Applied topics in conic finance and credit risk

Michielon, M.

Publication date
2024
Document Version
Final published version

APPLIED TOPICS IN CONIC FINANCE AND CREDIT RISK

MATTEO MICHIELON

UNIVERSITEIT VAN AMSTERDAM

APPLIED TOPICS IN CONIC FINANCE AND CREDIT RISK

MATTEO MICHIELON

UNIVERSITEIT VAN AMSTERDAM

ISBN: 978-94-6496-193-5

APPLIED TOPICS IN CONIC FINANCE AND CREDIT RISK

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op donderdag 3 oktober 2024, te 16.00 uur
door

MATTEO MICHIELON
geboren te Portogruaro

Dedicata a mio padre, Guido.

PROMOTIECOMMISSIE

PROMOTORES:

prof. dr. P.J.C. Spreij	Universiteit van Amsterdam

COPROMOTOR:

dr. A. Khedher	Universiteit van Amsterdam

OVERIGE LEDEN:

prof. dr. B.D. Kandhai	Universiteit van Amsterdam
dr. R. Pietersz	ABN AMRO Bank N.V.
prof. dr. W. Schoutens	Katholieke Universiteit Leuven
prof. dr. M. Vanmaele	Universiteit Gent
prof. dr. ir. M.H. Vellekoop	Universiteit van Amsterdam
prof. dr. ir. E.M.M. Winands	Universiteit van Amsterdam

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

CONTENTS

1 Introduction 1

I Applications in modeling financial markets under conic finance assumptions

2 From bid-ask credit default swap quotes to risk-neutral default probabilities using distorted expectations 15
- 2.1 Abstract 15
- 2.2 Introduction 15
- 2.3 Basic notions and valuation of CDSs 18
  - 2.3.1 Dynamics of the survival probabilities 20
- 2.4 Bid-ask pricing via distorted expectations 23
- 2.5 CDSs in a two-price economy 25
  - 2.5.1 A calibration example 29
- 2.6 Conclusion 31
- 2.A A remark on the monotonicity of CDS prices 33
- 2.B Tables 35

3 Liquidity-free implied volatilities: An approach using conic finance 37
- 3.1 Abstract 37
- 3.2 Introduction 37
- 3.3 Pricing in a two-price economy 40
- 3.4 Liquidity-free option implied volatilities 43
  - 3.4.1 Implied volatilities with the Wang transform 46
- 3.5 Example 50
- 3.6 Conclusion 57
- 3.A Proof of Theorem 2 58
- 3.B Derivation of the Bachelier option pricing formulae 60
- 3.C A remark on a property of the Wang transform 60

4 Neural network empowered liquidity pricing in a two-price economy under conic finance settings using multilayer perceptrons 63
- 4.1 Abstract 63
- 4.2 Introduction 64
- 4.3 Bid-ask pricing with distorted expectations 69
  - 4.3.1 Conic pricing of European options with the Wang transform 71
  - 4.3.2 Conic Monte Carlo 73
  - 4.3.3 Hybrid distortions 74
- 4.4 Machine learning techniques in quantitative finance 77
- 4.5 Neural networks 80
  - 4.5.1 Multilayer perceptrons 80

4.6 (Deep) Conic neural networks 83
4.6.1 A motivating example 83
4.6.2 CNNs in use 86
4.6.3 Sensitivities: Learning (the) Greek(s) 95
4.6.4 A warning on the choice of the activation function 99
4.6.5 Deep CNNs 101
4.7 Neural-network-driven LSV calibration 103
4.7.1 LSV essentials: A primer 104
4.7.2 Leverage function: A neural network approximation 105
4.7.3 Calibration 106
4.7.4 Application 108
4.8 Conclusion 114
4.A Replication of the conic SABR model by means of a CNN 116

II Wasserstein distances and proxy credit curves

5 Proxying credit curves via Wasserstein distances 121
5.1 Abstract 121
5.2 Introduction 121
5.3 Background information 124
5.3.1 A brief overview of Wasserstein distances 124
5.3.2 A brief overiew of CDSs 125
5.4 CDS proxy curves using Wasserstein square distances 127
5.5 Suitability of other metrics 132
5.6 Examples 133
5.7 Conclusion 139

6 On Wasserstein distances, barycenters, and the cross-section methodology for proxy credit curves 141
6.1 Abstract 141
6.2 Introduction 141
6.3 Literature review 144
6.3.1 The intersection methodology 144
6.3.2 The cross-section methodology 147
6.4 Wasserstein distances and the cross-section methodology 148
6.4.1 Wasserstein cross-intersection methodology on spreads 149
6.4.2 Wasserstein cross-intersection methodology on hazard rates 150
6.4.3 Global Wasserstein distance optimization 150
6.5 Comparison of the different methodologies 153
6.5.1 Data preprocessing 153
6.5.2 Results 155
6.6 Conclusion 161
6.A The cross-section methodology on hazard rates 163

7 Summary 165
8 Samenvatting (in Dutch) 169

9 Author contributions 173
Bibliography 175
10 Acknowledgements 185

1 INTRODUCTION

This PhD thesis is structured into two distinct yet interrelated parts, each delving into domains of current practical relevance in quantitative finance.

In the first part of the thesis (i.e., Part i), Chapter 2 introduces a methodology to compute risk-neutral default probabilities from *bid* and *ask* market quotes of *credit default swap* (CDS) transactions. By applying the theory of Chapter 2 further, in Chapter 3 we illustrate how to calculate implied volatility smiles from bid and ask options prices offering, under mild assumptions, a framework to derive theoretical risk-neutral option market prices without relying on any mid-quote approximation. Additionally, Chapter 4 explores two methodologies based on *neural networks* for valuing contingent claims in markets with direction-dependent prices, on the one hand improving the computational efficiency of bid and ask option price calculations, and enhancing flexibility in *local stochastic volatility* (LSV) models on the other hand.

The second part of the thesis (i.e., Part ii), like Chapter 2, focuses on CDSs. This time, however, quoted CDSs serve as inputs to estimate the default probabilities for those counterparties of financial institutions lacking reliable (or simply available) CDS market data. Chapter 5 illustrates, given a set of probability distributions implied from CDSs sharing similar characteristics, how to measure their *Wasserstein distances* in order to compute the respective *Wasserstein barycenters*. Building on these results, Chapter 6 provides a technique exploiting the cross-sectional information from the available CDS market quotes to construct CDS proxy credit curves.

Overall, the focus of the research conducted in this thesis lies in addressing applied topics within the financial sector, also motivated by the regulatory landscape shaping the industry.

Part i is devoted to a focused exploration, from a practical standpoint, of the *conic finance* theory Cherny and Madan (2010), which investigates the intricate interplay of *bid-ask spreads* in modern financial markets. That is, bid and ask prices are key indicators in security trading, with the bid (ask) price referring to the highest (lowest) price market agents are willing to pay (accept) to enter a given transaction. The difference between these two prices is what is called bid-ask spread which, basically, represents the cost for executing a transaction, and that reflects the dynamics of supply and demand. The bid-ask spread can be also interpreted as the profit margin for market makers, i.e., entities facilitating trading by ("continuously") quoting buy and sell prices for a wide range of

financial securities, so ensuring liquidity in financial markets. From a market agent's perspective, a smaller (wider) spread tends to indicate higher (lower) liquidity. Bid-ask spreads are also useful for gauging the expected volatility of a given market, with higher (lower) volatility typically leading to wider (narrower) bid-ask spreads as a consequence of the increase (decrease) in market uncertainty. Furthermore, bid-ask spreads represent a form of *friction* in security markets, by reducing (increasing) potential profits (losses) for buyers and/or sellers. For instance, in markets where bid-ask spreads are wide, traders might require security prices to substantially move in their favor before they can realize a gain, avoid a loss, or break even. Thus, bid-ask spreads act as frictions in trading activities by increasing the costs and complexity of asset exchanges, therefore negatively affecting market efficiency and transactional flows.

The conic finance theory is dedicated to understanding the dynamics and implications of bid and ask spreads across a diverse spectrum of domains such as, but not limited to, capital calculations Madan (2009); Madan (2012), the valuation of options and structured products Guillaume and Schoutens (2013); Madan and Schoutens (2017); Madan and Schoutens (2012); Madan et al. (2019); Marquet and Schoutens (2018) (and other complex financial instruments such as, *inter alia*, contingent convertibles Madan and Schoutens (2011a)), valuation adjustments van Bakel et al. (2020); Madan and Schoutens (2016b), and portfolio theory Madan (2016). The focus of conic finance is that of providing insights into the pricing of financial instruments (and into other areas of financial mathematics as well) in this nuanced context, and offers a set of novel and valuable tools that can be used to navigate the complexities of markets with frictions. In particular we investigate, under reasonable assumptions, how it is possible to extract, given bid and ask prices of specific derivative contracts, their risk-neutral counterparts. Furthermore, we also delve into the application of machine learning (and, specifically, of neural networks) to perform accurate and efficient pricing in the presence of bid-ask spreads.

The initial segment of this thesis, primarily encompassing Chapters 2 and 3, is dedicated to exploring the relationship between bid and ask quotes for a given financial product and their inherently non-observable risk-neutral counterparts. The reason for this investigation can be traced back to common financial modeling techniques, which frequently rely on the *law of one price*. Notably, this often-used principle to describe financial markets diverges from the observed market dynamics as, in practice, financial instruments are commonly quoted by means of their bid and ask prices, contingent on whether contracts are sold or bought, respectively. That is, financial markets are ruled by, in jargon, the *law of two prices*. However, in order to facilitate the modeling of the variables financial transactions depend on, it is often common practice to assume specific dynamics for the underlying process(es) under a predetermined risk-neutral measure, and to further hypothesize that each agreement can be bought and sold at the same price, the latter computed as (discounted) expectation of the future

cashflow(s) under the chosen pricing measure. Therefore, it is essential to ensure the coherence between the selected modeling framework and the market data utilized for calibration purposes. Nonetheless, due to the fact that risk-neutral market quantities (i.e., prices) are not directly observable in the market, their associated model parameters cannot be implied from the available market quotes right away. In response to this theoretical but also practical limitation, it is common practice to approximate the unobservable risk-neutral prices with their mid-point counterparts (i.e., by assuming the former equal to the average of the relevant bid and ask prices that can be observed in the market), as doing so provides an extremely intuitive and simple *modus operandi*. However, being the modeling choice just described an approximation (and, thus, not necessarily a coherent and faithful representation of reality), it is then natural to ask whether the aforementioned risk-neutral values could be, by means of a viable methodology, calculated.

In this respect Chapter 2 delves into the role of risk-neutral default probabilities and their link to CDS market quotes. The rationale behind this research choice is given by the fact that risk-neutral default probabilities are used as inputs in several credit risk models, as well as in valuation adjustment calculations, and it is also a known fact that risk-neutral default probabilities are in principle different from their mid-point counterparts calculated using mid CDS market quotes as inputs. From a practical perspective, the CDS market can be considered as one of (if not the) most extensive repository of data that can be employed to strip out survival probabilities for a plethora of different entities. This is because of the contractual nature of CDS contracts, which facilitate the transfer of credit risk between the parties involved in these bilateral transactions. That is, a CDS is a financial derivative contract that allows an investor to offset their credit risk associated with a particular investment or debt instrument. In a CDS agreement, one party (i.e., the protection buyer) makes periodic payments to a second party (i.e., the protection seller) in exchange for protection against a credit event related to a given (typically debt-issuing) reference entity. If a (contract-specific) credit event for the reference entity materializes within a give time frame, as for instance the latter defaulting on their obligations, the protection seller is obliged to compensate the protection buyer; see White (2013). For this reason, CDSs can be also interpreted, in some sense, as insurance policies safeguarding against the occurrence of credit events like, amongst others, bankruptcies or credit rating downgrades, contingent upon the individual stipulations embedded within each contract. Therefore, CDS market quotes can be used in the estimation of risk-neutral default probabilities, and this can be achieved through a relatively straightforward procedure which has become, through time, the predominant approach in the industry following the standardization of CDS contracts implemented in the aftermath of the 2007–2008 financial crisis as documented, e.g., in Markit (2009a); Markit (2009b).

The central objective of Chapter 2 revolves around the development of an approach capable of computing risk-neutral default probabilities solely from the observable (i.e., quoted) bid and ask CDS prices available in the market. This method distinguishes itself by abstaining from the use of their mid-point counterparts as inputs, thereby guaranteeing full theoretical consistency with the risk-neutral pricing paradigm. In pursuit of this research goal, one must navigate through the intricacies of modeling financial markets where the conventional law of one price is not in place, thus shifting from a one-price economy to a two-price economy, while ensuring that the new modeling framework remains consistent with the risk-neutral valuation principles.

A theory that allows maintaining the formerly-stated consistency is that known as conic finance, first introduced in Cherny and Madan (2010) following the seminal paper Cherny and Madan (2009). This approach is rooted in the idea of reweighting the relevant risk-neutral measure through the application of (properly calibrated) concave distortion functions, depending on (at least) one parameter. The core premise behind this framework is to describe a financial market as an abstract entity where agents can engage in transactions in a direction-dependent manner. The prices so generated, i.e., denoting bids and asks, are computed by means of assessing the "expected performance" of the relevant financial contracts. This initial evaluation builds upon the concept of (static) *acceptability index*, i.e., a mathematical tool that can be used to weigh the "quality" of contingent claims based on their (expected) terminal payoff(s). From a practical standpoint, this involves calculating bid and ask prices for financial contracts adopting Choquet expectations Choquet (1954) (relative to a calibrated distorted probability measure, as already mentioned) as natural valuation functionals. Using distorted expectations leads to the loss of the additivity property typical of probability measures, thereby making non-additive measure theory (see, e.g., Denneberg (1994); Wang and Klir (2009)) the natural domain providing the tools to be used in this framework with the aim of modeling markets with frictions.

Within the settings of conic finance valuation, as expounded in Chapter 2, we present a methodology that offers a robust procedure for calculating risk-neutral default probabilities directly from bid and ask CDS market quotes. These risk-neutral default probabilities can subsequently be readily transformed into risk-neutral CDS quotes whenever the need arises. Our work is specifically tailored within the context of *reduced-form models* (as, e.g., Jarrow and Turnbull (1995) and Duffie and Singleton (1999)). That is, following the reduced-form modeling paradigm, we make the explicit modeling choice of specifying the functional form of the probability distribution function describing the behavior of the default time of a given CDS reference entity. Specifically, we adopt a well-established choice and opt for defining the default probability distribution function for a given entity by means of a Poisson process characterized by a piece-wise constant term structure. This choice aligns seamlessly with the standardized specifications outlined by the International Swaps and Derivatives

Association (ISDA) model, which is recognized as an industry benchmark in this field; see White (2013). That is, the choice made embodies an industry-accepted manner for characterizing the functional form of default probability distributions, thus ensuring alignment and compatibility with the current established market practices.

From a practical and computational perspective, the methodology outlined in Chapter 2 requires solving a non-linear system in two equations and two unknowns for each CDS maturity (given a selected reference entity), which can be easily achieved by means of standard numerical techniques. This requirement stems from the fundamental assumption (and, in practice, observation) that, for a given CDS transaction, both bid and ask prices can be observed. Consequently, the model must possess the capacity to exactly replicate both of these prices at the same time, thus giving rise to a set of two equations (which are non-linear due to the non-linearity of bid and ask prices according to the conic finance paradigm). In parallel, the selection of distortion functions featuring a single degree of freedom, such as the *minmaxvar* distortion Cherny and Madan (2009) or the *Wang* transform Wang (2000), results in the presence of one unknown hazard rate and one unknown distortion parameter for each CDS maturity. Consequently, this leads to the cumulative number of free parameters within the system aggregating to two. In Chapter 2 we prove that, for each CDS maturity, the aforementioned system inherently admits a unique solution under reasonable assumptions easily applicable to the CDS market, and computing its (maturity-dependent) solution is feasible through standard numerical techniques. It is also worth mentioning that, from a theoretical standpoint, the prerequisites for securing the uniqueness of the solution of the system are readily met in practical scenarios due to the liquidity levels of (single-name) CDS markets Junge and Trolle (2015). The methodology introduced in Chapter 2 additionally provides a mechanism for computing the term structure of the *implied liquidity* parameter of CDSs as a bonus effect, in the spirit of the approach outlined in Corcuera et al. (2012). This offers a straightforward tool for quantifying the liquidity in CDS markets, akin to the role of implied volatilities within the domain of option pricing.

The insights and results presented in Chapter 2 provide the basis for our further study elaborated in Chapter 3. This subsequent chapter widens the scope of what has been outlined in Chapter 2, extending the applicability of the liquidity-related results discussed previously within a broader financial context. Specifically, the content of Chapter 3 considers the scenario where the risk-neutral price of a contingent claim depends on a single unobservable parameter, assuming that the former strictly increases with respect to (positive) variations in the latter. Within this extended financial environment, appropriate liquidity levels and model flexibility allow to compute a unique value for the model parameter itself, as well as for the market liquidity value (i.e., distortion parameter), making it possible to replicate the observed bid and ask market prices. Once again,

this couple of parameters can be determined numerically by solving a two-by-two non-linear system, mirroring the approach described in Chapter 2. This allows the principles and considerations introduced in Chapter 2 to remain pertinent and adaptable as they find suitable application also in Chapter 3, *mutatis mutandis*.

The theoretical results of Chapter 3 have then a direct and very relevant application within option markets. In fact, the theory and observations available therein are directly applied to construct (risk-neutral) implied volatility smiles for (European) options on different underlying assets. This is possible as option prices are (strictly) monotonically increasing with respect to implied volatilities. This suggests, therefore, to investigate the differences between risk-neutral and mid implied volatilities, which is backed by a twofold reason. That is, on the one hand, implied volatilities are used, as inputs, in several pricing models. Further, on the other hand, implied volatilities are often used as explanatory variables in econometric analyses, as well as to explain specific behaviors observed in financial markets An et al. (2014); Cao et al. (2020); Doran et al. (2013); Fu et al. (2016); Sornette et al. (2018); Szakmary et al. (2003). The results obtained in Chapter 3 show how, close to the at-the-money point, risk-neutral and mid implied volatility smiles almost coincide. This is of course expected as there bid-ask spreads are in general tight and, thus, often negligible. Despite these considerations still hold for highly liquid options such as options on stock indices (in general) also far from the at-the-money point, changes unfold when less liquid underlyings are considered. In particular, the data illustrates how mid implied volatilities might largely overestimate or underestimate their risk-neutral counterparts, potentially influencing model outcomes, contingent upon the specific context considered.

The results from Chapter 2 serve as a solid foundation for the developments available in Chapter 3, and they offer both practical and theoretical insights in market liquidity. From a theoretical standpoint, the material contained in Chapters 2 and 3 suggests that, under reasonable assumptions, consistency with the risk-neutral paradigm in model calibration processes can be achieved by observing that modeling is not (always) strictly bound to using mid quotes for calibration purposes. This observation is backed by the practical ease and computational efficiency of deriving risk-neutral quotes and parameters from bid and ask prices if the assumptions on which the theory outlined in Chapters 2 and 3 builds are satisfied. From a practical angle, financial institutions can find value in this approach in the sense that they can evaluate whether the input data and, consequently, calibrated parameters used in pricing and risk models (amongst others) under risk-neutral settings require any kind of *ex post* adjustment. This might involve, for instance, the setting up of model reserves or addressing market price uncertainty, which are particularly relevant considerations in the current highly regulated financial landscape EBA (2015).

The third (and last) chapter of this first part of this thesis, i.e., Chapter 4, aims to explore innovations in machine learning within the area of pricing contingent claims in a two-price economy, again under the conic finance paradigm Cherny and Madan (2010). That is, in this research study we harness the ductility of neural networks (and, specifically, of *multilayer perceptrons*) to model liquidity dynamics in markets with frictions. This research encompasses two distinct, yet interrelated, aspects.

Firstly, we investigate the potential role neural networks can have with respect to replicating bid and ask (model) prices of financial derivatives. The results illustrate how this technology, if properly set up, is not only capable of leading to accurate valuation of contingent claims, but also to provide advancements in the efficient computation of their associated sensitivities, known as Greeks. Notably, this methodology is able to offer extraordinary computational speed outpacing classical methods such as (conic) Monte Carlo simulations by orders of magnitude, as well as to accurately replicate model prices generated by means of distorted (Choquet) expectations. Additionally, in this research a versatile approach for sensitivities calculations under conic assumptions is introduced and, moreover, we highlight that the work presented in Chapter 4 can accommodate different valuation models as soon as the latter can be used for valuation purposes under conic settings. Overall, this malleable and model-agnostic approach allows to enhance computational efficiency while retaining a high level of numerical precision in bid-ask derivatives pricing.

The research conducted in Chapter 4, furthermore, explores the possibilities to value contingent claims in a two-price economy in the context of LSV models, extending the modeling framework by means of machine leaning methods. That is, we leverage on Cuchiero et al. (2020), where the local volatility component of an LSV model is approximated through the flexible implementation of a (combination of) neural network(s). Thus, eliminating the need for a rigid parametric functional form for the Dupire's formula Dupire (1994) set *a priori*. In particular, we adapt the research work available in Cuchiero et al. (2020) to the context of bid-ask pricing under conic market assumptions, which not only results in an enhancement of the versatility and adaptability of pricing models in this area, but also contributes to the resilience of valuation frameworks against the technical challenges posed by local volatility dynamics. Moreover, we introduce the concept of *hybrid distortion function*, that straightforwardly allows to define arbitrary families of distortions, potentially improving model calibration accuracy. Besides, we introduce an extension of the SABR model Hagan et al. (2002) (namely, the *conic SABR model*), based on the Wang transform Wang (2000), which still allows for (approximated) analytical formulae for bid and ask (European) option prices.

In essence, this last research work aims to provide advancements that intertwine the domains of conic finance and machine learning, leveraging on ap-

plications of neural networks to enhance both accuracy and performances of valuation models in markets with frictions.

The second part (i.e., Part ii) of this thesis is dedicated to a comprehensive examination of *credit risk* in a context where limited CDS market data poses significant obstacles for the estimation of the risk of default of a wide range of counterparties. In Part ii we aim to provide different methodologies for the credit risk (expressed in terms of default probabilities) of counterparties when faced with the constraint of insufficient (or incomplete) CDS information. Our approach is intrinsically holistic, as we consider a broad spectrum of strategies in order to highlight the advantages and disadvantages linked to each of them by means of accurate analyses based on real market data. For our applications we mainly leverage on the theoretical concepts of Wasserstein distance and Wasserstein barycenter (see, amongst others, Agueh and Carlier (2011), (Panaretos and Zemel, 2020, Ch. 1 and 2), (Villani, 2003, Ch. 7) and (Villani, 2009, Ch. 6)) to capture the complex relationships among CDS data points and to better assess the inherent embedded creditworthiness from a probabilistic angle. By embedding an approach based on transport theory with the aim of measuring the likelihood of default, the methodologies outlined in the second part of the thesis provide alternative ways to measure and assess credit risk in the presence of CDS data constraints and/or market uncertanties. By using Wasserstein distances and Wasserstein barycenters, the results available in Part ii distinguish themselves from many of the currently-adopted techniques in the sense that they approach the estimation of credit risk under *sub-par* circumstances adopting a probability-driven rather than data-driven viewpoint.

The rationale behind the examination of alternative methodologies aimed to construct proxy credit curves based on CDS market quotes in a risk-neutral environment stems from the critical need of financial institutions to accurately assess and quantify the default risk of counterparties, which has major applications, e.g., in valuation adjustment calculations (Green, 2015, Ch. 4). Despite, from a theoretical perspective, it is easily possible to derive risk-neutral default probabilities from CDS market quotes (see also Chapter 2), the practical reality often reveals an insufficiency in the availability (and/or reliability) of CDS data for a substantial number of entities. This is particularly relevant in the case of financial institutions like banks, which often extend financing to small and medium sized enterprises and businesses. Furthermore, due to the intrinsic illiquidity of CDS markets, as observed in Junge and Trolle (2015), there is an increasing demand for the development of robust and resilient techniques that can serve as CDS "data generators", whenever required. These approaches are indispensable for situations where CDS quotes are unreliable or, even worse, simply not available. The challenge of generating trustworthy and stable CDS data in these circumstances has assumed paramount importance given its potential impact on credit risk assessment and management (and, from a banking perspective,

in valuation adjustment calculations). The comprehensive exploration available in Chapters 5 and 6 provides an in-depth analysis on possible alternatives to construct proxy credit curves when CDS data is either unreliable or scarce, highlighting both benefits and drawbacks of the different methodologies taken into consideration.

The most straightforward manner to develop a proxy credit curve methodology is that given by the so-called *intersection methodology* EBA (2013) (see also Chourdakis et al. (2013a) and Sourabh et al. (2018) for further details). This simple approach is based on the idea of first bucketing the available CDS quotes according to their characteristics (i.e., in terms of rating, region, sector, etc.), and to calculate their averages afterwards. However, simply performing data manipulation on CDS data clearly ignores the intrinsic probabilistic information embedded in the market quotes considered. In fact, as from CDS quotes it is possible to compute the default probabilities of the relevant reference entities (see also Chapter 2), it is reasonable to attempt to directly include probabilistic information within the proxy credit curve methodology as well.

In Chapter 5 the above-introduced intersection methodology is revised from a probabilistic perspective. In particular, therein the idea of using Wasserstein distances and Wasserstein barycenters for the purpose of constructing credit curves is investigated. This means that, instead of approaching the task of generating proxy credit curves purely from a data perspective, in Chapter 5 a technique based on optimal transport is considered, analyzed, and finally implemented. The reason why the usage of Wasserstein distances has been proposed finds its origin in its intuitive definition. Specifically, the Wasserstein distance allows by construction to measure the minimal amount of "work" that is necessary in order to redistribute the probability masses of one probability distribution for restoring a second one. Further, the choice of using Wasserstein distances for this purpose is also dictated by a very important and intuitive property they have, i.e., that of having geodesics which are shape-preserving. This means that, implicitly, Wasserstein distances do not ignore the shape of the distributions they deal with, while this is not always the case for other metrics between probability distributions such as the total variation distance (Villani, 2009, Ch. 6) and the (square) Hellinger distance (Le Cam, 1986, Ch. 4), to name but two.

The methodology proposed in Chapter 5 assumes that the default time linked to each CDS entity is derived from a Poisson process with a flat hazard rate term structure. This assumption, which is well known due to its applications in CDS quoting conventions (see White (2013)), allows to obtain analytical formulae for the (flat) hazard rates of the proxy curves which not only are trivial to compute, but which also carry a very intuitive interpretation. That is, the hazard rate derived from a bucket of CDS quotes can be easily calculated as the harmonic mean of the hazard rates stripped from the bucketed CDSs considered. This suggests that, in order to proxy credit curves, hazard rates should be averaged by using harmonic rather than arithmetic means. The results obtained by implementing

the methodology relying on real market data illustrate that failing to incorporate probabilistic information when proxying credit curves from CDSs might result in overestimating credit risk. This is in line with the fact that harmonic averages are bounded from above by their arithmetic counterparts.

On top of the intersection methodology there is another well-known approach to proxy credit curves from CDS quotes. This is known in the literature as *cross-section methodology*, initially introduced in Chourdakis et al. (2013a). This method aims to resolve a major drawback related to the intersection methodology, which is given by the fact that some buckets might contain only a handful of quotes or, in the worst case, be empty. This may therefore result in either proxy curves that are not faithful, that cannot be even computed given the lack of available data, or that do not preserve monotonicity with respect to credit ratings. The approach Chourdakis et al. (2013a) assumes that the (logarithms) of the quoted CDSs can be decomposed as a product of multiple components (i.e., parameters), where each of them is linked to a given characteristic of the relevant CDS contract (i.e., to its rating, region, sector, etc.). By performing a linear regression with categorical variables the jointly-calibrated parameters are then obtained. This allows to overcome (many of) the deficiencies of the intersection methodology highlighted above, including preserving monotonicity with respect to ratings (at least in most of the cases). That is, practical evidences suggest that the approach Chourdakis et al. (2013a) tends to guarantee that, if two proxy curves relate to the same region, sector, etc., but relate to different ratings, then higher ratings correspond to lower CDS proxy spreads, as it is natural to expect. We also recall that the cross-section methodology Chourdakis et al. (2013a) has been extended in Sourabh et al. (2018) by including equity returns and volatilities as regression quantities to enhance the explicative power of the independent variables.

Chapter 6 presents a hybrid calibration procedure aiming to combine the idea of an intersection methodology based on Wasserstein barycenters, as illustrated in Chapter 5, with that of the cross-sectional approaches Chourdakis et al. (2013a) and Sourabh et al. (2018). In particular, the proposal outlined in Chapter 6 is based on a two-step calibration routine. That is, first a bucketing process takes place and proxy CDS spreads are computed by means of the approach outlined in Chapter 5 for each of the relevant buckets. Afterwards, the cross section methodology is applied to the previously-calculated proxy spreads. Also in this case, in line with what has been observed in Chapter 5, the results seem to indicate that leaving out the probabilistic information hidden in (but easily retrievable from) the available CDS quotes to construct proxy credit curves might lead to overestimating default risk. Further, the research outcomes obtained empirically illustrate that the cross-section methodologies Chourdakis et al. (2013a) and Sourabh et al. (2018) provide two malleable frameworks to back out cross-sectional information across CDS quotes.

In summary, this thesis is structured into two parts, each dedicated to currently relevant domains of financial mathematics. In particular, the research conducted in this dissertation seeks to make a meaningful contribution to the field of quantitative finance with the auspices of equipping financial practitioners with some new and versatile tools and insights to address certain complexities of contemporary financial markets.

Part i embarks on a focused exploration of specific applications within the conic finance paradigm, aiming to address some challenges related to the intricate dynamics of bid and ask prices in markets with frictions. In particular, Chapter 2 resulted in Michielon et al. (2021a), Chapter 3 in Michielon et al. (2021b), while Chapter 4 in Michielon et al. (2023).

In Part ii, we undertake an in-depth analysis of credit risk assessment, employing Wasserstein distances and barycenters to uncover the latent probabilistic information hidden in CDS markets in order to deliver reliable and robust default probability computations based on an intuitive rationale. More precisely, Chapter 5 resulted in Michielon et al. (2022b), and Chapter 6 in Michielon et al. (2022a).

Each chapter of the thesis, potentially up to minor (and mainly formatting-related) modifications, is a direct reproduction of an either published or working paper. Therefore, each of the chapters is fully self-contained. For this reason, no dedicated effort has been made with the aim of keeping the notation consistent throughout the thesis at hand, this way guaranteeing consistency with the already published or submitted material available in this document.

Part I

APPLICATIONS IN MODELING FINANCIAL MARKETS UNDER CONIC FINANCE ASSUMPTIONS

2 FROM BID-ASK CREDIT DEFAULT SWAP QUOTES TO RISK-NEUTRAL DEFAULT PROBABILITIES USING DISTORTED EXPECTATIONS

2.1 ABSTRACT

Risk-neutral default probabilities can be implied from credit default swap (CDS) market quotes. In practice, mid CDS quotes are used as inputs, as their risk-neutral counterparts are not observable. We show how to imply risk-neutral default probabilities from bid and ask quotes directly by means of formulating the CDS calibration problem to bid and ask market quotes within the conic finance framework. Assuming the risk-neutral distribution of the default time to be driven by a Poisson process we prove, under mild liquidity-related assumptions, that the calibration problem admits a unique solution that also allows to jointly calculate the implied liquidity of the market.

2.2 INTRODUCTION

Risk-neutral default probabilities play a crucial role in modeling (counterparty) credit risk, as for instance in valuation adjustment calculations, and an approach that is often followed is that of computing them starting from credit default swap (CDS) market quotes. This article aims to relate risk-neutral default probabilities and CDS quotes in a two-price economy within the conic finance paradigm by means of providing a methodology to extract the former from bid and ask quotes directly, i.e., without relying on any mid price approximation. Before explaining our contribution in detail, we first provide a review of the relevant literature concerning bid-ask pricing and conic finance.

Bid-ask pricing can be modeled, in a consistent manner with risk-neutral valuation, in different manners. A possible way to do so is that of transforming the risk-neutral measure via appropriate concave distortion functions as per Cherny and Madan (2009). This approach, known as conic finance and introduced in Cherny and Madan (2010), is based on the idea of modeling illiquid markets as abstract entities accepting, at zero cost, a convex cone of random variables containing the non-negative cashflows. By balancing risks and rewards to assess

the "quality" or "expected performance" of contingent claims via the concept of *index of acceptability*, this framework allows to use Choquet expectations Choquet (1954) as building blocks for computing bid and ask prices.

The conic modeling framework, which has triggered extensive research and of which several applications are available in Madan and Schoutens (2016a), employs a static notion of index of acceptability, which allows to choose amongst cashflows, at the valuation date, based on their (cumulative) expected terminal value. This idea has been further extended by Bielecki et al. (2014) and later by Biagini and Bion-Nadal (2014), amongst others, to a dynamic setup, where *dynamic indices of acceptability* are defined in a multi-period setting. Dynamic acceptability indices allow to re-assess the initial classification of the traded cashflows on the basis of the latest information available, consistently over time, in the sense that future preferences are conforming with the current ones. This has lead to the possibility of pricing and hedging in a *dynamic conic finance* framework for finite probability spaces in a discrete-time setting as in Bielecki et al. (2013), where time-dependent bid and ask prices of contingent claims, potentially including dividends and transaction costs, are expressed in terms of dynamic indices of acceptability, and where a (dynamically-consistent) version of the First Fundamental Theorem of Asset Pricing is provided in terms of no-good-deal conditions. For a unified framework for the time-consistency between dynamic risk measures and dynamic performance measures in discrete time, see Bielecki et al. (2018), while a survey concerning the time-consistency property of dynamic risk and performance measures is available in Bielecki et al. (2017).

Security prices do not only depend on the direction of the transaction but also on the size of the order, and different approaches can be considered to include this additional liquidity charge in the relevant pricing equations. Bion-Nadal (2009) introduces, in continuous time, an approach providing dynamic bid and ask processes for contingent claims which include both the aforementioned liquidity effect, as well as transaction costs. By replacing scale-invariance with sub-scale-invariance Rosazza Gianin and Sgarra (2013) develop a dynamic framework, in a continuous-time setup and given a general probability space, that also captures this additional liquidity cost and that allows to value financial securities in terms of *g-expectations* Peng (1997) (comparisons between the definitions of g- and Choquet expectations are available in Chen et al. (2005) and Chen et al. (2013)). Again on general probability spaces, both liquidity and transaction costs can be included within a dynamic conic finance approach where pricing is based on g-expectations as in Bielecki et al. (2015).

Within the conic finance paradigm different studies deal with credit-related topics. Eberlein et al. (2012) show that, if assets and liabilities are marked at the bid and at the ask, respectively, then the potential accounting profitability of a firm induced by its own credit quality deterioration is eliminated. These ideas are further applied in Madan and Schoutens (2016b) in the case of credit and debit valuation adjustments. Madan (2014) proposes an approach to estimate the

parameters of risk acceptability of CDSs and their time dependence, and applies the methodology to a period including, but not limited to, the 2008 financial crisis. Therein, the industry practice of taking mid CDS quotes to proxy their risk-neutral counterparts is adopted and, for each CDS, a flat hazard rate term structure is considered in the calibration. Further, within the dynamic conic finance framework, bid and ask price processes for CDSs are constructed in Bielecki et al. (2013) and in Bielecki et al. (2015).

A methodology that allows to jointly calibrate a CDS model to bid and ask market quotes and to imply risk-neutral default probabilities without computing them from mid quotes is not yet available. In the present article we provide an approach to tackle this problem within the conic finance paradigm. The economic rationale behind our research question is given by the fact that, while in practice model parameters are usually calibrated starting from mid quotes as proxies for their risk-neutral counterparts, in reality a security trades neither at the risk-neutral nor at the mid price, but instead either at the bid or at the ask, depending on the direction of the trade. Thus, one might then want to be able to include in a simple manner the liquidity effect within their CDS pricing equations, for instance as CDS markets for single name CDSs not being amongst the most liquid; see Junge and Trolle (2015). Given our aim of extracting risk-neutral probabilities from the currently-observed bid and ask CDS quotes, a static approach to conic finance suffices. Moreover, this allows to easily define a term structure for the liquidity level of the CDS market, and also to restate the bid-ask calibration problem in terms of recursively solving a nonlinear constrained system. In the case of CDSs, modeling the default time via a reduced-form model by explicitly specifying the functional form of its distribution is a popular choice. In particular, we consider the case of a Poisson process driving the dynamics of the credit event, of which the standard International Swaps and Derivatives Association (ISDA) model White (2013), which is a common choice amongst financial practitioners, is a possible specification. In these settings, we show that the bid-ask CDS calibration process has, under some mild assumptions, a unique solution. Further, the methodology proposed here allows to jointly strip *implied liquidity* parameters for CDS markets in the spirit of Corcuera et al. (2012) with a term structure. To the best of our knowledge, this is the first attempt to calibrate a credit model using Choquet expectations to bid and ask CDSs quotes directly without relying on approximating their risk-neutral counterparts with the respective mid quotes and, thus, our contribution is novel.

This paper is organized as follows. In Section 2.3 we recall the basics of CDS valuation in the risk-neutral framework, how Poisson processes can be used to model the default time for CDS valuation purposes, as well as their calibration to market data. In Section 2.4 we briefly recall how pricing via distorted expectations works, while in Section 2.5, we introduce the CDS bid-ask calibration problem in the settings of Cherny and Madan (2010). We show that the problem

admits, under simple assumptions, a unique solution. We also provide a calibration example, based on the standard ISDA model, which is a special case of Poisson-based CDS model. Section 2.6 concludes.

