such that:

(i) $\mathbb{E}\,\Xi$ can be computed at a very low cost by a deterministic method (closed form, numerical analysis method like *PDE*),

(ii) the random variable $X - \Xi$ can be simulated with the same complexity as X,

(iii) the variance $\mathrm{Var}(X - \Xi) < \mathrm{Var}(X)$.

Then, the random variable

$$X' = X - \Xi + \mathbb{E}\,\Xi$$

can be simulated at the same cost as X,

$$\mathbb{E}\,X' = \mathbb{E}\,X = m \quad \text{and} \quad \mathrm{Var}(X') = \mathrm{Var}(X - \Xi) < \mathrm{Var}(X).$$

A random variable Ξ satisfying (i)-(ii)-(iii) is called a *control variate* for X.

Variant (pseudo-control variate)

- When $X \geqslant 0$, one often replace (iii) by

$(iii)' \qquad 0 \leqslant \Xi \leqslant X$

- It is NOT a control variate but it often works...

A useful tool: Jensen's inequality

Theorem (Jensen's Inequality)

Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ be a random variable and let $g : \mathbb{R} \to \mathbb{R}$ be a convex function. Suppose X and $g(X)$ are integrable.

(a) Let $\mathcal{B}$ be a sub-σ-field $\mathcal{A}$,

$$g\big(\mathbb{E}(X\,|\,\mathcal{B})\big) \leqslant \mathbb{E}\big(g(X)\,|\,\mathcal{B}\big) \qquad \mathbb{P}\text{-}a.s.$$

(b) In particular considering if $\mathcal{B} = \{\emptyset, \Omega\}$

$$g\big(\mathbb{E}(X)\big) \leqslant \mathbb{E}\,g(X).$$

Remark. To memorize, think to the simplest convex function i.e. $|\,.\,|$

Basket option

- $(X_t^{1,x_1}, \ldots, X_t^{d,x_d})_{t \geqslant 0}$ models the price of d positive traded risky assets.
- The payoff of a basket option reads

$$h_T = \left(\sum_{i=1}^{d} \alpha_i X_T^{i,x_i} - K\right)^+, \quad \alpha_i > 0, \quad \sum_i \alpha_i = 1.$$

- By convexity

$$0 \leqslant e^{\sum_{1 \leqslant i \leqslant d} \alpha_i \log(X_T^{i,x_i})} \leqslant \sum_{i=1}^{d} \alpha_i X_T^{i,x_i}$$

so that

$$0 \leqslant k_T := \left(e^{\sum_{1 \leqslant i \leqslant d} \alpha_i \log(X_T^{i,x_i})} - K\right)^+ \leqslant h_T.$$

- It suggests to set $\Xi = k_T$ as a pseudo-control variate.

Phase I: in search for Ξ (in a B-S model)

- In a correlated d-dimensional Black-Scholes model:

$$X_t^{i,x_i} = x_i \exp\left(\left(r - \frac{\sigma_{i.}^2}{2}\right)t + \sum_{j=1}^{q} \sigma_{ij} W_t^j\right),\ t \in [0, T],\ x_i > 0,\ i = 1:d,$$

where $\sigma_{i.}^2 = \sum_{j=1}^{q} \sigma_{ij}^2, \quad i = 1, \ldots, d$. Then

$$e^{\sum_{1 \leqslant i \leqslant d} \alpha_i \log(X_T^{i,x_i})} \stackrel{d}{=} e^{aZ+b},\ a,\ b \in \mathbb{R},\ Z \stackrel{d}{=} \mathcal{N}(0;1).$$

- More precise (elementary) computations yield

$$e^{-rT}\mathbb{E}\, k_T = \mathrm{Call}_{BS}\left(\prod_{i=1}^{d} x_i^{\alpha_i} e^{-\frac{1}{2}\left(\sum_{i=1:d} \alpha_i \sigma_{i.}^2 - \alpha^* \sigma\sigma^* \alpha\right)T}, K, r, \sqrt{\alpha^*\sigma\sigma^*\alpha}, T\right)$$

where, the celebrated Black-Scholes formula is given by

$$\mathrm{Call}_{BS}(x_0, K, r, \sigma, \tau) = x_0 \Phi_0(d_1) - e^{-r\tau} K \Phi_0(d_2)$$

$$\text{with} \qquad d_1 = \frac{\log\left(\frac{x_0}{K}\right) + (r + \frac{\sigma^2}{2})\tau}{\sigma\sqrt{\tau}}, \quad d_2 = d_1 - \sigma\sqrt{\tau}.$$

