

Portfolio replication and least squares Monte Carlo with application to insurance risk management

Citation for published version (APA):

Schweizer, J. (2016). *Portfolio replication and least squares Monte Carlo with application to insurance risk management*. [Doctoral Thesis, Maastricht University]. Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20161007js>

Document status and date:

Published: 01/01/2016

DOI:

[10.26481/dis.20161007js](https://doi.org/10.26481/dis.20161007js)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

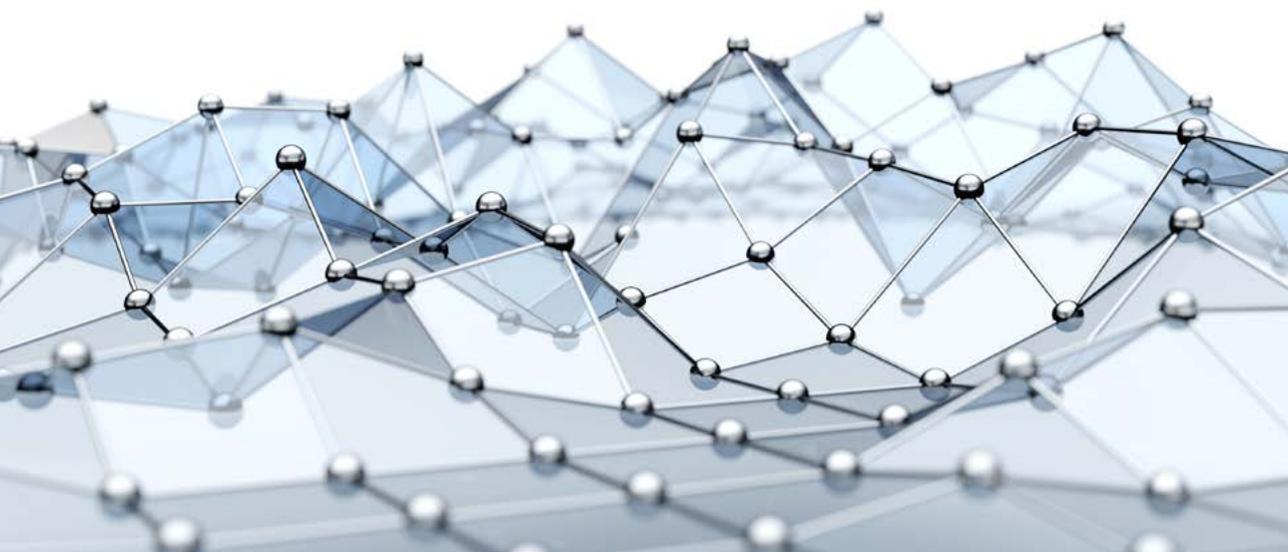
providing details and we will investigate your claim.

Download date: 25 Apr. 2024

Portfolio Replication and Least Squares Monte Carlo

with Application to Insurance Risk
Management

Janina Schweizer



© Janina Schweizer, Maastricht 2016

Cover picture: © Robert Kneschke / Fotolia

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission in writing from the author.

This book was typeset by the author using L^AT_EX.

Published by Universitaire Pers Maastricht

ISBN: 978 94 6159 585 0

Printed in The Netherlands by Datawyse Maastricht



Portfolio Replication and Least Squares Monte Carlo with Application to Insurance Risk Management

DISSERTATION

to obtain the degree of Doctor at
Maastricht University,
on the authority of the Rector Magnificus,
Prof. dr. R.M. (Rianne) Letschert,
in accordance with the decision of the Board of Deans,
to be defended in public
on Friday, October 7th, 2016 at 12.00 o'clock

by

Janina Schweizer

Supervisor:

Prof. dr. A.A.J. Pelsser

Co-Supervisor:

Dr. E.A. Beutner

Assessment Committee:

Prof. dr. P.C. Schotman (Chair)

Prof. dr. J.M. Schumacher (Tilburg University)

Prof. dr. J.R.Y.J. Urbain

Prof. dr. R. Werner (Augsburg University)

This research was financially supported by Allianz SE.

To my parents, who always see the best in me.

Acknowledgements

“Don’t Panic.”

– Douglas Adams, *The Hitchhiker’s Guide to the Galaxy*

ACKNOWLEDGEMENTS

During my second Bachelor year in International Business Economics at Maastricht University I started to think about my next steps. I fancied Econometrics, but was uncertain whether my theoretical background would suffice to take the step of pursuing a Masters in Econometrics. However, I was convinced that if I wanted to obtain a better understanding of Economics and particularly Finance, improving my technical skills would be crucial. I decided to give it a shot, knowing that a challenging year would lie ahead of me¹. And, I was right. But the journey had just started. Midway through the Masters I considered a Ph.D. However, I wanted to combine theory and practice, making sure that the topic I would be working on, would be of practical relevance. Half a year later I found myself sitting in the office of the central risk management function at Allianz SE in Munich, from where I would work for the next couple of years on Replicating Portfolios as an aspiring external Maastricht University Ph.D. candidate. A journey started, of which I had no idea what it would bring. Pursuing a Ph.D. is like attempting to climb a mountain, but there is no clear road on how to get to the top of it. Sometimes, you can see the tip of the mountain behind some misty clouds, but then you hit a detour and have to find a new path up the mountain. I started to realize that doing your Ph.D. is much more than acquiring new knowledge and finding solutions to new problems. It is your chance to grow as a person, to take a lifelong lesson of not giving up just because a bumpy road lies ahead, to re-focus and to believe in yourself. This journey would have been nothing without the people that supported me throughout and I would like to take the opportunity to thank them.

First of all, I want to thank my wonderful supervisors Antoon Pelsser and Eric Beutner for their continuous support and for enabling me to pursue a Ph.D. while gaining practical experience in the industry.

¹Special thanks goes to Dries Vermeulen for his support in making this possible.

Antoon, thank you for showing me to focus on my strengths and to always view the glass as being half full, rather than half empty. Eric, thank you for your incredible patience when it came to my questions. I learned a lot climbing up that mountain.

My gratitude goes to my former boss, Tobias Herwig, and my former colleagues at Allianz SE, particularly Deepak Pandey, Christian Brünger and Tobias Knupfer. Tobias H., thank you for all the freedom you gave me in my research while reminding me of what is really relevant in practice. Deepak, thank you for listening and helping me out when I got stuck. I had so much fun working with you. Christian and Tobias K., I learned so much working together with you. I would also like to thank Axel Seeman for helping me getting acquainted with the existing theory on Replicating Portfolios. My thanks also goes to Marco Hauck and Tom Wilson for allowing me to pursue my Ph.D. at Allianz. Finally, I want to thank all colleagues of the former Middle Office that accompanied me during my journey, especially Markus Hannemann, Christoph Renner, Svenja Hager, Beate Delp, to only name a few.

Furthermore, I would like to thank my (former) fellow Ph.D. students Anne Balter, Jan Natolski, Oana Floroui, Sally Shen. Anne, thank you so much for answering the plenty of questions I had regarding the Ph.D. procedure. Jan, I very much enjoyed our discussions on Replicating Portfolios.

My special thanks goes to Ralf Werner. I very much appreciated our talks on Replicating Portfolios and it always gave me a new perspective. Also, I would like to thank Hans Schumacher, for carefully reading a first version of my first paper for the Netspar Pension day. Your feedback was highly appreciated and very much helped to improve the thesis.

ACKNOWLEDGEMENTS

I would like to express my appreciation to the reading committee, Peter Schotman, Hans Schumacher, Jean-Pierre Urbain and Ralf Werner. Thank you for your careful reading of the thesis and your very valuable comments.

Moreover, I want to thank my friends, Lena Schneckenburger, Yamina Agharbi, Jasmine Ngai, Tatjana Vidusenko, among others, for making this journey much more fun. Uli Gerlach, you were the first to encourage me to take this route, and while I may not have been grateful at times for that, I am surely now. Moreover, I would like to mention Wun Hop Kuen Do TSV Waldtrudering, in which I train. You have shown me once again that hard work does pay off.

Finally, I want to thank my family for helping me through that journey. I thank my parents for their unconditional support and for always understanding me. I think, I have still not been very successful in explaining the content of this thesis to you, but trust me, you were a major contribution. Special thanks goes to my boyfriend Jens Erlewein who has endured me in the stressful times, dealing with a highly perfectionist girlfriend (who is unfortunately just not that perfect) and who has never stopped believing in me. Thank you for sharing your values with me and helping me to keep focus on what is really important. And, I know I am going to regret saying this, but you were right.

Without any of you, none of this would have been possible. Thank you.

Janina Schweizer

Munich, June 2016

Contents

Acknowledgements	vii
Contents	x
1 Introduction	1
2 Fast Convergence of Regress-Later Estimates in Least Squares Monte Carlo	9
2.1 Introduction	10
2.2 Mathematical Model for RN and RL	14
2.2.1 Regress-Now	17
2.2.2 Regress-Later	22
2.3 Convergence Rates for RL with sieves	26
2.4 Orthonormal piecewise linear functions as sieves	37
2.5 Conclusion	44
2.A Appendix: Proofs	45
3 Theory and Validation of Replicating Portfolios in Insurance Risk Management	55
3.1 Introduction	57

CONTENTS

3.2	Mathematical model for RPs	63
3.3	General asymptotic convergence of RPs	73
3.4	Orthonormal basis of piecewise linear functions	78
3.5	Application to path-dependent insurance policies	84
3.5.1	The Grosen and Jørgensen (2000) profit-sharing model	87
3.5.2	Outlook: Replication with generalized Asian op- tions	92
3.6	Conclusion	97
3.A	Appendix: Proofs	99
4	The Difference between LSMC and Replicating Portfo- lio in Insurance Liability Modeling	105
4.1	Introduction	106
4.2	Regression model for LSMC and PFREP	112
4.2.1	Least Squares Monte Carlo	115
4.2.2	Portfolio Replication	120
4.3	Impact of the zero projection error in RPs	126
4.3.1	Function fitting versus Portfolio Replication	126
4.3.2	Upper limit of R-square: LSMC vs PFREP	142
4.3.3	Asymptotic covariance with fixed truncation pa- rameter	146
4.3.4	Asymptotic measure independence in PFREP	152
4.3.5	Asymptotic convergence	165
4.4	Path-dependent and high-dimensional target functions	170
4.5	Conclusion	180
5	Conclusion	183
	Bibliography	191

CONTENTS

Valorization	201
Curriculum Vitae	209

List of Figures

2.1	Regress-Later convergence plot with K up to 30.	42
2.2	Regress-Later convergence plot with $K = 5$ fixed.	42
3.1	Nested stochastic simulation problem.	58
3.2	The Replicating Portfolio approach.	71
3.3	Components of piecewise linear function.	83
3.4	Empirical convergence rate with piecewise linear functions.	92
3.5	Empirical convergence rate for replication with Asians.	95
3.6	Comparison of variation with asset process at time $t = 10$	96
3.7	Comparison of variation with asset process at time $t = 5$	96
3.8	Comparison of variation with asset process at time $t = 1$	97
3.9	Illustration of goodness of fit of the replicating portfolio.	97
4.1	Nested stochastic simulation problem.	107
4.2	Illustration of the LSMC approach.	119
4.3	Payoff function at maturity for a call with maturity $T = 2$	122
4.4	Pricing function at $t = 1$ for a call with maturity $T = 2$	122
4.5	Illustration of the Replicating Portfolio approach	124
4.6	Noisy regression in LSMC (Example 13).	131
4.7	Regression in portfolio replication (Example 13).	131

LIST OF FIGURES

4.8	Noisy regression in LSMC (Example 14).	133
4.9	Regression in portfolio replication (Example 14).	133
4.10	LSMC fit for $N = 1\,200$ (Example 15).	135
4.11	LSMC fit for $N = 1\,000\,000$ (Example 15).	135
4.12	Missing risk factors regression in portfolio replication (Example 16).	136
4.13	Missing risk factors regression in LSMC (Example 16).	136
4.14	Illustration of LSMC fit at $t = 4$ with calibration at time 1 (Example 17).	140
4.15	LSMC fit at $t^* = 1$ given calibration at $t^* = 1$ (Example 18).	141
4.16	LSMC fit at $t = 3$ given calibration at $t^* = 1$ (Example 18).	141
4.17	LSMC fit with calibration on correct measure (Example 22).	157
4.18	LSMC fit with calibration on shifted normal (Example 22).	157
4.19	LSMC fit with calibration on correct measure (Example 23).	158
4.20	LSMC fit with calibration on uniform (Example 23).	158
4.21	Replication fit with calibration on correct measure (Ex- ample 23).	159
4.22	Replication fit with calibration on uniform (Example 23).	159
4.23	Payoff function of X (Example 24).	160
4.24	LSMC fit with calibration on uniform (Example 24).	163
4.25	Replication fit with calibration on uniform (Example 24).	163
4.26	LSMC fit with calibration on set three (Example 24).	164
4.27	LSMC fit with calibration on set four (Example 24).	164
4.28	Replication fit with calibration on set three (Example 24).	165
4.29	Replication fit with calibration on set four (Example 24).	165

4.30	Goodness of fit of the replicating portfolio (Example 28).	179
4.31	Comparison of variation with asset process at $t = 10$ (Example 28).	179
4.32	Comparison of variation with asset process at $t = 5$ (Example 28).	180
4.33	Comparison of variation with asset process at $t = 1$ (Example 28).	180

List of Tables

3.1	Fund-specific asset-liability interaction view at time t	88
3.2	Parameters for Example 12.	91
3.3	Parameters for replication with generalized Asians.	94
4.1	Parameters for Example 15.	133
4.2	Eigenvalues of $\Sigma_{lsmc} - \Sigma_{RP}$ for different K	152
4.3	Calibration sets for Example 24	161
4.4	Results for Example 24	162
4.5	Comparison portfolio replication versus LSMC.	181

Chapter 1

Introduction

“Our comforting conviction that the world makes sense rests on a secure foundation: our almost unlimited ability to ignore our ignorance.”

– Daniel Kahneman

Ever since the introduction of the Solvency II directive in 2009 the insurance industry has been busy developing appropriate quantitative risk models for calculating their risk capital requirements. While the Solvency II framework offers a standard model to the calculation of risk capital requirements, particularly large insurers opt for an internal model to better represent their business structure. This has triggered the demand for risk management solutions customized to the needs of the insurance industry. Particularly the enforcement of a market-consistent risk model (see Article 76, The European Parliament and The Council, 2009) has challenged insurers to come up with valuation techniques of their balance sheet items. Insurance policies are generally not traded such as other financial products, meaning that market prices are not available. Moreover, many insurance policies exhibit complex dynamic structures as they contain embedded options and guarantees coming from minimum guarantees, policyholder participations and surrender options. For these reasons closed-form solutions to their value are mostly not available. This especially applies to life insurance products, where very long maturities such as 60 years, for example, are nothing unusual. As a consequence numerical methods involving Monte Carlo techniques for estimating the value of insurance liabilities have gained much attention.

In the Solvency II context the relevant measure for the solvency capital requirement is “ [...] the Value-at-Risk of the basic own funds of an insurance or reinsurance undertaking subject to a confidence level of 99.5% over a one-year period.” (see SCR 1.9, EIOPA, 2012). The numerical calculation of the VaR requires the derivation of the empirical distribution of the value of the insurer’s own funds in one year. For that, real-world scenarios are constructed which simulate all

underlying risk factors up to one year¹. As closed-form solutions to the value of the liabilities are in principle not available, inner valuation scenarios must be generated per outer scenario node to obtain an estimate of the value of the liability at each outer node. The described approach is often referred to as full Monte Carlo simulation or nested stochastic Monte Carlo (see, for example, Bauer et al., 2009, for a detailed description). The difficulty with nested stochastic Monte Carlo is the resulting simulation effort, which is usually very high and in many cases infeasible in practice. For example, for a large insurance company with many underlying risk factors it is common to use 30,000-50,000 outer scenarios. Additionally, around 1,000 inner scenarios should at least be applied, resulting in a total simulation effort of 30-50 million scenarios. Depending on the size and complexity of the insurance company running such a high number of scenarios on asset-liability projection tools takes far too long and timely relevant risk capital reporting will thus not be possible. This especially applies to life insurance products, which are typically characterized by long maturities and embedded options². For that reason alternative methods have been explored, which combine approximation methods with Monte Carlo techniques with the ambition to yield accurate risk capital figures within a reasonable time frame.

The “nesting” in the full Monte Carlo simulation approach arises from the necessity to estimate the value of the liabilities at each outer node. The value of the liabilities at each outer node basically corresponds to an unknown conditional expectation function across time under some relevant probability measure. The simulation effort of the full Monte Carlo simulation approach could be significantly reduced,

¹In practice insurers often refrain from rolling their balance sheet forward one year, but rather consider the current balance sheet and apply instantaneous shocks.

²It should be mentioned that the issues and solutions presented in this thesis mainly pertain to life insurance products and P&C products similar to life.

if approximations to the conditional expectation function replaced the inner valuation simulations. The proxy function to the conditional expectation function would then be valued at each outer scenario node, thereby significantly reducing the simulation effort to the number of outer scenarios plus the calibration effort for constructing the proxy.

(Life) Insurance contracts largely exhibit features similar to (exotic) financial products. For example, many insurance policies are combinations of minimum guarantees and bonus options, which in principle is similar to the combination of a Zero Coupon Bond and a call option, where we assume for simplicity that the payoff is paid out at maturity of the contract regardless of survival or death of the policyholder. Naturally, insurers have thus turned to the financial pricing literature in search for solutions to the (life) insurance valuation problem. As a result two concepts are now widely applied for (life) insurance liabilities, both ultimately yielding an approximation to the unknown liability valuation function: Least Squares Monte Carlo (LSMC) and static portfolio replication.

LSMC originates from the pricing of Bermudan and American options, for which closed-form solutions are not available. The main difficulty in Bermudan and American option pricing is the calculation of the continuation value, which is basically an unknown conditional expectation function across time. This is reminiscent of the (life) insurance valuation problem, where the interest is to estimate an unknown conditional expectation function in a single-period context³. LSMC offers a solution to the estimation of unknown conditional expectation functions by combining Monte Carlo simulation with least squares

³It should be remarked that (life) insurance policies may also have features similar to American-style options. For example, the option to surrender on the contract is often given. However, these options often come with a high penalty. For simplicity American-style features are neglected in this paper and focus is put on the estimation of an unknown conditional expectation in a single-period context.

regression. The LSMC technique is widely applied and extensively discussed in the financial literature. Examples may be found in Carriere (1996), Longstaff and Schwartz (2001), Tsitsiklis and Van Roy (2001), Clement et al. (2002), Stentoft (2004), Glasserman and Yu (2004b), Egloff et al. (2007), Belomestny (2011), Gerhold (2011) and Zanger (2013). These papers on LSMC have in common that the unknown conditional expectation function at a time point t is approximated through least squares regression of the value function at time $T > t$ against basis functions at the earlier time point t . Glasserman and Yu (2004b) term this LSMC approach “regression now” and introduce an alternative LSMC approach termed “regression later”. The “regression now” version of the LSMC approach is well-known and its properties have been widely examined. Stentoft (2004), for example, derives the convergence rate of the (Regress-Now⁴) LSMC approach to the valuation of American options.

The Regress-Later version of LSMC has gained less attention than its Regress-Now sibling. While with Regress-Now an approximation to the value function at time t is attained in one step, Regress-Later achieves a representation of the conditional expectation function at time t in two steps. First the value function at time T is approximated by regressing the value function at time T against basis functions at the same time point. Subsequently the time t conditional expectation function is applied to the approximation function, which requires the calculation of the conditional expectation function of each basis term. As will later be seen Regress-Later corresponds to portfolio replication. Glasserman and Yu (2004b) show that in single-period problems Regress-Later yields more accurate estimates than Regress-Now. Similar observations are also reported by Broadie and Cao (2008)

⁴In this thesis we mostly use the terminology Regress-Now and Regress-Later instead of “regression now” and “regression later”.

and Bender and Steiner (2012). Clearly the Regress-Later estimator offers potential advantages, but its properties have not yet been sufficiently investigated to get a more comprehensive understanding of its advantages and disadvantages.

Chapter 2, jointly written with Eric Beutner and Antoon Pelsser, contributes to the discussion on LSMC Regress-Now versus LSMC Regress-Later by analyzing the properties of the Regress-Later estimator in terms of its asymptotic convergence rate in single-period problems. This seems to be the first initiative to derive convergence rates for the Regress-Later estimator. Our analysis reveals that Regress-Later is a non-standard regression problem, where the variance of the regression error converges to zero, which is fundamentally different from Regress-Now. As a direct consequence Regress-Later potentially converges faster than at N^{-1} , where N is the sample size. This is in contrast to Regress-Now, which can never converge faster than at N^{-1} . Furthermore, it is shown that nonparametric assumptions often applied in the derivation of convergence rates for Regress-Now estimators (see, for instance Newey, 1997; Stentoft, 2004) can be relaxed when considering Regress-Later estimators. For a basis of non-overlapping one-dimensional piecewise linear functions the explicit convergence rate is derived, which exceeds N^{-1} .

As earlier mentioned static portfolio replication is an alternative concept that insurers employ in approximating the value function of their (life) insurance liabilities. The concept of statically replicating a contingent claim is prominent in the financial literature. The principle of static replication is to construct a portfolio of securities that mirrors the terminal payoff of a target security in every possible state of the world. By the no-arbitrage condition, if the payoff of the target security is perfectly replicated, the replication automatically matches the security's value at all times before maturity, implying that they have

the same market-consistent price. Madan and Milne (1994) propose a general mathematical framework for the static replication of a class of contingent claims, where finite second moment contingent claims are modeled as elements of a separable Hilbert space. In Chapter 3, jointly written with Antoon Pelsser and Eric Beutner, the same idea is exploited to obtain a static replicating portfolio of a general contingent claim. We show that the static replicating portfolio combines least squares regression with Monte Carlo simulation. Recall that a replicating portfolio is understood to mirror the terminal payoff of the target contingent claim. Then, a replicating portfolio is obtained through least squares regression of the time T value against basis functions valued at the same time point. The same principle has already been presented when introducing the LSMC Regress-Later approach of Glasserman and Yu (2004b). In fact the theoretical construct for replicating portfolios exactly corresponds to LSMC Regress-Later. The asymptotic convergence properties for replicating portfolios are thus directly taken from the convergence results of Regress-Later estimators. The time t value function of the replicating portfolio is obtained by valuing its replicating instruments, which is then used as the proxy to the time t conditional expectation function of the target contingent claim. It is shown that convergence of the replicating portfolio to the terminal payoff function implies convergence of the time t value function of the replicating portfolio to the time t value function of the target contingent claim. Multi-dimensional piecewise linear functions on non-overlapping hypercubes are constructed as basis and its explicit convergence rate for a class of multi-dimensional target functions is derived. The findings of Chapter 2 conclude that the replicating portfolio concept in (life) insurance is a theoretically sound construct. Nonetheless practical challenges remain, particularly in view of finding a replicating portfolio for path-dependent insurance policies. A proposal

is made on how to approach such problems for a particular well-known path-dependent life insurance contract.

Chapter 2 gives the underlying theory of LSMC approaches to the approximation of unknown conditional expectation functions. Chapter 3 provides the link to portfolio replication. Chapter 4, which is joint work with Antoon Pelsser, sheds light on the differences between LSMC (Regress-Now) and portfolio replication (LSMC Regress-Later). Advantages and disadvantages of both methods are revealed. It is found that portfolio replication offers multiple advantages stemming from the fact that portfolio replication is a non-standard regression problem. However, the curse of dimensionality problem is for replicating portfolios more striking than for LSMC (Regress-Now) when it comes to path-dependent insurance policies. This renders the construction of a replicating portfolio of a path-dependent insurance policy a more challenging task. We give a proposal of how this issue could be addressed, but stress that further research is required.

Chapter 5 concludes the thesis and provides an overview of its limitations.

Although the notation is applied as uniformly as possible across the chapters, slight deviations occur where customizations to each chapter are sensible. As a consequence, the relevant notation for each chapter is defined within each respective chapter. Whenever applicable, proofs as well as additional information is contained in the appendices at the end of the respective chapter.

Chapter 2

Fast Convergence of Regress-Later Estimates in Least Squares Monte Carlo

Many problems in financial engineering involve the estimation of conditional expectations. Often Least Squares Monte Carlo techniques are used for the estimation, whenever the valuation of the conditional expectation with pure simulation methods is too costly. Unlike the standard implementation where the value function is regressed on a set of basis functions valued at the beginning of the time interval, the Regress-Later method regresses the value function on a set of basis functions valued at the end of the interval. The conditional expectation across the interval is then computed exactly for each basis function.

Whilst the existing literature offers derivations for the convergence rate of conventional methods (see, for example, Stentoft, 2004), the asymptotic properties of the Regress-Later method have not yet been investigated. In this chapter we provide sufficient conditions under which we derive the convergence rate of Regress-Later estimators. We show that the Regress-Later method is capable of converging significantly faster than conventional methods and provide an explicit example, which achieves a convergence rate faster than N^{-1} . Achieving faster convergence speed provides a strong motivation for using Regress-Later methods in Least Squares Monte Carlo algorithms¹.

2.1 Introduction

The Least Squares Monte Carlo (LSMC) technique is widely applied in the area of Finance to estimate conditional expectations across a time interval. Under LSMC the cross-sectional information inherent in the simulated data is exploited to obtain approximating functions to conditional expectations through performing least squares regressions on the simulated data. Examples may be found in Carriere (1996), Longstaff and Schwartz (2001), Tsitsiklis and Van Roy (2001), Clement et al.

¹This chapter is based on Beutner et al. (2013)

(2002), Stentoft (2004), Glasserman and Yu (2004a), Egloff et al. (2007), Belomestny (2011), Gerhold (2011) and Zanger (2013), who discuss approaches to LSMC with application to American/Bermudan option pricing; see also Broadie and Glasserman (1997) who apply simulation based methods and a dynamic programming algorithm to American option pricing. We further point to Belomestny et al. (2010), who discuss the convergence of regression-based Monte Carlo algorithms in the context of general stochastic optimal control problems, and Schoenmakers et al. (2013), who apply regression-based Monte Carlo methods for deriving dual estimates of Bermudan options. The papers on LSMC have in common that the conditional expectation at time t is approximated through least squares regression of the value function at a time point $T > t$ against basis functions at the earlier time point t . This approach to the estimation of conditional expectations has been termed “regression now” by Glasserman and Yu (2004b). Here we will use the expression Regress-Now. In the same paper Glasserman and Yu (2004b) introduce an alternative approach that they called “regression later” (throughout this paper Regress-Later). In Regress-Later the value function at a time point T is approximated through LSMC techniques by basis functions that are measurable with respect to the information available at time T . Moreover, the basis functions in Regress-Later are selected such that the conditional expectation can be computed exactly. The conditional expectation of the time T value function is then derived by evaluating the basis functions contained in the approximation function. In this paper, we shall show that the Regress-Later method is fundamentally different from the Regress-Now technique. But before we briefly review recent contributions to the literature.

Glasserman and Yu (2004b) show that the Regress-Later approach offers advantages compared to Regress-Now. They compare the prop-

erties of the coefficient estimates given that both approximations yield a linear combination of the same basis functions. Their results suggest that in a single-period problem the Regress-Later algorithm yields a higher coefficient of determination and a lower covariance matrix for the estimated coefficients; see also Broadie and Cao (2008) who report similar observations. This implies that with Regress-Later potentially a better fit is attained with more accurate coefficient estimates. The results depend on more restrictive conditions on the basis functions as these are required to fulfill the martingale property. However, for many financial applications it is reasonable to expect that such a basis exists. Bender and Steiner (2012) use LSMC to numerically approximate the conditional expectations involved in estimating backward stochastic differential equations. They consider the Regress-Later algorithm and combine it with martingale basis functions as suggested in Glasserman and Yu (2004b). Their empirical case studies suggest that Regress-Later with martingale basis functions achieves a better numerical approximation at lower computational costs compared to traditional LSMC. The empirical results all show faster convergence rates for the Regress-Later algorithm combined with martingale basis functions compared to traditional LSMC. Gobet and Zineb (2013) combine the Regress-Now and Regress-Later technique by essentially constructing control variates with Regress-Later for the estimation of conditional expectations with Regress-Now type of LSMC. In the context of Monte Carlo integration Gobet and Surana (2014) introduce an adaptive algorithm to estimate the coefficients of the basis representation of the value function. It should be remarked that the authors do not apply regression methods to estimate the coefficients.

In this paper we shall shed more light on the advantages offered by Regress-Later as observed in Glasserman and Yu (2004b), Broadie and Cao (2008) and Bender and Steiner (2012) by analyzing the properties

of Regress-Later in terms of its convergence rate. As it seems to be the first attempt to derive convergence rates for Regress-Later estimators we restrict ourselves to single-period problems. It is worth stressing that the estimation of conditional expectations in single-period problems is of interest in its own. One example is the derivation of risk measures based on the distribution of conditional expectations of random variables at a single time point, where closed form solutions for the conditional expectation are not available and its estimation with pure simulation methods is too costly. Our analysis will reveal that, as mentioned above, Regress-Later is fundamentally different from Regress-Now. Firstly, because Regress-Later can and does achieve a convergence rate for the mean-square error that is faster than N^{-1} ; cf. Section 2.3. Here and throughout N is the sample size. We shall present an example where the convergence is indeed much faster than N^{-1} ; cf. Section 2.4. This is in sharp contrast to Regress-Now that can never converge faster than N^{-1} . We provide explanations for both facts, i.e. the bound N^{-1} for Regress-Now and the faster convergence rate for Regress-Later. It will turn out that the latter is a consequence of the fact that Regress-Later is a non-standard regression problem, because the variance of the noise term converges to zero. Secondly, we shall explain that the conditions needed to derive convergence rates for Regress-Later estimators are much weaker than the typical assumptions used in the literature for Regress-Now estimators; an exception is the recent work by Zanger (2013). This has to do with the fact that for Regress-Now estimators reasonable conditions stemming from nonparametric statistics were employed in the literature whereas for Regress-Later estimators we should definitely use parametric assumptions. Thereby, we will easily obtain approximations of the value function on non-compact intervals; see the discussion in Section 2.3. Apart from these fundamental differences we will also present several

examples which show that the functions to be approximated in Regress-Now may differ in nature compared to Regress-Later. Furthermore, we explain that the nonparametric assumptions that were applied in deriving convergence rates for Regress-Now estimators (see, for instance, Newey (1997)) can be weakened. These relaxed assumptions allow us to approximate the value function on the entire real line by a Regress-Now estimator and not only on a compact domain.

The structure of this paper is as follows. Section 2.2 introduces the general LSMC estimator with sieve and distinguishes between its Regress-Now and Regress-Later applications. In Section 2.3 the asymptotic convergence rate for Regress-Later estimators is derived under conditions that allow to approximate the value function on non-compact intervals. Moreover, similar conditions are applied to Regress-Now estimators while a motivation is given for when these conditions may be applicable for the Regress-Now technique. We conclude this section by providing explanations for the different convergence rates of Regress-Now and Regress-Later estimators. Section 2.4 introduces an orthonormal basis based on piecewise linear functions and derives the explicit convergence rate for Regress-Later with that basis. Section 2.5 concludes. The proofs of all auxiliary results are presented in the appendix.

2.2 Mathematical Model for Regress-Now and Regress-Later

As described in the introduction Regress-Now and Regress-Later are simulation based techniques to estimate conditional expectations. Often they are combined with series or sieve estimation, where the number of regressors in the least squares estimation is not fixed and finite;

for an overview on series and sieve estimation one may refer to Chen (2007). In this section, we describe the mathematical model that is used throughout and explain the Regress-Now and Regress-Later approaches within this model.

We start with our mathematical model. Let $Z = \{Z(t), 0 \leq t \leq T\}$ be a d -dimensional stochastic process with $d \in \mathbb{N}$ defined on some filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0 \leq t \leq T}, \tilde{\mathbb{P}})$. We denote the filtration generated by Z by $\{\mathcal{F}_t\}_{0 \leq t \leq T}$. The measure $\tilde{\mathbb{P}}$ denotes some probability measure equivalent to the true probability measure \mathbb{P} . We leave $\tilde{\mathbb{P}}$ generally unspecified when developing our model, but provide the reader with an interpretation of the mathematical model for selecting $\tilde{\mathbb{P}}$ just before Subsection 2.2.1. The paths $Z(\cdot, \omega)$ of Z given by $t \rightarrow Z(t, \omega)$, $t \in [0, T]$, are assumed to lie in some function space $\mathbb{D}_d[0, T]$ consisting of functions mapping from $[0, T]$ to \mathbb{R}^d , and we consider Z as a random function. If $d = 1$ we just write $\mathbb{D}[0, T]$ and \mathbb{R} . We assume that the payoff X is \mathcal{F}_T -measurable and that for every ω in the sample space Ω the payoff $X(\omega)$ of the contingent claim X can be written as $g_T(A_T(Z(\cdot, \omega)))$, where A_T is a known (measurable) functional mapping from $\mathbb{D}_d[0, T]$ to \mathbb{R}^ℓ and g_T is a known Borel-measurable function that maps from \mathbb{R}^ℓ to \mathbb{R} . This basically means that the payoff function X depends only on finitely many characteristics of the stochastic paths of the underlying process. These characteristics are comprised in the functional mapping A_T . The notation is very powerful for our purposes later on, and we illustrate it here with an example.

Example 1. (*Asian option*) Let Z be one-dimensional and $X = (\int_0^T Z_1(u) du - K)^+$, where K is the strike price. Then X does only depend on $\int_0^T Z_1(u) du$. Thus, $A_T(f) = \int_0^T f(u) du$ for every function $f \in \mathbb{D}[0, T]$ and therefore $\ell = 1$.

Note that at time T it suffices to observe the time average of the stochastic process rather than the whole path. Further examples that

highlight the idea behind the notation are given in Sections 2.2.1 and 2.2.2.

In the relevant literature, it has become standard to restrict attention to square-integrable random variables; (see e.g. Stentoft, 2004; Bergstrom, 1985; Madan and Milne, 1994; Longstaff and Schwartz, 2001). We do the same here, that is we assume $g_T \in L_2(\mathbb{R}^\ell, \mathcal{B}(\mathbb{R}^\ell), \tilde{\mathbb{P}}^{A_T(Z)})$ (implying that X is square-integrable, because $X(\omega) = g_T(A_T(Z(\cdot, \omega)))$) where $\mathcal{B}(\mathbb{R}^\ell)$ denotes the Borel σ -algebra on \mathbb{R}^ℓ , and $\tilde{\mathbb{P}}^{A_T(Z)}$ denotes the probability measure on \mathbb{R}^ℓ induced by the mapping $A_T(Z)$. Recall that $L_2(\mathbb{R}^\ell, \mathcal{B}(\mathbb{R}^\ell), \tilde{\mathbb{P}}^{A_T(Z)})$ is a Hilbert space with inner product

$$\int_{\mathbb{R}^\ell} h_1(u)h_2(u) d\tilde{\mathbb{P}}^{A_T(Z)}(u) = \mathbb{E}_{\tilde{\mathbb{P}}} [h_1(A_T(Z))h_2(A_T(Z))]$$

and norm

$$\sqrt{\int_{\mathbb{R}^\ell} h_1(u)h_1(u) d\tilde{\mathbb{P}}^{A_T(Z)}(u)} = \sqrt{\mathbb{E}_{\tilde{\mathbb{P}}} [h_1^2(A_T(Z))]}.$$

As already mentioned the quantity of interest is $\mathbb{E}_{\tilde{\mathbb{P}}} [X|\mathcal{F}_t]$ where $\tilde{\mathbb{P}}$ denotes a probability measure. If we take $\tilde{\mathbb{P}} = \mathbb{Q}$ where \mathbb{Q} is the equivalent risk-neutral probability measure, then $\mathbb{E}_{\mathbb{Q}} [D(t, T)X|\mathcal{F}_t]$, where $D(t, T)$ is the discount factor for the period t to T , corresponds to the time t price of X . As a further example for the importance of $\mathbb{E}_{\tilde{\mathbb{P}}} [X|\mathcal{F}_t]$ take $\tilde{\mathbb{P}} = \mathbb{P}$, where \mathbb{P} is the true probability measure. Then, $\mathbb{E}_{\tilde{\mathbb{P}}} [X|\mathcal{F}_t]$ is the best L_2 -approximation to X that is measurable w.r.t. the σ -field \mathcal{F}_t . In Sections 2.3 and 2.4 we will use $\tilde{\mathbb{P}}$ and leave it unspecified. Regress-Now with sieves and Regress-Later with sieves are two different simulation-based approaches to obtain an approximation to the conditional expectation of X at time t . We outline both approaches in the following subsections.

2.2.1 Regress-Now

We first describe the Regress-Now approach which is currently more popular. To describe the Regress-Now approach, we assume that the quantity of interest, $\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t]$, can be written as

$$g_{0,t}(A_t(Z)) = \mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t], \quad 0 \leq t < T,$$

where A_t is a known (measurable) functional mapping from $\mathbb{D}_d[0, t]$ to \mathbb{R}^s and $g_{0,t}$ is an unknown Borel-measurable function that maps from \mathbb{R}^s to \mathbb{R} . Here, $\mathbb{D}_d[0, t]$ is the restriction of $\mathbb{D}_d[0, T]$ to the interval $[0, t]$.

Remark 2.1. The notation $g_{0,t}(A_t(Z))$ is used to emphasize that the function $g_{0,t}$ is generally unknown. Thus, we use the convention that a subscript ‘0’ indicates the true but unknown parameter. In contrast, note that $g_T(A_T(Z))$ refers to the payoff function, which is known in a simulation-based model as the simulation is controlled by the modeler.

We give a few examples below for $g_{0,t}$ and A_t that serve to illustrate the notation and concept. In these examples we take $\tilde{\mathbb{P}} = \mathbb{Q}$ to emphasise the pricing aspect of conditional expectations and for convenience we assume that the discount factor is equal to 1.

Example 2. (*European call with Regress-Now*) Let Z be one-dimensional and consider an European call. Then $X = (Z_1(T) - K)^+$, where K is the strike price. Moreover, $\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t]$ does only depend on $Z_1(t)$. Hence, we can take $A_t(f) = f(t)$ for every function $f \in \mathbb{D}[0, t]$ and therefore $s = 1$.

Example 3. (*European basket option with Regress-Now*) Consider a d -dimensional European basket option of the type $X = (\sum_{i=1}^d Z_i(T) - K)^+$, where K is the strike price. In general $\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t]$ depends on $\mathbf{Z}(t) = (Z_1(t), \dots, Z_d(t))$ and not only on $\sum_{i=1}^d Z_i(t)$. Then $A_t(f) =$

$f(t)$ for every function $f \in \mathbb{D}_d[0, t]$ and therefore $s = d$. We give an example that shows our claim. Consider two assets $Z_1(t)$ and $Z_2(t)$, $t = 0, 1, 2$, that are independent under \mathbb{Q} with

$$\begin{aligned} \mathbb{Q}(Z_1(0) = 10) &= 1; & \mathbb{Q}(Z_1(1) = 12) &= \mathbb{Q}(Z_1(1) = 6) = 0.5; \\ \mathbb{Q}(Z_1(2) = 14|Z_1(1) = 12) &= \mathbb{Q}(Z_1(2) = 8|Z_1(1) = 12) = 0.5, \\ \mathbb{Q}(Z_1(2) = 6|Z_1(1) = 6) &= 1, \end{aligned}$$

and

$$\begin{aligned} \mathbb{Q}(Z_2(0) = 10) &= 1; & \mathbb{Q}(Z_2(1) = 12) &= \mathbb{Q}(Z_2(1) = 6) = 0.5; \\ \mathbb{Q}(Z_2(2) = 14|Z_2(1) = 12) &= \mathbb{Q}(Z_2(2) = 8|Z_2(1) = 12) = 0.5; \\ \mathbb{Q}(Z_2(2) = 9|Z_2(1) = 6) &= \mathbb{Q}(Z_2(2) = 1|Z_2(1) = 6) = 0.5. \end{aligned}$$

Take $X = (Z_1(2) + Z_2(2) - K)^+$ with $K = 10$. We are interested in the conditional expectation at time $t = 1$, i.e. $\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_1]$. For the case where $Z_1(1) + Z_2(1) = 18$ we obtain the following results

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}}[X|Z_1(1) = 12, Z_2(1) = 6] &= 6.25 \text{ and} \\ \mathbb{E}_{\mathbb{Q}}[X|Z_1(1) = 6, Z_2(1) = 12] &= 7. \end{aligned}$$

We immediately see that knowing the sum $Z_1(1) + Z_2(1)$ does not suffice to determine the conditional expectation at time $t = 1$ as for $Z_1(1) + Z_2(1) = 18$ the conditional expectation can either be 6.25 or 7.

Example 4. (*Asian option with Regress-Now*) Let Z be one-dimensional and $X = \left(\int_0^T Z_1(u) du - K\right)^+$, where K is again the strike price. Then $\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t]$ does only depend on $\int_0^t Z_1(u) du$ and $Z_1(t)$. Thus, $A_t(f) = \left(\int_0^t f(u) du, f(t)\right)$ for every function $f \in \mathbb{D}[0, t]$ and therefore $s = 2$.

Example 5. (*Mildly path-dependent option with Regress-Now*) Let Z be one-dimensional, let X be a function of $Z(u)$, $u < T$, and $Z(T)$, i.e. $X = g_T(Z(u), Z(T))$ for some function g_T and suppose that the expectation $\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t]$ depends only on $Z(t)$ for $t < u$. Then $A_t(f) = f(t)$ for every function $f \in \mathbb{D}[0, t]$ and therefore $s = 1$.

The above examples illustrate the notation used for the Regress-Now model. We contrast the Regress-Now examples to their Regress-Later counterparts in Subsection 2.2.2.

In the following we describe how the Regress-Now *with sieves* approach estimates $g_{0,t}$. The description is rather detailed, because we will use it in Section 2.3 to explain the different convergence rates for Regress-Now and Regress-Later.

Recall that the square-integrability of X implies that $\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t]$ is square-integrable as well. Hence, we also have that

$g_{0,t} \in L_2(\mathbb{R}^s, \mathcal{B}(\mathbb{R}^s), \tilde{\mathbb{P}}^{A_t(Z)})$. Since the space $L_2(\mathbb{R}^s, \mathcal{B}(\mathbb{R}^s), \tilde{\mathbb{P}}^{A_t(Z)})$ is separable, $g_{0,t}$ is expressible in terms of a countable orthonormal basis $\{e_k^{now}\}_{k=1}^{\infty}$

$$g_{0,t} = \sum_{k=1}^{\infty} \alpha_k^{now} e_k^{now};$$

see, for instance, Bogachev (2007, Corollary 4.2.2 and Corollary 4.3.4). Because $g_{0,t}(A_t(Z))$ is the projection of X , the coefficients are given as

$$\alpha_k^{now} = \mathbb{E}_{\tilde{\mathbb{P}}}[X e_k^{now}(A_t(Z))].$$

Thus, in particular, we have

$$g_{0,t}(A_t(Z)) = \sum_{k=1}^{\infty} \alpha_k^{now} e_k^{now}(A_t(Z)); \tag{2.1}$$

and, as usual, we define the projection error $p_{0,t}(A_T(Z))$ by

$$p_{0,t}(A_T(Z)) := X - g_{0,t}(A_t(Z))$$

which implies the well-known representation

$$X = g_{0,t}(A_t(Z)) + p_{0,t}(A_T(Z)).$$

Notice also that by construction $g_{0,t}(A_t(Z))$ and $p_{0,t}(A_T(Z))$ are orthogonal, i.e.

$$\mathbb{E}_{\tilde{\mathbb{P}}} [g_{0,t}(A_t(Z))p_{0,t}(A_T(Z))] = 0.$$

Under the Regress-Now approach one estimates the unknown function $g_{0,t}$ through its representation in Equation (2.1) by generating data under $\tilde{\mathbb{P}}$. Equation (2.1) involves infinitely many parameters, which leaves a direct estimation infeasible. A standard solution to solve such nonparametric estimation problems is estimation with sieves. With sieve the model is estimated through finite-dimensional representations, which grow in complexity as the sample size increases and thereby yield the true outcome in the limit. For Equation (2.1) this implies that with sieves we approximate $g_{0,t}$ by

$$g_{0,t}^K := \sum_{k=1}^K \alpha_k^{now} e_k^{now} = (\boldsymbol{\alpha}_K^{now})^T \mathbf{e}_K^{now},$$

where $\boldsymbol{\alpha}_K^{now} = (\alpha_1, \dots, \alpha_K)^T$, $\mathbf{e}_K^{now} = (e_1^{now}, \dots, e_K^{now})^T$, and T denotes transpose. Thus, a superscript T means transpose and it should be easy to distinguish it from the terminal time T . This results in an approximation error $a_{0,t}^K$ for $g_{0,t}$ given by

$$a_{0,t}^K := g_{0,t} - g_{0,t}^K.$$

Notice that we have $\mathbb{E}_{\mathbb{P}} [g_{0,t}^K(A_t(Z))a_{0,t}^K(A_t(Z))] = 0$ by construction. By definition the approximation error $a_{0,t}^K$ converges to zero as $K \rightarrow \infty$ in L_2 . Moreover, we can now write X as

$$X = g_{0,t}^K(A_t(Z)) + a_{0,t}^K(A_t(Z)) + p_{0,t}(A_T(Z)).$$

From the last equation we can clearly see that the difference between X and $g_{0,t}^K(A_t(Z))$ results from two sources: an approximation error and a projection error.

Now, given a (simulated) sample of size N denoted by $((x_1, A_t(z_1)), \dots, (x_N, A_t(z_N)))$ it is natural to estimate $g_{0,t}^K$ by the ‘sample projection’

$$\hat{g}_{0,t}^K = \arg \min_{g \in \mathcal{H}_K} \frac{1}{N} \sum_{n=1}^N (x_n - g(A_t(z_n)))^2,$$

where $\mathcal{H}_K := \left\{ g : \mathbb{R}^s \rightarrow \mathbb{R} \mid g = \sum_{k=1}^K \alpha_k e_k^{now}, \alpha_k \in \mathbb{R} \right\}$. Thus, we have

$$\hat{g}_{0,t}^K = (\hat{\boldsymbol{\alpha}}_K^{now})^T \mathbf{e}_K^{now},$$

with

$$\hat{\boldsymbol{\alpha}}_K^{now} = \left((\mathbf{E}_K^{now})^T \mathbf{E}_K^{now} \right)^{-1} (\mathbf{E}_K^{now})^T \mathbf{X},$$

where $\mathbf{X} = (x_1, \dots, x_N)^T$ and \mathbf{E}_K^{now} is an $N \times K$ matrix with the n^{th} row equal to $\mathbf{e}_K^{now}(A_t(z_n))$, $n = 1, \dots, N$. Notice that $\hat{\boldsymbol{\alpha}}_K^{now}$ corresponds to the usual least squares estimator from a regression of X against K basis functions valued at time t .

Remark 2.2. Note that the same structure for Regress-Now is used as suggested in Glasserman and Yu (2004b), i.e. the payoff function valued at time T is regressed against basis functions valued at an earlier time point, which is $t < T$ here. A different approach could be to first determine an estimate of the conditional expectation function, e.g. by taking the discounted average of m simulations of the payoff function, which we write here as $1/m \sum_{i=1}^m \tilde{x}_i$ with \tilde{x}_i denoting the discounted payoff in scenario i . Then, this estimate is used in regression against basis functions valued at time t . The challenge is here to find an appropriate m such that the total simulation effort is still defensible. In this research we stick to the structure of Glasserman and Yu (2004b), where independent replications of the payoffs are directly regressed against values of the basis functions at time t .