2.3 BASIC NOTIONS AND VALUATION OF CDSS

A CDS is a bilateral derivative contract which involves the transfer of the credit risk arising from bonds or other forms of debt issued, amongst others, by corporates, municipalities, or sovereign states. Thus, a CDS is a sort of insurance policy, as it provides the protection buyer, who might or might not own the underlying credit[1], with protection against a *credit event*. The formal definition of a credit event is contract-specific and complex from the legal angle. Therefore, from here onward, the expressions *credit event* and *default* will be used to refer to a set of circumstances that trigger the protection payment.

A CDS contract involves two parties, i.e., a *protection buyer* and a *protection seller*. The protection seller commits to compensate for the (potential) loss of the counterparty if a default event for the reference entity occurs within a predetermined time frame. A CDS can be therefore seen as a derivative contract where the underlying is the default time of the issuing entity. CDSs were initially mainly physically settled: if default event occurs, then the protection buyer delivers one of the defaulted bonds of the reference entity to the protection seller, in return for its par value. However, due to the size of the CDS market it might happen that, should there not be enough supply of defaulted bonds in the market, an auction is conducted to determine what the recovery value of the defaulted bond is. In this case the CDS contract is, thus, cash settled, and this is nowadays the most common settlement practice (for further details refer to ISDA (2003)). The standardization process of credit derivatives led by the ISDA, see ISDA (1998); ISDA (2003); ISDA (2014), has introduced conventions on the way these contracts are traded. These conventions can be region specific: for example, some conventions for North-American CDSs (*CDS big bang*; see Markit (2009b)) might differ from those of European CDSs (*CDS small bang*; see Markit (2009a)). Before the CDS standardization process, in a similar fashion to interest rate swaps, CDSs used to be quoted at par, i.e., the coupon rate was defined such that the contract had zero value at inception, for both parties. However, CDSs have now standard coupons and, as a consequence, a non-zero entry cost called *upfront payment*, which is payed on the cash settlement date and that is usually quoted as a percentage (i.e., as *points upfront*) of the notional amount. The upfront payment can be interpreted as an amount reflecting the difference in value between a par CDS and one with a given standard coupon.

1 As CDSs do not require the buyer of the contract to hold the insured asset.

CDSs can be used to estimate the default probabilities of a wide range of issuing entities by means of appropriate pricing models calibrated using the available CDS market data. These implied probabilities can be used, for instance, as inputs in various valuation adjustment calculations, which makes CDS useful for hedging purposes; see (Green, 2015, Ch. 4 and Ch. 12).

We introduce now the essential notations and conventions that define CDSs; refer, for instance, to White (2013) for a detailed overview. In a standard CDS contract the *CDS dates* are the semi-annual termination dates of the CDS, and fall on 20 March and 20 December of each year.[2] From the *protection effective date* (i.e., t_p) the protection starts; this date is generally defined as the valuation date plus one day. The *cash settlement date* (i.e., t_s) is when any upfront payments are made, and can be lagged by a few business days compared to the valuation date (the standard ISDA model defines it as the valuation date plus three business days; see White (2013)). The *accrual start dates* (i.e., $s_1, \dots, s_N$) are used as starting points for calculating the coupon payments. This increasing sequence contains all the CDS dates before the maturity date, with s_1 set as the previous CDS date before the protection effective date. This is because holding a CDS over a coupon payment entails paying or receiving the full coupon payment amount. The *accrual end dates* (i.e., $e_1, \dots, e_N$) are the dates used as end points for calculating the premium payments, with e_N the maturity of the contract. Premium payments are made by the protection seller to the protection buyer at the *payment dates* (i.e., $t_1, \dots, t_N$). We denote with LGD the loss-given-default expressed per unit of notional (i.e., one minus the recovery rate), that we assume to be constant, and with $\Delta(t, s)$ the year fraction between t and s ($t < s$); see White (2013) for further details concerning day-count conventions. In particular, we use the shorthand notation Δ_i instead of $\Delta(s_i, e_i)$. Further, N denotes the notional amount, τ the default time, and $\mathbb{1}_{\{\cdot\}}$ the indicator function.

The protection leg is the contingent payment the protection seller makes to the protection buyer. Despite in practice there is usually a lag between the default time and the protection payment, modeling-wise at τ the protection seller is assumed to pay to the counterparty the amount

$$\text{LGD} \cdot N \cdot \mathbb{1}_{\{t_p \leqslant \tau \leqslant e_N\}}. \tag{1}$$

The premium leg is defined as the series of payments the protection buyer makes to the counterparty until either a credit event occurs or the contract expires. We denote its fixed coupon, per unit of notional, with C. The amount paid by the protection buyer to the protection seller on each payment date t_i is given by

$$C \cdot N \cdot \Delta_i \cdot \mathbb{1}_{\{\tau > e_i\}}. \tag{2}$$

2 Before 20 December 2015 the frequency of the CDS roll dates was quarterly instead of semi-annual, with resulting termination dates falling on 20 March, 20 June, 20 September and 20 December of each year; see ISDA (2015).

In the case of a credit event, the protection buyer pays to the counterparty the *accrued coupon*, i.e., if $\tau \in [s_i, e_i]$, the accrued coupon payed upon default equals

$$C \cdot N \cdot \Delta(s_i, \tau) \cdot \mathbb{1}_{\{s_i \leqslant \tau \leqslant e_i\}}. \tag{3}$$

On a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$, with $\mathbb{P}$ the real-word probability measure, we denote with $\mathbb{Q}$ a risk-neutral measure, with $\mathbb{E}^{\mathbb{Q}}(\,\cdot\,)$ the expectation, at valuation date, with respect to $\mathbb{Q}$, with $DF(t)$ the discount factor from t to valuation date, while with $\mathbb{Q}_S(t)$ the survival probability of the reference entity until time t, i.e., $\mathbb{Q}_S(t) \coloneqq \mathbb{Q}(\tau > t)$. From here onward we assume, without loss of generality, unit notionals.

For the protection buyer the value of a CDS equals the value of its protection leg minus the one of its premium leg. In symbols

$$PV^{prot} = LGD \cdot \mathbb{E}^{\mathbb{Q}}\left(DF(\tau) \cdot \mathbb{1}_{\{t_p \leqslant \tau \leqslant e_N\}}\right), \tag{4}$$

while the value of the premium leg is given by

$$PV^{prem} = C \sum_{i=1}^{N} \mathbb{E}^{\mathbb{Q}}\left(DF(t_i) \cdot \Delta_i \cdot \mathbb{1}_{\{\tau > e_i\}} + DF(\tau) \cdot \Delta(s_i, \tau) \cdot \mathbb{1}_{\{s_i \leqslant \tau \leqslant e_i\}}\right). \tag{5}$$

The present value of the CDS, from the perspective of the protection buyer, is defined as

$$PV^{CDS} \coloneqq PV^{prot} - PV^{prem}. \tag{6}$$

2.3.1 Dynamics of the survival probabilities

There are different approaches to model the dynamics of the default time of an issuing entity, and the category of *reduced-form* (or *intensity*) models is one of these. In reduced-form models the probability distribution of the credit event is modeled directly; two well-known illustrations of models belonging to this class are, amongst others, Jarrow and Turnbull (1995), where a discrete Poisson bankruptcy process is presented, and Duffie and Singleton (1999), where the risk-free discounting short-rate process is augmented with an instantaneous intensity process to account for credit risk. Reduced-form models are fundamentally different, for instance, from *structural* (or *firm-value*) models, which characterize defaults as consequences of events such as the value of a firm being too low for covering its liabilities, of which the so called Merton's 1974 firm-value model Merton (1974) is an illustration. This idea has been later extended in Black and Cox (1976), where default occurs when the value of the firm's asset falls below a given threshold level, and that is considered the first prototype of the so called *first passage time* models. While the main advantage of structural models is that of their consistency with the capital structure of the firm, they require

firm-specific information that is not necessarily easily available. Thus, the main difference between the reduced-form and the structural approaches is given by the fact that default is something exogenous in the former, while endogenous in the latter. The idea of modeling the default probability distribution directly as done in reduced-form models allows, at least in theory, to simplify the problem tractability, as modeling the default event per se is easier than modeling the economic situations that might cause it. This often makes reduced-form models preferable to structural ones for practical applications, as done in this article, given that we explicitly consider the CDS market as source of information. For more details concerning different approaches to credit risk modeling the reader can refer, amongst others, to Bielecki and Rutkowski (2010).

In the context of reduced-form models a possible approach to introduce a term structure for the distribution of the default time is that of defining the survival probability $\mathbb{Q}_S$ via

$$\mathbb{Q}_S(t) \coloneqq e^{-\int_0^t \lambda(s)\,ds}, \tag{7}$$

where the deterministic function $\lambda : [0, +\infty) \to (0, +\infty)$ is called *hazard rate* (or *default intensity*) function.

Assuming that K CDS quotes for a given reference entity with the same fixed coupon are available in the market, maturing respectively at $e_{N_1}, \ldots, e_{N_K}$ where $e_{N_1} < \ldots < e_{N_K}$, a possible way to define the hazard rate function with a term structure is given by setting

$$\lambda(t) \coloneqq \begin{cases} l_1(\lambda_1; t) & \text{if } 0 \leqslant t \leqslant e_{N_1} \\ l_2(\lambda_1, \lambda_2; t) & \text{if } e_{N_1} < t \leqslant e_{N_2} \\ \vdots & \\ l_K(\lambda_1, \ldots, \lambda_K; t) & \text{if } e_{N_{K-1}} < t \end{cases}, \tag{8}$$

where $\lambda_i > 0$, $l_i(\lambda_1, \ldots, \lambda_i; \cdot)$ is deterministic and continuous, and with the function $l_i(\lambda_1, \ldots, \lambda_{j-1}, \cdot, \lambda_{j+1}, \ldots, \lambda_i; t)$ increasing, for $1 \leqslant j \leqslant i$. The parameters $\lambda_1, \ldots, \lambda_K$ are those that, once set, specify the distribution of the default time. Common specifications for (8) are, among others, piecewise-constant and piecewise-linear.[3] The first option provides the simplest assumption possible concerning the behavior of the hazard rate function across CDS maturities, and it as well results in better numerical stability compared to its piecewise-linear counterpart; further, modeling the default time via this simple approach is often enough for practical applications such as for its usage in several credit valuation adjustment calculations; see (Green, 2015, Sec. 4.4). Moreover, note that (6) is

3 That is, given K positive values $\bar{\lambda}_1, \ldots, \bar{\lambda}_K$, in the piecewise-constant case, for every i, $l_i(\lambda_1, \ldots, \lambda_i; t) \equiv \bar{\lambda}_i$. For the piecewise-linear case, on the other hand, we have $l_1(\lambda_1; t) \equiv \bar{\lambda}_1$, $l_i(\lambda_1, \ldots, \lambda_i; t) = \bar{\lambda}_{i-1} + \frac{\bar{\lambda}_i - \bar{\lambda}_{i-1}}{e_{N_i} - e_{N_{i-1}}} \cdot (t - e_{N_{i-1}})$ for $2 \leqslant i \leqslant K - 1$, while $l_K(\lambda_1, \ldots, \lambda_K; t) \equiv \bar{\lambda}_K$.

a model-independent relationship which assumes interest rates being independent from the default time. In this case, see Brigo et al. (2010), CDS models can be calibrated to match CDS quotes exactly, and the resulting implied default probabilities calculated using different model specifications are expected to be in line with each other. In Brigo et al. (2010) this fact is illustrated by taking into account Lehman Brothers CDSs during different periods between August 2007 until the bank files for bankruptcy in September 2008. In particular, therein a comparison between the default probabilities implied using the Analytically-Tractable First Passage (AT1P) model and the intensity model with hazard rate function defined as per (8) in a piecewise-constant manner is provided (the AT1P model is a first-passage time structural model where default events are triggered by a firms' assets value hitting a deterministic threshold). The results show that the two models, despite their differences in terms of specifications, produce extremely close default probabilities (i.e., the largest difference observed at the calibration maturities is of the order of 0.8%). Hence, as we are interested in implying default probabilities at the valuation time, this further justifies the choice of the modeling approach we have followed: if little model risk is linked to the model specifications used to extract the default probabilities, then model simplicity and tractability should be encouraged.

To calibrate the model parameters, we denote with UF_i^{bid} (UF_i^{ask}) the bid (ask) upfront premium of the i^{th} quoted CDS contract. Their mid counterparts are denoted as UF_i^{mid}. The values $\lambda_1, \ldots, \lambda_K$ are computed to match the quoted CDS market values. Risk-neutral premia are not observable, and they are usually proxied with their mid counterparts.

Due to quoting convention, the first upfront premium is defined such that the equality

$$PV_1^{CDS}(\lambda_1) + Acc = DF(t_s) \cdot UF_1^{mid} \tag{9}$$

holds, where $PV_1^{CDS}(\lambda_1)$ denotes the present value of the first CDS, as a function of λ_1, and where Acc equals $DF(t_s) \cdot C \cdot \Delta(s_1, t_p)$. We can solve for $\lambda_1 > 0$ such that (9) is satisfied.[4]

Then, we can consider the second upfront premium. By using the value of λ_1 computed above, one can imply $\lambda_2 > 0$ such that

$$PV_2^{CDS}(\lambda_2) + Acc = DF(t_s) \cdot UF_2^{mid}, \tag{10}$$

4 The higher the values reached by the hazard rate function, the higher the chances are that there will be a default. Thus, the more the protection seller wants to be paid to sell insurance. One would then intuitively expect $PV_1^{CDS}(\lambda_1), \ldots, PV_K^{CDS}(\lambda_K)$ to be strictly increasing in $\lambda_1, \ldots, \lambda_K$, respectively. In 2.A we show that, for common coupon and LGD values, the value of the i^{th} CDS calculated using the setup outlined in the current section is strictly increasing in λ_i when $i > 1$, and that the same holds when $i = 1$, at least when λ_1 belongs to an interval wide enough for practical purposes. The strict monotonicity of $PV_i^{CDS}(\lambda_i)$ guarantees that, if $Acc - DF(t_s) \cdot UF_i^{mid} \in PV_i^{CDS}([0, +\infty))$, the equation $PV_i^{CDS}(\lambda_i) + Acc = DF(t_s) \cdot UF_i^{mid}$ admits a unique solution and, as a consequence, that the CDS calibration problem is well-defined. From here onwards we will always assume this to be the case.

with $PV_2^{CDS}(\lambda_2)$ the present value of the second CDS as a function of λ_2.

By proceeding inductively for the remaining indices this procedure allows to define a term structure for the default probabilities that is in line, via (8), with the mid quotes "observed" in the market.

Note that no choice of the hazard rate function in (7) comes without problems. For instance, the simple possible choice of assuming piecewise-constant hazard rates, which is at the base of the so called standard ISDA model, can produce negative hazard rates under specific market circumstances; see (Green, 2015, Ch. 4.4.3). Therefore, depending on the market conditions, some functional forms for the hazard rates can be more suitable than others. It is thus necessary to assume, for the chosen functional form of the hazard rate function, that risk-neutral quoted values allow the model to be properly specified, as well as the calibration problem under one-price settings to be successful.

2.4 BID-ASK PRICING VIA DISTORTED EXPECTATIONS

An index of acceptability is a map $\alpha : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \to [0, +\infty]$ aiming to measure the quality of random cashflows, i.e., for a given contingent claim X the higher the value of $\alpha(X)$, the higher X is ranked. We say that X is acceptable by the market at level γ whenever $\alpha(X) \geqslant \gamma$. An index of acceptability α is expected to satisfy some basic properties. Namely, if both X and X' are acceptable at level γ, then also $\lambda \cdot X + (1-\lambda) \cdot X'$ for $\lambda \in [0,1]$ is (quasi-concavity property). α is assumed to be monotonic, i.e., if $X \geqslant X'$ then $\alpha(X) \geqslant \alpha(X')$, as well as scale-invariant, i.e., $\alpha(\lambda \cdot X) = \alpha(X)$ for every $\lambda > 0$. Lastly, α is assumed to satisfy the Fatou property, which means that, if $(X_n)_n$ is a sequence of random variables such that, for every n, $|X_n| \leqslant 1$ and $\alpha(X_n) \geqslant \gamma$, then if $(X_n)_n$ converges in probability to a random variable X, also $\alpha(X) \geqslant \gamma$. It can be proven, see Cherny and Madan (2009), that given an index of acceptability α, for every $x \geqslant 0$ there exists a set $\mathfrak{Q}_x$ of probability measures absolutely continuous with respect to $\mathbb{P}$ such that

$$\alpha(X) = \sup \left\{ x \geqslant 0 : \inf_{\mathbb{Q} \in \mathfrak{Q}_x} \mathbb{E}^{\mathbb{Q}}(X) \geqslant 0 \right\} \tag{11}$$

and, further, if $x \leqslant x'$ then $\mathfrak{Q}_x \subseteq \mathfrak{Q}_{x'}$.

A *coherent risk measure* is a functional $\rho : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \to [0, +\infty]$ that satisfies the transitivity, sub-additivity, positively homogeneity and monotonicity properties; see (Madan and Schoutens, 2016a, Ch. 4.1).[5] It can be shown, see Delbaen

5 ρ is said to be transitive (or translation-invariant) when $\rho(X + \lambda) = \rho(X) + \lambda$ for every $\lambda \in \mathbb{R}$, sub-additive when $\rho(X + X') \leqslant \rho(X) + \rho(X')$, positively homogeneus when $\rho(\lambda \cdot X) = \lambda \cdot \rho(X)$ for every $\lambda > 0$, and monotonic when $\rho(X) \leqslant \rho(X')$ if $X \leqslant X'$. Note that the definition of coherent risk measure introduced in Artzner et al. (1999) differs from the one provided here in the sense that, in Artzner et al. (1999), cash-invariance reads $\rho(X + \lambda) = \rho(X) - \lambda$, where $\lambda \in \mathbb{R}$, while monotonicity as $\rho(X) \geqslant \rho(X')$ when $X \leqslant X'$ (refer to (Grabisch and Ridaoui, 2016, Sec.

(2002), that a coherent risk measure can be identified with a functional of the form $\sup_{Q\in\mathfrak{Q}} \mathbb{E}^{Q}(X)$, where $\mathfrak{Q}$ is a set of probability measures absolutely continuous with respect to $\mathbb{P}$. Therefore, the level of acceptability of a cashflow X can be rewritten in terms of coherent risk measures, i.e., as

$$\alpha(X) = \sup\{x \geqslant 0 : \rho_x(-X) \leqslant 0\}, \tag{12}$$

where $(\rho_x)_{x\geqslant 0}$ is a family of coherent risk measures such that $\rho_x(-X) \leqslant \rho_{x'}(-X)$ whenever $x \leqslant x'$. From this, it then follows that $\alpha(X) \geqslant \gamma$ if and only if $\rho_\gamma(-X) \leqslant 0$.[6] Note that, for every $x \geqslant 0$, one can define the *acceptability set* associated with α as $\mathcal{A}_x := \{X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) : \rho_x(-X) \leqslant 0\}$. It then follows that $(\mathcal{A}_x)_{x\geqslant 0}$ is a family of convex cones, each containing the non-negative random variables, with size decreasing in x. Thus, given an index of acceptability and a family of coherent risk measures $(\rho_x)_{x\geqslant 0}$, for every acceptability level x we obtain a convex cone $\mathcal{A}_x$ of contingent claims that are acceptable for the market, from which the term *conic finance* originates.

The (asymmetric) Choquet integral of X with respect to a *non-additive probability* μ is defined as

$$(C)\int_\Omega X\,d\mu := \int_{-\infty}^{0} \mu(X \geqslant t) - 1\,dt + \int_0^{+\infty} \mu(X \geqslant t)\,dt, \tag{13}$$

whenever it exists; see (Denneberg, 1994, Ch. 5). Choquet integration provides a natural extension to the Lebesgue integral able to deal with non-additive probabilities, as if μ in (13) is σ-additive, then (13) would reduce to a Lebesgue integral; see Mesiar et al. (2010).

We denote with $\psi(\,\cdot\,)$ a concave distortion from $[0,1]$ to $[0,1]$, i.e., a concave function such that $\psi(0) = 0$ and $\psi(1) = 1$, where $\psi(Q)(A) := \psi(Q(A))$, for every Q-measurable set A; note that the distorted probability measure just defined is not, in general, additive. Let $(\psi_x)_{x\geqslant 0}$ be an increasing family of concave distortion functions, and assume a risk-neutral measure $Q \in \bigcap_{x\geqslant 0} \mathfrak{Q}_x$.[7] We recall, see Delbaen (2002); Grabisch and Ridaoui (2016), that the functional ρ_x such that $X \mapsto (C)\int_\Omega X\,d\psi_x(Q)$ defines a coherent risk measure. This is because the (asymmetric) Choquet integral with respect to any non-additive measure guarantees the transitivity, positive homogeneity and monotonicity properties to be satisfied; see (Denneberg, 1994, Prop. 5.1). Further, the distorted probability measure $\psi_x(Q)$ is a submodular[8] set function, see (Denneberg, 1994, Ex. 2.1),

4.2.1) for some remarks concerning these differences). Given that we consider here coherent risk measures within the conic finance paradigm, we adopt therefore the definition outlined in Madan and Schoutens (2016a).

6 If $\rho_\gamma(-X) \leqslant 0$, by (12) it follows that $\alpha(X) \geqslant \gamma$. On the other hand, assume that $\rho_\gamma(-X) > 0$. Then, $\rho_x(-X) \geqslant \rho_\gamma(-X) > 0$ when $x \geqslant \gamma$, from which $\alpha(X) < \gamma$, contradiction.

7 Given that $\mathfrak{Q}_x \subseteq \mathfrak{Q}_{x'}$ when $x' \geqslant x$, it is sufficient to assume that a risk-neutral measure Q belongs to $\mathfrak{Q}_0$.

8 A non-additive probability μ is said to be submodular (or concave) if, for every μ-measurable sets A and A', it results that $\mu(A \cup A') + \mu(A \cap A') \leqslant \mu(A) + \mu(A')$.

which guarantees subadditivity; see (Denneberg, 1994, Th. 6.3). Thus, as suggested in Cherny and Madan (2009), we can employ functionals of this form as tools for modeling indices of acceptability via the relationship

$$\alpha(X) = \sup\left\{x \geqslant 0 : (C)\int_{\Omega} -X \, d\psi_x(\mathbb{Q}) \leqslant 0\right\}. \tag{14}$$

Indices of acceptability defined as in (14) are named *operational indices of acceptability*; see Cherny and Madan (2009).

We now assume that the market considers acceptable only the cashflows with an acceptability level of, at least, γ. The market is willing to buy X, which we assume to pay off at T, at a price b if and only if $\alpha(X - \mathrm{DF}(T)^{-1} \cdot b) \geqslant \gamma$ (recall footnote 6), i.e., if and only if $b \leqslant -\mathrm{DF}(T) \cdot (C)\int_{\Omega} -X \, d\psi_\gamma(\mathbb{Q})$. It follows that, if the market considers acceptable all the cashflows with a level of acceptability of at least γ, then the (γ-dependent) bid price of X, denoted as $\mathrm{bid}_\gamma(X)$, would equal

$$\mathrm{bid}_\gamma(X) = -\mathrm{DF}(T) \cdot (C)\int_{\Omega} -X \, d\psi_\gamma(\mathbb{Q}). \tag{15}$$

Denoting the ask price of X given a level of acceptability γ as $\mathrm{ask}_\gamma(X)$, by observing that $\mathrm{ask}_\gamma(X) = -\mathrm{bid}_\gamma(-X)$, from (15) it follows that

$$\mathrm{ask}_\gamma(X) = \mathrm{DF}(T) \cdot (C)\int_{\Omega} X \, d\psi_\gamma(\mathbb{Q}). \tag{16}$$

Thus, if the distribution function of X, as well as its bid or ask prices, are available, one can compute the level of γ needed to obtain the quoted price.

2.5 CDSS IN A TWO-PRICE ECONOMY

Given a parametric family of distortion functions $(\psi_\gamma)_{\gamma \geqslant 0}$, one can set a term structure for the liquidity parameter γ by assigning a value γ_i to each maturity e_{N_i}. These values can be then interpolated, once the model has been calibrated, if one wants to calculate bid and ask prices for non-quoted maturities. We still assume a Poisson process as in Section 2.3.1 driving the risk-neutral dynamics of τ. We denote with $\tilde{X}_i^{\mathrm{CDS}}$ the sum of the cashflows of the i^{th} CDS where all the cashflows are deferred to maturity, i.e., $\mathrm{cashflow}(t) \mapsto \mathrm{cashflow}(t) \cdot \frac{\mathrm{DF}(t)}{\mathrm{DF}(e_{N_i})}$. The bid and ask prices of $\tilde{X}_i^{\mathrm{CDS}}$ are denoted as $\mathrm{bid}_i^{\mathrm{CDS}}$ and $\mathrm{ask}_i^{\mathrm{CDS}}$, respectively.

We start with the first CDS. We need then to solve for $\lambda_1 > 0$ and $\gamma_1 > 0$ such that

$$\begin{cases} \mathrm{bid}_1^{\mathrm{CDS}}(\lambda_1, \gamma_1) + \mathrm{Acc} = \mathrm{DF}(t_s) \cdot \mathrm{UF}_1^{\mathrm{bid}} \\ \mathrm{ask}_1^{\mathrm{CDS}}(\lambda_1, \gamma_1) + \mathrm{Acc} = \mathrm{DF}(t_s) \cdot \mathrm{UF}_1^{\mathrm{ask}} \end{cases}, \tag{17}$$

with

$$\mathrm{DF}(t_s)\cdot \mathrm{UF}_1^{\mathrm{bid}} < \mathrm{PV}_1^{\mathrm{CDS}}(\lambda_1) + \mathrm{Acc} < \mathrm{DF}(t_s)\cdot \mathrm{UF}_1^{\mathrm{ask}}, \tag{18}$$

where the constraint (18) guarantees that the risk-neutral price of the CDS lies between its corresponding bid and ask prices.

By proceeding in a similar way as done in Section 2.3.1, at every step we need to solve a system of the form

$$\begin{cases} \mathrm{bid}_i(\lambda_i,\gamma_i) + \mathrm{Acc} = \mathrm{DF}(t_s)\cdot \mathrm{UF}_i^{\mathrm{bid}} \\ \mathrm{ask}_i(\lambda_i,\gamma_i) + \mathrm{Acc} = \mathrm{DF}(t_s)\cdot \mathrm{UF}_i^{\mathrm{ask}} \end{cases}, \tag{19}$$

with

$$\mathrm{DF}(t_s)\cdot \mathrm{UF}_i^{\mathrm{bid}} < \mathrm{PV}_i^{\mathrm{CDS}}(\lambda_i) + \mathrm{Acc} < \mathrm{DF}(t_s)\cdot \mathrm{UF}_i^{\mathrm{ask}}. \tag{20}$$

Above, λ_i represents the implied hazard rate for the i^{th} maturity, while γ_i the corresponding *implied liquidity* in the sense of Corcuera et al. (2012).

The problem of determining whether a (potentially unique) solution for this constrained non-linear system will be addressed in this section. We start by simplifying the notation in the constrained system above by rewriting it as

$$\begin{cases} \mathrm{bid}(\lambda,\gamma) = b \\ \mathrm{ask}(\lambda,\gamma) = a \end{cases}, \tag{21}$$

with

$$b < \mathrm{PV}^{\mathrm{CDS}}(\lambda) < a. \tag{22}$$

In both (21) and (22) we have set $b \coloneqq \mathrm{DF}(t_s)\cdot \mathrm{UF}_i^{\mathrm{bid}} - \mathrm{Acc}$ and $a \coloneqq \mathrm{DF}(t_s)\cdot \mathrm{UF}_i^{\mathrm{ask}} - \mathrm{Acc}$.

We provide now three lemmas that, under some mild assumptions related to the liquidity of the market, will be used in Theorem 1 to prove the existence and the uniqueness of a solution for the constrained non-linear system (21). We start by assuming that the quoted bid and ask prices of the chosen CDS are within the interval of possible risk-neutral prices that can be obtained by changing the parameter λ. In practice, this technical condition translates into the possibility of being able to calibrate the risk-neutral parameter λ to match bid and ask market quotes, respectively, from which Lemma 1 follows.

Assumption 1. *The inequalities* $\inf_{\lambda>0} \mathrm{PV}^{\mathrm{CDS}}(\lambda) < b$ *and* $\sup_{\lambda>0} \mathrm{PV}^{\mathrm{CDS}}(\lambda) > a$ *hold.*

Lemma 1. *Under Assumption 1, there exists an interval* $[\lambda_b,\lambda_a]$ *such that there is equivalence between* $b \leqslant \mathrm{PV}^{\mathrm{CDS}}(\lambda) \leqslant a$ *and* $\lambda \in [\lambda_b,\lambda_a]$.

Proof. $\mathrm{PV}^{\mathrm{CDS}}(\lambda)$ is an increasing and continuous function of λ. From Assumption 1 the result follows. □

We now introduce a second assumption that guarantees that, for λ in a given range, theoretical bid-ask spreads can exceed the observed one.[9] Intuitively, Assumption 2 is a technical condition stating that, for every fixed λ in $[\lambda_a, \lambda_b]$, it is always possible to construct a bid and an ask price that reflect the bid-ask spread observed in the market. Lemmas 2 and 3 follow.

Assumption 2. *For every* $\lambda \in [\lambda_b, \lambda_a]$ *there exists* $\gamma > 0$ *such that* $\text{ask}(\lambda, \gamma) - \text{bid}(\lambda, \gamma) = a - b$.

Lemma 2. *Under Assumptions 1 and 2, for every* $\lambda \in [\lambda_b, \lambda_a]$ *there exists a unique* $\gamma > 0$ *such that* $\text{ask}(\lambda, \gamma) - \text{bid}(\lambda, \gamma) = a - b$.

Proof. Fix $\lambda \in [\lambda_b, \lambda_a]$. By Assumption 2 there exists (at least) one $\gamma > 0$ such that $\text{ask}(\lambda, \gamma) - \text{bid}(\lambda, \gamma) = a - b$. Assume there exists γ_* and γ^* such that $\text{ask}(\lambda, \gamma_*) - \text{bid}(\lambda, \gamma_*) = \text{ask}(\lambda, \gamma^*) - \text{bid}(\lambda, \gamma^*) = a - b$, with $\gamma_* < \gamma^*$. The ask price is an increasing function of γ, while the opposite holds for the bid. Therefore, $\text{ask}(\lambda, \gamma_*) < \text{ask}(\lambda, \gamma^*)$ and $\text{bid}(\lambda, \gamma_*) > \text{bid}(\lambda, \gamma^*)$. Then $a - b = \text{ask}(\lambda, \gamma_*) - \text{bid}(\lambda, \gamma_*) < \text{ask}(\lambda, \gamma^*) - \text{bid}(\lambda, \gamma^*) = a - b$, contradiction. □

Lemma 3. *Under Assumptions 1 and 2, for every* $\lambda \in [\lambda_b, \lambda_a]$ *the function such that* $\lambda \mapsto \gamma(\lambda)$, *where* $\text{ask}(\lambda, \gamma(\lambda)) - \text{bid}(\lambda, \gamma(\lambda)) = a - b$, *is continuous.*

Proof. Fix $\bar{\lambda}$ in $[\lambda_b, \lambda_a]$ and let $(\lambda_n)_n$ be a sequence in $[\lambda_b, \lambda_a]$ that converges to $\bar{\lambda}$. We define $\phi(\lambda, \gamma) \coloneqq \text{ask}(\lambda, \gamma) - \text{bid}(\lambda, \gamma)$. We proceed in steps.

Claim (i): The sequence $(\gamma(\lambda_n))_n$ is bounded. Say this is not the case. Then, there exists a subsequence $(\gamma(\lambda_{n_k}))_k$ that diverges to $+\infty$. $(\lambda_{n_k})_k$ converges to $\bar{\lambda}$, as subsequence of a convergent sequence, and ϕ is continuous in both arguments. Therefore, $\lim_k \phi(\bar{\lambda}_{n_k}, \gamma(\lambda_{n_k})) = \phi(\bar{\lambda}, +\infty) = a - b$, as $\phi(\bar{\lambda}_{n_k}, \gamma(\lambda_{n_k}))$ always equals $a - b$, by construction. By Assumption 2 there exists $\bar{\gamma} > 0$ such that $\phi(\bar{\lambda}, \bar{\gamma}) = a - b$. Therefore, as ϕ is increasing in its second argument, it follows that $a - b = \phi(\bar{\lambda}, \bar{\gamma}) < \phi(\bar{\lambda}, +\infty) = a - b$, contradiction.

Claim (ii): The sequence $(\gamma(\lambda_n))_n$ has limit. As this sequence is bounded, it admits a convergent subsequence. Say there are two subsequences, namely $(\gamma(\lambda_{n_k}))_k$ and $(\gamma(\lambda_{n_h}))_h$, that converge to γ_* and γ^*, respectively, where $\gamma_* < \gamma^*$. Then $(\lambda_{n_k})_k$ and $(\lambda_{n_h})_h$ both converge to $\bar{\lambda}$, as subsequencies of the same

9 Observe that, when $\gamma = 0$, then bid and ask prices reduce to the ones calculated with respect to $\mathbb{Q}$ and that, for a given λ, the function $\text{ask}(\lambda, \gamma) - \text{bid}(\lambda, \gamma)$ is strictly increasing in γ. When $\gamma \to +\infty$, then $\psi_\gamma(\mathbb{Q})$ approximate the distribution that assigns zero to the null sets and one to any other set. $\text{ask}(\lambda, \gamma) - \text{bid}(\lambda, \gamma)$ can be rewritten, see Section 2.5.1, as $\text{ask}^{\text{prot}}(\lambda, \gamma) - \text{bid}^{\text{prem}}(\lambda, \gamma) - \text{bid}^{\text{prot}}(\lambda, \gamma) + \text{ask}^{\text{prem}}(\lambda, \gamma)$, where the superscripts identify the two legs of the contract. Ignoring discount factors for simplicity, $\text{ask}^{\text{prot}}(\lambda, \gamma)$ has magnitude of the order of $\text{LGD} \cdot \psi_\gamma(\mathbb{Q})(t_p \leqslant \tau \leqslant e_N) = \text{LGD}$; see (4). Further, as $(C)\int_\Omega -X \, d\mu = (C)\int_\Omega X \, d\bar{\mu}$ with $\bar{\mu}$ denoting the dual measure of μ, see (Denneberg, 1994, Prop. 5.1), from $\text{bid}(X) = -\text{ask}(-X)$ it follows that $\text{bid}^{\text{prot}}(\lambda, \gamma) = -\text{LGD} \cdot \overline{\psi_\gamma(\mathbb{Q})}(t_p \leqslant \tau \leqslant e_N) = 0$. Therefore, for extreme values of γ the theoretical bid-ask spread $\text{ask}(\lambda, \gamma) - \text{bid}(\lambda, \gamma)$ reaches high values due to its positive components $\text{ask}^{\text{prot}}(\lambda, \gamma)$ and $\text{ask}^{\text{prem}}(\lambda, \gamma)$, and to $\text{bid}^{\text{prem}}(\lambda, \gamma)$ being below its counterpart calculated when $\gamma = 0$. Thus, for practical purposes, this assumption is in general satisfied.

convergent sequence. So we obtain that $a - b = \lim_k \phi(\lambda_{n_k}, \gamma(\lambda_{n_k})) = \phi(\bar{\lambda}, \gamma_*) < \phi(\bar{\lambda}, \gamma^*) = \lim_h \phi(\lambda_{n_h}, \gamma(\lambda_{n_h})) = a - b$, contradiction (the first and the last equalities follow from the definitions of $(\lambda_{n_k})_k$ and $(\lambda_{n_h})_h$, respectively, the second and the penultimate equalities from the continuity of ϕ, while the inequality from ϕ being increasing in its second argument). Then, every convergent subsequence of $(\gamma(\lambda_n))_n$ has the same limit. As $(\gamma(\lambda_n))_n$ is bounded, then it admits limit.[10]

Claim (iii): The limit of $(\gamma(\lambda_n))_n$ is $\gamma(\bar{\lambda})$. Denote $\lim_n \gamma(\lambda_n)$ as $\bar{\gamma}$. Observe that ϕ is continuous in both arguments. Furthermore, the sequence $(\phi(\lambda_n, \gamma(\lambda_n)))_n$ is constant by construction, i.e., it always equals $a - b$. Therefore, it converges to $a - b$. Its limit is $\phi(\bar{\lambda}, \bar{\gamma})$, as ϕ is continuous. Due to Lemma 2, there exists a unique $\gamma(\bar{\lambda})$ such that $\phi(\bar{\lambda}, \gamma(\bar{\lambda})) = a - b$. So, $\bar{\gamma} = \gamma(\bar{\lambda})$. □

We now can, under Assumptions 1 and 2, use Lemmas 1, 2 and 3 to prove that the calibration problem (21) has a unique solution. Therefore, Theorem 1 guarantees that, under the hypotheses considered, risk-neutral default probabilities can be implied in a unique manner from bid and ask CDS quotes without relying on their mid counterparts.