Phase II: Joint simulation of $e^{-rT}(k_T, h_T)$

- Write in an explicit way

$$(k_T, h_T) = \Big(\psi(Z_{1:q}), \varphi(Z_{1:q})\Big), \quad Z_{1:q} \stackrel{d}{=} \mathcal{N}(0; I_q)$$

- The resulting pointwise estimator of the premium is given, with obvious notations, by

$$\mathrm{Call}_{BS}\left(\prod_{i=1}^{d} x_i^{\alpha_i} e^{-\frac{1}{2}\left(\sum_{i=1:d}\alpha_i\sigma_{i.}^2-\alpha^*\sigma\sigma^*\alpha\right)T}, K, r, \sqrt{\alpha^*\sigma\sigma^*\alpha}, T\right)+\frac{e^{-rT}}{M}\sum_{m=1}^{M} h_T^{(m)} - k_T^{(m)}$$

where

$$h_T^{(m)} - k_T^{(m)} = \varphi(Z_{1:q}^{(m)}) - \psi(Z_{1:q}^{(m)}) \quad \text{are independent copies of } h_T - k_T$$

and the $Z_{1:q}^{(m)}$ are simulated by the Box-Muller method (or the polar or Zigurat method. . .).

Asian option: Kemna-Vorst control variate

- Asian (path-dependent...) payoff $h_T = \varphi\left(\frac{1}{T-T_0}\int_{T_0}^{T} X_t^x dt\right) \geqslant 0$.
- Then if φ is increasing, Jensen's Inequality applied to the probability $\mathbf{1}_{[T_0,T]}\frac{dt}{T-T_0}$ (no interest rate) and the convex function exp yields

$$h_T \geq k_T = \varphi\left(exp\left(\frac{1}{T-T_0}\int_{T_0}^{T}\log(X_t^x)dt\right)\right) \geqslant 0.$$

- If $X_t^x = x\, e^{-\frac{\sigma^2}{2}t+\sigma W_t}$ (B-S!) then $\log X_t^x$ is a Gaussian process.
- Closed form for $\mathbb{E}\,\varphi(X_T^x)$ implies closed form for $\mathbb{E}\,k_T$.
- $\Xi = k_T$(pseudo- conrol variate) and hopefully $\mathrm{Var}(h_T - k_T) < \mathrm{Var}(h_T)$.
- Efficient if joint simulation of (h_T, k_T) is possible $i.e.$ of $\left(\int_{T_0}^{T} W_t dt, \int_{T_0}^{T} f(W_t)dt\right)$ is. But needs time discretization (see book).

Pros: Can be coded with the payoff in the pricer.

Cons: Needs a closed form hence brain time.

Dynamic control variate

- Let X, $X' \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ satisfying

$$\mathbb{E}\,X = \mathbb{E}\,X' = m_X \in \mathbb{R}\setminus\{0\}, \quad \mathrm{Var}(X), \mathrm{Var}(X'), \mathrm{Var}(X - X') > 0.$$

- **Question:** Which random vector (distribution. . .) is more appropriate?
- **Naive answer:** *if both X and X' can be simulated with a similar complexity*, the one with the lowest variance. . . *i.e.*

$$X \quad \text{if} \quad \mathrm{Var}(X) < \mathrm{Var}(X'), \quad X' \ \text{ otherwise}$$

$$\boxed{\text{Could we provide a more ambitious answer ?}}$$

- Like keeping the best from both?
- Let us start by a typical example from finance.

Parity equations in an *AOA* market (short background)

- Market satisfying *AOA* with (at least)

  – one *risky* asset $(S_t)_{t\in[0,T]}$ (with $S_0 = s_0$) defined on $(\Omega, \mathcal{A})$.

  – a *riskless* asset $S^0_t = e^{-rt}$ (for the sake of simplicity).