2.2.2 Regress-Later

In the previous section we discussed the Regress-Now approach. Regress-Later proceeds as follows to approximate the quantity of interest, i.e. $\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t], 0 \leq t < T$. Approximate first the payoff X by basis functions, mathematically speaking random variables, for which the calculation of the conditional expectation at time t is exact. Then, given the linear representation of X through basis functions, apply the operator $\mathbb{E}_{\tilde{\mathbb{P}}}[\cdot|\mathcal{F}_t]$ to these basis functions. The approach takes advantage of the linearity of the expectation operator. Note that the two-step approach is advantageous if for the payoff function X basis functions exist that can easily be evaluated under the conditional expectation. For the case where $\tilde{\mathbb{P}} = \mathbb{Q}$ this implies that closed-form solutions for the price of the basis function must be readily available. We introduce a very simple but effective basis function in Section 2.4.

We now describe the Regress-Later approach and address differences to the Regress-Now approach.

Recall that we assume $X = g_T(A_T(Z))$ with A_T a known (measurable) functional mapping from $\mathbb{D}_d[0, T]$ to \mathbb{R}^ℓ and g_T a known Borel-measurable function that maps from \mathbb{R}^ℓ to \mathbb{R} . The examples below shall illustrate the meaning of g_T and A_T . They may also be compared to their counterparts in Section 2.2.1.

Example 6. (*European call with Regress-Later*) Let Z be one-dimensional and consider an European call. Then $X = (Z_1(T) - K)^+$, where K is the strike price. Then, X does only depend on $Z_1(T)$. Therefore, we can take $A_T(f) = f(T)$ for every function $f \in \mathbb{D}[0, T]$ and hence $\ell = 1$. Moreover, g_T is given by $g_T(x) = (x - K)^+$.

Example 7. (*European basket option with Regress-Later*) Consider the d -dimensional European basket option of Example 3. Then we can take $A_T(f) = \sum_{i=1}^d f_i(T)$ for every function $f \in \mathbb{D}[0, T]$ and therefore $\ell = 1$. Compare with Example 3 where we had $s = d$. Moreover, g_T is again given by $g_T(x) = (x - K)^+$ with K the strike price.

Example 8. (*Asian option with Regress-Later*) This corresponds to Example 1. For the readers' convenience and to contrast it with Example 4 we repeat it here. Let Z be one-dimensional and $X = (\int_0^T Z_1(u) du - K)^+$, where K is the strike price. Then X does only depend on $\int_0^T Z_1(u) du$. Thus, $A_T(f) = \int_0^T f(u) du$ for every function $f \in \mathbb{D}[0, T]$ and therefore $\ell = 1$. Compare with Example 4 where we had $s = 2$. Again, we have $g_T(x) = (x - K)^+$.

Example 9. (*Mildly path-dependent option with Regress-Later*) Let X be as in Example 5. Then $A_T(f) = (f(u), f(T))$ for every function $f \in \mathbb{D}[0, T]$ and therefore $\ell = 2$. Recall that we had $s = 1$ in Example 5.

The above examples illustrate the notation that is applied to the Regress-Later approach. They also show that there may be fundamental differences between Regress-Now and Regress-Later. As already mentioned in the introduction the functions to be approximated in Regress-Now may differ in nature compared to Regress-Later. Recall that in Regress-Now the unknown function $g_{0,t}(A_t(Z))$ is approximated while in Regress-Later the known function $g_T(A_T(Z))$ is initially of interest. Although the ultimate goal in both Regress-Now and Regress-Later is to approximate the conditional expectation function, the approximation approaches are very different. As Examples 7, 8, 9 and their Regress-Now counterparts show the dimensionality of the function to be approximated under the same problem set-up may differ between Regress-Now and Regress-Later. The dimensionality of the function to be approximated may be one decision criterion in choosing between Regress-Now and Regress-Later. In Section 2.3 we investigate the speed of convergence of both estimators and provide a strong argument for using Regress-Later estimators.

In the following we describe the Regress-Later estimation *with sieves*. Remember that we assume square-integrability of the payoff function, meaning that $g_T \in L_2(\mathbb{R}^\ell, \mathcal{B}(\mathbb{R}^\ell), \tilde{\mathbb{P}}^{A_T(Z)})$. Hence, by the same argument as in Section 2.2.1,

$$X = g_T(A_T(Z)) = \sum_{k=1}^{\infty} \alpha_k^{lat} e_k^{lat}(A_T(Z)),$$

where $\{e_k^{lat}\}_{k=1}^{\infty}$ is a countable orthonormal basis of $L_2(\mathbb{R}^\ell, \mathcal{B}(\mathbb{R}^\ell), \tilde{\mathbb{P}}^{A_T(Z)})$. Notice that by construction the projection error is zero in contrast to the Regress-Now approach. The coefficients α_k^{lat} are given by

$$\alpha_k^{lat} = \mathbb{E}_{\tilde{\mathbb{P}}} [X e_k^{lat}(A_T(Z))].$$

Using the same sieve methodology as for Regress-Now the value function

$$g_T = \sum_{k=1}^{\infty} \alpha_k^{lat} e_k^{lat}$$

is approximated by a finite number of regressors, i.e.

$$g_T^K = \sum_{k=1}^K \alpha_k^{lat} e_k^{lat} = (\boldsymbol{\alpha}_K^{lat})^T \mathbf{e}_K^{lat},$$

where $\boldsymbol{\alpha}_K^{lat} = (\alpha_1^{lat}, \dots, \alpha_K^{lat})^T$ and $\mathbf{e}_K^{lat} = (e_1^{lat}, \dots, e_K^{lat})^T$. This finite-dimensional representation can then be estimated by generating data under $\tilde{\mathbb{P}}$. Defining the approximation error a_T^K as usual by $a_T^K := g_T - g_T^K$ we obtain the representation

$$X = g_T^K(A_T(Z)) + a_T^K(A_T(Z)), \quad (2.2)$$

which, as already mentioned, does not contain a projection error. Notice also that $\mathbb{E}_{\tilde{\mathbb{P}}} [g_T^K(A_T(Z))a_T^K(A_T(Z))] = 0$. It should be emphasized again that the approximation error converges to zero as $K \rightarrow \infty$ in L_2 . As for Regress-Now with sieves given a (simulated) sample of size N denoted by $(x_1, A_T(z_1)), \dots, (x_N, A_T(z_N))$ it is natural to estimate g_T^K by the ‘sample projection’ leading to

$$\hat{g}_T^K = (\hat{\boldsymbol{\alpha}}_K^{lat})^T \mathbf{e}_K^{lat},$$

with

$$\hat{\boldsymbol{\alpha}}_K^{lat} = \left((\mathbf{E}_K^{lat})^T \mathbf{E}_K^{lat} \right)^{-1} (\mathbf{E}_K^{lat})^T \mathbf{X},$$

where $\mathbf{X} = (x_1, \dots, x_N)^T$ and \mathbf{E}_K^{lat} is an $N \times K$ matrix with the n^{th} row equal to $\mathbf{e}_K^{lat}(A_T(z_n))$, $n = 1, \dots, N$. Notice that $\hat{\boldsymbol{\alpha}}_K^{lat}$ corresponds

to the usual least squares estimator from a regression of X against K basis functions valued at time T .

2.3 Convergence Rates for Regress-Later with sieves

In this section we derive convergence rates for Regress-Later with sieves and comment on convergence rates for Regress-Now with sieves. We start with the analysis of Regress-Later estimators. Our method of proof follows Newey (1997). Its presentation follows in part Hansen (2014). However, the conditions we impose are different from the conditions in Newey (1997) which have, for instance, also been applied by Stentoft (2004). To understand this note that Newey (1997) takes a nonparametric approach to estimating a conditional expectation that is unknown. We exemplify this with Assumption 3 in Newey (1997). This assumption with $d = 0$ (not to be confused with the d we use here for the dimension of Z) would read as follows for Regress-Later

There are $\gamma > 0$, α_K^{lat} s.t.

$$\sup_{x \in D} |g_T(x) - g_T^K(x)| = \sup_{x \in D} \left| g_T(x) - (\alpha_K^{lat})^T \mathbf{e}_K^{lat}(x) \right| = O(K^{-\gamma_{lat}}) \quad (2.3)$$

as $K \rightarrow \infty$, where D is the domain of g_T .

Note that Condition (2.3) is independent of the probability measure $\tilde{\mathbb{P}}$. From a nonparametric point of view this makes perfectly sense, because, if it is fulfilled, the convergence rate is the same whatever the true probability measure. However, in the context of LSMC *we do know* $\tilde{\mathbb{P}}$, because it is the measure used in the simulation and it is controlled by the user. Thus, it is legitimate to relax Assumption

3 in Newey (1997). Additionally, Condition (2.3) implicitly requires that g_T is bounded or that D is compact, unless g_T is, for instance, in the span of the \mathbf{e}_K^{lat} . In the context of American option pricing Stentoft (2004) circumvents the problem by explicitly neglecting far in-the-money and far out-of-the-money tails of the distribution. Although this is a reasonable assumption in the context of American options obtaining results on the whole domain is surely welcomed in other areas of application. As we know $\tilde{\mathbb{P}}$ in a simulation-based framework, we will replace Assumption 3 in Newey (1997) by the following condition

Assumption 2.1. There are $\gamma_{lat} > 0$, $\boldsymbol{\alpha}_K^{lat}$ s.t.

$$\begin{aligned} & \sqrt{\mathbb{E}_{\tilde{\mathbb{P}}} \left[(g_T(A_T(Z)) - (\boldsymbol{\alpha}_K^{lat})^T \mathbf{e}_K^{lat}(A_T(Z)))^4 \right]} \\ &= \sqrt{\int_{\mathbb{R}^\ell} (g_T(u) - (\boldsymbol{\alpha}_K^{lat})^T \mathbf{e}_K^{lat}(u))^4 d\tilde{\mathbb{P}}^{A_T(Z)}(u)} \\ &= \sqrt{\int_{\mathbb{R}^\ell} a_T^K(u)^4 d\tilde{\mathbb{P}}^{A_T(Z)}(u)} = O(K^{-\gamma_{lat}}). \end{aligned}$$

Notice that Assumption 2.1 does not require that g_T is bounded or that its domain is compact. From a nonparametric point of view Assumption 2.1 is unsatisfactory, because the $O(K^{-\gamma_{lat}})$ term on the right-hand side is not independent of $\tilde{\mathbb{P}}$. However, with regard to LSMC it is worth stressing once again that we know the data generating process, i.e. $\tilde{\mathbb{P}}$, A_T and g_T , and that therefore Assumption 2.1 can, for instance, be checked by considering the behavior of g_T in the tails. Moreover, it is also worth pointing out that

$$\sqrt{\int_{\mathbb{R}^\ell} (g_T(u) - g_T^K(u))^4 d\tilde{\mathbb{P}}^{A_T(Z)}(u)} = O(K^{-\gamma_{lat}})$$

implies

$$\sqrt{\int_{\mathbb{R}^\ell} (g_T(u) - g_T^K(u))^4 d\widehat{\mathbb{P}}^{A_T(Z)}(u)} = O(K^{-\gamma_{lat}})$$

whenever $\widehat{\mathbb{P}}^{A_T(Z)}$ has a bounded density w.r.t. $\tilde{\mathbb{P}}^{A_T(Z)}$.

We now state our second assumption to derive the convergence rate for Regress-Later with sieves.

Assumption 2.2. $((X_1, A_T(Z_1)), \dots, (X_N, A_T(Z_N)))$ are i.i.d.

To formulate our theorem on the convergence rate of Regress-Later with sieves we define the net $\tilde{h}^{lat} : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{R}$ by

$$\tilde{h}^{lat}(N, K) := \frac{1}{N} \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(\left(\mathbf{e}_K^{lat}(A_T(Z)) \right)^T \mathbf{e}_K^{lat}(A_T(Z)) \right)^2 \right].$$

We can now state our main theorem for the convergence of Regress-Later estimators.

Theorem 2.1. *Let Assumptions 2.1 and 2.2 be satisfied. Additionally, assume that there is a sequence $\mathcal{K} : \mathbb{N} \rightarrow \mathbb{N}$ such that*

$$\tilde{h}^{lat}(N, \mathcal{K}(N)) \rightarrow 0 \text{ as } N \rightarrow \infty. \quad (2.4)$$

Then

$$\mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) \right)^2 \right] = O_{\tilde{\mathbb{P}}}(\mathcal{K}(N)^{-\gamma_{lat}}).$$

Please notice that the convergence rate in Theorem 2.1 is completely determined by the speed of the approximation and Condition (2.4), i.e. the growth rate of $\mathcal{K}(N)$. Below, we will see that this is in sharp contrast to Regress-Now. Moreover, this fact makes it possible that we may obtain a convergence rate that is faster than N^{-1} . We will

comment on this at the end of this section.

Before giving the proof of Theorem 2.1 two further remarks are in order. First, notice that Condition (2.4) restricts the growth rate of \mathcal{K} when compared to N , because for fixed N the net $\tilde{h}^{lat}(N, K)$ is increasing in K . Second, if Condition (2.4) holds, then we also have

$$\begin{aligned} \widehat{h}^{lat}(N, \mathcal{K}(N)) &:= \frac{1}{N} \mathbb{E}_{\widehat{\mathbb{P}}} \left[\left((\mathbf{e}_{\mathcal{K}(N)}^{lat}(A_T(Z)))^T \mathbf{e}_{\mathcal{K}(N)}^{lat}(A_T(Z)) \right)^2 \right] \\ &\rightarrow 0 \text{ as } N \rightarrow \infty \end{aligned}$$

whenever $\widehat{\mathbb{P}}^{A_T(Z)}$ has a bounded density w.r.t. $\tilde{\mathbb{P}}^{A_T(Z)}$.

The proof of Theorem 2.1 is based on the following two lemmas whose proofs are given in the Appendix.

Lemma 2.1. *If Condition (2.4) and Assumption 2.2 hold, we have*

$$\left\| \frac{1}{N} (\mathbf{E}_{\mathcal{K}(N)}^{lat})^T \mathbf{E}_{\mathcal{K}(N)}^{lat} - I_{\mathcal{K}(N)} \right\|_F = o_{\widehat{\mathbb{P}}}(1), \quad (2.5)$$

where $\|\cdot\|_F$ is the Frobenius norm and $I_{\mathcal{K}(N)}$ denotes the $\mathcal{K}(N) \times \mathcal{K}(N)$ identity matrix. Moreover,

$$\lambda_{\min} \left(\frac{1}{N} (\mathbf{E}_{\mathcal{K}(N)}^{lat})^T \mathbf{E}_{\mathcal{K}(N)}^{lat} \right) \xrightarrow{\widehat{\mathbb{P}}} 1, \quad (2.6)$$

where $\lambda_{\min}(\mathbf{A})$ denotes the smallest eigenvalue of a matrix \mathbf{A} .

Lemma 2.2. *If Condition (2.4) and Assumptions 2.1 and 2.2 hold, we have*

$$(\widehat{\boldsymbol{\alpha}}_{\mathcal{K}(N)}^{lat} - \boldsymbol{\alpha}_{\mathcal{K}(N)}^{lat})^T (\widehat{\boldsymbol{\alpha}}_{\mathcal{K}(N)}^{lat} - \boldsymbol{\alpha}_{\mathcal{K}(N)}^{lat}) = o_{\widehat{\mathbb{P}}}(\mathcal{K}(N)^{-\gamma_{lat}}).$$

Lemma 2.1 shows that the sample second moment matrix of the basis functions converges to the identity matrix in the Frobenius norm. Lemma 2.2 considers the convergence of the estimation error of the estimated coefficients. We now give the proof of the above theorem.

Proof of Theorem 2.1. Observe first that Assumption 2.1 implies

$$\mathbb{E}_{\tilde{\mathbb{P}}}\left[\left(g_T(A_T(Z)) - (\boldsymbol{\alpha}_K^{lat})^T \mathbf{e}_K^{lat}(A_T(Z))\right)^2\right] \leq O(K^{-\gamma_{lat}}) \quad (2.7)$$

by Cauchy-Schwarz and that

$$\begin{aligned} \mathbb{E}_{\tilde{\mathbb{P}}}\left[e_k^{lat}(A_T(Z)) \left(g_T(A_T(Z)) - (\boldsymbol{\alpha}_K^{lat})^T \mathbf{e}_K^{lat}(A_T(Z))\right)\right] &= 0, \\ k &= 1, \dots, K. \end{aligned}$$

Hence,

$$\begin{aligned} &\mathbb{E}_{\tilde{\mathbb{P}}}\left[\left(g_T(A_T(Z)) - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z))\right)^2\right] \\ &= \mathbb{E}_{\tilde{\mathbb{P}}}\left[\left(g_T(A_T(Z)) - (\boldsymbol{\alpha}_{\mathcal{K}(N)}^{lat})^T \mathbf{e}_{\mathcal{K}(N)}^{lat}(A_T(Z))\right)^2\right] \\ &\quad + \mathbb{E}_{\tilde{\mathbb{P}}}\left[\left((\boldsymbol{\alpha}_{\mathcal{K}(N)}^{lat})^T \mathbf{e}_{\mathcal{K}(N)}^{lat}(A_T(Z)) - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z))\right)^2\right] \\ &\leq O(\mathcal{K}(N)^{-\gamma_{lat}}) + (\hat{\boldsymbol{\alpha}}_{\mathcal{K}(N)}^{lat} - \boldsymbol{\alpha}_{\mathcal{K}(N)}^{lat})^T (\hat{\boldsymbol{\alpha}}_{\mathcal{K}(N)}^{lat} - \boldsymbol{\alpha}_{\mathcal{K}(N)}^{lat}) \\ &= O(\mathcal{K}(N)^{-\gamma_{lat}}) + o_{\tilde{\mathbb{P}}}(\mathcal{K}(N)^{-\gamma_{lat}}) \\ &= O_{\tilde{\mathbb{P}}}(\mathcal{K}(N)^{-\gamma_{lat}}), \end{aligned} \quad (2.8)$$

where the inequality follows from (2.7) and the second equality from Lemma 2.2. \square

The first equality in (2.8) nicely illustrates that the sieve estimator is subject to two errors: an *approximation error*

$$\mathbb{E}_{\tilde{\mathbb{P}}}\left[\left(g_T(A_T(Z)) - (\boldsymbol{\alpha}_{\mathcal{K}(N)}^{lat})^T \mathbf{e}_{\mathcal{K}(N)}^{lat}(A_T(Z))\right)^2\right],$$

and an *estimation error*

$$\mathbb{E}_{\tilde{\mathbb{P}}}\left[\left((\boldsymbol{\alpha}_{\mathcal{K}(N)}^{lat})^T \mathbf{e}_{\mathcal{K}(N)}^{lat}(A_T(Z)) - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z))\right)^2\right].$$

It is worth emphasizing once more that for Regress-Later both are entirely driven by the speed of the approximation error and the growth rate of $\mathcal{K}(N)$ only. The fact that the estimation error is entirely driven by the speed of the approximation error and the growth rate of $\mathcal{K}(N)$ is a result of the fact that Equation (2.2) describes a nonstandard regression problem. Indeed, as N increases the variance of the noise term, i.e. $a_T^K(A_T(Z))$, converges to zero. We further comment on this at the end of this section.

Let us now discuss convergence rates for Regress-Now with sieves. We argued above that Regress-Later with sieves need not be considered as a nonparametric problem, because we do know g_T , A_T and $\tilde{\mathbb{P}}$. We therefore argued that we may replace the assumptions typically imposed in nonparametric settings by weaker ones. The situation is (slightly) different for Regress-Now with sieves. There we are interested in $g_{0,t}$ which is not given. Although, $g_{0,t}$ depends only on g_T , $A_T(Z)$, $\tilde{\mathbb{P}}$ and the information generated by Z up to time t , the problem of assessing certain properties to $g_{0,t}$ might be rather complicated so that one would tend to consider the problem as nonparametric even though $\tilde{\mathbb{P}}$ is known. In such a case, convergence rates and conditions needed to obtain these rates may be taken from Theorem 1 in Newey (1997).

However, if we have some knowledge about $g_{0,t}$ (which is not unlikely since the problem is not nonparametric), then we may weaken the assumptions imposed in Newey (1997) similar as we did above for Regress-Later with sieves.

We now briefly outline the Regress-Now approach under assumptions similar to Assumption 2.1 and Condition (2.4).

Assumption 2.3. There are $\gamma_{now} > 0$, $\boldsymbol{\alpha}_K^{now}$ s.t.

$$\begin{aligned} & \sqrt{\mathbb{E}_{\tilde{\mathbb{P}}} \left[(g_{0,t}(A_t(Z)) - (\boldsymbol{\alpha}_K^{now})^T \mathbf{e}_K^{now}(A_t(Z)))^4 \right]} \\ &= \sqrt{\int_{\mathbb{R}^s} (g_{0,t}(u) - (\boldsymbol{\alpha}_K^{now})^T \mathbf{e}_K^{now}(u))^4 d\tilde{\mathbb{P}}^{A_t(Z)}(u)} \\ &= \sqrt{\int_{\mathbb{R}^s} a_{0,t}^K(u)^4 d\tilde{\mathbb{P}}^{A_t(Z)}(u)} = O(K^{-\gamma_{now}}). \end{aligned}$$

We also assume that

Assumption 2.4. $((X_1, A_t(Z_1)), \dots, (X_N, A_t(Z_N)))$ are i.i.d. and $\mathbb{E}_{\tilde{\mathbb{P}}} \left[(p_{0,t}(A_T(Z)))^2 | A_t(Z) \right] = \sigma^2$.

Introduce the net $\tilde{h}^{now} : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{R}$ by

$$\tilde{h}^{now}(N, K) := \frac{1}{N} \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left((\mathbf{e}_K^{now}(A_t(Z)))^T \mathbf{e}_K^{now}(A_t(Z)) \right)^2 \right].$$

We can now state

Theorem 2.2. *Let Assumptions 2.3 and 2.4 be satisfied. Additionally, assume that there is a sequence $\mathcal{K} : \mathbb{N} \rightarrow \mathbb{N}$ such that*

$$\tilde{h}^{now}(N, \mathcal{K}(N)) \rightarrow 0 \text{ as } N \rightarrow \infty. \quad (2.9)$$

Then

$$\mathbb{E}_{\tilde{\mathbb{P}}}\left[\left(g_{0,t}(A_t(Z)) - \hat{g}_{0,t}^{\mathcal{K}(N)}(A_t(Z))\right)^2\right] = O_{\tilde{\mathbb{P}}}\left(\frac{\mathcal{K}(N)}{N} + \mathcal{K}(N)^{-\gamma_{now}}\right). \quad (2.10)$$

The result corresponds to Theorem 1 in Newey (1997), but requires weaker assumptions; see also Stentoft (2004) for a similar result, where a nonparametric setting was used for the pricing of American options. The result differs from the result of Newey (1997) and Stentoft (2004) as the convergence speed, $\mathcal{K}(N)/N + \mathcal{K}(N)^{-\gamma_{now}}$, in (2.10) is not independent of $\tilde{\mathbb{P}}$. Notice the appearance of the term $\mathcal{K}(N)/N$ in the convergence rate of the Regress-Now estimator. This term does not appear in the convergence rate of the Regress-Later estimator. We further explain this difference at the end of this section.

The proof of Theorem 2.2 is based on the following two lemmas. The first lemma is very similar to Lemma 2.1. Its proof follows along the lines of the proof of Lemma 2.1 and is therefore omitted. The second lemma is different from its counterpart for Regress-Later and it explains why we obtain the term $\mathcal{K}(N)/N$ in Equation (2.10). Its proof is given in the appendix.

Lemma 2.3. *If Condition (2.9) and Assumption 2.4 hold, we have*

$$\left\|\frac{1}{N}(\mathbf{E}_{\mathcal{K}(N)}^{now})^T \mathbf{E}_{\mathcal{K}(N)}^{now} - I_{\mathcal{K}(N)}\right\|_F = o_{\tilde{\mathbb{P}}}(1),$$

where $\|\cdot\|_F$ is again the Frobenius norm. Moreover,

$$\lambda_{\min}\left(\frac{1}{N}(\mathbf{E}_{\mathcal{K}(N)}^{now})^T \mathbf{E}_{\mathcal{K}(N)}^{now}\right) \xrightarrow{\tilde{\mathbb{P}}} 1,$$

where $\lambda_{\min}(\mathbf{A})$ denotes again the smallest eigenvalue of a matrix \mathbf{A} .

Lemma 2.4. *If Condition (2.9) and Assumptions 2.3 and 2.4 hold, we have*

$$(\hat{\boldsymbol{\alpha}}_{\mathcal{K}(N)}^{now} - \boldsymbol{\alpha}_{\mathcal{K}(N)}^{now})^T (\hat{\boldsymbol{\alpha}}_{\mathcal{K}(N)}^{now} - \boldsymbol{\alpha}_{\mathcal{K}(N)}^{now}) = O_{\mathbb{P}} \left(\frac{\mathcal{K}(N)}{N} \right) + o_{\mathbb{P}}(\mathcal{K}(N)^{-\gamma_{now}}).$$

We now give the proof of the above theorem.

Proof of Theorem 2.2. Similar to the proof of Theorem 2.1 observe first that Assumption 2.3 implies

$$\mathbb{E}_{\mathbb{P}} \left[\left(g_{0,t}(A_t(Z)) - (\boldsymbol{\alpha}_K^{now})^T \mathbf{e}_K^{now}(A_t(Z)) \right)^2 \right] \leq O(K^{-\gamma_{now}}) \quad (2.11)$$

by Cauchy-Schwarz. Moreover,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} \left[e_k^{now}(A_t(Z)) \left(g_{0,t}(A_t(Z)) - (\boldsymbol{\alpha}_K^{now})^T \mathbf{e}_K^{now}(A_t(Z)) \right) \right] &= 0, \\ k &= 1, \dots, K. \end{aligned}$$

Hence,

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}} \left[\left(g_{0,t}(A_t(Z)) - \hat{g}_{0,t}^{\mathcal{K}(N)}(A_t(Z)) \right)^2 \right] \\ &= \mathbb{E}_{\mathbb{P}} \left[\left((\boldsymbol{\alpha}_{\mathcal{K}(N)}^{now})^T \mathbf{e}_{\mathcal{K}(N)}^{now}(A_t(Z)) - \hat{g}_{0,t}^{\mathcal{K}(N)}(A_t(Z)) \right)^2 \right] \\ &\quad + \mathbb{E}_{\mathbb{P}} \left[\left(g_{0,t}(A_t(Z)) - (\boldsymbol{\alpha}_{\mathcal{K}(N)}^{now})^T \mathbf{e}_{\mathcal{K}(N)}^{lat}(A_t(Z)) \right)^2 \right] \\ &\leq (\hat{\boldsymbol{\alpha}}_{\mathcal{K}(N)}^{now} - \boldsymbol{\alpha}_{\mathcal{K}(N)}^{now})^T (\hat{\boldsymbol{\alpha}}_{\mathcal{K}(N)}^{now} - \boldsymbol{\alpha}_{\mathcal{K}(N)}^{now}) + O(\mathcal{K}(N)^{-\gamma_{now}}) \\ &= O_{\mathbb{P}} \left(\frac{\mathcal{K}(N)}{N} \right) + O(\mathcal{K}(N)^{-\gamma_{now}}), \end{aligned} \quad (2.12)$$

where the inequality follows from (2.11) and the second equality from Lemma 2.4. \square

As for Regress-Later the first equality in (2.12) illustrates that also Regress-Now is subject to two errors: an *approximation error*

$$\mathbb{E}_{\mathbb{P}} \left[\left(g_{0,t}(A_t(Z)) - (\boldsymbol{\alpha}_{\mathcal{K}(N)}^{now})^T \mathbf{e}_{\mathcal{K}(N)}^{lat}(A_t(Z)) \right)^2 \right],$$

and an *estimation error*

$$\mathbb{E}_{\mathbb{P}} \left[\left((\boldsymbol{\alpha}_{\mathcal{K}(N)}^{now})^T \mathbf{e}_{\mathcal{K}(N)}^{now}(A_t(Z)) - \hat{g}_{0,t}^{\mathcal{K}(N)}(A_t(Z)) \right)^2 \right].$$

Notice that as for Regress-Later the approximation error is also driven by the speed of the approximation error and the growth rate of $\mathcal{K}(N)$. However, note also the difference compared to Regress-Later: The estimation error is driven by the ratio of $\mathcal{K}(N)$ to N and the approximation error. The difference can also be seen from the following equations where we omitted the superscripts *now* and *lat*

$$\begin{aligned} (\hat{\boldsymbol{\alpha}}_K - \boldsymbol{\alpha}_K) &= (\mathbf{E}_K^T \mathbf{E}_K)^{-1} \mathbf{E}_K^T (\mathbf{X} - \mathbf{E}_K \boldsymbol{\alpha}_K) \\ &= (\mathbf{E}_K^T \mathbf{E}_K)^{-1} \mathbf{E}_K^T ((\mathbf{X} - \mathbf{E} \boldsymbol{\alpha}) + (\mathbf{E} \boldsymbol{\alpha} - \mathbf{E}_K \boldsymbol{\alpha}_K)) \\ &= (\mathbf{E}_K^T \mathbf{E}_K)^{-1} \mathbf{E}_K^T (\mathbf{p} + \mathbf{a}^K), \end{aligned}$$

where \mathbf{E} is an infinite-dimensional matrix containing all basis functions and $\boldsymbol{\alpha}$ is the true (infinite-dimensional) parameter vector. Here \mathbf{p} gives the projection error which is zero for Regress-Later. However, it is unequal zero for Regress-Now. Moreover, for both Regress-Now and Regress-Later the variance of the approximation error a^K converges to zero. However, as can be seen from the proof of Lemma 2.4 in the Appendix 2.A, it is the projection error that contributes the rate $\mathcal{K}(N)/N$ to the estimation error in Regress-Now.

The absence of the term $\mathcal{K}(N)/N$ in the mean-square error of Regress-Later makes it plausible that the Regress-Later estimator may potentially converge faster than the Regress-Now estimator. We deliberately

state here “potentially” as the ultimate convergence rate depends on the γ_{now} and γ_{lat} which are problem-dependent. In particular, the choice of basis plays an important role in the determination of γ_{now} and γ_{lat} . However, it is clear that the Regress-Now convergence rate can never be faster than N^{-1} . This follows simply from the fact that the best we can hope for is that $g_{0,t}$ is contained in the span of finitely many basis functions. Then the approximation error vanishes and we are left with the rate N^{-1} . In contrast, in Regress-Later if Condition (2.4) is fulfilled with $\mathcal{K}(N) \propto N^a$ for some $0 < a < 1$, then the convergence rate for Regress-Later equals $N^{-a \gamma_{lat}}$. We can see that for the right combination of a and γ_{lat} it is possible to achieve a convergence rate that is even faster than N^{-1} . An example will be provided in Section 2.4.

We finally comment on the fact that the discussed convergence rates pertain to slightly different problems. The speed of convergence for the Regress-Now estimator refers to convergence to the conditional expectation function $g_{0,t}$. On the contrary, the discussed convergence rate for the Regress-Later estimator pertains to convergence to the payoff function X . As discussed in Section 2.2.2 in Regress-Later we achieve the approximation to the conditional expectation function by applying the conditional expectation operator to the estimated payoff function, $\hat{g}_T^{\mathcal{K}(N)}$. We thereby do not incur a projection error as long as the conditional expectations of the basis functions have closed-form solutions. We can show that the convergence rate of the conditional expectation of the Regress-Later estimator to the conditional expectation of X , i.e. $g_{0,t}(A_t(Z))$ is implied by the convergence of the Regress-Later estimator to X . More explicitly we have

$$\begin{aligned}
 & \mathbb{E}_{\hat{\mathbb{P}}} \left[\left(g_{0,t}(A_t(Z)) - \mathbb{E}_{\hat{\mathbb{P}}} \left[\hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) | \mathcal{F}_t \right] \right)^2 \right] \\
 &= \mathbb{E}_{\hat{\mathbb{P}}} \left[\left(\mathbb{E}_{\hat{\mathbb{P}}} \left[X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) | \mathcal{F}_t \right] \right)^2 \right] \\
 &\leq \mathbb{E}_{\hat{\mathbb{P}}} \left[\mathbb{E}_{\hat{\mathbb{P}}} \left[\left(X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) \right)^2 | \mathcal{F}_t \right] \right] \\
 &= \mathbb{E}_{\hat{\mathbb{P}}} \left[\left(X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) \right)^2 \right],
 \end{aligned}$$

where the first inequality follows from Jensen's inequality for conditional expectations and the last equality uses the projection law of expectations.

2.4 Orthonormal piecewise linear functions as sieves

In this section, we show that the convergence rate in mean-square for Regress-Later can indeed be faster than N^{-1} . To present this claim in a very simple set-up and to avoid technicalities that are of no relevance for our claim we consider a compact interval. The convergence rate depends on the properties of the basis. Typically applied bases are polynomials. Here we consider a basis consisting of piecewise linear functions, for which both the construction as well as the analysis in view of establishing convergence rates simplifies. Importantly, the suggested basis is by construction orthogonal and can be easily set up. Moreover, the results of Theorem 2.1 can be explicitly calculated for the piecewise linear functions as the derived γ_{lat} applies to a large class of functions.

We have pointed out earlier that in the Regress-Later approach the computation of the conditional expectation of the basis terms must be readily available. When it is and when it is not possible to compute these conditional expectations depends on both the structure of the basis function and the complexity of $A_T(Z)$. We will later address this for the piecewise linear functions.

We now outline the Regress-Later estimation with piecewise linear functions as sieves. Let $D = [a_1, a_2] \subset \mathbb{R}$ denote the support of $g_T(A_T(Z))$. We construct an orthonormal basis on $L_2(D, \mathcal{B}(D), \tilde{\mathbb{P}}^{A_T(Z)})$ based on non-overlapping linear functions. We require the following assumption in order to construct our basis functions

Assumption 2.5. $g_T(A_T(Z))$ has a density w.r.t. Lebesgue measure which is a positive continuous function on D .

Then, the domain D is chopped into K intervals, $[b_k, b_{k+1})$, $k = 1, \dots, K + 1$, such that $\tilde{\mathbb{P}}^{A_T(Z)}(b_k \leq A_T(Z) < b_{k+1}) = 1/K$, $\forall k = 1, \dots, K$. Assumption 2.5 ensures that as the truncation parameter K grows, the intervals can be made arbitrarily small and cover each probability $1/K$. Define K non-overlapping indicator functions

$$\mathbb{1}_k^{lat}(u) := \begin{cases} 1 & \text{if } u \in [b_k, b_{k+1}) \\ 0 & \text{otherwise} \end{cases}$$

for $k = 1, \dots, K$. By construction the indicator functions are orthogonal. On each interval two basis functions are now defined:

$$\begin{aligned} e_{0k}^{lat}(u) &:= C_{0k} \mathbb{1}_k^{lat}(u) \\ e_{1k}^{lat}(u) &:= C_{1k} \mathbb{1}_k^{lat}(u)(u - c_k), \end{aligned}$$

where C_{0k} , C_{1k} and c_k are chosen such that $e_{0k}(A_T(Z))$ and $e_{1j}(A_T(Z))$ are orthonormal $\forall k, j$. Hence,

$$\begin{aligned} C_{0k} &= \sqrt{K}, \\ C_{1k} &= 1/\sqrt{\mathbb{E}_{\mathbb{P}} [\mathbf{1}_k^{lat}(A_T(Z))(A_T(Z) - c_k)^2]} \text{ and} \\ c_k &= K \mathbb{E}_{\mathbb{P}} [\mathbf{1}_k^{lat}(A_T(Z))A_T(Z)]. \end{aligned}$$

By construction we then have the following orthonormality results

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} [e_{0k}^{lat}(A_T(Z))e_{0j}^{lat}(A_T(Z))] &= \delta_{kj} \\ \mathbb{E}_{\mathbb{P}} [e_{1k}^{lat}(A_T(Z))e_{1j}^{lat}(A_T(Z))] &= \delta_{kj} \\ \mathbb{E}_{\mathbb{P}} [e_{0k}^{lat}(A_T(Z))e_{1j}^{lat}(A_T(Z))] &= 0, \end{aligned}$$

where δ_{kj} denotes the Kronecker delta.

Assumption 2.6. g_T is twice continuously differentiable on (a_1, a_2) and there is a $B < \infty$ such that $\sup_{u \in (a_1, a_2)} |g_T''(u)| \leq B$.

Lemma 2.5. *If Assumptions 2.5 and 2.6 hold, the deterministic approximation error vanishes as $K \rightarrow \infty$;*

$$\sqrt{\mathbb{E}_{\mathbb{P}} \left[(g_T(A_T(Z)) - g_T^K(A_T(Z)))^4 \right]} = O(K^{-4}).$$

Proof. See Appendix 2.A. □

Consequently, Assumption 2.1 is satisfied with $\gamma_{lat} = 4$. For Equation (2.4) we obtain

$$\begin{aligned} &\tilde{h}^{lat}(N, \mathcal{K}(N)) \\ &= \frac{1}{N} \mathbb{E}_{\mathbb{P}} \left[\left(\sum_{k=1}^{\mathcal{K}(N)} (e_{0k}^{lat}(A_T(Z))^2 + e_{1k}^{lat}(A_T(Z))^2) \right)^2 \right] \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{N} \mathbb{E}_{\tilde{\mathbb{P}}} \left[\sum_{\ell=1}^{\mathcal{K}(n)} \sum_{j=1}^{\mathcal{K}(n)} (e_{0\ell}^{\text{lat}}(A_T(Z))^2 + e_{1\ell}^{\text{lat}}(A_T(Z))^2) \right. \\
 &\quad \left. (e_{0j}^{\text{lat}}(A_T(Z))^2 + e_{1j}^{\text{lat}}(A_T(Z))^2) \right] \\
 &= \frac{1}{N} \mathbb{E}_{\tilde{\mathbb{P}}} \left[\sum_{k=1}^{\mathcal{K}(N)} (e_{0k}^{\text{lat}}(A_T(Z))^2 + e_{1k}^{\text{lat}}(A_T(Z))^2)^2 \right] \\
 &= \frac{1}{N} \sum_{k=1}^{\mathcal{K}(N)} \mathbb{E}_{\tilde{\mathbb{P}}} \left[\mathcal{K}(N)^2 \mathbf{1}_k^{\text{lat}}(A_T(Z)) + 2\mathcal{K}(N) \mathbf{1}_k^{\text{lat}}(A_T(Z)) \right. \\
 &\quad \left. C_{1k}^2 (A_T(Z) - c_k)^2 + C_{1k}^4 \mathbf{1}_k^{\text{lat}}(A_T(Z)) (A_T(Z) - c_k)^4 \right] \\
 &= \frac{1}{N} \sum_{k=1}^{\mathcal{K}(N)} \left(\mathcal{K}(N) + 2\mathcal{K}(N) + \right. \\
 &\quad \left. C_{1k}^4 \mathbb{E}_{\tilde{\mathbb{P}}} \left[\mathbf{1}_k^{\text{lat}}(A_T(Z)) (A_T(Z) - c_k)^4 \right] \right) \\
 &\leq \frac{3\mathcal{K}(N)^2}{N} + \frac{\mathcal{K}(N)}{N} \max_k \left(\frac{\mathbb{E}_{\tilde{\mathbb{P}}} \left[\mathbf{1}_k^{\text{lat}}(A_T(Z)) (A_T(Z) - c_k)^4 \right]}{\left(\mathbb{E}_{\tilde{\mathbb{P}}} \left[\mathbf{1}_k^{\text{lat}}(A_T(Z)) (A_T(Z) - c_k)^2 \right] \right)^2} \right). \tag{2.13}
 \end{aligned}$$

Assumption 2.5 ensures that there is enough variation on each arbitrary interval such that the denominator in (2.13) is greater than zero. Moreover, it makes sure that the last term in the last line of Equation (2.13) does not grow faster than the first term in the last line of Equation (2.13). Moreover, the particular growth rate can be determined.

Lemma 2.6. *If Assumption 2.5 is satisfied the following result holds*

$$\max_{1 \leq k \leq \mathcal{K}(N)} \frac{\mathbb{E}_{\tilde{\mathbb{P}}} \left[\mathbf{1}_k^{\text{lat}}(A_T(Z)) (A_T(Z) - c_k)^4 \right]}{\left(\mathbb{E}_{\tilde{\mathbb{P}}} \left[\mathbf{1}_k^{\text{lat}}(A_T(Z)) (A_T(Z) - c_k)^2 \right] \right)^2} \leq O(\mathcal{K}(N)).$$

Proof. See Appendix 2.A. □

Hence, by combining Lemma 2.6 and Equation (2.13)

$$\tilde{h}^{lat}(N, \mathcal{K}(N)) \leq O\left(\frac{\mathcal{K}(N)^2}{N}\right).$$

A sufficient condition for $\tilde{h}^{lat}(N, \mathcal{K}(N)) \rightarrow 0$ as $N \rightarrow \infty$ is that $\mathcal{K}(N) \propto N^a$, and with $a < 1/2$ Condition (2.4) in Theorem 2.1 holds. Now Theorem 2.1 is applicable and gives the convergence rate in mean-square

$$\mathbb{E}_{\tilde{\mathbb{P}}}\left[\left(g_T(A_T(Z)) - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z))\right)^2\right] = O_{\tilde{\mathbb{P}}}(\mathcal{K}(N)^{-4}). \quad (2.14)$$

We immediately see that for $\mathcal{K}(N) \propto N^a$ and choosing a only slightly smaller than $1/2$ we almost achieve a convergence rate of N^{-2} , which is considerably faster than the conventional Monte Carlo rate of N^{-1} .

We now look at the Regress-Later estimator with orthonormal piecewise linear functions as sieve for a specific underlying random variable.

Brownian Motion

We consider a Brownian Motion $W(T)$ as the underlying for a function $g_T(W(T))$ fulfilling the necessary conditions for Equation (2.14). We consider a compact domain $[a_1, a_2]$ for $W(T)$ and define the K intervals $[b_k, b_{k+1}]$, $k = 1, \dots, K$ such that each covers probability $1/K$. The conditional normal density is then $\varphi(w|D) = \varphi(z)/\mathbb{P}(W(T) \in D)$ with $\varphi(w) = \exp(-w^2/(2T))/\sqrt{2\pi T}$ the normal density. Assumption 2.5 is immediately satisfied. Thus, Lemma 2.6 applies. Then, again choosing $\mathcal{K}(N) \propto N^a$ with a only slightly smaller than $1/2$ produces a mean square error that converges in probability almost at rate N^{-2} .

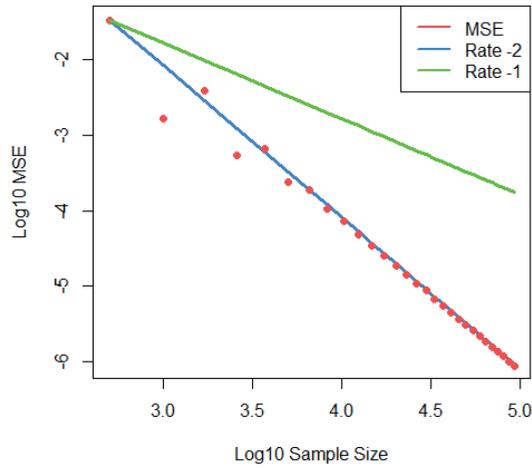


Figure 2.1: Regress-Later convergence plot with K up to 30.

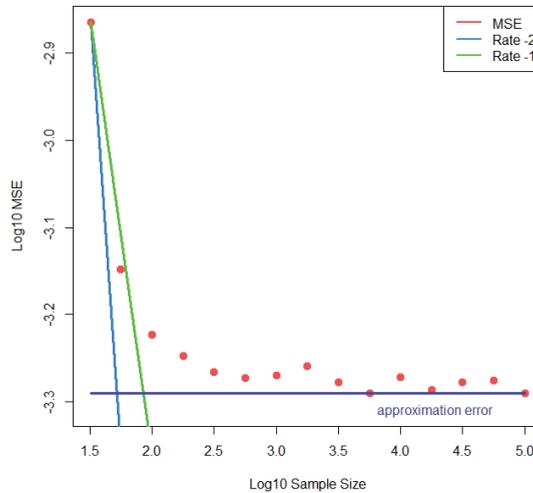


Figure 2.2: Regress-Later convergence plot with $K = 5$ fixed.

Figure 2.1 gives the convergence for $X = \tanh(W(10))$ for K up to 30 and $N = 100 K^{2.01}$. The logarithm of the mean-square error is plotted against the logarithm of the sample size. The green line indicates a convergence at rate N^{-1} while the blue line indicates a convergence at N^{-2} . The actual mean-square error of the data is represented by the red dots. We see that we can already achieve the fast convergence rate of N^{-2} in finite samples. Figure 2.2 gives the mean square error for the payoff function where the number of basis functions is fixed at $K = 5$ and only the sample size grows up to 10^5 . The example illustrates that the mean square error does not converge further if only the sample size is increased. While the sampling error decreases with the growth of the sample size the approximation error only converges when the number of basis functions grows.

Remark 2.3. (On the computation of the conditional expectation of piecewise linear functions)

The Regress-Later method relies on basis functions, for which the conditional expectation function is fairly easily computed. For piecewise linear functions we can easily see that these are (approximately) equal to linear combinations of call options. The piecewise linear function has two components, the piecewise constant given by $e_{0k}(A_T(Z)) = \mathbb{1}_k(A_T(Z))C_{0k}$ and the piecewise linear part $e_{1k}(A_T(Z)) = \mathbb{1}_k(A_T(Z))C_{1k}(A_T(Z) - c_k)$. In an interval $[b_k, b_{k+1})$ we thus have three components: C_{0k} , $\mathbb{1}_k(A_T(Z))A_T(Z)$ and $-\mathbb{1}_k(A_T(Z))c_k$. The first and third component can be perfectly expressed in terms of long and short cash-or-nothing digital options. The second component can be perfectly replicated by a portfolio of long and short calls and cash-or-nothing digital options. A digital option pays out some fixed amount if the option expires in-the-money, i.e. when the underlying exceeds the strike value. It is approximately statically replicated by the combination of a long and short call where the strikes differ by a

very small amount. Consequently, the piecewise linear functions are (approximately) equal to portfolios of call options, for which closed-form solutions are available given the structure of the underlying is not (highly) path-dependent. Here, the underlying is $A_T(Z)$. Note that for a specific problem $A_T(Z)$ is not unique, meaning that it is reasonable to expect that $A_T(Z)$ can be chosen such that the conditional expectation of call options on it can be computed.

2.5 Conclusion

In this paper the discussion on Regress-Later estimators is picked up and addressed in comparison to Regress-Now estimators, which are currently more popular. Both estimators refer to LSMC solutions. Clarification is given on the functionality of Regress-Now and Regress-Later estimators based on several examples. Examples have been discussed that help to better understand the differences of Regress-Now and Regress-Later. The estimation approach for both estimators is outlined and the regression error for each is specified. It is shown that in Regress-Later the involved regression is nonstandard as the regression error corresponds to the approximation error, which vanishes in the limit. In contrast, the regression error of Regress-Now estimators contains an approximation and a projection error. While the approximation error vanishes in the limit the projection error is not eliminated. This leads to different convergence rates for Regress-Now and Regress-Later estimators. The current literature addresses convergence rates for Regress-Now estimators in a nonparametric setting. In this paper it is shown that the problem specification for Regress-Later is not nonparametric. This allows to relax the conditions typically necessary in nonparametric problems solved with sieve. Moreover, it is indicated that a nonparametric problem specification may also apply

to Regress-Now estimators, which then similarly allows for weaker conditions. A specific basis is constructed based on piecewise linear functions and the Regress-Later convergence rate with this basis is derived explicitly. The result shows that Regress-Later estimators can be constructed such that they converge faster than the more often applied Regress-Now estimators.