Theorem 1. *Under Assumptions 1 and 2, there exists a solution of the constrained non-linear system* (21)*, and it is unique.*

Proof. Consider the interval $[\lambda_b, \lambda_a]$ as per Lemma 1. There exists a unique γ_b such that $\text{ask}(\lambda_b, \gamma_b) - \text{bid}(\lambda_b, \gamma_b) = a - b$. Observe that $\text{bid}(\lambda_b, \gamma_b) < \text{PV}^{\text{CDS}}(\lambda_b) = b$, so $b < \text{ask}(\lambda_b, \gamma_b) < a$.

Similarly, consider λ_a. There exists a unique γ_a such that $\text{ask}(\lambda_a, \gamma_a) - \text{bid}(\lambda_a, \gamma_a) = a - b$. Because $a = \text{PV}^{\text{CDS}}(\lambda_a) < \text{ask}(\lambda_a, \gamma_a)$, it follows that $b < \text{bid}(\lambda_a, \gamma_a) < a$.

The functions $\text{ask}(\lambda, \gamma)$, $\text{bid}(\lambda, \gamma)$, and – see Lemma 3 – $\gamma(\lambda)$, are continuous in λ. Thus, there exists $\bar{\lambda} \in (\lambda_b, \lambda_a)$ and corresponding $\bar{\gamma}$ such that $\text{ask}(\bar{\lambda}, \bar{\gamma}) = a$ and $\text{bid}(\bar{\lambda}, \bar{\gamma}) = b$. By virtue of Lemma 2 the pair $(\bar{\lambda}, \bar{\gamma})$ satisfying (21) is unique. □

Note that to obtain the existence and uniqueness result of Theorem 1 we have relied on the fact that, for each given maturity, the model describing the risk-neutral default distribution has a single free parameter, i.e., the hazard rate corresponding to the maturity considered. Therefore, considering the distortion parameter related to that maturity as additional degree of freedom allows the calibration problem to be defined, up to the constraint, by two equations and two unknowns. If more complex models with additional parameters were to be used, then the problem should have been approached in a least-square sense,

10 Here, we have used the following elementary result: if a bounded real sequence has the property that all its convergent subsequences converge to the same real limit, then the sequence itself also converges to it; see (Abbott, 2015, Ex. 2.5.5).

and the best possible outcome would have been that of finding an unique minimum. This very favorable situation, however, would not necessarily guarantee observed market quotes to be matched by the model, and therefore would as well not guarantee the implied risk-neutral distribution to be the "true" one.

2.5.1 A calibration example

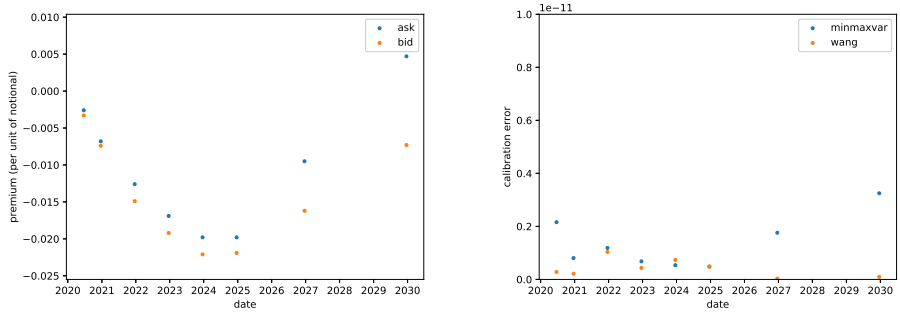
The simplest possible manner to specify model (8) consists in defining it as a piecewise-constant function, as done in the ISDA CDS standard model commonly used in practice, which is based on the approach of O'Kane and Turnbull (2003). Note that there is no information available on the hazard rate level between CDS maturities. Therefore, these specifications provide the smallest possible set of assumptions concerning the default intensity process and is a common choice amongst financial practitioners.

To compute bid and ask prices, one would need to approximate Choquet integrals numerically. In order to do so, a simple approximation of $(C)\int_\Omega X\,d\mu$ can be performed, see (Wang and Klir, 2009, Ch. 11.5), as follows. Given a partition of Ω as $\bigcup_{i=1}^{M} A_i$ choose, for every i, $x_i \in X(A_i)$. Let σ denote a permutation of $\{1,\dots,M\}$ such that $x_{\sigma(1)} \leqslant \dots \leqslant x_{\sigma(M)}$. Then, $(C)\int_\Omega X\,d\mu$ can be then approximated as

$$\sum_{i=1}^{M} (x_{\sigma(i)} - x_{\sigma(i-1)}) \cdot \mu\left(\bigcup_{k=i}^{M} A_k\right), \tag{23}$$

where $x_0 \coloneqq 0$. In the case of a CDS, one can then set a grid (for instance, daily), namely $A_1 \coloneqq \{\tau \in [0, d_1]\}, \dots, A_M \coloneqq \{\tau \in [d_{M-1}, d_M]\}$, where M denotes the total number of points (i.e., dates) in the grid, and set $x_i \coloneqq \tilde{X}^{\mathrm{CDS}}|_{\tau=d_i}$ (recall that, using the notation introduced in Section 2.5, the superscript tilde indicates that cashflows are deferred at the maturity of the CDS contract considered).

We recall, see Eberlein et al. (2012), that the bid and ask prices of a contingent claim X can be calculated as $\mathrm{bid}(X) = \mathrm{bid}(X^+) - \mathrm{ask}(X^-)$ and $\mathrm{ask}(X) = \mathrm{ask}(X^+) - \mathrm{bid}(X^-)$, respectively, where the X^+ (X^-) denotes the positive (negative) part of X. We denote with $\tilde{X}^{\mathrm{prot}}$ ($\tilde{X}^{\mathrm{prem}}$) the protection (premium) leg of $\tilde{X}^{\mathrm{CDS}}$. From (6), and by noting that with our conventions X^+ coincides with $\tilde{X}^{\mathrm{prot}}$ and X^- with $\tilde{X}^{\mathrm{prem}}$, it follows that $\mathrm{bid}(\tilde{X}^{\mathrm{CDS}}) = -\mathrm{ask}(-\tilde{X}^{\mathrm{prot}}) - \mathrm{ask}(\tilde{X}^{\mathrm{prem}})$ and that $\mathrm{ask}(\tilde{X}^{\mathrm{CDS}}) = \mathrm{ask}(\tilde{X}^{\mathrm{prot}}) + \mathrm{ask}(-\tilde{X}^{\mathrm{prem}})$. Therefore, in principle it is sufficient to separately calculate the ask prices of the (signed) CDS legs only.



(a) (b)

Figure 1: Bid and ask CDS upfront premia used in the calibration example, panel (a), and calibration errors, panel (b). In particular, the calibration errors of panel (b) represent aggregated figures, i.e., each point corresponds to the sum of the calibration error for the bid quote and of that of the related ask quote.

As an example, we consider the specifications of the standard ISDA model, i.e., we assume a piecewise-constant hazard rate function. We take into account a set of market quotes for a BBB European financial institution with maturities 6 months and 1, 2, 3, 4, 5, 7 and 10 years, respectively, as of 13 February 2020. The recovery rate equals 40%, and the coupon 1%. Discounting performed with OIS EUR curve.

Figure 1a represents the bid and ask quoted upfront premia, expressed per unit of notional, while Figure 1b the aggregated calibration errors, i.e., each value represents the sum of the bid and ask calibration errors, respectively.

In this example we consider two common choices to define the family of distortion function, i.e., the *minmaxvar* distortion Cherny and Madan (2009), defined via

$$\psi_\gamma(x) \coloneqq 1 - \left(1 - x^{\frac{1}{1+\gamma}}\right)^{1+\gamma}, \tag{24}$$

and the *Wang* distortion Wang (2000), defined by setting

$$\psi_\gamma(x) \coloneqq \Phi\left(\Phi^{-1}(x) + \gamma\right), \tag{25}$$

with $\Phi(\cdot)$ denoting the cumulative distribution function of a standard normal random variable; in both (24) and (25) it is assumed that $x \in [0,1]$ (in the case of the latter, right and left limit should be considered for 0 and 1, respectively) and that $\gamma \geqslant 0$. Other examples of families of distortion functions are outlined, for instance, in (Madan and Schoutens, 2016a, Ch. 4.7) and in (Föllmer and Schied, 2016, Ch. 4.6). For each of the two choices we have made in terms of the distortion function, Figure 2a shows the piecewise-constant hazard rate function, while Figure 2b the linearly interpolated distortion parameter.

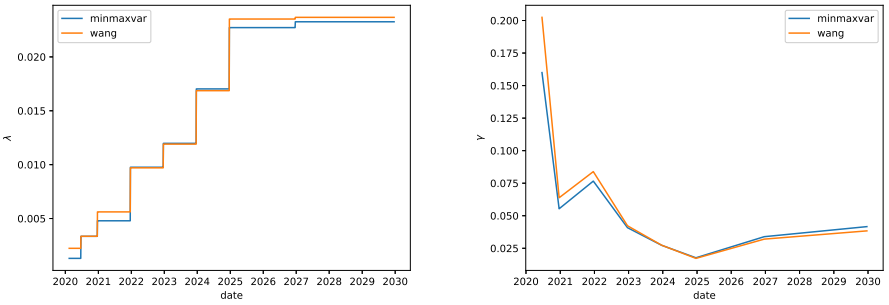


Figure 2: Implied parameters provided as a result of the calibration procedure to the bid and ask CDS quotes of Figure 1a: hazard rates (i.e., λ) are depicted in panel (a), while distortion parameters (i.e., γ) in panel (b).

For each of the two choices of the distortion function we have made, the minimum of the γ parameter in Figure 2b lies in proximity of the 5Y CDS, where the latter is usually the most liquid maturity. We also note how the pattern of the implied distortion parameter in Figure 2b follows that of the (relative) bid-ask CDS premium spread available in Table 1 (last column therein). Note that, as illustrated by Figures 2a and 2b results obtained using the minmaxvar and Wang transforms are very similar, indicating little model risk. For completeness, the CDS input quotes of Figure 1a, the implied hazard rates of Figure 2a and the implied distortion parameters of Figure 2b have been reported in 2.B; see Tables 1, 2a and 2b, respectively.

2.6 CONCLUSION

In this article we have considered the problem of calibrating a CDS model to the available bid and ask quotes within the conic finance paradigm of Cherny and Madan (2010). In particular, in the context of reduced-form models, we have considered the default time as modeled by a Poisson process. The bid-ask calibration problem requires to iteratively solve a constrained non-linear system in two equations and two unknowns. We have showed that, under reasonable assumption for practical purposes, the calibration problem admits a unique solution. We have as well illustrated, with a practical example based on real market data, how the calibration to bid and ask CDS quotes works under the specifications of the standard CDS ISDA model and by considering two different choices for the the distortion function. In both the cases considered, as expected from

the theory, the model could be calibrated to exactly match the observed market quotes. Despite our work outlined in Section 2.5 is specific to CDSs, the fact that financial instruments trade neither at the risk-neutral nor at the mid price apply to all contingent claims. Therefore, being able to fit valuation models solely to bid and ask quotes in such a way that risk-neutral parameters are implied as a result of the calibration routines is a desirable model feature that allows to drop the common assumption of equating risk-neutral and mid prices without additional ones being added. Hence, investigating how to calibrate models to bid and ask quotes without relying on mid quote approximations is a field on which further research is encouraged.

2.A A REMARK ON THE MONOTONICITY OF CDS PRICES

We consider here the i^{th} CDS outlined in Section 2.3.1, i.e., the one maturing at e_{N_i}, and we denote with $N(i)$ the number of coupon periods related to it.

From (5), the present value of its premium leg can be rewritten as

$$C \sum_{j=1}^{N(i)} \left(DF(t_j) \cdot \Delta_j \cdot \mathbb{Q}_S(e_j) + \mathbb{E}^{\mathbb{Q}} \left(DF(\tau) \cdot \Delta(s_j, \tau) \cdot \mathbb{1}_{\{s_j \leqslant \tau \leqslant e_j\}} \right) \right). \tag{26}$$

We define $j(i) := \min \left\{ j : e_j > e_{N_{i-1}},\ 1 \leqslant j \leqslant N(i) \right\}$, with the convention that $j(i) = 1$ if $i = 1$. If λ_i increases, from (7) and (8) it follows that $\mathbb{Q}_S(e_j)$ strictly decreases for each $j \geqslant j(i)$, leaving the others, if any, unchanged. We can also rewrite the present value of the protection leg, see (4), minus the accrual payments in (26), as

$$\begin{aligned} \sum_{j=1}^{N(i)} \mathbb{E}^{\mathbb{Q}} \left(DF(\tau) \cdot \big(LGD - C \cdot \Delta(s_j, \tau) \big) \cdot \mathbb{1}_{\{\max(t_p, s_j) \leqslant \tau \leqslant e_j\}} \right) \\ - \mathbb{E}^{\mathbb{Q}} \left(DF(\tau) \cdot C \cdot \Delta(s_1, \tau) \cdot \mathbb{1}_{\{0 \leqslant \tau \leqslant t_p\}} \right), \end{aligned} \tag{27}$$

due to $s_j = e_{j-1}$ whenever $j > 1$.

If $i > 1$, when λ_i increases then $\mathbb{Q}(\max(t_p, s_j) \leqslant \tau \leqslant e_j)$ strictly increases for each $j \geqslant j(i)$, leaving the other probabilities with $j < j(i)$, as well as $\mathbb{Q}(0 \leqslant \tau \leqslant t_p)$, unchanged. Thus, if the condition

$$LGD > C \cdot \max_{j(i) \leqslant j \leqslant N(i)} \Delta(s_j, e_j) \tag{28}$$

holds, then (27) strictly increases if λ_i increases.

In practice, condition (28) is verified for usual values of LGD and C: for instance, if the often-standard value for LGD of 60% is chosen and $C = 5\%$, then the right-hand side of (28) would be equal, up to day-count rounding, to $5\% \cdot 0.25 = 1.25\%$, due to the quarterly payments of each CDS contract. A graphical illustration is provided in Figure 3.

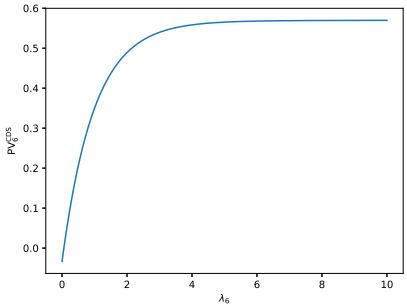


Figure 3: Present value of the 6^{th} CDS used for the calibration example provided in Section 2.5.1 (5Y maturity) with risk-neutral default probabilities implied via the minmaxvar distortion, as a function of λ_6. Notional assumed unitary.

Note that, when $i = 1$, the second summand in (27) is, in general, negligible. This is because the event $\{0 \leqslant \tau \leqslant t_p\}$ means observing a default between the valuation date and the protection start date of the CDS, where the latter is usually one day after the former. With a good approximation $DF(\tau)$ equals 1, as when $0 \leqslant \tau \leqslant t_p$ the year fraction between the valuation date and the default time is almost zero. Further, we can bound $\Delta(s_1, \tau)$ from above, up to day-count rounding, with 0.25. Thus, an approximate upper bound for $\mathbb{E}^{\mathbb{Q}}\left(DF(\tau) \cdot C \cdot \Delta(s_1, \tau) \cdot \mathbb{1}_{\{0 \leqslant \tau \leqslant t_p\}}\right)$ is given by $C \cdot 0.25 \cdot \mathbb{Q}(0 \leqslant \tau \leqslant t_p)$.

To give an idea about the magnitude of this term, if we consider as a simple case a piecewise-constant hazard rate functional form for (8), we than have that $\mathbb{Q}(0 \leqslant \tau \leqslant t_p) = 1 - e^{-\lambda_1 \cdot \Delta(0, t_p)}$. If λ_1 increases by an amount δ, by using a first order Taylor expansion we obtain that $\mathbb{Q}(0 \leqslant \tau \leqslant t_p)$ increases by approximately $\delta \cdot \Delta(0, t_p)$. Thus, if λ_1 increases by δ then the change in $\mathbb{E}^{\mathbb{Q}}\left(DF(\tau) \cdot C \cdot \Delta(s_1, \tau) \cdot \mathbb{1}_{\{0 \leqslant \tau \leqslant t_p\}}\right)$ is approximately bounded from above by $C \cdot 0.25 \cdot \delta \cdot \Delta(0, t_p)$. Again, assume C equals 5% and that t_p occurs one day after the valuation date. Using the Act/360 day-count convention we obtain that this amount equals $1.25\% \cdot \delta \cdot \frac{1}{360}$, which is negligible when δ not too large; see Figure 4a for an example.

When $i = 1$, if $\lambda_1 = 0$ then the present value of the protection leg would be zero, making the value of the contract negative. When λ_1 increases, the present value of the contract increases as well, and for λ_1 large enough it would reach a positive value. However, when λ_1 diverges to $+\infty$, then a default would occur while the contract is being signed, which would make the value of the contract drop. Therefore, the monotonicity would be guaranteed, when $i = 1$ and when usual coupon and LGD amounts are considered, on an interval $[0, \tilde{\lambda}_1]$, which is usually wide enough for practical applications. This is illustrated in Figure 4b.

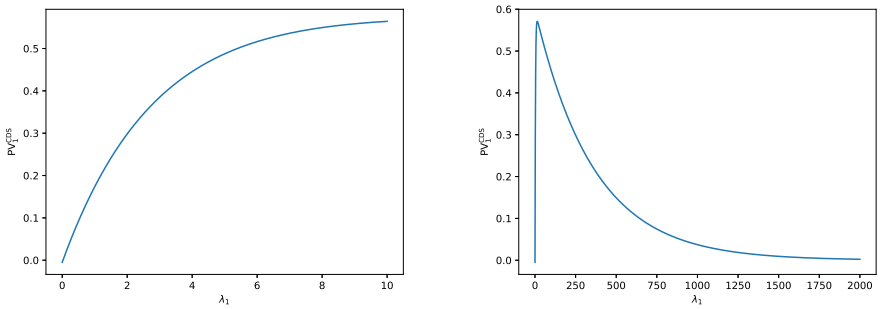


Figure 4: Present value of the first CDS used for the calibration example provided in Section 2.5.1 (6M maturity) with risk-neutral default probabilities implied via the min-maxvar distortion, as a function of λ_1 for the same range for λ_1 as in Figure 3, panel (a), and when λ_1 diverges, panel (b). Notional assumed unitary.

Therefore, for parameters usually considered in practice, assuming that the price of the i^{th} CDS strictly increases when λ_i increases (eventually within an interval that is large enough for practical applications in the case $i = 1$) is a reasonable assumption, that we have used throughout the paper.

2.B TABLES

Tenor	UF_i^{bid}	UF_i^{ask}	UF_i^{mid}	$\lvert(UF_i^{ask} - UF_i^{bid})/UF_i^{mid}\rvert$
6M	-0.0033	-0.0026	-0.0030	23.73%
1Y	-0.0074	-0.0068	-0.0071	8.45%
2Y	-0.0149	-0.0126	-0.0138	16.73%
3Y	-0.0192	-0.0169	-0.0181	12.74%
4Y	-0.0221	-0.0198	-0.0210	10.98%
5Y	-0.0219	-0.0198	-0.0209	10.07%
7Y	-0.0162	-0.0095	-0.0129	52.14%
10Y	-0.0073	0.0047	-0.0013	932.08%

Table 1: Bid and ask CDS premia depicted in Figure 1a, rounded to the basis-point digit. For comparison purposes, the mid CDS premia (UF^{mid}), as well as the absolute value of the bid-ask CDS premium spreads with respect to mid CDS premia have been reported.

Tenor	$\lambda_i^{\text{minmaxvar}}$	λ_i^{Wang}
6M	0.001302	0.002235
1Y	0.003348	0.003353
2Y	0.004790	0.005615
3Y	0.009759	0.009705
4Y	0.011977	0.011906
5Y	0.017028	0.016865
7Y	0.022719	0.023514
10Y	0.023259	0.023671

(a)

Tenor	$\gamma_i^{\text{minmaxvar}}$	γ_i^{Wang}
6M	0.159982	0.202429
1Y	0.055352	0.063821
2Y	0.076611	0.083899
3Y	0.040828	0.042300
4Y	0.027284	0.027351
5Y	0.017690	0.017241
7Y	0.033905	0.032033
10Y	0.041636	0.038361

(b)

Table 2: Calibrated parameters: hazard rates λ, panel (a), and distortion parameters γ, panel (b). Implied hazard rates and distortion parameters are depicted in Figures 2a and 2b, respectively.

3 LIQUIDITY-FREE IMPLIED VOLATILITIES: AN APPROACH USING CONIC FINANCE

3.1 ABSTRACT

We consider the problem of calculating risk-neutral implied volatilities of European options without relying on option mid prices but solely on bid and ask prices. We provide an approach, based on the conic finance paradigm, that allows to uniquely strip risk-neutral implied volatilities from bid and ask quotes, and that does not require restrictive assumptions. Our methodology also allows to jointly calculate the implied liquidity of the market. The idea outlined in this paper can be applied to calculate other implied parameters from bid and ask security prices as soon as their theoretical risk-neutral counterparts are strictly increasing with respect to the former.

3.2 INTRODUCTION

The implied volatility of an option is defined as the value of the volatility of the underlying asset which, when used as input in a given pricing model, returns a theoretical value matching the current market price of the option considered. In this article we propose a methodology which allows to compute risk-neutral implied volatilities of European options.[1] This is accomplished without relying on any mid quote approximations. Instead, our approach can be applied starting from bid and ask quotes directly, and we outline how to use our technique under both Black-Scholes and Bachelier modeling settings.

The concept of implied volatility is relevant for different reasons. First of all, Black-Scholes (but also Bachelier) implied volatilities are important *quoting conventions* in financial markets. They are therefore useful as benchmarks for the calibration of option pricing models. Nonetheless, several applications of the notion of implied volatility have been investigated outside the valuation framework, and, more precisely, in forecasting analysis: there are studies investigating the use of implied volatilities to predict, amongst other things, realized

1 Here options are always assumed to have European-style exercise. Thus, the exercise type will be often omitted, for brevity.

volatilities Szakmary et al. (2003), asset returns An et al. (2014); Fu et al. (2016) and financial market bubbles Sornette et al. (2018). Moreover, implied volatility spreads, i.e., the differences in call and put implied volatilities, have been used to forecast option returns Doran et al. (2013) and equity premia Cao et al. (2020). Thus, the more accurately one can calculate implied volatilities from option prices, especially far from the at-the-money point where liquidity is lower, the better.

In practical applications and analyses, implied volatilities are often calculated starting from mid option prices, as for instance in Ulrich and Walther (2020). It is a known fact that the risk-neutral price of a contract lies within an interval with lower and upper bounds given by the bid and the ask prices, respectively. However, the risk-neutral price, in general, does not coincide with the mid, despite the latter is, usually, employed as a proxy for the former. To be able to calculate the correct risk-neutral implied volatility of an option without relying on mid market approximations, one would need to model option prices in a two-price economy. One possible approach to do so is that of conic finance, introduced in Cherny and Madan (2010). This allows to evaluate bid and ask prices of contingent claims by recognizing that, in an economy, risk cannot be fully eliminated. Therefore, markets should quote based on the notions of (static) *index of acceptability* and *coherent risk measure*, consistently with the risk-neutral paradigm. By characterizing the structure of the contingent claims that are considered acceptable by the market, computing bid and ask prices can be performed by means of Choquet expectations Choquet (1954) of the relevant terminal payoffs with respect to distorted versions of the risk-neutral distribution of the underlying asset. This static approach to conic finance has found disparate practical applications. These applications range from exotic and structured products Guillaume and Schoutens (2013); Madan and Schoutens (2012) to contingent convertibles Madan and Schoutens (2011a), from capital calculations Madan (2009); Madan (2012) to credit valuation adjustments van Bakel et al. (2020); Madan and Schoutens (2016b), and again from hedging insurance risk Carr et al. (2016) to implied liquidity Corcuera et al. (2012). For a better overview of the applications of conic finance, see Madan and Schoutens (2016a). We observe that the approach to conic finance based on static indices of acceptability can be extended to a time-dependent framework, as in Bielecki et al. (2013), via the notion of *dynamic index of acceptability* Bielecki et al. (2014). However, as our aim is that of extracting market information from the currently-available market data (i.e., European option prices), a static approach to conic finance already suits our needs.

Our approach to imply risk-neutral volatilities without relying on any mid quote approximation is based on the conic finance theory of Cherny and Madan (2010) and, in particular, on Michielon et al. (2021a), where a methodology to imply risk-neutral default probability distributions from bid and ask credit default swaps (CDSs) is outlined. Note that our methodology can be used to compute

risk-neutral market-implied quantities from quoted bid and ask prices of any type of contingent claim, provided that some basic assumptions are satisfied. In particular, in the specific case of European options, our methodology requires some technical conditions to be fulfilled concerning the liquidity level of the market and the infima and suprema of the option prices with respect to changes in the volatility parameter. We observe that the fact that European option prices are strictly increasing with respect to the volatility parameter is essential for our technique to be applied (therefore, for other products, to imply a given model parameter a monotonicity condition needs to be satisfied; see Theorem 2 for the technical details).

Within the conic finance framework the concept of conic implied volatility has been introduced in (Madan and Schoutens, 2016a, Ch. 5.4.3). Therein it is illustrated how, given market bid and ask quotes, a distortion function, and a liquidity level, one can compute the implied volatilities that allow to price back the observed bid and ask prices, named *conic implied volatilities*. This approach is different from that of calculating *bid* and *ask implied volatilities*, that is, the implied volatilities that allow to match quoted bid and ask option prices under risk-neutral settings. Bid implied volatilities are lower than ask implied volatilities given that bid prices are below their ask counterparts. On the contrary, (Madan and Schoutens, 2016a, Ch. 5.4.3) show that this condition does not need to hold when conic implied volatilities are calculated. Note that the technique proposed in (Madan and Schoutens, 2016a, Ch. 5.4.3) is outlined in the specific case of options with European exercise features under Black-Scholes specifications for underlyings paying continuous dividends. However, it can be also applied, for instance, in the case of Bachelier specifications, and it is not restricted to the equity asset class only. Our idea is fundamentally different from both the above, as in our approach we imply a single volatility given a bid and an ask, and not an implied volatility per quote (i.e., one for the bid and one for the ask) as done in the standard case and in (Madan and Schoutens, 2016a, Ch. 5.4.3). This because we are interested in computing implied volatilities which can be interpreted as risk-neutral ones. Further, our approach still allows to compute implied volatility spreads, as the methodology can be followed for calls and puts separately. Our method guarantees that implied risk-neutral volatilities and liquidity levels, the latter in the spirit of Corcuera et al. (2012), can be uniquely determined. Further, the methodology outlined here is also simple from a computational perspective, as it only requires to solve a (constrained) non-linear system with two equations and two unknowns.

We highlight here that it is not our intention to advocate the usage of Black-Sholes (or Bachelier) settings in financial modeling. The reason why we provide a method to strip *liquidity-free* implied risk-neutral volatilities is that option prices are often quoted in terms of Black-Scholes (or Bachelier) volatilities. In addition, both Black-Scholes and Bachelier settings can be seen as option price interpolators, and the implied volatilities they generate are often benchmark in-

puts in several pricing models. Therefore, their accurate calculation, which we show can be performed without relying on mid quote approximations, is of key importance in financial modeling.

This paper is organized as follows. Section 3.3 provides a brief introduction to the theory of conic finance. Section 3.4 outlines how to compute risk-neutral implied volatilities starting from bid and ask option prices using the conic finance theory. In particular, it highlights how to do so in the case distortions are modeled as Wang transforms Wang (2000) by recalling the conic Black Scholes formulae of (Madan and Schoutens, 2016a, Ch. 5.4) and by providing conic Bachelier option pricing formulae. Section 3.5 provides an illustration of the methodology outlined in this article, while Section 3.6 concludes. The proof of Theorem 2 can be found in Appendix 3.A. For completeness, in Appendix 3.B the derivation of risk-neutral Bachelier option pricing formulae is available, while in Appendix 3.C a remark on a property of the Wang transform is provided.

3.3 PRICING IN A TWO-PRICE ECONOMY

The theory of conic finance introduced in Cherny and Madan (2009) is based on the idea that, in financial markets, risks cannot be fully hedged. Therefore, positions are taken after having weighted the possible risks and rewards connected to the instruments traded in the market. Hence, financial markets are modeled as abstract counterparties that allow trades to take place after they have passed some sort of "quality assessment". To do so, financial markets would need a machinery to perform this appraisal. In Cherny and Madan (2009) this is based on the concept of index of acceptability. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a functional $\alpha : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \to [0, +\infty]$ which assigns higher (lower) values to random variables that are expected to perform better (worse) is what an index of acceptability is. Cherny and Madan (2009) impose some technical conditions on the notion of index of acceptability, otherwise their class would be too wide for being of practical use. In particular, if two random cashflows are acceptable at a given level γ (i.e., if their level of acceptability is at least γ), then the same applies to any convex combination of them. Moreover, indices of acceptability are assumed to be monotonic: if a random cashflow always outperforms a second one, then the former would be better ranked than the latter. Indices of acceptability are also supposed to be scale-invariant, i.e., the expected performance of a cashflow X is the same of that of λX, for every $\lambda > 0$. Finally, the technical Fatou property needs to be satisfied: given $(X_n)_n$ a sequence of random cashflows such that, for every n, $|X_n| \leqslant 1$ and $\alpha(X_n) \geqslant \gamma$, then if $(X_n)_n$ converges in probability to a random cashflow X, then also $\alpha(X) \geqslant \gamma$. Cherny and Madan (2009) prove that, provided an index of acceptability α, for every $x \geqslant 0$ there

exists a set $\mathfrak{Q}_x$ of probability measures absolutely continuous with respect to $\mathbb{P}$ such that $x \leqslant x'$ implies $\mathfrak{Q}_x \subseteq \mathfrak{Q}_{x'}$, and that

$$\alpha(X) = \sup\left\{x \geqslant 0 : \inf_{Q \in \mathfrak{Q}_x} \mathbb{E}^Q(X) \geqslant 0\right\}.$$

The concept of index of acceptability can be then linked to that of coherent risk measure, i.e., a map $\rho : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \to [0, +\infty]$ that is transitive, sub-additive, positively homogeneous and monotonic, see (Madan and Schoutens, 2016a, Ch. 4.1). Delbaen (2002) show that a coherent risk measure can be identified with a functional such that $X \mapsto \sup_{Q \in \mathfrak{Q}} \mathbb{E}^Q(X)$, where the set $\mathfrak{Q}$ contains measures that are absolutely continuous with respect to $\mathbb{P}$. The concepts of index of acceptability and that of coherent risk measure can be tied together via the relationship

$$\alpha(X) = \sup\{x \geqslant 0 : \rho_x(-X) \leqslant 0\}, \tag{29}$$

where $(\rho_x)_{x \geqslant 0}$ is a family of coherent risk measures such that $\rho_x(-X) \leqslant \rho_{x'}(-X)$ whenever $x \leqslant x'$. From this it follows that $\alpha(X) \geqslant \gamma$ is equivalent to $\rho_\gamma(-X) \leqslant 0$. For this reason, indices of acceptability with the aforementioned properties are often called *coherent indices of acceptability*.

We now recall the definition of (asymmetric) Choquet integral Choquet (1954) (we will, from here onwards, always omit the word "asymmetric", as symmetric Choquet integrals are not relevant in this framework, and we refer the interested reader to (Denneberg, 1994, Ch. 7)). For a *non-additive probability* μ and a random variable X, the Choquet integral is defined as

$$(C)\int_\Omega X \, d\mu := \int_{-\infty}^0 \mu(X \geqslant t) - 1 \, dt + \int_0^{+\infty} \mu(X \geqslant t) \, dt. \tag{30}$$

In (30), the integrals on the right-hand side should be interpreted as improper Riemann integrals. Therefore, they both exist given that their arguments are monotonic functions, which guarantees that the sets of their discontinuities have a Lebesgue measure of zero. Note, however, that their sum does not necessarily exist[2]: see (Denneberg, 1994, Ch. 5) for a detailed treatment of Choquet integrals.

Let a risk-neutral measure $\mathbb{Q} \in \bigcap_{x \geqslant 0} \mathfrak{Q}_x$. A *distortion function* is a function from $[0, 1]$ to $[0, 1]$ that maps 0 to 0 and 1 to 1. For a concave distortion function $\psi(\cdot)$ we denote with $\psi(\mathbb{Q})(A)$ the (potentially non-additive) probability measure that assigns to each measurable set A the probability mass $\psi(\mathbb{Q}(A))$. Given $(\psi_x)_{x \geqslant 0}$ an increasing family of concave distortion functions, the map ρ_x such that $X \mapsto (C)\int_\Omega X \, d\psi_x(\mathbb{Q})$ defines a coherent risk measure. Hence, as per Cherny and Madan (2009), functionals of this form can be employed to describe indices of acceptability by setting

$$\alpha(X) := \sup\left\{x \geqslant 0 : (C)\int_\Omega -X \, d\psi_x(\mathbb{Q}) \leqslant 0\right\}. \tag{31}$$

2 If X is non-negative(positive), then (30) is guaranteed to be well-defined.

The tools just introduced can be now used to characterize direction-dependent pricing in financial markets.

Assume that a threshold of at least γ has been set by the market for a given contingent claim to be considered acceptable and, thus, tradable. We assume a constant risk-free rate r[3] and consider a contingent claim X with a terminal payoff at time T. The market is then willing to buy X at a price b if and only if $\alpha(X - e^{-rT}b) \geqslant \gamma$. This is equivalent, given the assumption that the market evaluates the performance of contingent claims by means of Choquet integrals, to the condition $b \leqslant -e^{-rT}(C)\int_\Omega -X\, d\psi_\gamma(Q)$. Thus, the bid price of X, $\text{bid}_\gamma(X)$, equals

$$\text{bid}_\gamma(X) = -e^{-rT}(C)\int_\Omega -X\, d\psi_\gamma(Q). \tag{32}$$

As the ask price of X, $\text{ask}_\gamma(X)$, equals $-\text{bid}_\gamma(-X)$, from (32) it then immediately follows that

$$\text{ask}_\gamma(X) = e^{-rT}(C)\int_\Omega X\, d\psi_\gamma(Q). \tag{33}$$

Note that should more than one risk-neutral measure exist, then one would need to choose which risk-neutral measure to use within formulae (32) and (33). Further, we observe that, in the formulae provided above, the choice of the distortion function provides the modeler with a degree of freedom to describe the liquidity dynamics of the market. In addition, to different values of the distortion parameter γ there correspond different market liquidity specifications. This framework reminds that of modeling preferences towards risk by means of utility functions. Despite utility theory characterizes agents' behavior from a micro-economic perspective (i.e., the individual preferences of each agent) while conic finance describes risk attitudes of financial markets, there are some similarities between the two approaches worth of attention. In particular, in utility theory the modeler has to choose the functional form of the utility function to be used. This is similar to the conic finance case, where a choice related to the distortion function also has to be made. Further, in utility theory one has to choose the parameter(s) of the utility function in order to describe the level of risk aversion (or risk tolerance) of an agent. Similarly, on the conic finance side, the behavior of the market is further described by the distortion parameter γ. In addition, we also point out that Choquet integrals are common tools in decision theory. In particular, we recall the results of Schmeidler (1989) which characterize choices under uncertainty, the latter in the sense of Knight (1921), in terms of Choquet integrals (for a representation result concerning Choquet integrals, see Schmeidler (1986), on which Schmeidler (1989) is based on). We further highlight that Choquet integrals can be also applied to option pricing problems

3 Note that the constant risk-free rate assumption has been made only for consistency with the fact that, in this article, we consider Black-Scholes and Bachelier models. However, in the case of a time-dependent risk-free rate, all the steps outlined from here onwards would still hold.

under uncertainty by means of Choquet Brownian motions, introduced in Kast et al. (2014), as done in Driouchi et al. (2015).

3.4 LIQUIDITY-FREE OPTION IMPLIED VOLATILITIES

From here onwards we consider European options only, and we assume to be either within the Black-Scholes or the Bachelier framework. This because we are interested in backing out either log-normal or normal implied volatilities.

We denote with $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ a filtered probability space, and with with $(X_t)_{t\in[0,T]}$ the process representing a "generic" underlying. In particular, by introducing an adjusted risk-neutral drift $r - \alpha$, one can define Black-Scholes dynamics via the stochastic differential equation (SDE) given, under the risk-neutral measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ on $\mathcal{F}$, by

$$dX = (r - \alpha)Xdt + \sigma X dW. \tag{34}$$

In (34), σ denotes the volatility term, and $(W_t)_{t\in[0,T]}$ a Brownian motion adapted to the filtration $(\mathcal{F}_t)_{t\in[0,T]}$. The parameter α can be defined according to the asset class considered. For instance, setting $\alpha = 0$ outlines the standard Black-Scholes framework on a non-dividend paying underlying, setting $\alpha = q$ with q denoting the continuous dividend yield corresponds to the Black-Scholes framework for an underlying paying continuous dividends, while setting $\alpha = r$ corresponds to the Black model for futures options; see (Haug, 2007, Ch. 1.1.6) for further possible specifications.