  – The/A *risk neutral probability*, denoted $\mathbb{P}$ satisfies

$$\widetilde{S}_t = e^{-rt} S_t \qquad \text{is a } \mathbb{P}\text{-martingale.}$$

- If $h_T \geqslant 0$ is a *European payoff*, (we already used that)

$$V_0 = e^{-rT}\mathbb{E}\, h_T.$$

Vanilla Call-Put parity

$$\mathrm{Call}_0 = e^{-rT}\mathbb{E}\big((S_T - K)^+\big) \quad \text{and} \quad \mathrm{Put}_0 = e^{-rT}\mathbb{E}\big((K - S_T)^+\big).$$

Hence

$$\begin{aligned}\mathrm{Call}_0 - \mathrm{Put}_0 &= \mathbb{E}\left[e^{-rT}\big((S_T - K)^+ - (K - S_T)^+\big)\right]\\ &= \mathbb{E}\left[e^{-rT}(S_T - K)\right]\\ &= \mathbb{E}\,\widetilde{S}_T - e^{-rT}K\end{aligned}$$

so that, by the martingale property,

$$\boxed{\mathrm{Call}_0 - \mathrm{Put}_0 = s_0 - e^{-rT}K}$$

since $(\widetilde{S}_t)_{t\in[0,T]}$ is a $\mathbb{P}$-martingale.

$$\mathrm{Call}_0 = \mathbb{E}\,X = \mathbb{E}\,X'$$

with

$$X := e^{-rT}(S_T - K)^+ \quad \text{and} \quad X' := e^{-rT}(K - S_T)^+ + s_0 - e^{-rT}K.$$

- $(K - S_T)^+ + s_0 e^{rT} - K$ is called a *synthetic call*.
- **Complexity.** As $s_0 e^{rT} - K$ is instantaneous to compute

$$\text{Complexity}(X, X') \simeq \text{Complexity}(X) \simeq \text{Complexity}(S_T).$$

Asian Call-Put parity

- At $T_0 \in [0, T)$ starts the *averaging* interval $[T_0, T]$.

$$\mathrm{Call}_0 = e^{-rT}\mathbb{E}\left(\left(\frac{1}{T-T_0}\int_{T_0}^{T} S_t dt - K\right)^+\right)$$

$$\mathrm{Put}_0 = e^{-rT}\mathbb{E}\left(\left(K - \frac{1}{T-T_0}\int_{T_0}^{T} S_t dt\right)^+\right).$$

- Using that $\widetilde{S}_t = e^{-rt}S_t$ is a $\mathbb{P}$-martingale and Fubini's theorem yield

$$\boxed{\mathrm{Call}_0^{As} - \mathrm{Put}_0^{As} = s_0\frac{1-e^{-r(T-T_0)}}{r(T-T_0)} - e^{-rT}K.}$$

so that

$$\mathrm{Call}_0^{As} = \mathbb{E}\, X = \mathbb{E}\, X'$$

with

$$X := e^{-rT}\left(\frac{1}{T-T_0}\int_{T_0}^{T} S_t dt - K\right)^+$$

$$X' := s_0\frac{1-e^{-r(T-T_0)}}{r(T-T_0)} - e^{-rT}K + e^{-rT}\left(K - \frac{1}{T-T_0}\int_{T_0}^{T} S_t dt\right)^+.$$

Complexity.

$$\mathrm{Complexity}(X, X') \approx \mathrm{Complexity}(X)$$

since computing once $s_0\frac{1-e^{-r(T-T_0)}}{r(T-T_0)} - e^{-rT}K$ is negligible w.r.t. simulating $\displaystyle\int_{T_0}^{T} S_t dt$.

Remark. In both cases, the key is the $\mathbb{P}$-*martingale property* of $\widetilde{S}_t$.

Variance reduction by linear regression

▷ General problem For every $\lambda \in \mathbb{R}$, set

$$\boxed{X^\lambda := (1-\lambda)X + \lambda X' \qquad \text{satisfies} \qquad \mathbb{E}(X^\lambda) = m}$$

$$\boxed{\text{Minimizing the (convex) function} \qquad \lambda \longmapsto \Phi(\lambda) := \mathrm{Var}(X^\lambda)}$$

▷ Reformulation with Ξ.

