Fast Convergence of “Regress-later” Series Estimators

New Thinking in Finance, London

Eric Beutner, Antoon Pelsser, Janina Schweizer

Maastricht University & Kleynen Consultants

12 February 2014

Solvency II challenges insurers to establish appropriate risk models

- Financial institutions are required to enhance their understanding of the risks they are taking.
- Solvency II Pillar I requires insurers to compute accurate risk capital figures.
- Extracting risk figures from balance sheets involves complicated computation:
  - Generate “outer” scenarios for 1 year VaR calculation
  - For each outer scenario calculate price of all items on balance sheet
  - Full Monte-Carlo leads to simulation-in-simulation

Solvency II challenges insurers to establish appropriate risk models

- Financial institutions are required to enhance their understanding of the risks they are taking.
- Solvency II Pillar I requires insurers to compute accurate risk capital figures.
- Extracting risk figures from balance sheets involves complicated computation:
  - Generate "outer" scenarios for 1 year VaR calculation
  - For each outer scenario calculate price of all items on balance sheet
  - Full Monte-Carlo leads to simulation-in-simulation

Solution

Map balance sheet items into simplified functions which can computed analytically. This removes the "inner" simulation.

Least-Squares Monte-Carlo

- This is a similar method as the Least-Squares Monte-Carlo (LSMC) widely used in finance.
- For example for pricing of American-style options with MC. Use regression method to estimate the "continuation value"
- LSMC also used for numerical solution of Backward Stochastic Differential Equations (BSDE's).
- LSMC techniques are used here to compute the solution process Y_t and the "gradient process" Z_t.
- Convergence of LSMC has been studied in the literature (e.g. Stentoft, 2004).

Agenda

Very Brief Literature Review

- Madan and Milne (1994): static replication in Hilbert space.
- Longstaff and Schwartz (2001): LSMC for American option pricing.
- Stentoft (2004): convergence of LSMC **Regress-Now** estimator.
- Glasserman and Yu (2002) suggest **Regress-Later** estimator.

Hilbert space theory

- W_t, $0 \leq t \leq T$, is an underlying random process.
- X_T gives the payoff or value function) at time T contingent on W_T.
- Consider target function of the form $g(W_t) := \mathbb{E}(X_T | W_t)$.
- Space of all square-integrable payoff functions is given by Hilbert space $L_2(\Omega, \mathcal{F}, \mathbb{P})$.
- From Hilbert space theory the function $g()$ can be written as

$$g(W_t) = \sum_{k=0}^{\infty} \alpha_k e_k(W_t)$$

.

- Infinite dimensional space with **countable** basis, e.g. monomials.

How to estimate $g()$?

- Non-parametric estimation problem in infinite(!) dimensional space
- Econometricians have studied this class of estimation problems
- Approximate true function as sequence of finite-dimensional sums.
- Challenge of two limits: truncation K and sample size N.
- Solution: theory for **sieve estimators** and/or **empirical processes** gives conditions and convergence rates.
- Alternative: literature on training of neural networks.

Method of sieve gives a two-step estimator

Approximation

The finite-dimensional linear sieve is

$$H_K := \Big\{ g : \Omega \to \mathbb{R},$$

$$g_K(y^t) = \sum_{k=1}^{K} \alpha_k e_k(y^t) : \ \alpha_1, ..., \alpha_K \in \mathbb{R} \Big\}$$

with $\dim(H_K) = K \to \infty$ slowly as $N \to \infty$. The series estimator of $g(W_t)$ is then

Estimation

$$\hat{g} = \arg \min_{g \in H_K} \frac{1}{N} \sum_{i=1}^{N} \left(g(W_t) - g_K(W_t) \right)^2 .$$

Two estimators should be distinguished

Regress-now

Estimate $g(W_t) = \mathbb{E}[X_T | W_t]$ as

$$\hat{g}(W_t) = \boldsymbol{e}^{\boldsymbol{K}}(\boldsymbol{W_t})'\hat{\alpha}_{\text{now}}.$$

- Directly fits the pricing function.
- Applies a smoothing before estimation.
- Is model-dependent: changing the pricing measure yields a new pricing function.

Two estimators should be distinguished

Regress-now

Estimate $g(W_t) = \mathbb{E}[X_T | W_t]$ as

$$\hat{g}(W_t) = \boldsymbol{e}^{\boldsymbol{K}}(\boldsymbol{W_t})'\hat{\alpha}_{\text{now}}.$$

- Directly fits the pricing function.
- Applies a smoothing before estimation.
- Is model-dependent: changing the pricing measure yields a new pricing function.

Regress-later

Est. $\hat{g}(W_T) = \boldsymbol{e}^{\boldsymbol{K}}(\boldsymbol{W_T})'\hat{\alpha}_{\text{lat}}$,

$$\hat{g}(W_t) = \mathbb{E}[\boldsymbol{e}^{\boldsymbol{K}}(\boldsymbol{W_T})' | W_t]\hat{\alpha}_{\text{lat}}$$

- First fits the payoff function.
- Compute cond.exp. of basis analytically.
- Is model-independent; changing the pricing measure does not affect the composition of the fitting function.

Assumptions and Conditions

Usual non-parametric assumption:

- The maximal approximation error, $\left|g(y^T) - g_K(y^T)\right|$, must diminish with $\mathcal{O}(K^{-\gamma})$
- Weak assumption, that does not depend on measure $\mathbb{P}$.