To take into account the possibility that the underlying asset can reach negative values, e.g., in the case of rates and oil prices[4], option prices can also be quoted in terms of Bachelier (i.e., normal) implied volatilities. Therefore, in a similar manner as per (34) and using the same notation conventions, one can define the Bachelier SDE as

$$dX = (r - \alpha)Xdt + \sigma dW. \tag{35}$$

Via (35) we have chosen to describe a generalized variant, in the sense of (Haug, 2007, Ch. 1.1.6), of the "contemporary" version of the Bachelier model as in (Musiela and Rutkowski, 2005, Ch. 3.3). Note, however, that in the literature sometimes the SDE corresponding to the Bachelier model slightly differs from that outlined in (35), for instance by not considering the drift term (see (Haug, 2007, Ch. 1.3.1)). In any case, independently on the exact specifications of the Bachelier SDE considered, all the Bachelier-related calculations available in this article can be performed in the same manner, up to minor rearrangements.

4 The reader can refer to `https://www.cmegroup.com/content/dam/cmegroup/notices/clearing/2020/04/Chadv20-152.pdf` for a note of the Chicago Mercantile Exchange concerning the possible use of the Bachelier formula due to the negative oil prices observed in 2020.

For a given strike K and maturity T, we denote with $\mathcal{C}$ and with $\mathcal{P}$ the prices of a call and a put option written on X with such strike and maturity. However, when considering the Black-Scholes model (34) (Bachelier model (35)), we will use $\mathcal{C}^{BS}$ ($\mathcal{C}^{B}$) and $\mathcal{P}^{BS}$ ($\mathcal{P}^{B}$), instead. Note that, depending, on the context, we will make option prices explicitly depend on specific parameters only, as it will be clear in the next sections. This to keep the notation as light as possible. In any case, the dependency on both strike and maturity will be always omitted, as redundant in our context.

In practice, implied volatilities, either normal or log-normal, are backed out from call and put options separately. More precisely, implied volatilities for call options are computed starting from the mid prices of these options and, similarly, the same applies to put implied volatilites. However, in principle, one would like to compute the real risk-neutral implied volatilities, without relying on approximating risk-neutral prices by their mid counterparts. A similar problem to this has been analyzed in Michielon et al. (2021a). Therein it is shown that, in the case of CDSs, under mild assumption concerning the liquidity level of the market and the characteristics of the default time process, it is possible to strip risk-neutral default probabilities from bid and ask CDS quotes directly in a unique manner. The considerations available in Michielon et al. (2021a) are now extended to a more general setup. In particular, the methodology we highlight here is quite general. That is, it can be applied to any contingent claim whose risk-neutral price depends on a single unknown parameter provided that the former is strictly increasing with respect to the latter, and as soon as two basic additional conditions are satisfied. That is, the range of theoretical prices obtainable by changing the free parameter, as well as the liquidity level of the market, should be "wide enough", as we will explain more technically in Theorem 2.

Let Y be a contingent claim, and denote with $\mathrm{PV}(Y(\lambda))$ its risk-neutral price, assumed dependent on an unknown parameter λ. Further, let b and a denote its quoted bid and ask prices, respectively. The main result available in Michielon et al. (2021a) is recalled in Theorem 2 in a more general fashion. The proof of Theorem 2 follows from Lemmas 1, 2, 3 and Theorem 1 in Michielon et al. (2021a), as the steps outlined therein can be followed in the same manner. For completeness, we have provided the aforementioned results, adapted to the more general context considered in the present article, in Appendix 3.A.

Theorem 2. *Let Y be a contingent claim whose price depends on a parameter $\lambda > 0$ such that the risk-neutral price of Y is strictly increasing with respect to λ. Assume that $\inf_{\lambda>0} \mathrm{PV}(Y(\lambda)) < b$ and that $\sup_{\lambda>0} \mathrm{PV}(Y(\lambda)) > a$. This uniquely identifies an interval $[\lambda_a, \lambda_b]$ such that $\lambda \in [\lambda_b, \lambda_a]$ if and only if $\inf_{\lambda>0} \mathrm{PV}(Y(\lambda)) < b$ and $\sup_{\lambda>0} \mathrm{PV}(Y(\lambda)) > a$. Moreover, assume that for every $\lambda \in [\lambda_b, \lambda_a]$ there exists*

$\gamma > 0$ *such that* $\text{ask}(Y(\lambda),\gamma) - \text{bid}(Y(\lambda),\gamma) = a - b$. *Then, the constrained non-linear system*

$$\begin{cases} \text{bid}(Y(\lambda),\gamma) = b \\ \text{ask}(Y(\lambda),\gamma) = a \end{cases} \tag{36}$$

with

$$b < \text{PV}(Y(\lambda)) < a$$

admits a solution, which is also unique.

We observe that both call and put option prices are strictly increasing with respect to the volatility of the underlying.[5] Denote with $b_{\mathcal{C}}$ ($b_{\mathcal{P}}$) and with $a_{\mathcal{C}}$ ($a_{\mathcal{P}}$) the quoted bid and ask price of the call (put), respectively. Ideally, one would aim to find an implied risk-neutral volatility σ and two distortion parameters $\gamma_{\mathcal{C}}$ and $\gamma_{\mathcal{P}}$ such that the equalities

$$\begin{cases} \text{bid}(\mathcal{C}(\sigma),\gamma_{\mathcal{C}}) = b_{\mathcal{C}} \\ \text{ask}(\mathcal{C}(\sigma),\gamma_{\mathcal{C}}) = a_{\mathcal{C}} \\ \text{bid}(\mathcal{P}(\sigma),\gamma_{\mathcal{P}}) = b_{\mathcal{P}} \\ \text{ask}(\mathcal{P}(\sigma),\gamma_{\mathcal{P}}) = a_{\mathcal{P}} \end{cases} \tag{37}$$

with the constraints

$$\begin{cases} b_{\mathcal{C}} < \text{PV}(\mathcal{C}(\sigma)) < a_{\mathcal{C}} \\ b_{\mathcal{P}} < \text{PV}(\mathcal{P}(\sigma)) < a_{\mathcal{P}} \end{cases}$$

are satisfied. Notwithstanding, it is in general not possible to solve (37) due to the obvious lack of degrees of freedom. Therefore, separate volatility and liquidity parameters should be used for calls and puts, as done in practice to compute call-put volatility spreads. In particular, one can solve

$$\begin{cases} \text{bid}(\mathcal{C}(\sigma),\gamma) = b_{\mathcal{C}} \\ \text{ask}(\mathcal{C}(\sigma),\gamma) = a_{\mathcal{C}} \end{cases}$$

with

$$b_{\mathcal{C}} < \text{PV}(\mathcal{C}(\sigma)) < a_{\mathcal{C}},$$

and obtain that a unique solution, in virtue of Theorem 2, exists, denoted with $(\sigma_{\mathcal{C}},\gamma_{\mathcal{C}})$. Similarly, one can solve

$$\begin{cases} \text{bid}(\mathcal{P}(\sigma),\gamma) = b_{\mathcal{P}} \\ \text{ask}(\mathcal{P}(\sigma),\gamma) = a_{\mathcal{P}} \end{cases}$$

5 Note that the conditions $\inf_{\lambda>0} \text{PV}(Y(\lambda)) < b$ and that $\sup_{\lambda>0} \text{PV}(Y(\lambda)) > a$ can be made more explicit in the case of European options. In particular, for a call option it results that $\inf_{\sigma>0} \mathcal{C}(\sigma) = e^{-rT}(e^{(r-\alpha)T}X_0 - K)^+$, while $\sup_{\sigma>0} \mathcal{C}(\sigma) = e^{-\alpha T}X_0$. Similarly, for a put option it results that $\inf_{\sigma>0} \mathcal{P}(\sigma) = e^{-rT}(K - e^{(r-\alpha)T}X_0)^+$, and that $\sup_{\sigma>0} \mathcal{P}(\sigma) = e^{-rT}K$.

with

$$b_{\mathcal{P}} < \mathrm{PV}\left(\mathcal{P}(\sigma)\right) < a_{\mathcal{P}}$$

separately, and again obtain a unique solution, due to Theorem 2, denoted as $(\sigma_{\mathcal{P}}, \gamma_{\mathcal{P}})$. We call $\sigma_{\mathcal{C}}$ and $\sigma_{\mathcal{P}}$ the liquidity-free call and put implied volatilities, respectively. The quantities $\gamma_{\mathcal{C}}$ and $\gamma_{\mathcal{P}}$ denote the implied liquidity levels of the market for calls and puts, respectively.

The approach proposed here is different from that proposed in (Madan and Schoutens, 2016a, Ch. 5.4.3). Therein, the notion of conic (Black-Scholes) implied volatility is introduced. In particular, for a fixed (and known) distortion parameter, one can then imply a volatility for the bid and one for the ask. However, our approach allows to simultaneously imply both the distortion parameter and the implied volatility directly, without therefore relying on an initial estimation procedure for the distortion itself. This because our goal is that of computing implied volatilities that can be interpreted as risk-neutral ones.

In time series analysis call-put volatility spreads, as outlined in Section 3.2, can be used as regression variables for forecasting analysis. By taking into account the approach outlined here one would not only have the possibility to introduce liquidity-free call-put volatility spreads in the regression model considered, but also to take into account the implied distortion (i.e., liquidity) parameters as regression variables as well. Potentially, this could enhance the explanatory power of the regression models used for prediction purposes (see van Bakel et al. (2020) for an illustration of the explanatory power of the distortion parameter as far as liquidity is concerned).

3.4.1 Implied volatilities with the Wang transform

The choice of the distortion function to be used in the bid-ask calibration problem is arbitrary, provided that it is concave. Consequently, different possibilities are available, see Cherny and Madan (2009) and (Madan and Schoutens, 2016a, Ch. 4.7). However, for distributions of normal or log-normal random variables, which are often employed in financial applications, the Wang transform Wang (2000), which is defined as

$$\psi_\gamma(x) \coloneqq \Phi(\Phi^{-1}(x) + \gamma) \tag{38}$$

with $\Phi(\,\cdot\,)$ denoting the cumulative distribution function of a standard normal random variable, is a convenient choice. This because (38) still allows to obtain closed-form solutions for call and put option prices, see (Madan and Schoutens, 2016a, Ch. 5.4).[6] Therefore, under both Black-Scholes (34) and Bachelier (35) settings, exact formulae can be used to calculate bid and ask option prices via

6 See Appendix 3.C for a remark concerning how the Wang transform can be a useful tool as soon as the distribution of a normal random variable is transformed via a non-decreasing and left-continuous function.

the Wang transform.[7] Thus, our procedure to back out implied volatilities (and, consequentially, implied distortion parameters) can be easily implemented, with the advantage that it does not require to compute the integrals (32) and (33) numerically should the Wang transform be used.

In the case of the Black-Scholes framework, one obtains that the risk-neutral price of a call option is given by

$$\mathcal{C}^{BS}(\alpha) = e^{-\alpha T} X_0 \Phi(d_+) - e^{-rT} K \Phi(d_-),$$

where

$$d_+ := \frac{\ln\left(\frac{X_0}{K}\right) + (r - \alpha + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}},$$

and with

$$d_- := d_+ - \sigma\sqrt{T} = \frac{\ln\left(\frac{X_0}{K}\right) + (r - \alpha - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}.$$

Further,

$$\mathcal{P}^{BS}(\alpha) = e^{-rT} K \Phi(-d_-) - e^{-\alpha T} X_0 \Phi(-d_+).$$

(Madan and Schoutens, 2016a, Ch. 5.4.2) obtain that, by considering the Wang transform under Black-Scholes settings, bid and ask prices for European calls and puts can be computed as $\text{bid}_\gamma\left(\mathcal{C}^{BS}(\alpha)\right) = \mathcal{C}^{BS}\left(\alpha + \frac{\gamma\sigma}{\sqrt{T}}\right)$, $\text{ask}_\gamma\left(\mathcal{C}^{BS}(\alpha)\right) = \mathcal{C}^{BS}\left(\alpha - \frac{\gamma\sigma}{\sqrt{T}}\right)$, $\text{bid}_\gamma\left(\mathcal{P}^{BS}(\alpha)\right) = \mathcal{P}^{BS}\left(\alpha - \frac{\gamma\sigma}{\sqrt{T}}\right)$ and, finally, with $\text{ask}_\gamma\left(\mathcal{P}^{BS}(\alpha)\right) = \mathcal{P}^{BS}\left(\alpha + \frac{\gamma\sigma}{\sqrt{T}}\right)$.

We now provide similar relationships in the case the Bachelier model (35) is considered.[8] Let $F^{\mathbb{Q}}_{X_T}(\cdot)$ denote the time-T risk-neutral distribution of the underlying asset. First of all we recall that for European vanilla options, if the underlying can reach negative values, in line with (Madan and Schoutens, 2016a, Ch. 5.5) the following formulae can be used to calculate bid and ask European option prices:

$$\text{bid}_\gamma(\mathcal{C}) = e^{-rT} \int_K^\infty (x - K)\, d\psi_\gamma\left(F^{\mathbb{Q}}_{X_T}(x)\right), \tag{39}$$

$$\text{ask}_\gamma(\mathcal{C}) = e^{-rT} \int_K^\infty (K - x)\, d\psi_\gamma\left(1 - F^{\mathbb{Q}}_{X_T}(x)\right), \tag{40}$$

7 Some other cases where the Wang transform produces analytical option prices formulae are those of the Sprenkle, Boness and Samuelson models (see (Haug, 2007, Ch. 1.31, 1.32 and 1.33)). However, note that computing the Wang transform is computationally expensive, as this requires the evaluation of both the cumulative distribution and quantile functions of a standard normal random variable. Therefore, for large datasets and when the Wang transform does not guarantee analytical formulae to exists, then other choices for the distortion function might be more convenient (see (Madan and Schoutens, 2016a, Ch. 4.7) for an overview).

8 Risk-neutral call and put option pricing formulae for the Bachelier model are available in Appendix 3.B, for completeness.

$$\text{bid}_\gamma(\mathcal{P}) = e^{-rT}\int_{-\infty}^{K}(x-K)\,d\psi_\gamma\left(1-F^{Q}_{X_T}(x)\right), \tag{41}$$

and

$$\text{ask}_\gamma(\mathcal{P}) = e^{-rT}\int_{-\infty}^{K}(K-x)\,d\psi_\gamma\left(F^{Q}_{X_T}(x)\right). \tag{42}$$

Observe that under both the Black-Scholes and Bachelier specifications (34) and (35) continuous probability density functions for the terminal risk-neutral distribution of the underlying asset are available. Therefore, the relationships (39), (40), (41) and (42) can be interpreted as both Riemann-Stieltjes and Lebesgue-Stieltjes integrals.

Under the Bachelier dynamics (35) the risk-neutral distribution of the underlying, at time T, is normal with mean $\bar{\mu}$ and variance $\bar{\sigma}^2$ as per (45) and (46) in Appendix 3.B. If we consider a Wang transformation with distortion parameter γ we obtain that, at time T, the underlying X_T is still normally distributed with the same variance $\bar{\sigma}^2$, but this time with mean given by $\bar{\mu}_- := \bar{\mu} - \gamma\bar{\sigma}$, see Wang (2000). Therefore, we can apply relationship (39) and obtain that

$$\begin{aligned}
\text{bid}_\gamma\left(C^B\right) &= e^{-rT}\int_{K}^{\infty}(x-K)\,d\psi_\gamma\left(F^{Q}_{X}(x)\right)\\
&= e^{-rT}\int_{K}^{\infty}\frac{x-K}{\bar{\sigma}\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\bar{\mu}_-}{\bar{\sigma}}\right)^2}\,dx\\
&= e^{-rT}\int_{\frac{K-\bar{\mu}_-}{\bar{\sigma}}}^{\infty}\frac{\bar{\mu}_- + \bar{\sigma}x - K}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx\\
&= e^{-rT}\left[(\bar{\mu}_- - K)\int_{\frac{K-\bar{\mu}_-}{\bar{\sigma}}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx\right.\\
&\qquad\left.-\bar{\sigma}\int_{\frac{K-\bar{\mu}_-}{\bar{\sigma}}}^{\infty}\frac{-x}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx\right]\\
&= e^{-rT}\left[(\bar{\mu}_- - K)\Phi\left(\frac{\bar{\mu}_- - K}{\bar{\sigma}}\right) + \bar{\sigma}\phi\left(\frac{\bar{\mu}_- - K}{\bar{\sigma}}\right)\right].
\end{aligned}$$

We can now calculate the call ask price via (40). First we observe, see Wang (2000), that

$$\begin{aligned}
\psi_\gamma\left(1-F^{Q}_{X_T}(x)\right) &= \psi_\gamma\left(1-\Phi\left(\frac{x-\bar{\mu}}{\bar{\sigma}}\right)\right)\\
&= \psi_\gamma\left(\Phi\left(\frac{\bar{\mu}-x}{\bar{\sigma}}\right)\right)\\
&= \Phi\left(\frac{\bar{\mu}-x+\gamma\bar{\sigma}}{\bar{\sigma}}\right).
\end{aligned} \tag{43}$$

By setting $\bar{\mu}_+ \coloneqq \bar{\mu} + \gamma\bar{\sigma}$ we obtain that

$$\begin{aligned}
\mathrm{ask}_\gamma\left(\mathcal{C}^B\right) &= e^{-rT}\int_K^\infty (K-x)\,d\psi_\gamma\left(1-F^Q_{X_T}(x)\right)\\
&= e^{-rT}\int_K^\infty \frac{x-K}{\bar{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\bar{\mu}_+}{\bar{\sigma}}\right)^2} dx\\
&= e^{-rT}\left[(\bar{\mu}_+ - K)\Phi\left(\frac{\bar{\mu}_+ - K}{\bar{\sigma}}\right) + \bar{\sigma}\phi\left(\frac{\bar{\mu}_+ - K}{\bar{\sigma}}\right)\right].
\end{aligned}$$

We now calculate the ask price of an European put option via (42). It results that

$$\begin{aligned}
\mathrm{ask}_\gamma\left(\mathcal{P}^B\right) &= e^{-rT}\int_{-\infty}^K (K-x)\,d\psi_\gamma\left(F^Q_{X_T}(x)\right)\\
&= e^{-rT}\int_{-\infty}^K \frac{K-x}{\bar{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\bar{\mu}_-}{\bar{\sigma}}\right)^2} dx\\
&= e^{-rT}\int_{-\infty}^{\frac{K-\bar{\mu}_-}{\bar{\sigma}}} \frac{K-\bar{\mu}_- - \bar{\sigma}x}{\bar{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx\\
&= e^{-rT}\left[(K-\bar{\mu}_-)\int_{-\infty}^{\frac{K-\bar{\mu}_-}{\bar{\sigma}}} \frac{1}{\bar{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx\right.\\
&\qquad \left. +\bar{\sigma}\int_{-\infty}^{\frac{K-\bar{\mu}_-}{\bar{\sigma}}} \frac{-x}{\bar{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx\right]\\
&= e^{-rT}\left[(K-\bar{\mu}_-)\Phi\left(\frac{K-\bar{\mu}_-}{\bar{\sigma}}\right) + \bar{\sigma}\phi\left(\frac{K-\bar{\mu}_-}{\bar{\sigma}}\right)\right].
\end{aligned}$$

Recalling (43), the bid price of the put can be calculated using (41), from which it follows that

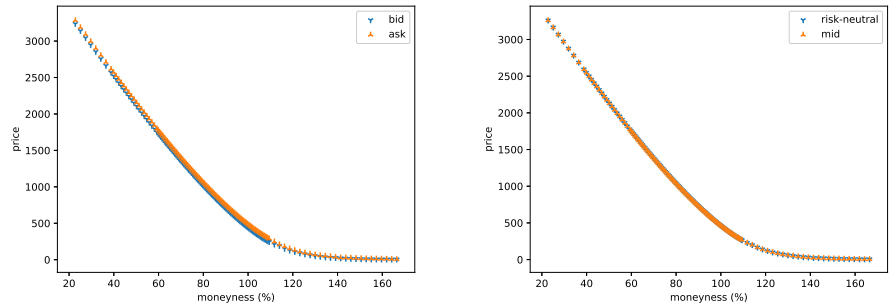
$$\begin{aligned}
\mathrm{bid}_\gamma\left(\mathcal{P}^B\right) &= e^{-rT}\int_{-\infty}^K (x-K)\,d\psi_\gamma\left(1-F^Q_{X_T}(x)\right)\\
&= e^{-rT}\int_{-\infty}^K \frac{K-x}{\bar{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\bar{\mu}_+}{\bar{\sigma}}\right)^2} dx\\
&= e^{-rT}\left[(K-\bar{\mu}_+)\Phi\left(\frac{K-\bar{\mu}_+}{\bar{\sigma}}\right) + \bar{\sigma}\phi\left(\frac{K-\bar{\mu}_+}{\bar{\sigma}}\right)\right].
\end{aligned}$$

To summarize, with reference to the notation introduced in Appendix 3.B, it results that $\mathrm{bid}_\gamma\left(\mathcal{C}^B\right) = \mathcal{C}^B(\bar{\mu}_-)$, $\mathrm{ask}_\gamma\left(\mathcal{C}^B\right) = \mathcal{C}^B(\bar{\mu}_+)$, $\mathrm{bid}_\gamma\left(\mathcal{P}^B\right) = \mathcal{P}^B(\bar{\mu}_+)$, while $\mathrm{ask}_\gamma\left(\mathcal{P}^B\right) = \mathcal{P}^B(\bar{\mu}_-)$.

3.5 EXAMPLE

Here we show how liquidity-free implied volatilities can be extracted from bid and ask prices. In particular, we consider European options on four different underlyings, i.e., European call options on the S&P 500 index, European put options on the FTSE MIB index, European call options on UBS shares, and European put options on Deutsche Telekom shares. For each of the cases considered we compute, for a given maturity (not kept unchanged for all the underlyings), bid and ask prices, risk-neutral and mid prices, absolute liquidity spreads, relative liquidity spreads, implied risk-neutral and mid volatilities, as well as implied distortion parameters. All the aforementioned calculations have been performed for all the quoted options available for which both bid and ask prices could be retrieved.[9] The Wang transform has been chosen as distortion in all the cases analyzed.

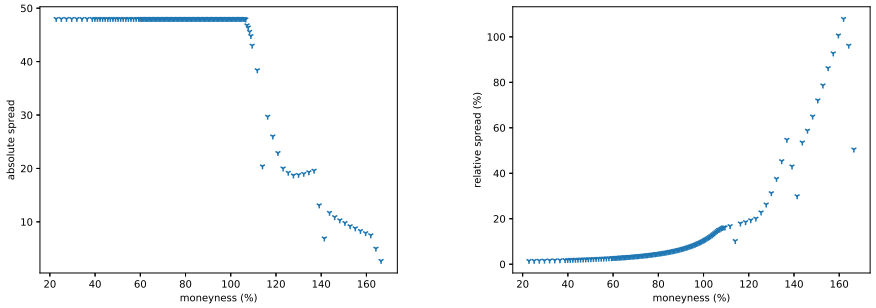
We start by considering European call options on the S&P 500 index, for which a wide range of strikes is available. These options are very liquid, as illustrated by Figures 5a and 6a (note that the relative bid-ask spreads for deep out-of-the-money options in Figure 6b are large due to those options having small market value). This results in risk-neutral and mid prices that are very close to each other, as shown in Figure 5b.



(a) (b)

Figure 5: Bid and ask prices for European call options on the S&P 500 index expiring in 886 days (Options Price Reporting Authority), panel (a), and their corresponding risk-neutral and mid counterparts, panel (b).

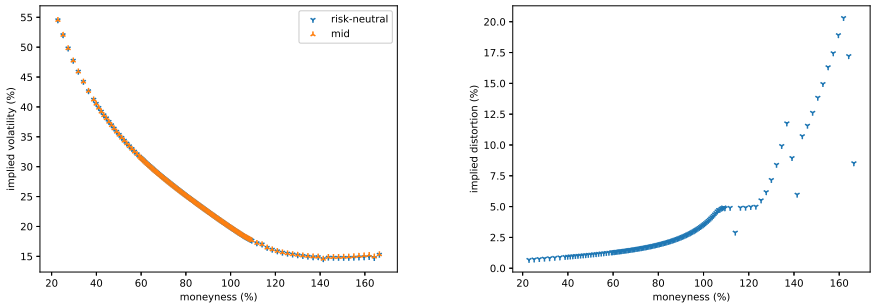
9 In this section plots have been constructed with respect to moneyness, defined here as the ratio between a given strike price and the value of the underlying.



(a) (b)

Figure 6: Absolute bid-ask spreads for the options considered in Figure 5, panel (a), and their relative counterparts (calculated with respect to mid prices), panel (b).

Also the risk-neutral and mid implied volatility smiles, see Figure 7a, are basically overlapping, as expected. The implied distortion parameters, illustrated in Figure 7b, closely follow the trend of the relative bid-ask spreads of Figure 6b.



(a) (b)

Figure 7: Risk-neutral and mid implied volatilities, panel (a), and implied liquidity levels, panel (b), for the options considered in Figure 5.

We now consider European put options on the FTSE MIB index.

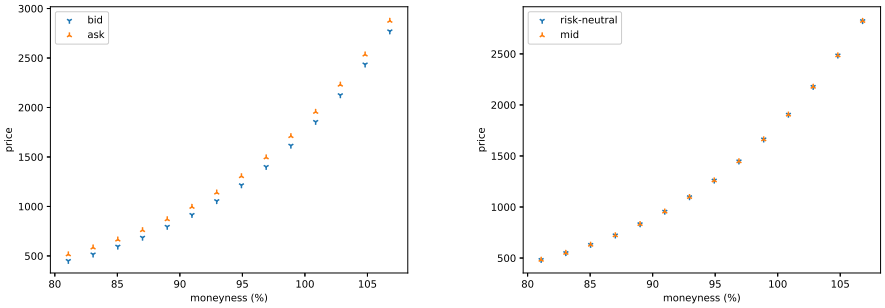


Figure 8: Bid and ask prices for European put options on the FTSE MIB (Milan Stock Exchange) expiring in 249 days, panel (a), and their corresponding risk-neutral and mid counterparts, panel (b).

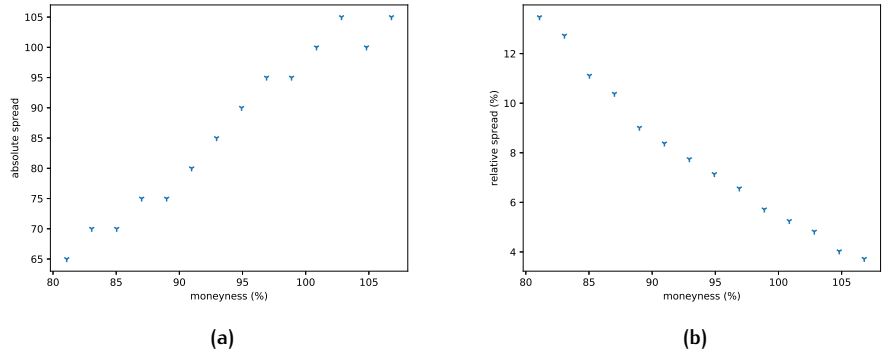


Figure 9: Absolute bid-ask spreads for the options considered in Figure 8, panel (a), and their relative counterparts (calculated with respect to mid prices), panel (b).

In this case fewer strikes are traded compared to the S&P 500 case. However, as Figure 8a illustrates, these options are still very liquid; see also Figures 9a and 9b. This is further confirmed by the low levels of the implied liquidity parameter of Figure 10b.

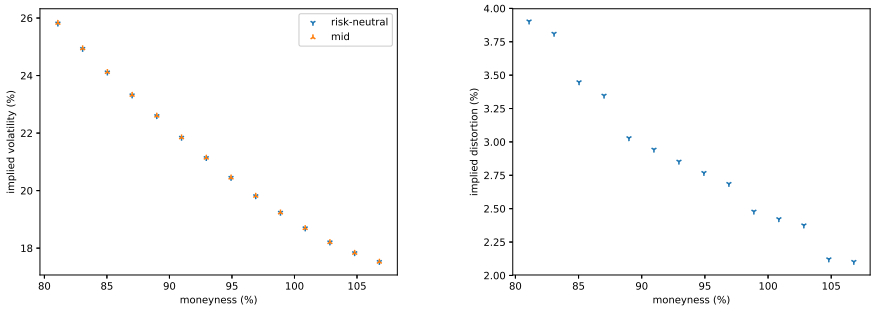


Figure 10: Risk-neutral and mid implied volatilities, panel (a), and implied liquidity levels, panel (b), for the options considered in Figure 8.

We therefore still obtain risk-neutral implied volatilities and prices that are closely approximated by their mid counterparts; see Figures 10a and 8b, respectively. We now consider European call options on UBS; see Figure 11. As it is

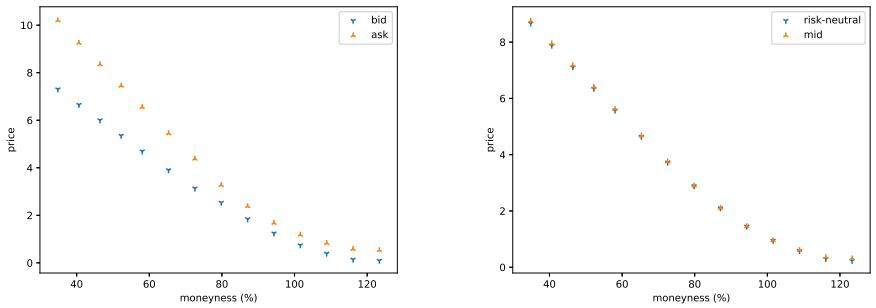


Figure 11: Bid and ask prices for European call options on UBS (Eurex) expiring in 345 days, panel (a), and their corresponding risk-neutral and mid counterparts, panel (b).

clear from Figures 11a, 12a, 12b and 13b, these options are less liquid than those considered in the two cases above (i.e., those on the S&P 500 and the FTSE MIB indices, respectively).

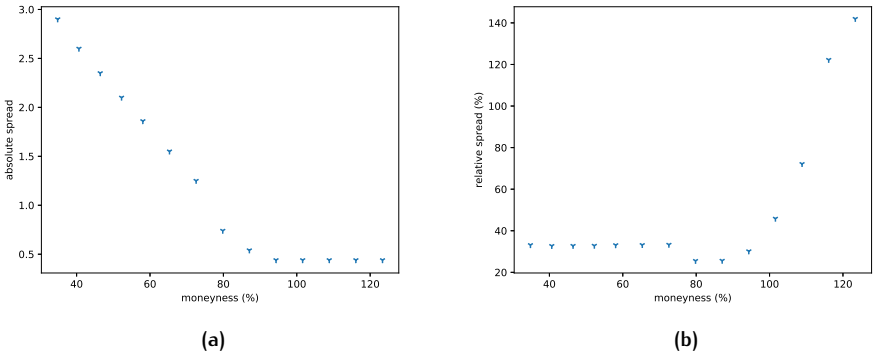


Figure 12: Absolute bid-ask spreads for the options considered in Figure 11, panel (a), and their relative counterparts (calculated with respect to mid prices), panel (b).

Therefore, this results in risk-neutral and implied volatility smiles that, for deep out-of-the-money, but especially for deep in-the-money options, exhibit non-negligible differences, with mid implied volatilities overestimating their risk-neutral counterparts up to 2-3% in the former case, and up to 9-10% in the latter case; see Figure 13a. Note that deep in-the-money and out-of-the-money options have small vegas, which leads to risk-neutral and mid option prices being close to each other; see Figure 11b.

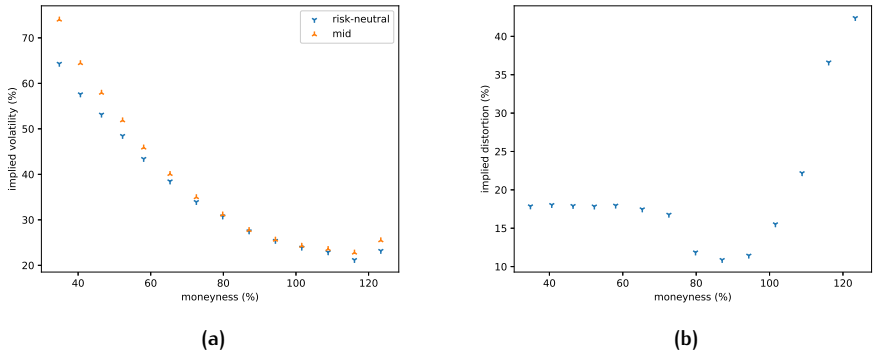


Figure 13: Risk-neutral and mid implied volatilities, panel (a), and implied liquidity levels, panel (b), for the options considered in Figure 11.

As the last case we consider that of European put options on Deutsche Telekom; see Figure 14a. Also in these circumstances liquidity is not as high as in the cases of options on the S&P 500 and FTSE MIB indices. This is illustrated by the high

levels of bid-ask spreads displayed in Figures 15a and 15b, and reiterated by the high implied liquidity levels of Figure 16b.

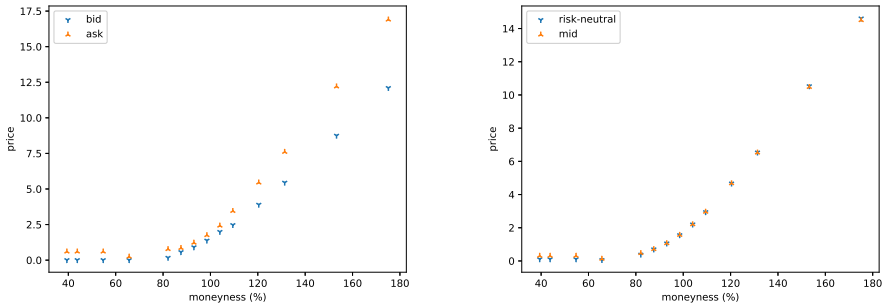


Figure 14: Bid and ask prices for European put options on Deutsche Telekom (Eurex) expiring in 345 days, panel (a), and their corresponding risk-neutral and mid counterparts, panel (b).

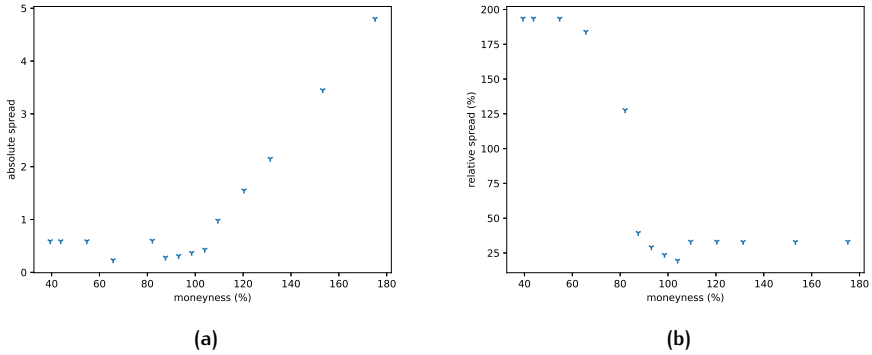


Figure 15: Absolute bid-ask spreads for the options considered in Figure 14, panel (a), and their relative counterparts (calculated with respect to mid prices), panel (b).

Due to the low liquidity for both in-the-money and out-of-the-money options for this particular underlying, differences between risk-neutral and mid implied volatilities are considerable; see Figures 16a. In the former case we observe mid implied volatilities underestimating their risk-neutral counterparts up to 9-10%, while in the latter case mid implied volatilities overestimate risk-neutral ones, with differences up to 14-15%.

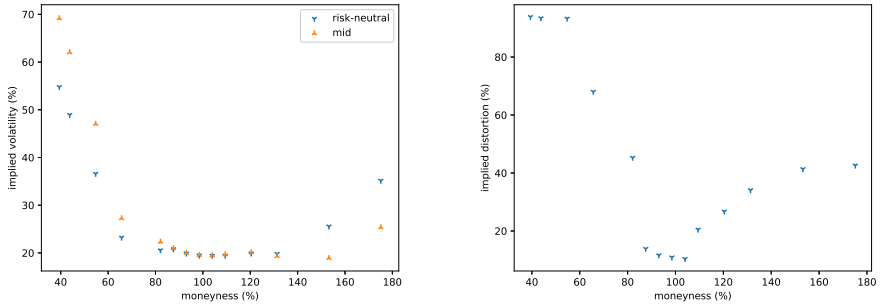


Figure 16: Risk-neutral and mid implied volatilities, panel (a), and implied liquidity levels, panel (b), for the options considered in Figure 14.

Also in this case mid prices are good proxies for their risk-neutral counterparts; see Figure 14b. This is again due to the fact that close to the at-the-money point liquidity is high, and far from it, even if liquidity decreases, options are not very sensitive to volatility changes.