2.A Appendix: Proofs

Proof of Lemma 2.1. We have

$$\begin{aligned} & \left\| \frac{1}{N} (\mathbf{E}_{\mathcal{K}(N)}^{lat})^T \mathbf{E}_{\mathcal{K}(N)}^{lat} - I_{\mathcal{K}(N)} \right\|_F^2 \\ &= \sum_{j=1}^{\mathcal{K}(N)} \sum_{\ell=1}^{\mathcal{K}(N)} \left(\frac{1}{N} \sum_{n=1}^N e_j^{lat}(A_T(z_n)) e_\ell^{lat}(A_T(z_n)) \right. \\ & \quad \left. - \mathbb{E}_{\tilde{\mathbb{P}}} [e_j^{lat}(A_T(z_n)) e_\ell^{lat}(A_T(z_n))] \right)^2 \end{aligned}$$

Therefore,

$$\begin{aligned} & \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left\| \frac{1}{N} (\mathbf{E}_{\mathcal{K}(N)}^{lat})^T \mathbf{E}_{\mathcal{K}(N)}^{lat} - I_{\mathcal{K}(N)} \right\|_F^2 \right] \\ &= \frac{1}{N} \sum_{j=1}^{\mathcal{K}(N)} \sum_{\ell=1}^{\mathcal{K}(N)} \mathbb{V}\text{ar}_{\tilde{\mathbb{P}}} [e_j^{lat}(A_T(Z)) e_\ell^{lat}(A_T(Z))] \\ &\leq \frac{1}{N} \sum_{j=1}^{\mathcal{K}(N)} \sum_{\ell=1}^{\mathcal{K}(N)} \mathbb{E}_{\tilde{\mathbb{P}}} \left[(e_j^{lat}(A_T(Z)) e_\ell^{lat}(A_T(Z)))^2 \right] \\ &= o(1). \end{aligned}$$

Now, (2.5) follows by Markov's inequality. Since $I_{\mathcal{K}(N)}$ is the identity matrix we have:

$$\lambda_{\min} \left(\frac{1}{N} (\mathbf{E}_{\mathcal{K}(N)}^{lat})^T \mathbf{E}_{\mathcal{K}(N)}^{lat} \right) - 1 = \lambda_{\min} \left(\frac{1}{N} (\mathbf{E}_{\mathcal{K}(N)}^{lat})^T \mathbf{E}_{\mathcal{K}(N)}^{lat} - I_{\mathcal{K}(N)} \right).$$

The result now follows from the fact that the smallest eigenvalue of a matrix is bounded above by its Frobenius norm and that therefore (2.5) implies (2.6). \square

Proof of Lemma 2.2. By the standard representation of the empirical error $(\hat{\boldsymbol{\alpha}}_K^{lat} - \boldsymbol{\alpha}_K^{lat})$ for least squares estimators it follows that

$$(\hat{\boldsymbol{\alpha}}_K^{lat} - \boldsymbol{\alpha}_K^{lat}) = \left((\mathbf{E}_K^{lat})^T \mathbf{E}_K^{lat} \right)^{-1} (\mathbf{E}_K^{lat})^T \mathbf{a}_T^K.$$

Putting $\mathbf{B}_{\mathcal{K}(N)} = \left((1/N) (\mathbf{E}_{\mathcal{K}(N)}^{lat})^T \mathbf{E}_{\mathcal{K}(N)}^{lat} \right)$ we have by the above representation for the empirical error

$$\hat{\boldsymbol{\alpha}}_{\mathcal{K}(N)}^{lat} - \boldsymbol{\alpha}_{\mathcal{K}(N)}^{lat} = \mathbf{B}_{\mathcal{K}(N)}^{-1} \frac{1}{N} (\mathbf{E}_{\mathcal{K}(N)}^{lat})^T \mathbf{a}_T^{\mathcal{K}(N)}.$$

Then

$$\begin{aligned} & (\hat{\boldsymbol{\alpha}}_{\mathcal{K}(N)}^{lat} - \boldsymbol{\alpha}_{\mathcal{K}(N)}^{lat})^T (\hat{\boldsymbol{\alpha}}_{\mathcal{K}(N)}^{lat} - \boldsymbol{\alpha}_{\mathcal{K}(N)}^{lat}) \\ &= \frac{1}{N^2} (\mathbf{a}_T^{\mathcal{K}(N)})^T \mathbf{E}_{\mathcal{K}(N)}^{lat} \mathbf{B}_{\mathcal{K}(N)}^{-1} \mathbf{B}_{\mathcal{K}(N)}^{-1} (\mathbf{E}_{\mathcal{K}(N)}^{lat})^T \mathbf{a}_T^{\mathcal{K}(n)} \\ &\leq \frac{1}{N^2} \left(\lambda_{\max} (\mathbf{B}_{\mathcal{K}(N)}^{-1}) \right)^2 (\mathbf{a}_T^{\mathcal{K}(N)})^T \mathbf{E}_{\mathcal{K}(N)}^{lat} (\mathbf{E}_{\mathcal{K}(N)}^{lat})^T \mathbf{a}_T^{\mathcal{K}(N)}, \end{aligned} \tag{2.15}$$

where $\lambda_{\max}(\mathbf{A})$ denotes the largest eigenvalue of a matrix \mathbf{A} . Notice that by Assumption 2.2

$$\frac{1}{N^2} \mathbb{E}_{\tilde{\mathbb{P}}} \left[(\mathbf{a}_T^{\mathcal{K}(N)})^T \mathbf{E}_{\mathcal{K}(N)}^{lat} (\mathbf{E}_{\mathcal{K}(N)}^{lat})^T \mathbf{a}_T^{\mathcal{K}(N)} \right]$$

$$\begin{aligned}
 &= \frac{1}{N} \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(\mathbf{a}_T^{\mathcal{K}(N)} (A_T(Z)) \right)^2 \left(\mathbf{e}_{\mathcal{K}(N)}^{lat} \right)^T \mathbf{e}_{\mathcal{K}(N)}^{lat} \right] \\
 &\leq \frac{1}{N} \sqrt{\mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(\mathbf{a}_T^{\mathcal{K}(N)} (A_T(Z)) \right)^4 \right]} \sqrt{\mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(\left(\mathbf{e}_{\mathcal{K}(N)}^{lat} \right)^T \mathbf{e}_{\mathcal{K}(N)}^{lat} \right)^2 \right]},
 \end{aligned} \tag{2.16}$$

where we used the Cauchy-Schwarz inequality. Using Assumption 2.1 and Condition (2.4) we get from (2.16)

$$\frac{1}{N^2} \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(\mathbf{a}_T^{\mathcal{K}(N)} \right)^T \mathbf{E}_{\mathcal{K}(N)}^{lat} \left(\mathbf{E}_{\mathcal{K}(N)}^{lat} \right)^T \mathbf{a}_T^{\mathcal{K}(N)} \right] = o(\mathcal{K}(N)^{-\gamma_{lat}}).$$

By Markov's inequality it follows

$$\frac{1}{N^2} \left(\mathbf{a}_T^{\mathcal{K}(N)} \right)^T \mathbf{E}_{\mathcal{K}(N)}^{lat} \left(\mathbf{E}_{\mathcal{K}(N)}^{lat} \right)^T \mathbf{a}_T^{\mathcal{K}(N)} = o_{\tilde{\mathbb{P}}}(\mathcal{K}(N)^{-\gamma_{lat}}).$$

Since $\lambda_{\max}(\mathbf{B}_{\mathcal{K}(N)}^{-1}) = (\lambda_{\min}(\mathbf{B}_{\mathcal{K}(N)}))^{-1}$ Equation (2.6) implies that $\lambda_{\max}(\mathbf{B}_{\mathcal{K}(N)}^{-1}) = O_{\tilde{\mathbb{P}}}(1)$. Putting everything together we get that (2.15) is indeed $o_{\tilde{\mathbb{P}}}(\mathcal{K}(N)^{-\gamma_{lat}})$. \square

Proof of Lemma 2.4. By the standard representation of the empirical error $(\hat{\boldsymbol{\alpha}}_K^{now} - \boldsymbol{\alpha}_K^{now})$ for least squares estimators it follows that

$$(\hat{\boldsymbol{\alpha}}_K^{now} - \boldsymbol{\alpha}_K^{now}) = \left((\mathbf{E}_K^{now})^T \mathbf{E}_K^{now} \right)^{-1} (\mathbf{E}_K^{now})^T (\mathbf{a}_{0,t}^K + \mathbf{p}_{0,t}).$$

Putting $\mathbf{B}_{\mathcal{K}(N)} = \left((1/N) \left(\mathbf{E}_{\mathcal{K}(N)}^{now} \right)^T \mathbf{E}_{\mathcal{K}(N)}^{now} \right)$ we have again by the above representation for the empirical error

$$\hat{\boldsymbol{\alpha}}_{\mathcal{K}(N)}^{now} - \boldsymbol{\alpha}_{\mathcal{K}(N)}^{now} = \mathbf{B}_{\mathcal{K}(N)}^{-1} \frac{1}{N} \left(\mathbf{E}_{\mathcal{K}(N)}^{now} \right)^T \left(\mathbf{p}_{0,t} + \mathbf{a}_{0,t}^{\mathcal{K}(N)} \right).$$

Then

$$\begin{aligned}
 & (\hat{\boldsymbol{\alpha}}_{\mathcal{K}(N)}^{now} - \boldsymbol{\alpha}_{\mathcal{K}(N)}^{now})^T (\hat{\boldsymbol{\alpha}}_{\mathcal{K}(N)}^{now} - \boldsymbol{\alpha}_{\mathcal{K}(N)}^{now}) \\
 &= \frac{1}{N^2} \left(\mathbf{a}_{0,t}^{\mathcal{K}(N)} + \mathbf{p}_{0,t} \right)^T \mathbf{E}_{\mathcal{K}(N)}^{now} \mathbf{B}_{\mathcal{K}(N)}^{-1} \mathbf{B}_{\mathcal{K}(N)}^{-1} \left(\mathbf{E}_{\mathcal{K}(N)}^{now} \right)^T \left(\mathbf{a}_{0,t}^{\mathcal{K}(N)} + \mathbf{p}_{0,t} \right) \\
 &\leq \frac{1}{N^2} \left(\lambda_{max} \left(\mathbf{B}_{\mathcal{K}(N)}^{-1} \right) \right)^2 \left(\mathbf{a}_{0,t}^{\mathcal{K}(N)} + \mathbf{p}_{0,t} \right)^T \mathbf{E}_{\mathcal{K}(N)}^{now} \left(\mathbf{E}_{\mathcal{K}(N)}^{now} \right)^T \\
 &\quad \left(\mathbf{a}_{0,t}^{\mathcal{K}(N)} + \mathbf{p}_{0,t} \right),
 \end{aligned}$$

where $\lambda_{max}(\mathbf{A})$ denotes the largest eigenvalue of a matrix \mathbf{A} . By Assumption 2.4 we have

$$\begin{aligned}
 & \frac{1}{N^2} \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(\mathbf{a}_{0,t}^{\mathcal{K}(N)} + \mathbf{p}_{0,t} \right)^T \mathbf{E}_{\mathcal{K}(N)}^{now} \left(\mathbf{E}_{\mathcal{K}(N)}^{now} \right)^T \left(\mathbf{a}_{0,t}^{\mathcal{K}(N)} + \mathbf{p}_{0,t} \right) \right] \\
 &= \frac{1}{N} \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(\mathbf{a}_{0,t}^{\mathcal{K}(N)}(A_t(Z)) + p_{0,t}(A_T(Z)) \right)^2 \left(\mathbf{e}_{\mathcal{K}(N)}^{now} \right)^T \mathbf{e}_{\mathcal{K}(N)}^{now} \right].
 \end{aligned} \tag{2.17}$$

Now, notice that

$$\begin{aligned}
 & \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(p_{0,t}(A_T(Z)) \right)^2 \left(\mathbf{e}_{\mathcal{K}(N)}^{now} \right)^T \mathbf{e}_{\mathcal{K}(N)}^{now} \right] \\
 &= \mathbb{E}_{\tilde{\mathbb{P}}} \left[\mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(p_{0,t}(A_T(Z)) \right)^2 \left(\mathbf{e}_{\mathcal{K}(N)}^{now} \right)^T \mathbf{e}_{\mathcal{K}(N)}^{now} \mid A_t(Z) \right] \right] \\
 &= \mathbb{E}_{\tilde{\mathbb{P}}} \left[\mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(p_{0,t}(A_T(Z)) \right)^2 \mid A_t(Z) \right] \left(\mathbf{e}_{\mathcal{K}(N)}^{now} \right)^T \mathbf{e}_{\mathcal{K}(N)}^{now} \right] \\
 &= \sigma^2 \mathcal{K}(N).
 \end{aligned}$$

Moreover, since $\mathbb{E}_{\tilde{\mathbb{P}}} [p_{0,t}(A_T(Z)) \mid A_t(Z)] = 0$ and since $a_{0,t}^{\mathcal{K}(N)} = \sum_{\ell=\mathcal{K}(N)+1}^{\infty} \alpha_{\ell}^{now} e_{\ell}^{now}(A_t(Z))$ implying that $\mathbb{E}_{\tilde{\mathbb{P}}} [a_{0,t}^{\mathcal{K}(N)}(A_t(Z)) \mid A_t(Z)] = a_{0,t}^{\mathcal{K}(N)}(A_t(Z))$, we obtain

$$\begin{aligned}
 & \mathbb{E}_{\tilde{\mathbb{P}}} \left[a_{0,t}^{\mathcal{K}(N)}(A_t(Z)) p_{0,t}(A_T(Z)) \left(\mathbf{e}_{\mathcal{K}(N)}^{now} \right)^T \mathbf{e}_{\mathcal{K}(N)}^{now} \right] \\
 &= \mathbb{E}_{\tilde{\mathbb{P}}} \left[\mathbb{E}_{\tilde{\mathbb{P}}} [p_{0,t}(A_T(Z)) \mid A_t(Z)] a_{0,t}^{\mathcal{K}(N)}(A_t(Z)) \left(\mathbf{e}_{\mathcal{K}(N)}^{now} \right)^T \mathbf{e}_{\mathcal{K}(N)}^{now} \right]
 \end{aligned}$$

$$= 0.$$

Hence,

$$\begin{aligned}
 (2.17) &= \frac{1}{N} \mathbb{E}_{\mathbb{P}} \left[\left(a_{0,t}^{\mathcal{K}(N)}(A_t(Z)) \right)^2 \left(\mathbf{e}_{\mathcal{K}(N)}^{\text{now}} \right)^T \mathbf{e}_{\mathcal{K}(N)}^{\text{now}} \right] + \frac{\sigma^2 \mathcal{K}(N)}{N} \\
 &\leq \frac{1}{N} \sqrt{\mathbb{E}_{\mathbb{P}} \left[\left(a_{0,t}^{\mathcal{K}(N)}(A_t(Z)) \right)^4 \right]} \sqrt{\mathbb{E}_{\mathbb{P}} \left[\left(\left(\mathbf{e}_{\mathcal{K}(N)}^{\text{now}} \right)^T \mathbf{e}_{\mathcal{K}(N)}^{\text{now}} \right)^2 \right]} \\
 &\quad + \frac{\sigma^2 \mathcal{K}(N)}{N}
 \end{aligned}$$

where we used the Cauchy-Schwarz inequality. Using Assumption 2.3 and (2.9) we have that (2.17) is $o(\mathcal{K}(N)^{-\gamma_{\text{now}}}) + O(\mathcal{K}(N)/N)$. The remaining steps are now as in the proof of Theorem 2.1. \square

Proof of Lemma 2.5. Let f be the density on D . Then $m := \min_{u \in D} f(u) > 0$ and $M := \max_{u \in D} f(u) < \infty$. We approximate the coefficients α_{0k} and α_{1k} by $g_T(c_k)/\sqrt{K}$ and $g'_T(c_k)/C_{1k}$, respectively. By a first order Taylor expansion around c_k with Lagrange's form of the remainder term, i.e.

$$g_T(u) = g_T(c_k) + g'_T(c_k)(u - c_k) + \frac{1}{2} g''_T(\xi)(u - c_k)^2, \quad \xi \in [u, c_k],$$

we obtain

$$\begin{aligned}
 &\sqrt{K} \alpha_{0k} \\
 &= K \mathbb{E}_{\mathbb{P}} [g_T(A_T(Z)) \mathbb{1}_k(A_T(Z))] \\
 &= K \int_{b_k}^{b_{k+1}} \left(g_T(c_k) + g'_T(c_k)(u - c_k) + \frac{1}{2} g''_T(\xi)(u - c_k)^2 \right) f(u) \, du \\
 &= g_T(c_k) + \frac{K}{2} \int_{b_k}^{b_{k+1}} g''_T(\xi)(u - c_k)^2 f(u) \, du,
 \end{aligned}$$

and

$$\begin{aligned}
 & C_{1k} \alpha_{1k} \\
 &= C_{1k}^2 \mathbb{E}_{\bar{\mathbb{P}}} [g_T(A_T(Z))(A_T(Z) - c_k) \mathbb{1}_k(A_T(Z))] \\
 &= C_{1k}^2 \left\{ \int_{b_k}^{b_{k+1}} \left(g_T(c_k) + g'_T(c_k)(u - c_k) + \frac{1}{2} g''_T(\xi)(u - c_k)^2 \right) \right. \\
 &\quad \left. (u - c_k) f(u) du \right\} \\
 &= g'_T(c_k) + \frac{C_{1k}^2}{2} \int_{b_k}^{b_{k+1}} g''_T(\xi)(u - c_k)^3 f(u) du.
 \end{aligned}$$

The following bounds will be helpful in the remainder of the proof

$$\frac{1}{KM} \leq (b_{k+1} - b_k) \leq \frac{1}{Km}. \quad (2.18)$$

They follow from the fact that by definition $1/K = \int_{b_k}^{b_{k+1}} f(u) du$ and the trivial inequalities $m(b_{k+1} - b_k) \leq \int_{b_k}^{b_{k+1}} f(u) du \leq M(b_{k+1} - b_k)$. Moreover

$$\begin{aligned}
 \max_{1 \leq k \leq K} C_{1k}^2 &= \frac{1}{\min_{1 \leq k \leq K} \mathbb{E}_{\bar{\mathbb{P}}} [\mathbb{1}_k(A_T(Z))(A_T(Z) - c_k)^2]} \\
 &\leq \frac{12}{m(b_{k+1} - b_k)^3} \\
 &\leq \frac{12(KM)^3}{m}, \quad (2.19)
 \end{aligned}$$

where the second inequality follows from (2.18) and the first inequality from the fact

$$\begin{aligned}
 \mathbb{E}_{\bar{\mathbb{P}}} [\mathbb{1}_k(A_T(Z))(A_T(Z) - c_k)^2] &= \int_{b_k}^{b_{k+1}} (u - c_k)^2 f(u) du \\
 &\geq m \int_{b_k}^{b_{k+1}} (u - c_k)^2 du \\
 &= \frac{m}{3} [(b_{k+1} - c_k)^3 - (b_k - c_k)^3] \\
 &\geq \frac{m}{12} (b_{k+1} - b_k)^3,
 \end{aligned}$$

because $(b_{k+1} - c_k)^3 - (b_k - c_k)^3$ as a function of c_k is minimized at $c_k = (b_{k+1} + b_k)/2$.

For the fourth moment of the approximation error we now obtain with $B := \sup_{u \in D} |g_T''(u)|$

$$\begin{aligned}
 & \mathbb{E}_{\tilde{\mathbb{P}}} \left[(g_T(A_T(Z)) - g_T^K(A_T(Z)))^4 \right] \\
 &= \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(g_T(A_T(Z)) - \sum_{k=1}^K (\alpha_{0k} e_{0k}(A_T(Z)) + \alpha_{1k} e_{1k}(A_T(Z))) \right)^4 \right] \\
 &= \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(\sum_{k=1}^K \left(g_T(A_T(Z)) - \alpha_{0k} \sqrt{K} - \alpha_{1k} C_{1k}(A_T(Z) - c_k) \right) \right. \right. \\
 &\quad \left. \left. \mathbb{1}_k(A_T(Z)) \right)^4 \right] \\
 &= \mathbb{E}_{\tilde{\mathbb{P}}} \left[\sum_{k=1}^K \left(g_T(A_T(Z)) - \alpha_{0k} \sqrt{K} - \alpha_{1k} C_{1k}(A_T(Z) - c_k) \right)^4 \right. \\
 &\quad \left. \mathbb{1}_k(A_T(Z)) \right] \\
 &= \sum_{k=1}^K \int_{b_k}^{b_{k+1}} \left(g_T(u) - g_T(c_k) - g_T'(c_k)(u - c_k) \right. \\
 &\quad \left. - \frac{K}{2} \int_{b_k}^{b_{k+1}} g_T''(\xi)(v - c_k)^2 f(v) dv \right. \\
 &\quad \left. - \frac{C_{1k}^2}{2} \left(\int_{b_k}^{b_{k+1}} g_T''(\xi)(v - c_k)^3 f(v) dv \right) (u - c_k) \right)^4 f(u) du \\
 &\leq 27 \sum_{k=1}^K \left(\int_{b_k}^{b_{k+1}} \left(\frac{1}{2} g_T''(\xi)(u - c_k)^2 \right)^4 f(u) du \right. \\
 &\quad \left. + \int_{b_k}^{b_{k+1}} \left(\frac{K}{2} \int_{b_k}^{b_{k+1}} g_T''(\xi)(v - c_k)^2 f(v) dv \right)^4 f(u) du \right. \\
 &\quad \left. + \int_{b_k}^{b_{k+1}} \left(\frac{C_{1k}^2}{2} \left(\int_{b_k}^{b_{k+1}} g_T''(\xi)(v - c_k)^3 f(v) dv \right) (u - c_k) \right)^4 f(u) du \right)
 \end{aligned}$$

$$\begin{aligned}
 &= 27 \sum_{k=1}^K \left(\frac{1}{16} \int_{b_k}^{b_{k+1}} g_T''(\xi)^4 (u - c_k)^8 f(u) du \right. \\
 &\quad + \frac{K^4}{16} \int_{b_k}^{b_{k+1}} \left(\int_{b_k}^{b_{k+1}} g_T''(\xi) (v - c_k)^2 f(v) dv \right)^4 f(u) du \\
 &\quad \left. + \frac{C_{1k}^8}{16} \int_{b_k}^{b_{k+1}} \left(\int_{b_k}^{b_{k+1}} g_T''(\xi) (v - c_k)^3 f(v) dv \right)^4 (u - c_k)^4 f(u) du \right) \\
 &\leq \frac{27}{16} K \left(\frac{1}{K} B^4 \max_{1 \leq k \leq K} (b_{k+1} - b_k)^8 + K^4 \frac{1}{K^5} B^4 \left(\max_{1 \leq k \leq K} (b_{k+1} - b_k)^2 \right)^4 \right. \\
 &\quad \left. + \frac{1}{K^5} \max_{1 \leq k \leq K} (C_{1k}^8) B^4 \left(\max_{1 \leq k \leq K} (b_{k+1} - b_k)^3 \right)^4 \max_{1 \leq k \leq K} (b_{k+1} - b_k)^4 \right) \\
 &\leq \frac{54}{16} \cdot B^4 \cdot \frac{1}{m^8 K^8} + \frac{27}{16} \frac{1}{K^4} \frac{12^4 M^{12} K^{12}}{m^4} \cdot B^4 \cdot \frac{1}{K^{12} m^{12}} \frac{1}{K^4 m^4} \\
 &= O\left(\frac{1}{K^8}\right),
 \end{aligned}$$

where we used that $g_T(c_k) + g_T'(c_k)(u - c_k)$ corresponds to the first order Taylor expansion of $g_T(u)$ around c_k . The first inequality follows from Loève's c_r -inequality (see Hansen, 2014) and the third makes use of (2.18) and (2.19). Lemma 2.5 follows immediately. \square

Proof of Lemma 2.6. Let m and M as in the proof of Lemma 2.5. Let b_k and b_{k+1} be in $[a_1, a_2]$ with $b_k < b_{k+1}$ and let $c_k \in [b_k, b_{k+1}]$. Then

$$\int_{b_k}^{b_{k+1}} (u - c_k)^4 f(u) du \leq M \int_{b_k}^{b_{k+1}} (u - c_k)^4 du \leq M (b_{k+1} - b_k)^5.$$

Moreover, from the proof of Lemma 2.5 we know

$$\int_{b_k}^{b_{k+1}} (u - c_k)^2 f(u) du \geq \frac{m}{12} (b_{k+1} - b_k)^3.$$

Therefore

$$\begin{aligned} \frac{\int_{b_k}^{b_{k+1}} (u - c_k)^4 f(u) \, du}{\left(\int_{b_k}^{b_{k+1}} (u - c_k)^2 f(u) \, du \right)^2} &\leq \frac{M(b_{k+1} - b_k)^5}{\left(\frac{m}{12}\right)^2 (b_{k+1} - b_k)^6} \\ &= C \frac{1}{(b_{k+1} - b_k)}, \end{aligned}$$

where $C := M/(m/12)^2$. Using the left hand inequality in (2.18) we get

$$\frac{\int_{b_k}^{b_{k+1}} (u - c_k)^4 f(u) \, du}{\left(\int_{b_k}^{b_{k+1}} (u - c_k)^2 f(u) \, du \right)^2} \leq C \cdot M \cdot K.$$

□

Chapter 3

Theory and Validation of Replicating Portfolios in Insurance Risk Management

In the previous chapter we have provided the convergence rate of the Regress-Later estimator. In this chapter we draw the link between Regress-Later and portfolio replication in the context of obtaining approximations to the value function of insurance contingent claims.

The Solvency II framework challenges insurers to evaluate and manage their embedded balance sheet risks appropriately. However, insurances hold balance sheet items, for which closed-form solutions and market prices are not available. As pointed out in Chapter 1 pure Monte Carlo valuation requires nested simulations, which are too time-intensive. Therefore, methods that project these balance sheet items into functional representations, which simplify and enhance risk analysis, have been suggested. Among these, replicating portfolios are widely applied in practice, though their validity and properties have not been fully examined yet. In this chapter we correct this shortcoming and propose a mathematical framework within which the asymptotic properties of replicating portfolios are analyzed. In fact, we will see that the replicating portfolio estimator corresponds to the Regress-Later estimator in Least Squares Monte Carlo, which allows to transfer the results of Chapter 2 to the replicating portfolio problem. It is shown that the replicating portfolio problem is mathematically well-defined and asymptotically converges to the true solution. Hence, this chapter provides a general mathematical validation for replicating portfolios applied in insurance. A typical path-dependent insurance policy is discussed within the framework and the practical challenges therein are revealed¹.

¹This chapter is based on Beutner et al. (2015)

3.1 Introduction

The aftermath of the global financial crisis triggered the need for higher awareness and acknowledgment of sound risk management in the financial services sector. Stricter requirements from regulators in view of the introduction of Basel III and Solvency II are pushing banks, insurances and pension funds to increase their understanding in terms of the risks they are taking and to implement appropriate risk management mechanisms. Financial institutions rely on financial and economic models for conducting their businesses. In a risk-management context the models for calculating risk should be accurate and fast in terms of the calculation of risk figures such that the rapid pace of market environments is matched. In insurance risk management insurers are faced with the challenge to quickly revalue their liabilities under economic stress scenarios based on fair market valuation (see Article 76, The European Parliament and The Council, 2009). Typically insurance liabilities exhibit features such as options and guarantees comparable to standard financial products, but are generally not closed-form valuable. Numerical methods, such as Monte Carlo techniques, to estimate the value of the liabilities therefore become inevitable. For solvency capital calculations the insurer requires the distribution of the liabilities at a future time point, typically one year. Procedures known as “nested simulation” or “full stochastic Monte Carlo simulation” take a full simulation approach, which is illustrated in Figure 3.1 (see also Bauer et al., 2009, for a description). In the nested simulation approach first all risk factors are simulated on the outer scenarios (shock scenarios) up to the time point, at which the value of the portfolio shall be calculated, which is t in Figure 3.1. For insurance risk reporting purposes t is typically one year. We remark that in practice most insurers approximate the one-year horizon by instantaneous shocks.

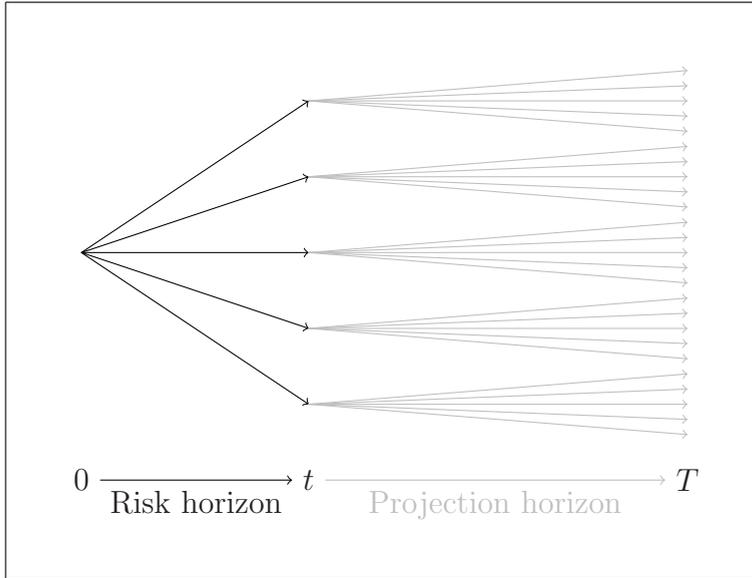


Figure 3.1: Nested stochastic simulation problem.

Inner scenarios are constructed at each node at point t to estimate the value of the portfolio conditional on the information at time t . The average of the discounted simulated values from one node at time t to T gives an estimate of the market-consistent value of the liabilities at time t given the information at time t . Ultimately the empirical distribution of the fair values at time t is obtained, which is required to derive risk figures².

Due to the complexity and size of an insurer's liabilities running asset-liability projection models to re-value the liabilities based on simulations under different market conditions and other risk factor realizations is a timely exercise and is, given current computing power, infeasible. Therefore, most insurers turn to approximation methods

²Note that for the purpose described the shock horizon is typically simulated under the true probability measure, while the projection horizon is simulated under the risk-neutral measure.

in combination with Monte Carlo techniques for estimating the value of their liabilities. Essentially, the objective is to find approximations of the time t conditional expectation, which gives the liability value conditional on information at time t . The exercise calls for apt methods that suitably fit the liabilities to an approximating function. “Portfolio replication” and “Least-Squares-Monte-Carlo” (LSMC) have become buzz words for the Solvency II internal model methodologies and are generally known as liability proxy modeling techniques (see e.g. Bauer et al., 2009; Oechslin et al., 2007). These models use static representations of the liabilities as proxies in risk-capital calculations. In this paper we address the method of replicating portfolios and show that it mathematically corresponds to a specific LSMC type, termed “regression later” by Glasserman and Yu (2004b), which has been developed in the context of American option pricing. We now briefly review the related literature on the valuation of life insurance policies as well as on static replication and LSMC.

In the literature on the valuation of life insurance policies many authors have drawn parallels between life insurance policies and (exotic) financial options. For several insurance policies it has been shown that the payoff of an insurance policy can be decomposed into financial instruments, which then allows to derive solutions for the fair value from option pricing theory. Pioneers of this work are Brennan and Schwartz (1976) and Boyle and Schwartz (1977) for the valuation of unit-linked contracts with guarantees. Pelsser and Schrager (2004) discuss regular premium unit-linked policies and show the similarities to Asian options. Grosen and Jørgensen (2000) establish a model for profit-sharing (also referred to as participating) life insurance policies and conclude that these consist of a zero coupon bond, a bonus option and a surrender option. Further contributions dealing with the valuation of life insurance policies are, among others, Bacinello (1993), Ekern and

Persson (1996), Bauer et al. (2006), Bacinello (2001) and Tanskanen and Lukkarinen (2003). While all these authors contribute to the fair valuation concept of life insurance policies, in this paper we put focus on achieving static representations of the liability payoff through a portfolio of replicating instruments, which has not fully been exploited by the previous authors. The valuation of life insurance policies is then shifted to the valuation of the instruments making up the replicating portfolio.

The principle of static replication is to construct a portfolio of securities that mirrors the terminal payoff of a target security in every possible state of the world. In contrast to dynamic replication the portfolio weights of the static replicating portfolio do not change with changes in market conditions. By the no-arbitrage condition, if the payoff of the target security is perfectly replicated, the replication automatically matches the security's value at all times before maturity, implying that they have the same market-consistent price. Breeden and Litzenberger (1978) show that path-independent securities can be statically replicated through a portfolio of vanilla calls and puts. Static replication has later on largely gained relevance in finding static hedging opportunities for exotic options to overcome the risks and costs associated with dynamic hedging strategies. Derman et al. (1995), Carr and Bowie (1994), Carr and Chou (1997) and Carr et al. (1998), among others, are good examples. Madan and Milne (1994) significantly contribute to the static replication literature by formulating a general mathematical framework, within which general contingent claims can be statically replicated. The contingent claims are only required to have finite variance. Contingent claims are then modeled in a Hilbert space and the static replication problem is solved by constructing a countable orthonormal basis. Static replications have also gained attention in life insurance, particularly in the context of hedging

and risk management, where insurance liabilities must be repeatedly valued under diverse stress scenarios, but closed-form solutions are not available. Pelsser (2003) leverages the static portfolio replication concept to derive hedging strategies with swaptions for life insurance policies with guaranteed annuity options. If the value of the replicating instruments is readily available, the value of the replicating portfolio under different economic scenarios is quickly determined. For risk management purposes taking the replicating portfolio as a proxy to the true liabilities then speeds up risk calculations tremendously. This feature has naturally been exploited in the context of Solvency II. Oechslin et al. (2007) provide first guidance in how to set up replicating portfolios for life insurance liabilities. Recently, Natolski and Werner (2014) discuss and compare several approaches to the construction of replicating portfolios in life insurance. Chen and Skoglund (2012), Daul and Vidal (2009), Kalberer (2012), Koursaris (2011b), Koursaris (2011c) and Burmeister (2007), for example, address the construction of replicating portfolios in life insurance from a more practical point of view and give recommendations.

The concept of statically representing contingent claims has also been exploited in the pricing of Bermudan and American options, for which no closed-form solutions exist. The main difficulty in Bermudan and American option pricing is the calculation of the continuation value, i.e. a conditional expectation across timesteps. A popular idea is to estimate the continuation value going backward through cross-sectional regression on Monte Carlo simulated paths. As the approach combines Monte Carlo simulation of the paths of the underlying and regression of the value function against basis functions of the underlying it is commonly termed Least Squares Monte Carlo (LSMC). Examples may be found in Carriere (1996), Longstaff and Schwartz (2001), Tsitsiklis and Van Roy (2001), Clement et al. (2002), Stentoft (2004), Glasserman

and Yu (2004b), Egloff et al. (2007), Belomestny (2011), Gerhold (2011) and Zanger (2013). Andreatta and Corradin (2003), Bacinello et al. (2009) and Bacinello et al. (2010) apply the LSMC approach to the valuation of life insurance policies with surrender options. While the static representations discussed by all of these authors are similar to the replicating portfolio technique, they differ in one important aspect: They immediately estimate the valuation function rather than the payoff function of the contingent claim. This approach is termed “regression now” by Glasserman and Yu (2004b) and is contrasted to “regression later” or Regress-Later LSMC estimation (see also Beutner et al., 2013), which pertains to a static representation of the payoff function and is the subject of this paper.

In this paper we focus on statically replicating the payoff of a liability through basis functions. We provide a mathematical validation for the construction of static replicating portfolios to general contingent claims, which may depend on financial as well as non-financial risk factors. We only require the contingent claims to have finite second moments. This allows to model the contingent claims in a separable Hilbert space as for example also discussed in Madan and Milne (1994). Accordingly, a perfect replicating portfolio exists by the countable basis representation in the Hilbert space. It will become clear that the replicating portfolio approach discussed combines Monte Carlo simulation with least squares regression as in LSMC. In fact, we show that the replicating portfolio estimator corresponds to the LSMC Regress-Later estimator in Beutner et al. (2013), which has first been proposed by Glasserman and Yu (2004b). We may therefore use the terminologies portfolio replication and LSMC Regress-Later interchangeably. As will be shown, LSMC with Regress-Later produces a non-standard least squares estimator, which asymptotically converges to the true result. The mathematical framework is largely based on Beutner et al. (2013),

from which also the asymptotic convergence results are taken. We introduce multi-dimensional piecewise linear functions as a simple, but very effective basis, for which we derive the explicit convergence rate. Typical replicating instruments in practice are standard financial instruments. While piecewise linear functions may therefore at first seem uncommon as replicating instruments, we provide a simple relation between piecewise linear functions and call options. Accordingly a replicating portfolio of piecewise linear functions is approximately equal to a portfolio of calls. Finally we consider a complex path-dependent insurance policy, apply the replicating portfolio theory to it and discuss the challenges along the way.

The structure of this paper is as follows. Section 3.2 introduces the mathematical framework for replicating portfolios of general finite variance contingent claims. Section 3.3 discusses the asymptotic convergence result for replicating portfolios, which is taken from the convergence theory of Regress-Later estimators in Beutner et al. (2013). In Section 3.4 the basis of piecewise linear functions is derived and the asymptotic convergence rate is explicitly calculated. These sections present the general foundation for replicating portfolios. In Section 3.5 the replicating portfolio framework is applied to a participating life insurance policy as presented in Grosen and Jørgensen (2000). Section 3.6 concludes.

3.2 Mathematical model for Replicating Portfolios

In this section we give the mathematical model and the estimation approach for the construction of replicating portfolios. Both the model

and the notation largely follow Beutner et al. (2013), which we repeat here.

We assume a frictionless arbitrage-free financial market with finite time horizon T . Let $Z = \{Z(t), 0 \leq t \leq T\}$ be a d -dimensional stochastic process with $d \in \mathbb{N}$ defined on some filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0 \leq t \leq T}, \tilde{\mathbb{P}})$, where $\{\mathcal{F}_t\}_{0 \leq t \leq T}$ denotes the filtration generated by Z . The measure $\tilde{\mathbb{P}}$ denotes some probability measure equivalent to the true probability measure \mathbb{P} . Recall from Section 3.1 that in the nested stochastic simulation framework the projection horizon is typically simulated under the pricing measure \mathbb{Q} as the market-consistent price at time t is of interest. We deliberately leave $\tilde{\mathbb{P}}$ unspecified in our model, which allows for freedom in choosing a measure for the construction of the replicating portfolio. Our model formulation is sufficiently general to take $\tilde{\mathbb{P}} = \mathbb{Q}$. We specifically point that case out, where appropriate and later on provide a more detailed interpretation for choosing $\tilde{\mathbb{P}} = \mathbb{Q}$, where \mathbb{Q} again denotes the pricing measure. We interpret Z to be the ultimate d -dimensional random driver, on which the cash flows of an insurance contingent claim depend. Insurance liabilities typically generate (finitely) many stochastic payoffs at multiple time points on a finite time horizon that depend on finitely many underlying risk drivers, both of financial as well as non-financial nature. We remark that in principle our model allows to account for both financial and non-financial risk factors, i.e. Z may account for any risk. We do not further specify Z , but refer the reader to Bauer et al. (2010), for example, for a full generic probability model for life insurance contracts.

We denote the terminal payoff at time T of an insurance contingent claim by X . With terminal cash flow we refer to the definition in Oechslein et al. (2007) and Madan and Milne (1994). Accordingly the terminal cash flow is the sum of all cash flows over time accumulated

at the money market account. The paths $Z(\cdot, \omega)$ with $\omega \in \Omega$, of Z given by $t \rightarrow Z(t, \omega)$, $t \in [0, T]$, are assumed to lie in some function space $\mathbb{D}_d[0, T]$ consisting of functions mapping from $[0, T]$ to \mathbb{R}^d , and we consider Z as a random function. We sometimes abbreviate $Z(\cdot, \omega)$ by $Z(\omega)$. Recall that the payoff function X is driven by Z . We assume that the payoff X is \mathcal{F}_T -measurable and we want to write X in terms of Z . However, as insurance contingent claims are typically path-dependent and generate multiple cash flows over time the payoff X at time T depends on the paths of $Z(\cdot, \omega)$. Thus, we define a process, denoted by $A_T(Z)$, which carries all the information on the paths of the d -dimensional stochastic process Z from time 0 to T that is relevant for the contingent claim X . Now we can write for every ω in the sample space Ω the payoff $X(\omega)$ of the contingent claim X as $g_T(A_T(Z(\cdot, \omega)))$, where A_T is a known (measurable) functional mapping from the function space $\mathbb{D}_d[0, T]$ to \mathbb{R}^ℓ and g_T is a known Borel-measurable function that maps from \mathbb{R}^ℓ to \mathbb{R} . This basically means that for a payoff function X that depends on finitely many characteristics of the stochastic path of the underlying process it suffices to observe these characteristics, which are stored in A_T . The functional mapping A_T thus comprises all those characteristics of the stochastic path of the d -dimensional underlying that determine the contingent claim. We denote the dimensionality of A_T by ℓ , which is driven by the dependence structure on the d -dimensional process Z and the number of characteristics on the stochastic path that are required to determine X . Note that if we were only interested in plain vanilla contingent claims at time T it would suffice to observe the stochastic process Z at time T . For Z one-dimensional we would then have $\ell = 1$. We exemplify the meaning of A_T with two examples, which also demonstrate that A_T is not unique, but depends on the specification of the modeler.

Example 10. (*Asian Option*) Consider a discrete Asian option on a stock. The payoff X of the contingent claim depends on all past stock values prior to maturity. $A_T(Z)$ must now comprise all the information of the underlying driver such that X is specified. We may choose $A_T(Z)$ as the time average over the past stock values, which suffices to calculate the payoff X . To better see this let Z be one-dimensional and take $X = (\sum_{s=1}^T Z_1(s) - K)^+$, where K is the strike price. Then X does only depend on $\sum_{s=1}^T Z_1(s)$. Thus, $A_T(f) = \sum_{s=1}^T f(s)$ for every function $f \in \mathbb{D}[0, T]$ and therefore $\ell = 1$. Alternatively, we may also choose to take $A_T(f) = (f(1), \dots, f(T))$ and then $\ell = T$.

Example 11. (*Profit-sharing contract*) Consider a profit-sharing contract, in which on a yearly basis interest is credited to the policyholder's account. A minimum crediting rate is guaranteed and additional profit is shared depending on the specification of the bonus credited. Then the contingent claim X at maturity T is given by $X = L_0 \prod_{t=1}^T (1 + r_G + r_B(t))$, where L_0 is the initial value of the policy, r_G denotes the minimum guarantee rate and $r_B(t)$ is the bonus credited at time t . We leave $r_B(t)$ unspecified and only note that it may, for example, be linked to some external reference index or to the performance of the underlying asset portfolio. Straightforwardly, we can define $A_T(f) = \prod_{t=1}^T (1 + r_G + f(t))$ for every function $f \in \mathbb{D}[0, T]$ and therefore $\ell = 1$. As a result the dimensionality of the problem is only one, but the specification of $A_T(Z)$ is complex. The basis given a complicated $A_T(Z)$ may then be difficult to price. In that respect, we can also specify $A_T(f) = (f(1), \dots, f(T))$ and therefore $\ell = T$.

We now introduce the mathematical framework for replicating contingent claims. As in Beutner et al. (2013) we restrict attention to finite second moment contingent claims and refer to the relevant related literature, in which the same assumption is applied (see e.g.

Stentoft, 2004; Bergstrom, 1985; Madan and Milne, 1994; Longstaff and Schwartz, 2001). Thus we assume that the contingent claim X has finite mean and variance, which allows to model it as an element of a Hilbert space (see also Madan and Milne, 1994). More specifically, we assume that g_T belongs to the functional space $L_2(\mathbb{R}^\ell, \mathcal{B}(\mathbb{R}^\ell), \tilde{\mathbb{P}}^{A_T(Z)})$, where $\mathcal{B}(\mathbb{R}^\ell)$ denotes the Borel σ -algebra on \mathbb{R}^ℓ , and $\tilde{\mathbb{P}}^{A_T(Z)}$ denotes the probability measure on \mathbb{R}^ℓ induced by the mapping $A_T(Z)$. Now, $L_2(\mathbb{R}^\ell, \mathcal{B}(\mathbb{R}^\ell), \tilde{\mathbb{P}}^{A_T(Z)})$ is a separable Hilbert space with inner product

$$\int_{\mathbb{R}^\ell} h_1(u)h_2(u) d\tilde{\mathbb{P}}^{A_T(Z)}(u) = \mathbb{E}_{\tilde{\mathbb{P}}} [h_1(A_T(Z))h_2(A_T(Z))]$$

and norm

$$\sqrt{\int_{\mathbb{R}^\ell} h_1(u)h_1(u) d\tilde{\mathbb{P}}^{A_T(Z)}(u)} = \sqrt{\mathbb{E}_{\tilde{\mathbb{P}}} [h_1^2(A_T(Z))]}$$

(Bogachev, 2007). Recall that a Hilbert space simply abstracts the finite-dimensional geometric Euclidean space to infinite dimensions (Conway, 1985). It is a well-known result that a separable Hilbert space has a countable orthonormal basis, in terms of which its elements may be expressed; see for instance Bogachev (2007, Corollary 4.2.2 and Corollary 4.3.4). Then, X can be written as

$$X = g_T(A_T(Z)) = \sum_{k=1}^{\infty} \alpha_k e_k(A_T(Z)), \quad (3.1)$$

where $\{e_k\}_{k=1}^{\infty}$ is a countable orthonormal basis of $L_2(\mathbb{R}^\ell, \mathcal{B}(\mathbb{R}^\ell), \tilde{\mathbb{P}}^{A_T(Z)})$ with coefficients α_k given by

$$\alpha_k = \mathbb{E}_{\tilde{\mathbb{P}}} [X e_k(A_T(Z))]. \quad (3.2)$$

The theoretically perfect replicating portfolio of the payoff X is thus given by the linear combination of the basis functions spanning the space as specified in Equation (3.1). We stress that the representation in Equation (3.1) holds for any contingent claim with finite mean and variance where we can write X as $g_T(A_T(Z))$, which belongs to the previously specified L_2 Hilbert space.

We try to estimate the payoff X through its representation in Equation (3.1) by simulating data for Z under $\tilde{\mathbb{P}}$. However, the representation in Equation (3.1) involves the estimation of infinitely many parameters, which leaves a direct estimation infeasible. Consequently, finite-dimensional approximations with a truncated basis $\{e_k\}_{k=1}^K$, $K < \infty$, are used instead. For Equation (3.1) this implies that we can approximate g_T by

$$g_T^K = \sum_{k=1}^K \alpha_k e_k = (\boldsymbol{\alpha}_K)^T \mathbf{e}_K,$$

where $\boldsymbol{\alpha}_K = (\alpha_1, \dots, \alpha_K)^T$ and $\mathbf{e}_K = (e_1, \dots, e_K)^T$. Defining the approximation error a_T^K straightforwardly by $a_T^K := g_T - g_T^K$ we obtain the representation

$$X = g_T^K(A_T(Z)) + a_T^K(A_T(Z)). \quad (3.3)$$

This gives the regression equation for our estimation problem where a_T^K represents the regression error. Note that the approximation error and its variance,

$$\mathbb{E}_{\tilde{\mathbb{P}}} \left[(a_T^K(A_T(Z)))^2 \right] = \sum_{k=K+1}^{\infty} \alpha_k^2 \rightarrow 0 \quad (3.4)$$

converge to zero as $K \rightarrow \infty$. Notice also that

$$\mathbb{E}_{\mathbb{P}} [g_T^K(A_T(Z)) a_T^K(A_T(Z))] = 0$$

by the orthogonality of the basis. Now, given a (simulated) sample of size N denoted by $(x_1, A_T(z_1)), \dots, (x_N, A_T(z_N))$ it is natural to estimate g_T^K through least squares regression leading to the estimator

$$\hat{g}_T^K = (\hat{\boldsymbol{\alpha}}_K)^T \mathbf{e}_K \quad (3.5)$$

with

$$\hat{\boldsymbol{\alpha}}_K = \left((\mathbf{E}_K)^T \mathbf{E}_K^{lat} \right)^{-1} (\mathbf{E}_K^{lat})^T \mathbf{X}, \quad (3.6)$$

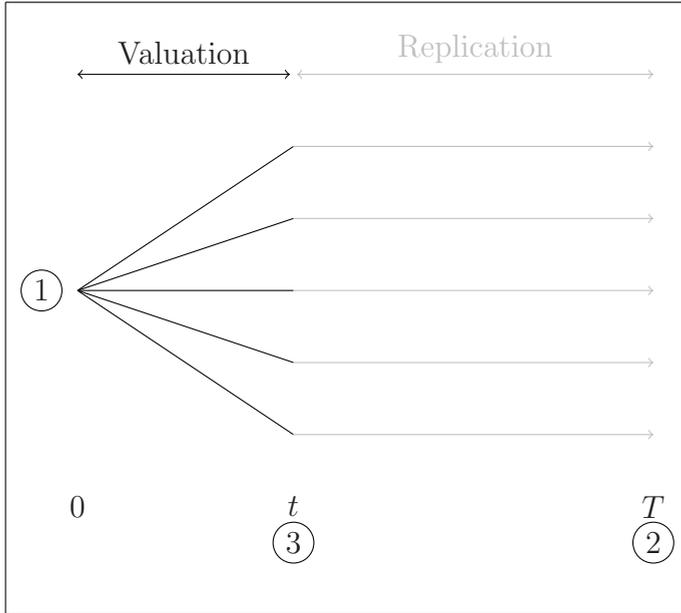
where $\mathbf{X} = (x_1, \dots, x_N)^T$ and \mathbf{E}_K is an $N \times K$ matrix with the n^{th} row equal to $\mathbf{e}_K(A_T(z_n))$, $n = 1, \dots, N$.