For convenience, set

$$\Xi := X - X' \qquad \text{so that} \qquad X^\lambda = X - \lambda\,\Xi.$$

Then, the convex function

$$\Phi(\lambda) = \lambda^2 \mathrm{Var}(\Xi) - 2\,\lambda\,\mathrm{Cov}(X, \Xi) + \mathrm{Var}(X)$$

atteins its minimum value at $\lambda_{\min}$ with

$$\boxed{\lambda_{\min} := \frac{\mathrm{Cov}(X, \Xi)}{\mathrm{Var}(\Xi)} = 1 - \frac{\mathrm{Cov}(X', \Xi)}{\mathrm{Var}(\Xi)}}$$

and

$$\boxed{\sigma^2_{\min} := \mathrm{Var}(X^{\lambda_{\min}}) = \mathrm{Var}(X) - \frac{(\mathrm{Cov}(X, \Xi))^2}{\mathrm{Var}(\Xi)} = \mathrm{Var}(X') - \frac{(\mathrm{Cov}(X', \Xi))^2}{\mathrm{Var}(\Xi)}}$$

Note that $\mathrm{Cov}(X, \Xi) = 0$ iff $\mathrm{Var}(X) = \mathrm{Cov}(X, X')$ iff $X = X'$. Hence

$$\boxed{\sigma^2_{\min} < \min\left(\mathrm{Var}(X), \mathrm{Var}(X')\right).}$$

More precisely,if $\rho_{X,\Xi}$ denotes the correlation coefficient of X and Ξ

$$\boxed{\sigma^2_{\min} = \mathrm{Var}(X)(1-\rho^2_{X,\Xi}) = \mathrm{Var}(X')(1-\rho^2_{X',\Xi}).}$$

A more symmetric expression for $\mathrm{Var}(X^{\lambda_{\min}})$ is

$$\begin{aligned}\sigma^2_{\min} &= \frac{\mathrm{Var}(X)\mathrm{Var}(X')(1-\rho^2_{X,X'})}{(\mathrm{Var}(X)-\mathrm{Var}(X'))^2+2\sqrt{\mathrm{Var}(X)\mathrm{Var}(X')}(1-\rho_{X,X'})}\\
&\leqslant \sqrt{\mathrm{Var}(X)\mathrm{Var}(X')}\,\frac{1+\rho_{X,X'}}{2}.\end{aligned}$$

How to use that ?

Back to Examples

- Vanilla Call-Put parity:

$$X = e^{-rT}(S_T - K)^+ \quad \text{and} \quad \Xi := e^{-rT}S_T - s_0.$$

and

$$\mathrm{Complexity}(X, \Xi) \approx \mathrm{Complexity}(X).$$

- Asian Call-Put parity:

$$\Xi := \frac{e^{-rT}}{T - T_0}\int_{T_0}^{T} S_t\,dt - s_0\frac{1 - e^{-r(T-T_0)}}{r(T - T_0)}.$$

and

$$\mathrm{Complexity}(X, \Xi) \simeq \mathrm{Complexity}(X) \simeq \mathrm{Complexity}\left(\int_{T_0}^{T} S_t\,dt\right).$$

- Antithetic variables: if, furthermore, $X \stackrel{d}{=} X'$, then

$$\mathbb{E}\, X\Xi = \mathbb{E}\, X^2 - \mathbb{E}\, XX'$$

and

$$\begin{aligned}\mathbb{E}\,\Xi^2 = \mathbb{E}\,(X - X')^2 &= \mathbb{E}\, X^2 - 2\mathbb{E}\, XX' + \mathbb{E}(X')^2 \\ &= 2\left(\mathbb{E}\, X^2 - \mathbb{E}\, XX'\right) = 2\,\mathbb{E}\, X\Xi.\end{aligned}$$

Hence

$$\lambda_{\min} = \frac{1}{2},\; X^{\lambda_{\min}} = \frac{X + X'}{2} \text{ and } \sigma^2_{\min} = \tfrac{1}{2}(\mathrm{Var}(X) + \mathrm{Cov}(X, X')) < \mathrm{Var}(X).$$

Example: $X = f(Z)$ and $Z \stackrel{d}{=} L - Z$, $(L \in \mathbb{R})$,

$$X = f(Z), \qquad \Xi := f(Z) - f(L - Z) \quad \text{with} \quad \mathbb{E}\,\Xi = 0.$$

Warning! If $\mathrm{Complexity}(Z) \ll \mathrm{Complexity}(f(Z)\,|\,Z)$ then

$$\mathrm{Complexity}(X, \Xi) \approx 2\times\mathrm{Complexity}(X)\ldots$$

[To be continued. . .]