Overall, we see that for very liquid instruments mid implied volatilities are very well approximated by their risk-neutral counterparts, as expected. When liquidity decreases, however, for in-the-money and out-of-the-money options implied risk-neutral volatilities might differ noticeably from mid volatilities. Predictably, this does not have a considerable impact on option prices. This is because close to the at-the-money point liquidity is in general high, making risk-neutral and mid implied volatilities close to each other. On the other hand, for in-the-money and out-of-the-money options liquidity can considerably affect volatilities. However, these options have low vegas, which makes their prices not very sensitive to changes in the volatility of the underlying. Nonetheless, whether risk-neutral and mid prices are close to each other is beside the point: for both risk-neutral and mid prices there is no liquidity in the market, so from a trading perspective only bid and ask prices matter. What we are interested in is assessing whether implied volatilities can be extracted from traded option quotes in a consistent manner with the risk-neutral framework. In particular, what we believe is important is to assess how considering bid and ask prices as a starting point instead of their mid counterparts can affect the shape of the volatility smile. As we have seen in the examples considered, computing implied volatilities from bid and ask prices instead of from mid prices in some cases can have a large impact on the implied volatility figures, and this can have implications in different contexts. As an example, if a smile model is calibrated by means of a least-square approach to the available implied volatilities, then differences

as those observed in Figures 13a and 16a would result in risk-neutral and mid volatility smiles with different shapes, as in-the-money and out-of-the-money implied volatilities would affect the calibration as a whole. Furthermore, when simulation models for over-the-counter derivatives, as for instance credit models, calibrated to implied volatilities are used, then choosing to input risk-neutral rather than mid implied volatilities might have a non-marginal impact for running contracts which, trough time, due to market movements ended up being in-the-money or out-of-the-money. Therefore, the examples considered, as well as the theoretical consistency of the methodology outlined in this article with the risk-neutral paradigm (paradigm that is not satisfied when mid prices are considered), make liquidity-free implied implied volatilities a potentially useful tool in financial modeling.

3.6 CONCLUSION

In this article we have considered the problem of computing implied volatilities from bid and ask European option prices directly, i.e., without relying on mid price approximations. The methodology we have outlined relies on the conic finance framework of Cherny and Madan (2009). Based on the results of Michielon et al. (2021a) it is possible, given the bid and ask prices of an option, to imply both the risk-neutral volatility and the liquidity level of the market at the same time. In particular, in the case of Black-Scholes and Bachelier specifications, this procedure results particularly efficient when the Wang transform is used, as the latter allows to analytically compute bid and ask option prices. In the case of the Bachelier model, these analytical formulae have been provided. The methodology outlined in this article relies on some intuitive and simple assumptions concerning the liquidity level of the market and the wideness of the range of option prices with respect to changes in the volatility parameter. A potential application for the technique we propose is that of constructing liquidity-free implied volatilitiy surfaces (and, consequently, corresponding implied liquidity surfaces at the same time). These liquidity-free implied volatility surfaces could be used as calibration inputs for different models under risk-neutral settings in a consistent manner.

3.A PROOF OF THEOREM 2

In order to prove Theorem 2, we provide two technical lemmas. The results presented here are based on Michielon et al. (2021a), and they have been made available here to make the article self-contained. If we assume that the quoted bid and ask prices of the contingent claim Y lie within the interval of all the possible risk-neutral prices that can be obtained by changing the parameter λ, and that for every λ in a given interval it is always possible to find a distortion parameter γ such that the observed bid-ask spread can be reproduced, Lemma 4 follows.

Lemma 4. *Let Y be a contingent claim whose price depends on a parameter $\lambda > 0$ such that the risk-neutral price of Y is strictly increasing with respect to λ. Assume that the inequalities $\inf_{\lambda>0} \mathrm{PV}(Y(\lambda)) < b$ and $\sup_{\lambda>0} \mathrm{PV}(Y(\lambda)) > a$ hold. Then, there exists an interval $[\lambda_b, \lambda_a]$ such that there is equivalence between $b \leqslant \mathrm{PV}(Y(\lambda)) \leqslant a$ and $\lambda \in [\lambda_b, \lambda_a]$. Further, assume that for every $\lambda \in [\lambda_b, \lambda_a]$ there exists $\gamma > 0$ such that $\mathrm{ask}(Y(\lambda), \gamma) - \mathrm{bid}(Y(\lambda), \gamma) = a - b$. Then, such γ is unique.*

Proof. The existence of the interval $[\lambda_b, \lambda_a]$ immediately follows from $\mathrm{PV}(Y(\lambda))$ being an increasing and continuous function of λ.

We now consider a fixed $\lambda \in [\lambda_b, \lambda_a]$. There exists at least one $\gamma > 0$ such that $\mathrm{ask}(Y(\lambda), \gamma) - \mathrm{bid}(Y(\lambda), \gamma) = a - b$. To derive a contradiction, suppose that there exist γ_* and γ^* satisfying the relationships $\mathrm{ask} Y(\lambda), \gamma_* - \mathrm{bid}(Y(\lambda), \gamma_*) = \mathrm{ask}(\lambda, \gamma^*) - \mathrm{bid}(\lambda, \gamma^*) = a - b$, where $\gamma_* < \gamma^*$. We recall that the ask (bid) price of $Y(\lambda)$ is an increasing (decreasing) function of γ. From this it follows that $\mathrm{ask}(\lambda, \gamma_*) < \mathrm{ask}(\lambda, \gamma^*)$, and that $\mathrm{bid}(\lambda, \gamma_*) > \mathrm{bid}(\lambda, \gamma^*)$. Therefore, we obtain that $a - b = \mathrm{ask}\lambda, \gamma_* - \mathrm{bid}(\lambda, \gamma_*) < \mathrm{ask}\lambda, \gamma^* - \mathrm{bid}(\lambda, \gamma^*) = a - b$, which leads to a contradiction. □

We now prove the following continuity-related result.

Lemma 5. *Under the hypotheses of Lemmas 4, for every $\lambda \in [\lambda_b, \lambda_a]$ the function such that $\lambda \mapsto \gamma(\lambda)$, where $\mathrm{ask}(Y(\lambda), \gamma(\lambda)) - \mathrm{bid}(Y(\lambda), \gamma(\lambda)) = a - b$, is continuous.*

Proof. We consider $\bar{\lambda}$ in $[\lambda_b, \lambda_a]$ fixed, as well as a sequence $(\lambda_n)_n$ in $[\lambda_b, \lambda_a]$ converging to $\bar{\lambda}$. Further, let $\phi(\lambda, \gamma) \coloneqq \mathrm{ask}(Y(\lambda), \gamma) - \mathrm{bid}(Y(\lambda), \gamma)$. We proceed in three steps.

(i) We first prove that the sequence $(\gamma(\lambda_n))_n$ is bounded. By contradiction, assume this is not the case. It then follows that there exists a subsequence $(\gamma(\lambda_{n_k}))_k$ of $(\gamma(\lambda_n))_n$ diverging to $+\infty$. The sequence $(\lambda_{n_k})_k$ converges to $\bar{\lambda}$, as it is a subsequence of a convergent sequence. Further, recall that ϕ is continuous in both arguments. Therefore, it results that $\lim_k \phi(\bar{\lambda}_{n_k}, \gamma(\lambda_{n_k})) = \phi(\bar{\lambda}, +\infty) = a - b$, as $\phi(\bar{\lambda}_{n_k}, \gamma(\lambda_{n_k}))$ always equals $a - b$, by construction. As assumed in Lemma 4 there exists $\bar{\gamma} > 0$ such that $\phi(\bar{\lambda}, \bar{\gamma}) = a - b$. Due to ϕ being increasing

in its second argument, we obtain that $a - b = \phi(\bar{\lambda}, \bar{\gamma}) < \phi(\bar{\lambda}, +\infty) = a - b$, contradiction.

(ii) We now prove that the sequence $(\gamma(\lambda_n))_n$ admits limit. As $(\gamma(\lambda_n))_n$ is bounded, it has a convergent subsequence. Assume that there exist two subsequences, denoted as $(\gamma(\lambda_{n_k}))_k$ and $(\gamma(\lambda_{n_h}))_h$, that converge to γ_* and γ^*, respectively, with $\gamma_* < \gamma^*$. The sequences $(\lambda_{n_k})_k$ and $(\lambda_{n_h})_h$ are subsequencies of the same convergent sequence. Therefore, they both converge to $\bar{\lambda}$. We then obtain that $a - b = \lim_k \phi(\lambda_{n_k}, \gamma(\lambda_{n_k})) = \phi(\bar{\lambda}, \gamma_*) < \phi(\bar{\lambda}, \gamma^*) = \lim_h \phi(\lambda_{n_h}, \gamma(\lambda_{n_h})) = a - b$, contradiction (the first and the last equalities are due to the definitions of $(\lambda_{n_k})_k$ and $(\lambda_{n_h})_h$, respectively, the second and the penultimate equalities follow from the continuity of ϕ, while the inequality is a result of ϕ being increasing in its second argument). This means that every convergent subsequence of $(\gamma(\lambda_n))_n$ has the same limit. As $(\gamma(\lambda_n))_n$ is bounded, it admits a limit itself (recall that if all the convergent subsequences of a bounded sequence converge to the same real limit, then the sequence itself also converges to the same limit as well).

(iii) Lastly, we now prove that the limit of $(\gamma(\lambda_n))_n$ is $\gamma(\bar{\lambda})$. We denote $\lim_n \gamma(\lambda_n)$ as $\bar{\gamma}$, and recall that ϕ is continuous in both arguments. The sequence $(\phi(\lambda_n, \gamma(\lambda_n)))_n$ is constant by construction, as it always equals $a - b$, and thus it converges to $a - b$. Its limit equals $\phi(\bar{\lambda}, \bar{\gamma})$ due to the continuity of ϕ is continuous. As a consequence of Lemma 4, there exists a unique $\gamma(\bar{\lambda})$ such that the equality $\phi(\bar{\lambda}, \gamma(\bar{\lambda})) = a - b$ holds. From this it follows that $\bar{\gamma} = \gamma(\bar{\lambda})$. □

Now, as a consequence of the lemmas above, Theorem 2 can be proven.

Proof of Theorem 2. Consider the interval $[\lambda_b, \lambda_a]$ as per Lemma 4. It follows that there exists a unique γ_b such that the equality $\text{ask}(\lambda_b, \gamma_b) - \text{bid}(\lambda_b, \gamma_b) = a - b$ holds. As $\text{bid}(\lambda_b, \gamma_b) < \text{PV}(\lambda_b) = b$, we obtain that $b < \text{ask}(\lambda_b, \gamma_b) < a$.

Analogously, consider λ_a. There exists a unique γ_a for which the equality $\text{ask}(Y(\lambda_a), \gamma_a) - \text{bid}(Y(\lambda_a), \gamma_a) = a - b$ is satisfied. Because $a = \text{PV}(Y(\lambda_a)) < \text{ask}(Y(\lambda_a), \gamma_a)$, it results that $b < \text{bid}(Y(\lambda_a), \gamma_a) < a$.

The functions $\text{ask}(Y(\lambda), \gamma)$, $\text{bid}(Y(\lambda), \gamma)$ and $\gamma(\lambda)$ are continuous in λ (the latter statement is a consequence of Lemma 5). Therefore, there exists $\bar{\lambda} \in (\lambda_b, \lambda_a)$ and a corresponding $\bar{\gamma}$ such that $\text{ask}\big(Y(\bar{\lambda}), \bar{\gamma}\big) = a$ and such that $\text{bid}\big(Y(\bar{\lambda}), \bar{\gamma}\big) = b$. From Lemma 4 the pair $(\bar{\lambda}, \bar{\gamma})$ satisfying (36) is unique. □

3.B DERIVATION OF THE BACHELIER OPTION PRICING FORMULAE

Equation (35) describes the dynamics of an Ornstein-Uhlenbeck process. Thus, (35) admits a solution which, at time T, is of the form

$$X_T = e^{(r-\alpha)T} X_0 + \sigma \int_0^T e^{(r-\alpha)(T-t)}\, dW_t. \tag{44}$$

From (44) it follows that, under $\mathbb{Q}$, X_T is normally distributed and, further, its mean and variance can be calculated analytically. More precisely, we have that its mean equals

$$\bar{\mu} := e^{(r-\alpha)T} X_0, \tag{45}$$

while its variance

$$\bar{\sigma}^2 := \frac{\sigma^2 (e^{2(r-\alpha)T} - 1)}{2(r-\alpha)}. \tag{46}$$

The present value of an European call option written on X, with maturity T and strike price K, is given by

$$\begin{aligned} \mathcal{C}^B(\bar{\mu}) &= e^{-rT} \mathbb{E}^{\mathbb{Q}} \left((X_T - K)^+ \right) \\ &= e^{-rT} \int_K^{+\infty} \frac{x-K}{\bar{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\bar{\mu}}{\bar{\sigma}}\right)^2} dx \\ &= e^{-rT} \left[(\bar{\mu} - K) \Phi\left(\frac{\bar{\mu}-K}{\bar{\sigma}} \right) + \bar{\sigma} \phi\left(\frac{\bar{\mu}-K}{\bar{\sigma}} \right) \right]. \end{aligned} \tag{47}$$

One can then retrieve the value of the put option in a similar manner:

$$\begin{aligned} \mathcal{P}^B(\bar{\mu}) &= e^{-rT} \mathbb{E}^{\mathbb{Q}} \left((K - X_T)^+ \right) \\ &= e^{-rT} \int_{-\infty}^K \frac{K-x}{\bar{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\bar{\mu}}{\bar{\sigma}}\right)^2} dx \\ &= e^{-rT} \left[(K - \bar{\mu}) \Phi\left(\frac{K-\bar{\mu}}{\bar{\sigma}} \right) + \bar{\sigma} \phi\left(\frac{K-\bar{\mu}}{\bar{\sigma}} \right) \right]. \end{aligned} \tag{48}$$

Observe that (47) and (48) satisfy the generalized put-call parity relationship

$$e^{-\alpha T} X_0 - e^{-rT} K + \mathcal{P} - \mathcal{C} = 0,$$

see (Haug, 2007, Ch. 1.2.1).

3.C A REMARK ON A PROPERTY OF THE WANG TRANSFORM

As noted in Wang (2000), transforming a (log)normal random variable via the Wang transform allows to still obtain a (log)normal random. However, we

point out that this property still holds every time a normal random variable is transformed by means of a non-decreasing and left-continuous function. To show this, assume that a random variable X is normally distributed with mean μ and variance σ^2 with respect to a probability measure $\mathbb{P}$. Further, let $f : \mathbb{R} \to \mathbb{R}$ be a non-decreasing function. Denote with $f^{-1}(\cdot)$ its inverse should $f(\cdot)$ be strictly increasing, or its pseudo-inverse otherwise. In the latter case this means, given $y \in \mathbb{R}$, that

$$f^{-1}(y) \coloneqq \inf\{x \in \mathbb{R} : f(x) > y\}. \tag{49}$$

Let Z be a standard normal random variable. We obtain that[10]

$$\begin{aligned}\mathbb{P}(f(X) \leqslant u) &= \mathbb{P}(f(\mu + \sigma Z) \leqslant u) \\ &= \mathbb{P}\left(Z \leqslant \frac{f^{-1}(u) - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{f^{-1}(u) - \mu}{\sigma}\right).\end{aligned} \tag{50}$$

Therefore, by applying the Wang transform with distortion parameter γ to (50), which we denote with $\psi_\gamma(\cdot)$, it follows that

$$\psi_\gamma(\mathbb{P}(f(X) \leqslant u)) = \Phi\left(\frac{f^{-1}(u) - \mu + \gamma\sigma}{\sigma}\right). \tag{51}$$

Thus, (51) shows that the distribution of X is invariant with respect to the Wang transform, up to upgrading its mean from μ to $\mu - \gamma\sigma$.

10 We recall the following property arising from (49): given y and z in $\mathbb{R}$, $f(z) \leqslant y$ if and only if $z \leqslant f^{-1}(y)$. First of all, observe that $f^{-1}(y)$ in (49) can be rewritten as $\sup\{x \in \mathbb{R} : f(x) \leqslant y\}$. Therefore, if $f(z) \leqslant y$, then $z \in \{x \in \mathbb{R} : f(x) \leqslant y\}$, from which $z \leqslant f^{-1}(y)$. On the other hand, if $z \leqslant f^{-1}(y)$, there exists an increasing sequence $(x_n)_n$ in $\{x \in \mathbb{R} : f(x) \leqslant y\}$ such that $\lim_n x_n = f^{-1}(y)$. Therefore, for every n, $f(x_n) \leqslant y$. For the left-continuity of f it follows that $f(f^{-1}(y)) = \lim_n f(x_n) \leqslant y$. As $z \leqslant f^{-1}(y)$, given that f is non-decreasing it results that $f(z) \leqslant y$.

4 NEURAL NETWORK EMPOWERED LIQUIDITY PRICING IN A TWO-PRICE ECONOMY UNDER CONIC FINANCE SETTINGS USING MULTILAYER PERCEPTRONS

4.1 ABSTRACT

Due to the nowadays extraordinarily demanding performance requirements in financial engineering, and the availability of enormous datasets, machine learning techniques have become more and more popular in quantitative finance with the aim of boosting computational efficiency and exploiting patterns from market-observable quantities, among other things. However, so far almost no investigation has been carried out to link the usage of machine learning technologies with pricing and hedging in a two-price economy. This latter area is of great practical interest, as in financial markets the law of one price does not hold and (derivative) contracts are traded in a direction-dependent manner. In this respect, in the last decade the developments of the conic finance paradigm have made many steps forwards. However, at the time of writing no attempts have been made yet to explore the capabilities of machine learning models to value and risk manage financial derivatives under conic settings. This article represents the first attempt to apply machine learning techniques in conic financial markets specifically for valuation and risk-management purposes. In particular, in the manuscript at hand multilayer perceptrons are put into use to model liquidity in financial markets in two different contexts. That is, on the one hand this paper illustrates how the use of neural networks within a two-price economy allows to obtain extremely accurate pricing and Greeks at lightning fast speed, potentially enhancing computational performances by orders of magnitude compared to classical approaches such as (conic) Monte Carlo. In this respect, this article also illustrates how to calculate sensitivities under conic settings by means of neural networks, and investigates the potential additional benefits and/or drawbacks of using deep neural networks as well. The methodology proposed here for this purpose is agnostic of the underlying valuation model, and it easily adapts to all models suitable for pricing in conic financial markets. On the other hand, this article also investigates the possibility of valuing contingent claims, in the presence of bid-ask spreads under conic assumptions, using local stochas-

tic volatility models where the local volatility component is approximated by means of a (combination of) neural networks. This to avoid employing a parametric form for the local volatility function of the model. Moreover, as far as distortion functions are concerned, in this manuscript we also show how it is possible to generate hybrid families of distortion functions to better fit the implied liquidity of the market. All the research outlined in the current article is novel. Over and above, to the best of our knowledge our work is the first to introduce the usage of neural networks for valuation and risk-management purposes within the conic two-price paradigm, and no doubt significantly contributes to the developments of high-performance computing in markets with frictions, as well as to further enhance the flexibility of local stochastic volatility modeling.

4.2 INTRODUCTION

The article at hand investigates two methodologies, based on neural networks, that allow to value and risk manage financial derivatives in markets with frictions.

With the constantly increasing complexity of valuation models (and models in general) used within the financial industry, and the need for high-performance calculations and versatile modeling frameworks, machine learning is carving out more and more importance in financial applications. Furthermore, due to the native flexibility belonging to these techniques as approximation tools, their expanding popularity in this sector has further risen. In this paper we propose two techniques, based on (potentially deep) neural networks, allowing to compute prices and sensitivities of financial derivatives in a *two-price* economy. This assuming financial markets can be described according to the *conic finance* paradigm of Cherny and Madan (2010). In particular, this article illustrates two techniques leveraging the use of neural networks to price liquidity in markets with frictions. That is, it first illustrates how to use (vector-valued) *multilayer perceptrons* to allow to jointly calculate bid and ask prices of contingent claims, as well as their sensitivities. The methodology is showed to be robust, lightning fast, and to be able to produce results which are very accurate. Furthermore, we propose a novel enhancement in local stochastic volatility (LSV) modeling by adapting the framework proposed by Cuchiero et al. (2020), which entails approximating the local volatility component of an LSV model by means of a (combination of) neural networks. In the article at hand, we extend the technique outlined in Cuchiero et al. (2020) to cope with markets with frictions by means of conic Monte Carlo simulations, as well as provide further enhancements at risk-neutral calibration level.

In financial modeling it is often common proceeding to evaluate derivative positions under a chosen risk-neutral measure. However, this is of course not in

line with what is observed in financial markets, where the the law of one price does not hold. Namely, in reality contingent claims are traded at prices depending on whether market agents have the intention to either buy or sell a given contract. Therefore, this introduces frictions in financial markets, which raises the need for a valuation framework able to replicate the dynamics observed. This is a peculiarly important feature especially as far as illiquid markets are concerned. In this respect the conic finance approach of Cherny and Madan (2009) comes in handy. This is because the aforementioned theory elegantly allows to extend a given risk-neutral valuation framework to bid-ask pricing by noticing that the latter can be obtained by distorting the risk-neutral distribution of the underlying asset(s) considered by means of concave distortion functions[1] from $[0, 1]$ onto itself. Thus, by loosing the (numerable) additivity property of probability measures, (asymmetric) Choquet integrals Choquet (1954) (see also (Denneberg, 1994, Ch. 5) and (Wang and Klir, 2009, Ch. 11)) represent the prototypical valuation functionals to be used within this context.

The conic finance theory, since the writing of the seminal paper Cherny and Madan (2009), has been very successful, with applications in the most diverse areas. That is, this modeling setting has been employed for the valuation of structured products Madan and Schoutens (2012) (as in the case of constant proportion portfolio insurances Marquet and Schoutens (2018), to name but one), in capital and valuation adjustment calculations van Bakel et al. (2020); Madan and Schoutens (2011b); Madan and Schoutens (2016b); Madan (2009); Madan (2012), in portfolio theory Madan (2016), option pricing Madan and Schoutens (2017); Guillaume and Schoutens (2013) (including the case of cryptocurrency markets Madan et al. (2019)), as well as in many other areas. In particular, we point out the importance of the conic finance modeling environment as it allows to calculate implied liquidity amounts, which provide a "universal" and intuitive way to measure the liquidity of the market, in an analogous way as implied volatilities measure the market expected variability in the underlying asset values Corcuera et al. (2012); Albrecher et al. (2013); Guillaume (2015); Michielon et al. (2021b). For an extensive collection of applications of the conic finance theory the interested reader can refer to Madan and Schoutens (2016a).

For financial institutions performing bid and ask pricing in an efficient, robust, reliable and realistic manner is imperative from many points of view. To name but a few, real-time valuation and quoting of financial derivatives are crucial activities for a conspicuous amount of financial firms (both on the buy and sell sides), as well as the related risk-management duties, which are in general computationally more expensive than the former due to the often required need of using finite difference approximations to compute the Greeks of interest. Not to mention (sensitivity- and mark-to-market-based) end-of-day reporting tasks to be used for risk limits and capital allocation. Besides, financial institutions also

1 Here often referred to solely as "distortions", for brevity.

need to compute bid-ask reserves and, equally important (if not more so), within prudent accounting (whenever relevant), they should mark their assets and liabilities at the bid and ask prices, respectively; see Madan and Schoutens (2011a). All these requirements, to which many others could be added, even strengthen the need for high-performance computing, as well as for valuation models which faithfully render observed market behaviors, in markets with frictions. However, the fact that markets are ruled by the *law of two prices* introduces additional challenges as far as speed, accuracy and calibration are concerned. That is, under the settings of Cherny and Madan (2010), calibration and pricing routines would need an additional step, i.e., the calibration within this frame of reference of the distortion function to the observed bid and ask market quotes. This, on top of an initial calibration under risk-neutral settings. And, not to forget, the estimation of Greeks needs to be performed for both bid and ask prices, respectively, basically doubling the number of function evaluations required. Therefore, in a more and more technology-driven financial environment, innovation and development are strongly bound to their computational counterparts, making high performance computing and low-latency key success factors for banks, trading houses, and basically almost all firms operating in capital markets. Machine learning algorithms often yield remarkable advancements in this domain.

The use of machine learning techniques has found applications in almost all disciplines, ranging from medical diagnosis to speech and image recognition, as well as from product recommendations for marketing purposes to self driving cars, to name but a few (see, e.g., Johri et al. (2020) for an overview). And of course, also the financial industry has, in the last decade, proven to be a fertile ground in this respect. To start with, applications that affect retail consumers and the society as a whole include, amongst other things, financial fraud detection Awoyemi et al. (2017); Sadgali et al. (2019) and anti money laundering Domashova and Mikhailina (2021); Chen et al. (2018). And it is natural to expect that financial modeling and quantitative analysis are not an exception in this respect. That is, different methodologies driven by machine learning algorithms have flourished in the last years to ameliorate (computational) speed (almost) without compromising accuracy; see, e.g., De Spiegeleer et al. (2018); Davis et al. (2021) (a broad overview of the most relevant (and up-to-date) innovations in the field can be found in Section 4.4). However, to the best of our knowledge, there is currently no literature addressing how to perform accurate, realistic and efficient bid and ask pricing in financial markets where contingent claims are traded in a direction-dependent manner under conic settings while using machine learning techniques. The article at hand provides two novel and pioneering methodologies, based on neural networks, that represent the first attempt to bridge this gap in the literature. The results outlined in the document at hand clearly illustrate, both qualitatively and qualitatively, how the usage of machine learning techniques can really be adopted also when deviating from the risk-neutral valuation framework. More in detail, in this paper we inves-

tigate two applications of machine learning (in particular, neural networks and, even more specifically, multilayer perceptrons) that can be used to value and risk manage financial contracts in a two-price economy.

The choice of adopting neural networks is driven and motivated by several factors. To start with, neural networks offer a high degree of architectural flexibility as they allow for arbitrary (up to memory constraints of course) numbers of layers and nodes, as well as for the possibility of choosing different neuron activation functions by selecting from a plethora of possible functional relationships to capture the non-linearities between inputs and outputs. Further, due to their common usage in fields such as image recognition (and related areas), which require processing billions of pixels at a lightning fast speed, neural network implementations naturally allow for GPU acceleration features, which enable to speed up calculations by orders of magnitude compared to CPU architectures Oh and Jung (2004). And the fact that neural networks are basically defined as (sequential) compositions of (potentially vector-valued) linear and non-linear function further benefits in this respect. Moreover, neural networks allow for easy multi-valued regression implementations Borchani et al. (2015), which is one possible way of interpreting bid-ask pricing. Last but not least, the choice of using neural networks and, in particular, multilayer perceptrons, is further supported by the literature; see, e.g., Morelli et al. (2004); Bishop (2006), amongst others.

This article offers a compelling new perspective to bid ask pricing that adds significant value to the ongoing research in efficient computing and realistic contingent claim valuation in markets with frictions, bringing to the table several novel and meaningful contributions that shed new light on the usage of neural networks within the conic finance paradigm of Cherny and Madan (2010). That is, (i) as far as we can tell the manuscript at hand is the first paper providing a fully neural network based architecture for derivative valuation in a two-price economy under the conic framework of Cherny and Madan (2010). Additionally, one of the techniques we have developed (ii) allows to replicate theoretical bid and ask prices with a high level of accuracy. Further, being the approach proposed based on neural networks, it allows to (iii) build a model-agnostic framework. That is, a setting compatible with bid and ask prices (and also able to provide the relevant sensitivities) generated by any model (and/or numerical technique) suitable for the pricing architecture of Cherny and Madan (2010). Moreover, the aforementioned approach (iv) allows for extraordinary computational performance improvements that can reach orders of magnitude compared to standard techniques such as, e.g., (conic) Monte Carlo simulation. This bid-ask calibration procedure also provides, at the same time, a fast and efficient risk-neutral model calibration as bonus benefit. We also (v) extend the aforementioned technique to compute sensitivities of potentially any order (for both liquidity or any other market factors), and additionally demonstrate how the culmination of "learning the Greeks" provides an auxiliary benefit, that is, a po-

tential reduction in the size of the training set. Furthermore, the work outlined in this paper also (vi) provides novel developments in local stochastic volatility modeling not only by enhancing the framework developed in Cuchiero et al. (2020), where it is proposed to model local volatility functions by means of (combinations of) neural networks, but also by extending it to bid-ask pricing by means of conic Monte Carlo techniques. On the side, (vii) we also introduce the conic version of the SABR Hagan et al. (2014) model, based on the Wang distortion Wang (2000), which still allows for analytical bid and ask pricing formulae. Yet, (viii) we provide simple techniques to generate arbitrary families of hybrid distortion functions that can be used as decision making tools for conic modeling purposes, and that can potentially allow for an improved bid-ask calibration fit.

The manuscript at hand is organized as follows. Section 4.3 provides the necessary tools and essential notions related to conic pricing under the paradigm of Cherny and Madan (2009). In particular, therein Section 4.3.1 recalls how to perform bid-ask valuation using the Wang transform Wang (2000) (with Section 4.3.1 introducing the conic SABR model), while Section 4.3.2 illustrates how to estimate bid and ask prices of contingent claims by means of simulations following the conic Monte Carlo paradigm (Madan and Schoutens, 2016a, Ch. 5.3). Furthermore, Section 4.3.3 delineates a simple methodology to generate arbitrary families of distortion functions. The notions introduced in Section 4.3 will be required for the applications provided in the current paper and have been provided to make the document at hand self-contained. Section 4.4 presents a literature review concerning the most recent and relevant applications of machine learning in financial engineering and quantitative analysis. This allows to clearly illustrate where the work proposed in the document at hand stands, at the time of writing, in relation to the available literature. Section 4.5 introduces the main concepts related to neural networks and, in particular, to multilayer perceptrons (see Section 4.5.1), as our approaches build upon these notions. The two approaches investigated in this paper are presented in Sections 4.6 and 4.7. In particular, starting with the former, in Section 4.6 we illustrate how to perform bid and ask pricing by means of a (vectorized) neural network regression. More in detail, Section 4.6.1 provides a motivating example that shows, under simple settings, how this technique is inherently intuitive, and recognizes vector-valued multilayer perceptrons as a natural tool to be employed. Going forward, Section 4.6.2 further elaborates in this respect, also providing a methodology to calibrate to bid and ask prices of European options in an arbitrage-free manner (see Section 4.6.2 therein). Real world illustrations exemplifying the potency of the methodology in the context of rough volatility modeling are available in Section 4.6.2. Section 4.6.3 illustrates how the suggested methodology can be extended to compute Greeks, and motivates the proposal by means of explanatory and elucidative instances. Section 4.6.4 provides some practical considerations concerning the choice of the activation function(s) to be used, while Section 4.6.5

relates to the extension of the methodologies proposed in the context of deep learning. On the other hand, in Section 4.7 our work related to LSV modeling can be found. That is, therein Section 4.7.1 provides the basic notions related to LSV models, while Section 4.7.2 illustrates how it is possible to approximate the local volatility component of an LSV model by means of (a combination of) neural networks following the work of Cuchiero et al. (2020). Moreover, Section 4.7.3 outlines how, under risk-neutral settings, LSV models equipped with a neural network component can be calibrated to the available market observable prices. Further, Section 4.7.4 provides an LSV calibration example using real market data and, in particular, in Section 4.7.4 the aforementioned technique is extended to a two-price economy by means of conic Monte Carlo techniques. Section 4.8 concludes. Finally, an illustration of the methodology described in Section 4.6 is provided in the case of the conic SABR model within Appendix 4.A[2].

4.3 BID–ASK PRICING WITH DISTORTED EXPECTATIONS: AN INTUITIVE INTRODUCTION

We start by considering a probability space $(\Omega, \mathcal{F}, \mathbb{Q})$, where $\mathbb{Q}$ denotes a chosen pricing (i.e., risk-neutral) probability measure. Furthermore, we consider a contingent claim, paying out at a given (known) future time, which we represent with the random variable X. For simplicity, and for the ease of notation as well, from here onwards (and without loss of generality) we will assume the time value of money to be zero (otherwise, discounted prices could be considered, instead).[3]

If the contingent claim X trades under the law of one price (still dependent on the probability measure $\mathbb{Q}$), then it can be bought or sold for the same amount $\mathbb{E}^{\mathbb{Q}}(X) = \int_\Omega X \, d\mathbb{Q}$, which can be rewritten as

$$\mathbb{E}^{\mathbb{Q}}(X) = \int_{-\infty}^{0} \mathbb{Q}(X \geqslant x) - 1 \, dx + \int_{0}^{+\infty} \mathbb{Q}(X \geqslant x) \, dx, \tag{52}$$

whenever the two integrals in (52) are not both infinite of opposite sign.

Assume now that a concave (distortion) function $\psi : [0,1] \to [0,1]$ such that $\psi(0) = 0$ and $\psi(1) = 1$ is available. One can then note, after re-weighting the $\mathbb{Q}$-probabilities in (52) by means of $\psi(\,\cdot\,)$, that

$$\int_{-\infty}^{0} \psi(\mathbb{Q}(X \geqslant x)) - 1 \, dx + \int_{0}^{+\infty} \psi(\mathbb{Q}(X \geqslant x)) \, dx \geqslant \mathbb{E}^{\mathbb{Q}}(X). \tag{53}$$

2 Note that the aim of this section is solely that of illustrating how the methodology available in this paper is model-agnostic and can therefore be adapted to different models and pricing techniques.

3 This will be assumed to be always the case.

Therefore, the left-hand-side of (53), named (asymmetric) Choquet integral of X with respect to the distorted probability measure $\psi \circ \mathbb{Q}$ (see (Denneberg, 1994, Ch. 5)), can be naturally interpreted as an ask price for X. This is because, via (53), higher (lower) weights are given to the high (low) realizations of X. From here onwards we will denote the left-hand-side of (53) as $(C)\int_\Omega X \, d\psi \circ \mathbb{Q}$. By recognizing that buying X is equivalent to sell $-X$, one than obtains that $-(C)\int_\Omega -X \, d\psi \circ \mathbb{Q}$ can be interpreted as a natural candidate for the bid price of X. This is because, in this case, lower (higher) weights are given to high (low) payoffs of X.

However, the definitions just given for bid and ask prices are not "operational" yet. That is, one would need the distortion function to depend on (at least) one parameter, in such a way that changing the value(s) of the latter would allow to replicate the bid and ask spread observed in the market (at least in a least-square sense should this not be possible in an exact manner). To this purpose Cherny and Madan (2009) propose to consider an increasing family of distortion functions $(\psi_\lambda)_{\lambda \geqslant 0}$.[4] As a consequence, one can then obtain a wide range of λ-dependent bid and ask prices.[5] By assuming the dynamics of the underlying under $\mathbb{Q}$ are fully specified and known, one can then, by changing the values of λ, attempt to calibrate the theoretical conic prices to their market-observable counterparts by means of the relationships

$$\text{ask}_\lambda(X) = (C)\int_\Omega X \, d\psi_\lambda \circ \mathbb{Q} \tag{54}$$

and

$$\text{bid}_\lambda(X) = -(C)\int_\Omega -X \, d\psi_\lambda \circ \mathbb{Q} \tag{55}$$

where, for clarity, we emphasize that in (54) and (55) the same λ is used. We recall that the parameter λ and the Choquet representations of bid and ask prices are strongly linked to the concepts of acceptability indices and coherent risk measures. We refer the reader to Cherny and Madan (2009); Cherny and Madan (2010) and (Madan and Schoutens, 2016a, Ch. 4) for a more detailed elucidation of the matter.

Albeit the formulae presented above are sufficient to perform bid and ask pricing of contingent claims (up to rescaling with the appropriate discount factors) we also point out, to ease intuition, that $\text{ask}_\lambda(X) = \text{ask}_\lambda(X^+) - \text{bid}_\lambda(X^-)$, and that $\text{bid}_\lambda(X) = \text{bid}_\lambda(X^+) - \text{ask}_\lambda(X^-)$ (here, we denote with X^+ and with X^- the positive and negative parts of X, respectively). That is, the first equality suggests that, to obtain the ask price of X, we need to "overprice" its positive part and

4 From $[0,1]$ to $[0,1]$ and such that $\psi_0(\,\cdot\,)$ coincides with the identity function. Also, in this context the term "increasing" should be interpreted as pointwise.

5 From here on, whenever the symbol λ is used, it would always refer to a distortion parameter ranging in $[0,+\infty)$.

"underprice" its negative part. Vice versa, to obtain its bid price we need to "underprice" its positive part and "overprice" its negative part. For further details we refer the interested reader to Eberlein et al. (2012).

4.3.1 Conic pricing of European options with the Wang transform

In the applications we provide in the article at hand great use has been made of the Wang transform (see, e.g., Section 4.6.2). We highlight here the reason why this concave distortion function plays an important role in the valuation of contingent claim in a direction-dependent trading environment.

The Wang transform was initially introduced in Wang (2000) to value both financial and insurance risks using Choquet expectations. In particular, in the framework outlined in Wang (2000) it is shown how this distortion function allows for an alternative construction of the Black-Scholes option pricing model, while being consistent with the Bühlman economic premium principle, as well as with the capital asset pricing model. In symbols, the Wang transform reads

$$\psi_\lambda(x) := \Phi(\Phi^{-1}(x) + \lambda), \tag{56}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal random variable, and $\Phi^{-1}(\cdot)$ indicates its inverse. As it is clear from (56), the Wang transform shifts the quantile of standard normal random variables by the positive[6] amount λ.

Definition (56) (see also (Madan and Schoutens, 2016a, Ch. 4.7.5)) is very useful in practice. In particular, the Wang transform has the convenient property that, if a random variable is lognormally distributed, then by compounding the distribution function of the latter with the Wang transform still a lognormal distribution is obtained (with adjusted mean). Similar considerations also hold for the normal case. In addition, it has been recently shown, see Michielon et al. (2021b), that this property still holds whenever a normal random variable is transformed by means of a non-decreasing function. Therefore, the use of the Wang transform has become very popular in conic pricing, as it allows to still obtain analytical formulae for bid and ask prices of European options under both Bachelier (see Michielon et al. (2021b)) and Black-Scholes (see (Madan and Schoutens, 2016a, Ch. 5.4)) settings.[7] Therefore, from this we can argue that the Wang transform is somehow the natural distortion to be used in option pricing given that vanilla options are often quoted in terms of their (lognormal or normal) implied volatilities.