As previously explained the truncation of the linear basis representation for g_T leads to an approximation error. The approximation error decreases when K grows. However, as K grows more parameters need to be estimated and it is intuitive that a larger sample is required. Consequently, the truncation parameter K should be an increasing function of the sample size and we write $K = \mathcal{K}(N)$. Then, the approximation error vanishes in the limit as $N \rightarrow \infty$. This relation produces a nonstandard regression problem as the regression error in Equation (3.3), which is given by the approximation error, and its variance decrease to zero for $N \rightarrow \infty$.

Until now we have focused on the construction of replicating portfolios for contingent payoffs in insurance without addressing the valuation problem. The described estimation procedure so far constructs a portfolio consisting of basis functions $\{e_k(A_T(Z))\}_{k=1}^K$ that replicates the payoff X . From an insurance risk management perspective we are

ultimately interested in the (market-consistent) value at time t . Hence, the ultimate objective is to estimate unknown conditional expectations of the form $\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t]$. If we take $\tilde{\mathbb{P}}$ equal to the pricing measure and discount X to time t the conditional expectation corresponds to the time t price of X . We proceed as follows to approximate the quantity of interest, i.e. $\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t], 0 \leq t < T$. We have seen that the contingent payoff X can be estimated by a finite linear combination of basis functions, mathematically speaking random variables. We choose a basis for which the calculation of the time t value is either exact meaning that the analytical solution is known or it can be quickly and fairly accurately estimated through numerical integration. Then, given the linear representation of X through basis functions, the operator $\mathbb{E}_{\tilde{\mathbb{P}}}[\cdot|\mathcal{F}_t]$ is applied to these basis functions. The approach takes advantage of the linearity of the expectation operator. For the case where $\tilde{\mathbb{P}} = \mathbb{Q}$ the basis terms must thus be priced. The clear advantage of the approach is that given a basis representation with a fairly easily priceable basis, the value of the target variable is quickly re-priced under different time t -realizations of the risk factors of the value function $g_{0,t}$, where t is typically equal to 1 in insurance. This is exactly the idea of static replicating portfolios. Of course, that requires the availability of such basis functions. Also note that the replication of the contingent payoff and its valuation is completely separated. In the first step the contingent payoff is replicated by a linear combination of basis terms. In the second step the contingent payoff's value is estimated by valuing the basis terms. In Section 3.3 we show that the measure for replication may even be different from the measure for valuation as long as the Radon-Nikodym derivative is bounded.

Recall the initial “nested stochastic simulation” problem in Figure 3.1. We can now draw a different picture. In Figure 3.2 we illustrate the replicating portfolio approach that has been outlined in this section.



1. Monte Carlo simulation of Z and X on the paths $[0, T]$. Determine $A_T(Z)$.
2. Regression of X against basis functions of $A_T(Z)$.
3. Estimate $\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t]$ by $\mathbb{E}_{\tilde{\mathbb{P}}}[\hat{g}_T(A_T(Z))|\mathcal{F}_t]$.

Figure 3.2: The Replicating Portfolio approach.

First the payoffs at time T are simulated. Then, the payoff function is replicated through regression of the payoff values against the basis functions valued at time T . Next, the conditional expectation function at time $t < T$ is estimated through applying the conditional expectation operator to the basis functions making up the previously obtained replicating portfolio. The replicating portfolio technique combines Monte Carlo simulation of the liabilities over the full projection horizon with regression against explanatory variables, which are also simulated on the full projection horizon. In that sense we can view the replicating portfolio approach as a Least Squares Monte Carlo technique. The resulting replicating portfolio serves as the estimator for the payoff X of the target function. The conditional expectation of X is then estimated by applying the conditional expectation operator to the basis functions contained in the replicating portfolio. In a risk-neutral setting this corresponds to pricing the replicating instruments. The replicating portfolio approach outlined exactly corresponds to a two-step LSMC estimator, which has first been suggested by Glasserman and Yu (2004b) in the context of American option pricing and has been labeled “regression later”. Beutner et al. (2013) discuss this estimator in a single-period setting and use the terminology “Regress-Later”. While technically the LSMC Regress-Later and the replicating portfolio approach are identical, replicating portfolios are mostly understood as portfolios of financial instruments, meaning that the basis is given by a series of financial instruments. We will come back to the topic of choosing a basis in Section 3.4. In the remainder of this paper we focus on the first step, i.e. the replication problem, and only address the valuation where necessary. The key take aways from this section are:

- The replicating portfolio problem is theoretically well defined.

- The replicating portfolio approach corresponds to LSMC with Regress-Later.
- The construction of the replicating portfolio of the time T terminal payoff of the target function is separated from the problem of valuing its time $t < T$ conditional expectation function.

3.3 General asymptotic convergence of Replicating Portfolios

In the previous section we established the mathematical framework within which replicating portfolios are well defined. We now want to validate the replicating portfolio as a method for approximating contingent claims by analyzing its asymptotic properties. The result shown in this section holds for a general multi-dimensional contingent claim as specified in Section 3.2. As previously argued the replicating portfolio technique essentially corresponds to the LSMC Regress-Later approach suggested by Glasserman and Yu (2004b). Beutner et al. (2013) derive the convergence rate of the Regress-Later estimator. We repeat here the theorem and its necessary assumptions, but refer the reader to Beutner et al. (2013) for the corresponding proofs and further technical details.

From Section 3.2 we know that the estimation approach of X consists of an approximation and an estimation step. The approximation is necessary as the infinite-dimensional representation of X in Equation (3.1) cannot be directly estimated. We hence also expect that the overall convergence rate is driven by the convergence of the approximation and the estimation error. In deriving the convergence rate the following condition on the convergence of the approximation error is imposed.

Assumption 3.1. There are $\gamma > 0$, $\boldsymbol{\alpha}_K$ s.t.

$$\begin{aligned}
 & \sqrt{\mathbb{E}_{\tilde{\mathbb{P}}} \left[(g_T(A_T(Z)) - (\boldsymbol{\alpha}_K)^T \mathbf{e}_K(A_T(Z)))^4 \right]} \\
 &= \sqrt{\int_{\mathbb{R}^\ell} (g_T(u) - (\boldsymbol{\alpha}_K)^T \mathbf{e}_K(u))^4 d\tilde{\mathbb{P}}^{A_T(Z)}(u)} \\
 &= \sqrt{\int_{\mathbb{R}^\ell} a_T^K(u)^4 d\tilde{\mathbb{P}}^{A_T(Z)}(u)} = O(K^{-\gamma}). \tag{3.7}
 \end{aligned}$$

This condition ensures that the square root of the fourth moment of the approximation error decreases at rate $K^{-\gamma}$. Intuitively the better a basis is suited for approximating the contingent claim X the faster we expect the approximation error to converge. This should be captured by the parameter γ . We remark that Assumption 3.1 depends on the probability measure $\tilde{\mathbb{P}}$. As in a simulation-based framework the data generating process of X is known, Assumption 3.1 is feasible.

Furthermore, we assume that the stochastic paths of the pair $(X, A_T(Z))$ are sampled independently and identically.

Assumption 3.2. $((X_1, A_T(Z_1)), \dots, (X_N, A_T(Z_N)))$ are i.i.d.

We now state the theorem.

Theorem 3.1. (Theorem 3.1 in Beutner et al. (2013)) *Let Assumptions 3.1 and 3.2 be satisfied. Additionally, assume that there is a sequence $\mathcal{K} : \mathbb{N} \rightarrow \mathbb{N}$ and $\tilde{h} : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{R}$ such that*

$$\tilde{h}(N, \mathcal{K}(N)) := \frac{1}{N} \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left((\mathbf{e}_{\mathcal{K}(N)})^T \mathbf{e}_{\mathcal{K}(N)} \right)^2 \right] \rightarrow 0 \text{ as } N \rightarrow \infty. \tag{3.8}$$

Then

$$\mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) \right)^2 \right] = O_{\tilde{\mathbb{P}}}(\mathcal{K}(N)^{-\gamma}). \tag{3.9}$$

Proof. See Beutner et al. (2013) □

Equation (3.8) determines the growth relation of the truncation parameter K and the sample size N . It ensures that the truncation parameter K does not grow too fast in relation to the sample size N . When only N grows the estimation error converges, but the approximation error a_T^K does not vanish. If K grows too fast N may be insufficient to estimate K parameters. To let both the approximation error and the estimation error converge N and K must grow at the same time and in a balanced relation. We illustrate the relation of the approximation error and the estimation error by writing the mean-square error as

$$\begin{aligned}
 & \mathbb{E}_{\hat{\mathbb{P}}} \left[\left(X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) \right)^2 \right] \\
 &= \mathbb{E}_{\hat{\mathbb{P}}} \left[\left(\underbrace{X - g_T^{\mathcal{K}(N)}(A_T(Z))}_{\text{approximation error}} \right)^2 \right] \\
 &+ \mathbb{E}_{\hat{\mathbb{P}}} \left[\left(\underbrace{g_T^{\mathcal{K}(N)}(A_T(Z)) - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z))}_{\text{estimation error}} \right)^2 \right]. \tag{3.10}
 \end{aligned}$$

We clearly see that the total error of the estimator may be split into an approximation and an estimation error. For the errors individually the approximation error converges as K grows while the estimation error converges as N grows. Since K is defined as a function of N both converge by letting $N \rightarrow \infty$.

The discussed general convergence rate pertains to the convergence of the replicating portfolio to the true payoff function X . We have explained in Section 3.2 that ultimately insurers use replicating portfolios to quickly obtain an estimate of the time t value of its liabilities under different scenarios for the underlying risk drivers. As explained

in Section 3.2 we achieve the approximation to the conditional expectation function by applying the conditional expectation operator to the estimated payoff function, $\hat{g}_T^{\mathcal{K}(N)}$. We can show that the ultimate estimator given by the conditional expectation of the estimator of X does not converge slower than at the rate derived for the convergence of $\hat{g}_T^K(A_T(Z))$. More explicitly we have

$$\begin{aligned}
 & \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(\mathbb{E}_{\tilde{\mathbb{P}}} [X | \mathcal{F}_t] - \mathbb{E}_{\tilde{\mathbb{P}}} \left[\hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) | \mathcal{F}_t \right] \right)^2 \right] \\
 &= \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(\mathbb{E}_{\tilde{\mathbb{P}}} \left[X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) | \mathcal{F}_t \right] \right)^2 \right] \\
 &\leq \mathbb{E}_{\tilde{\mathbb{P}}} \left[\mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) \right)^2 | \mathcal{F}_t \right] \right] \\
 &= \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) \right)^2 \right],
 \end{aligned}$$

where the first inequality follows from Jensen's inequality and the last equality uses the projection law of expectations.

At the beginning of Section 3.2 we pointed out that there is some flexibility in the specification of the measure $\tilde{\mathbb{P}}$. Convergence in mean-square under one measure implies convergence in mean square under another measure as long as the Radon-Nikodym derivative of the second measure exists and is bounded. We can exploit this observation to show that we may even use a measure $\tilde{\mathbb{P}}$ different from the valuation measure \mathbb{Q} for constructing the replicating portfolio, but calculate the conditional expectation of the replicating portfolio under the valuation measure \mathbb{Q} as long as \mathbb{Q} has a bounded density with respect to $\tilde{\mathbb{P}}$. Thus, let us assume that for the Radon-Nikodym derivative $q(A_T(Z)) := d\mathbb{Q}/d\tilde{\mathbb{P}}$ we have $|q(A_T(Z))| \leq M$, $M < \infty$. Let $\hat{g}_T^{\mathcal{K}(N)}$ be the replicating

portfolio under $\tilde{\mathbb{P}}$. Then,

$$\begin{aligned} & \mathbb{E}_{\mathbb{Q}} \left[\left(X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) \right)^2 \right] \\ &= \mathbb{E}_{\tilde{\mathbb{P}}} \left[q(A_T(Z)) \left(X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) \right)^2 \right] \\ &\leq M \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) \right)^2 \right]. \end{aligned}$$

Given the assumption on the Radon-Nikodym derivative holds the valuation error under a measure \mathbb{Q} different from the calibration measure $\tilde{\mathbb{P}}$ converges in mean-square if the estimation error of the terminal payoff X converges under $\tilde{\mathbb{P}}$.

$$\begin{aligned} & \mathbb{E}_{\mathbb{Q}} \left[\left(\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t] - \mathbb{E}_{\mathbb{Q}} \left[\hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) | \mathcal{F}_t \right] \right)^2 \right] \\ &\leq \mathbb{E}_{\mathbb{Q}} \left[\left(X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) \right)^2 \right] \\ &\leq M \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) \right)^2 \right]. \end{aligned}$$

The asymptotic analysis conducted in this section shows that the replicating portfolio estimator for the value of X converges to the true value in the limit. Consequently, the replicating portfolio approach is a valid method to the approximation of the time t value of a contingent claim X . We can deduce further valuable information from the results in this section. The convergence rate depends on γ and the necessary growth relation of K and N , which is derived from $\tilde{h}(N, \mathcal{K}(N))$. We can now already give some guidance for the selection of a basis. A good basis (for a particular contingent claim) is characterized by

- A large value for γ .
- A fast allowed growth rate of K in relation to N .

We will discuss a convenient orthonormal basis fulfilling these characteristics in Section 3.4.

3.4 Orthonormal basis of piecewise linear functions

In practical applications finding a basis is a crucial exercise. From the previously described theory it is clear that a Hilbert basis must be constructed that spans the space. Madan and Milne (1994) point out that a Hilbert space basis may be difficult to construct as it requires knowledge on the stochastic processes of the underlying risk drivers. As we have previously explained, in a Monte Carlo framework for generating the payoffs of a contingent claim the underlying stochastic processes are controlled by the user. In this section we introduce a simple, but powerful basis given by a sequence of multi-dimensional piecewise linear functions that is applicable to a large class of contingent claims and offers significant advantages. One strong argument in favour of piecewise linear functions is that both the construction as well as the analysis in view of establishing convergence rates simplifies. Importantly, the suggested basis is by construction orthogonal and can be easily set up. As we will see in the following, under some mild conditions the convergence rate for replicating a large class of target functions with orthonormal piecewise functions can be explicitly calculated.

For simplicity we will consider a compact domain in the following. Although insurance claims may theoretically be unlimited, it is a fair assumption to expect that payoffs are cut off at a certain level as the insurance company cannot pay out more. Thus, we can assume a compact domain D^ℓ on which the payoff function needs to be replicated.

Recall that the dimensionality of the problem is driven by the dimensionality of the chosen $A_T(Z)$, which is denoted by ℓ . Consequently, we focus on replicating the payoff on the compact domain D^ℓ . For values outside of the compact domain cut-off values may be defined and the replication is simply given by the cut-off value.

We construct an orthonormal basis on $L_2(D^\ell, \mathcal{B}(D^\ell), \tilde{\mathbb{P}}^{A_T(Z)})$ based on non-overlapping hypercubes. $\tilde{\mathbb{P}}^{A_T(Z)}$ denotes here the probability measure on $\mathcal{B}(D^\ell)$ induced by the mapping $A_T(Z)$. On each dimension the support is split in K non-overlapping intervals of equal length. This results in K^ℓ orthogonal hypercubes. On each hypercube a multi-dimensional piecewise linear function is constructed. Let us write $A_T(Z) = (Y_1, \dots, Y_\ell)^T := \mathbf{Y}$. We thus define

$$\begin{aligned} e_{0k}(\mathbf{Y}) &:= C_{0k} \mathbf{1}_k(\mathbf{Y}) \\ \mathbf{e}_k(\mathbf{Y}) &:= \mathbf{C}_k \mathbf{1}_k(\mathbf{Y})(\mathbf{Y} - \mathbf{c}_k), \end{aligned}$$

with $\mathbf{Y} = (Y_1, \dots, Y_\ell)^T$, $\mathbf{C}_k = \text{diag}(C_{1k}, \dots, C_{\ell k})$, $\mathbf{c}_k = (c_{1k}, \dots, c_{\ell k})^T$,

$$\begin{aligned} C_{jk} &= \left(\mathbb{E}_{\tilde{\mathbb{P}}} [\mathbf{1}_k(\mathbf{Y}) (Y_j - c_{jk})^2] \right)^{-1/2}, \quad \forall j = 1, \dots, \ell, \\ C_{0k} &= \left(\mathbb{E}_{\tilde{\mathbb{P}}} [\mathbf{1}_k(\mathbf{Y})] \right)^{-1/2} \end{aligned}$$

denote the normalizing constants and

$$c_{jk} = C_{0k}^2 \mathbb{E}_{\tilde{\mathbb{P}}} [\mathbf{1}_k(\mathbf{Y}) Y_j]$$

are chosen such that in each hypercube the linear terms are orthogonal to the constant term. Thus, by construction the constant part e_{0k} is orthogonal to each linear term in $\mathbf{e}_k(\mathbf{Y})$, however the linear terms in the same hypercube are not orthogonal to each other. The approximation

of the ℓ -dimensional function $g_T(\mathbf{u})$ is

$$g_T^K(\mathbf{u}) = \sum_{k=1}^{K^\ell} (\alpha_{0k} e_{0k} + (\boldsymbol{\alpha}_k)^T \mathbf{e}_k(\mathbf{u})), \quad (3.11)$$

where the superscript K denotes the number of intervals considered per dimension. The total number of hypercubes is straightforwardly given by K^ℓ with $2K$ basis functions and $\ell + 1$ parameters to be estimated per hypercube. Note that the number of parameters to be estimated does not grow linearly in K . Take K intervals per dimension. This results in $K^\ell(\ell + 1)$ parameters to be estimated. Now we increase the number of intervals per dimension by 1. Hence, we chop the support in each dimension in $(K + 1)$ intervals. The total number of parameters to be estimated is then $(K + 1)^\ell(\ell + 1)$. In fact, only for $\ell = 1$ the growth is linear in K . Note that the exact number of parameters to be estimated is simply proportional to the number of hypercubes since the term $(\ell + 1)$ is constant. The convergence is thus driven by the total number of hypercubes, which we denote by $\mathcal{K} := K^\ell$ in the following. We require the following assumptions.

Assumption 3.3. $A_T(Z)$ has a joint density w.r.t. Lebesgue measure which is a positive continuous function on D^ℓ .

Assumption 3.3 ensures that as the truncation parameter \mathcal{K} grows, the hypercubes can be made arbitrarily small.

Assumption 3.4. The target function is twice continuously differentiable and there is an $a < \infty$ such that

$$\|(\mathbf{H}g_T(\mathbf{u}))\|_{\max} = \max_{i,j=1,\dots,\ell} |h_{ij}| \leq a,$$

where $(\mathbf{H}g_T(\mathbf{u}))$ denotes the $\ell \times \ell$ Hessian matrix with its elements denoted by h_{ij} .

Now we can look at the fourth approximation error in 3.1 and obtain the below lemma.

Lemma 3.1. *If Assumptions 3.3 - 3.4 hold, $\gamma = 4/\ell$*

Proof. See Appendix 3.A. □

The ultimate convergence rate depends on how fast the number of hypercubes may grow with the sample size. Let us denote the number of hypercubes as a function of the sample size by $\mathcal{K}(N)$.

Lemma 3.2.

$$\tilde{h}(N, \mathcal{K}(N)) = O\left(\frac{\mathcal{K}(N)^2}{N}\right). \quad (3.12)$$

Proof. See Appendix 3.A. □

Now Theorem 3.1 is applicable and gives the convergence in mean-square

$$\mathbb{E}_{\tilde{\mathbb{P}}}\left[\left(g_T(A_T(Z)) - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z))\right)^2\right] = O_{\tilde{\mathbb{P}}}\left(\mathcal{K}(N)^{-4/\ell}\right). \quad (3.13)$$

A sufficient condition is that $\mathcal{K}(N) \propto N^a$, which is met for $a < 1/2$. Choosing a only slightly smaller than $1/2$ achieves a convergence rate of almost $N^{-2/\ell}$. In other words the convergence in each dimension is almost as fast as N^{-2} , which is twice as fast as the pure Monte Carlo rate of N^{-1} .

Two remarks are in order. First of all, the replicating portfolio method suffers from the curse of dimensionality problem if the multi-dimensional basis is constructed by the tensor product of the univariate basis, as we have shown here for the basis of piecewise linear functions. Lemma 3.1 confirms that the convergence rate slows down with increasing complexity in terms of the dimensionality of the problem. Second,

typically life insurance policies are high-dimensional path-dependent functions. Consider a payoff that depends on the (discrete) path of an underlying over a 30-year horizon and $\ell = 30$. The convergence would still be as fast as $N^{-0.07}$, which is considerable for a 30-dimensional function.

Remark 3.1. (On the practicality of piecewise linear functions)

At first sight piecewise linear functions may in practice be rather unusual. We now briefly show that piecewise linear functions are in fact (approximately) equal to linear combinations of call options. We consider the piecewise linear function in a hypercube, which has $\ell + 1$ components: the piecewise constant given by $e_{0k}(\mathbf{Y}) = \mathbb{1}_k(\mathbf{Y})C_{0k}$ and the piecewise linear parts $e_{jk}(\mathbf{Y}) = \mathbb{1}_k(\mathbf{Y})C_{jk}(Y_j - c_{jk})$, $j = 1, \dots, \ell$. Let a hypercube now be defined by the intervals $[b_{j,k}, b_{j,k+1})$ per dimension $j = 1, \dots, \ell$ and $k = 1, \dots, K$. In a hypercube the payoff of e_{0k} is C_{0k} , which is illustrated in the first displayed figure in Figure 3.3. The linear part, e_{jk} , has two components, $\mathbb{1}_k(\mathbf{Y})Y_j$ and $-\mathbb{1}_k(\mathbf{Y})c_{jk}$, which are given in the second and third displayed figures of Figure 3.3. The first and third component in Figure 3.3 can be perfectly expressed in terms of long and short cash-or-nothing digital options, which in turn can be approximately replicated by calls. The second component in Figure 3.3 can be perfectly replicated by a portfolio of long and short calls and cash-or-nothing digital options, which again in turn can be approximately replicated by calls only. Thus, in total the piecewise linear function can approximately be expressed in terms of calls. A digital option pays out some fixed amount if the option expires in-the-money, i.e. when the underlying exceeds the strike value. It is approximately statically replicated by the combination of a long and short call where the strikes differ by a very small amount. In particular, the static replicating portfolio for a digital cash-or-nothing option with strike $b_{j,k}$ paying \$ 1 if the option expires in-the-money is

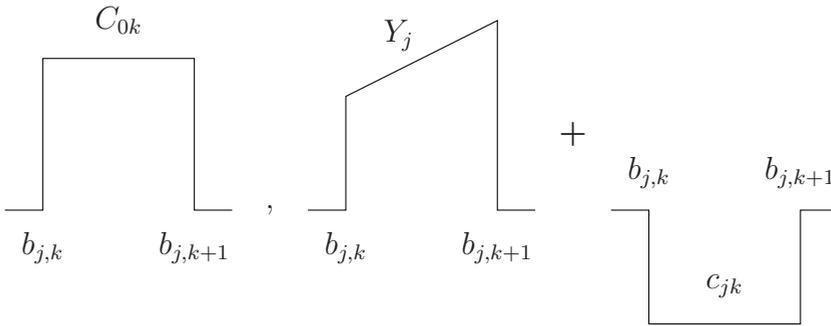


Figure 3.3: Components of piecewise linear function.

- long $\frac{1}{\epsilon}$ European calls at strike $b_{j,k}$,
- short $\frac{1}{\epsilon}$ European calls at strike $b_{j,k} + \epsilon$.

The hedging error can be made arbitrarily small by choosing ϵ . Note that as $\epsilon \rightarrow 0$ the representation in terms of calls converges pointwise to the payoff of the digital option. To see this write the payoff of the hedge as a function of Y_j :

$$f(Y_j) := \frac{1}{\epsilon} \left((Y_j - b_{j,k})^+ - (Y_j - b_{j,k} - \epsilon)^+ \right).$$

We look at

$$\lim_{\epsilon \rightarrow 0} f(Y_j) = \begin{cases} \lim_{\epsilon \rightarrow 0} 0 = 0 & \text{if } Y_j \leq b_{j,k} \\ \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \left((Y_j - b_{j,k}) - (Y_j - b_{j,k} - \epsilon)^+ \right) = 1 & \text{otherwise.} \end{cases}$$

The result follows as for any $Y_j > b_{j,k}$ we eventually have $Y_j > b_{j,k} + \epsilon$ as $\epsilon \rightarrow 0$.

3.5 Application to path-dependent insurance policies

In Section 3.2 we have shown that the Hilbert space model explained is consistent in achieving an approximation to the general contingent payoff X at time T under mild regularity assumptions such as limitation to finite variance contingent claims. This gives us the mathematical foundation for replicating portfolios. Complex problems that involve path-dependency can mathematically be handled since an orthonormal basis spanning the space must exist. The complexity of the problem depends on the number and type of risk drivers underlying the insurance policy. Within our model this is captured by the choice for $A_T(Z)$. In Examples 10 and 11 we have shown that $A_T(Z)$ is not unique. The important takeaway was that no matter how we choose $A_T(Z)$ it must contain all the information required to identify X . Going back to the examples we see that there is a trade-off between dimensionality and complexity of structure when choosing $A_T(Z)$. For a path-dependent insurance policy we can either choose the state vector such that it captures the path-dependency or by including each element on the path. While a higher dimensionality of $A_T(Z)$ slows down the convergence rate, a path-dependent $A_T(Z)$ complicates the easiness to price the basis. Note that a Hilbert space admits more than one basis. Now the difficult part in finding a good replication to a contingent payoff is to identify a sufficiently rich basis in combination with $A_T(Z)$ that at the same time lives up to our goal of expressing the target payoff X in a simplified way, i.e. in terms of the linear combination of “easier-to-work-with” basis functions, and for which the convergence rate is fast.

As has been shown in prior literature (see the discussion in Section 3.1) many insurance policies have a similar structure to well-known

financial instruments and a basis for replication is readily available. Consider first traditional life insurance policies that have neither guarantees nor embedded options. In terms of the financial risk the structure of these policies is very simple as a fixed amount is paid upon survival or death. With deterministic survival rates the expected cash flows can then simply be replicated through zero coupon bonds. In general, for insurance payoffs that are path-independent replicating portfolios are rather straightforward. However, many insurance contracts are highly path-dependent due to profit-sharing schemes and/or premium re-investments. Replicating such contracts becomes far more difficult. In this section we discuss the construction of replicating portfolios as described in Section 3.2 for a common participating life insurance policy and offer a discussion on the above mentioned issue. Focus is hereby put on the financial risk of insurance policies that exhibit path-dependency.

In participating life insurance contracts, also known as with-profit contracts, realized profits on the life insurance company's assets as well as technical profits are partially credited to the policyholders' account. The (legal) requirement to share profits also arises as traditionally premiums are calculated based on a conservative technical interest rate. The profits realized from conservative premiums are, at least partially, distributed back to the policyholder. Typically participating contracts offer an interest rate guarantee, which means that the policyholder profits from the upside development of the life insurance company's assets, but does not fully share in on the downside. Past returns on the life insurance company's assets are smoothed before being periodically credited to the policyholders' account. The intention of the insurer is to generate for its customers a stable, competitive and low-volatility return over time.

We consider the following contract specifications. At time 0 the policyholder acquires a participating insurance contract of nominal value L_0 through a single-premium payment. At initiation of the contract the insurance company invests the policyholders' premiums in the financial market and commits to periodically crediting interest to the policyholders' account according to its specified profit-sharing policy linked to the realized market return on its assets or some external reference index. Each period at least a guaranteed return of r_G is credited to the policyholder's account. With T the maturity of the contract the minimum guarantee at maturity is $G = L_0(1 + r_G)^T$. Commonly insurance companies have a profit sharing policy, according to which each period a potential bonus return is additionally credited to the policyholder's account. Rather than directly paying out the bonus, bonuses are accumulated and return on return is earned in the subsequent periods. Thus, each period the interest credited to the policyholders' account is the greater of either the minimum guarantee rate and the participating rate pertaining to the period. At maturity of the contract the insurer settles its obligation through a single payment to the policyholder. We define the rate credited to the policyholder's account at time t as

$$r_C(t) := r_G + r_B(t), \tag{3.14}$$

where $r_B(t)$ is the bonus rate given by

$$r_B(t) := \max(0, \delta R(t) - r_G) \tag{3.15}$$

with δ the distribution ratio. $R(t)$ denotes the reference rate for profit sharing at time t , which needs to be specified. The insurance company may, for instance, base this rate on an external reference index. In the Netherlands, for example, most insurance company's base $R(t)$ on the

moving average of the so-called u-rate, which is an average of different parts that are in turn calculated as weighted averages of an effective return on a basket of government bonds (see Pelsser and Plat, 2009). In Pelsser and Plat (2009) it is pointed out that $R(t)$ may also be a moving average of m -year swap rates. Profit sharing may also be based on the performance and profits of the insurance company. Grosen and Jørgensen (2000) offer here a prominent profit-sharing scheme, which we will elaborate on in more detail.

At maturity the value of the policyholder's account is then

$$L(T) = L_0 \prod_{t=1}^T (1 + r_C(t)) = L_0 \prod_{t=1}^T (1 + r_G + r_B(t)) \quad (3.16)$$

assuming that the policyholder receives $L(T)$ regardless of whether he is alive or not. $L(T)$ is similar to a floating rate note, except that the periodic coupons are not paid out, but are re-invested and accumulated up to the time to maturity. In each period t , $1 \leq t \leq T$, the accumulated notional rolls up with a variable rate $r_C(t)$. Thus, $L(T)$ can be seen as an exotic floating rate note. Note that the payoff $L(T)$ is the payoff function at time T . We furthermore remark that $L(t)$ at $t < T$ is the policyholder amount that has been accumulated up to time t and should not be confused with the value function of $L(T)$ at time t .

3.5.1 The Grosen and Jørgensen (2000) profit-sharing model

We now elaborate on the well-known participating contract modeled in Grosen and Jørgensen (2000). The crediting strategy derived in Grosen and Jørgensen (2000) is based on the simple asset-liability interaction view presented in Table 3.1. We largely follow here the notation used

in Grosen and Jørgensen (2000). We denote the market value at time t of the assets backing the contract by the stochastic process $Z(t)$, where this is an abbreviation of $Z(t, \cdot)$ with $t \in [0, T]$ of Z given by $\omega \rightarrow Z(t, \omega)$, $\omega \in \Omega$. We let $L(t)$ be the policy reserve, i.e. the book value of liabilities, and $B(t)$ is the bonus reserve calculated by the difference of $Z(t)$ and $L(t)$. As we will see $L(t)$ and $B(t)$ are functions of the asset process $Z(t)$. In the Grosen and Jørgensen (2000) model

Table 3.1: Fund-specific asset-liability interaction view at time t

Assets	Liabilities
$Z(t)$	$L(t)$
	$B(t)$

the bonus rate at time t is given by

$$\begin{aligned}
 r_B(t) &= \max \left(0, \delta \left(\frac{B(t-1)}{L(t-1)} - \lambda \right) - r_G \right) \\
 &= \max \left(0, \delta \left(\frac{Z(t-1) - L(t-1)}{L(t-1)} - \lambda \right) - r_G \right). \quad (3.17)
 \end{aligned}$$

The ratio of bonus reserves to policy reserves, $B(t-1)/L(t-1)$, is the buffer ratio. If the buffer ratio exceeds the target buffer ratio λ , then a positive fraction δ of the excess buffer is shared with the policyholder. Comparing Equation (3.17) with (3.15) we see that in the Grosen and Jørgensen (2000) model $R(t)$ is defined as

$$R(t) = \left(\frac{B(t-1)}{L(t-1)} - \lambda \right) = \left(\frac{Z(t-1) - L(t-1)}{L(t-1)} - \lambda \right). \quad (3.18)$$

It follows that $L(t)$ is given by

$$L(t) = L(t-1)(1 + r_C(t)) = L_0 \prod_{s=1}^t (1 + r_C(s)). \quad (3.19)$$

We again assume here that the policyholder receives $L(T)$ regardless of whether the policyholder is alive or not at maturity. The payoff at maturity is floored at $L_0(1 + r_G)^T$, which is the minimum guarantee. In that respect the insurance policy consists of a risk-free zero coupon bond with maturity T and face value equal to the maturity guarantee, $L_0(1 + r_G)^T$, and a bonus option. We can also again look at $L_0 \prod_{t=1}^T (1 + r_G + r_B(t))$ as an exotic floating rate note, where periodic coupons depend on the development of the bonus account and are accumulated up to maturity. The bonus account in turn is driven by the development of the underlying's asset performance and the consequent evolution of the liability fund. Clearly the payoff $L(T)$ is highly path-dependent.

We now want to write the terminal payoff $L(T)$ as a function of Z as $g_T(A_T(Z))$. In replicating the terminal payoff the first step is to define $A_T(Z)$. Looking at Equation (3.19) we can identify two extreme cases for $A_T(Z)$:

1. $A_T(Z) = \prod_{s=1}^T (1 + r_C(s))$,
2. $A_T(Z) = \{r_C(2), \dots, r_C(T)\}$.

Recall that the crediting rate $r_C(s) \forall s \in [0, T]$ is a function of the stochastic process Z . In the first case A_T is one-dimensional and the payoff $L(T)$ can be written as $g_T(A_T(Z)) = L_0 A_T(Z)$. As basis we consider the piecewise linear functions introduced in the previous section. While the problem is only one-dimensional, obtaining the time t value of the payoff function requires pricing the piecewise linear functions, which is complicated given a complicated $A_T(Z)$. A closed-form solution is not available. Thus, choosing the first option for $A_T(Z)$ transfers the problem of valuing the payoff function $L(T)$ to the problem of valuing the basis terms of the replicating portfolio. The first option is thus not constructive. In the second case $A_T(Z)$ is $(T - 1)$ -dimensional,

but each element is simple enough in its structure³. Choosing the second option for $A_T(Z)$ does not pose the basis valuation problem as in the first option, but results in a much higher dimensional replication problem. This requires a very extensive basis as the following example demonstrates.

Example 12. (*Empirical example: Curse of dimensionality*)

We consider a simple example for the second above choice of $A_T(Z)$. We model the underlying asset process simply as a geometric Brownian motion;

$$Z(t) = Z_0 \exp \left((\mu - 0.5 \sigma^2)t + \sigma W(t) \right). \quad (3.20)$$

$W(t)$ denotes the standard Brownian motion at time t . The parameters are given in Table 3.2. z_{\min} and z_{\max} denote the truncation values for simulating z in

$$Z(t) = Z(s) \exp \left((\mu - 0.5 \sigma^2)s + \sigma \sqrt{s}z \right), \quad s < t.$$

For the calibration we moreover choose to sample z from the uniform distribution on the pre-defined compact interval. We now want to find a replication with piecewise linear functions. The simulation procedure is as follows. For a fixed K the coefficients $\{\alpha_k\}_{k=1}^K$ are calculated based on the sample N , where the theoretical relation between the sample size and the number of hypercubes is used, i.e. $N \propto K^{\ell/a}$ with a only slightly smaller than $1/2$. To obtain the out-of-sample fit the basis functions are simulated on a sample m_{OoS} . By multiplying the estimated coefficients with the out-of-sample basis the fitting function is obtained based on the sample m_{OoS} . The mean square error for a

³To see why it is $(T - 1)$ -dimensional note that $r_C(1) = r_G + \max(0, \delta((Z_0 - L_0)/L_0 - \lambda) - r_G)$. Since Z_0 and L_0 are known, $r_C(1)$ is known and we omit it in $A_T(Z)$.

given K is then the average squared difference of the true payoffs on the sample m_{OS} and the payoffs from the estimated payoff function on the same sample. This process is repeated m_{MC} times to decrease the sampling error in the coefficient estimates. The final mean-square error for a given K is then the average over the m_{MC} collected mean-square errors.

Table 3.2: Parameters for Example 12.

δ	λ	r_G	μ	σ	Z_0
0.75	0.05	0.02	0.04	0.2	1.05
L_0	$\max K$	c	z_{\min}	z_{\max}	T
1	15	10	-2	2	3

We illustrate the empirical convergence rate for a 2-dimensional Grosen-Jorgensen payoff in Figure 3.4. Recall that the number of parameters to be estimated is proportional to the number of hypercubes K^ℓ . We have constructed up to $K^\ell = 15^2 = 225$ hypercubes with in total $K^\ell(\ell + 1) = 15^2 * 3 = 675$ parameters to be estimated. The logarithm of the mean-square error is plotted against the logarithm of the sample size. The green line indicates the convergence at rate N^{-1} . The red dots give the mean-square error data points of the example. The theoretical convergence rate for the 2-dimensional case is N^{-1} , which, as can be seen from the Figure, is not yet fully achieved constructing piecewise linear functions on 225 hypercubes. The number of hypercubes grows quickly in K requiring a large sample size and ultimately leading to a too high simulation effort.

The LSMC technique suffers from the curse of dimensionality problem. The challenge in constructing $A_T(Z)$ such that each piece of information on the path is considered as a driver lies in the high dimensionality of the problem. Consider a 30-dimensional Grosen-Jorgensen

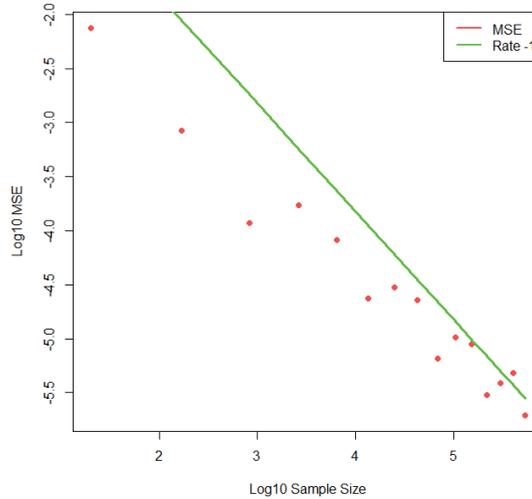


Figure 3.4: Empirical convergence rate with piecewise linear functions.

style contract, which is realistic for life insurance participating contracts that typically run for a long time horizon. Suppose we want to chop each dimension in 5 intervals. This gives us a total of $K^\ell = 5^{30} \approx 9.31 * 10^{20}$ hypercubes with piecewise linear functions constructed on each. The required sample size must then be proportional to 5^{60} . The example illustrates that the problem is infeasible in terms of the simulation effort.

3.5.2 Outlook: Replication with generalized Asian options

The previous analysis has revealed that while the replication approach explained in Sections 3.2 and 3.3 is theoretically sound, it poses significant practical challenges for complicated path-dependent insurance payoffs. Using vanilla-style basis functions, i.e. functions on path-

independent risk drivers, to replicate path-dependent insurance claims has the advantage that the basis is easy to price. We have seen in the previous section and from Example 12 that the disadvantage is the size of the basis required to reasonably replicate the payoff. Hence, the curse of dimensionality problem quickly dominates. We now attempt to look at a replication approach using path-dependent instruments that are still fairly easy to price. For that purpose we allow to deviate from our initial replicating portfolio framework and focus on an empirical analysis.

Recall that the Grosen-Jorgensen payoff depends on the paths of the underlying asset process, which we denote here again by the random process Z . Let us investigate to what extent weighted averages of the paths of the asset process help to fairly well explain the original Grosen-Jorgensen payoff. We have worked before with piecewise linear functions and have shown that these are similar to options. Here we choose generalized Asian-style options as instruments for replicating a Grosen-Jorgensen payoff. We define the k^{th} replicating instrument as follows.

$$e_k(\mathbf{Z}) = \max(0, \mathbf{a}_k^T \mathbf{Z} - d_k), \quad (3.21)$$

where $\mathbf{Z} = (Z(1), \dots, Z(T-1))^T$ refers to the underlying asset process over time, $\mathbf{a}_k = (a_1, \dots, a_{T-1})^T$ is a series of coefficients for the calculation of the weighted average and d_k is the strike. Moreover, for the first replicating instrument we take

$$e_0(\mathbf{Z}) = \mathbf{a}_0^T \mathbf{Z}$$

Table 3.3: Parameters for replication with generalized Asians.

δ	λ	r_G	μ	σ	Z_0
0.75	0.1	0.02	0.08	0.2	100
L_0	z_{\min}	z_{\max}	$\max K$	c	T
100	$-\infty$	∞	10	100	11

with $\mathbf{a}_k = (a_{0,1}, \dots, a_{0,T-1})^T$ a $(T - 1)$ -vector of coefficients. We let the replicating portfolio be given by

$$r_P(\mathbf{Z}) := \sum_{k=0}^K e_k(\mathbf{Z}) \tag{3.22}$$

and sequentially optimize for the parameters $\{\mathbf{a}_k, d_k\}$, $k = 0, \dots, K$ by minimizing the sum of squared errors. The starting value for the construction of the K^{th} generalized Asian option is the $(K - 1)^{\text{th}}$ result from the previous estimation. For $K = 0$ the optimized coefficients are then simply the least squares regression coefficients. We stress that while both the Grosen-Jorgensen payoff function and the replicating instruments we define here are functions of the random process Z , their structure is very different.

We attempt to find a good fit to the original Grosen-Jorgensen payoff, particularly in high-dimensional cases, without running into the curse of dimensionality problem. Consider a realistic Grosen-Jorgensen contract with maturity $T = 11$, which means that we are faced with a 10-dimensional problem. The parameterization is given in Table 3.2 and the asset process is specified in Example 12. Figure 3.5 gives the empirical convergence rate in terms of the mean square error for K up to 10. The logarithm of the mean-square error is again plotted against the logarithm of the sample size. Two reference rates are indicated by the light blue line (convergence at $2/\ell$, i.e. 0.2) and the dark blue line

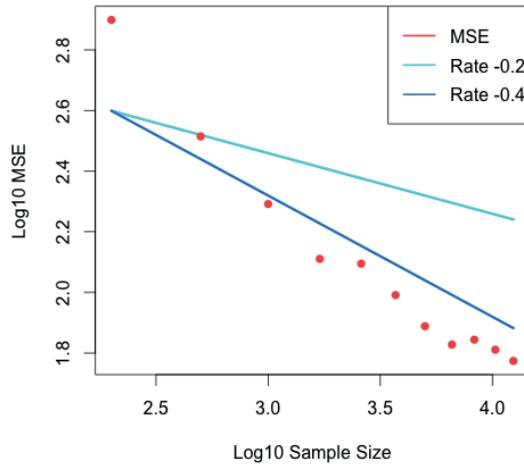


Figure 3.5: Empirical convergence rate for replication with Asians.

(convergence at 0.4). From the red dots it can be seen that the empirical convergence rate exceeds $2/\ell$, which is the theoretical convergence rate for a replication with piecewise linear functions. This shows that the replication with generalized Asian options is very promising.

Figures 3.6, 3.7 and 3.8 illustrate the variation of the target function and the replicating portfolio consisting of 10 generalized Asian options with the underlying asset process at different time points. In blue the true Grosen-Jorgensen payoff is plotted against the asset value at the respective time point; the analogue is plotted in red for the replicating portfolio. We see that the replicating portfolio mirrors the behavior of the target function very closely. The R-square from regressing the replicating portfolio against the target function is 99.64% and we illustrate the goodness of fit in Figure 3.9, where the replicating portfolio result with 10 generalized Asian options is plotted against the target Grosen-Jorgensen payoff. Note that the plots and the R-square are all based on out-of-sample data, i.e. data that has not been used

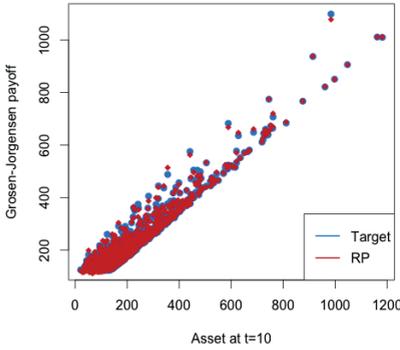


Figure 3.6: Comparison of variation with asset process at time $t = 10$.

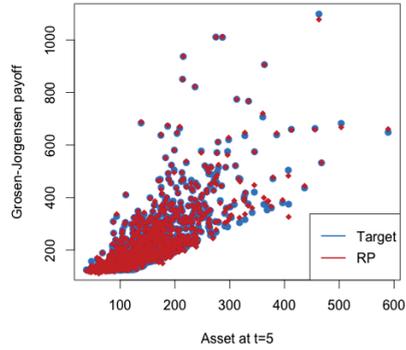


Figure 3.7: Comparison of variation with asset process at time $t = 5$.

for the calibration of the replicating portfolio. Overall we can conclude that a remarkably good fit has been achieved with a small number of basis terms.

The example shows that generalized Asian options are a promising alternative for replicating complicated path-dependent functions with still priceable instruments. Our attempt in this section was to offer an outlook for future research in this direction. The dependency structure of generalized Asian options on the random process differs from the dependency structure of the original Grosen-Jorgensen payoff function. It has not been discussed whether a perfect replicating portfolio can in fact be achieved when the number of replicating instruments grows to infinity. However, what has been shown is that with a reasonable number of replicating instruments a remarkable (out-of-sample) fit can be achieved, leaving a small error that may even be tolerable in practice.

Much further research is required to investigate the optimal basis selection for path-dependent payoffs and for overcoming the curse

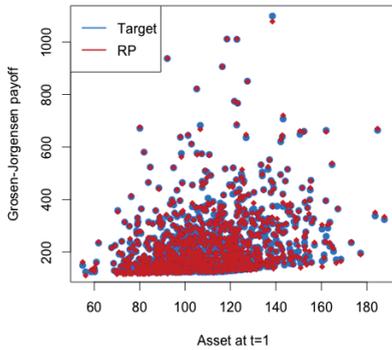


Figure 3.8: Comparison of variation with asset process at time $t = 1$.

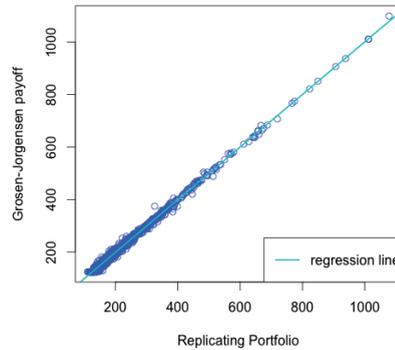


Figure 3.9: Illustration of goodness of fit of the replicating portfolio.

of dimensionality problem. Using generalized Asian options as we have done in this section is a promising attempt to mitigate between the curse of dimensionality problem and the priceability of the basis. While the empirical analysis shows promising results, future research is required to develop the theory.