The naive approach

- Let (X_k, X'_k), $k \geqslant 1$, be (simulated...) independent copies of (X, X'), $\Xi_k = X_k - X'_k$
- Set, for (a large enough *fixed*) integer $M \geqslant 1$:

$$V_M := \frac{1}{M}\sum_{k=1}^{M} \Xi_k^2 \qquad C_M := \frac{1}{M}\sum_{k=1}^{M} X_k \Xi_k$$

$$\lambda_M := \frac{C_M}{V_M}$$

- Then $\lambda_M \to \lambda_{\min}$ $\mathbb{P}$-$a.s.$ so that

$$\frac{1}{M}\sum_{k=1}^{M} X_k^{\lambda_M} = \overline{X}_M - \lambda_M \overline{\Xi}_M \xrightarrow{a.s.} \mathbb{E}X = m_X \quad \text{as} \quad M \to +\infty.$$

- **Cons**: biased estimator since $\mathbb{E}\left(\lambda_M \overline{\Xi}_M\right) \neq 0$.

The "batch" approach (Glasserman)

- Compute a rough estimate $\lambda_{M'}$ of $\lambda_{\min}$ on a (small) sample of size $M' \ll M$ using the above formulas.
- Let (X_k, X'_k), $k \geqslant 1$, be (simulated...) NEW independent copies of (X, X'), $\Xi_k = X_k - X'_k$ (hence independent of $\lambda_{M'}$). Then compute

$$\frac{1}{M}\sum_{k=1}^{M} X_k^{\lambda_{M'}} = \overline{X}_M - \lambda_{M'}\,\overline{\Xi}_M \xrightarrow{a.s.} \mathbb{E}X = m_X \quad \text{as} \quad M \to +\infty.$$

- This time $\mathbb{E}\left(\lambda_{M'}\,\overline{\Xi}_M\right) = 0$ and there is no more bias.

Recursive implementation

- We admit that
$$\lambda_M = \frac{C_M}{V_M} \xrightarrow{L^2(\mathbb{P})} \lambda_{\min}.$$
- Note that (C_M) and (V_M) can be computed recursively from (C_{M-1}, V_{M-1}) and (X_M, X'_M).
- Then set for every $k \geqslant 1$,

$$\boxed{X''_k = X_k - \lambda_{k-1}\Xi_k = (1-\lambda_{k-1})X_k + \lambda_{k-1}X'_k.}$$

- Let $\mathcal{F}_M := \sigma(X_1, X_1', \ldots, X_M, X_M')$, $M \geqslant 1$, denote the filtration of the simulation. For every $k \geqslant 1$,

$$\begin{aligned}
\mathbb{E}\big(X_k'' \,|\, \mathcal{F}_{k-1}\big) &= \mathbb{E}(X_k \,|\, \mathcal{F}_{k-1}) - \lambda_{k-1}\mathbb{E}(\Xi_k \,|\, \mathcal{F}_{k-1}) = m \\
\mathbb{E}\big((X_k'' - m)^2 \,|\, \mathcal{F}_{k-1}\big) &= \mathrm{Var}(X_k^\lambda)_{|\lambda=\lambda_{k-1}} \\
&= \Phi(\lambda_{k-1}) \overset{k\to\infty}{\longrightarrow} \Phi(\lambda_{\min}) = \min_\lambda \Phi(\lambda).
\end{aligned}$$

- Let $\overline{X}''_M = \dfrac{1}{M}\displaystyle\sum_{k=1}^{M} X_k''$. Standard arguments from martinagle theory show

Theorem (Unbiased estimator and *CLT*)

(a) $\overline{X}''_M \to m$ as $M \to +\infty$.

(b) $\sqrt{M}(\overline{X}''_M - m) \xrightarrow{\mathcal{L}} \mathcal{N}(0; \sigma^2_{\min})$ as $M \to +\infty$.

- Keep in mind that:

  – In our "Parity" examples Ξ is involved in the simulation of X.