6 Note that in the context of conic pricing λ is assumed to be non-negative (with the case $\lambda = 0$ reducing to the risk-neutral case). However, in other frameworks the Wang transform can be also considered for negative values of the distortion parameter (note that for $\lambda > 0$ the Wang transform is concave, while when $\lambda < 0$ it is convex).

7 And, of course, in all those situations where (log)normal distributions relate to the relevant options pricing formulae; see, e.g., (Haug, 2007, Ch. 1.3) for some examples.

Interlude: Conic SABR

The observations we made In Section 4.3.1 concerning the Wang distortion allow to make some interesting considerations concerning the SABR model Hagan et al. (2002); Hagan et al. (2014). This model is very popular in the financial industry and it has applications in different asset classes: mainly rates, but also applications of the SABR model (potentially with adaptations) can be found in foreign-exchange van der Stoep et al. (2015) and equity (Overhaus et al., 2007, Ch. 2.1.2) markets, amongst others.

For an underlying process $(Y_t)_{t\geqslant 0}$, the standard SABR definition reads

$$\begin{cases} dY_t = \sigma_t Y_t^{\beta}\, dW_t \\ d\sigma_t = \alpha\sigma_t dZ_t \\ d\langle W_t, Z_t\rangle = \rho dt \end{cases}, \tag{57}$$

where $\alpha > 0$, $\rho \in [-1,1]$ (i.e., the correlation coefficient between the two Brownian motions $(W_t)_{t\geqslant 0}$ and $(Z_t)_{t\geqslant 0}$), and where $\beta \in [0,1]$. In Hagan et al. (2014) it is shown that, up to higher order effects, the SABR specifications (57) still allow to use the Black formula for pricing European options, but this time with a "twicked" implied volatility dependent on a set of (calibrated) parameters and strikes.[8] That is, for a fixed maturity T and a given strike price K, the SABR model (57) allows for the use of the Black formula with input volatility given by

$$\sigma^{\text{SABR}}(K,T) \coloneqq \frac{x}{y}\cdot\frac{z}{\chi(z)}, \tag{58}$$

where

$$x \coloneqq \sigma_0\left\{1+\left[\frac{(1-\beta)^2}{24}\frac{\sigma_0^2}{(Y_0K)^{1-\beta}}+\frac{1}{4}\frac{\rho\beta\alpha\sigma_0}{(Y_0K)^{(1-\beta)/2}}+\frac{2-3\rho^2}{24}\alpha^2\right]T\right\},$$

$$y \coloneqq (Y_0K)^{(1-\beta)/2}\left[1+\frac{(1-\beta)^2}{24}\ln^2\left(\frac{Y_0}{K}\right)+\frac{(1-\beta)^4}{1920}\ln^4\frac{Y_0}{K}\right],$$

$$z \coloneqq \frac{\alpha}{\sigma_0}(Y_0K)^{(1-\beta)/2}\ln\left(\frac{Y_0}{K}\right),$$

$$\chi(z) \coloneqq \ln\left(\frac{\sqrt{1-2\rho z+z^2}+z-\rho}{1-\rho}\right),$$

and with σ_0 the initial value of the volatility.

By virtue of the aforementioned properties of the Wang transform outlined in Section 4.3.1 we obtain that, under conic settings and SABR risk-neutral model

8 Note that, in practice, the parameter β is often not calibrated but, instead, exogenously set equal to one (zero) in the case of lognormal (normal) SABR specifications, or to $\frac{1}{2}$ for "in-between" dynamics.

specifications, call and put option prices can be still computed analytically. That is, for a given liquidity level λ and having defined $\delta(=\delta(K,T;\lambda)) \coloneqq \frac{\lambda\sigma^{SABR}(K,T)}{\sqrt{T}}$, it holds that

$$\text{bid}_\lambda(\mathcal{C}(K,T)) = Y_0 e^{-\delta T}\Phi(d_1) - K\Phi(d_2), \tag{59}$$

$$\text{ask}_\lambda(\mathcal{C}(K,T)) = Y_0 e^{\delta T}\Phi(d_1) - K\Phi(d_2), \tag{60}$$

$$\text{bid}_\lambda(\mathcal{P}(K,T)) = K\Phi(-d_2) - Y_0 e^{\delta T}\Phi(-d_1), \tag{61}$$

and

$$\text{ask}_\lambda(\mathcal{P}(K,T)) = K\Phi(-d_2) - Y_0 e^{-\delta T}\Phi(-d_1), \tag{62}$$

where

$$d_1 \coloneqq \frac{\ln\left(\frac{Y_0}{K}\right) + \left(-\delta + \frac{\sigma^{SABR}(K,T)}{2}\right)T}{\sigma^{SABR}(K,T)\sqrt{T}}$$

and with

$$d_2 \coloneqq d_1 - \sigma^{SABR}(K,T)\sqrt{T}.$$

In formulae (59), (60), (61) and (62) $\mathcal{C}(K,T)$ ($\mathcal{P}(K,T)$) denotes the price of a European call (put) option with strike K and expiry T.

4.3.2 Conic Monte Carlo

We now briefly describe how to perform bid and ask pricing of financial derivatives, within the conic finance framework, by means of simulations. This methodology will be used in the numerical examples provided later on in this paper (see Section 4.7).

Under risk-neutral settings, given a (potentially path-dependent) contingent claim paying out at a fixed future time, we assume that N scenarios $\omega_1,\dots,\omega_N$ have been sampled to obtain the corresponding payoffs $x_1 \coloneqq X(\omega_1),\dots,x_N \coloneqq X(\omega_N)$. One would just need to average the latter to obtain the current value of the contract (again, up to discounting). This is because each outcome is considered equally likely.

In a two price economy, on the other hand, the realizations of the contingent claim across the different states of nature need to be weighted differently. Intuitively, giving more (less) weight to high (low) payoffs allows to obtain an ask (a bid) price for X (see beginning of Section 4.3). That is, assume that the outcomes $x_1,\dots,x_N$ are, without loss of generality, ordered in a non-decreasing way. To calculate the bid and ask prices of X one needs to weight the sampled realizations in an alternative fashion. That is, for each $i \in \{1,\dots,N\}$ and a given liquidity level $\lambda \in [0,+\infty)$, the probability to be assigned to x_i equals

$$p_i^{\text{bid}_\lambda} \coloneqq \psi_\lambda\left(\frac{i}{N}\right) - \psi_\lambda\left(\frac{i-1}{N}\right). \tag{63}$$

Contrariwise, that to be used to compute the ask price should equal

$$p_i^{\mathrm{ask}_\lambda} \coloneqq \psi_\lambda\left(\frac{N-i+1}{N}\right) - \psi_\lambda\left(\frac{N-i}{N}\right). \tag{64}$$

Thus, denoting with $x \coloneqq (x_1, \ldots, x_N)$, $p^{\mathrm{bid}_\lambda} \coloneqq (p_1^{\mathrm{bid}_\lambda}, \ldots, p_N^{\mathrm{bid}_\lambda})$ and $p^{\mathrm{ask}_\lambda} \coloneqq (p_1^{\mathrm{ask}_\lambda}, \ldots, p_N^{\mathrm{ask}_\lambda})$, to value in a two-price economy one just needs to compute bid and ask prices by means of the scalar products

$$\mathrm{bid}_\lambda(X) = \langle x, p^{\mathrm{bid}_\lambda}\rangle \tag{65}$$

and

$$\mathrm{ask}_\lambda(X) = \langle x, p^{\mathrm{ask}_\lambda}\rangle, \tag{66}$$

respectively.[9] For the derivations of formulae (65) and (66), which basically depend on a discretized version of the Choquet integral (see (Wang and Klir, 2009, Ch. 11.5)), we refer the reader to (Madan and Schoutens, 2016a, Ch. 5.2).

4.3.3 Hybrid distortions

A distortion function needs to satisfy some basic properties; see Section 4.3. That is, it should map zero to zero and one to one (to guarantee that null and almost-sure events are still treated as such), and it also needs to be concave. Furthermore, a family of distortion functions $(\psi_\lambda)_{\lambda\geqslant 0}$ needs to be increasing in λ, and such that $\psi_0(\,\cdot\,)$ coincides with the identity of $[0,1]$ onto itself. Common families of concave distortion functions are the Minmaxvar distortion, the Wang distortion, the Maxvar distortion and the Minvar distortion; see (Madan and Schoutens, 2016a, Ch. 4.7) for an overview.

In practical applications, the choice of the distortion function is somehow arbitrary. For instance, as seen in Section 4.3.1 (and Section 4.3.1 therein), the Wang transform is often used for European option pricing as it allows to obtain, under normal or lognormal dynamics, analytical pricing formulae. However, to the best of our knowledge, there is currently no methodology available providing some rules of thumb for distortion function selection. And, further, the ultimate aspect that a modeler is interested in is given by how well a given distortion function allows to replicate the bid ask spreads observed in the market. Thus, in this section we propose some simple techniques that allow to generate arbitrary families of distortions, which generalize the currently available ones, and that can be used as test functions to choose whether one distortion function is better than another, or whether something "in-between" would fit even better. Therefore, one might seek to combine the characteristics and specifications of different

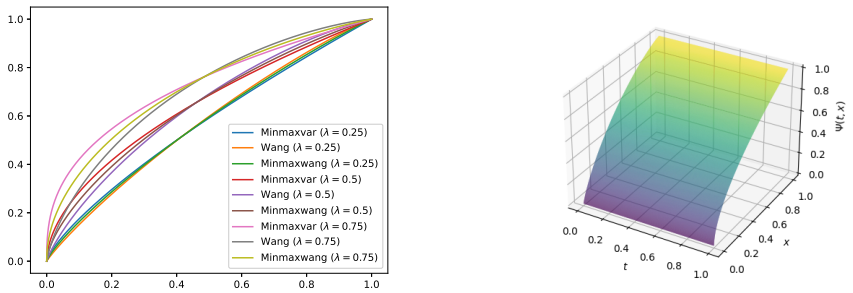
9 From (63) and (64) it is obvious that the case where $\psi_\lambda(\,\cdot\,)$ equals the identity function the standard uniform weighting case is retrieved.

distortion families to introduce more flexibility to the modeling framework. And here we provide some intuitive ways to do so.

Initially, given N (λ-dependent) distortion functions $\psi_\lambda^1(\cdot), \ldots, \psi_\lambda^N(\cdot)$, observe that any convex linear combination of them is still a concave distortion function. This means that, if we consider the combined distortion

$$\tilde{\psi}_\lambda(x) := \sum_{i=1}^{N} \alpha_i \psi_\lambda^i(x) \tag{67}$$

such that, for every $i \in \{1, \ldots, N\}$, $\alpha_i \geqslant 0$ and with $\sum_{i=1}^{N} \alpha_i = 1$, we still obtain a concave distortion. Furthermore, it also trivially follows that the family $(\tilde{\psi}_\lambda)_{\lambda \geqslant 0}$ is increasing with respect to λ and that $\tilde{\psi}_0(\cdot)$ coincides with the identity function on $[0, 1]$.



(a) (b)

Figure 17: Graphical depiction of the Minmaxwang distortion given different levels of the distortion parameter λ, panel (a), and of the continuous mapping $H(\cdot, \cdot)$ transforming the Minmaxvar distortion into the Wang distortion, panel (b).

As an illustration of this, we take into account the Minmaxvar[10] and the Wang distortions, i.e., we set $N = 2$ and assume that $\psi_\lambda^1(\cdot) := \psi_\lambda^{\text{Minmaxvar}}(\cdot)$, and that $\psi_\lambda^2(\cdot) := \psi_\lambda^{\text{Wang}}(\cdot)$, with $\alpha_1 := 1 - t$ and $\alpha_2 := t$ (for $t \in [0, 1]$). Intuitively, this way we have created a family of distortions that, continuously, transforms the Minmaxvar distortion into the Wang distortion (and the other way round). That is, we are simply considering the straight-line homotopy between $\psi_\lambda^{\text{Minmaxvar}}(\cdot)$ and $\psi_\lambda^{\text{Wang}}(\cdot)$, i.e., the map $H : [0, 1] \times [0, 1] \to [0, 1]$ given by $H(t, x) := (1 - t) \cdot \psi_\lambda^{\text{Minmaxvar}}(x) + t \cdot \psi_\lambda^{\text{Wang}}(x)$ such that $H(0, x) = \psi_\lambda^{\text{Minmaxvar}}(x)$, while $H(1, x) = \psi_\lambda^{\text{Wang}}(x)$. In particular, we name the "average" distortion, obtainable when $t = 0.5$, from here onwards, Minmaxwang, for the ease of terminology (and

10 Defined as $\psi_\lambda^{\text{Minmaxvar}}(x) := 1 - \left(1 - x^{\frac{1}{\lambda+1}}\right)^{\lambda+1}$; see Cherny and Madan (2009).

t-Minmaxwang when $t \neq 0.5$). The Minmaxwang distortion is illustrated in Figure 17a, while the homotopy $H(\cdot, \cdot)$ in Figure 17b.

As Figure 17a illustrates, the Minmaxvar distortion tends to give higher (lower) weights to "low" ("high") probabilities, compared to the Wang distortion, especially for high values of λ. Thus, the Minmaxwang defines, in some sense, a more balanced choice.

The advantage of considering combined distortions as proposed here is that it makes it possible to have one[11] help parameter that allows the calibration routine for more flexibility. Given that every distortion function on the right-hand-side of (67) can be obtained by setting the appropriate parameters to zero, then it is guaranteed that (67) will never worsen the calibration results for any set of distortion functions initially chosen.

Remark 1. With reference to (Madan and Schoutens, 2016a, Ch. 4.7), we recall that two desirable properties of a distortion function are that

$$\lim_{x \to 0^+} \psi'_\lambda(x) = +\infty, \tag{68}$$

and that

$$\lim_{x \to 1^-} \psi'_\lambda(x) = 0. \tag{69}$$

One can then trivially observe that, in order for condition (69) to be verified for $\tilde{\psi}_\lambda(\cdot)$, then (69) needs to be verified for all the distortions $\psi^1_\lambda(\cdot), \ldots, \psi^N_\lambda(\cdot)$. On the other hand, for (68) to be verified by $\tilde{\psi}_\lambda(\cdot)$, then it is enough that one of the distortion functions $\psi^1_\lambda(\cdot), \ldots, \psi^N_\lambda(\cdot)$ is such that (68) holds.

We now propose another way of generating hybrid distortions. That is, we observe that the composition of the N distortion functions $\psi^1_\lambda(\cdot), \ldots, \psi^N_\lambda(\cdot)$ is still a valid distortion function. Therefore, if we start by taking into account, as done above, the Minmaxvar and Wang distortions, then we can define the Minmaxvar-Wang distortion as $\psi^{\text{Wang}}_\lambda \circ \psi^{\text{Minmaxvar}}_\lambda(\cdot)$. However, note that defining distortions using the methodology just highlighted is not a commutative operation. Therefore, we could also opt for the Wang-Minmaxvar distortion by inverting the order of the composition, i.e., by taking into account $\psi^{\text{Minmaxvar}}_\lambda \circ \psi^{\text{Wang}}_\lambda(\cdot)$. Examples, for different values of λ of the Minmaxvar-Wang and the Wang-Minmaxvar distortions are depicted in Figure 18.

11 Or more, if additional distortions are also considered.

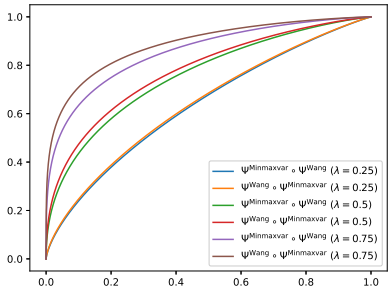


Figure 18: Minmaxvar-Wang and Wang-Minmaxvar distortions, given different levels of λ.

From the relationships between the Minmaxvar-Wang and the Wang-Minmaxvar distortions observed in Figure 18 we therefore expect, assuming similar calibration accuracy, that to the Wang-Minmaxvar distortion would correspond lower implied liquidity levels compared to the Minmaxvar-Wang and, therefore, different levels of liquidity deltas. These considerations should be taken into account in liquidity hedging considerations.

Remark 2. Let us consider here the simple case given by the combination of two distortion functions

$$\hat{\psi}_\lambda(\,\cdot\,) \coloneqq \psi^1_\lambda \circ \psi^2_\lambda(\,\cdot\,).$$

It then results, for $x \in [0,1]$, that

$$\hat{\psi}'_\lambda(x) = (\psi^1_\lambda)'(\psi^2_\lambda(x)) \cdot (\psi^2_\lambda)'(x).$$

By definition, both $\psi^1_\lambda(\,\cdot\,)$ and $\psi^2_\lambda(\,\cdot\,)$ map zero to zero and one to one. Therefore, we obtain that condition (68) is satisfied whenever one of the derivatives of $\psi^1_\lambda(\,\cdot\,)$ or $\psi^2_\lambda(\,\cdot\,)$ diverges at zero (provided that the other, if any, is non-zero). Similarly, condition (69) is verified whenever one of the derivatives of $\psi^1_\lambda(\,\cdot\,)$ or $\psi^2_\lambda(\,\cdot\,)$ tends to zero at one (provided that the other, if any, is finite).

4.4 MACHINE LEARNING TECHNIQUES IN QUANTITATIVE FINANCE: A (BRIEF) LITERATURE REVIEW

The scientific literature is teeming with applications of machine learning in financial engineering and quantitative finance. We provide here a literature review concerning some important (recent) developments in the field, as well as some additional ones to be used as general references. Of course, given the exceptional amount of research produced in the last decade in this respect, this

list is far from being exhaustive. However, we believe it to be rich enough to provide a clear idea of where the frameworks illustrated in the article at hand can be placed compared to the published material until now.

As some very relevant examples of applications of machine learning in financial engineering we recall De Spiegeleer et al. (2018), where machine learning techniques are applied to pricing problems within a risk-neutral framework (see also Davis et al. (2021) for specific applications of gradient boosting in the same area and Buehler et al. (2019) for hedging considerations). In particular, in De Spiegeleer et al. (2018) it is illustrated how the development of classical machine learning techniques can come in handy in several financial areas such as valuation and calibration whenever speed becomes essential, at the price of slightly loosing in accuracy (despite the latter remaining satisfactory from an operational angle). It does not come as a surprise that also various applications of machine learning techniques are available as far as algorithmic trading strategies are concerned. For instance, in Yang et al. (2015) regression-based strategies are used to attempt to characterize the specific behavior of (algorithmic) traders. This domain of study aims to estimate the impact automation has in financial markets (which bears great significance from a regulatory perspective). We also recall Madan and Sharaiha (2020), where regression methodologies and different performance indicators such as Sharpe ratios and acceptability indices (the latter, thus, linked to the concept of distorted expectations) are combined with the aim of improving the predictive power of algorithmic trading strategies. For a survey refer to Rundo et al. (2019); see also Dixon et al. (2020); Ni et al. (2021).

In the context of neural networks, we mention Ferguson and Green (2018) and Morelli et al. (2004), where neural networks are applied for derivative valuation in an economy where the law of one price holds, and where it is shown how neural networks are excellent tools in terms of structural flexibility that can offer performance speed-ups of orders of magnitude compared to classical pricing techniques (see also Hernandez (2017b); Hernandez (2017a)). Notably, the research proposed in Morelli et al. (2004) leverages on the natural capabilities of neural networks to approximate non-linear functions to accurately replicate option prices and sensitivities (due to the universal approximator property of neural networks). In particular, therein it is shown how the degree of accuracy and the speed improvements produced by the usage of neural networks are very relevant for practical use. Furthermore, in Morelli et al. (2004) it is also shown how the multilayer perceptron architecture[12] for neural networks stands out in terms of performances. Besides, still in this area we recall Liu et al. (2019); Liu et al. (2019) and Ye and Zhang (2019). Therein, the computational added value in pricing and implied volatility calculations that can be obtained via the implementation of neural networks is investigated and commented on in a positive

12 This type of architecture, which is the one used in the applications that can be found in the article at hand, is outlined in Section 4.5.1.

manner given the performance enhancements accomplished. For more specific applications of neural networks (and, in particular, of the class of feed-forward neural networks, to which multilayer perceptrons belong to) in quantitative finance and, in particular, in the case of (stochastic) local volatility modeling, we refer to Cuchiero et al. (2020); Chataigner et al. (2022) and Wang et al. (2022), amongst others. Primarily, the common thread of these research papers is that of attempting not to rely on a specific parametric functional form for the modeling of the local volatility function (and not even on its classical non-parametric Dupire's counterpart Dupire (1994)). But, rather, to deploy a machine learning approximator for the local volatility model component able to perform the task of matching vanilla option prices and "interpolating" them in an arbitrage-free manner. For a general overview related to various applications of neural networks in the banking industry and further applications see, amongst others, Huang et al. (2020).

One of the main advantages of performing calculations using neural networks is that given by their extraordinary computational performances. This is because, in a nutshell, a neural network is nothing but a (nested) combination of linear and non-linear functions. This, therefore, makes them by construction potentially efficient tools with a broad range of applications. In particular, as in nowadays financial markets speed and performances play a key and pivotal role, several authors have investigated the improvements arising from the utilization of GPUs in quantitative analysis compared to the use of CPUs. And there are different reasons to conduct research in this area. The predominant factor behind this is given by the massive parallelization of computations that can be achieved with the former compared to the latter. That is, GPUs are architectures with thousands of small processing units enabling for multiple simultaneous operations. This compared to common CPUs which consist of only a handful (compared to GPUs) of cores. Therefore, whenever calculations can be broken down into several small and independent pieces, GPUs offer tremendous performance improvements compared to CPUs. Further, being GPUs optimized for graphical processing, they are much more efficient than CPUs when it comes to handling vectorized and matrix operations, which are core in neural network architectures.[13] Over the last few years there has been a growing body of research on this subject domain. We refer the reader to Crépey et al. (2019), Crépey and Dixon (2019) and Albanese et al. (2020), where the usage of neural networks (and other machine learning techniques) and the related GPU features are investigated in the context of valuation adjustment calculations for performance enhancement. These articles are of importance as, nowadays, financial institutions are required to calculate different valuation adjustments (such as CVA, FVA, KVA and MVA,

13 There are also other factors that play a role in this case such as, for instance, memory bandwidth. However, we do not discuss them here as a detailed tractation of this topic falls beyond the aim of the paper at hand. For a benchmark analysis of different processing architectures in deep learning refer to Wang et al. (2019).

amongst others). These adjustments are in general computationally intensive as they require, in principle, to be able to value (potentially-netted) portfolio values throughout a given time frame. Therefore, this introduces several and severe computational challenges. An illustration of this is given by the computations of XVA Greeks (relevant, e.g., for MVA considerations), which might become prohibitively expensive with standard techniques. Therefore, the combination of machine learning and hardware architectures such as GPUs offers a potentially exceptionally powerful set of tools to help managing the performance standards required nowadays. As a result, the interest of several researchers in investigating these possibilities further follows.

However, as far as conic finance is concerned, at the time of writing and to the best of our knowledge, Chopra (2020) represents the only attempt to use machine learning in a conic economy. However, the research conducted therein mainly applies to pairs trading and index performance tracking, which is outside the domain we are dealing with in the manuscript at hand.

4.5 NEURAL NETWORKS

Neural networks are machine learning tools that can be used in different areas, such as classification and regression. In particular, their name derives from the attempt to detect mathematical relationships between the interactions of the components of biological systems (Bishop, 2006, Ch. 5). Amongst all kinds of neural networks, the multilayer perceptron is the most popular and successful type of network architecture; see, amongst others, (Bishop, 2006, Ch. 5). In this section, therefore, we will provide the basic notions that relate to this kind of neural network, as they represent the building blocks needed for the methodologies outlined in this paper. Note that, from here onwards, whenever we refer to a neural network, we will always refer to multilayer perceptron.

4.5.1 Multilayer perceptrons

There are different types of neural networks, and one of the most successful ones is the *feed-forward* neural network architecture, also known as *multilayer perceptron*. Very attractive due to its approximation properties, in this network the information only travels in one direction, from inputs to outputs, from which the nomenclature "feed-forward" originates. In a nutshell, a feed-forward neural network is a nested composition of multiple regression models, each one continuous and non-linear, with affine transformations in between such layers. That is, we start by considering $K \in \mathbb{N} \setminus \{0\}$. Consequently, we consider $d_0, \dots, d_K \in \mathbb{N} \setminus \{0\}$, and $h : \mathbb{R} \longrightarrow \mathbb{R}$. For every $k \in \{1, \dots, K\}$, let $A_k : \mathbb{R}^{d_{k-1}} \longrightarrow \mathbb{R}^{d_k}$ be an affine function such that $x \mapsto w_k x + \beta_k$, with

$w_k \in \mathbb{R}^{d_k} \times \mathbb{R}^{d_{k-1}}$ and $\beta_k \in \mathbb{R}^{d_k}$ (here, the w_k's are called *weights*, while the β_k's *biases*). A function $NN : \mathbb{R}^{d_0} \longrightarrow \mathbb{R}^{d_K}$, such that

$$NN(x) = A_K \circ \Phi_{K-1} \circ \cdots \circ \Phi_1 \circ A_1(x), \tag{70}$$

is called a *feed-forward neural network*, where all the functions of the form $\Phi_k : \mathbb{R}^{d_k} \longrightarrow \mathbb{R}^{d_k}$ ($k \in \{1, \ldots, K-1\}$) in (70) are called *activation functions*, and are defined by means of the relationship

$$\Phi_k(x_1, \ldots, x_{d_k}) \coloneqq (h(x_1), \ldots, h(x_{d_k})).$$

The number K is the number of *layers* in the network, with $K-1$ the number of *hidden layers*. Often, terminology-wise, the number K is also called *depth* of the neural network (a neural network with more than one hidden layer is called *deep neural network*). The dimensions of the hidden layers are represented by $d_1, \ldots, d_{K-1}$ and, furthermore, d_0 and d_K represent the dimensions of the input and output layers, respectively. In Figure 19 we represent a schematic graphical depiction of a feed-forward neural network architecture with $d_0 = 2$, $d_1 = 3$, and with $d_2 = 1$.

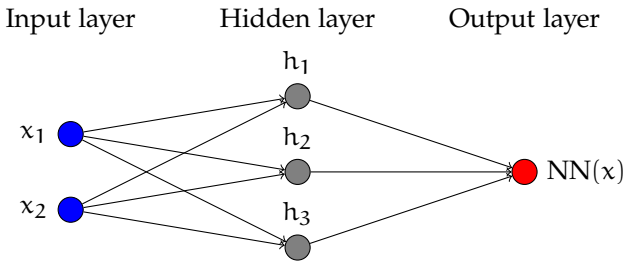


Figure 19: Sample graphical depiction of a feed-forward neural network with one hidden layer.

In order to calibrate (or, in data analytics jargon, to "train") a neural network, one would need to apply an algorithm to calculate the optimal (in some sense) weights and biases (i.e., the *trainable parameters*). When referring to the trainable parameters we shall refer to $\theta = (\theta_1, \ldots, \theta_K)$, where $\theta_k \coloneqq (w_k, \beta_k)$, for $k \in \{1, \ldots, K\}$. In the training process we are given a set x of inputs of size N, and a set of corresponding outputs, y. The set containing both inputs and outputs is also referred to as the *training set*. The set of inputs is mapped to an output by means of the transformation $NN(\theta, x)$, where θ indicates the dependence on the training parameters. A loss (or error) function is then employed to measure

the distance between the neural network output $NN(\theta, x)$ and the target output y. In particular, the calibration error $\epsilon(\theta)$ is defined as[14]

$$\epsilon(\theta) \coloneqq \sum_{i=1}^{N} \|NN(\theta, x_i) - y_i\|^2, \tag{71}$$

which needs to be minimized to determine the set of optimal parameters. Thereafter, a sequence of weights is constructed by considering the iterative formula

$$\theta_{n+1} \coloneqq \theta_n - \nu \cdot \nabla \epsilon(\theta_n), \tag{72}$$

where in (72) the parameter $\nu > 0$ is often called *learning rate*. Relationship (72) is also known in the literature in the context of error *back-propagation*; see (Bishop, 2006, Ch. 5.3). The iterations stop whenever a given error criteria is satisfied. Relationship (72) is updated based, at each iteration, on a training sample chosen randomly. This is also known as *stochastic gradient descend*; see (Bishop, 2006, Ch. 5.2.4). Note that the learning rate parameter in (72) can be in principle defined in a non-constant manner (see, e.g., Bottou and Bousquet (2007)), but in practice the common choice, at least most of the times, is that of defining it as constant.

There are of course different alternatives to be used to estimate the optimal weights (refer to equation (71) for the quantity to be minimized) for the neural network. We name here a few, for completeness. A very common choice in weight optimization is the so-called *adaptive moment estimation* (Adam), introduced in Kingma and Ba (2014), which is a computationally efficient methodology for stochastic optimization based on (adaptive) estimates of lower moments. This technique is often a common choice for many practical applications. Given the popularity and the performances of the Adam optimization procedure, in (some of) the examples provided in this paper (see Section 4.6) we will use this technique. However, for completeness we mention here some other very popular choices (without of course the aim of being exhaustive given the enormous amount of literature available in this respect), as the *ADADELTA* (adaptive learning rate algorithm) Zeier (2015), or the *ADAGRAD* (adaptive subgradient method) Duchi et al. (2011), amongst disparate others.

14 In this paper, we denote with $\|\cdot\|$ the Euclidean norm in $\mathbb{R}^N$. Again, from a notation perspective, we denote with $\cdot$ the product of a scalar and a vector, and we also denote, for the ease of notation, with x^2 ($x \in \mathbb{R}^N$) the scalar product $\langle x, x \rangle$. Further, observe that, sometimes, (71) is defined slightly differently, for convenience. That is, it is often rescaled by $\frac{1}{2}$ (as in (Bishop, 2006, Ch. 5)) as doing so allows for simpler back-propagation formulae. Or, for instance, sometimes the average of the error components is taken into account (as in Kingma and Ba (2014)). Note, however, that as soon as monotonic transformations of the error function are taken into account, results are in principle not affected.

4.6 (DEEP) CONIC NEURAL NETWORKS: LIQUIDITY PRICING AND HEDGING WITH VECTOR-VALUED MULTILAYER PERCEPTRONS

In this section we in-depth illustrate a methodology that allows to compute bid and ask prices in a two price economy under the conic finance paradigm of Cherny and Madan (2010). In particular, we will outline how neural networks can be used for bid-ask calibration purposes. In the current section, from here onwards we will call every neural network to be used to produce prices of contingent claims in a two price economy a *conic neural network*, and we will denote it with the acronym CNN[15].

4.6.1 A motivating example

We first start by taking into account a simple example to provide a step-by-step illustration of how to calibrate (and, therefore, price) in a two-price economy under conic finance settings by using CNNs.

We begin with a very simple framework solely for illustrative purposes. That is, we consider an economy such that, in a risk-neutral context, the dynamics of a selected (non-dividend paying) underlying asset of interest follow the specifications of the Black-Scholes model. Assume that, for the underlying chosen, a set of European call options, all with the same maturity, are quoted. Furthermore, we also hypothesize that the bid and ask prices of these quoted options under these simplistic settings are consistent with the conic paradigm. In particular, we assume that they can all be obtained, starting from their risk-neutral prices, by means of applying the Wang distortion Wang (2000) (see also Section 4.3.1) with the same constant distortion parameter. Thus, we are in a simple economy where both risk-neutral and bid-ask pricing can be performed by means of analytical formulae [Ch. 5.4]Madan and Schoutens (2016a) (see also Section 4.3.1). Besides, assume that the underlying price $Y_0 = 10$, that the time to maturity $T = 1$, that the constant volatility parameter $\sigma = 25\%$, and that the distortion parameter $\lambda = 12.5\%$. Also, it is supposed that the discount rate is zero and that the strike prices $K_1, \dots, K_{11}$ of the quoted options belong to $\{7.5, 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12, 12.5\}$, i.e., that they correspond to the 75% to 125% option moneyness levels (equally spaced by 0.5 currency units). We denote the risk-neutral option prices as $\mathcal{C}_1, \dots, \mathcal{C}_{11}$, while their corresponding bid and ask prices as $b_1, \dots, b_{11}$ and $a_1, \dots, a_{11}$, respectively. In this case, $b_1, \dots, b_{11}$ and $a_1, \dots, a_{11}$ represent our (synthetic) market dataset.

15 Not to be confused with convolutional neural networks, which are not relevant within this framework.

In order to train the neural network that will be used for model calibration, one needs to proceed per steps (we assume here that the market data, i.e., the underlying price Y_0 and the zero risk-free rate are fixed and static, for illustration purposes).[16] That is, first of all one needs to select a large-enough intervals where the (to-be-estimated) volatility and liquidity parameters are assumed to belong. In this case, we consider $[0, 50\%]$ for the volatility, $[0, 25\%]$ for the distortion. We also consider the range $[7, 13]$ for the strike price. One can then specify a grid $0 =: \sigma_1 < \ldots < \sigma_N := 50\%$ for the volatility parameter, $0 =: \lambda_1 < \ldots < \lambda_N := 25\%$ for the distortion parameter, and $\tilde{K}_1 := 7, \ldots, \tilde{K}_N =: 13$[17] for the strike price (with respect to this example we select 25 equally-distant points in all the cases considered, that is, $N = 25$).[18] For each pair $(\sigma_i, \lambda_j, K_k)$ (here, i, j and k all range in $\{1, \ldots, 25\}$) one can then generate the corresponding theoretical bid and ask prices by using the analytical bid and ask option pricing formulae (Madan and Schoutens, 2016a, Ch. 5.3). Afterwards, the training of the CNN can take place.[19] In this case, therefore, we are training a neural network which attempts to approximate the function, from $[0, +\infty)^3$ to $[0, +\infty)^2$, such that $(\sigma, \lambda, K) \mapsto (\text{bid}_\lambda(\mathcal{C}(\sigma, K)), \text{ask}_\lambda(\mathcal{C}(\sigma, K)))$, where $\text{bid}_\lambda(\mathcal{C}(\sigma, K))$ ($\text{ask}_\lambda(\mathcal{C}(\sigma, K))$) represents the bid (ask) price of a call option $\mathcal{C}$ with, under Black-Scholes settings, an implied volatility of σ, a distortion parameter of λ, and a strike price equal to K. Therefore, the first component of the neural network, i.e., $\text{CNN}(\cdot, \cdot, \cdot)_1$, denotes the bid price, while the second one, i.e., $\text{CNN}(\cdot, \cdot, \cdot)_2$, the ask price, instead.

Once the training of the neural network has terminated, our model approximator is ready for the calibration phase. Thus, we first consider $\text{CNN}(\cdot, 0, \cdot)$ and perform the minimization

$$\min_{\sigma \geqslant 0} \sum_{i=1}^{11} \left(\mathcal{C}_i - \frac{1}{2}(\text{CNN}(\sigma, 0, K_i)_1 + \text{CNN}(\sigma, 0, K_i)_2) \right)^2. \tag{73}$$

Note that in (73) we have considered the average of the bid and the ask prices, calculated using the CNN, after having set the distortion parameter equal to zero. Observe that, theoretically speaking, both of them should coincide with

16 Note, however, that this assumption is by no mean restrictive, as it would be sufficient to enlarge the training dataset of the CNN by considering different values for the underlying and for the discount rate.

17 Here we use the tilde superscript for the training grid in the strike direction to distinguish this dataset from the "market-observables" $K_1, \ldots, K_{11}$.

18 We remark that, in principle, also considering (random and) uniformly distributed parameters is a possible choice, as well as considering a different amount of training points for the volatility compared to that of the distortion parameter or to that of the strike price. Decisions should always be made on the base of experimentation depending on the model specifications considered. Note, however, that as far as the distortion parameter is concerned, zero should always belong to the grid, as needed for the preliminary risk-neutral calibration as it will be shown later.

19 In all the examples provided in the paper at hand, the Adam optimizer (see Section 4.5.1, as well as Kingma and Ba (2014) for further details), has been used.

the corresponding risk-neutral prices of the options taken into account. Thus, in practice one could have just considered $\mathrm{CNN}(\sigma, 0, K)_1$ or $\mathrm{CNN}(\sigma, 0, K)_2$ directly, instead of their average $\frac{1}{2}(\mathrm{CNN}(\sigma, 0, K)_1 + \mathrm{CNN}(\sigma, 0, K)_2)$. However, due to numerical approximations, $\mathrm{CNN}(\sigma, 0, K)_1$ and $\mathrm{CNN}(\sigma, 0, K)_2$ are never exactly the same, so for instance one could choose to approximate the risk-neutral price with the average $\frac{1}{2}(\mathrm{CNN}(\sigma, 0, K)_1 + \mathrm{CNN}(\sigma, 0, K)_2)$.[20]

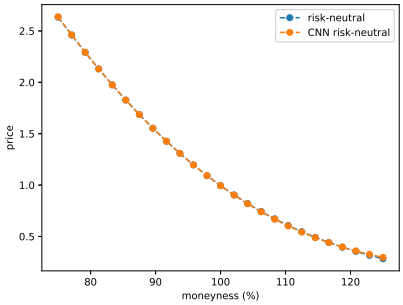


Figure 20: Synthetic risk-neutral prices approximated by $\mathrm{CNN}(\cdot, 0, \cdot)$ by means of $\frac{1}{2}(\mathrm{CNN}(\cdot, 0, \cdot)_1 + \mathrm{CNN}(\cdot, 0, \cdot)_2)$.