3.6 Conclusion

In this paper, a general mathematical framework for statically replicating contingent claims through orthonormal basis functions has been established. Importantly, the static replication only requires to “invest” in a countable set of basis functions. Portfolio replication is a two-step approach to estimating the value of a contingent claim. The first step pertains to the construction of an approximation to the target function itself, i.e. the payoff function. By applying the conditional expectation operator to the basis functions of a replicating portfolio an estimate of the value function is obtained. The focus of this paper pertains to

the first step which is achieved by combining Monte Carlo simulation with least squares regression. In that respect the replicating portfolio approach belongs to the category of Least Squares Monte Carlo estimators. In LSMC two types have to be distinguished: Regress-Now and Regress-Later. In Regress-Later the target function at time T is regressed against basis functions valued at the same time point in order to construct a linear combination of basis terms that replicates the payoff of the target function. This procedure exactly corresponds to the construction of replicating portfolios and we conclude that LSMC Regress-Later and portfolio replication are the same thing. Based on the results for LSMC Regress-Later the asymptotic convergence rate of a replicating portfolio has been presented.

Multi-dimensional piecewise linear functions have been introduced as a convenient orthonormal basis, for which the convergence rate has been explicitly calculated. The developed tools have then been applied to a difficult path-dependent example often found in life insurance practice: a participating life insurance product, where the return rate credited to the policyholder depends on the performance of the insurance company's assets. Based on the participating life insurance product the practical challenges of portfolio replication have been revealed. More explicitly, it has been shown that the replicating portfolio is subject to the curse of dimensionality problem if a "naive" basis is chosen in high-dimensional problems. With "naive" basis we mean a multivariate basis, which is simply given as the tensor product of the univariate basis. A very promising alternative basis has been suggested, with which the curse of dimensionality problem may be overcome. The empirical results are very encouraging. Much room for research is left in analyzing eligible basis and their performance for sets of target functions, particularly pertaining to path-dependent target functions.

3.A Appendix: Proofs

Proof of Lemma 3.1. We first calculate the coefficients α_{0k} and $\boldsymbol{\alpha}_k$ $\forall k = 1, \dots, K^\ell$ exploiting the Taylor expansion of $g_T(\mathbf{u})$ in every hypercube H_k

$$g_T(\mathbf{u}) = g_T(\mathbf{c}_k) + (\mathbf{D}g_T(\mathbf{c}_k))^T (\mathbf{u} - \mathbf{c}_k) + R(\mathbf{u}), \mathbf{u} \in H_k$$

where $\mathbf{D}g_T(\mathbf{c}_k)$ denotes the first partial derivatives evaluated at \mathbf{c}_k . The remainder is specified as

$$R(\mathbf{u}) = \frac{1}{2}(\mathbf{u} - \mathbf{c}_k)^T (\mathbf{H}g_T(\boldsymbol{\xi})) (\mathbf{u} - \mathbf{c}_k), \boldsymbol{\xi} \in H_k,$$

where H_k denotes the k^{th} hypercube and \mathbf{H} the Hessian matrix. The following bounds are now helpful. We write the compact domain D^ℓ as

$$D^\ell = [d_{1,1}, d_{1,2}], \times \dots \times [d_{\ell,1}, d_{\ell,2}].$$

For the construction of K^ℓ orthogonal hypercubes the domain D^ℓ is cut on each dimension $i = 1, \dots, \ell$ in K equal length, non-overlapping intervals. Thus for each interval we have

$$(b_{i,k+1} - b_{i,k}) = \frac{(d_{i,2} - d_{i,1})}{K} := \frac{L_i}{K}, \forall i = 1, \dots, \ell.$$

Straightforwardly,

$$(u_i - c_{ik}) \leq (b_{i,k+1} - b_{i,k}) = O\left(\frac{1}{K}\right).$$

Then $R(\mathbf{u}) = O(1/K^2)$.

In the following we set $\mathbf{Y} := (A_T(Z))^T$. We now make use of the above Taylor expansion for $g_T(\mathbf{Y})$ and write

$$\begin{aligned} C_{0k}\alpha_{0k} &= C_{0k}\mathbb{E}_{\tilde{\mathbb{P}}}[e_{0k}(\mathbf{Y})g_T(\mathbf{Y})] \\ &= g_T(\mathbf{c}_k) + \frac{1}{2}C_{0k}^2\mathbb{E}_{\tilde{\mathbb{P}}}\left[\mathbf{1}_k(\mathbf{Y})(\mathbf{Y} - \mathbf{c}_k)^T(\mathbf{H}g_T(\boldsymbol{\xi}))(\mathbf{Y} - \mathbf{c}_k)\right] \end{aligned} \quad (3.23)$$

and

$$\begin{aligned} &C_k\boldsymbol{\alpha}_k \\ &= C_k\left(\mathbb{E}_{\tilde{\mathbb{P}}}\left[e_k(\mathbf{Y})(e_k(\mathbf{Y}))^T\right]\right)^{-1}\mathbb{E}_{\tilde{\mathbb{P}}}[e_k(\mathbf{Y})g_T(\mathbf{Y})] \\ &= C_k\left(\mathbb{E}_{\tilde{\mathbb{P}}}\left[e_k(\mathbf{Y})(e_k(\mathbf{Y}))^T\right]\right)^{-1} \\ &\quad \left(C_k\mathbb{E}_{\tilde{\mathbb{P}}}\left[\mathbf{1}_k(\mathbf{Y})(\mathbf{Y} - \mathbf{c}_k)(Dg_T(\mathbf{c}_k))^T(\mathbf{Y} - \mathbf{c}_k)\right] + \mathbb{E}_{\tilde{\mathbb{P}}}[e_k(\mathbf{Y})R(\mathbf{Y})]\right) \\ &= C_k\left(C_k\mathbb{E}_{\tilde{\mathbb{P}}}\left[\mathbf{1}_k(\mathbf{Y})(\mathbf{Y} - \mathbf{c}_k)(\mathbf{Y} - \mathbf{c}_k)^T\right]C_k\right)^{-1} \\ &\quad \left(C_k\mathbb{E}_{\tilde{\mathbb{P}}}\left[\mathbf{1}_k(\mathbf{Y})(\mathbf{Y} - \mathbf{c}_k)(\mathbf{Y} - \mathbf{c}_k)^T\right](Dg_T(\mathbf{c}_k)) + \right. \\ &\quad \left. C_k\mathbb{E}_{\tilde{\mathbb{P}}}\left[\mathbf{1}_k(\mathbf{Y})(\mathbf{Y} - \mathbf{c}_k)R(\mathbf{Y})\right]\right) \\ &= (Dg_T(\mathbf{c}_k)) + \frac{1}{2}\left(\mathbb{E}_{\tilde{\mathbb{P}}}\left[\mathbf{1}_k(\mathbf{Y})(\mathbf{Y} - \mathbf{c}_k)(\mathbf{Y} - \mathbf{c}_k)^T\right]\right)^{-1} \\ &\quad \mathbb{E}_{\tilde{\mathbb{P}}}\left[\mathbf{1}_k(\mathbf{Y})(\mathbf{Y} - \mathbf{c}_k)(\mathbf{Y} - \mathbf{c}_k)^T(\mathbf{H}g_T(\boldsymbol{\xi}))(\mathbf{Y} - \mathbf{c}_k)\right]. \end{aligned} \quad (3.24)$$

Recall that the above coefficients α_{0k} and $\boldsymbol{\alpha}_k$ for $k = 1, \dots, K^\ell$ of the approximating function g_T^K are optimal in the L^2 sense, i.e. they are chosen such that g_T^K minimizes $\mathbb{E}_{\tilde{\mathbb{P}}}[(g_T(\mathbf{Y}) - g_T^K(\mathbf{Y}))^2]$. By construction the K^ℓ hypercubes do not overlap and in sum cover the whole domain D^ℓ , on which we consider the target function $g_T(\mathbf{Y})$. As a result the coefficients α_{0k} and $\boldsymbol{\alpha}_k$ are also locally optimal, i.e. they minimize the squared deviation of the target and the approximating function on the

respective hypercube H_k . Let us now define the coefficients

$$\tilde{\alpha}_{0k} := (\mathbf{C}_{0,k})^{-1} g_T(\mathbf{c}_k)$$

$$\tilde{\boldsymbol{\alpha}}_k := (\mathbf{C}_k)^{-1} (\mathbf{D}g_T(\mathbf{c}_k))^T,$$

where we have only retained the first part of the equations for α_{0k} and $\boldsymbol{\alpha}_k$ in (3.23) and (3.24), respectively. Note that on every hypercube H_k the coefficients $\tilde{\alpha}_{0k}$ and $\tilde{\boldsymbol{\alpha}}_k$ are suboptimal, meaning that they may lead to a higher error in the mean-square sense than the optimal coefficients α_{0k} and $\boldsymbol{\alpha}_k$. For simplicity we replace α_{0k} and $\boldsymbol{\alpha}_k$ by the suboptimal coefficients in the remainder of the proof, where we want to derive the convergence rate for the fourth moment of the approximation error. Our line of reasoning is as follows. The approximation error using the true optimal coefficients must converge at least as fast as the approximation error using the suboptimal coefficients. Hence we provide for an upper bound by using the suboptimal coefficients.

Let f be the density on D^ℓ and consider the bounds $m_c := \min_{\mathbf{u} \in D^\ell} f(\mathbf{u}) > 0$ and $M_c := \max_{\mathbf{u} \in D^\ell} f(\mathbf{u}) < \infty$. We now turn to the evaluation of the fourth moment of the approximation error. Using the coefficients $\tilde{\alpha}_{0k}$ and $\tilde{\boldsymbol{\alpha}}_k$ we get in each hypercube H_k

$$\begin{aligned} & \left\{ \int_{H_k} \left\{ g_T(\mathbf{c}_k) + (\mathbf{D}g_T(\mathbf{c}_k))^T (\mathbf{u} - \mathbf{c}_k) \right. \right. \\ & \quad \left. \left. + R(\mathbf{u}) - \tilde{\alpha}_{0k} C_{0,k} - \tilde{\boldsymbol{\alpha}}_k \mathbf{C}_k (\mathbf{u} - \mathbf{c}_k) \right\}^4 f(\mathbf{u}) d\mathbf{u} \right\} \\ & = \int_{H_k} R(\mathbf{u})^4 f(\mathbf{u}) d\mathbf{u} \leq M_c \int_{H_k} R(\mathbf{u})^4 d\mathbf{u} = O\left(\frac{1}{K^{8+\ell}}\right). \end{aligned}$$

We stress again that if the convergence is as fast as $1/(K^{8+\ell})$ in each hypercube using the suboptimal coefficients $\tilde{\alpha}_{0k}$ and $\tilde{\boldsymbol{\alpha}}_k$, it must be at least as fast using the optimal coefficients. Summing over all

hypercubes we get for the fourth moment of the approximation error

$$\begin{aligned} \mathbb{E}_{\tilde{\mathbb{P}}}\left[\left(g_T(\mathbf{Y}) - g_T^K(\mathbf{Y})\right)^4\right] &\leq \sum_{k=1}^{\mathcal{K}} O\left(\frac{1}{K^{8+\ell}}\right) \leq \mathcal{K}O\left(\frac{1}{\mathcal{K}^{\frac{8+\ell}{\ell}}}\right) \\ &= O\left(\frac{1}{\mathcal{K}^{8/\ell}}\right) \end{aligned}$$

Lemma 3.1 follows immediately.

Remark 3.2. We want to remark that the proof given in this paper could also be used for the one-dimensional case. In Beutner et al. (2013) a slightly more extensive proof for the one-dimensional case is given. While in this paper the proof is simplified by using suboptimal coefficients in Beutner et al. (2013) the optimal coefficients are used. \square

Proof of Lemma 3.2. Let f again be the density on D^ℓ . Then $m_c := \min_{\mathbf{u} \in D^\ell} f(\mathbf{u}) > 0$ and $M_c := \max_{\mathbf{u} \in D^\ell} f(\mathbf{u}) < \infty$. For simplicity we again take $\mathbf{Y} := (A_T(Z))^T$. Moreover, we establish the following bounds, which we will need for the subsequent derivation. Recall that

$$(b_{i,k+1} - b_{i,k}) = \frac{L_i}{K}.$$

We define the upper and lower bounds by

$$\frac{L_{\min}}{K} \leq (b_{i,k+1} - b_{i,k}) \leq \frac{L_{\max}}{K}, \quad \forall i = 1, \dots, \ell.$$

Then,

$$\mathbb{E}_{\tilde{\mathbb{P}}}[\mathbf{1}_k(\mathbf{Y})] = \int_{H_k} f(\mathbf{u}) d\mathbf{u} \leq M_c \prod_{i=1}^{\ell} (b_{i,k+1} - b_{i,k}) \leq \frac{M_c (L_{\max})^\ell}{\mathcal{K}}$$

By the same token we have

$$C_{0k}^2 \leq \frac{\mathcal{K}}{m_c(L_{\min})^\ell}$$

Moreover, we need an upper bound for C_{ik}^2 . Observe that

$$\begin{aligned} & \mathbb{E}_{\tilde{\mathbb{P}}} [\mathbb{1}_k(\mathbf{Y})(Y_{ik} - c_{ik})^2] \\ &= \int_{H_k} (u_i - c_{ik})^2 f(\mathbf{u}) du_1 \dots du_\ell \\ &\geq \frac{m_c}{3} [(b_{i,k+1} - c_{ik})^3 - (b_{i,k} - c_{ik})^3] \prod_{j=1, j \neq i}^{\ell} (b_{j,k+1} - b_{j,k}) \\ &\geq \frac{m_c}{12} (b_{i,k+1} - b_{i,k})^3 \prod_{j=1, j \neq i}^{\ell} (b_{j,k+1} - b_{j,k}) \\ &\geq \frac{m_c}{12} \left(\frac{L_i}{K}\right)^3 \frac{(L_{\min})^{\ell-1}}{K^{\ell-1}} \geq \frac{m_c(L_{\min})^\ell}{12 \mathcal{K} K^2} \end{aligned}$$

where the result follows since $(b_{i,k+1} - c_{ik})^3 - (b_{i,k} - c_{ik})^3$ as a function of c_{ik} is minimized at $c_{ik} = (b_{i,k+1} + b_{i,k})/2$. Straightforwardly,

$$\begin{aligned} \max_{1 \leq i \leq \ell, 1 \leq k \leq K} C_{ik}^2 &\leq \frac{1}{\min_{i,k} \mathbb{E}_{\tilde{\mathbb{P}}} [\mathbb{1}_k(\mathbf{Y})(Y_{ik} - c_{ik})^2]} \\ &\leq \frac{12 \mathcal{K} K^2}{m_c(L_{\min})^\ell}. \end{aligned}$$

Now we can look at $\tilde{h}(N, \mathcal{K}(N))$

$$\begin{aligned} & \tilde{h}(N, \mathcal{K}(N)) \\ &= \frac{1}{N} \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(\sum_{k=1}^{\mathcal{K}(N)} (e_{0k}(A_T(Z))^2 + (e_k(A_T(Z)))^T e_k(A_T(Z))) \right)^2 \right] \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{N} \sum_{k=1}^{\mathcal{K}(N)} \mathbb{E}_{\tilde{\mathbb{P}}} \left[(e_{0k}(A_T(Z))^2 + (\mathbf{e}_k(A_T(Z)))^T \mathbf{e}_k(A_T(Z)))^2 \right] \\
 &= \frac{1}{N} \sum_{k=1}^{\mathcal{K}(N)} \mathbb{E}_{\tilde{\mathbb{P}}} \left[C_{0k}^4 \mathbf{1}_k(\mathbf{Y}) + \sum_{i=1}^{\ell} C_{ik}^4 (Y_i - c_{ik})^4 \mathbf{1}_k(\mathbf{Y}) \right. \\
 &\quad + 2 \sum_{i=1}^{\ell} C_{0k}^2 C_{ik}^2 (Y_i - c_{ik})^2 \mathbf{1}_k(\mathbf{Y}) \\
 &\quad \left. + 2 \sum_{i=1}^{\ell} \sum_{j=1, j \neq i}^{\ell} C_{ik}^2 C_{jk}^2 (Y_i - c_{ik})^2 (Y_j - c_{jk})^2 \mathbf{1}_k(\mathbf{Y}) \right] \\
 &\leq \frac{\mathcal{K}(N)}{N} \left(\max C_{0k}^2 + \ell \max_i (C_{ik}^4 (b_{i,k+1} - b_{i,k})^4) \mathbb{E}_{\tilde{\mathbb{P}}}[\mathbf{1}_k(\mathbf{Y})] \right. \\
 &\quad + 2\ell \max_i (C_{ik}^2 (b_{i,k+1} - b_{i,k})^2) \\
 &\quad \left. + 2 \ell(\ell - 1) \max_{i,j} (C_{ik}^2 C_{jk}^2 (b_{i,k+1} - b_{i,k})^2 (b_{j,k+1} - b_{j,k})^2) \mathbb{E}_{\tilde{\mathbb{P}}}[\mathbf{1}_k(\mathbf{Y})] \right) \\
 &\leq \frac{\mathcal{K}(N)}{N} \left(\frac{\mathcal{K}}{m_c(L_{\min})^\ell} + \ell \left(\frac{12 \mathcal{K} K^2}{m_c(L_{\min})^\ell} \right)^2 \left(\frac{L_{\max}}{K} \right)^4 \frac{M_c(L_{\max})^\ell}{\mathcal{K}} \right. \\
 &\quad + 2\ell \frac{12 \mathcal{K} K^2}{m_c(L_{\min})^\ell} \left(\frac{L_{\max}}{K} \right)^2 \\
 &\quad \left. + 2\ell(\ell - 1) \left(\frac{12 \mathcal{K} K^2}{m_c(L_{\min})^\ell} \right)^2 \left(\frac{L_{\max}}{K} \right)^4 \frac{M_c(L_{\max})^\ell}{\mathcal{K}} \right) \\
 &= \frac{\mathcal{K}(N)^2}{N} \left(\frac{1}{m_c(L_{\min})^\ell} + \frac{144\ell(L_{\max})^{4+\ell} M_c}{m_c^2(L_{\min})^{2\ell}} + \frac{24\ell(L_{\max})^2}{m_c(L_{\min})^\ell} \right. \\
 &\quad \left. + \frac{288\ell(\ell - 1) M_c(L_{\max})^{4+\ell}}{m_c^2(L_{\min})^{2\ell}} \right). \tag{3.25}
 \end{aligned}$$

□

Chapter 4

The Difference between LSMC and Replicating Portfolio in Insurance Liability Modeling

In this chapter we want to shed light on the differences between portfolio replication (also from the previous chapters known as Least Squares Monte Carlo with Regress-Later) and Least Squares Monte Carlo (with Regress-Now). Both are prominent and widely applied techniques for approximating the value of life insurance liabilities for risk management purposes. We show that while both are variants of regression-based Monte Carlo methods, they differ in one significant aspect. While the replicating portfolio approach only contains an approximation error, which converges to zero in the limit, in LSMC additionally a projection error is present, which cannot be eliminated. This difference has several (practical) implications, which are subsequently elaborated and underlined by simple, but compelling examples. This chapter thereby provides an overview of the advantages and disadvantages of one method over the other and may serve as guidance when choosing a method. It is found that the replicating portfolio technique enjoys numerous advantages stemming from the fact that the regression problem is non-standard and is therefore an attractive model choice¹.

4.1 Introduction

The Solvency II framework requires insurers to appropriately evaluate and manage embedded balance sheet risks. In the context of calculating risk figures insurers are challenged to revalue their liabilities under economic stress scenarios based on fair market valuation principles (see Article 76, The European Parliament and The Council, 2009). Particularly for life insurance liabilities, which contain embedded options and guarantees coming from policyholder participations, minimum guarantees and surrender options, this leaves the insurer with a strenuous task. As a consequence numerical methods involving Monte Carlo

¹This chapter is based on Pelsser and Schweizer (2015)

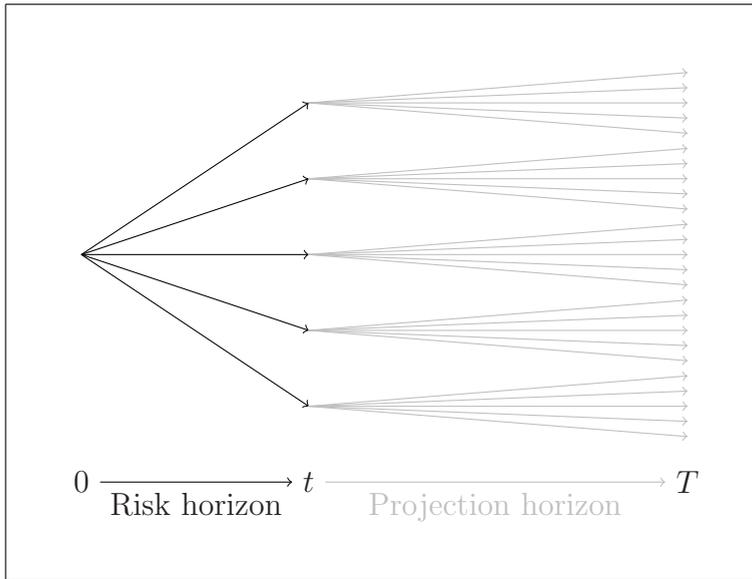


Figure 4.1: Nested stochastic simulation problem.

techniques for estimating the value of the liabilities have gained much attention. Procedures known as “nested simulation” or “full stochastic Monte Carlo simulation” take a full simulation approach, from which the empirical distribution of the liability values at the relevant point in time t is obtained. In insurance risk reporting t typically corresponds to one year. Based on the empirical distribution the estimate for the t year Value at Risk (VaR) can be derived, which is the Solvency II relevant risk figure. The nested simulation approach is illustrated in Figure 4.1, where the first simulation set from time 0 to time t represents the real-world scenarios over the risk horizon and the second set from time t to time T gives the risk-neutral scenarios for the estimation of the value at time t ; see also Bauer et al. (2009) and Beutner et al. (2015). Due to the scale and scope of a typical insurer’s life liabilities the nested stochastic simulation approach is computationally inefficient and regarding relevant reporting on the risk situation of the insurance

company a too timely exercise. For that reason alternative methods have been explored, which combine approximation methods with Monte Carlo techniques with the ambition to yield accurate risk capital figures within a reasonable time frame. Major discussions among practitioners evolve around two of these methods, largely known as Portfolio Replication and Least Squares Monte Carlo (LSMC) (see e.g. Bauer et al., 2009; Oechslin et al., 2007; Natolski and Werner, 2014). In this paper we want to shed light on the differences between these two approaches and the practical consequences resulting from that.

LSMC originates from the idea to estimate the continuation value of an American option through cross-sectional regression on Monte Carlo simulated paths. By going backward in time the American option price can thus be determined. Examples for LSMC in the context of American option pricing may be found in Carriere (1996), Longstaff and Schwartz (2001), Tsitsiklis and Van Roy (2001), Clement et al. (2002), Stentoft (2004), Glasserman and Yu (2004b), Egloff et al. (2007), Belomestny (2011), Gerhold (2011) and Zanger (2013). Andreatta and Corradin (2003) and Bacinello et al. (2009, 2010) apply the LSMC approach to the valuation of life insurance policies with surrender options. Devineau and Chauvigny (2011) show how the LSMC method can be extended to obtain a portfolio of replicating assets consisting of standard financial instruments. All these authors have in common that the static representations constructed immediately estimate the valuation function rather than the payoff function of the contingent claim. In the context of the insurance problem of estimating the risk capital at time t this means that the LSMC method yields an approximation function for the conditional expectation function at time t . This allows to quickly obtain an empirical distribution of the time t value under different real-world scenarios, from which the risk capital figure can then be extracted.

Glasserman and Yu (2004b) are the first to offer a different perspective on the LSMC method. They term LSMC techniques that directly estimate the valuation function “regression now” and propose a slightly different approach termed “regression later”. In “regression later” first the terminal payoff of the contingent claim is approximated by a linear combination of basis functions. The approximation to the valuation function at time t is then attained by evaluating the basis functions under the conditional expectation operator at time t . Both LSMC types, Regress-Now and Regress-Later, have further been investigated in Beutner et al. (2013). Moreover, in Beutner et al. (2015) it has been shown that the LSMC Regress-Later approach corresponds to the replicating portfolio technique. The principle of static replication is to construct a portfolio of financial instruments that mirrors the terminal payoff function of a target random variable. The static replicating portfolio is perfect if it replicates the target payoff in every possible state of the world. By the no-arbitrage condition, if the payoff of the target security is perfectly replicated, the replication automatically matches the security’s value at all times prior to maturity implying that they have the same market-consistent price. Given a replicating portfolio to the payoff of a contingent claim consists of instruments for which its values are readily available, the time t value can be quickly determined under different real-world scenarios, which again allows to extract risk capital figures. Naturally this feature has been exploited in the risk management of life insurance liabilities. Pelsser (2003) leverages the static portfolio replication concept to derive hedging strategies with swaptions for life insurance policies with guaranteed annuity options. Oechslein et al. (2007) consider how to set up replicating portfolios for life insurance liabilities in a more generalized approach. Recently, Natolski and Werner (2014) discuss and compare several approaches to the construction of replicating portfolios in life insurance. Chen and

Skoglund (2012), Daul and Vidal (2009), Kalberer (2012), Koursaris (2011b,c), and Burmeister (2007), for example, address the construction of replicating portfolios in life insurance from a more practical point of view and give recommendations. Taking the replicating portfolio as a proxy to the true liability payoff or the LSMC estimator as a proxy to the liability value at time t speeds up risk calculations tremendously. Thus, both methods fulfill the target to enable risk capital calculations for a life insurance portfolio. The straightforward question is then which method to use and why. The current literature offers little insight with regards to what the essential differences of these methods and their advantages over the other are. Glasserman and Yu (2004b) compare the properties of the coefficient estimates given that the approximations attained with LSMC Regress-Now and with LSMC Regress-Later yield a linear combination of the same basis functions. Their results suggest that in a single-period problem the LSMC Regress-Later algorithm yields a higher coefficient of determination and a lower covariance matrix for the estimated coefficients; see also Broadie and Cao (2008) who report similar observations. Beutner et al. (2013) remark that the functions to be approximated in LSMC Regress-Now may differ in nature compared to LSMC Regress-Later. Examples are provided which underline the observation. Several practitioners have touched upon a qualitative assessment of the advantages and disadvantages of particular proxy techniques, including LSMC and portfolio replication; see for example Koursaris (2011a), Morrison (2008), Hörig and Leitschkis (2012) and Hörig et al. (2014). While all these authors contribute to the discussion on the differences between LSMC and portfolio replication, no structured framework is provided to explain the observations. We attempt to close this gap with this paper.

In this paper we want to give insight on the fundamental differences between LSMC and portfolio replication. As has already been pointed out the replicating portfolio estimator corresponds to LSMC Regress-Later. When we use the brief terminology “LSMC” we refer to the Regress-Now type. Both are regression-based Monte Carlo methods, but we will accentuate that the one is a function fitting method while the other is truly a portfolio replication approach. As we will see this allows us to implement a simple measure in portfolio replication as a valuable indicator for the quality of the replicating portfolio. First the mathematical models for both approaches are presented, based on which the fundamental difference between the two methods is pinned down. Then we will elaborate on the consequences that follow from the difference between these methods. We will illustrate our conclusions with straightforward examples, which are simple, but compelling. Finally, we will address the challenges that arise for path-dependent insurance products.

The structure of this paper is as follows. In Section 4.2 we repeat the mathematical framework for LSMC and portfolio replication, which is largely taken from Beutner et al. (2013). We will highlight the mathematical difference between these two models, which builds the basis for the sections to follow. In Section 4.3 we elaborate on the consequences that result from the difference between LSMC and portfolio replication. In Section 4.4 the challenges for path-dependent payoff functions are addressed. Section 4.5 concludes.

4.2 The regression model for LSMC and Portfolio Replication

In this section we give the mathematical model and the estimation approach for the LSMC and the portfolio replication technique. We will see that the approaches are very similar but differ in one significant aspect. Both the model and the notation largely follow Beutner et al. (2013), which we repeat here.

Life insurance liabilities commonly generate several stochastic payoffs at different time points on a finite time horizon. The stochastic payoffs are typically driven by finitely many underlying risk drivers, which may be both of financial as well as non-financial nature. For our model we fix a finite time horizon T . We denote the terminal payoff of an insurance contingent claim at time T by X , which is driven by a d -dimensional stochastic process Z . We define the terminal cash flow as the sum of all cash flows over time $[0, T]$ accumulated at the money market account to the time point T . This is in line with the definitions in Oechslein et al. (2007) and Madan and Milne (1994). Let us now define the underlying dynamics of the contingent payoff X . Consider $Z = \{Z(t), 0 \leq t \leq T\}$ to be a d -dimensional stochastic process with $d \in \mathbb{N}$ defined on some filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0 \leq t \leq T}, \tilde{\mathbb{P}})$. We denote the filtration generated by Z by $\{\mathcal{F}_t\}_{0 \leq t \leq T}$. The measure $\tilde{\mathbb{P}}$ denotes some probability measure equivalent to the true probability measure \mathbb{P} . We interpret Z to be the ultimate d -dimensional random driver, on which the cash flows of an insurance contingent claim depend. We do not further specify Z , but remark that in principle it may account for both financial and non-financial risks. The paths $Z(\cdot, \omega)$ with $\omega \in \Omega$, of Z given by $t \rightarrow Z(t, \omega)$, $t \in [0, T]$, are assumed to lie in some function space $\mathbb{D}_d[0, T]$ consisting of functions mapping from $[0, T]$ to \mathbb{R}^d , and we consider Z as a random function. Recall that the

payoff function X is driven by Z . We assume that the payoff X is \mathcal{F}_T -measurable and we want to write X in terms of Z . However, as insurance contingent claims are typically path-dependent and generate multiple cash flows over time the payoff X at time T depends on the paths of $Z(\cdot, \omega)$. Thus, we define a process, denoted by $A_T(Z)$, which carries all the information on the paths of the d -dimensional stochastic process Z from time 0 to T that is relevant for the contingent claim X . We denote the dimensionality of A_T by ℓ_T , which is driven by the dependence structure on the d -dimensional process Z and the number of characteristics on the stochastic path that are required to determine X . Now we can write for every ω in the sample space Ω the payoff $X(\omega)$ of the contingent claim X as $g_T(A_T(Z(\cdot, \omega)))$, where A_T is a known (measurable) functional mapping from the function space $\mathbb{D}_d[0, T]$ to \mathbb{R}^{ℓ_T} and g_T is a known Borel-measurable function that maps from \mathbb{R}^{ℓ_T} to \mathbb{R} . Note that if we were only interested in plain vanilla contingent claims at time T it would suffice to observe the stochastic process Z at time T . But as insurance liabilities are often path-dependent, we need the information on the process of the underlying risk factors over time that is relevant for the contingent claim X , which we store in $A_T(Z)$.

The characterization of $A_T(Z)$ is subject to the specification of the modeler. Take the example of an Asian option with maturity T , where X gives the payoff of the Asian option at its maturity date T . In order to get the payoff it suffices to observe the time average of the underlying over the run-time of the Asian option. This information would be stored in $A_T(Z)$ and we would have $\ell_T = 1$. Alternatively, we may also observe the values of the underlying at each time point, which we would store in $A_T(Z)$. Then $\ell_T = T$. From this example we can see that $A_T(Z)$ is not unique, but depends on the choice of the modeler. We will return to this topic in Section 4.4.

As in Beutner et al. (2013) we restrict attention to finite second moment contingent claims and refer to the relevant related literature, in which the same assumption is applied (see e.g. Stentoft, 2004; Bergstrom, 1985; Madan and Milne, 1994; Longstaff and Schwartz, 2001). Thus we assume that the contingent claim X has finite mean and variance, which allows to model it as an element of a Hilbert space (see also Madan and Milne, 1994). More specifically, we assume that g_T belongs to the functional space $L_2(\mathbb{R}^{\ell_T}, \mathcal{B}(\mathbb{R}^{\ell_T}), \tilde{\mathbb{P}}^{A_T(Z)})$, where $\mathcal{B}(\mathbb{R}^{\ell_T})$ denotes the Borel σ -algebra on \mathbb{R}^{ℓ_T} , and $\tilde{\mathbb{P}}^{A_T(Z)}$ denotes the probability measure on \mathbb{R}^{ℓ_T} induced by the mapping $A_T(Z)$. Now, $L_2(\mathbb{R}^{\ell_T}, \mathcal{B}(\mathbb{R}^{\ell_T}), \tilde{\mathbb{P}}^{A_T(Z)})$ is a separable Hilbert space with inner product

$$\int_{\mathbb{R}^{\ell_T}} h_1(u)h_2(u) d\tilde{\mathbb{P}}^{A_T(Z)}(u) = \mathbb{E}_{\tilde{\mathbb{P}}} [h_1(A_T(Z))h_2(A_T(Z))]$$

and norm

$$\sqrt{\int_{\mathbb{R}^{\ell_T}} h_1(u)h_1(u) d\tilde{\mathbb{P}}^{A_T(Z)}(u)} = \sqrt{\mathbb{E}_{\tilde{\mathbb{P}}} [h_1^2(A_T(Z))]}$$

(Bogachev, 2007). Recall that a Hilbert space simply abstracts the finite-dimensional geometric Euclidean space to infinite dimensions (Conway, 1985). The theory for constructing the LSMC and the portfolio replication estimates is largely driven by the fact that under the restriction to finite variance contingent claims the payoff X is an element of a separable Hilbert space. This allows to express it in terms of a countable orthonormal basis. We will elaborate on the details in Sections 4.2.1 and 4.2.2, where the least squares regression models for LSMC and replicating portfolios, respectively, are presented.

Recall our initial problem of calculating risk figures. An insurer that needs to calculate the risk capital for its life insurance portfolio is ultimately interested in obtaining the empirical distribution for the

values of X at the risk horizon $t \leq T$, where t typically corresponds to one year in the Solvency II framework. Basically, the insurer is interested in the expectation of X conditional on information at time t . The nested stochastic simulation approach discussed in Section 4.1 is one path to obtain a solution to the problem. However, as previously pointed out the simulation effort is too high and in that respect the nested simulation approach is infeasible. LSMC and portfolio replication both reduce the simulation effort by requiring a smaller amount of inner simulations in Figure 4.1 to obtain an approximating function to the conditional expectation of X . However, they differ very much in the way the approximating function is constructed. While in LSMC an approximating function to $\mathbb{E}_{\mathbb{P}}[X|\mathcal{F}_t]$ is directly yielded through a least squares regression, portfolio replication focuses on approximating the payoff function X instead. This approximation is also obtained through least squares regression, but with different regressors than in LSMC. Given the approximating function for X its conditional expectation is estimated by applying the conditional expectation operator to the approximating function. This implies that regressors for the approximation to X must be chosen, for which the conditional expectation is either exact or can be quickly and fairly accurately estimated through numerical integration. Taking the above into account we will explain in the following two sections the least squares approaches for constructing the LSMC and the portfolio replication estimates.

4.2.1 Least Squares Monte Carlo

To this day the Least Squares Monte Carlo (LSMC) method has received much attention in the academic literature, particularly in the context of estimating the continuation value in American option pricing; see for example Longstaff and Schwartz (2001), Tsitsiklis

and Van Roy (2001) and also Stentoft (2004). Calculating risk capital figures for life insurance portfolios poses a similar problem to the extent that an unknown conditional expectation function must be estimated. Therefore, the LSMC method has also found its appeal in insurance risk modeling. Importantly, in LSMC the estimation of the conditional expectation function is achieved in one step by exploiting the cross-sectional information in Monte Carlo simulations and regressing across time using least squares. To describe the LSMC approach, we assume that the quantity of interest, $\mathbb{E}_{\mathbb{P}}[X|\mathcal{F}_t]$, can be written as

$$g_{0,t}(A_t(Z)) = \mathbb{E}_{\mathbb{P}} [X|\mathcal{F}_t], \quad 0 \leq t < T, \quad (4.1)$$

where A_t is a known (measurable) functional mapping from $\mathbb{D}_d[0, t]$ to \mathbb{R}^{ℓ_t} and $g_{0,t}$ is an unknown Borel-measurable function that maps from \mathbb{R}^{ℓ_t} to \mathbb{R} . Here, $\mathbb{D}_d[0, t]$ is the restriction of $\mathbb{D}_d[0, T]$ to the interval $[0, t]$ and ℓ_t denotes the dimensionality of $A_t(Z)$.

Remark 4.1. We use $g_{0,t}(A_t(Z))$ to denote the expected time t value of X , which is generally unknown. The subscript “0” is deliberately used to contrast the conditional expectation as an unknown function from the payoff function $g_T(A_T(Z))$, which is known in a simulation-based model as the simulation is controlled by the modeler.

In the following we describe the LSMC approach for estimating $g_{0,t}$. Recall that the square-integrability of X implies that $\mathbb{E}_{\mathbb{P}}[X|\mathcal{F}_t]$ is square-integrable as well. Hence, we also have that $g_{0,t} \in L_2(\mathbb{R}^{\ell_t}, \mathcal{B}(\mathbb{R}^{\ell_t}), \tilde{\mathbb{P}}^{A_t(Z)})$, which is again a separable Hilbert space. It is a well-known result that a separable Hilbert space has a countable orthonormal basis, in terms of which its elements may be expressed; see for instance Bogachev (2007, Corollary 4.2.2 and Corollary 4.3.4).

Then we can write $g_{0,t}$ as

$$g_{0,t} = \sum_{k=1}^{\infty} \beta_k v_k,$$

where $\{v_k\}_{k=1}^{\infty}$ is a countable orthonormal basis of the Hilbert space, in which $g_{0,t}$ lies. Because $g_{0,t}$ is the projection of X , the coefficients are given as

$$\beta_k = \mathbb{E}_{\tilde{\mathbb{P}}}[\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t] v_k(A_t(Z))] = \mathbb{E}_{\tilde{\mathbb{P}}}[X v_k(A_t(Z))]. \quad (4.2)$$

Thus, in particular, we have

$$g_{0,t}(A_t(Z)) = \sum_{k=1}^{\infty} \beta_k v_k(A_t(Z)). \quad (4.3)$$

and, as usual, we define the projection error $p_{0,t}$ by

$$p_{0,t}(A_t(Z)) := X - g_{0,t}(A_t(Z)). \quad (4.4)$$

The LSMC approach tries to estimate the unknown function $g_{0,t}$ through its representation in Equation (4.3) by generating data under $\tilde{\mathbb{P}}$. However, Equation (4.3) involves infinitely many parameters, which leaves a direct estimation infeasible. Consequently, finite-dimensional approximations with a truncated basis $\{v_k\}_{k=1}^K$, $K < \infty$, are used instead. For Equation (4.3) this implies that with sieves we approximate $g_{0,t}$ by

$$g_{0,t}^K := \sum_{k=1}^K \beta_k v_k = (\boldsymbol{\beta}_K)^T \mathbf{v}_K, \quad (4.5)$$

where $\boldsymbol{\beta}_K = (\beta_1, \dots, \beta_K)^T$, $\mathbf{v}_K = (v_1, \dots, v_K)^T$, and T denotes transpose. Thus, a superscript T means transpose and it should be easy to distinguish it from the terminal time T . This results in an approximation error $a_{0,t}^K$ for $g_{0,t}$ given by

$$a_{0,t}^K := g_{0,t} - g_{0,t}^K, \quad (4.6)$$

Notice that we have $\mathbb{E}_{\mathbb{P}}[g_{0,t}^K(A_t(Z))a_{0,t}^K(A_t(Z))] = 0$ by construction². By definition the approximation error $a_{0,t}^K$ converges to zero as $K \rightarrow \infty$. We can now write the following regression equation

$$X = g_{0,t}^K(A_t(Z)) + a_{0,t}^K(A_t(Z)) + p_{0,t}(A_T(Z)), \quad (4.7)$$

where the sum of the approximation and the projection error represents the regression error. Now, given a (simulated) sample of size N denoted by $((x_1, A_t(z_1)), \dots, (x_N, A_t(z_N)))$ it is natural to estimate $g_{0,t}^K$ by the ‘sample projection’

$$\hat{g}_{0,t}^K = \arg \min_{g \in \mathcal{H}_K} \frac{1}{N} \sum_{n=1}^N (x_n - g(A_t(z_n)))^2,$$

where $\mathcal{H}_K := \left\{ g : \mathbb{R}^{\ell_t} \rightarrow \mathbb{R} \mid g = \sum_{k=1}^K \beta_k v_k, \beta_k \in \mathbb{R} \right\}$. This corresponds to the least squares estimation of the above regression equation, i.e. from regressing the time T payoff of the contingent claim X against K explanatory variables valued at time t . Thus, we have

$$\hat{g}_{0,t}^K = \left(\hat{\boldsymbol{\beta}}_K \right)^T \mathbf{v}_K, \quad (4.8)$$

²We also remark that in the case where the basis includes a constant $\mathbb{E}_{\mathbb{P}}[a_{0,t}^K(A_t(Z))] = 0$.

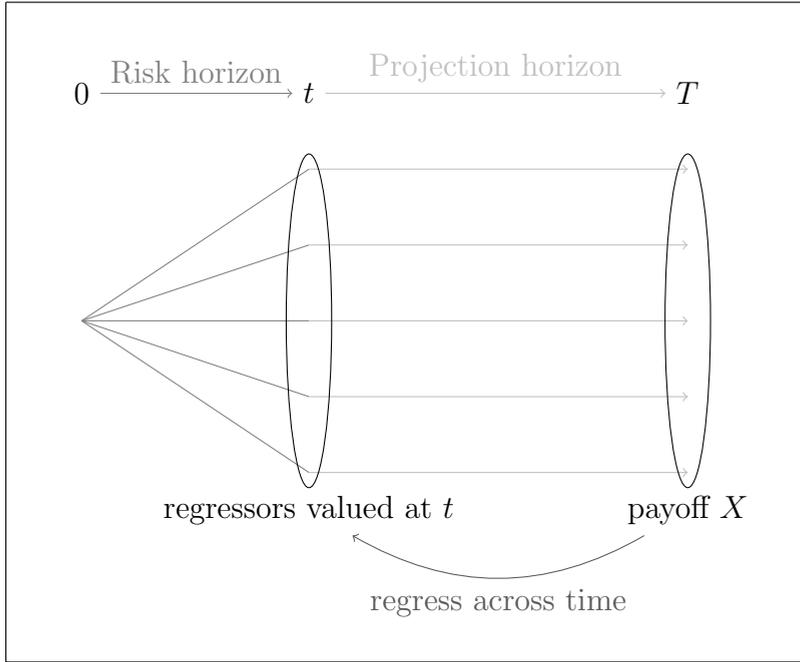


Figure 4.2: Illustration of the LSMC approach.

with

$$\hat{\beta}_K = \left((\mathbf{V}_K)^T \mathbf{V}_K \right)^{-1} (\mathbf{V}_K)^T \mathbf{X},$$

where $\mathbf{X} = (x_1, \dots, x_N)^T$ and \mathbf{V}_K is an $N \times K$ matrix with the n^{th} row equal to $\mathbf{v}_K(A_t(z_n))$, $n = 1, \dots, N$.

We illustrate the LSMC approach in Figure 4.2. Based on calibration scenarios the LSMC estimator is constructed by regressing the payoff function X against regressors valued at time t . The least squares regression approach naturally provides thereby an estimate for the conditional expectation function $\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t]$. Given this estimate the distribution of time t values over real-world scenarios constructed on the risk horizon can be obtained.

Naturally the LSMC estimator is subject to an error. More specifically, the LSMC estimator $\hat{g}_{0,t}^K$ involves three sources of error resulting from an approximation, a projection and an estimation error. This can also nicely be seen from Equation (4.5), which gives the regression equation. The regression error here consists of the approximation and the projection error. The estimation error arises from estimating the coefficients of the regression equation based on a finite sample. While the approximation error vanishes for $K \rightarrow \infty$ and the estimation error for $N \rightarrow \infty$, the projection error cannot be eliminated in the limit. The nonzero projection error arises from projecting the cash flows across the time interval $[t, T]$. To better see the impact of the projection error on the estimation result, consider the estimation error of the coefficients,

$$\begin{aligned}(\hat{\beta}_K - \beta_K) &= \left((\mathbf{V}_K)^T \mathbf{V}_K \right)^{-1} (\mathbf{V}_K)^T (\mathbf{X} - \mathbf{V}_K \beta_K) \\ &= \left((\mathbf{V}_K)^T \mathbf{V}_K \right)^{-1} (\mathbf{V}_K)^T \left((\mathbf{X} - \mathbf{V} \beta) + \right. \\ &\quad \left. (\mathbf{V} \beta - \mathbf{V}_K \beta_K) \right) \\ &= \left((\mathbf{V}_K)^T \mathbf{V}_K \right)^{-1} (\mathbf{V}_K)^T (\mathbf{p}_{0,t} + \mathbf{a}_{0,t}^K)\end{aligned}$$

Observe that the projection error can in fact only be eliminated by regressing the payoff X valued at time T against regressors valued at the same time point. This brings us to the replicating portfolio approach, which we address in the following section.

4.2.2 Portfolio Replication

In the previous section we have discussed the LSMC approach, which obtains an estimate to the time t value of a contingent claim by regressing the payoffs at time T resulting from a Monte Carlo simulation

sample against basis functions valued at time t . In contrast, in this section we are first interested in constructing an estimate to the payoff function X , i.e. we construct a static replicating portfolio to the payoff function. Then, given the linear representation of X through basis functions, apply the operator $\mathbb{E}_{\tilde{\mathbb{P}}}[\cdot|\mathcal{F}_t]$ to these basis functions. The approach takes advantage of the linearity of the expectation operator. Note that the two-step approach is advantageous if for the payoff function X basis functions are used whose conditional expectation is easily obtained. For the case where $\tilde{\mathbb{P}} = \mathbb{Q}$ with \mathbb{Q} denoting the risk-neutral measure this implies that closed-form solutions for the price of the basis functions must be readily available. The replicating portfolio approach corresponds to the LSMC Regress-Later approach first discussed in Glasserman and Yu (2004b); see also Beutner et al. (2013).

Remember that we assume square-integrability of the payoff function, meaning that $g_T \in L_2(\mathbb{R}^{\ell_T}, \mathcal{B}(\mathbb{R}^{\ell_T}), \tilde{\mathbb{P}}^{A_T(Z)})$. Hence, by the same argument as in Section 4.2.1,

$$X = g_T(A_T(Z)) = \sum_{k=1}^{\infty} \alpha_k e_k(A_T(Z)), \quad (4.9)$$

where $\{e_k\}_{k=1}^{\infty}$ is a countable orthonormal basis of $L_2(\mathbb{R}^{\ell_T}, \mathcal{B}(\mathbb{R}^{\ell_T}), \tilde{\mathbb{P}}^{A_T(Z)})$.

We use a different notation for the coefficients and the basis functions than in Section 4.2.1 to emphasize that in general the basis functions chosen for LSMC may differ from the ones used in portfolio replication, the reason being that the functions to be approximated in LSMC and in portfolio replication may differ in nature. Recall that in LSMC we directly estimate the conditional expectation function while in portfolio replication the approximation refers to the payoff function. Take the example of a call option. The payoff has a kinked structure,

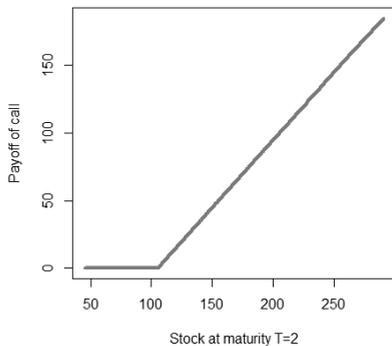


Figure 4.3: Payoff function at maturity for a call with maturity $T = 2$.

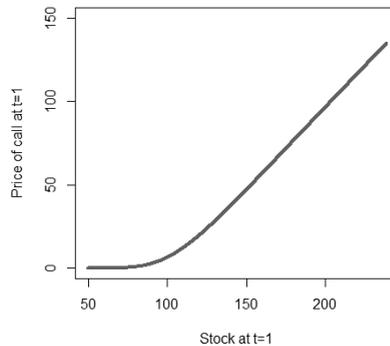


Figure 4.4: Pricing function at $t = 1$ for a call with maturity $T = 2$.

but the conditional expectation function is smooth (see Figures 4.3 and 4.4). Thus, for that specific example polynomials are a convenient basis in LSMC to approximate the smooth conditional expectation function, while for the payoff function piecewise linear functions are, for instance, more appropriate in order to replicate the kink.