  Hence the complexity of the simulation process *is not* increased.

  – Updating λ_M and (the empirical mean) $\overline{X}''_M$ is (almost) costless
  Hence,

  The complexity remains the same.

  Warning ! This not rue for antithetic. . .

- When the complexity is doubled, the method is efficient iff

$$(\star) \qquad \sigma^2_{\min} < \frac{1}{2}\min(\mathrm{Var}(X), \mathrm{Var}(X')).$$

- Antithetic method:
  - $\mathrm{Var}(X) = \mathrm{Var}(X')$
  - $\sigma^2_{\min} = \frac{1}{2}(\mathrm{Var}(X) + \mathrm{Cov}(X, X'))$
  - hence $(\star)$ is satisfied iff $\mathrm{Cov}(X, X') < 0$
  - Ok e.g. if $X = \varphi(Z)$, $\varphi \uparrow$ or $\downarrow$ and $Z \stackrel{d}{=} T(Z)$, $T \downarrow$, $X' = \varphi(T(Z))$.

Vanilla Black-Scholes Calls: regular versus synthetic

Of course a toy-exercise !

▷ Model parameters:

$$\boxed{T = 1,\ s_0 = 100,\ r = 5\,\%,\ \sigma = 20\,\%,\ K = 90, \ldots, 120.}$$

▷ *MC* parameter:

$$\boxed{M = 10^6}$$

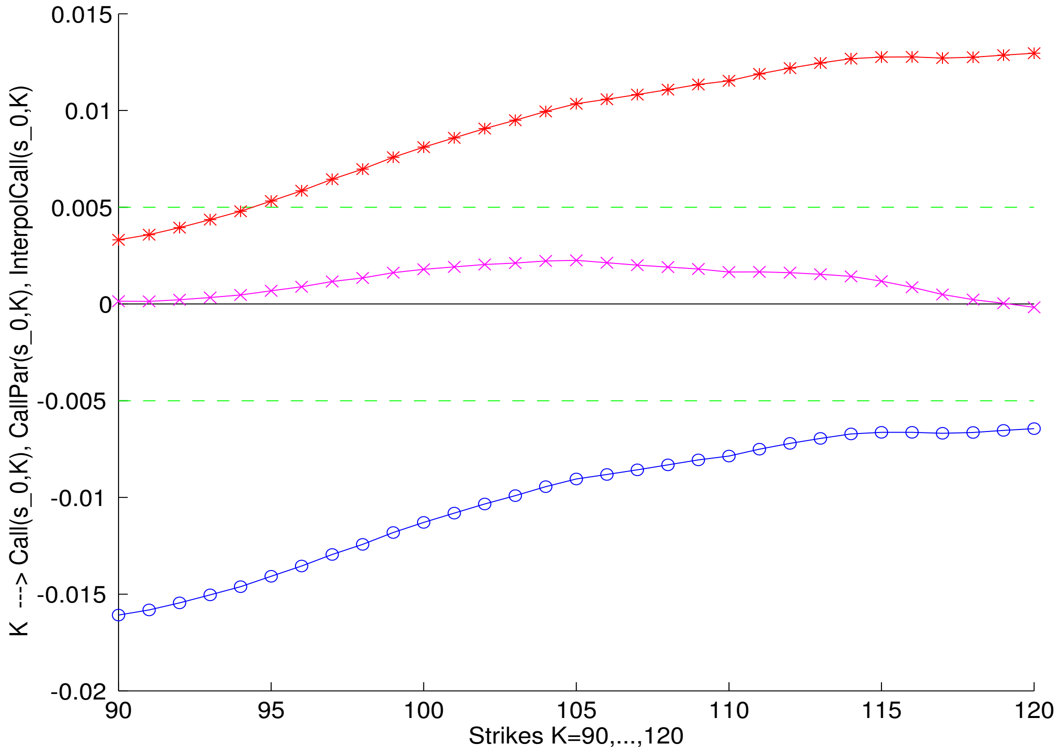

Figure: Black-Scholes Calls: *Error= Reference BS−(MC Premium).*
$K = 90, \ldots, 120$. –o–o–o– *Crude Call (X).* –∗–∗–∗– *Synthetic Parity Call (X').*
–×–×–×– *Interpolated synthetic Call (X'').*