Figure 20 illustrates how $\mathrm{CNN}(\cdot, 0, \cdot)$ approximates the synthetic risk-neutral prices $\mathcal{C}_1, \ldots, \mathcal{C}_{11}$. Once an optimal σ has been estimated as described in the former step (which we denote with $\bar{\sigma}$), the bid-ask calibration routine can finally take place. This means that the second minimization step that aims to calculate the implied liquidity parameter can at this point start. That is, the minimization

$$\min_{\lambda \geqslant 0} \sum_{i=1}^{11} \left((\mathrm{bid}(\mathcal{C}_i) - \mathrm{CNN}(\bar{\sigma}, \lambda, K_i)_1)^2 + ((\mathrm{ask}(\mathcal{C}_i) - \mathrm{CNN}(\bar{\sigma}, \lambda, K_i)_2)^2 \right) \tag{74}$$

should be performed. Once an estimate for the optimal distortion parameter, denoted with $\bar{\lambda}$, has been found, then the two-price model has been calibrated, and it can thus be used for valuation and sensitivity calculation purposes. A graphical depiction of the calibrated bid and ask prices computed by means of the CNN has been provided in Figure 21, which shows how the prices obtained by means of the CNN basically overlap with those of the synthetic market data set.

20 Also observe that, as far as the liquidity parameter is concerned, for the CNN zero is the left-most point that belongs to the training dataset (see footnote 18). Therefore, one could also argue that considering the average $\frac{1}{2}(\mathrm{CNN}(\sigma, \epsilon, K)_1 + \mathrm{CNN}(\sigma, \epsilon, K)_2)$ for ϵ small would likewise be an appropriate choice. This to be sure that the approximated risk-neutral price belongs to the interpolation area of the CNN. However, from the practical experiments we have performed, it seems that both choices perform well enough with marginal and negligible differences.

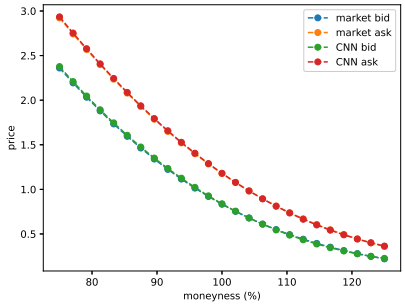


Figure 21: Synthetic bid and ask prices approximated by $\mathrm{CNN}(\bar{\sigma}, \bar{\lambda}, \cdot)$.

Please note that, as far as the distortion parameter is concerned, in practice one would need to estimate a different distortion parameter per strike (similar to what is done with respect to the implied volatility, which is computed in a strike-dependent manner). I.e., one would need to perform, for every i, the minimization

$$\min_{\lambda \geqslant 0} \left((\mathrm{bid}(\mathcal{C}_i) - \mathrm{CNN}(\bar{\sigma}, \lambda, K_i)_1)^2 + ((\mathrm{ask}(\mathcal{C}_i) - \mathrm{CNN}(\bar{\sigma}, \lambda, K_i)_2)^2 \right),$$

given that implied liquidity amounts are, in general, not constant across the strike direction. This will be shown in Section 4.6.2.[21]

4.6.2 CNNs in use

In this section we provide relevant applications of the idea concerning CNNs proposed in Section 4.6.1. In particular, from here onwards we will delineate some important and relevant use cases to showcase the remarkable potential of this methodology. That is, we will highlight how the use of neural networks in quantitative analysis expresses its best in situations where more standard procedures still require a considerable amount of time to provide the desired results. This is particularly true in the case of Monte Carlo methods which, despite different techniques can be adopted to improve their performances remain, generally speaking, computationally expensive.

From here onwards we will always assume, unless differently specified, that a given CNN has been previously trained. As already done in Section 4.6.1, without loss of generality we will consider the training with respect to a set of

21 Note that, in principle, one could also separate the calibration in two parts and determine a λ for the bid by minimizing $(\mathrm{bid}(\mathcal{C}_i) - \mathrm{CNN}(\bar{\sigma}, \lambda, K_i)_1)^2$, and one for the ask by minimizing $(\mathrm{ask}(\mathcal{C}_i) - \mathrm{CNN}(\bar{\sigma}, \lambda, K_i)_2)^2$. However, this would be slightly inconsistent, from a theoretical perspective, with (54) and (55), which state that the same value of the distortion parameter λ should be considered for the bid and the ask prices of the same contingent claim.

parameters θ, the distortion λ, as well as the strike price given that, in this article, we are dealing with options. However, extending the methodology to the case where the training step is performed with respect to different market data sets (and/or parameters) is similar. Thus, the steps listed here should not be viewed as limiting in any way.

Given $\theta = (\theta_1, \dots, \theta_N)$ we consider M possible values for θ_1, as well as for the remaining $N-1$ parameters in θ, M possible values for the distortion parameter λ, and M possible values for the strike price K. That is, we construct a grid of points $\{\theta_1^1, \dots, \theta_1^M\}$ for θ_1, until $\{\theta_N^1, \dots, \theta_N^M\}$ for θ_N, together with M points for the distortion parameter $\{\lambda^1, \dots, \lambda^M\}$ and for the strike price $\{K^1, \dots, K^M\}$ too.[22] For each of the $(N+2)^M$ possible parameter combinations $(\theta_1^{i_1}, \dots, \theta_N^{i_N}, \lambda^{i_{N+1}}, K^{i_{N+2}})$, where the indices $i_1, \dots, i_{N+2}$ range in $\{1, \dots, M\}$, one should generate option prices with a given chosen model. The CNN can be then trained following the step-by-step procedure described in Section 4.6.1. As already mentioned, note that the CNN is agnostic of the technique used to produce contract payoffs, so the latter can therefore be chosen at will.

A fully-fledged arbitrage-free calibration routine with CNNs: An illustration

We highlight here an approach to perform bid and ask calibration of European call[23] options by means of vector-valued multilayer perceptrons. The steps provided in this section provision a thorough and meticulous routine to perform conic calibration using CNNs.

To start we assume, given a fixed maturity T, that L call options $\mathcal{C}_1, \dots, \mathcal{C}_L$, with strike prices $K_1 < \dots < K_L$, respectively, are quoted in the market for a given underlying. We denote the value of the underlying asset, at the current time, by Y_0.[24] We denote with $b_1, \dots, b_L$ and with $a_1, \dots, a_L$ the corresponding bid and ask option prices quoted in the market. We now illustrate the procedure to be followed, step by step.

(i) Foremost, recall that calibrating to bid and ask quotes under conic finance settings first requires to perform a risk-neutral calibration as initial mandatory step. We highlight that, if for a given option bid and ask prices are quoted, then (any of) their risk-neutral prices would lie in-between them, but would not necessarily coincide with their average (i.e., with the mid price). As a common way to compare and quote European options is by

22 Observe that, in this case, we have chosen an equal number of points per variable. However, in principle this is not necessary. Furthermore, one can also choose whether to consider equally-spaced grids or not. One important remark is that λ^0 should always be equal to zero, as this is necessary due to the fact that training in a conic framework first requires calibrating under risk-neutral settings, circumstance that can be retrieved by setting $\lambda^0 = 0$ (see footnotes 18 and 20).

23 The case concerning put options is analogous.

24 From here onwards, for brevity and convenience, the index i will be always assumed to be ranging in $\{1, \dots, L\}$.

means of their implied volatilities, we consider here the framework introduced in Michielon et al. (2021b). That is, for each of the quoted options, we compute the correspondent *liquidity-free* implied volatility and related distortion parameter. This means that, in symbols, for the i^{th} option we solve the system

$$\begin{cases} \text{bid}^{BS}_{\lambda_i}(\sigma_i, K_i) = b_i \\ \text{ask}^{BS}_{\lambda_i}(\sigma_i, K_i) = a_i \end{cases} \tag{75}$$

with respect to σ_i and λ_i. In (75), $\text{bid}^{BS}_{\lambda_i}(\cdot, \cdot)$ and $\text{ask}^{BS}_{\lambda_i}(\cdot, \cdot)$ denote the conic Black-Scholes bid and ask functionals, assuming the Wang transform is used. Further, σ_i and λ_i represent the Black-Scholes implied volatility and the Wang implied distortion values for the option considered, respectively. The non-linear system in two equations and two unknowns (75) is guaranteed (under almost-always satisfied assumptions) to have a unique solution. We denote the so-calculated implied volatility and liquidity amounts with $\bar{\sigma}_i$ and $\bar{\lambda}_i$, respectively;

(ii) Once, as prescribed in (i), we have successfully managed to compute the liquidity-free implied volatilities $\bar{\sigma}_1, \dots, \bar{\sigma}_L$, we can then calculate their corresponding risk-neutral Black-Scholes prices, denoted as $BS(\bar{\sigma}_1, K_1), \dots, BS(\bar{\sigma}_L, K_L)$, respectively;

(iii) Note that the L risk-neutral call prices $BS(\bar{\sigma}_1, K_1), \dots, BS(\bar{\sigma}_L, K_L)$ are not automatically guaranteed to be arbitrage-free. Therefore, for the relevant arbitrage-free conditions to be satisfied, we apply the option data preprocessing *filtering methodology* proposed in Moussa (2018), which provides an approach for adjusting option data when no-arbitrage conditions are violated. This means that option prices, when no-arbitrage conditions are infringed, get updated in an arbitrage-free manner with respect to the other options taken into account in the dataset until no-arbitrage bounds hold again. Observe that applying this procedure, and therefore updating risk-neutral prices whenever necessary, is both theoretically and practically justified. This since risk-neutral prices are neither tradable nor observable. Therefore, as soon as they lie within their corresponding bid and ask counterparts in an arbitrage-free way, applying Moussa (2018) is a legitimate step;

(iv) After the filtering procedure has been applied, the actual calibration can take place. First of all, assume a distortion parameter of zero as, under these circumstances, the calibration needs to be performed at risk-neutral level first. That is, one needs to perform the minimization procedure

$$\min_{\theta \in \Theta} \sum_{i=1}^{L} \left(BS(\bar{\sigma}_i, K_i) - \frac{1}{2}(CNN(\theta, 0, K_i)_1 + CNN(\theta, 0, K_i)_2) \right)^2,$$

where Θ indicates a large-enough set containing the "true" risk-neutral model parameters. In order to find a good estimate for the parameters in θ one can for instance use a direction set method (Press et al., 2007, Ch. 10.7) or a (quasi-)Newton method (Press et al., 2007, Ch. 10.9), depending on the circumstances, after having estimated the initial guess by means, e.g., of a genetic algorithm as Kaelo and Ali (2007). We denote the risk-neutral parameters estimated as just described by $\bar{\theta}$;

(v) Now that $\bar{\theta}$ has been estimated and thus the CNN is calibrated under risk-neutral settings, the bid-ask calibration routine begins. That is, for each i, the implied liquidity parameter is computed by performing the minimization

$$\min_{\lambda_i \geqslant 0} \left(\left(b_i - \mathrm{CNN}(\bar{\theta}, \lambda_i, K_i)_1 \right)^2 + \left(a_i - \mathrm{CNN}(\bar{\theta}, \lambda_i, K_i)_2 \right)^2 \right). \tag{76}$$

By experimenting we observe that, at least as far as the Wang transform is concerned, using $\frac{1}{2}\bar{\sigma}_i$ as initial guess for a local minimization routine seems to work well.

At the current point all the parameters of the CNN have been estimated in such a way that the CNN is capable of pricing back (in a least-square sense) the quoted options. This now makes it suitable for pricing and risk-management purposes. In particular, bid and ask prices of non-quoted options can be calculated by feeding the CNN with the appropriate strike and liquidity levels (the liquidity levels of the market can be interpolated from the implied ones, for instance linearly or by using a monotonic interpolation Fritsch and Carlson (1980)).

A real world example

For illustrative purposes, we will perform experiments concerning options pricing based on the rough Bergomi model Bayer et al. (2016). Rough volatility models, based on the idea of rough volatility introduced in Gatheral et al. (2018), are becoming more and more popular in equity volatility modeling. However, they are in general particularly expensive to simulate. In our tests we take into account the rough Bergomi model as (i) rough volatility modeling is still a relatively new framework in financial engineering. Thus, we believe that considering an up-to-date financial modeling environment as benchmark adds value to the analysis provided here. Further, this choice (ii) provides a valuable example to illustrate the computational advantage coming from the use of neural networks for valuation and sensitivities calculation purposes in a market with frictions. And besides, (iii) to the best of our knowledge no examples of conic Monte Carlo combined with rough volatility modeling are available in the literature.

Therefore, we believe this is a fit-for-purpose choice to illustrate the methodology described in the article at hand. We want to highlight, as already mentioned in Section 4.4, that the methodology proposed in this article is model-agnostic. I.e., if another model were to take the place of the rough Bergomi model, then the methodology could be applied in a very similar manner. Further, the same considerations hold as far as the payoffs are concerned, as well as for the pricing technique[25] chosen. Therefore, the choice made here is, by no mean, neither restrictive nor specific.[26]

In this example, we will provide an illustration of the calibration routine described in Section 4.6.2. In particular, in this case we will follow the modeling framework of Buehler (2010), given that we consider, in our example, equity options on potentially dividend-paying underlying assets. That is, we start by assuming that a given underlying asset $(Y_t)_{t\geqslant 0}$ pays (cash) dividends at the dividend dates $\tau_1 < \ldots < \tau_D$; we denote these amounts as $d_1, \ldots, d_D$.[27] We denote with D_t the sum $\sum_{j:\tau_j > t} d_j$. From Buehler (2010) it follows, under the same pricing measure used for valuation purposes, that

$$Y_t = (F_t - D_t) \cdot U_t + D_t, \tag{77}$$

where in (77) $(U_t)_{t\geqslant 0}$ is a non-negative local martingale (under the same pricing measure) starting at one, and where $F_t := Y_t(\sum_{j:0<\tau_j\leqslant t} d_j)$. To evaluate a call option with maturity T one would then need to compute the amount

$$\mathcal{C}_Y(K,T) = (F_T - D_T) \cdot \mathcal{C}_U\left(\frac{K - D_T}{F_T - D_T}, T\right), \tag{78}$$

where $\mathcal{C}_Y(K,T)$ indicates that the option with strike price K and maturity T should be considered as having $(Y_t)_{t\geqslant 0}$ as underlying. Contrariwise, its counterpart $\mathcal{C}_U\left(\frac{K-D_T}{F_T-D_T}, T\right)$ should be interpreted as an option on $(U_t)_{t\geqslant 0}$, with the same maturity as the former but, this time, with an appropriately adjusted strike price.

As far as the rough Bergomi model Bayer et al. (2016) is concerned, we recall here its functional form. Given relationship (78), the rough Bergomi model can be then used to describe the dynamics of the process $(U_t)_{t\geqslant 0}$ ($U_0 = 1$). That is, the following set of equations should be used

25 As soon as compatible with the conic finance paradigm.

26 In Appendix 4.A, by way of illustration and to provide further evidence to support the generality of the techniques introduced in this paper, we show how prices, implied liquidity amounts, as well as liquidity deltas calculated using the conic SABR model of Section 4.3.1 can be accurately replicated by means of a properly trained CNN.

27 Recall that, throughout this article, for simplicity and without loss of generality we assume the time value of money to be zero (otherwise, discounted prices could be considered, instead).

$$\begin{cases} U_t = \mathcal{E}\left(\int_0^t \sqrt{v_u}\, dW_u\right) \\ v_u = \sigma_0^2 \cdot \mathcal{E}\left(\eta\sqrt{2H}\int_0^u (u-s)^{H-\frac{1}{2}}\, dZ_s\right) \\ d\langle W_t, Z_t\rangle = \rho dt \end{cases}, \tag{79}$$

with $\sigma > 0$, $\eta > 0$, $H \in \left(0, \frac{1}{2}\right)$ the Hurst exponent, $\rho \in [-1, 1]$ the correlation between the two Brownian motions $(W_t)_{t\geqslant 0}$ and $(Z_t)_{t\geqslant 0}$, and with $\mathcal{E}(\cdot)$ denoting the stochastic exponential. Observe that the second equation in (79) involves a Volterra process, which is linked to the concept of fractional Brownian motion Mandelbrot and van Ness (1968).

For performing path-wise simulations under rough Bergomi dynamics (in a risk-neutral environment), a popular choice is that of discretizing the integral related to the Volterra process within the second equation in (79) by means of the hybrid scheme Bennedsen et al. (2017) (see also McCrickerd and Pakkanen (2018) for additional details), at step j of N, given by

$$\int_0^{\frac{j}{N}} (u-s)^{H-\frac{1}{2}}\, dZ_s \approx \int_{\frac{j-1}{N}}^{\frac{j}{N}} \left(\frac{j}{N} - s\right)^{H-\frac{1}{2}} dZ_s + \sum_{l=2}^{j} \left(\frac{g_l}{N}\right)^{H-\frac{1}{2}} \left(Z_{\frac{j-l+1}{N}} - Z_{\frac{j-1}{N}}\right),$$

where

$$g_l \coloneqq \left(\frac{l^{H+\frac{3}{2}} - (l-1)^{H+\frac{1}{2}}}{H+\frac{3}{2}}\right)^{\frac{1}{H+\frac{1}{2}}}.$$

Bid-ask pricing according to the specifications of the rough Bergomi model can be performed using the conic Monte Carlo approach described in Section 4.3.2. Moving forward, we will refer to the rough Bergomi model under conic settings as to the conic rough Bergomi model, for simplicity and to avoid ambiguities.

We now present some results with the aim of illustrating how CNNs perform in practice. In particular, we consider (European) equity options maturing in one year on a dividend-paying asset and, therefore, we follow the modeling approach (78). As far as the strike range is concerned, we take into account all the available options with moneyness level in-between the 75% and 125% range.[28] The reason for this is given by the fact that, for deeply in-the-money or out-of-the-money options, often only bids or asks are quoted, depending on whether the market trend is that of shorting or longing specific option positions. Therefore, we focus our analysis within the moneyness area where, given the data available, the observable market information is more reliable.

28 In line with the example provided in Section 4.6.1.

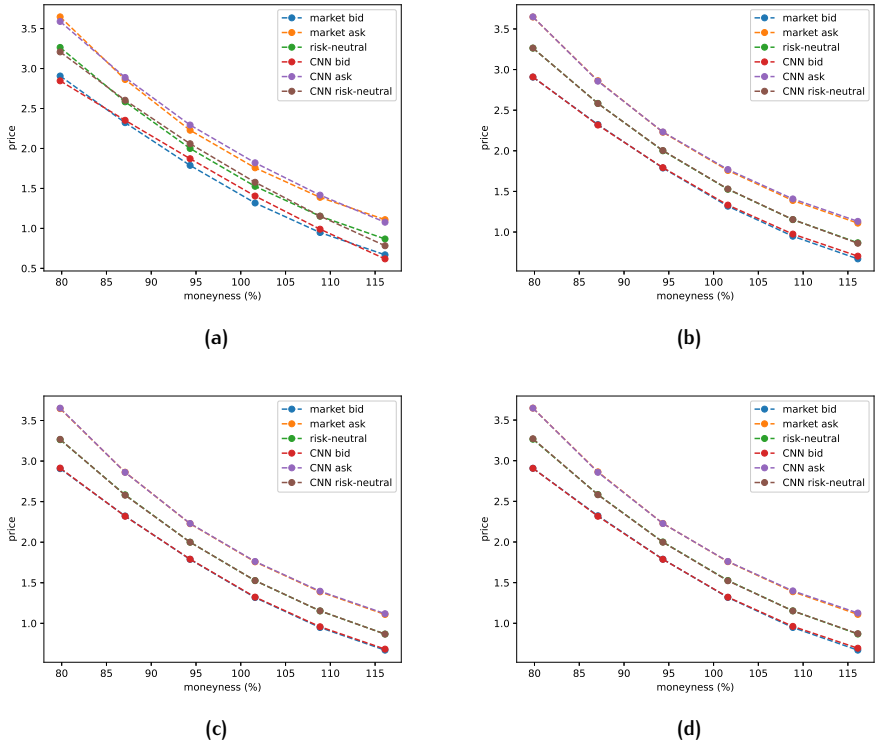


Figure 22: Replication of quoted bid and ask option prices, as well as their risk-neutral counterparts, using the conic rough Bergomi model and its CNN extension. Risk-neutral prices have been calculated by considering $\lambda = 0$. The number of points used in the CNN equals 5 in panel (a), 10 in panel (b), 15 in panel (c), and 20 in panel (d). The number of points should be interpreted per training parameter.

Note, however, that this choice is by no means restrictive, as considering more (or fewer) options would just require feeding the CNN with an extended (shrinked) dataset. Everything else would work exactly the same manner. In order to train the CNN we consider as input ranges those that can be obtained by taking into account the calibrated parameters of the original model to which a buffer of $\pm 100\%$ (relative) is added. For each of the chosen intervals, the CNN is trained by taking into account a different number of points. This in order to show how, when a sufficiently high number of input points in the training dataset is considered, the CNN allows to accurately replicate the quantities of interest, either implied or market-observable.

We start by taking into account the replication of the (calibrated) model prices given the input dataset. This is illustrated in Figures 22a to 22d. In particular, these illustrations show how the CNN, by increasing the number of inputs,

smoothly and accurately replicates the option prices quoted in the market and the respective model prices computed with the (calibrated) conic rough Bergomi model.

In terms of execution, by testing valuation performances using a CNN versus conic Monte Carlo simulation we observe a speed up, for options maturing in one year, of approximately 75 times, which becomes even more impressive for larger maturities (i.e., approximately 150 times in the case of a two-year expiry).

As Figures 22a to 22d illustrate, the conic Bergomi model can be flawlessly replicated by the CNN approach. In particular, the least squares error obtained between model and CNN prices is of the order of 10^{-5}, which is more than satisfactory. However, by increasing the number of points, this amount can be decreased at will.[29]

We now consider another important aspect related to conic pricing, i.e., the notion of *implied liquidity* introduced in Corcuera et al. (2012). To intuitively explain this concept recall that the implied volatility is some sort of forward-looking forecast measure of the expected volatility of a given underlying that can be "implied" from option quotes. That is, it somehow represents and quantifies the market's overall expectation of the variability of the underlying within a given horizon. Similarly, the implied liquidity (also known as *implied distortion* given its definition in mathematical terms, see (74) and (76)) for a given option contract represents, assuming market agents trade options according to the conic finance paradigm Cherny and Madan (2010), the implicit willingness to trade a given contract. This amount is, therefore, essential as far as hedging considerations are concerned and therefore needs to be taken into account wisely. Graphical depictions of how implied liquidity parameters for different strikes can be replicated using the CNN are available in Figures from 23a to 23d.

As Figures 23a to 23d illustrate, the implied liquidity parameters produced by the conic rough Bergomi model can be accurately replicated by their CNN counterparts, in a similar manner it was observed in the case of (bid, ask and risk-neutral) option prices as per Figures 22a to 22d. That is, as soon as enough points are considered in the training set, implied liquidity amounts can be replicated with a very good accuracy. We again stress here the importance of being able, in a market with frictions, to compute liquidity parameters accurately.

29 Of course, up to numerical limitations and boundaries. Observe that some neural network implementations are optimized for single-precision floating-point arithmetic rather than for double-precision floating-point arithmetic. This in order to reduce memory requirements and overhead, as well as improving training performances, given the often colossal size of the training data they are required to deal with. Therefore, in this case the calibration differences obtained are not so far from machine precision and can therefore be considered more than accurate. In any case, enriching the training dataset would allow for potentially even closer model replication.

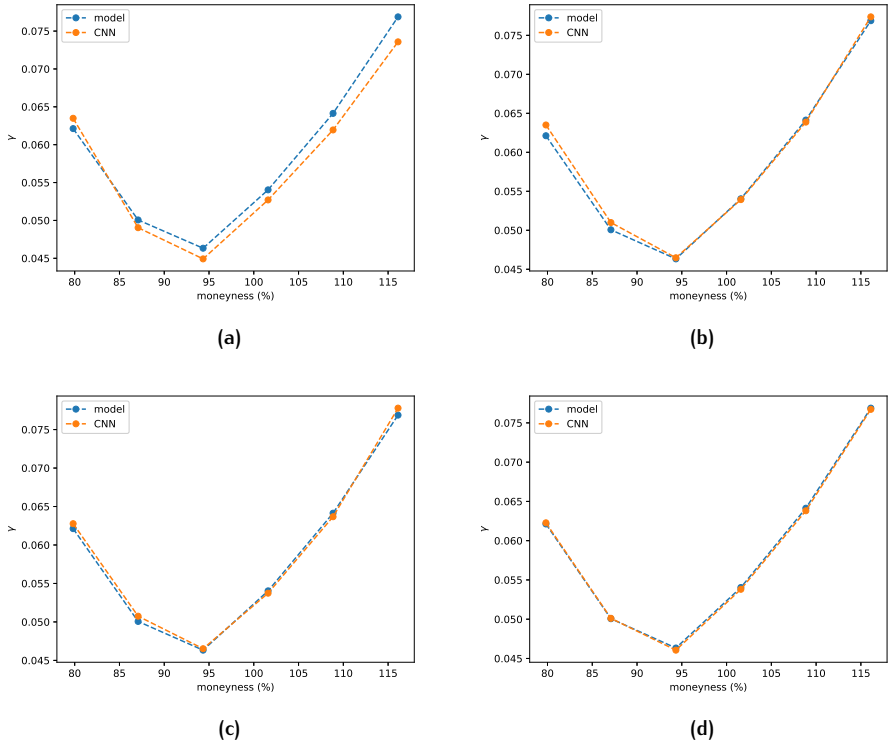


Figure 23: Replication of implied liquidity amounts computed using the conic rough Bergomi model by means of its CNN extension. The number of points used in the CNN equals 5 in panel (a), 10 in panel (b), 15 in panel (c), and 20 in panel (d). The number of points should be interpreted per training parameter.

The reason for this choice is given by the fact that, the higher the implied liquidity (i.e., the lower the liquidity of the market) for a given contract, the higher the costs a market agent would incur should they decide to trade in that specific security. And this independently of the direction of the transaction. Therefore, this can seriously affect hedging considerations and, to further emphasize the importance of this aspect, we illustrate this with a simple elucidating example. Say a financial institution is largely exposed to a given contract on their trading book that is considered particularly risky and, therefore, needs to be risk-managed carefully. For this reason, the financial institution might decide to liquidate (part) of it. However, if it happens that the implied liquidity of the contract raises considerably, this might imply that the liquidation process cannot take place, unless the financial institution is willing to report undesirable losses due to the large bid-ask spreads. Therefore, instead of liquidating the contract, it might then be chosen to (for instance delta) hedge it somehow, by entering

a position in the underlying itself, or by transacting an alternative highly correlated security appropriately. Additionally, the implied liquidity can also affect business decisions. Entering a given contract might imply having, for instance, a large liquidity exposure. This might make a financial institution opt for positions in more liquid contracts due to the fact that, in times of financial distress, the first choice might cause high liquidation costs (i.e., losses). A practical example of these considerations is given by the credit default swap (CDS) market during the COVID-19 pandemics, which caused the widening of CDS spreads for a wide range of reference entities; see, e.g., Hasan et al. (2023). As far as CDSs are concerned, under standard market assumptions implied liquidity and corresponding risk-neutral default probabilities can be exactly calculated following the procedure delineated in Michielon et al. (2021a). Therefore, accurately computing and risk-managing liquidity sensitivities is also a highly desirable model feature. In Section 4.6.3 we will show how a CNN is also able to accurately calculate sensitivities, of which liquidity deltas represent a special instance.

4.6.3 Sensitivities: Learning (the) Greek(s)

We now discuss how to calculate Greeks using CNNs. That is, the approach provided in this section will allow to calculate Greeks of bid and ask prices by allowing the CNN to be trained, at the same time, to "learn the Greeks" as well. Also in this case we highlight that this methodology is general and, again, model agnostic. In particular, as the research proposed in this article relates to the conic finance paradigm Cherny and Madan (2010), we will focus our attention to the computations of liquidity Greeks (i.e., liquidity deltas in this case).[30] We want however to bring to the forefront that this methodology not only can be extended to higher order liquidity Greeks, but also to any (potentially high-order as well) sensitivity related to any of the parameters that play a role in the modeling framework. This can be easily achieved as soon as the training framework is rich enough. This means that the considerations and the approach described here is by no means specific, and it easily allows to be generalized. However, this would make the notation glutted and saturated, which we aim to avoid for the sake of clear and concise exposition.

In Sections 4.6.1 and 4.6.2 we illustrated how using CNNs requires approximating, by means in this case of a multilayer perceptron, a function such that $(\theta, \lambda; K)$ is mapped to $(\mathrm{bid}_\lambda(\theta; K), \mathrm{ask}_\lambda(\theta; K))$. However, one might want to extend the aforementioned considerations to calculate sensitivities as well. This will be now illustated. Note that here, for the sake of exposition, we will show

30 We make prominent that we have chosen to take into account liquidity delta sensitivities only because, to the best of our knowledge, research developments in liquidity management using distorted expectations are still at an embryonic phase. Therefore, from a practical perspective being able to account for first order sensitivities in this respect should suffice.

how to calculate Greeks of the first order only. However, the methodology can be trivially extended to higher order sensitivities as well.

Having said that, the steps are the following. Once set

$$\nabla \mathrm{bid}_\lambda(\theta, K) \coloneqq \left(\frac{\partial \mathrm{bid}_\lambda(\theta, K)}{\partial \theta_1}, \dots, \frac{\partial \mathrm{bid}_\lambda(\theta, K)}{\partial \theta_N}, \frac{\partial \mathrm{bid}_\lambda(\theta, K)}{\partial \lambda} \right)$$

and, similarly,

$$\nabla \mathrm{ask}_\lambda(\theta, K) \coloneqq \left(\frac{\partial \mathrm{ask}_\lambda(\theta, K)}{\partial \theta_1}, \dots, \frac{\partial \mathrm{ask}_\lambda(\theta, K)}{\partial \theta_N}, \frac{\partial \mathrm{ask}_\lambda(\theta, K)}{\partial \lambda} \right),$$

we consider the map

$$(\theta, \lambda, K) \mapsto (\mathrm{bid}_\lambda(\theta, K), \mathrm{ask}_\lambda(\theta, K) | \nabla \mathrm{bid}_\lambda(\theta, K) | \nabla \mathrm{ask}_\lambda(\theta, K)), \tag{80}$$

where in (80) the symbol "|" denotes the juxtaposition of vectors. By performing the training of the CNN as illustrated in Section 4.6.2 (in this case the derivatives above can be computed, e.g., by means of finite differences) one can then obtain a CNN that also allows to compute sensitivities, simultaneously.

In the case of the (conic) rough Bergomi model we provide some illustrations in Figures 24a to 24d. As the graphs illustrate, in the case of delta computations a lower number of points seems enough to produce a satisfactory accuracy in contrast to the case of option prices (compare with Figures 22a to 22d) and implied liquidity parameters (compare with Figures 23a to 23d). The reason for this is given by the fact that, contrary to the former two cases (i.e., market values and implied distortions), here we are calculating "differences in market values". Therefore, if the valuation of the CNN slightly overestimate (underestimates) a specific quantity, it is in any case natural to expect that these misalignments tend to somehow compensate each other, resulting in acceptable approximation errors for the deltas. This is, as expected, what it seems to emerge from Figures 24a to 24d. From a computational perspective, the performances to calculate Greeks for the examples considered here provide speed-ups of the order from 150 to 300 times (for maturities between one and two years) compared to forward finite differences.

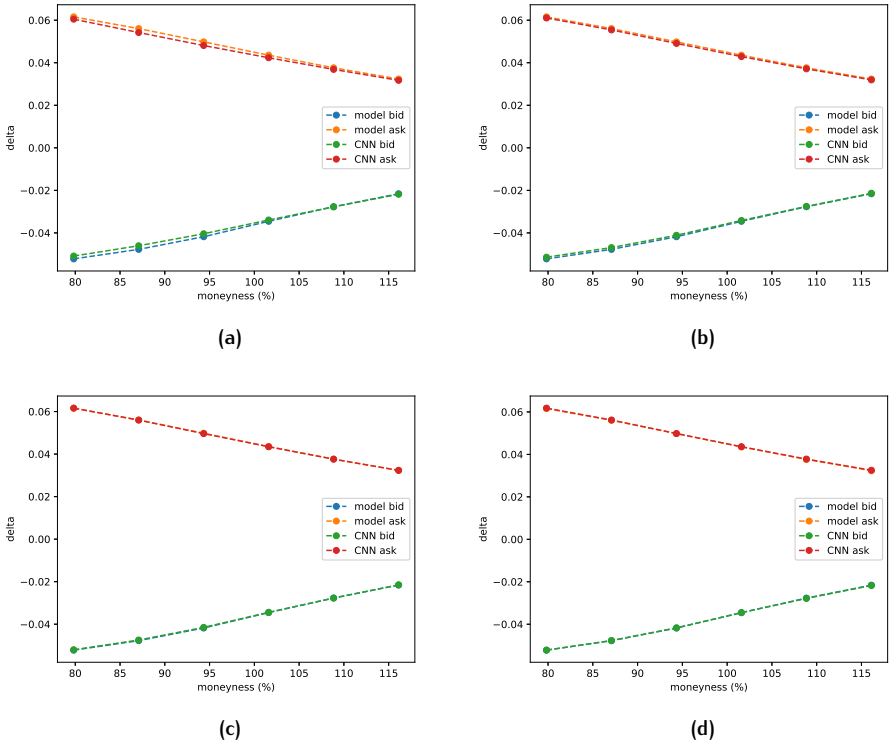


Figure 24: Replication of implied liquidity delta amounts computed using the rough Bergomi model by means of its CNN extension. The number of points used in the CNN equals 5 in panel (a), 10 in panel (b), 15 in panel (c), and 20 in panel (d). The number of points should be interpreted per training parameter. Liquidity delta amounts have been multiplied by 1%, in line with what is often done in risk monitoring, e.g., for vega calculations.

However, for analysis and comparison purposes, we also want to further comment on the methodology. In fact, one might also observe that the majority of the common neuron activation functions introduced in neural networks to account for nonlinearities between inputs and outputs are infinitely differentiable.[31] Therefore, it might be interesting to further justifying the modeling choice proposed in this section (i.e., "learning the Greeks") with that of computing sensitivities by means of finite difference approximations of a CNN.

31 To be precise, note that some activation functions are infinitely differentiable almost everywhere, such as the *rectified linear unit* (also known as *ReLU* activation) or the *softsign* functions (amongst others). That is, these examples of activation functions, which are very popular in practical applications, are not differentiable at zero. See Table 3 for an overview of (some) common choices for neuron activation functions.

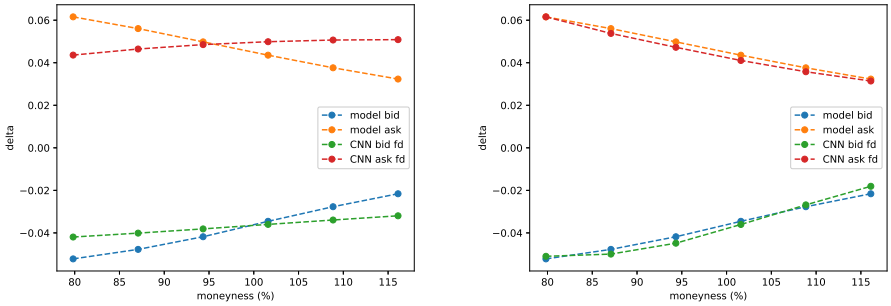


Figure 25: Attempt to replicate liquidity delta amounts computed using the rough Bergomi model by means of its CNN extension where finite differences are used. The number of points used in the CNN equals 5 in panel (a) and 10 in panel (b). In both panels (a) and (b) the label "model bid (ask)" refers to the liquidity delta amounts computed using the conic Bergomi model, while the label "CNN bid (ask) fd" to their counterparts calculated using the CNN where sensitivities are computed by means of (forward) finite differences. The number of points should be interpreted per training parameter. Liquidity delta amounts have been multiplied by 1%, in line with what is often done in risk monitoring, e.g., for vega calculations.

For this reason, therefore, we compare the two approaches in order to provide the reader with a qualitative analysis that, depending on their needs and on the speed/accuracy trade-off required, might tip the scales in favor of one approach or the other. That is, we observe that one could compute liquidity deltas by means of the CNN using a simple finite difference approach. For this reason, to compare the two possible approaches and draw appropriate conclusions, we consider in this case the liquidity delta computed this time using the CNN where a first-order forward finite difference scheme is applied; see Figures 25a and 25b for an illustration.

As Figures 25a and 25b show, applying a finite difference scheme to the neural network directly does not provide results which are as accurate as those generated using the methodology provided in this section; c.f. Figures 24a and 24b. That is, it is clear how performing sensitivity calculations by means of finite differences requires more points in the training set to obtain the desired accuracy. The underlying reason for this phenomenon is given by the fact that if two functions are close, in some sense, their derivatives are not necessarily so. This leads

to the differences observed.[32] Therefore, the comparison just performed clearly provides supporting evidence to justify the approach followed.

An additional benefit of following this approach is given by the fact that computing Greeks by a finite difference approach requires double[33] the time compared to that of pricing an option, in contrast to a marginal increase in the other case which requires, on the other hand, a sensitivity-augmented data set. Notwithstanding, the augmented dateset needs to be generated once only.