The coefficients α_k are given by

$$\alpha_k = \mathbb{E}_{\tilde{\mathbb{P}}} [X e_k(A_T(Z))]. \quad (4.10)$$

As for LSMC the representation of X in Equation (4.9) involves infinitely many parameters, which leaves a direct estimation infeasible. Consequently, the right-hand side of Equation (4.9) is truncated to a finite number K ;

$$g_T^K = \sum_{k=1}^K \alpha_k e_k = (\boldsymbol{\alpha}_K)^T \mathbf{e}_K, \quad (4.11)$$

where $\boldsymbol{\alpha}_K = (\alpha_1, \dots, \alpha_K)^T$ and $\mathbf{e}_K = (e_1, \dots, e_K)^T$. Defining the approximation error a_T^K as usual by $a_T^K := g_T - g_T^K$ we obtain the representation

$$X = g_T^K(A_T(Z)) + a_T^K(A_T(Z)). \quad (4.12)$$

This gives the regression equation for the replicating portfolio problem, where a_T^K represents the regression error³. Now given a (simulated) sample of size N denoted by $(x_1, A_T(z_1)), \dots, (x_N, A_T(z_N))$ we estimate g_T^K by least squares regression leading to

$$\hat{g}_T^K = (\hat{\boldsymbol{\alpha}}_K)^T \mathbf{e}_K, \quad (4.13)$$

with

$$\hat{\boldsymbol{\alpha}}_K = \left((\mathbf{E}_K)^T \mathbf{E}_K \right)^{-1} (\mathbf{E}_K)^T \mathbf{X}, \quad (4.14)$$

where $\mathbf{X} = (x_1, \dots, x_N)^T$ and \mathbf{E}_K is an $N \times K$ matrix with the n^{th} row equal to $\mathbf{e}_K(A_T(z_n))$, $n = 1, \dots, N$. Notice that $\hat{\boldsymbol{\alpha}}_K$ corresponds to the usual least squares estimator from a regression of X against K basis functions valued at time T . Recall that in Regress-Now, in contrast, X is regressed against basis functions valued at time t .

We illustrate the replicating portfolio approach in Figure 4.5. Based on calibration scenarios the replicating portfolio estimator is constructed by regressing the payoff function X against regressors valued at the same time point T . The least squares regression approach naturally provides thereby an estimate for the payoff function X since $\mathbb{E}_{\mathbb{P}}[X|\mathcal{F}_T] = X$. Given this estimate the time t value of the regressors must be determined to get an estimate for the conditional expectation

³We remark again that in the case where the basis includes a constant $\mathbb{E}_{\mathbb{P}}[a_T^K(A_T(Z))] = 0$.

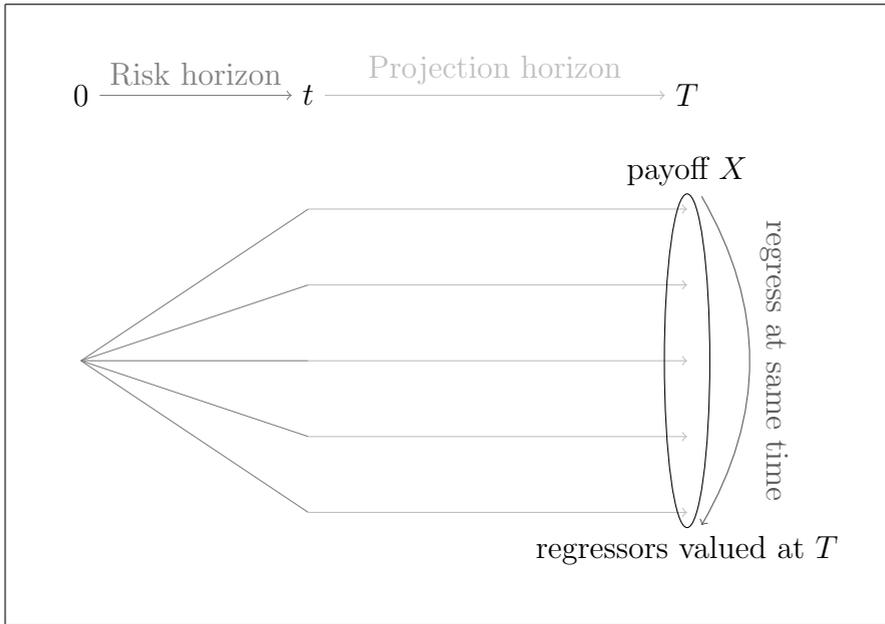


Figure 4.5: Illustration of the Replicating Portfolio approach

function $E_{\tilde{p}}[X|\mathcal{F}_t]$. This in turn can then be used to obtain an empirical distribution of the time t values at the risk horizon t in order to extract risk figures.

Just like the LSMC estimator the replicating portfolio estimator is also subject to an error. However, the replicating portfolio estimator \hat{g}_T^K involves only two sources of error resulting from an approximation and an estimation error. The estimation error again arises from estimating the coefficients of the regression equation based on a finite sample and converges to zero as $N \rightarrow \infty$. To better see this we again consider the coefficient error

$$(\hat{\alpha}_K - \alpha_K) = \left((\mathbf{E}_K)^T \mathbf{E}_K \right)^{-1} (\mathbf{E}_K)^T (\mathbf{X} - \mathbf{E}_K \beta_K)$$

$$\begin{aligned}
 &= \left((\mathbf{E}_K)^T \mathbf{E}_K \right)^{-1} (\mathbf{E}_K)^T \left((\mathbf{X} - \mathbf{E}\boldsymbol{\alpha}) + \right. \\
 &\quad \left. (\mathbf{E}\boldsymbol{\alpha} - \mathbf{E}_K\boldsymbol{\alpha}_K) \right) \\
 &= \left((\mathbf{E}_K)^T \mathbf{E}_K \right)^{-1} (\mathbf{E}_K)^T \mathbf{a}_T^K.
 \end{aligned}$$

Remark 4.2. We remark again that the functions to be approximated with LSMC and portfolio replication differ. In LSMC we directly estimate the conditional expectation function while in portfolio replication the approximation to the conditional expectation function is obtained by applying the conditional expectation operator to the obtained proxy of the payoff function. This also implies that the error of the time t value in portfolio replication is not $a_T^K(A_T(Z))$, but $\mathbb{E}_{\mathbb{P}}[a_T^K(A_T(Z))|\mathcal{F}_t]$. Since the replicating portfolio is used in the Solvency II context as a proxy to the liability value in extreme scenarios, ensuring a very small error at time t is of utmost importance. We will later return to this point.

Compare the regression equation for LSMC (4.7) with the regression equation of the replicating portfolio (4.12). Clearly, the regression error of LSMC is composed of an approximation and a projection error, while the regression error of the replicating portfolio only contains an approximation error. Notice that for both methods the approximation error vanishes for $K \rightarrow \infty$. For the replicating portfolio this implies that the regression error converges to zero as the number of basis functions grows. The replicating portfolio approach is thus a non-standard regression problem. In contrast, even when the approximation error is zero, the LSMC regression error still contains the projection error. We will discuss the implications of the replicating portfolio being a non-standard regression problem in the next section.

4.3 Impact of the zero projection error in Portfolio Replication

In Section 4.2 we have outlined the Monte Carlo regression frameworks for constructing LSMC and replicating portfolio estimates. We have stressed that in LSMC the payoff function X at time T is regressed against basis functions valued at time $t < T$, while in portfolio replication it is regressed against basis functions valued at the same time point T . This subtle, but critical distinction leads to very different characterizations of the regression problem. The regression error of the replicating portfolio method only contains an approximation error, which converges to zero in the limit as more and more basis terms are included in the representation. The LSMC regression error also contains an approximation error, which vanishes in the limit, but, due to the time gap of the regressand and the regressors the regression error additionally contains a projection error. The difference in the composition of the regression error has several consequences that we want to illuminate throughout the subsequent sections.

4.3.1 Function fitting versus Portfolio Replication

We have earlier pointed out that two types of Least Squares Monte Carlo approaches are discussed in the literature: LSMC Regress-Now, which we have referred to as LSMC in this paper, and LSMC Regress-Later. Also, we have indicated that LSMC Regress-Later is actually portfolio replication and we have used this terminology throughout the paper. Now we want to take a closer look at the reason why the least squares regression framework for replicating portfolios in Section 4.2.2 is truly a replication approach and why the least squares regression framework for LSMC in Section 4.2.1 is not.

Let us first clarify the terms “replicating portfolio” and “function fitting”. A replicating portfolio of a target claim is a portfolio of instruments that has the same properties as the target. In line with the definitions in Oechslin et al. (2007) and Madan and Milne (1994) we consider a replicating portfolio as a portfolio of instruments that has the same terminal cash flow as the target. By construction we achieve this in the Hilbert space framework of Section 4.2, where the replicating portfolio of X is given by the infinite basis representation of Equation (4.9). The regression equation for X then involves an approximation error from truncating the basis to $K < \infty$. With function fitting we refer to the construction of a smooth function that best approximates observed data. Least squares regression in its standard form is a data fitting approach that focuses on finding a smooth curve that best explains the variation in observed data with random errors. Now, for both LSMC and portfolio replication we apply the least squares regression technique. However, for LSMC we approximate an unknown function based on noisy data while for portfolio replication we want to find an exact representation for the (known) payoff function based on simulated data points. Thus in LSMC we face a noisy regression while in portfolio replication the regression is non-noisy even when the approximation error is nonzero. To better see this we will next analyze the variance of the residuals in both LSMC and portfolio replication.

Let us consider the regression error in LSMC first, which is given by the sum of the approximation and the projection error, i.e. $a_{0,t}^K(A_t(Z)) + p_{0,t}(A_T(Z))$. For the variance of the regression error we obtain

$$\begin{aligned}
 & \text{Var} (a_{0,t}^K(A_t(Z)) + p_{0,t}(A_T(Z))) && (4.15) \\
 &= \text{Var} (a_{0,t}^K(A_t(Z))) + \text{Var} (p_{0,t}(A_T(Z))) \\
 &= \sum_{k=K+1}^{\infty} \beta_k^2 - (\mathbb{E}_{\mathbb{P}}[a_{0,t}^K(A_t(Z))])^2 + \mathbb{E}_{\mathbb{P}}[X^2] - \mathbb{E}_{\mathbb{P}}[(\mathbb{E}_{\mathbb{P}}[X|\mathcal{F}_t])^2]
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{k=K+1}^{\infty} \beta_k^2 - (\mathbb{E}_{\tilde{\mathbb{P}}}[a_{0,t}^K(A_t(Z))])^2 + \mathbb{E}_{\tilde{\mathbb{P}}}[X^2] - \sum_{k=1}^{\infty} \beta_k^2 \\
 &= \mathbb{E}_{\tilde{\mathbb{P}}}[X^2] - \sum_{k=1}^K \beta_k^2 - (\mathbb{E}_{\tilde{\mathbb{P}}}[a_{0,t}^K(A_t(Z))])^2, \tag{4.16}
 \end{aligned}$$

where we have exploited that

$$\mathbb{E}_{\tilde{\mathbb{P}}}[p_{0,t}(A_T(Z))v_k(A_t(Z))] = 0 \quad \forall k.$$

Notice that as the approximation error vanishes for $K \rightarrow \infty$ the variance of the regression error converges to the variance of the projection error, i.e.

$$\begin{aligned}
 \text{Var}(p_{0,t}(A_T(Z))) &= \mathbb{E}_{\tilde{\mathbb{P}}}[(p_{0,t}(A_T(Z)))^2] \\
 &= \mathbb{E}_{\tilde{\mathbb{P}}}[X^2] - \mathbb{E}_{\tilde{\mathbb{P}}}[(\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t])^2] \tag{4.17}
 \end{aligned}$$

$$= \mathbb{E}_{\tilde{\mathbb{P}}}[X^2] - \sum_{k=1}^{\infty} \beta_k^2. \tag{4.18}$$

Since we know that X is expressible in terms of an infinite orthonormal basis, i.e. $X = \sum_{k=1}^{\infty} \alpha_k e_k(A_T(Z))$, we can even write

$$\text{Var}(p_{0,t}(A_T(Z))) = \sum_{j=1}^{\infty} \alpha_j^2 - \sum_{k=1}^{\infty} \beta_k^2. \tag{4.19}$$

We also want to investigate the conditional variance of the regression error:

$$\begin{aligned}
 &\text{Var}(a_{0,t}^K(A_t(Z)) + p_{0,t}(A_T(Z)) | \mathcal{F}_t) \\
 &= \text{Var}(a_{0,t}^K(A_t(Z)) | \mathcal{F}_t) + \text{Var}(p_{0,t}(A_T(Z)) | \mathcal{F}_t) \\
 &\quad + 2 \text{Cov}(a_{0,t}^K(A_t(Z)), p_{0,t}(A_T(Z)) | \mathcal{F}_t) \\
 &= \mathbb{E}_{\tilde{\mathbb{P}}}[(p_{0,t}(A_T(Z)))^2 | \mathcal{F}_t]
 \end{aligned}$$

$$= \text{Var} [X | \mathcal{F}_t]. \quad (4.20)$$

This is the conditional variance of the target function X . Depending on the underlying stochastic processes and the structure of X it may well be that the conditional variance of the time T random payoff X varies with observations at time t . Therefore, in LSMC we may potentially deal with heteroskedastic residuals.

We repeat the analysis of the variance of the regression error for the replicating portfolio approach. Recall that the regression error in portfolio replication is given by $a_T^K(A_T(Z))$. For the variance we obtain

$$\begin{aligned} \text{Var} (a_T^K(A_T(Z))) &= \mathbb{E}_{\mathbb{P}} \left[(a_T^K(A_T(Z)))^2 \right] - (\mathbb{E}_{\mathbb{P}}[a_T^K(A_T(Z))])^2 \\ &= \sum_{k=K+1}^{\infty} \alpha_k^2 - (\mathbb{E}_{\mathbb{P}}[a_T^K(A_T(Z))])^2. \end{aligned} \quad (4.21)$$

Clearly the variance converges to zero in the limit for $K \rightarrow \infty$ as the perfect replicating portfolio is attained. Let us take a look at the conditional variance of the residual of the replicating portfolio problem:

$$\text{Var} (a_T^K(A_T(Z)) | \mathcal{F}_T) = 0. \quad (4.22)$$

The zero conditional variance of the residuals implies that there is no variation of the error at each observation of $A_T(Z)$. This actually makes sense as the residual simply reflects the approximation error, which is clearly defined at each observation of $A_T(Z)$. We can therefore understand the replicating portfolio approach as non-noisy even when the approximation error is nonzero. Summing up, in portfolio replication the conditional variance of the residuals is zero and the unconditional variance of the residuals converges to zero as the number of basis terms grows. Thus, the perfect replicating portfolio is attained that truly reproduces the terminal payoff X . Consequently, the least

squares regression approach underlying the replicating portfolio approach is not a typical regression approach of fitting a function through a cloud of data. In the following we give two simple examples which illustrate the non-standard regression problem in portfolio replication and the noisy regression problem in LSMC.

Example 13. (*Simple Brownian Motion*)

Let us consider the most simple example, where the approximation errors are zero for LSMC and portfolio replication. The payoff function is given by $X = W_T$ with W_T being a standard Brownian motion. As regressors we take W_t for LSMC and W_T for portfolio replication. Obviously, for portfolio replication a perfect fit is achieved. Consequently, the conditional expectation function $g_t(W_t) = W_t$ is also perfectly fit for any $t \leq T$. For LSMC the approximation error is zero, but we are still faced with a noisy regression due to the persistence of the projection error. The projection error is

$$p_{0,t}(W_T) = X - \mathbb{E}[X|\mathcal{F}_t] = W_T - W_t.$$

As Brownian motions have stationary independent increments the distribution of $(W_T - W_t)$ is independent of information at time t . Therefore we have

$$\begin{aligned}\text{Var}(W_T - W_t) &= \text{Var}(W_T - W_t|\mathcal{F}_t) \\ &= T - t.\end{aligned}$$

We illustrate this in Figures 4.6 and 4.7, where we have plotted the LSMC and the portfolio replication regression problem for the simple Brownian motion example with $t = 1$ and $T = 10$. Figure 4.6 gives the LSMC regression problem by plotting the regressand W_T against the regressor W_t . Least squares regression of W_T on W_t returns the function

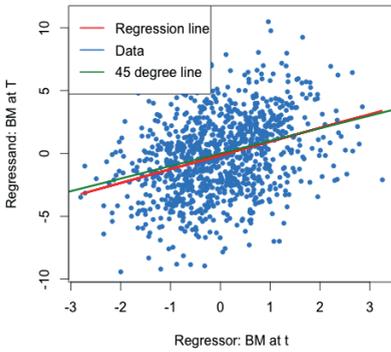


Figure 4.6: Noisy regression in LSMC (Example 13).

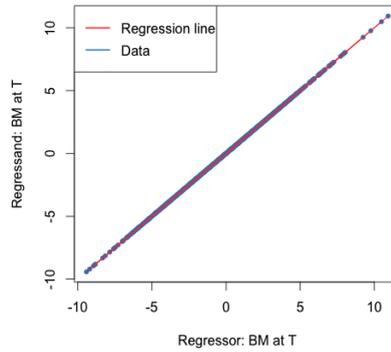


Figure 4.7: Regression in portfolio replication (Example 13).

that best fits the cloud of data (represented by the red regression line). By construction the best line is the conditional expectation $\mathbb{E}_{\mathbb{P}}[W_T|W_t]$ (represented by the 45 degree line plotted in green).

Example 14. (*Exponential function*)

We take a simple exponential function to be replicated

$$X = e^{\sigma W_T}$$

with W_T a standard Brownian motion. The conditional expectation is then

$$\mathbb{E}[X|\mathcal{F}_t] = e^{\sigma W_t + \frac{1}{2}\sigma^2(T-t)}.$$

We investigate the following LSMC and portfolio replication regression equations

$$X = \beta_0 + \beta_1 W_t + \epsilon_t$$

$$X = \alpha_0 + \alpha_1 W_T + \epsilon_T.$$

Recall that for LSMC the regression error ϵ_t consists of an approximation and a projection error, while for the replicating portfolio problem the non-standard regression error ϵ_T involves only an approximation error. We can clearly see this from Figures 4.8 and 4.9, where the regression line is given by the red line and the data is represented by the blue points⁴. For the example at hand the LSMC regression problem is heteroskedastic (Figure 4.8). Even if the approximation error was zero in LSMC, the projection error persists and the noisy regression would still be heteroskedastic. To see this consider the conditional variance of the projection error

$$\begin{aligned}\text{Var}(p_{0,t}(W_T)|\mathcal{F}_t) &= \mathbb{E} \left[e^{2\sigma W_T} | \mathcal{F}_t \right] - e^{2\sigma W_t + \sigma^2(T-t)} \\ &= e^{2\sigma W_t + \sigma^2(T-t)} \left(e^{\sigma^2(T-t)} - 1 \right),\end{aligned}$$

which clearly increases for larger values of the Brownian motion at time t .

Example 15. (*Artificial portfolio with perfect basis*)

In this example we construct a portfolio of puts and calls in the Black-Scholes framework. As basis we use the components that make up the payoff function, which ensures that at least theoretically the perfect representation for both portfolio replication and LSMC is available. Let X be as defined below

$$X = 100 - 2(K_1 - S(T))^+ + (S(T) - K_2)^+ - 2(S(T) - K_3)^+$$

⁴Note that the example serves to show that LSMC is a noisy regression approach, while portfolio replication is truly a replication approach. No conclusions on the quality of the results are drawn. In fact, for the simple exponential example at hand, LSMC and portfolio replication yield the same result (see also Example 20):

$$\alpha_0 = \beta_0 = \exp\left(\frac{1}{2}\sigma^2 T\right); \quad \alpha_1 = \beta_1 = \sigma \exp\left(\frac{1}{2}\sigma^2 T\right).$$

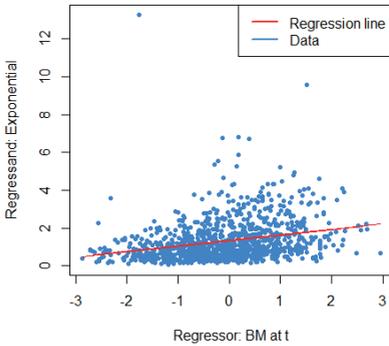


Figure 4.8: Noisy regression in LSMC (Example 14).

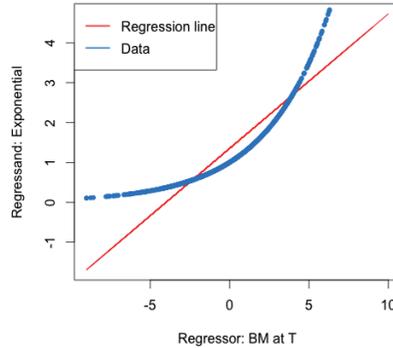


Figure 4.9: Regression in portfolio replication (Example 14).

Table 4.1: Parameters for Example 15.

t	T	μ	σ	r	S_0	N	m
1	5	0.08	0.2	0.02	100	1200	5000

$$+ (S(T) - K_4)^+ + 0.5(S(T) - K_5)^+ - 0.5(S(T) - K_6)^+ \quad (4.23)$$

with strikes $K_i = S_0 e^{(\mu - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}z_i}$ where $\{z_i\}_{i=1}^6 = \{-1.5, -0.5, 0, 1, 1.5, 2\}$. The parameters are defined in Table 4.1, where r is the risk-free rate, N is the sample size of the calibration set and m is the sample size for the out-of-sample set. Ultimately we want to find an approximation to the price of X at time t . We estimate the replicating portfolio by regressing the values of X against the basis and price the basis using the Black-Scholes formula in order to obtain the pricing function at time t . With LSMC an estimate of the pricing function at time t is directly obtained by regressing the discounted payoff X against the time t -prices of the basis. The calibration sample set is based on the risk-neutral measure here. We will come back to the

relevance of the measure in Section 4.3.4. As the correct price of the target function X is available in the Black-Scholes framework the LSMC and portfolio replication results can be assessed against it.

The optimal solution for the coefficients of the LSMC and replicating portfolio representation is

$$\boldsymbol{\alpha} = \boldsymbol{\beta} = (100, -2, 1, -2, 1, 0.5, -0.5)^T. \quad (4.24)$$

When estimating the replicating portfolio on a sufficiently diverse scenario set exactly these coefficients are obtained. Given the perfect replicating portfolio also the conditional expectation at any $t < T$ is perfectly obtained by pricing the basis terms. For LSMC we do not get the exact result for the coefficients although the perfect basis is available. Figure 4.10 illustrates the imperfect fit that results on a sample of size $N = 1200$. The dark blue line gives the target conditional expectation function while the turquoise line shows the LSMC solution. Increasing the sample size to $N = 1\,000\,000$ the LSMC solution approximates the true conditional expectation function very well (see Figure 4.11) and an R^2 of 99.99% is achieved. The estimated coefficients, though, are

$$\hat{\boldsymbol{\beta}} = (101.82, -2.10, 0.19, -0.15, -4.12, 8.65, -5.25)^T \quad (4.25)$$

and thus differ from the coefficients that would return the replicating portfolio. Clearly, LSMC is a function fitting method and not a portfolio replication method.

Example 16. (*Equity Swap*)

In this example we consider a simple equity swap with payoff at maturity T

$$X = S_2(T) - S_1(T),$$

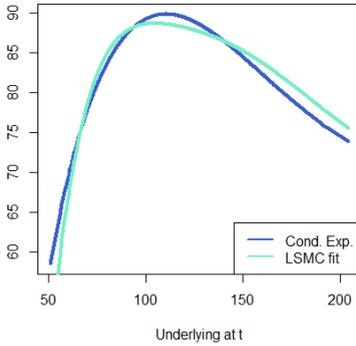


Figure 4.10: LSMC fit for $N = 1\,200$ (Example 15).

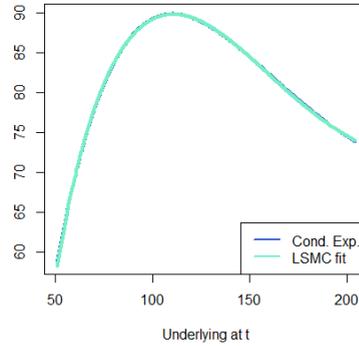


Figure 4.11: LSMC fit for $N = 1\,000\,000$ (Example 15).

where $S_1(T)$ and $S_2(T)$ are modeled as uncorrelated geometric Brownian motions

$$S_i(T) = S_i(0)e^{(\mu_i - \frac{1}{2}\sigma_i^2)T + \sigma_i W(T)}, \quad i = 1, 2$$

with parameters $\mu_1 = 0.08$, $\sigma_1 = 0.2$, $\mu_2 = 0.05$ and $\sigma_2 = 0.15$. The payoff X depends on the values of both assets $S_1(T)$ and $S_2(T)$. Its conditional expectation function at time t also requires the information of both assets at time t , $S_1(t)$ and $S_2(t)$. Let us now consider the construction of both replicating portfolio and LSMC estimates, where the risk factors are not correctly identified. In other words, the regression equation misses regressors constructed on relevant risk factors. The regression functions are specified for portfolio replication and LSMC, respectively as,

$$X = \alpha_0 + \alpha_1 S_1(T) + \epsilon_T$$

$$X = \beta_0 + \beta_1 S_1(t) + \epsilon_t.$$

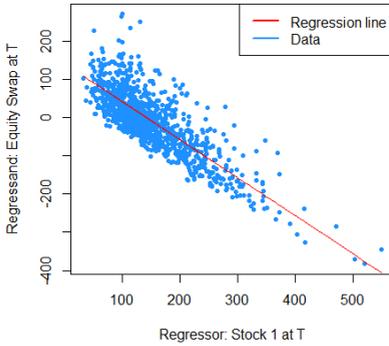


Figure 4.12: Missing risk factors regression in portfolio replication (Example 16).

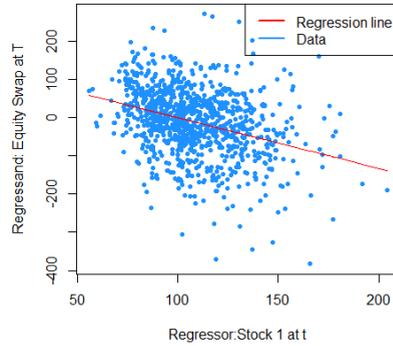


Figure 4.13: Missing risk factors regression in LSMC (Example 16).

Figures 4.12 and 4.13 illustrate the regression of the payoff function X against $S_1(T)$ in portfolio replication and $S_1(t)$ in LSMC. Both figures reveal noisy regressions. While for LSMC a noisy regression is not surprising, for portfolio replication this is not expected if all risk factors have been correctly identified. Consequently, risk factors must have been neglected in the replicating portfolio. Note that for LSMC this conclusion cannot be drawn as the regressions are always noisy. Regressing only against $S_1(t)$ still yields an estimated conditional expectation function, i.e. the expectation conditional on the smaller information set $S_1(t)$, but this is not the conditional expectation function of interest. For the replicating portfolio missing the information of $S_2(T)$ the resulting R^2 is 66.75%. For the LSMC regression it is 11.21%. The details on R^2 as a measure for the goodness of fit of both portfolio replication and LSMC will be explained in Section 4.3.2. Nonetheless, it is worthwhile to mention at this point that in LSMC it is usual to observe a low R^2 . In portfolio replication, in contrast, a low R^2 either signifies a large approximation error,

i.e. a larger number of basis functions is required to obtain a better replicating portfolio⁵, or, risk factors are missing, i.e. $A_T(Z)$ is not correctly identified.

The last example has shown that with the LSMC approach a conditional expectation is always estimated, it may just not be the one we are actually interested in. Due to the time gap of the regressand and the regressors the LSMC regression is noisy by construction. Detecting the issue of potentially having neglected relevant risk factors is therefore difficult. For portfolio replication the regression is not noisy given that all underlying risk factors of the payoff function have been identified. R^2 is a useful measure that provides important information on the approximation error of the regression in portfolio replication (see Section 4.3.2). A low R^2 may moreover be an indicator for missing risk factors.

So far we have delivered the argument that LSMC is a function fitting approach as its least squares regression is noisy. The least squares approach to portfolio replication is, in contrast, non-standard as the regression error converges to zero in the limit and the conditional variance of the residuals is zero. In that context there is one more argument why the least squares approach of Section 4.2.2 is truly a replicating portfolio approach while the least squares approach of Section 4.2.1 is not. In portfolio replication the payoff function at time T is approximated. The conditional expectation function at any $t < T$ is then obtained by calculating the time t value of the basis terms that make up the approximation of the target payoff function X . The better the replicating portfolio mirrors the payoff function

⁵This means that K should be increased, i.e. more basis terms built on the already identified risk factor. We remark that a higher K in principle also requires a larger sample size as more parameters need to be estimated. The relation of K and N is also addressed in Section 4.3.5 in the context of the asymptotic convergence properties of LSMC and portfolio replication.

at time T the better the fit to the conditional expectation functions at any time $t < T$. Straightforwardly this implicates a great amount of flexibility, particularly if the conditional expectation at several time points is of interest. With LSMC, in contrast, the conditional expectation at a particular $t^* < T$ is approximated by regressing basis terms valued at time t^* against the target payoff function X valued at time T . The result is an approximation of the conditional expectation at the particular time point t^* and does not necessarily imply an approximation of the conditional expectations at times $t < T$ with $t \neq t^*$. Consider the representations for X and $\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t]$

$$g_T(A_T(Z)) = \sum_{k=1}^{\infty} \alpha_k e_k(A_T(Z))$$
$$g_{0,t}(A_t(Z)) = \sum_{k=1}^{\infty} \beta_k v_k(A_t(Z)).$$

Moreover,

$$g_T(A_T(Z)) = g_{0,t}(A_t(Z)) + p_{0,t}(A_T(Z)).$$

Given the replicating portfolio of X we obtain $\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t]$ for any $t < T$ by taking the conditional expectation of the basis terms, i.e.

$$\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t] = \sum_{k=1}^{\infty} \alpha_k \mathbb{E}_{\tilde{\mathbb{P}}}[e_k(A_T(Z))|\mathcal{F}_t].$$

For the LSMC representation of the conditional expectation at a particular time point $t^* < T$, $g_{0,t^*}(A_{t^*}(Z)) = \sum_{k=1}^{\infty} \beta_k v_k(A_{t^*}(Z))$, the same holds for $t < t^*$ only if we can compute the conditional expectations of

the basis terms and the projection error, i.e.

$$\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t] = \sum_{k=1}^{\infty} \beta_k \mathbb{E}_{\tilde{\mathbb{P}}}[v_k(A_{t^*}(Z))|\mathcal{F}_t] + \mathbb{E}_{\tilde{\mathbb{P}}}[p_{0,t^*}(A_T(Z))|\mathcal{F}_t], \quad t < t^*.$$

It is to be expected that the calculation of the conditional expectation of the projection error is most likely not straightforward, particularly when considering that LSMC is used in applications, for which already the time t^* conditional expectation is not closed-form available. In order to get $\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t]$ for $t^* < t < T$ $g_{0,t^*}(A_{t^*}(Z))$ must be corrected by the time t conditional expectation of the projection error

$$\begin{aligned} \mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t] &= g_{0,t^*}(A_{t^*}(Z)) + \mathbb{E}_{\tilde{\mathbb{P}}}[p_{0,t^*}(A_T(Z))|\mathcal{F}_t], \quad t^* < t < T \\ &= g_{0,t^*}(A_{t^*}(Z)) + (\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t] - \mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_{t^*}]), \end{aligned}$$

where again the calculation of the conditional expectation of the projection errors is probably not straightforward. Moreover, it cannot simply be inferred that the LSMC representation at time t^* also holds at time t , $t > t^*$, by valuing the basis at time t . Thus given the time t^* coefficients $\{\beta_{k,t^*}\}_{k=1}^{\infty}$, which we denote with the subscript t^* , it cannot be inferred that

$$\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t] = \sum_{k=1}^{\infty} \beta_{k,t^*} v_k(A_t(Z)).$$

Example 17. (*Example 15 revisited: Artificial portfolio with perfect basis*)

Reconsider Example 15, for which a very good fit to the conditional expectation $\mathbb{E}[X|\mathcal{F}_1]$ has been found with LSMC. Using the estimated coefficients in (4.25) and the prices of the basis at time $t = 4$ the resulting fit to the conditional expectation at time $t = 4$ is assessed. Figure 4.14 highlights that the LSMC coefficients calibrated to the

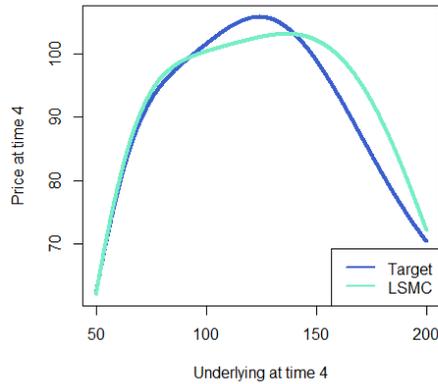


Figure 4.14: Illustration of LSMC fit at $t = 4$ with calibration at time 1 (Example 17).

conditional expectation at time 1 do not imply a good fit to the conditional expectation at a different time point. This is in contrast to a portfolio replication approach. Remember that with portfolio replication the correct coefficients as in (4.24) have been identified. Thus, automatically, the conditional expectation for any $t < T$ is also perfectly obtained by applying the conditional expectation operator to the replicating portfolio.

Example 18. (*LSMC with Hermite polynomials*)

The simple exponential payoff function from Example 14 is taken, for which the LSMC technique with a basis of Hermite polynomials is applied to approximate its conditional expectation function. Let $T = 5$ and $\sigma = 0.2$. We simulate 1000 paths of a Brownian motion, $\{W_{t^*}, W_T\}$ with $t^* = 1$, and consider the Hermite polynomials on $(W_{t^*}/\sqrt{t^*})$. With only $K = 5$ Hermite terms a reasonably good fit is achieved, which is visualized in Figure 4.15. However, taking the coefficients from the time $t^* = 1$ calibration and valuing the Hermite

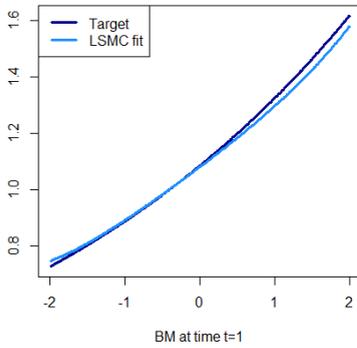


Figure 4.15: LSMC fit at $t^* = 1$ given calibration at $t^* = 1$ (Example 18).

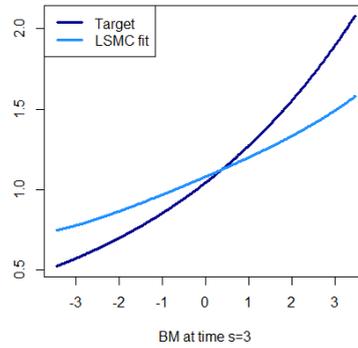


Figure 4.16: LSMC fit at $t = 3$ given calibration at $t^* = 1$ (Example 18).

polynomials at a different time point t , $t^* < t < T$, does not yield a good representation for the conditional expectation function at time t . Figure 4.16 illustrates this for $t = 3$. The example indicates that a good representation of the conditional expectation at a particular time point does not imply a similarly good representation of the conditional expectation at a different time point.

Summarizing we can generally infer the following. The least squares Monte Carlo approach of Section 4.2.2 is truly a replicating portfolio approach for the subsequent reasons.

- The better the replicating portfolio approximates the target payoff function, the better the resulting fit to the conditional expectation function at any time t , $t < T$. A perfect replicating portfolio thus implies a perfect fit to the conditional expectation function at any time t , $t < T$.

- This is linked to the fact that the least squares Monte Carlo approach of Section 4.2.2 is nonstandard resulting in a non-noisy regression.

LSMC is a function fitting method where the estimation of the conditional expectation function at a specific time point is attained by fitting a curve through noisy data. The LSMC representation of the conditional expectation at a distinct time point does not straightforwardly suggest a representation to the conditional expectation at a different time point.

4.3.2 Upper limit of R-square: LSMC versus Portfolio Replication

In this section we want to discuss the squared correlation coefficient R^2 as a measure for the goodness of fit. We will explain that the measure R^2 , which is typically calculated in a least squares regression, is not a useful measure in LSMC, but is meaningful for portfolio replication. In fact the R^2 in LSMC can never be 1 even if the conditional expectation function is perfectly fitted, unless $t = T$, which would not correspond to LSMC, but to portfolio replication. Intuitively we expect the target function X at time T to correlate more strongly with regressors valued at time T than with regressors valued at time t . This is more pronounced the greater the gap between the time points t and T . Thus, the observation that R^2 is not meaningful for LSMC is caused by the persistence of the projection error in LSMC.

In a first step and to illustrate our result we assume now that the approximation error is zero, i.e. we compare the LSMC and portfolio replication result with zero approximation error. From the LSMC

regression equation we then have

$$X = \mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t] + p_{0,t}(A_T(Z))$$

with $p_{0,t}$ the projection error. R^2 is defined as the explained sum of squares (SSE) over the total sum of squares (SST), which can be written as

$$R^2 = \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\text{SSR}}{\text{SST}},$$

where SSR is the residual sum of squares. By construction for portfolio replication with zero approximation error the R^2 is 1 as the residual sum of squares is zero. For portfolio replication this means in general that the higher the R^2 the smaller the approximation error and the closer the portfolio replication estimator is to the true result. For LSMC we get the following R^2

$$\begin{aligned} R_{lsmc}^2 &= \frac{\mathbb{E}_{\tilde{\mathbb{P}}} [(\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t] - \mathbb{E}_{\tilde{\mathbb{P}}}[X])^2]}{\mathbb{E}_{\tilde{\mathbb{P}}} [(X - \mathbb{E}_{\tilde{\mathbb{P}}}[X])^2]} \\ &= 1 - \frac{\mathbb{E}_{\tilde{\mathbb{P}}} [(X - \mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t])^2]}{\mathbb{E}_{\tilde{\mathbb{P}}} [(X - \mathbb{E}_{\tilde{\mathbb{P}}}[X])^2]} \\ &= 1 - \frac{\text{Var}(p_{0,t}(A_T(Z)))}{\text{Var}(X)} = 1 - \frac{\sum_{j=1}^{\infty} \alpha_j - \sum_{k=1}^{\infty} \beta_k^2}{\text{Var}(X)}, \end{aligned} \tag{4.26}$$

which is only equal to 1 if the projection error is zero or equivalently $\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t] = X$. By the definition of X and for $t < T$ this is not the case. Also note that Equation (4.26) is the upper bound for the R^2 that can be maximally attained in LSMC. It gives the R^2 when only the projection error remains as regression error, meaning that

the approximation error is zero and a perfect fit to the conditional expectation function $g_{0,t}$ is achieved.

Example 19. (*Simple Brownian Motion*)

We illustrate the result with the most simple example again, where the approximation errors are zero for both LSMC and portfolio replication. We take the same set-up as in Example 13. Obviously, for the replicating portfolio a perfect fit is achieved with an R^2 of 1. Consequently, the conditional expectation function $g_t(W_t) = W_t$ is also perfectly fit for any $t < T$. For LSMC the goodness of fit depends on the projection error, which is driven by the time gap between t and T . This can be directly seen from the R^2 , which is given by

$$R_{lsmc}^2 = 1 - \frac{(T-t)}{T} = \frac{t}{T}.$$

This is the highest R^2 that can be reached with the LSMC method of approximating the conditional expectation function W_t through regression of W_T on a basis valued at t .

Let us now explore the more general case, in which we allow a nonzero approximation error in both LSMC and portfolio replication. For LSMC we obtain the following R^2

$$\begin{aligned} R_{lsmc}^2 &= 1 - \frac{\mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(a_{0,t}^K(A_t(Z)) + p_{0,t}(A_T(Z)) \right)^2 \right]}{\text{Var}(X)} \\ &= 1 - \frac{\mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(a_{0,t}^K(A_t(Z)) \right)^2 \right] + \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(p_{0,t}(A_T(Z)) \right)^2 \right]}{\text{Var}(X)} \\ &= 1 - \frac{\sum_{k=K+1}^{\infty} \beta_k^2 + \mathbb{E}_{\tilde{\mathbb{P}}} [X^2] - \sum_{k=1}^{\infty} \beta_k^2}{\text{Var}(X)} \\ &= 1 - \frac{\sum_{j=1}^{\infty} \alpha_j^2 - \sum_{k=1}^K \beta_k^2}{\text{Var}(X)}, \end{aligned} \tag{4.27}$$

which is smaller than the R^2 of (4.26) unless $K \rightarrow \infty$, confirming again that (4.26) is the upper limit for R^2 in LSMC. For the R^2 of the least squares regression in portfolio replication we obtain

$$\begin{aligned} R_{RP}^2 &= 1 - \frac{\mathbb{E}_{\mathbb{P}} \left[(a_T^K(A_T(Z)))^2 \right]}{\text{Var}(X)} \\ &= 1 - \frac{\sum_{k=K+1}^{\infty} \alpha_k^2}{\text{Var}(X)}. \end{aligned} \tag{4.28}$$

Clearly the smaller the sum $\sum_{k=K+1}^{\infty} \alpha_k^2$ the higher the R^2 for portfolio replication. Since that sum is driven by the approximation error we see a direct link between the R^2 and the approximation error and can conclude that a higher R^2 indicates a smaller approximation error.

Example 20. (*Exponential function*)

For Example 14 an approximation error is present in both LSMC and portfolio replication. Based on a sample $N = 1000$ we obtain for the LSMC regression an R^2 of 0.077 while for the replicating portfolio we obtain an R^2 of 0.74. If we calculate the (in-sample) mean square error for the fit of both methods to the conditional expectation function $\exp(\sigma W_t + \frac{1}{2}\sigma^2(T-t))$ we obtain comparable results with an (in-sample) mean square error of approximately 0.004 for both methods. From that we can deduce that while both methods yield the same quality in terms of the goodness of fit to the conditional expectation function, the R^2 for LSMC does not reveal this and is misleading.

Note that calculating the upper R^2 limit in LSMC in (4.26) for a particular target function X involves the calculation of the variance of the projection error. For the applications, for which proxy methods such as portfolio replication and LSMC are used, we do not expect that the variance of the projection error is readily available. Without the upper limit judging an R^2 obtained for an LSMC representation

becomes difficult. The R^2 thus does not provide information on how good or bad the estimated representation is. Drawing conclusions on missing risk factors and/or basis terms is not straightforward. This is different for portfolio replication, where the upper limit of R^2 is always 1 indicating a perfect fit. Consequently, we can use R^2 as a simple, but very effective measure for assessing the quality of a replicating portfolio. Due to the direct link between R^2 and the approximation error a_T^K , we can say that the higher the R^2 the smaller the approximation error. Recall that in portfolio replication we have to evaluate the conditional expectation function in a second step by applying the conditional expectation operator to the replicating portfolio. The resulting error in the replicating portfolio proxy to the conditional expectation function $g_{0,t}$ is then $\mathbb{E}_{\tilde{\mathbb{P}}}[a_T^K(A_T(Z))|\mathcal{F}_t]$. By ensuring that a_T^K is small, we also ensure that $\mathbb{E}_{\tilde{\mathbb{P}}}[a_T^K(A_T(Z))|\mathcal{F}_t]$ is small. In that respect we can apply R^2 in portfolio replication as a warning signal for the quality of our proxy, i.e. only replicating portfolios with very high R^2 should be used. As we have seen in this section R^2 cannot be interpreted in the same way in LSMC.

4.3.3 Asymptotic covariance with fixed truncation parameter

Intuitively we expect basis functions valued at time T to be more strongly correlated with the target function X , which is also valued at time T . In contrast to that we expect basis functions valued at time $t < T$ to be less strongly correlated with the target function valued at time T . We have first addressed this in Section 4.3.1 where we have highlighted that in LSMC we deal with noisy regressions due to the time gap in the regressand and the regressors. The analysis of R^2 in Section 4.3.2 furthermore confirms the hypothesis. In this

section we derive the asymptotic covariance matrix for LSMC and portfolio replication for a fixed truncation parameter K . Given a fixed K the asymptotic distribution of $\hat{\boldsymbol{\alpha}}_K$ and $\hat{\boldsymbol{\beta}}_K$, respectively, is derived. Assume that the sampling schemes $((X_1, A_t(Z_1)), \dots, (X_N, A_t(Z_N)))$ and $((X_1, A_T(Z_1)), \dots, (X_N, A_T(Z_N)))$ are such that

$$\begin{aligned} \frac{1}{N} \left((\mathbf{V}_K)^T \mathbf{V}_K \right) &\xrightarrow{\mathbb{P}} C_{lsmc} \text{ and} \\ \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{v}_K(A_t(Z_i)) (a_{0,t}^K(A_t(Z_i)) + p_{0,t}(A_T(Z_i))) &\xrightarrow{d} N(0, \boldsymbol{\Sigma}_{lsmc}) \end{aligned} \quad (4.29)$$

and

$$\begin{aligned} \frac{1}{N} \left((\mathbf{E}_K)^T \mathbf{E}_K \right) &\xrightarrow{\mathbb{P}} C_{RP} \text{ and} \\ \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{e}_K(A_T(Z_i)) a_T^K(A_T(Z_i)) &\xrightarrow{d} N(0, \boldsymbol{\Sigma}_{RP}), \end{aligned} \quad (4.30)$$

where $N(0, \Sigma)$ denotes a normal distribution with mean 0 and covariance matrix Σ , $\xrightarrow{\mathbb{P}}$ denotes convergence in probability and \xrightarrow{d} denotes convergence in distribution. Then by the standard representation of the empirical error of least squares estimators and Slutsky's lemma it follows that

$$\begin{aligned} \sqrt{N} \left(\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K \right) &= \sqrt{N} \left((\mathbf{V}_K)^T \mathbf{V}_K \right)^{-1} (\mathbf{V}_K)^T (\mathbf{a}_{0,t}^K + \mathbf{p}_{0,t}) \\ &\xrightarrow{d} N(0, (C_{lsmc})^{-1} \boldsymbol{\Sigma}_{lsmc} (C_{lsmc})^{-1}), \end{aligned}$$

where $\mathbf{a}_{0,t}^K = (a_{0,t}(A_t(Z_1)), \dots, a_{0,t}(A_t(Z_N)))^T$ and $\mathbf{p}_{0,t} = (p_{0,t}(A_T(Z_1)), \dots, p_{0,t}(A_T(Z_N)))^T$. By the same argument

$$\sqrt{N} \left(\hat{\boldsymbol{\alpha}}_K - \boldsymbol{\alpha}_K \right) = \sqrt{N} \left((\mathbf{E}_K)^T \mathbf{E}_K \right)^{-1} (\mathbf{E}_K)^T \mathbf{a}_T^K$$

$$\xrightarrow{d} N(0, (C_{RP})^{-1} \Sigma_{RP} (C_{RP})^{-1})$$

where $\mathbf{a}_T^K = (a_T^K(A_T(Z_1)), \dots, a_T^K(A_T(Z_N)))^T$. Assume that the data $(X_i, A_t(Z_i))$, $i = 1, \dots, N$, are i.i.d., then by the orthogonality of $g_{0,t}(A_t(Z))$ and $a_{0,t}^K(A_t(Z)) + p_{0,t}(A_T(Z))$ the second part of conditions (4.29) holds and (4.29) holds with $C_{lsmc} = I_K$, where I_K denotes the $K \times K$ identity matrix, and

$$\Sigma_{lsmc} = \mathbb{E}_{\mathbb{F}} \left[(a_{0,t}^K(A_t(Z)) + p_{0,t}(A_T(Z)))^2 \mathbf{v}_K(A_t(Z)) (\mathbf{v}_K(A_t(Z)))^T \right]. \quad (4.31)$$

Similarly, if the data $(X_i, A_T(Z_i))$, $i = 1, \dots, n$, are i.i.d., then (4.30) holds with $C_{RP} = I$ and

$$\Sigma_{RP} = \mathbb{E}_{\mathbb{F}} \left[(a_T^K(A_T(Z)))^2 \mathbf{e}_K(A_T(Z)) (\mathbf{e}_K(A_T(Z)))^T \right]. \quad (4.32)$$

The two asymptotic covariance matrices (4.31) and (4.32) in the i.i.d. case basically differ by the terms $a_{0,t}^K(A_t(Z)) + p_{0,t}(A_T(Z))$ and $a_T^K(A_T(Z))$, because

$$\begin{aligned} \mathbb{E}_{\mathbb{F}} \left[\mathbf{v}_K(A_t(Z)) (\mathbf{v}_K(A_t(Z)))^T \right] &= \mathbb{E}_{\mathbb{F}} \left[\mathbf{e}_K(A_T(Z)) (\mathbf{e}_K(A_T(Z)))^T \right] \\ &= I_K. \end{aligned}$$

We stress that in principle the functions g_T and $g_{0,t}$ are quite different in various aspects, for example they may differ in their dimensionality, so that a general comparison of Σ_{lsmc} and Σ_{RP} may not be feasible. We will come back to the potential differences in the structures of g_T and $g_{0,t}$ in Section 4.4. However, if g_T and $g_{0,t}$ have a similar structure so that $a_{0,t}^K$ and a_T^K also have a similar structure, we expect the asymptotic covariance matrix of the LSMC estimator to be larger than the asymptotic covariance matrix of the replicating portfolio

estimator due to the projection term $p_{0,t}$ in the LSMC asymptotic covariance matrix. Hence, then it should hold that

$$\Sigma_{\text{RP}} \leq \Sigma_{\text{lsmc}} \quad (4.33)$$

meaning, by Loewner's ordering,

$$\Sigma_{\text{diff}} = \Sigma_{\text{lsmc}} - \Sigma_{\text{RP}} \quad (4.34)$$

is a positive semidefinite matrix (see Definition 1.1, Siotani, 1967). Thus, if the approximation errors in LSMC and portfolio replication have a similar structure, then we can expect the variance of the replicating portfolio estimator to be smaller than the variance of the LSMC estimator, meaning that with portfolio replication we can yield a more accurate estimate. We next empirically analyze the property using the same basis for LSMC and portfolio replication given a function where the payoff and the conditional expectation function are similar in their structure.