Within MC we **know** the data-generating process

- Use stronger assumption: $\sqrt{\mathbb{E}\left[\left(g(y^T) - g_K(y^T)\right)^4\right]} = \mathcal{O}(K^{-\gamma})$.
- Depends explicitly on measure $\mathbb{P}$.
- Define the net $h(K,N) := \frac{1}{N}\mathbb{E}\left[\left(\boldsymbol{e}^{\boldsymbol{K}\prime}\boldsymbol{e}^{\boldsymbol{K}}\right)^2\right]$, this is the variance of the finite-sample covariance matrix of the basis functions.
- Assume there is a sequence $K(N)$, such that $h(N, K(N)) \to 0$ for $N \to \infty$.

Difference in "regress-now" and "regress-later"

Theorem:

- Regress-later mean square error converges as: $\mathcal{O}_p(K(N)^{-\gamma})$
- Regress-now mean square error converges as: $\mathcal{O}_p(K/N + K^{-\gamma})$

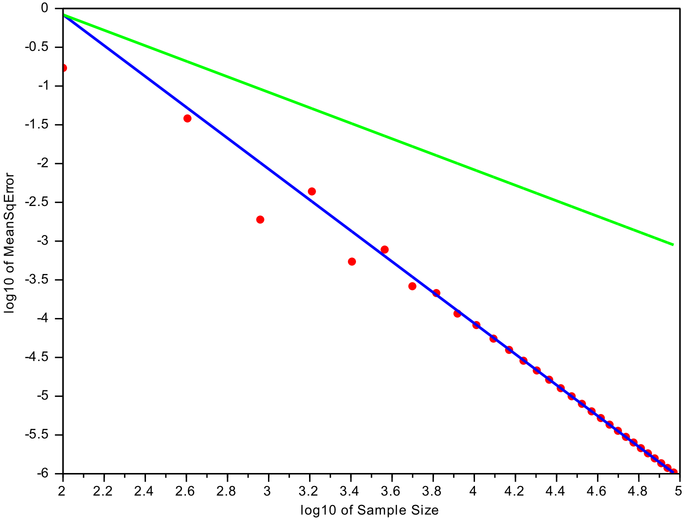
- Regress-now exhibits an additional error linked to projection of X_T on smaller filtration $\mathcal{F}_t$. Regress-later avoids this error.
- Regress-now asymptotically attains Stone's bound for optimal choice $K(N) = N^{\frac{1}{\gamma+1}}$, giving convergence of $\mathcal{O}_p(N^{\frac{-\gamma}{\gamma+1}})$.
- With Regress-Later we can "break through" Stone's bound and converge faster than $\mathcal{O}_p(N^{-1})$.

Piecewise linear functions give easy basis

- Chop domain into K intervals: $\{(b_1, b_2], (b_2, b_3], \ldots\}$. Consider on each interval a linear function as a basis-function.
- If $g()$ is twice differentiable, then $\gamma = 4$.
- Choose $K(N) = N^{0.499}$, then $h(N, K(N)) \to 0$.
- Convergence in mean square error thus $\mathcal{O}_p(N^{-1.996})$ which is considerably faster than MC convergence $\mathcal{O}_p(N^{-1})$.
- We conjecture that even faster convergence can be achieved for more optimised bases.

Fast convergence with Regress Later

Green line: $\mathcal{O}(N^{-1})$, blue line $\mathcal{O}(N^{-2})$.



Extension to multiple time-steps

- The result we have show here is for one single time-step.
- This is sufficient for risk calculations over one single horizon.

- For pricing American-style options, we should consider multiple time-steps.
- We have potential feed-forward of approximation errors in the algorithm.
- Results from previous regressions, are basis for next regression.
- Topic of ongoing research at the moment.

Extension to multivariate path-dependent claims

- Every contingent claim can be modelled as a function of d-dimensional stochastic process $\boldsymbol{W}(t) = (W_1(t), ..., W_d(t))'$; $0 \leq t \leq T$.
- "Mild" path-dependency is handled by adding "summary variables" (e.g. running maximum or partial average) as additional stochastic processes.
- In full generality, path-dependency can be handled by chopping up time and adding intermediate values as additional stochastic processes. (Same idea for construction of stochastic integral)
- The multivariate basis is given by the product of the univariate bases.
- Note: we still have a countable basis.

Naive multivariate basis does not work

Total number of parameters to be estimated for a replication up to maximum order K^*

$$\sum_{k^*=0}^{K^*} \binom{d+k^*-1}{k^*} = \binom{K^*+d}{K^*} = \frac{[K^*+d]!}{K^*!d!}.$$

Example:
$K^* = 2$, $d = 6 \times 10 = 60$, number of terms: 1891 $\Rightarrow$ **Curse of dimensionality!**

Solution:

- Only consider "mild" path-dependent products in low dimensions
- Possible alternative: sparse bases.
- Results about "universal approximation" in machine learning.

Conclusion

- Regression based LSMC very important for numerical algorithms in finance and insurance.
- Most implementations based on Regress-Now approach.
- We investigate Regress-Later, and shows that is has fundamentally different properties.
- We prove that it is possible to achieve fast convergence speeds with Regress-Later.
- Show explicit example of convergence in MSE of $\mathcal{O}(N^{-2})$.

Some References

- Glasserman, P. and Yu, B. (2002). Simulation for American Options: Regression now or Regression later? Monte Carlo and Quasi-Monte Carlo Methods.
- Longstaff, F. A. and Schwartz, E.S. (2001). Valuing American Options by Simulation: A simple least-squares approach. Review of Financial Studies, 14(1):113-47.
- Madan, D.B. and Milne, F. (1994). Contingent claims valued and hedged by pricing and investing in a basis. Mathematical Finance, 4(3):223-245.
- Newey, W.K. (1997). Convergence rates and asymptotic normality for series estimators. Journal of Econometrics, 79(1):147-168.
- Stentoft, L. (2004). Convergence of the Least Squares Monte Carlo Approach to American Option Valuation. Management Science, 50(9):1193-1203.