Example 21. (*Exponential with indicator functions*)

Let the target variable X be the payoff from a geometric Brownian motion at time T on a compact domain,

$$X = e^{-\frac{1}{2}\sigma^2 T + \sigma W(T)}; \quad W(T) \in [-2\sqrt{T}, 2\sqrt{T}]. \quad (4.35)$$

We construct an orthonormal basis on $L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ based on non-overlapping indicator functions. Consider the stochastic risk factor $W(T)$ with probability measure \mathbb{P} . The domain \mathbb{R} is chopped into K intervals, $\{[b_1, b_2), [b_2, b_3), \dots, [b_K, b_{K+1})\}$, such that $Pr(b_k \leq W(T) < b_{k+1}) = 1/K, \forall k = 1, \dots, K$. Define K non-overlap-

ping indicator functions

$$\mathbf{1}_k(W(T)) := \begin{cases} 1 & \text{if } W(T) \in [b_k, b_{k+1}) \\ 0 & \text{otherwise} \end{cases} \quad (4.36)$$

for $k = 1, \dots, K$. By construction the indicator functions are orthogonal. Hence,

$$\mathbb{E}_{\mathbb{P}} [\mathbf{1}_j(W(T))\mathbf{1}_l(W(T))] = \begin{cases} \frac{1}{K} & \text{if } j = l \\ 0 & \text{otherwise.} \end{cases} \quad (4.37)$$

Note that the set of indicator functions $\{\sqrt{K} \mathbf{1}_k(W_T)\}_{k=1}^{\infty}$ is a basis for the Hilbert space (see Theorem 7.8, Hunter, 2011). The approximation to $X = g_T(W(T))$ is then

$$g_T^K(W(T)) = \sqrt{K} \sum_{k=1}^K \alpha_k \mathbf{1}_k(W(T)) \quad (4.38)$$

with

$$\alpha_k = \sqrt{K} \frac{\Phi\left(\frac{b_{k+1}}{\sqrt{T}} - \sigma\sqrt{T}\right) - \Phi\left(\frac{b_k}{\sqrt{T}} - \sigma\sqrt{T}\right)}{\Phi(2) - \Phi(-2)}, \quad (4.39)$$

where $\Phi(\cdot)$ denotes the cumulative standard normal distribution function. From Equation (4.32) the expectations are estimated based on simulations of W_T with sample size 1 000 000 and parameters $\sigma = 0.2$, $T = 10$ and $t = 1$.

In LSMC the target variable to be replicated is the conditional expectation function $g_{0,t}(W(t))$,

$$g_{0,t}(W(t)) = e^{-\frac{1}{2}\sigma^2 t + \sigma W(t)}; \quad W(t) \in [-2\sqrt{t}, 2\sqrt{t}]. \quad (4.40)$$

Analogously to the portfolio replication case we construct a basis of indicator functions for the LSMC problem. The domain \mathbb{R} is chopped into K intervals, $\{[a_1, a_2), [a_2, a_3), \dots, [a_K, a_{K+1})\}$, such that $Pr(a_k \leq W(t) < a_{k+1}) = 1/K, \forall k = 1, \dots, K$. Define K non-overlapping indicator functions

$$\mathbb{1}_k(W(t)) := \begin{cases} 1 & \text{if } W(t) \in [a_k, a_{k+1}) \\ 0 & \text{otherwise} \end{cases} \quad (4.41)$$

for $k = 1, \dots, K$. By construction the indicator functions are orthogonal. The approximation to $g_{0,t}(W(t))$ is then

$$g_{0,t}^K(W(t)) = \sqrt{K} \sum_{k=1}^K \beta_k \mathbb{1}_k(W(t)) \quad (4.42)$$

with

$$\beta_k = \sqrt{K} \frac{\Phi\left(\frac{a_{k+1}}{\sqrt{t}} - \sigma\sqrt{t}\right) - \Phi\left(\frac{a_k}{\sqrt{t}} - \sigma\sqrt{t}\right)}{\Phi(2) - \Phi(-2)}. \quad (4.43)$$

The entries of the LSMC asymptotic covariance matrix in Equation (4.31) are estimated based on simulating 1,000,000 sample paths of the standard Brownian motion from time t to T . Table 4.2 gives the eigenvalues of $\Sigma_{\text{lsmc}} - \Sigma_{\text{RP}}$ for $K = 2, 5, 10$. The eigenvalues for $K = 50, 70$ have also been calculated, but to save space are not included in the table. The results indicate in every case that Σ_{diff} is positive semidefinite.

Table 4.2: Eigenvalues of $\Sigma_{lsmc} - \Sigma_{RP}$ for different K .

	$K = 2$	$K = 5$	$K = 10$
Eigenvalues	0.12374054	0.13279645	0.11529827
	0.09337982	0.12104551	0.11289384
		0.11074061	0.10271642
		0.09129928	0.09111993
		0.06819562	0.08367246
			0.07561172
			0.06906777
			0.06209503
			0.05424688
			0.04384077

4.3.4 Asymptotic measure independence in Portfolio Replication

In both LSMC and portfolio replication we are searching for the coefficients of the basis terms that make up their respective representations. Looking at Equations (4.2) and (4.10) the coefficients depend on the measure $\tilde{\mathbb{P}}$. In many cases it may be desirable to calibrate the representation under a different measure. For example, in order to sufficiently capture the tails of the target function, we may want to simulate more tail values of the underlying risk factors. Changing the measure, however, affects the result for the coefficients, meaning that we may not obtain the correct representation of the target function given a basis. In this section we show that the replicating portfolio method is asymptotically measure independent, but the LSMC result always depends on the chosen calibration measure. We will again see that the cause of this difference between LSMC and portfolio replication is linked to the non-zero projection error in LSMC.

Let us first discuss the portfolio replication case. Let \mathbf{E} be the orthonormal basis under $\tilde{\mathbb{P}}$. Let $\tilde{\mathbb{Q}}$ be a measure equivalent to $\tilde{\mathbb{P}}$. We

first assume that we can perfectly replicate the target payoff function X , meaning that

$$\mathbf{X} = \mathbf{E}\boldsymbol{\alpha}.$$

We want to investigate whether the coefficients $\boldsymbol{\alpha}$ can be found under both $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{Q}}$. Let us first calculate the coefficients under $\tilde{\mathbb{P}}$

$$\begin{aligned}\boldsymbol{\alpha}_{\tilde{\mathbb{P}}} &= \mathbb{E}_{\tilde{\mathbb{P}}} [\mathbf{E}^T \mathbf{X}] \\ &= \mathbb{E}_{\tilde{\mathbb{P}}} [\mathbf{E}^T \mathbf{E} \boldsymbol{\alpha}] \\ &= \mathbb{E}_{\tilde{\mathbb{P}}} [\mathbf{E}^T \mathbf{E}] \boldsymbol{\alpha} \\ &= \boldsymbol{\alpha}\end{aligned}$$

since $\mathbb{E}_{\tilde{\mathbb{P}}} [\mathbf{E}^T \mathbf{E}] = \mathbf{I}$ due to the orthonormality of the basis under $\tilde{\mathbb{P}}$, where \mathbf{I} is the identity matrix. Now, when we change the measure to $\tilde{\mathbb{Q}}$ the basis may not be orthonormal anymore. Hence, the coefficients are calculated as

$$\begin{aligned}\boldsymbol{\alpha}_{\tilde{\mathbb{Q}}} &= (\mathbb{E}_{\tilde{\mathbb{Q}}} [\mathbf{E}^T \mathbf{E}])^{-1} \mathbb{E}_{\tilde{\mathbb{Q}}} [\mathbf{E}^T \mathbf{X}] \\ &= (\mathbb{E}_{\tilde{\mathbb{Q}}} [\mathbf{E}^T \mathbf{E}])^{-1} \mathbb{E}_{\tilde{\mathbb{Q}}} [\mathbf{E}^T \mathbf{E} \boldsymbol{\alpha}] \\ &= (\mathbb{E}_{\tilde{\mathbb{Q}}} [\mathbf{E}^T \mathbf{E}])^{-1} \mathbb{E}_{\tilde{\mathbb{Q}}} [\mathbf{E}^T \mathbf{E}] \boldsymbol{\alpha} \\ &= \boldsymbol{\alpha}.\end{aligned}$$

Thus, when the perfect basis is available, the correct coefficients are obtained independent of the measure. Note that this does actually not depend on the orthonormality property of the basis, i.e. it also holds when \mathbf{E} is not orthonormal under neither $\tilde{\mathbb{P}}$ nor $\tilde{\mathbb{Q}}$. Now let us consider the portfolio replication case, where we have an approximation error,

i.e.

$$\mathbf{X} = \mathbf{E}^K \boldsymbol{\alpha}^K + \mathbf{a}_T^K,$$

where \mathbf{E}^K contains the truncated basis, i.e. K basis terms, and $\boldsymbol{\alpha}^K$ denotes the K true coefficients of the truncated basis terms; \mathbf{a}_T^K denotes the approximation error. We are looking for the coefficients $\boldsymbol{\alpha}^K$, which we again correctly obtain under $\tilde{\mathbb{P}}$,

$$\begin{aligned} \boldsymbol{\alpha}_{\tilde{\mathbb{P}}}^K &= \mathbb{E}_{\tilde{\mathbb{P}}} [(\mathbf{E}^K)^T \mathbf{X}] \\ &= \mathbb{E}_{\tilde{\mathbb{P}}} [(\mathbf{E}^K)^T (\mathbf{E}^K \boldsymbol{\alpha}^K + \mathbf{a}_T^K)] \\ &= \mathbb{E}_{\tilde{\mathbb{P}}} [(\mathbf{E}^K)^T \mathbf{E}^K \boldsymbol{\alpha}^K] + \mathbb{E}_{\tilde{\mathbb{P}}} [(\mathbf{E}^K)^T \mathbf{a}_T^K] \\ &= \boldsymbol{\alpha}^K, \end{aligned}$$

since by the orthonormality of the basis \mathbf{E}^K and \mathbf{a}_T^K are orthogonal and $\mathbb{E}_{\tilde{\mathbb{P}}} [(\mathbf{E}^K)^T \mathbf{E}^K] = \mathbf{I}$. Changing the measure to $\tilde{\mathbb{Q}}$ yields

$$\begin{aligned} \boldsymbol{\alpha}_{\tilde{\mathbb{Q}}}^K &= (\mathbb{E}_{\tilde{\mathbb{Q}}} [(\mathbf{E}^K)^T \mathbf{E}^K])^{-1} \mathbb{E}_{\tilde{\mathbb{Q}}} [(\mathbf{E}^K)^T \mathbf{X}] \\ &= (\mathbb{E}_{\tilde{\mathbb{Q}}} [(\mathbf{E}^K)^T \mathbf{E}^K])^{-1} \left(\mathbb{E}_{\tilde{\mathbb{Q}}} [(\mathbf{E}^K)^T \mathbf{E}^K] \boldsymbol{\alpha}^K \right. \\ &\quad \left. + \mathbb{E}_{\tilde{\mathbb{Q}}} [(\mathbf{E}^K)^T \mathbf{a}_T^K] \right) \\ &= \boldsymbol{\alpha}^K + (\mathbb{E}_{\tilde{\mathbb{Q}}} [(\mathbf{E}^K)^T \mathbf{E}^K])^{-1} \mathbb{E}_{\tilde{\mathbb{Q}}} [(\mathbf{E}^K)^T \mathbf{a}_T^K]. \end{aligned}$$

We see that when there is an approximation error changing the measure does not yield the correct coefficients. Only when $K \rightarrow \infty$, the approximation error converges to zero and $\boldsymbol{\alpha}_{\tilde{\mathbb{Q}}}^K \rightarrow \boldsymbol{\alpha}^K$. Hence, asymptotically the replicating portfolio technique is measure independent.

Let us now investigate the LSMC case. We denote the basis at time t by \mathbf{V} , which is orthonormal under $\tilde{\mathbb{P}}$. Then we can write

$$\mathbf{X} = \mathbf{V}^K \boldsymbol{\beta}^K + \mathbf{a}_{0,t}^K + \mathbf{p}_{0,t},$$

where \mathbf{V}^K denotes the truncated basis, $\mathbf{a}_{0,t}^K$ denotes the approximation error and $\mathbf{p}_{0,t}$ is the projection error. We again first calculate the coefficients under the measure $\tilde{\mathbb{P}}$.

$$\begin{aligned}
 \boldsymbol{\beta}_{\tilde{\mathbb{P}}}^K &= \mathbb{E}_{\tilde{\mathbb{P}}} [(\mathbf{V}^K)^T \mathbf{X}] \\
 &= \mathbb{E}_{\tilde{\mathbb{P}}} [(\mathbf{V}^K)^T (\mathbf{V}^K \boldsymbol{\beta}^K + \mathbf{a}_{0,t}^K + \mathbf{p}_{0,t})] \\
 &= \boldsymbol{\beta}^K + \mathbb{E}_{\tilde{\mathbb{P}}} [(\mathbf{V}^K)^T \mathbf{a}_{0,t}^K] + \mathbb{E}_{\tilde{\mathbb{P}}} [(\mathbf{V}^K)^T \mathbf{p}_{0,t}] \\
 &= \boldsymbol{\beta}^K
 \end{aligned}$$

since the approximation error and the basis terms up to K are orthogonal by construction and the projection error is orthogonal to each basis term at time t . Changing the measure to $\tilde{\mathbb{Q}}$ gives

$$\begin{aligned}
 \boldsymbol{\beta}_{\tilde{\mathbb{Q}}}^K &= (\mathbb{E}_{\tilde{\mathbb{Q}}} [(\mathbf{V}^K)^T \mathbf{V}^K])^{-1} \mathbb{E}_{\tilde{\mathbb{Q}}} [(\mathbf{V}^K)^T \mathbf{X}] \\
 &= (\mathbb{E}_{\tilde{\mathbb{Q}}} [(\mathbf{V}^K)^T \mathbf{V}^K])^{-1} \mathbb{E}_{\tilde{\mathbb{Q}}} [(\mathbf{V}^K)^T (\mathbf{V}^K \boldsymbol{\beta}^K + \mathbf{a}_{0,t}^K + \mathbf{p}_{0,t})] \\
 &= \boldsymbol{\beta}^K + (\mathbb{E}_{\tilde{\mathbb{Q}}} [(\mathbf{V}^K)^T \mathbf{V}^K])^{-1} \left(\mathbb{E}_{\tilde{\mathbb{Q}}} [(\mathbf{V}^K)^T \mathbf{a}_{0,t}^K] \right. \\
 &\quad \left. + \mathbb{E}_{\tilde{\mathbb{Q}}} [(\mathbf{V}^K)^T \mathbf{p}_{0,t}] \right).
 \end{aligned}$$

Even if the approximation error is zero, i.e. $K \rightarrow \infty$ we have

$$\boldsymbol{\beta}_{\tilde{\mathbb{Q}}} = \boldsymbol{\beta} + (\mathbb{E}_{\tilde{\mathbb{Q}}} [(\mathbf{V}^K)^T \mathbf{V}^K])^{-1} \mathbb{E}_{\tilde{\mathbb{Q}}} [(\mathbf{V}^K)^T \mathbf{p}_{0,t}].$$

Thus, even in the limit when the approximation error is zero, the projection error remains and changing the measure affects the coefficients obtained.

Summing up, as the approximation error vanishes, the replicating portfolio constructed with the least squares Monte Carlo method of Section 4.2.2 is perfect regardless of the measure used for calibration. For LSMC the situation is different. Even if the approximation error is zero, the projection error is nonzero since in LSMC $t < T$. Consequently,

even in the limit the LSMC estimator is measure-dependent. We will illustrate this result for LSMC and portfolio replication with several simple examples. Note that we refrain from orthonormalizing the basis as the examples are more intuitive using the non-orthonormalized basis. Nonetheless, we could, of course, orthonormalize these basis terms to be consistent with the presented theory. Moreover, it can easily be shown that the conclusions made above on the measure dependence also hold if a non-orthonormal basis is used.

Example 22. (*Simple Brownian Motion*)

First we again use the very simple Brownian motion case of Example 13, in which both for LSMC and portfolio replication the approximation errors are equal to zero. Recall that for this example we have $X = W_T$ and $\mathbb{E}[X|\mathcal{F}_t] = W_t$. For the LSMC approach the basis is W_t while for the portfolio replication approach the basis is W_T . We write down the following regression equations with constants

$$\begin{aligned}W_T &= \alpha_0 + \alpha_1 W_T \\W_T &= \beta_0 + \beta_1 W_t + p_{0,t}(W_T)\end{aligned}$$

The correct coefficients are $\alpha_0, \beta_0 = 0$ and $\alpha_1, \beta_1 = 1$. Instead of simulating the Brownian motions from the normal distribution, we run the least squares regression based on a sample of size 1000 simulated from the shifted normal distribution with mean $\mu = 5$. Hence, we simulate $W_t = \sqrt{t}Z$ and $W_T = W_t + \sqrt{T-t}Z$ with $Z \sim N(\mu, 1)$. For the portfolio replication approach the change of measure has no effect since the perfect replicating portfolio is still simply the Brownian motion at time T , i.e. W_T . Thus, we obtain the correct coefficients $\alpha_0 = 0$ and $\alpha_1 = 1$. However, the LSMC estimate gives $\hat{\beta}_0 = 15.65$ and $\hat{\beta}_1 = 0.88$. The coefficients make sense considering that the conditional expectation function under the shifted normal measure is

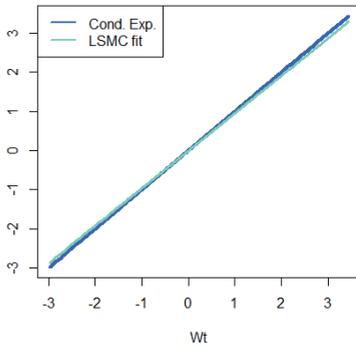


Figure 4.17: LSMC fit with calibration on correct measure (Example 22).

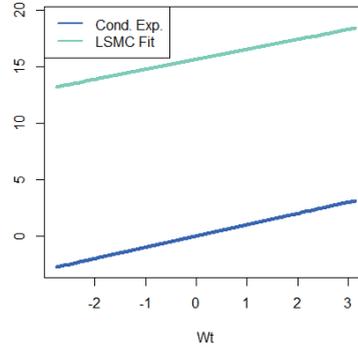


Figure 4.18: LSMC fit with calibration on shifted normal (Example 22).

now $g_t(W_t) = \sqrt{T-t}\mu + W_t$. After all though, the goal is to achieve a fit to the conditional expectation function under the original measure. Clearly, in LSMC we cannot easily switch to a different measure for calibrating the fitting function. We evaluate the out-of-sample fit of both regression approaches based on a sample that has not been used for calibration. Since with portfolio replication the correct coefficients were obtained, the out-of-sample fit is perfect. However, for LSMC the coefficients are biased due to the calibration based on the shifted normal distribution. Figure 4.17 gives the LSMC result calibrated based on the normal distribution. Figure 4.18 shows the fit for the LSMC estimation calibrated based on the shifted normal distribution.

Example 23. (*Exponential function*)

Let us take Example 14, but consider Z in $W_t = \sqrt{t}Z$ and $W_T = \sqrt{T}Z$ to be simulated from a truncated normal on $[-2, 2]$. We compare the goodness of fit for both LSMC and portfolio replication when calibrating under the truncated normal distribution and when calibrating under

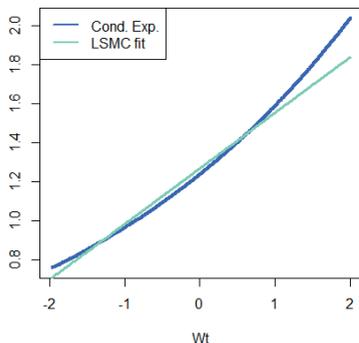


Figure 4.19: LSMC fit with calibration on correct measure (Example 23).

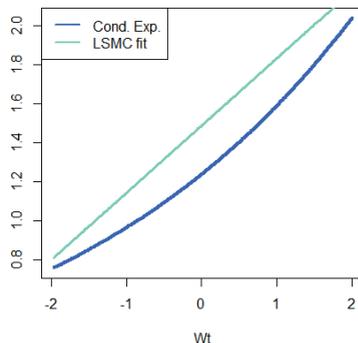


Figure 4.20: LSMC fit with calibration on uniform (Example 23).

the uniform on $[-2, 2]$. Note that for the case at hand in both LSMC and portfolio replication an approximation error is present.

Figures 4.19 and 4.20 compare the LSMC out-of-sample goodness of fit for the calibration on the true measure and on the uniform. Figures 4.21 and 4.22 are the analog for portfolio replication. We clearly see that for the example at hand both LSMC and portfolio replication are measure-dependent, meaning that the coefficient estimates depend on the measure we use for calibration. While we have already seen in the previous example that LSMC is measure-dependent, the measure dependence here for the replicating portfolio results from the approximation error.

Example 24. (*Artificial portfolio*)

In this example we construct a payoff function from a set of calls and puts. We define the target payoff function as

$$X = 100 - 2(K_1 - S(T))^+ + (S(T) - K_2)^+ - 2(S(T) - K_3)^+ \\ + (S(T) - K_4)^+ + 0.5(S(T) - K_5)^+ - 0.5(S(T) - K_6)^+$$

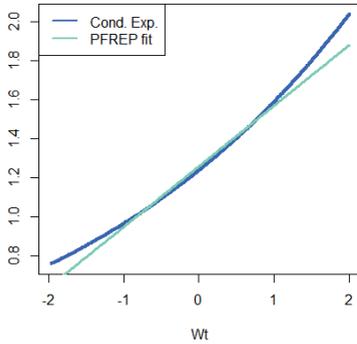


Figure 4.21: Replication fit with calibration on correct measure (Example 23).

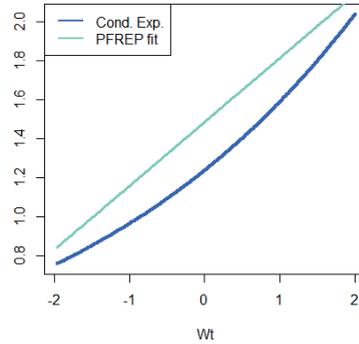


Figure 4.22: Replication fit with calibration on uniform (Example 23).

with K_1 to K_6 given by $\{20, 50, 100, 150, 200, 205\}$. Note that we have deliberately constructed the target function such that it changes in the tails at, for example, $S_T = 20$. This is also the reason why the strikes in this example differ from Example 15. We plot the payoff function for X in Figure 4.23. The underlying stock, denoted by $S(T)$ at time T , is modeled as a geometric Brownian motion

$$\begin{aligned}
 S(T) &= S_0 \exp \left(\left(\mu - \frac{1}{2} \sigma^2 \right) T + \sigma W(T) \right) \\
 &= S(t) \exp \left(\left(\mu - \frac{1}{2} \sigma^2 \right) (T - t) + \sigma (W(T) - W(t)) \right),
 \end{aligned}
 \tag{4.44}$$

where $W(\cdot)$ is a standard Brownian motion. The parameters are given in Table 4.1. We assume the Black-Scholes model and thus obtain the conditional expectation function $g_t(S(t))$ by pricing the calls and puts in X using the Black-Scholes formula. Note that we omit the subscript “0” in g_t as the conditional expectation function is known for the case

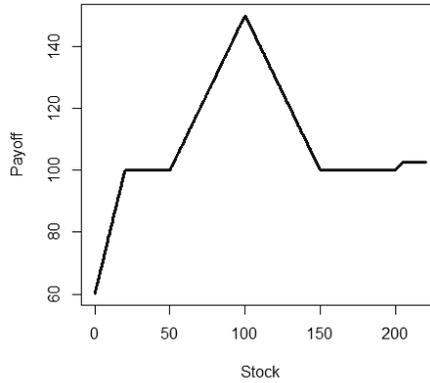


Figure 4.23: Payoff function of X (Example 24).

at hand. Ultimately with LSMC and portfolio replication we want to obtain an approximation of the Black-Scholes price of X .

As basis we choose eight terms consisting of a constant (zero-coupon bond), the underlying stock and a series of puts on the underlying stock with strikes $\{18, 48, 98, 148, 198, 203\}$. We want to investigate the measure dependence for both LSMC and portfolio replication using different measures for calibration. We consider five different calibration scenario sets, each of size N , which contain the paths for $\{S(t), S(T)\}$, based on the real-world probability measure \mathbb{P} , the risk-neutral measure \mathbb{Q} and the uniform measure. While for the calibration of the replicating portfolio we only need the values $S(T)$, for the calibration of the LSMC representation we require both. Under \mathbb{P} the stock $S(T)$ is modeled as in Equation (4.44). Changing to the equivalent measure \mathbb{Q} $S(T)$ is modeled as

$$S(s) = S_0 \exp \left(\left(r - \frac{1}{2} \sigma^2 \right) T + \sigma W(T) \right)$$

Table 4.3: Calibration sets for Example 24

Set 1: \mathbb{P}	N values generated from (4.44) with $S_0 = 100$.
Set 2: \mathbb{Q}	N values generated from (4.45) with $S_0 = 100$.
Set 3: \mathbb{P} mixed	$N - 400$ values generated from (4.44) with $S_0 = 100$ plus 200 values per $S_0 = 20$ and $S_0 = 150$.
Set 4: \mathbb{Q} mixed	$N - 400$ values generated from (4.45) with $S_0 = 100$ plus 200 values per $S_0 = 20$ and $S_0 = 150$.
Set 5: uniform	$S(t) \in [0, 250]$, $S(T) = S(t) * [0, 1.5]$.

$$= S(t) \exp \left(\left(r - \frac{1}{2} \sigma^2 \right) (T - t) + \sigma (W(T) - W(t)) \right), \quad (4.45)$$

where r is the risk-free rate. Under the uniform we simply simulate the stock values from the uniform. The sets are specified in Table 4.3. For sets one to four the same random numbers for $W(T) - W(t) = \sqrt{T - t} Z$, $Z \sim N(0, 1)$ are used to ensure that the difference in the sample truly comes from the difference between the measures \mathbb{P} and \mathbb{Q} . Set five is constructed such that the range on which the target function X varies the most is sufficiently captured. Note that with set five the assumption on measure equivalence is violated as the measure in set five has a different domain than \mathbb{P} and \mathbb{Q} . Even when violating this assumption, we will see that set five is helpful for our testing purposes. We assess the quality of fit based on an m -sized sample of paths for $S(t)$ and $S(T)$ that is sufficiently diverse to capture the range of values, on which X and $g_t(S(t))$ vary the most.

As an almost perfect basis is used for both the construction of the LSMC representation and the replicating portfolio the approximation error is small. Therefore, we expect the replicating portfolio approach to be rather measure-independent. As in LSMC a projection error

Table 4.4: Results for Example 24

		Fit to X		Fit to $g_t(S(t))$	
		MSE	R^2 ⁶	MSE	R^2 ⁷
RP	Set 1	58.49935	0.8203814	57.9955	0.6385629
	Set 2	18.26597	0.9439155	16.60685	0.9119326
	Set 3	1.473218	0.9954766	0.1499509	0.9992251
	Set 4	1.480922	0.9954529	0.09592363	0.9993492
	Set 5	1.905271	0.99415	0.2598456	0.9983761
LSMC	Set 1			3.73E+18	0.2276543
	Set 2			5.55E+16	0.2276543
	Set 3			9.621861	0.9645457
	Set 4			0.5070824	0.9964742
	Set 5			250.848	0.6685076

is additionally present we expect to see measure-dependence when calibrating the LSMC representation under different measures.

Table 4.4 summarizes the results for the out-of-sample MSE and R^2 for both LSMC and portfolio replication. The out-of-sample R^2 is here calculated as the R^2 from regressing the fitted function \hat{g}_t^K from LSMC and portfolio replication against the true function g_t . Note that for portfolio replication we additionally provide the measures for the goodness of fit to the payoff function X . For portfolio replication the best results are attained when the calibration set is sufficiently diverse to capture the full range, on which the target function varies the most. Therefore, a comparably good fit is achieved under scenario sets three to five. The resulting out-of-sample fit when using calibration set five, for example, is illustrated in Figure 4.25.

⁶For consistency (see the following footnote) calculated as the R^2 from regressing the fitted function \hat{g}_T^K against the true function g_T .

⁷Calculated as the R^2 from regressing the fitted function \hat{g}_t^K from LSMC and portfolio replication against the true function g_t .

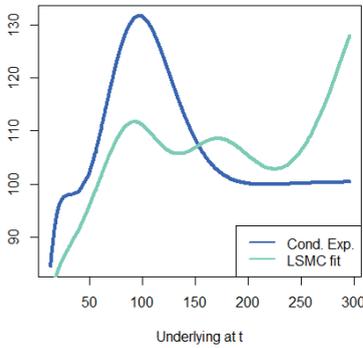


Figure 4.24: LSMC fit with calibration on uniform (Example 24).

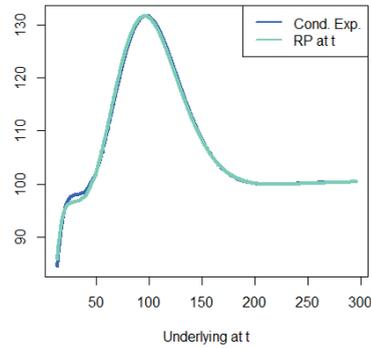


Figure 4.25: Replication fit with calibration on uniform (Example 24).

In LSMC the solution is measure-dependent. Clearly, calibrating under set five leads to a bad LSMC result (see turquoise fitting function in Figure 4.24). Nonetheless, a diverse scenario set is required in order to capture the tail behavior of the target function. Set four works best for the LSMC calibration. It is based on the risk-neutral measure, with which the conditional expectation function is calculated, and contains shock scenarios, which makes it more diverse compared to set two. Set three and four both perform better than set two although set three is based on measure \mathbb{P} . This is due to the fact that set three is much more diverse than set two and this information is needed to calibrate the function well in the tails. However, set four clearly outperforms set three in LSMC while in the replicating portfolio approach set three and four yield almost equal results. Set three and four are very similar in their structure as the same random numbers have been used and the only difference is their growth rate μ and r . But the coefficients in LSMC are measure-dependent. From Figures 4.26 and 4.27 it can be seen that the LSMC solution (illustrated by the turquoise line) differs

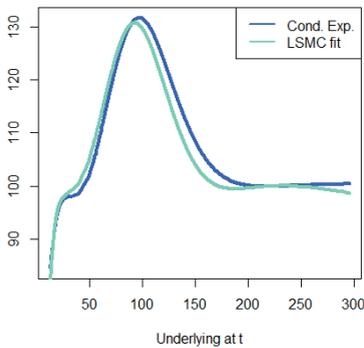


Figure 4.26: LSMC fit with calibration on set three (Example 24).

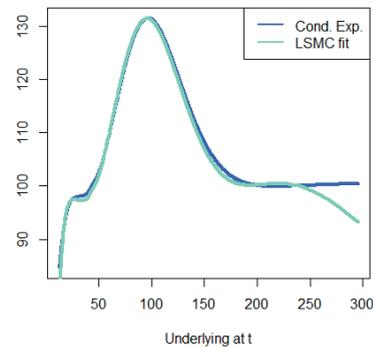


Figure 4.27: LSMC fit with calibration on set four (Example 24).

when calibrating under set three and four, respectively. For portfolio replication almost the same portfolios are achieved with sets three and four (see Figures 4.28 and 4.29).

We note that the last example is quite artificial in that we exactly know the range on which the target function varies and could construct the calibration and out-of-sample scenario sets accordingly. In practice this information is mostly not available. In that case we should decide for each risk factor on the range that is considered relevant. The calibration and out-of-sample scenario sets should then be created to sufficiently cover that range.

The following conclusions are drawn from the analysis and the examples.

- For LSMC use a sufficiently diverse calibration set based on the measure under which the conditional expectation function is calculated.

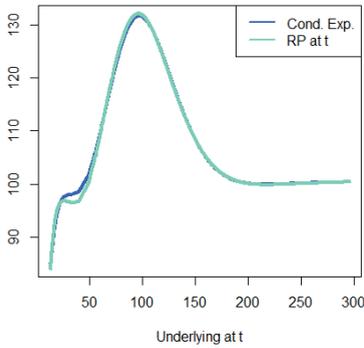


Figure 4.28: Replication fit with calibration on set three (Example 24).

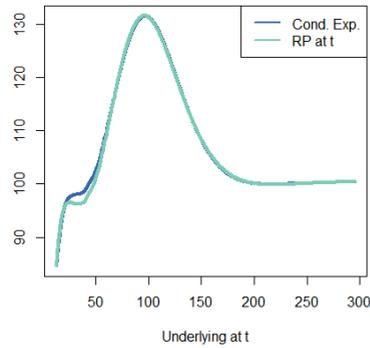


Figure 4.29: Replication fit with calibration on set four (Example 24).

- With portfolio replication there is much more flexibility in choosing a measure when we expect the approximation error to be rather small. The recommendation is then to use a diverse set that sufficiently covers the relevant range. Calibrations using the uniform distribution have shown good results. The measure for calibrating the replicating portfolio may thus differ from the measure under which we want to find the conditional expectation function.

4.3.5 Asymptotic convergence

In Section 4.2 we have given the mathematical model for LSMC and portfolio replication. We have seen that given a basis on the relevant risk factors of the respective target functions X and $g_{0,t}$ a perfect representation exists. However, the perfect representation may involve infinitely many basis terms, which complicates the problem of estimating the coefficients of the basis terms in finite samples. Therefore, the infinite representations are truncated to finite representations based on

a finite number of K basis terms. In order to analyze the asymptotic convergence to the true result both the truncation parameter K and the sample size N must grow simultaneously. In Beutner et al. (2013) the asymptotic convergence rate of LSMC Regress-Later, i.e. portfolio replication, is derived and compared to the asymptotic convergence rate of LSMC Regress-Now, i.e. LSMC in the terminology of this paper (see also Newey (1997) and Stentoft (2004)). In this section we briefly repeat the asymptotic convergence theorems and comment on the difference in the convergence rates for LSMC and portfolio replication. We refer to Beutner et al. (2013) for details and the proofs.

Let us first give the asymptotic convergence result for portfolio replication. Two assumptions are required.

Assumption 4.1. There are $\gamma_{RP} > 0$, α_K s.t.

$$\begin{aligned} & \sqrt{\mathbb{E}_{\tilde{\mathbb{P}}} [(g_T(A_T(Z)) - (\alpha_K)^T \mathbf{e}_K(A_T(Z)))^4]} \\ &= \sqrt{\int_{\mathbb{R}^\ell} (g_T(u) - (\alpha_K)^T \mathbf{e}_K(u))^4 d\tilde{\mathbb{P}}^{A_T(Z)}(u)} \\ &= \sqrt{\int_{\mathbb{R}^\ell} a_T^K(u)^4 d\tilde{\mathbb{P}}^{A_T(Z)}(u)} = O(K^{-\gamma_{RP}}). \end{aligned}$$

Assumption 4.1 controls the convergence of the approximation error.

Assumption 4.2. $((X_1, A_T(Z_1)), \dots, (X_N, A_T(Z_N)))$ are i.i.d.

Moreover, we define

$$\tilde{h}^{RP}(N, K) := \frac{1}{N} \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left((\mathbf{e}_K(A_T(Z)))^T \mathbf{e}_K(A_T(Z)) \right)^2 \right].$$

Notice that $\tilde{h}^{RP}(N, K)$ controls the growth rate of the truncation parameter K in relation to the sample size N . Intuitively it is clear that such a growth rate is required in order to ensure that the sample

size is sufficiently large to estimate a certain number of parameters. Now, we can give the theorem on the asymptotic convergence rate of the portfolio replication method.

Theorem 4.1. *Let Assumptions 4.1 and 4.2 be satisfied. Additionally, assume that there is a sequence $\mathcal{K} : \mathbb{N} \rightarrow \mathbb{N}$ such that*

$$\tilde{h}^{RP}(N, \mathcal{K}(N)) \rightarrow 0 \text{ as } N \rightarrow \infty. \quad (4.46)$$

Then

$$\mathbb{E}_{\tilde{\mathbb{P}}} \left[\left(X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) \right)^2 \right] = O_{\tilde{\mathbb{P}}} \left(\mathcal{K}(N)^{-\gamma_{RP}} \right).$$

Proof. See Beutner et al. (2013). □

Next, we present the asymptotic convergence theory for LSMC as stated in Beutner et al. (2013). We again require two assumptions.

Assumption 4.3. There are $\gamma_{lsmc} > 0$, β_K s.t.

$$\begin{aligned} & \sqrt{\mathbb{E}_{\tilde{\mathbb{P}}} \left[(g_{0,t}(A_t(Z)) - (\beta_K)^T \mathbf{v}_K(A_t(Z)))^4 \right]} \\ &= \sqrt{\int_{\mathbb{R}^s} (g_{0,t}(u) - (\beta_K)^T \mathbf{v}_K(u))^4 d\tilde{\mathbb{P}}^{A_t(Z)}(u)} \\ &= \sqrt{\int_{\mathbb{R}^s} a_{0,t}^K(u)^4 d\tilde{\mathbb{P}}^{A_t(Z)}(u)} = O(K^{-\gamma_{lsmc}}). \end{aligned}$$

Assumption 4.4. $((X_1, A_t(Z_1)), \dots, (X_N, A_t(Z_N)))$ are i.i.d. and $\mathbb{E}_{\tilde{\mathbb{P}}} \left[(p_{0,t}(A_T(Z)))^2 | A_t(Z) \right] = \sigma^2$.

Similarly as in portfolio replication we also define

$$\tilde{h}^{lsmc}(N, K) := \frac{1}{N} \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left((\mathbf{v}_K(A_t(Z)))^T \mathbf{v}_K(A_t(Z)) \right)^2 \right],$$

which controls the growth rate of K in relation to N . We can now state the theorem.

Theorem 4.2. *Let Assumptions 4.3 and 4.4 be satisfied. Additionally, assume that there is a sequence $\mathcal{K} : \mathbb{N} \rightarrow \mathbb{N}$ such that*

$$\tilde{h}^{lsmc}(N, \mathcal{K}(N)) \rightarrow 0 \text{ as } N \rightarrow \infty. \quad (4.47)$$

Then

$$\mathbb{E}_{\tilde{\mathbb{P}}}\left[\left(g_{0,t}(A_t(Z)) - \hat{g}_{0,t}^{\mathcal{K}(N)}(A_t(Z))\right)^2\right] = O_{\tilde{\mathbb{P}}}\left(\frac{\mathcal{K}(N)}{N} + \mathcal{K}(N)^{-\gamma_{lsmc}}\right). \quad (4.48)$$

Proof. See Beutner et al. (2013). □

The difference in the convergence rate of LSMC and portfolio replication depends on γ_{RP} and γ_{lsmc} . Moreover, the LSMC convergence rate additionally contains the term $\mathcal{K}(N)/N$, which is not present in portfolio replication. In Beutner et al. (2013) it is shown that this additional term in LSMC is driven by its non-zero projection error. The absence of the term $\mathcal{K}(N)/N$ in the mean-square error of portfolio replication makes it plausible that the replicating portfolio estimator may potentially converge faster than the LSMC estimator. We deliberately state here “potentially” as the ultimate convergence rate depends on the γ_{lsmc} and γ_{RP} which are problem-dependent. In particular, the choice of basis plays an important role in the determination of γ_{lsmc} and γ_{RP} . However, the LSMC convergence rate can never be faster than N^{-1} . This follows simply from the fact that the best we can hope for is that $g_{0,t}$ is contained in the span of finitely many basis functions. Then the approximation error vanishes and we are left with the rate N^{-1} . In contrast, in portfolio replication if Condition (4.46) is fulfilled

with $\mathcal{K}(N) \propto N^a$ for some $0 < a < 1$, then the convergence rate for the replicating portfolio equals $N^{-a \gamma_{RP}}$. We can see that for the right combination of a and γ_{RP} it is possible to achieve a convergence rate that is even faster than N^{-1} .

We want to remark on one further point. The discussed general convergence rates pertain to convergence to different functions. While in LSMC the convergence rate pertains to convergence to the unknown conditional expectation function $g_{0,t}(A_t(Z))$, the convergence rate for replicating portfolios pertains to convergence to the true payoff function X . Ultimately in the context of Solvency II insurers are interested in the time t value of its liabilities under different scenarios for the underlying risk drivers. While in LSMC we directly have this, in portfolio replication we achieve the approximation to the conditional expectation function by applying the conditional expectation operator to the estimated payoff function, $\hat{g}_T^{\mathcal{K}(N)}$. We can show that the ultimate estimator given by the conditional expectation of the estimator of X does not converge slower than at the rate derived for the convergence of $\hat{g}_T^{\mathcal{K}(N)}(A_T(Z))$. More explicitly we have

$$\begin{aligned} & \mathbb{E}_{\hat{\mathbb{P}}} \left[\left(\mathbb{E}_{\hat{\mathbb{P}}} [X | \mathcal{F}_t] - \mathbb{E}_{\hat{\mathbb{P}}} \left[\hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) | \mathcal{F}_t \right] \right)^2 \right] \\ &= \mathbb{E}_{\hat{\mathbb{P}}} \left[\left(\mathbb{E}_{\hat{\mathbb{P}}} \left[X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) | \mathcal{F}_t \right] \right)^2 \right] \\ &\leq \mathbb{E}_{\hat{\mathbb{P}}} \left[\mathbb{E}_{\hat{\mathbb{P}}} \left[\left(X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) \right)^2 | \mathcal{F}_t \right] \right] \\ &= \mathbb{E}_{\hat{\mathbb{P}}} \left[\left(X - \hat{g}_T^{\mathcal{K}(N)}(A_T(Z)) \right)^2 \right], \end{aligned}$$

where the first inequality follows from Jensen's inequality and the last equality uses the projection law of expectations.

4.4 Path-dependent and high-dimensional target functions

By now we have discussed several aspects of portfolio replication and LSMC, which have highlighted some of the advantages of one method over the other. What we have not addressed so far is that the problems in portfolio replication and LSMC may differ very much in nature. In portfolio replication the initial objective is to find the representation that best mirrors the payoff function. From that the representation of the conditional expectation function is derived. In LSMC only the conditional expectation function is approximated. Now, in many cases, the conditional expectation function differs in its structure, smoothness and dimensionality from the payoff function, where with smoothness we refer to the differentiability of the function. Particularly, in life insurance we may expect the payoff function to exhibit multiple kinks due to options and guarantees. Moreover, life insurance policies are often strongly path-dependent, which affects the dimensionality of the problem. The conditional expectation function typically “smooths” the payoff function in terms of its differentiability, but also lowers its dimensionality compared to a path-dependent kinked payoff function. The difference in the structure, smoothness and dimensionality of the target function to be approximated significantly affects the feasibility of the LSMC and portfolio replication method in practice. In this section we highlight this point by means of several examples. We will see that the major challenge in portfolio replication compared to LSMC pertains to the replication of path-dependent payoff functions.

Finding either an LSMC representation or a replicating portfolio for a particular target function X with conditional expectation function $\mathbb{E}_{\mathbb{P}}[X|\mathcal{F}_t]$ requires two important steps before calibration.

1. Identification of all risk factors that drive the target function, summarized by $A_t(Z)$ and $A_T(Z)$, respectively.
2. Choosing a basis built on $A_t(Z)$ in LSMC and choosing a basis built on $A_T(Z)$ in portfolio replication.

We will elaborate on these two in the remainder of this section. The first step in LSMC and portfolio replication is the identification of all risk factors that drive the target function, for which a basis representation shall be found. In portfolio replication this means that the risk factors of g_T must be identified, while in LSMC the risk factors driving $g_{0,t}$ must be determined. The complexity of finding the LSMC or replicating portfolio solution highly depends on the number and type of risk drivers underlying the target function. Recall from Section 4.2 that the dimensionality of $A_t(Z)$ and $A_T(Z)$ is denoted by ℓ_t and ℓ_T , respectively, which we view as an indicator for the complexity of the problem. The following examples illustrate the identification of $A_t(Z)$ and $A_T(Z)$, respectively (see also Beutner et al. (2013, 2015)).

Example 25. (*Asian Option*)

Let Z be one-dimensional and consider a discrete Asian option on a stock with

$$X = \max \left(\frac{1}{T} \sum_{s=1}^T Z_1(s) - K, 0 \right),$$

where K is the strike price. The payoff X of the contingent claim depends on all past stock values prior to maturity. $A_T(Z)$ must now comprise all the information of the underlying driver such that X is specified. We may choose $A_T(Z)$ as the time average over the past stock values, which suffices to calculate the payoff X . Then X does only depend on $\sum_{s=1}^T Z_1(s)$. Thus, $A_T(f) = \left(\sum_{s=1}^T f(s) \right)$ for every function $f \in \mathbb{D}[0, T]$ and therefore $\ell_T = 1$. Alternatively, we may also

choose to take into account the value of $Z_1(s)$ at each time point, i.e. $A_T(f) = (f(1), \dots, f(T))$ leading to $\ell_T = T$.

The conditional expectation function at time $t < T$, $\mathbb{E}_Q[X(T)|\mathcal{F}_t]$, only depends on $\sum_{s=1}^t Z_1(s)$ and $Z_1(t)$. Hence, $\ell_t = 2$.

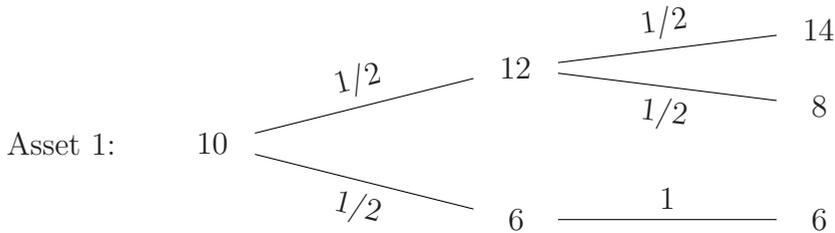
Example 26. (*European Basket Option*)

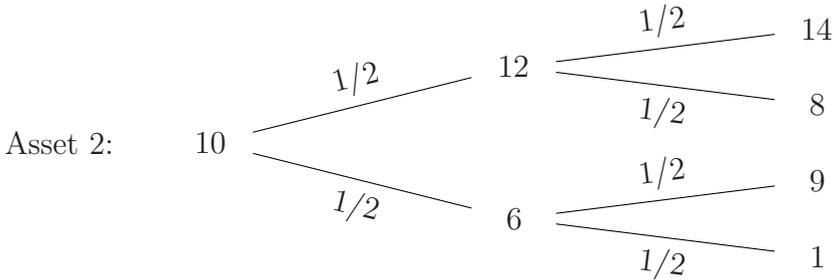
Consider a d -dimensional basket option of the type

$$X = \max\left(\sum_{i=1}^d Z_i(T) - K, 0\right),$$

where K is the strike price. In order to identify X it suffices to take $A_T(f) = \sum_{i=1}^d f_i(T)$ for every function $f \in \mathbb{D}[0, T]$ and therefore $\ell_T = 1$. Alternatively, we could also take $A_T(f) = f(t)$ for every function $f \in \mathbb{D}[0, T]$ leading to $\ell_T = d$.

Now, let us take a look at the conditional expectation function. In general $\mathbb{E}_{\mathbb{P}}[X(T)|\mathcal{F}_t]$ depends on $\mathbf{Z}(t) = (Z_1(t), \dots, Z_d(t))$ and not only on $\sum_{i=1}^d Z_i(t)$. Then $A_t(f) = f(t)$ for every function $f \in \mathbb{D}[0, t]$ and therefore $\ell_t = d$. We give an example that shows our claim. Consider two assets $Z_1(t)$ and $Z_2(t)$ that move across the time steps $t = 0, 1, 2$ as outlined in the trees below.





Take $X = (Z_1(2) + Z_2(2) - K)^+$ with $K = 10$. We are interested in the conditional expectation at time $t = 1$, i.e. $\mathbb{E}_{\mathbb{P}}[X|\mathcal{F}_1]$, for which we obtain the following results

$$\mathbb{E}_{\mathbb{P}}[X|Z_1(1) = 12, Z_2(1) = 12] = 12$$

$$\mathbb{E}_{\mathbb{P}}[X|Z_1(1) = 12, Z_2(1) = 6] = 6.25$$

$$\mathbb{E}_{\mathbb{P}}[X|Z_1(1) = 6, Z_2(1) = 12] = 7$$

$$\mathbb{E}_{\mathbb{P}}[X|Z_1(1) = 6, Z_2(1) = 6] = 2.5.$$

We immediately see that knowing the sum $Z_1(1) + Z_2(1)$ at time $t = 1$ does not suffice to uniquely determine the conditional expectation at time $t = 1$. In particular, for $Z_1(1) + Z_2(1) = 18$ the conditional expectation can either be 6.25 or 7. The European basket option example shows that while for LSMC the problem is d -dimensional, the replicating portfolio problem potentially is only one-dimensional.

Example 27. (*Profit-sharing contract*)

Consider a profit-sharing contract, in which on a yearly basis interest is credited to the policyholder's account. A minimum crediting rate is guaranteed and additional profit is shared depending on the specification of the bonus credited. Then the contingent claim X at maturity

T is given by

$$X = L_0 \prod_{s=1}^T (1 + r_G + r_B(s)), \quad (4.49)$$

where L_0 is the initial value of the policy, r_G denotes the minimum guarantee rate and $r_B(s)$ is the bonus credited at time s . Let $r_B(s) = (r_A(s) - r_G)^+$, where $r_A(s)$ denotes the time s return of some reference portfolio. Straightforwardly, we can define $A_T(f) = \prod_{s=1}^T (1 + r_G + f(s))$ for every function $f \in \mathbb{D}[0, T]$ and therefore $\ell_T = 1$. As a result the dimensionality of the problem is only one, but the specification of $A_T(Z)$ is complex. In that respect, we can also specify $A_T(f) = (f(1), \dots, f(T))$ and therefore $\ell_T = T$.

The conditional expectation function $\mathbb{E}_{\tilde{\mathbb{P}}}[X|\mathcal{F}_t]$ in general depends on $(\prod_{s=1}^t (1 + r_G + f(s)))$ and $r_B(t)$. Therefore $\ell_t = 2$.

The previous examples have shown that for the same problem statement the complexity of the LSMC and portfolio replication method in terms of the dimensionality of the problem may be quite different. Moreover, for the replicating portfolio technique we have stressed that in principle different $A_T(Z)$ can be constructed⁸. For a path-dependent insurance policy we can either choose the state vector such that it captures the path-dependency or by including each element on the path. There is, however, a major trade-off in choosing a lower-dimensional $A_T(Z)$ over a higher-dimensional $A_T(Z)$, which we want to point out next.

Given $A_t(Z)$ and $A_T(Z)$ are identified a suitable basis must be chosen, which is constructed on the underlying risk factors, i.e. on $A_t(Z)$ and $A_T(Z)$, respectively. In the replicating portfolio problem we

⁸While theoretically the same holds for $A_t(Z)$, we omit this as for $A_t(Z)$ we may always choose the lowest dimensionality since we do not face the pricing dilemma of the basis.

moreover require basis functions, for which the conditional expectation under the relevant measure can be fairly easy and quickly determined, preferably even closed-form. For the pure replication of the payoff function X the path-dependency of X may well be captured by either choice of $A_T(Z)$. However, once the conditional expectation of X should be obtained by applying the conditional expectation operator to the basis on the path-dependent $A_T(Z)$, the original dilemma of valuing X at time t is transferred to the problem of valuing the basis. Hence, the more complex the underlying risk factor $A_T(Z)$ the more difficult it will be to obtain a closed-form solution to the conditional expectation of that basis. Choosing a lower-dimensional, but more complex $A_T(Z)$ may therefore complicate the easiness to determine the time t value of the basis built on it. To that end, using vanilla-style basis functions, i.e. functions on path-independent risk drivers, to replicate path-dependent insurance claims has the disadvantage of producing a high dimensional $A_T(Z)$, but the advantage that the conditional expectation of the basis is easily available.

The LSMC method offers here two advantages over the portfolio replication method. First, its basis must not be valued under the conditional expectation operator, meaning that a complex structure for $A_t(Z)$ triggers no successive difficulties. Therefore, a low-dimensional $A_t(Z)$ with complex (path-dependent) structure can always be chosen. Second, as Example 27 has shown, $A_t(Z)$ is potentially lower-dimensional than $A_T(Z)$ if a composite (but low-dimensional) $A_T(Z)$ results in a too complex valuation of the basis built on it. As the discussion highlights finding a basis is a much easier task in LSMC than in portfolio replication.

Let us now consider the construction of a multivariate basis and show why the dimensionality ℓ_t and ℓ_T , respectively, matter. The linear sieve approximation to multivariate contingent claims is ob-

tained analogously to the univariate representation by constructing a tensor product space as described in Chen (2007). Accordingly, the multivariate orthonormal basis is constructed by the tensor product of the respective univariate basis. While the basis is still countable, it is much more elaborate. Truncating the basis representation at K in the univariate case would give K^ℓ basis terms in the multivariate case of dimension ℓ . Thus, the higher-dimensional $A_t(Z)$ in LSMC and $A_T(Z)$ in portfolio replication, the larger the basis. The curse of dimensionality problem quickly dominates. Consider the profit-sharing policy contract of Example 27. For a life insurance policy the terminal time point typically lies far in the future, say 30-60 years from now. Consider $T = 30$ and $K = 5$ basis terms per dimension. For portfolio replication $A_T(Z)$ is 30-dimensional resulting in $K^{\ell_T} = 5^{30} \approx 9.31 \times 10^{20}$ basis terms. In order to estimate such a tremendous number of coefficients an immense sample size is required. Hence, in terms of the simulation effort the problem becomes infeasible. Compare that to LSMC where $\ell_t = 2$ leading with $K = 5$ per dimension to $K^{\ell_t} = 5^2 = 25$ basis terms.

When it comes to path-dependent target functions the lower-dimensionality of the conditional expectation function and the indifference for the basis to be easily valued under the conditional expectation operator seems to give LSMC a competitive edge over portfolio replication. On the other hand the curse of dimensionality in portfolio replication hits in when $A_T(Z)$ is chosen such that each of its components is path-independent and the basis is built as the tensor product of the univariate basis. Therefore, solutions for portfolio replication may be found by deviating from the strict framework of building a basis. In the next example we discuss the construction of a replicating portfolio for a common path-dependent insurance contract.

Example 28. (*Grosen-Jorgensen profit-sharing contract*)

In Example 27 a general profit-sharing contract is discussed. A well-

known variant of profit-sharing contracts is the insurance contingent claim modelled in Grosen and Jørgensen (2000). Here the bonus return depends on the performance of the insurer's underlying asset portfolio. The contract's payoff at time T is as in (4.49) with the yearly bonus rate $r_B(s)$ defined as

$$r_B(s) = \max \left(0, \delta \left(\frac{Z(s-1)}{L(s-1)} - (1 + \lambda) \right) - r_G \right),$$

where $Z(s)$ gives the underlying's stochastic asset value at time s and $L(s)$ gives the value of the liabilities at time s . Note that from Equation (4.49) the liability value at time s is recursively calculated as

$$L(s) = L(s-1)(1 + r_G + r_B(s)).$$

Then, λ defines a buffer ratio and δ the fraction of the excess return that is shared with the policyholder. Clearly, the value of the liabilities at maturity depends on the performance of the underlying asset portfolio over time. Let us consider path-dependent basis functions on the yearly return of the asset portfolio, which captures much of the path-dependent dynamics of the Grosen-Jørgensen payoff, but is still different in its structure.

Consider a sequence of generalized Asian options on the asset process $Z(s)$ as basis, where the k^{th} basis is defined as follows.

$$e_k(\mathbf{Z}) = \max(0, \mathbf{a}_k^T \mathbf{Z} - d_k), \quad k = 1, \dots, K, \quad (4.50)$$

where $\mathbf{Z} = (Z(1), \dots, Z(T-1))^T$ refers to the underlying asset process over time, $\mathbf{a}_k = (a_{k,1}, \dots, a_{k,T-1})^T$ is a series of coefficients for the calculation of the weighted average and d_k is the strike. Moreover, for

the first basis term we take

$$e_0(\mathbf{Z}) = \mathbf{a}_0^T \mathbf{Z}$$

with $\mathbf{a}_k = (a_{0,1}, \dots, a_{0,T-1})^T$ a $(T - 1)$ -vector of coefficients. The parameters $\{\mathbf{a}_k, d_k\}$, $k = 0, \dots, K$ are determined by minimizing the sum of squared errors. The replicating portfolio R_P is then given by

$$R_P(\mathbf{Z}) := \sum_{k=0}^K e_k(\mathbf{Z}). \quad (4.51)$$

Clearly, the structure of the generalized Asian options does not fully identify the original Grosen-Jorgensen payoff. Nonetheless, as we will see, with the replicating portfolio of (4.51) the behaviour of the Grosen-Jorgensen payoff can be largely captured.

To empirically test the performance of generalized Asian options we consider a Grosen-Jorgensen payoff with maturity $T = 11$, $r_G = 0$, $L_0 = Z_0 = 100$, $\lambda = 0.1$ and $\delta = 0.75$. Let the asset process be given by a geometric Brownian motion

$$Z(s) = Z(s - 1)e^{(\mu - \frac{1}{2}\sigma^2) + \sigma(W(s) - W(s-1))}$$

In our example $\mu = 0.08$ and $\sigma = 0.16$. Based on a sample of size $N = 1000$ the coefficients of the replicating portfolio in (4.51) are globally optimized for a chosen number of basis terms K by minimizing the error sum of squares. For $K = 4$ a remarkably good fit is already achieved with an out-of-sample R^2 of 99.73%, which is illustrated in Figure 4.30, where the replicating portfolio result is plotted against the target Grosen-Jorgensen payoff. The out-of-sample R^2 is based on a sample $m = 1000$ that has not been used in the calibration of the replicating portfolio. Figures 4.31 to 4.33 illustrate the variation of the

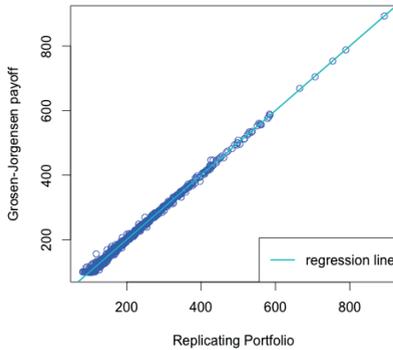


Figure 4.30: Goodness of fit of the replicating portfolio (Example 28).

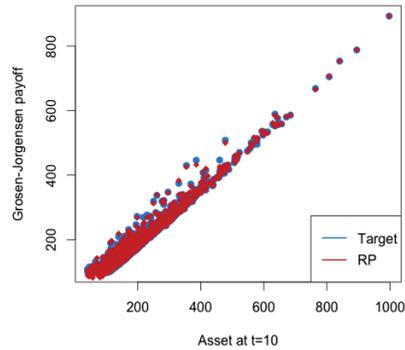


Figure 4.31: Comparison of variation with asset process at $t = 10$ (Example 28).

Grosen-Jorgensen payoff function and the replicating portfolio against the asset process at different time points. In blue the true Grosen-Jorgensen payoff is plotted against the asset value at the respective time point; the analogue is plotted in red for the replicating portfolio. Clearly, the replicating portfolio mirrors the behaviour of the target payoff very closely.

Recall that in the context of insurance risk capital calculations replicating portfolios of the liability payoffs are constructed in order to simplify the calculation of the liability value at the risk horizon. Therefore, the value of the replicating instruments making up the replicating portfolio must be readily available. Generalized Asian options are path-dependent and closed-form solutions to their value are normally not available. However, good approximations to the value of Asian options have been found, which makes them almost analytically priceable and justifies their use in portfolio replication. We refer the reader to Rogers and Shi (1995), for example.

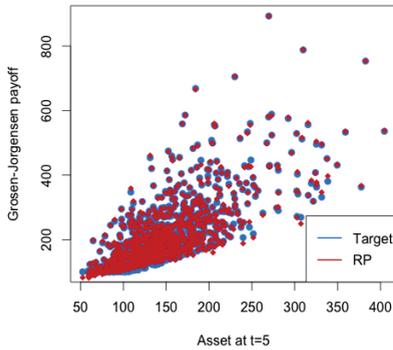


Figure 4.32: Comparison of variation with asset process at $t = 5$ (Example 28).

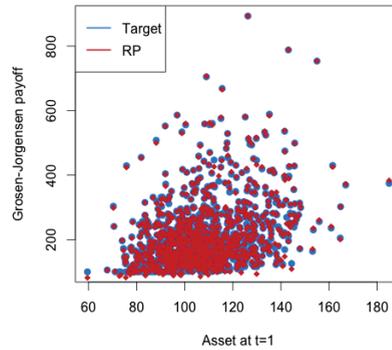


Figure 4.33: Comparison of variation with asset process at $t = 1$ (Example 28).

The previous example has shown that although portfolio replication is a more difficult problem when it comes to path-dependent insurance products, good solutions can be found and portfolio replication is feasible for such payoff functions.

4.5 Conclusion

In this paper two popular proxy techniques commonly applied in the risk management of life insurance policies for approximating unknown conditional expectation functions have been discussed. Their mathematical set-ups have been given and it has been shown that while both methods belong to the category of least squares Monte Carlo algorithms they work very differently. LSMC provides a direct approximation to the conditional expectation function and is a function fitting method. In portfolio replication a replicate of the terminal payoff function is constructed instead. This is then used to obtain a proxy to the conditional expectation function. The difference in

Table 4.5: Comparison portfolio replication versus LSMC.

Portfolio Replication (LSMC Regress-Later)	LSMC (LSMC Regress-Now)
<ul style="list-style-type: none"> • Non-noisy regression • By construction implies fit to the conditional expectation function at any $t < T$. • R^2 is a useful measure with an $R^2 = 1$ reflecting a perfect fit. • Result is asymptotically independent of the measure used for calibration. • Potentially faster convergence rate than N^{-1} can be achieved. • Path-dependent policies imply a higher dimensionality of the problem and finding a good basis is more challenging. • Choice of basis is critical. The conditional expectation of the basis must be readily available. 	<ul style="list-style-type: none"> • Noisy regression • Achieves only a fit to the conditional expectation function $\mathbb{E}_{\mathbb{P}}[X \mathcal{F}_t]$ for a particular $t < T$. • R^2 is not a useful measure and is always lower than 1. • Result depends on the measure chosen for calibration. • Convergence rate can never exceed N^{-1}. • Path-dependent policies do not imply a higher dimensionality of the problem in LSMC. Finding a basis is in principle easier. • Choice of basis is not limited by the easiness of calculating its conditional expectation. In principle any basis built on $A_t(Z)$ works.

the set-up of LSMC and portfolio replication has multiple practical consequences which have been illuminated using elementary examples. In that respect it has been shown that the performance of LSMC versus portfolio replication depends on several factors. These are summarized in Table 4.5. Clearly, portfolio replication enjoys multiple benefits such as potentially faster convergence than at rate N^{-1} , where N is the sample size, asymptotic measure independence and R^2 as a simple and meaningful measure for assessing the quality of the replicating portfolio. Its major challenge pertains to the replication of (strongly) path-dependent insurance policies. Using a “naive” multivariate basis constructed as the tensor product of the univariate bases quickly poses the curse of dimensionality problem. The LSMC technique does not suffer from the same problem and is easier to use for path-dependent payoffs compared to the replicating portfolio technique. In portfolio replication alternative basis constructions must be considered in order to overcome the curse of dimensionality problem. For a strongly path-dependent profit-sharing contract commonly encountered in insurance we have provided a solution approach that results in a very good replicating portfolio. Of course, much room for future research is left to explore the possibilities for replicating path-dependent insurance payoffs. Overall we can conclude that while portfolio replication is a more difficult problem when it comes to path-dependent payoff functions, we have revealed multiple advantages of portfolio replication which show that the challenge of finding a replicating portfolio is worthwhile.

Chapter 5

Conclusion

“The idea that the future is unpredictable is undermined every day by the ease with which the past is explained.”

– Daniel Kahneman

This chapter provides concluding remarks for this thesis. Conclusions specific to each chapter are embedded in the corresponding chapter, wherefore in this chapter details are omitted, but overall conclusions as well as general limitations of the thesis are identified.

This thesis addresses the challenge in insurance risk management to find approximations to the time t value of life insurance liabilities, where t is one year in the Solvency II framework (see SCR 1.9, EIOPA, 2012), in order to enable as well as speed up risk capital calculations. In calculating risk capital figures balance sheet items must be re-valued under various scenarios for the underlying risk factors. The objective to find approximations to the value function of life insurance liabilities arises as closed-form solutions are generally not available and full nested stochastic Monte Carlo methods often result in a too high simulation effort. The value function of liabilities at a time point t conditional on the realization of the underlying risk factors at that time point basically corresponds to a conditional expectation function across time. The target is thus to find approximations to unknown conditional expectation functions across time. Two concepts, borrowed from the financial literature, are commonly applied in insurance for achieving this goal: Least Squares Monte Carlo and portfolio replication.

In this thesis we have contributed to this topic threefold. In Chapter 2 it was addressed that the Least Squares Monte Carlo method offers two versions: Regress-Now and Regress-Later. The convergence rate of Regress-Later estimators has then been derived and compared to the convergence rate of Regress-Now estimators which has already been analyzed in existing literature. Chapter 3 discussed the replicating portfolio technique. It was shown that portfolio replication corresponds to Regress-Later, which provides a clear link between portfolio replication and LSMC. Moreover, the results of 3 provide a mathematical foundation for the application of replicating portfolios in insurance lia-

bility modeling. In Chapter 4 portfolio replication (i.e. Regress-Later) was compared to the conventional LSMC method (i.e. Regress-Now), revealing their advantages and disadvantages. We summarize the main contributions of this thesis in the following.

1. *Regress-Later potentially achieves faster convergence than Regress-Now.*

The mathematical frameworks for Regress-Now and Regress-Later were introduced in Chapter 2. While the convergence rate for Regress-Now has been analyzed in the literature (see, for example, Stentoft, 2004; Newey, 1997), the convergence analysis of Regress-Later has so far been missing. We have closed this gap by deriving the convergence rate of the Regress-Later estimator. It was shown that the Regress-Later estimator potentially converges faster than at N^{-1} , where N denotes the sample size. In comparison, the Regress-Now estimator can never converge faster than at N^{-1} . This feature makes Regress-Later an attractive choice when deciding on an approximation technique for unknown conditional expectation functions.

2. *The replicating portfolio problem is mathematically well-defined and asymptotically converges to the true solution.*

The results of Chapter 2 have been applied in Chapter 3, where it was argued that portfolio replication corresponds to Regress-Later. Exploiting the convergence theory of Chapter 2 it was then shown that the replicating portfolio asymptotically converges to the true target terminal payoff function at a time T , where T is the terminal time point. Moreover, given that the replicating portfolio converges to the terminal payoff function, its value function at an earlier time point $t < T$ converges to the value function of the target contingent claim. We thereby laid out

the elementary theoretical validation for applying replicating portfolios in insurance risk management.

- 3. The replicating portfolio technique offers numerous advantages and is therefore an attractive choice.*

In this thesis we have seen two methods for obtaining an estimate to the value function of insurance liabilities: LSMC with Regress-Now and portfolio replication (i.e. LSMC with Regress-Later). In Chapter 4 we addressed the differences between these two methods and the implications in practice. Numerous advantages of replicating portfolios have been identified, such as non-noisy regression, asymptotic measure independence and potentially faster convergence as well as more accurate estimates. On the other hand, it was shown that the replicating portfolio technique is challenging when it comes to path-dependent contingent claims. While we have given a proposal on how to address these problems, more advanced solutions to the curse of dimensionality problem in portfolio replication may be investigated in future research.

Overall this thesis sheds light on LSMC and replicating portfolio techniques in the context of approximating unknown conditional expectation functions of insurance liabilities across time. It emphasizes that portfolio replication is a mathematically sound concept and an attractive choice compared to LSMC (Regress-Now). The novelties presented in this thesis significantly contribute to a better understanding of the two proxy techniques under discussion. The results are not only interesting from a theoretical perspective, but also significantly contribute to the discussion among insurance risk managers having to select a proxy method for risk capital calculations. In the context of Solvency II insurers have to validate their internal models. The results

of this thesis are here potentially very useful for insurers using proxy techniques to their liabilities in their internal risk management models.

Of course, the analysis conducted in this thesis is subject to limitations and is far from being exhaustive. Several questions remain which provide interesting avenues for future research. The remainder of this chapter is dedicated to these.

1. *How to solve the curse of dimensionality problem?*

As indicated in Chapters 3 and 4 replicating high-dimensional insurance policies is particularly difficult as the curse of dimensionality problem arises. In LSMC (Regress-Now) the same problem is encountered, though less for path-dependent policies while the replicating portfolio technique is here still affected. For the approximation techniques to be feasible solutions to the curse of dimensionality problem are paramount. While proposals have been made in this thesis much room for a deeper analysis is left.

2. *Which basis to use?*

In this thesis we have used piecewise linear functions as basis, which are similar to combinations of call options. The advantage of piecewise linear functions is that their convergence rate can be explicitly calculated. Also, their similarity to call options justifies their usefulness in practice. Still, it would be of interest to investigate the performance of other basis functions in terms of their convergence rate and to identify a set of basis functions that works best for a set of target functions.

3. *Do Regress-Later estimators also potentially converge faster in multi-period models?*

In this thesis the convergence rate of Regress-Later estimators has been derived in single-period models. For the applications mentioned in this thesis single-period models are of interest.

However, Regress-Later may also be interesting for multi-period models as, for instance, relevant in Bermudan and American option pricing. The advantage of Regress-Later as presented in this thesis is driven by the fact that the target payoff function is known. This advantage is partially lost in the multi-period models required for Bermudan and American option pricing. Therefore, it would be very interesting to see whether overall faster convergence may still be achieved with Regress-Later, which would make Regress-Later an attractive alternative to the standard LSMC (Regress-Now) methods applied to Bermudan/American option pricing.

4. *What is the impact of the approximation error in proxy methods on risk capital estimates?*

The methods discussed in this thesis are approximation methods to the value function of insurance liabilities. The approximations are then used in the estimation of risk capital numbers, such as VaR. Quantifying the impact of the approximation error on the estimation of the risk capital is a very interesting research question.

5. *Incomplete market problem for life insurance liabilities*

In this thesis we have omitted considerations addressing the incomplete market problem that is faced when it comes to the valuation of life insurance liabilities. Thus, we have implicitly presumed that a valuation measure exists or that an assumption on the valuation measure is made. The concepts of this thesis may be extended by explicitly accounting for the market incompleteness problem to ensure model robustness.

6. *Decreasing the projection error in Regress-Now*

In this thesis the Regress-Now approach as suggested in Glasser-

man and Yu (2004b) has been investigated, i.e. independent replications of the payoff function valued at time T are regressed against basis functions valued at an earlier time point $t < T$. It has briefly been addressed in Chapter 2 that, alternatively, one could consider to construct (weak) estimates of the conditional expectation function through an appropriate aggregation of (few) inner simulations of the payoff, which are then regressed against basis functions at time t . It would be interesting to investigate if, how exactly, and to what extent the projection error in Regress-Now could thereby be reduced.

In conclusion, while there are a variety of open questions and possible options for future research, this thesis provides a first step towards better understanding the replicating portfolio technique as commonly applied in insurance risk management from both a theoretical as well as a practical perspective. It is not only shown that the replicating portfolio technique is a mathematically sound concept, but also that it is a very attractive choice when deciding on an approximation technique in life insurance liability modeling. The results of this thesis therefore contribute, theoretically and empirically, to the discussion among insurance risk managers on the choice of approximation techniques in life insurance risk management.

Bibliography

- Andreatta, G. and S. Corradin (2003). Valuing the surrender options embedded in a portfolio of italian life guaranteed participating policies: a Least Squares Monte Carlo Approach. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.145.1096>.
- Bacinello, A. R. (1993). Pricing equity-linked life insurance with endogenous minimum guarantees. *Insurance: Mathematics and Economics* 12(3), 245–257.
- Bacinello, A. R. (2001). Fair pricing of life insurance participating policies with a minimum interest rate guarantee. *ASTIN BULLETIN* 31(2), 257–297.
- Bacinello, A. R., E. Biffis, and P. Millosovich (2010). Regression-based algorithms for life insurance contracts with surrender guarantees. *Quantitative Finance* 10(9), 1077–1090.
- Bacinello, A. R., E. Biffis, and P. Millosovich (2009). Pricing life insurance contracts with early exercise features. *Journal of Computational and Applied Mathematics* 233(1), 27–35.

BIBLIOGRAPHY

- Bauer, D., D. Bergmann, and R. Kiesel (2010). On the risk-neutral valuation of life insurance contracts with numerical methods in view. *ASTIN Bulletin* 40(1), 65–95.
- Bauer, D., D. Bergmann, and A. Reuss (2009). Solvency II and nested simulations - a Least-Squares Monte Carlo approach. http://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi2/forschung/preprint-server/2009/0905_200905_solvency_preprint-server.pdf. Working Paper.
- Bauer, D., R. Kiesel, A. Kling, and J. Ruß (2006). Risk-neutral valuation of participating life insurance contracts. *Insurance: Mathematics and Economics* 39(2), 171–183.
- Belomestny, D. (2011). Pricing Bermudan options by nonparametric regression: optimal rates of convergence for lower estimates. *Finance and Stochastics* 15(4), 655–683.
- Belomestny, D., A. Kolodko, and J. Schoenmakers (2010). Regression methods for stochastic control problems and their convergence analysis. *SIAM Journal on Control and Optimization* 48(5), 3562–3588.
- Bender, C. and J. Steiner (2012). Least-squares Monte Carlo for backward SDEs. In *Numerical Methods in Finance*, Volume 12 of *Springer Proceedings in Mathematics*, pp. 257–289. Berlin: Springer Berlin Heidelberg.
- Bergstrom, A. R. (1985, April). The estimation of nonparametric functions in a Hilbert space. *Econometric Theory* 1(1), 7–26.
- Beutner, E., A. Pelsser, and J. Schweizer (2013). Fast convergence of Regress-Later estimates in Least Squares Monte Carlo. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2328709. Working Paper.

- Beutner, E., A. Pelsser, and J. Schweizer (2015). Theory and Validation of Replicating Portfolios in Insurance Risk Management. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2557368. Working Paper.
- Bogachev, V. I. (2007). *Measure Theory*, Volume 1. Berlin: Springer-Verlag.
- Boyle, P. P. and E. S. Schwartz (1977). Equilibrium prices of guarantees under equity-linked contracts. *The Journal of Risk and Insurance* 44(4), 639–660.
- Breeden, D. T. and R. H. Litzenberger (1978). Prices of state-contingent claims implicit in option prices. *The Journal of Business* 51(4), 621–651.
- Brennan, M. J. and E. S. Schwartz (1976). The pricing of equity-linked life insurance policies with an asset value guarantee. *Journal of Financial Economics* 3(3), 195–213.
- Broadie, M. and M. Cao (2008). Improved lower and upper bound algorithms for pricing American options by simulation. *Quantitative Finance* 8(8), 845–861.
- Broadie, M. and P. Glasserman (1997). Pricing American-style securities using simulation. *Journal of Economic Dynamics and Control* 21(8-9), 1323–1352.
- Burmeister, C. (2007). Portfolio replication - variable annuity case study. http://www.ermssymposium.org/2008/pdf/handouts/Q/Q5_burmeister.pdf.
- Carr, P. and J. Bowie (1994). Static simplicity. *Risk* 7(8), 45–49.
- Carr, P. and A. Chou (1997). Breaking barriers. *Risk* 10(9), 139–145.

BIBLIOGRAPHY

- Carr, P., K. Ellis, and V. Gupta (1998, 06). Static hedging of exotic options. *Journal of Finance* 53(3), 1165–1190.
- Carriere, J. F. (1996). Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics* 19(1), 19–30.
- Casa, F. D. and M. Gaffo (2013). Portfolio optimization via replication. *Insurance Risk*, 36–41.
- Chen, W. and J. Skoglund (2012). Cashflow replication with mismatch constraints. *The Journal of Risk* 14(4), 115–128.
- Chen, X. (2007, January). *Large Sample Sieve Estimation of Semi-Nonparametric Models*, Volume 6 of *Handbook of Econometrics*. Elsevier.
- Clement, E., D. Lamberton, and P. Protter (2002). An analysis of a least squares regression method for American option pricing. *Finance and Stochastics* 6(4), 449–471.
- Conway, J. B. (1985). *A course in functional analysis*, Volume 96. New York: Springer-Verlag.
- Daul, S. and E. G. Vidal (2009). Replication of insurance liabilities. *Risk Metrics* 9(1), 79–96.
- Davidson, C. (September 2011). The alternative route. *Life & Pension-Risk*, 14–17.
- Derman, E., D. Ergener, and I. Kani (1995). Static options replication. *The Journal of Derivatives* 2(4), 78–95.
- Devineau, L. and M. Chauvigny (2011). Replicating portfolios: Calibration techniques for the calculation of the Solvency II economic capital. *Bulletin Francais D'Actuariat* 11, 59–97.

- Egloff, D., M. Kohler, and N. Todorovic (2007). A dynamic look-ahead Monte Carlo algorithm for pricing Bermudan options. *The Annals of Applied Probability* 17(4), 1139–1171.
- EIOPA (2012). Technical specifications for the Solvency II valuation and Solvency Capital Requirements calculation (Part I). .
- Ekern, S. and S.-A. Persson (1996). Exotic unit-linked life insurance contracts. *The Geneva Papers on Risk and Insurance Theory* 21, 35–63.
- Gerhold, S. (2011). The Longstaff-Schwartz algorithm for Lévy models: results on fast and slow convergence. *The Annals of Applied Probability* 21(2), 589–608.
- Glasserman, P. and B. Yu (2004a). Number of paths versus number of basis functions in American option pricing. *The Annals of Applied Probability* 14(4), 2090–2119.
- Glasserman, P. and B. Yu (2004b). Simulation for American options: Regression now or regression later? In *Monte Carlo and Quasi-Monte Carlo Methods 2002*, pp. 213–226. Berlin Heidelberg: Springer-Verlag.
- Gobet, E. and K. Surana (2014). A new sequential algorithm for L_2 -approximation and application to Monte-Carlo integration. http://hal.archives-ouvertes.fr/docs/00/97/20/16/PDF/SequentialVariationReduction_final_HAL.pdf.
- Gobet, E. and T. B. Zineb (2013). Preliminary control variates to improve empirical regression methods. *Monte Carlo Methods and Applications* 19(4), 331–354.

BIBLIOGRAPHY

- Grosen, A. and P. L. Jørgensen (2000). Fair valuation of life insurance liabilities: The impact of interest rate guarantees, surrender options, and bonus policies. *Insurance: Mathematics and Economics* 26(1), 37–57.
- Hansen, B. (2014). Econometrics. <http://www.ssc.wisc.edu/~bhansen/econometrics/>. Online draft textbook.
- Hörig, M. and M. Leitschkis (2012). Solvency II proxy modelling via Least Squares Monte Carlo. <http://www.milliman.com/insight/insurance/Solvency-II-proxy-modelling-via-Least-Squares-Monte-Carlo/>.
- Hörig, M., M. Leitschkis, K. Murray, and E. Phelan (2014). An application of Monte Carlo proxy techniques to variable annuity business: A case study. <http://de.milliman.com/uploadedFiles/insight/2013/monte-carlo-proxy-techniques.pdf>.
- Hunter, J. K. (2011). Measure theory. University Lecture Notes.
- Kalberer, T. (2012). Stochastic determination of the value at risk for a portfolio of assets and liabilities. *Der Aktuar* 1, 12–22.
- Koursaris, A. (2011a). The advantages of Least Squares Monte Carlo. http://www.barrhibb.com/documents/downloads/The_Advantages_of_Least_Squares_Monte_Carlo.pdf.
- Koursaris, A. (2011b). A Least Squares Monte Carlo approach to liability proxy modelling and capital calculation. http://www.barrhibb.com/documents/downloads/Least_Squares_Monte_Carlo_Approach_to_Liability_Proxy_Modelling_and_Capital_Calculation.pdf.

- Koursaris, A. (2011c). A primer in replicating portfolios. http://www.barrhibb.com/documents/downloads/Primer_in_Replicating_Portfolios.pdf.
- Longstaff, F. A. and E. S. Schwartz (2001). Valuing American options by simulation: A simple least squares approach. *Review of Financial Studies* 14(1), 113–47.
- Madan, D. B. and F. Milne (1994). Contingent claims valued and hedged by pricing and investing in a basis. *Mathematical Finance* 4(3), 223–245.
- Morrison, S. (2008). Replicating Portfolios for economic capital: Replication or approximation? http://www.barrhibb.com/documents/downloads/Model_Insights_04_-_Replicating_Portfolios_for_Economic_Capital.pdf.
- Natolski, J. and R. Werner (2014). Mathematical analysis of different approaches for replicating portfolios. *European Actuarial Journal* 4(2), 411–435.
- Newey, W. K. (1997, July). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79(1), 147–168.
- Oechslin, J., O. Aubry, M. Aellig, A. Kaeppli, D. Broennimann, A. Tandonnet, and G. Valois (2007). Replicating embedded options. *Life & Pensions*.
- Pelsser, A. (2003). Pricing and hedging guaranteed annuity options via static option replication. *Insurance: Mathematics and Economics* 33(2), 283–296.

BIBLIOGRAPHY

- Pelsser, A. and R. Plat (2009). Analytical approximations for prices of swap rate dependent embedded options in insurance products. *Insurance: Mathematics and Economics* 44, 124–134.
- Pelsser, A. and J. Schweizer (2015). The Difference between LSMC and Replicating Portfolio in Insurance Liability Modeling. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2557383. Working Paper.
- Pelsser, A. A. and D. F. Schrage (2004). Pricing rate of return guarantees in regular premium unit linked insurance. *Insurance: Mathematics and Economics* 35(2), 369–398.
- Rogers, L. C. G. and Z. Shi (1995). The value of an Asian option. *Journal of Applied Probability* 32(4), 1077–1088.
- Schoenmakers, J., J. Zhang, and J. Huang (2013). Optimal dual martingales, their analysis, and application to new algorithms for Bermudan products. *SIAM Journal of Financial Mathematics* 4(1), 86–116.
- Siotani, M. (1967). Some application of Loewner's ordering on symmetric matrices. *Annals of the Institute of statistical Mathematics* 19(2), 245–259.
- Stentoft, L. (2004, September). Convergence of the Least Squares Monte Carlo approach to American option valuation. *Management Science* 50(9), 1193–1203.
- Tanskanen, A. J. and J. Lukkarinen (2003). Fair valuation of path-dependent participating life insurance contracts. *Insurance: Mathematics and Economics* 33(3), 595–609.

The European Parliament and The Council (2009). Directive 2009/138/EC of the European Parliament and of The Council of 25 November 2009. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:335:0001:0155:en:PDF>.

Tsitsiklis, J. and B. Van Roy (2001, July). Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks* 12(4), 694–703.

Zanger, D. Z. (2013). Quantitative error estimates for a least-squares Monte Carlo algorithm for American option pricing. *Finance and Stochastics* 17(3), 503–534.

Valorization

“Nothing is more practical than a good theory.”

– Kurt Lewin

This addendum is dedicated to the discussion of the economic and social value added by this dissertation as well as the opportunities of transferring the knowledge presented in this thesis into practical use. According to the National Valorization Committee knowledge valorization refers to “the process of creating value from knowledge, by making knowledge suitable and/or available for social (and/or economic) use and by making knowledge suitable for translation into competitive products, services, processes and new commercial activities”. This thesis has been supported by the largest German insurer, Allianz, who is a major stakeholder and a customer of the results of this thesis. In that respect knowledge valorization has been a continuous process throughout the construction of this thesis. In the following addendum I outline the knowledge valorization of this dissertation by discussing the economic relevance of portfolio replication¹, the stakeholders and target groups, the implications of the results for the industry, and the innovativeness of the research.

As the title of this thesis already suggests this work is mainly targeted at the insurance risk sector. Solvency II defines the regulatory framework for insurance supervision in the EU. It is a comprehensive concept that, among other aspects, lays down the capital requirements for the insurance industry and the quantitative rules for determining the risk capital. While the Solvency II framework offers a standard model to the calculation of risk capital requirements, particularly large insurers opt for an internal model to better represent the individual characteristics of their business. Part of the Solvency II requirement is the market-consistent valuation of the insurer’s own funds. This is a challenging task as insurance liabilities are typically not traded

¹Chapter 1 already to a large extent addresses the economic relevance and necessity to discuss portfolio replication in the context of insurance risk management, which is the topic of this dissertation. Therefore, part of the discussion in this chapter can also be found in Chapter 1.

and market values are therefore not immediately available. Moreover, many insurance products are complex in their structure and underlying dynamics and closed-form solutions to their value do not exist. Hence, numerical techniques are typically employed instead. Straightforwardly, the value of an insurance liability may simply be estimated through Monte Carlo simulation of all underlying risk factors from the risk horizon to the terminal time point, mostly the maturity time point, of the insurance policy (or a fund of pooled policies). The sample average then serves as the value estimate of the liability at the risk horizon. The problem now is, that an insurer requires the value of the liabilities under different possible economic risk factor realizations at the risk horizon in order to calculate risk capital figures as required by Solvency II. For a large insurance company, for example, it is common to consider 10000-50000 scenario realizations at the risk horizon. This means that the exercise of estimating the liability value at the risk horizon would need to be repeated 10000-50000 times. Using, for example, 1000 Monte Carlo simulations for the estimation of the liability value at each scenario realization at the risk horizon would result in a total simulation set of 10-50 million scenarios. Depending on the size and complexity of the insurance company this may quickly lead to an infeasible simulation size. The problem particularly arises for life insurance products where very long maturities such as 60 years are to be expected and minimum guarantee and profit-sharing mechanisms often additionally complicate their valuation. The problem statement has triggered the demand for alternative solutions to the valuation of (life) insurance liabilities. Commonly, approximation techniques are used in combination with Monte Carlo simulation in order to represent the valuation function of the liabilities. Given a proxy function to the value of the liabilities the estimation of the value under different economic scenarios at the risk horizon tremendously simplifies when the

proxy function is used as a substitute for the insurers liability portfolio. Among these proxy techniques, portfolio replication and Least Squares Monte Carlo (LSMC) are very popular and widely applied in the industry. Allianz, for example, employs portfolio replication for its life insurance business in order to represent market risk in its risk capital calculations (see e.g. Davidson, 2011)

Now, while both methods are already applied in insurance risk modeling, their properties have not been fully explored yet and many open questions remain. This particularly pertains to portfolio replication as a proxy method in the context of insurance liability modeling, which has so far received little attention in the academic literature compared to LSMC. LSMC is a well-known numerical technique that originates from American option pricing and has been widely discussed in the literature; see, among others, Carriere (1996), Longstaff and Schwartz (2001), Tsitsiklis and Van Roy (2001), Clement et al. (2002), Stentoft (2004), Glasserman and Yu (2004b), Egloff et al. (2007), Belomestny (2011), Gerhold (2011) and Zanger (2013). The asymptotic convergence theory for LSMC has, for example, been analyzed in Stentoft (2004), where it is shown that the LSMC estimator converges in the limit to the true value. Until now a theoretical foundation for replicating portfolios as a proxy technique has been missing in the academic literature. In that context the asymptotic convergence theory of the replicating portfolio technique has not been analyzed. Yet, understanding the asymptotic behaviour of an estimator is important when using a method. Another gap in the existing literature is a discussion on the advantages and disadvantages of the LSMC and the replicating portfolio technique and how they compare. While indications are given in the current literature (see Glasserman and Yu (2004b), Broadie and Cao (2008) and Bender and Steiner (2012)), a full-fledged discussion is presently missing. In practical applications considerations such as what quality

measures to use for the assessment of a proxy solution and what to be aware of in the set-up of the calibration scenarios are very important.

In conclusion, questions that insurance risk managers pose in choosing and setting up a proxy method in order to simplify the representation of their liabilities have not fully been answered yet by the existing literature. Apart from the internal aspiration to establish a solid internal risk model there are also external regulatory requirements. Insurers developing or extending their internal risk models within the Solvency II framework are subject to a regulatory internal model approval process, as a result of which the regulator will either accept or reject the proposed model or model changes. In order to get their internal risk models approved by the regulators, insurers have to perform a validation of their risk model. Consequently, insurers using the LSMC and/or portfolio replication technique also have to provide information regarding the validity of the method employed and why a particular method has been chosen over other alternatives.

This thesis closes the aforementioned gaps in the literature by analyzing the asymptotic properties of the replicating portfolio method, comparing the two methods under discussion and elaborating on implications in practice. In Chapters 2 and 3 it is shown that portfolio replication is a mathematically sound concept with a well-developed theoretical background. The asymptotic convergence of the replicating portfolio estimator to its true value is an important aspect with regards to the validity of the replicating portfolio method as a proxy technique for the representation of insurance liabilities. Chapter 4 discusses the differences between portfolio replication and LSMC as well as the implications of these differences for practical applications. The information provided supports the process of choosing one method over the other and stresses what to be aware of when using a method.

The results and methods presented in this thesis allow insurance risk managers, researchers and regulators to obtain a better understanding of the mathematical and asymptotic properties of the replicating portfolio technique, on the one hand, and, on the other hand, to see a clear relation to the currently more popular LSMC method. Understanding the advantages and disadvantages of these methods helps stakeholders to make better informed decisions on the choice of one method over the other. The information provided in this thesis thereby also supports insurance risk managers in explaining their model choice to the regulator. Moreover, aspects in the calibration and assessment of replicating portfolios are addressed in this thesis. For instance, in Chapter 4 the usefulness of a particular quality measure for the assessment of replicating portfolios is examined. Another example is the construction of scenarios for the calibration of the replicating portfolio. Both considerations support insurance risk managers in setting up an appropriate replicating portfolio approach.

The results of this thesis are mainly targeted at insurance risk managers and regulators in the context of Solvency II. This is because in that field proxy methods find direct application in risk capital calculations as they enable a simplified representation of insurance liabilities. However, the same idea may be leveraged to enable other interesting analysis in the insurance context. Casa and Gaffo (2013) discuss portfolio replication in the context of asset-liability management and portfolio optimization, where the replicating portfolio technique is leveraged to efficiently compare a series of different asset allocation opportunities. A careful implementation of the replicating portfolio technique is essential for the analysis to be meaningful. The results of this thesis therefore also find application in that research field. Additionally, there may be numerous other fields where the use of

proxy methods is of interest and the results of this thesis may be helpful.

As already mentioned the process of knowledge valorization outlined in this addendum is already realized by an application of the results of this thesis outside the academic world. It has been mentioned before that Allianz employs portfolio replication in its internal risk model, but also other insurers use the replicating portfolio or LSMC technique and can therefore profit from the outcomes of this thesis. This thesis also indicates various avenues for potential future research valued by the insurance industry. One option is the extension of the results of this thesis to multi-period problems. This is relevant in insurance due to dynamic lapse and surrender options, which make (life) insurance liabilities similar in its structure to American options. Furthermore, complex path-dependent insurance liabilities are more difficult to replicate and the construction of a appropriate replicating instruments is challenging. While a first step towards solving such problems is given in this thesis, much more research is required to find more general solutions. I look forward to extensions of this thesis in that direction as well as the development towards wider applicability of replicating portfolios in the insurance sector.

Curriculum Vitae

“Who in the world am I? Ah, that’s the great puzzle.”

– Lewis Carroll, *Alice in Wonderland*



Janina Schweizer was born on August 23, 1986 in Berlin, Germany. She received her Gymnasium diploma (Abitur) in 2006 in Germany, finishing as the second-best in class. After subsequently having spent one year in Chicago, USA, she studied Economics and Business Economics at Maastricht University receiving her B.Sc. in 2010. During her undergraduate studies she spent one semester as an exchange student at Hong Kong University of Science and Technology in Hong Kong. Subsequently, she studied Econometrics at Maastricht University, for which she obtained her M.Sc. with distinction (*cum laude*) in August 2011.

After graduation Janina started as an external Ph.D. candidate at the Department of Quantitative Economics in September 2011 under the supervision of Prof. dr. Antoon Pelsser and dr. Eric Beutner. At the same time she joined the central risk management department of the Allianz headquarter in Munich, where she worked on Replicating Portfolios used in the internal risk model under Solvency II. Her role at Allianz enabled her to gain highly valuable practical insight related to her academic research. The results of Janina's research are presented in this thesis. Janina presented her work at various international academic conferences, such as the World Congresses of the Bachelier Finance Society in Sydney and in Brussels, CEQURA (Center for quantitative risk analysis) conference in Munich and the Netspar Pension day.

In March 2015 Janina moved to the Finance function of Allianz Germany, where she worked on finance-related strategic projects for the CFO office. In July 2016 Janina assumed her new role as assistant to the CFO of Allianz Germany.

“Begin at the beginning,” the King said, very gravely, “and go on till you come to the end: then stop.”

– Lewis Carroll, *Alice in Wonderland*

Portfolio Replication and Least Squares Monte Carlo

with Application to Insurance Risk Management

The Solvency II framework requires insurers to market-consistently value their own funds. The task is challenging given that insurance liabilities are typically not traded financial instruments and closed-form solutions are mostly not available. One solution is to obtain an estimate of the future value of liabilities through pure Monte Carlo simulations, which, however, in risk-capital calculations quickly becomes too time-intensive. This thesis deals with Least Squares Monte Carlo (LSMC) approaches, Regress-Now and Regress-Later, that yield an approximation to the value of the insurance liabilities. The asymptotic properties of the methods are analyzed. It is shown that the Replicating Portfolio technique commonly applied by insurers, corresponds to LSMC with Regress-Later. Thereby a theoretical foundation for the Replicating Portfolio technique is provided. Lastly, advantages and disadvantages of Replicating Portfolio and LSMC (with Regress-Now) are discussed.