4.6.4 A warning on the choice of the activation function

The structure of a neural network depends on the choice of the neuron activation function(s) it uses to introduce nonlinearities in the learning process. Otherwise, without those, a neural network would be just a combination of affine functions, and therefore would remain such. The list of possible activation functions to be used within a neural network architecture is extensive. We report (some of) the most well-known possible candidates in Table 3. For studies concerning the comparison of activation functions in (deep) neural networks we refer the reader, e.g., to Szandała (2021).

Activation function	**Functional form**
Exponential	e^x
Hyperbolic tangent	$\tanh(x)$
GeLU (Gaussian error linear unit)	$x \cdot \Phi(x)$
ReLU (rectified linear unit)	$\max(x, 0)$
Sigmoid	$\frac{1}{1+e^{-x}}$
Softplus	$\ln(e^x + 1)$
Softsign	$\frac{x}{\lvert x\rvert+1}$
Swish	$\frac{x}{1+e^{-x}}$

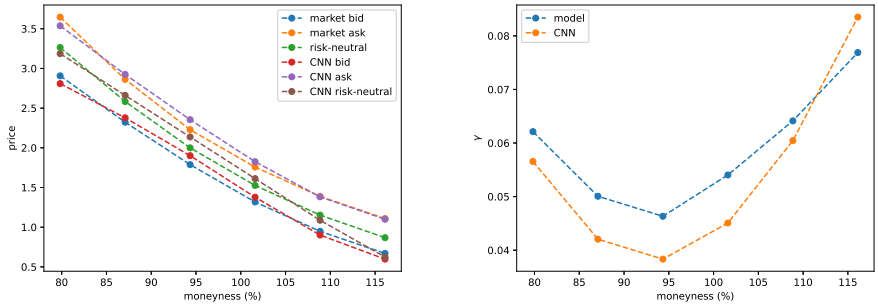
Table 3: A (non-exhaustive) list the most commonly used activation functions in neural networks, ordered lexicographically.

32 This can be explained by means of a simple illustrative example. Consider, for instance, the constant function $f(x) = 0$ and the periodic function $g(x) = \epsilon \cdot \sin(k \cdot x)$ for $\epsilon > 0$ "small" and $k \in \mathbb{Z}$. In this case we obtain that, for every x, $f(x)$ and $g(x)$ will never deviate for more than ϵ, so they will always be "close enough". However, while the derivative of $f(\cdot)$ is constant (i.e., zero), that of $g(\cdot)$ can potentially reach very high (or low) values, depending on the choice of k. Therefore, despite this example is somewhat "pathological", it provides the background for the methodological choice proposed in the current section.

33 If a forward or backward scheme is used. Otherwise, if additional precision is required and, therefore, if a central difference scheme is used, the extra computational cost triples. Of course, in the case of higher-order derivatives, the extra calculations required (model speed performance) further increase (deteriorate).

In particular, in common applications one of the most popular choices as far as the neuron activation function is concerned is the ReLU activation. This activation function, introduced for the first time in Fukushima (1975), is nowadays heavily used. This not only for its simplicity, which results in faster training of (deep) neural networks, but also because it (empirically) helps alleviating the vanishing gradient problem, which can occur in deep neural networks during backpropagation; see for instance Glorot et al. (2011).[34]

However, in the context of option pricing, models depend on a series of input parameters which are, often, positive. For instance, in the case of the rough Bergomi model (79) all the parameters except the correlation between the two model Brownian motions are positive. Furthermore, option premia are positive quantities and, besides, Choquet integrals are highly non-linear functionals. Therefore, in order to obtain good model replication, the neural network considered should be "significantly nonlinear" in the input parameters. Given however that model and market parameters are, in this situation, mainly non-negative, one would expect the ReLU function not to perform as well as in other applications. As it intuitively would not be able to introduce enough nonlinearities in the model. In order to test this hypothesis, we therefore train the CNN using the ReLU function as neuron activator, and plot the calibrated model prices and resulting implied distortions in Figures 26a and 26b, respectively, for comparison purposes.



(a) (b)

Figure 26: Replication of model prices, panel (a), and implied liquidity amounts, panel (b), by means of a CNN with ReLU activation function. In both cases 20 inputs points have been taken into account. The number of points should be interpreted per training parameter.

34 The vanishing gradient problem refers to the complication given by gradients becoming smaller and smaller (i.e., the gradient vector with respect to the weights of the neural network has a small magnitude with respect to a certain norm) as they propagate backwards through the network, which can potentially pose a significant hurdle heavily affecting the learning process.

By comparing Figures 26a and 26b with Figures 22d and 23d, respectively, one immediately observes that the performances of the training procedure, in the case the ReLU activation function is used, deteriorate. That is, model prices (and therefore implied distortions as a consequence) are largely off. This empirically confirms our initial conjecture about the usage of the ReLU activation function for bid-ask valuations in a conic framework.

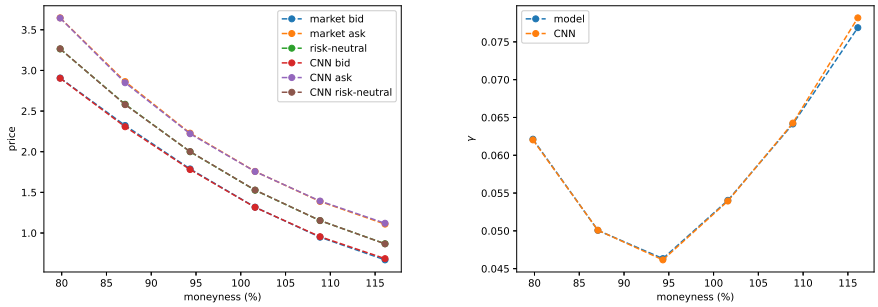
4.6.5 Deep CNNs

A deep neural network is nothing but a type of artificial neural network architecture that has been developed to solve complex tasks such as image and speech recognition, natural language processing, and more. However, at basic level, deep neural networks differ from standard neural networks by means of the number of hidden layers (see Section 4.5.1). The main idea for including more hidden layers in the model is that of allowing the neuron interconnections to replicate even more non-linear relationships between inputs and outputs. This with the aim of improving the performances of the neural network to solve both classification and regression problems. In this section we will provide some considerations concerning using deep neural networks to solve the regression problem of pricing in a conic framework.

From a practical angle taking into account a deep neural network requires adding (at least) one additional hidden layer to the architecture described in Section 4.5. In particular, as in the article at hand we are dealing with multilayer perceptrons (see Section 4.5.1), we will consider here a deep multilayer perceptron. That is, here we take into account a CNN with two hidden layers. From an operational perspective, both the calibration and the calculation of implied liquidity amounts can be performed following the procedure described in Section 4.6.2[35] and, in particular, all the steps available in Section 4.6.2 can be still followed. Figures 27a and 27b provide an illustration of the calibration accuracy and of the implied distortion amounts resulting from employing a deep CNN.

By comparing Figures 27a and 27b with their counterparts 22d and 23d, respectively, we observe that, qualitatively, both the CNN and its deep counterpart perform very well. In particular, also in the deep CNN case we obtain a least square error with an order of magnitude equal to 10^{-5}. As far as the implied liquidity parameter is concerned, we observe that the deep CNN produces slightly more accurate results for in-the-money and at-the-money options, and slightly worse results in the case of out-of-the-money options than its counterpart. However, we can say that results are equally accurate, up to minor and negligible differences. Therefore, for comparison purposes it comes natural to ask whether there would be circumstances where one approach would be more recommended than the other or the other way round.

35 As the procedure proposed therein is independent of the neural network architecture chose.



(a) (b)

Figure 27: Replication of model prices, panel (a), and implied liquidity amounts, panel (b), by means of a deep CNN. In both cases 20 inputs points have been taken into account. The number of points should be interpreted per training parameter.

The fact of considering a deep CNN has a straightforward advantage compared to the case of a CNN, i.e., more flexibility. That is, by having multiple hidden layers, a deep approach allows for more potential interactions (i.e., connections) between neurons belonging to different layers. Furthermore, an architecture including more layers also naturally allows to model neuron activators using different activation functions per layer. This makes it possible to have potentially infinitely many possibilities for modeling nonlinearities between inputs and outputs. However, this comes at cost. First of all, more degrees of freedom naturally require larger datasets for the training process. This is due to the fact that too much flexibility in the modeling framework also implies that the error minimization algorithm requires more steps to converge because of the higher presence of local minima. And therefore the necessity of more data points for the backpropagation approach to find its way towards the convergence point. Furthermore, despite in many cases relatively marginal due to GPU parallelization (unless an extremely high number of hidden layers is considered), computing output values using deep neural networks requires more time than in the standard case.[36] This is obvious given that, for each extra layer, an affine and a nonlinear functions need to be valued on a grid of points. This, therefore, can in principle slow down both the training procedure as well as the pricing operations. Therefore, whether one approach is better than the other depends on the circumstances. Nonetheless, we believe that, given the tests performed in the article at hand, in many cases the use of a single-layer CNN suffices. This is because CNNs have been shown to be able to reprice the market accurately, as well as to replicate implied liquidity amounts and sensitivities faithfully. However,

36 Despite this is partially compensated by the fact that, as a rule of thumb, when a neural network is extended with extra hidden layers, then often the number of nodes per layer is chosen to be smaller compared to the case of a single hidden layer.

the flexibility of deep CNN should not be overlooked, which can for sure come in handy in large scale computations.

4.7 NEURAL-NETWORK-DRIVEN LSV CALIBRATION IN A TWO-PRICE ECONOMY

LSV models have become increasingly popular in financial mathematics as a way to capture the complex dynamics of financial markets. LSV modeling provides an hybrid framework that allows to embed the advantages of both local and stochastic volatility, as the name obviously suggests.

Local volatility modeling makes it possible to replicate vanilla option prices by assuming that the volatility of the underlying asset is a deterministic function of both strike and time to expiry; see Dupire (1994). However, practical evidence seems to suggest that local volatility models produce forward skews which are mostly flat. This, therefore, might result in price misestimation for some financial securities. On the other hand, stochastic volatility models, see, e.g., Heston (1993), form a class of financial models attempting to capture the erratic nature of the volatility in financial markets. This is achieved by supposing that the volatility is a random process itself, varying over time. Unlike local volatility models, which assume that volatility is a deterministic function of the underlying asset price, the possibility of volatility evolving randomly over time can better capture the complex dynamics of financial markets. However, this does not guarantee that pure stochastic volatility models are always able to reproduce the term structure of the volatility smile accurately enough. It is worth pointing out that, despite many stochastic volatility models allow to closely replicate quoted vanilla prices in an almost indistinguishable manner, they might result in very different prices should exotic derivatives be taken into account (see Schoutens et al. (2003)). This can therefore be a non-negligible source of model risk. By combining the two aforementioned approaches at the same time, LSV models have been developed to better capture the dynamics of option prices, which are particularly sensitive to changes in volatility. In particular, LSV models allow to better capture forward smile risk compared to their local volatility counterparts; see Mazzon and Pascucci (2017). LSV models have been successfully used in a wide range of applications, including option pricing, hedging, and risk management. They have been particularly useful in pricing exotic options, which are often difficult to be valued accurately using more simplistic models. For this reason, LSV modeling has found applications in different asset classes such as equities and foreign exchange, to name but two. Therefore, the rationale behind the increasing usage of this modeling approach is given by its ability to better capture the complex dynamics of financial markets and to provide more accurate pricing and risk management tools. By allowing for extra flexibility in modeling

volatility, LSV models can better capture the true risk of financial instruments and provide more reliable pricing and Greeks.

The popularity of LSV models in the financial industry, as well as the continuously-growing interest of researchers in applications of machine learning in quantitative analysis, have driven research aiming to improve volatility modeling by means of neural networks; see Section 4.4 for relevant references. However, to the best of our knowledge, at the time of writing, no literature investigating the applications of neural networks within LSV modeling in markets with frictions in the sense of Cherny and Madan (2010) seems to be available. Therefore, this section aims to investigate this new research direction. In particular, we consider the approach of Cuchiero et al. (2020), where the SABR model, embedded with a local volatility component, is taken into account. Therein, the local volatility factor of the model is approximated by means of a (combination of) neural network(s). In this section we show how it is possible to extend such a modeling template to calibrate to bid and ask market quotes by means of conic Monte Carlo (see Section 4.3.2). By comparing the results obtained with respect to the conic SABR model introduced in Section 4.3.1 we show how adding a local volatility term "parametrized" using neural network remarkably improves the accuracy of the model.[37]

4.7.1 LSV essentials: A primer

We now outline the essential concepts and notions related to LSV modeling. In particular, we provide the basic notations, definitions and equations that are relevant to introduce the framework of Cuchiero et al. (2020). For further details concerning LSV models and for an in-depth treatment of the topic refer, amongst others, to Guyon and Henry-Labordère (2011).

To briefly introduce LSV modeling we start from the local volatility approach of Dupire (1994). In particular, for a given underlying price process $(Y_t)_{t\geqslant 0}$, Dupire (1994) proposes to model its evolution through time by means of

$$dY_t = \sigma(Y_t, t)Y_t\,dW_t, \tag{81}$$

where $(W_t)_{t\geqslant 0}$ is a one-dimensional Brownian motion, and with $\sigma(\cdot,\cdot)$ denoting a time-dependent deterministic function contingent on the underlying spot price. It can be shown, see Dupire (1994) (but also (Gatheral, 2006, Ch. 1), amongst many others), that in order to exactly match the quoted European call

37 We recall that, in this paper, for simplicity and without loss of generality, the risk-free rate is assumed to be zero. Also, underlying prices will be always interpreted as cleansed from dividends, and all the calculations will be intended under a given pricing (i.e., risk-neutral) measure Q.

option prices on the considered underlying, then $\sigma(\cdot, \cdot)$ needs to satisfy the so-called *Dupire's formula* given by

$$\sigma^{\text{Dupire}}(K, T) = \sqrt{\frac{\frac{\partial \mathcal{C}(K,T)}{\partial T}}{\frac{1}{2}K^2 \frac{\partial^2 \mathcal{C}(K,T)}{\partial K^2}}}, \tag{82}$$

where in (82) $\mathcal{C}(K, T)$ denotes the call price of an European option with strike K expiring at T.[38]

We also recall that a stochastic volatility model, given the same underlying asset $(Y_t)_{t \geqslant 0}$, has the form

$$dY_t = \alpha_t Y_t \, dW_t. \tag{83}$$

However, the volatility component in (83), i.e., $(\alpha_t)_{t \geqslant 0}$ is here given by a stochastic process.

Given the local and stochastic volatility specifications as per (81) and (83), respectively, the ultimate goal behind LSV models is that of providing an hybrid template aiming to get "the best of both worlds". That is, an LSV model for $(Y_t)_{t \geqslant 0}$ can be then defined by means of

$$dY_t = \alpha_t L(Y_t, t) Y_t \, dW_t, \tag{84}$$

with $(\alpha_t)_{t \geqslant 0}$ a stochastic process for the volatility as in (83), and with with the factor $L(\cdot, \cdot)$, often named *leverage function*, aiming to represent the local volatility component of the model. It can be shown, see Guyon and Henry-Labordère (2011), that in order to match quoted (European) call option prices, the leverage function needs to satisfy the relationship

$$L(K, T) = \frac{\sigma^{\text{Dupire}}(K, T)}{\sqrt{\mathbb{E}^{\mathbb{Q}}(\alpha_T^2 | Y_T = K)}}. \tag{85}$$

4.7.2 Leverage function: A neural network approximation

Despite (85) provides an expression for the leverage function under LSV dynamics as per (84), in practice dealing with such an expression is not a trivial task for a number of reasons. For instance, inside (85) the Dupire's formula (82) appears. Unless one is able to compute the derivatives in (82) exactly, as it would be the case should algorithmic differentiation techniques been used (see Henrard (2017) for an overview), then the choice of assuming a parametric form for (82) is often the one followed in practice as proposed, e.g., in Carmona and Nadtochiy (2008). The reason why this is, in many cases, the preferred choice,

38 Observe that (82) can also be expressed in terms of implied volatilities rather than option prices. This is a convenient feature as, often, implied volatilities are used as quoting conventions. This expression can be found, for instance, in (Gatheral, 2006, Ch. 1).

is that the Dupire's formula is very sensitive to the values of the second derivative at the denominator in (82) which, due to numerical approximations, can cause severe pricing inaccuracies. Also, adopting an a-priori fully-parametric approach somehow implies additional assumptions and model choices in the framework. Furthermore, we also highlight that the denominator in (85) cannot, in general, be computed analytically. Therefore, this introduces an additional source of inaccuracy, leading to further undesirable numerical instabilities.

The approach provided in Cuchiero et al. (2020) introduces a novel idea as far as LSV modeling is concerned. That is, considering M maturities for vanilla call options on a given underlying $(Y_t)_{t\geqslant 0}$, denoted as $T_1 < \ldots < T_M$ (with the convention that $T_0 \coloneqq 0$), in Cuchiero et al. (2020) it is assumed that the leverage function (85) is of the form

$$L(K, T, \theta) \coloneqq 1 + \sum_{i=1}^{M} NN_i(K, \theta_i) \cdot \mathbb{1}_{[T_{i-1}, T_i)}(T) \tag{86}$$

where, for every $i \in \{1, \ldots, M\}$, $NN_i(\cdot, \cdot)$ denotes a feed-forward neural network with one hidden layer and a certain number of nodes, dependent on the set of (to-be-calibrated) weights $\theta_i \in \Theta_i \subseteq \mathbb{R}^{l_i}$ (for some $l_i > 0$); θ denotes the collection of all the aforementioned weights, while $\mathbb{1}_{(\cdot)}$ the indicator function. Considering the neural networks $NN_1(\cdot, \cdot), \ldots, NN_M(\cdot, \cdot)$ with the aim of modeling the leverage function allows to introduce a high degree of flexibility in the model. In fact, by appropriate learning, due to the natural resilience properties of neural networks provided by the wide range of neuron activation functions and choices in the number of nodes, a framework as the one considered here enables to model very arbitrary shapes for the local volatility function. This, in principle, provides a much more general and adaptable setting than just considering a parametric family dependent on a handful of parameters.

4.7.3 Calibration

Once the functional form for the leverage function has been set as done by means of (86), one needs to set up a calibration routine for both the leverage function and the parameters belonging to the stochastic volatility component of (84), i.e., those related to the process $(\alpha_t)_{t\geqslant 0}$. In particular, in Cuchiero et al. (2020) it is proposed to perform the calibration in two steps. That is, after having calibrated the "stand-alone" stochastic volatility process (83), then the calibration of the leverage function (85) can take place with the aim of compensating for the calibration inaccuracies that the pure stochastic volatility approach produces (as much as possible). Here, therefore, we assume that the parameters of the process $(\alpha_t)_{t\geqslant 0}$ are known (i.e., they have already been calibrated) and we will focus on the calibration of the local volatility component as outlined in Cuchiero et al. (2020).

Considering the framework provided in Section 4.7.2 given by M option maturities, we assume that, for the i^{th} maturity T_i, J_i options are quoted. We denote their strikes and prices by $K_{i,j}$ and $\mathcal{C}_{i,j}$, respectively ($j \in \{1, \dots, J_i\}$).[39] The corresponding model prices computed with the LSV approach described in Section 4.7.2 will be denoted by $\mathcal{C}^{LSV}_{i,j}(\theta_i)$. Assuming the calibration is performed in a least-squares sense, then it is necessary to perform the weighted minimization problem given by

$$\min_{\theta_i \in \Theta_i} \sum_{j=1}^{J_i} w_{i,j} (\mathcal{C}^{LSV}_{i,j}(\theta_i) - \mathcal{C}_{i,j})^2, \tag{87}$$

where the factors $w_{i,j}$ denote some positive weights. The optimization task is then performed iteratively; that is, one should start by minimizing (87) for the first maturity before attempting to solve (87) for the second maturity (the second step, therefore, is indirectly based on the calibrated parameters for the first maturity that have been calculated at the former step). And so forth, inductively.

We now consider maturity T_i as fixed[40] and observe that, given that the model price $\mathcal{C}^{LSV}_{i,j}(\theta_i)$ actually equals $\mathbb{E}^{\mathbb{Q}}((Y_{T_i}(\theta_i) - K_{i,j})^+)$, one can then describe the discrepancy between model and market prices for the j^{th} strike (i.e., the calibration error) by means of the random variable

$$\epsilon_{i,j}(\theta_i)(\omega) := (Y_{T_i}(\theta_i)(\omega) - K_{i,j})^+ - \mathcal{C}_{i,j},$$

from which it follows that

$$\mathcal{C}^{LSV}_{i,j}(\theta_i) - \mathcal{C}_{i,j} = \mathbb{E}^{\mathbb{Q}}(\epsilon_{i,j}(\theta_i)).$$

This means that (87), in practice, translates into minimizing

$$\min_{\theta_i \in \Theta_i} \sum_{j=1}^{J_i} w_{i,j} (\mathbb{E}^{\mathbb{Q}}(\epsilon_{i,j}(\theta_i)))^2. \tag{88}$$

In Cuchiero et al. (2020) it is proposed, for reducing the number of paths required for calibration and pricing purposes, to use a control variate approach. That is, on top of considering the random variable $\epsilon_{i,j}(\theta_i)$, we can then consider a second (θ_i-dependent) random variable $Z_{i,j}(\theta_i)$ and define the control variate counterpart of $\epsilon_{i,j}(\theta_i)$, i.e., $\epsilon^{CV}_{i,j}(\theta_i)$, by means of the relationship

$$\epsilon^{CV}_{i,j}(\theta_i) := \epsilon_{i,j}(\theta_i) + \alpha_{i,j}(Z_{i,j}(\theta_i) - \mathbb{E}^{\mathbb{Q}}(Z_{i,j}(\theta_i))), \tag{89}$$

39 From here onwards, for simplicity, the index i will be always assumed to range in $\{1, \dots, M\}$, while the index $j(= j(i))$ in $\{1, \dots, J_i\}$, unless otherwise stated.

40 Here, we make it explicit that the all the values we are dealing with are dependent on the set of parameters θ_i.

which has zero mean by construction and variance[41], after having chosen $\alpha_{i,j}$ equal to $\frac{-\mathrm{Cov}(\epsilon_{i,j}(\theta_i), Z_{i,j}(\theta_i))}{\mathrm{Var}(Z_{i,j}(\theta_i))}$, given by

$$\mathrm{Var}(\epsilon^{CV}_{i,j}(\theta_i)) = (1 - \mathrm{Corr}^2(\epsilon_{i,j}(\theta_i)), Z_{i,j}(\theta_i)) \cdot \mathrm{Var}(\epsilon_{i,j}(\theta_i)) \leqslant \mathrm{Var}(\epsilon_{i,j}(\theta_i)). \quad (90)$$

However, as in general it is not possible to compute the numerator and the denominator of (90) analytically, one would need to use their sample counterparts. The choice of the random variables $Z_{i,j}(\theta_i)$'s, as well as that of the weights $w_{i,j}$'s, will be made explicit in Section 4.7.4. The estimation of the parameter θ_i can be performed by means of gradient descent; see Cuchiero et al. (2020).

4.7.4 Application

In this section we provide an applicative example of how to calibrate an LSV pricing model to European option prices under risk-neutral settings. The results available here will then be the foundations for calibrating to bid and ask market prices, by means of conic Monte Carlo (see Section 4.3.2). In particular, we employ the neural network approximation approach for the leverage function outlined in Sections 4.7.2 and 4.7.3. To start with, following Cuchiero et al. (2020), as LSV model we chose the LSV variation of the SABR model (57) given by

$$\begin{cases} dY_t = \sigma_t Y_t L(t, Y_t) dW_t \\ d\sigma_t = \alpha \sigma_t dZ_t \\ d\langle W_t, Z_t \rangle = \rho dt \end{cases}, \quad (91)$$

where the notation used in (91) has been already outlined in Sections 4.3.1 and 4.7.2. Choosing SABR specifications as done in Cuchiero et al. (2020) would also allow to assess the advantages of considering the framework taken into account here compared to the pure Conic SABR model (see Section 4.3.1). Note that, in this example, we take into account SABR specifications consistent with lognormal volatilities, i.e., the parameter β is exogenously set equal to one. On a side note, in order to calibrate the leverage function, we will work with log-coordinates (see Cuchiero et al. (2020)).

For LSV calibration purposes, Cuchiero et al. (2020) propose to split the calibration procedure into two steps. That is, in the first part of the calibration, the goal is that of calibrating the SABR parameters only. This means that, in this case, we set $L(\cdot, \cdot) \equiv 1$, and therefore consider the model as a pure stochastic volatility one. In particular, given the SABR specifications (91), we aim here to find a set of SABR parameters that, as good as possible, fit the "observable" risk-neutral prices for all the maturities available in the dataset. The rationale behind

41 Observe that, to follow this approach, the newly introduced variable $Z_{i,j}(\theta_i)$ needs to be square integrable. Also, we will denote with $\mathrm{Var}(\cdot)$ the variance of a given random variable, while with $\mathrm{Corr}(\cdot, \cdot)$ and $\mathrm{Cov}(\cdot, \cdot)$ the correlation and covariance between two stochastic quantities.

this initial step is that, once the SABR parameters have been calibrated, then the local volatility component of the model will be fine-tuned at a later moment. This to compensate for the miscalibrations caused by the fact that the stand-alone SABR model (57) cannot perfectly fit, at the same time, all the options quoted for all the possible strikes and maturities. However, from a practical perspective, we propose here not to use a simulation-based approach to calibrate the SABR parameters as done in Cuchiero et al. (2020) but, rather, to use the (approximated) analytical SABR volatilities as per (58). The underlying reason for this choice is twofold, i.e., using analytical formulae for computing option prices under the SABR dynamics (57) is not only more performing than using a Monte Carlo approach, but it also avoids introducing unnecessary simulation noise within the calibration procedure.

In symbols, following the conventions introduced in Section 4.7.3, assuming M maturities $T_1 < \ldots < T_M$ and J_i options quoted per maturity ($i \in \{1, \ldots, M\}$) with strike prices $K_{i,j}$ ($j \in \{1, \ldots, J_i\}$), then the calibration task for the pure SABR component of the LSV model (91) (in implied volatility terms) boils down to the minimization problem

$$\min_{\sigma_0, \alpha, \rho} \sum_{i=1}^{M} \sum_{j=1}^{J_i} w_{i,j} \left(\sigma_j^{\text{SABR}}(K_{i,j}, T_i) - \sigma^{\text{market}}(K_{i,j}, T_i) \right)^2, \tag{92}$$

where in (92) $\sigma^{\text{market}}(K_{i,j}, T_i)$ denotes the market implied volatility corresponding to maturity T_i and to strike $K_{i,j}$, while the weights $w_{i,j}$'s are defined by means of

$$w_{i,j} := 1 - \frac{|K_{i,j} - K_i^{\text{ATM}}|}{K_i^{\text{ATM}}}. \tag{93}$$

In (93), K_i^{ATM} represents the at-the-money point for the i^{th} maturity. That is, defining the weights as per (93) aims to give less and less importance to options which are far from the at-the-money point.[42] This choice is justified by the fact that, solely with the SABR parameters, it is not possible to have a satisfactory calibration for all maturities and strikes simultaneously as already highlighted above. Therefore, here we attempt to fit the volatility surface backbone the best way possible given the set of SABR parameters available. In order to achieve this we give less and less importance to the points that departure from the at-the-money levels related to their respective maturities. We then attempt to fit, in a least-square sense, those implied volatilities the best way possible. In contrast, in Cuchiero et al. (2020) it is proposed to calibrate the SABR parameters to the first maturity only. Despite the choice of how to calibrate the SABR parameters remains in any case arbitrary (and considering that the local volatility component of the model should, in principle, be able to compensate for the initial

42 This is sufficient, in practice, given the dataset that we consider, as we take into account strikes that are usually not larger than twice the at-the-money point.

calibration errors), by experimentation it results that attempting to fit the SABR parameters to somehow reproduce the volatility backbone, whenever possible, seems to outperform the choice proposed in Cuchiero et al. (2020).

Once the SABR parameters have been calibrated following the methodology highlighted above, then the calibration of the local volatility component of the LSV SABR model (i.e., the leverage function) can take place, at risk-neutral level, following the procedure highlighted in Section 4.7.3.

We now provide an in-depth example of the calibration routine just outlined. We recall that, in a two-price economy under conic finance settings, the calibration is a two-step process. That is, first the parameters describing the risk-neutral dynamics of the underlying process need to be calibrated. Afterwards, the calibration to the observed bid and ask prices can take place. In this example, as proxy for the risk-neutral prices, we consider their mid-counterparts. The results of the second step, i.e., those related to the bid-ask calibration, will be outlined in Section 4.7.4.

We start by considering, under risk-neutral settings, the SABR model (57), to be calibrated maturity per maturity. That is, we first calibrate the SABR model to the smile corresponding to each maturity. Afterwards, given that now that the benchmark calibration has been performed, we proceed with the second step. That is, we calibrate the actual SABR LSV model outlined in Section 4.7.3 (under risk-neutral settings). As the final step for the risk-neutral calibration, we now add the local volatility component to the SABR model and fit the model to the market data by means of approximating its local volatility component with a neural network as illustrated in Section 4.7.2.

In order to provide a transparent analysis of the calibration results, we now quantify and compare the errors obtained using a pure SABR calibration versus the case where a local volatility component is introduced in the model. The results of this comparison are available in Table 4.

T	SABR Max. Err.	SABR Avg. Err.	LSV Max. Err.	LSV Avg. Err.
0.17	0.0042	0.0019	0.0021	0.0010
0.33	0.0048	0.0026	0.0024	0.0004
0.47	0.0054	0.0027	0.0006	0.0002
0.62	0.0055	0.0025	0.0010	0.0006
0.89	0.0055	0.0021	0.0010	0.0003
1.04	0.0063	0.0020	0.0022	0.0008
Agg.	0.0063	0.0023	0.0024	0.0005

Table 4: Statistics related to the calibration errors when comparing the pure SABR calibration (see "SABR Max. Err." and "SABR Avg. Err." columns) with its LSV extension (see "LSV Max. Err." and "LSV Avg. Err." columns). Results are expressed per maturity. The last row (i.e, "Agg.") reports the relevant quantities at aggregate level.

These statistics reveal that the average absolute error in the SABR calibration is 23 basis points, while in the SABR LSV approach this error is reduced to 5 basis points. This is a very significant improvement, as the results indicate that the (average) calibration errors reduce by a factor of (at least) four when a local volatility component is also taken into account. Consequently, this firmly establishes robust evidence in favor of the effectiveness of the analyzed methodology, which, by relinquishing its reliance on analytical formulas, achieves significantly greater accuracy. This is not only important from a pure pricing perspective, but also from a risk management angle, amongst others.

Finally, for making the error analysis even more complete, we provide additional details concerning the implied volatility squared errors of the SABR model without and with a local volatility component. The results of this analysis are presented in Figures 28 and 29, respectively. The outcomes clearly illustrate how introducing a local volatility component within the SABR calibration allows to obtain (squared) errors in volatility space which are one order of magnitude lower than their counterparts computed using a pure SABR approach. In particular, from the empirical analysis provided it is clear that introducing a leverage function in the calibration allows to drastically reduce calibration errors, as it is outmost evident close to the first maturity considered. We also observe that, with the exception of a handful of points, the implied volatility squared error in the case of the LSV calibration (see Figure 29) is smoother that its pure SABR counterpart (see Figure 28). And it is also worth remarking that, even when the peaks observed in Figure 29 are taken into account, the values they correspond to are, in any case, smaller than their counterparts available in Figure 28.

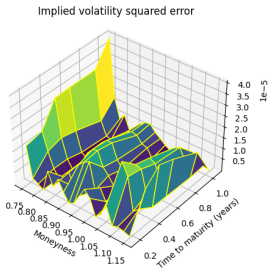


Figure 28: Surface of the squared calibration errors in implied volatility terms corresponding to the SABR model.

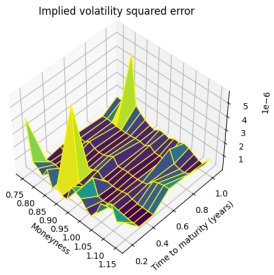


Figure 29: Surface of the squared calibration errors in implied volatility terms corresponding to the LSV SABR model.

Results: Bid-ask calibration

In this section, we present the results obtained by extending the calibration LSV SABR to conic two-price settings. The methodology outlined in Section 4.3.2,

as well as the results obtained Section 4.7.4, serve as foundations for this extension. In particular, combining a neural network calibration of an LSV model with conic Monte Carlo is, to the best of our knowledge, a novel technique yet to be experimented. The results outlined here are a continuation of those available in Section 4.7.4. That is, after having performed the calibration of the LSV model at risk-neutral level we can, by means of conic Monte Carlo (see Section 4.3.2), perform a bid-ask calibration given real market data. Furthermore, we also present a selection of calibration results coming from the application of some of the hybrid families of distortion functions introduced in Section 4.3.3. For the sake of conciseness, we exhibit here, from a graphical perspective, only the calibration results pertaining to the first and to the last maturity.

In Figure 30, we present the bid and ask implied volatilities obtained following the pure conic SABR calibration outlined in Section 4.3.1. Additionally, for the purpose of comparison, we include the market bid and ask implied volatilities, as well as the market mid-price implied volatilities.

An examination of Figure 30 reveals an often poor performance of the calibration method, which seems particularly evident for options with low moneyness levels and short maturities. The plots in Figure 30 illustrate a significant bid-ask spread which the calibration fails to capture, as evidenced by the high repricing errors for certain options, where the calibrated bid implied volatility differ by approximately 800 basis points from the market bid implied volatility.

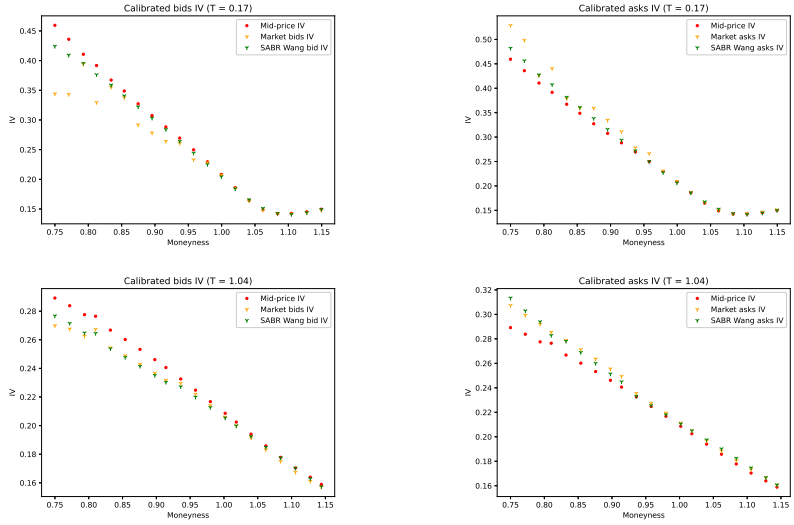


Figure 30: Calibrated bid (left panels) and ask (right panels) implied volatilities using the conic SABR model for the first (top panels) and the last (bottom panels) maturities in the dataset.

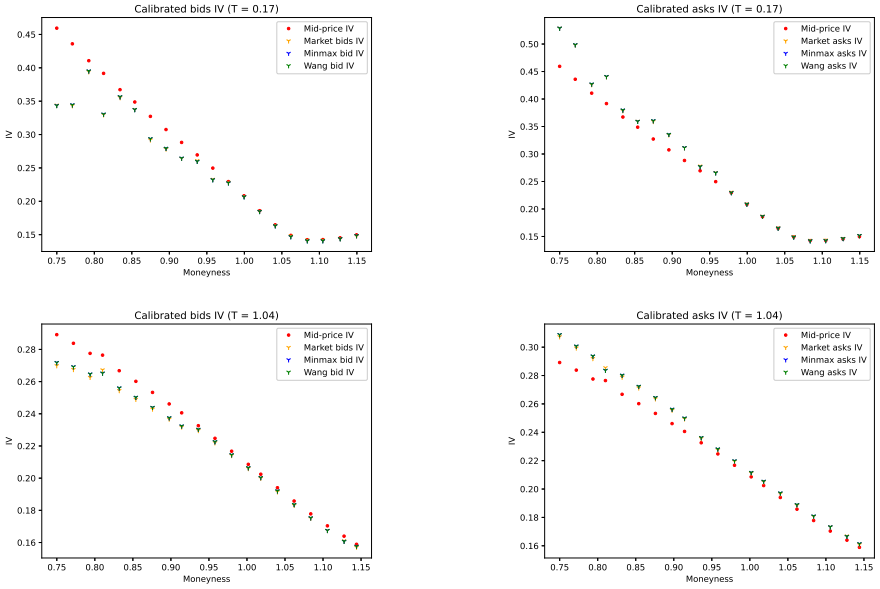


Figure 31: Calibrated bid (left panels) and ask (right panels) implied volatilities using the LSV SABR model for the first (top panels) and the last (bottom panels) maturities in the dataset.

For comparison purposes, in Figure 31 we present the results obtained by calibrating the LSV SABR model by means of conic Monte Carlo given the same dataset considered so far. Furthermore, in Figure 31 we present bid and ask implied volatilities obtained through the LSV SABR calibration using both the Minmaxvar and the Wang distortion functions.[43] The results demonstrate a crystal clear improvement in terms of calibration accuracy.

We now also present, for completeness, some calibration results obtained by employing the novel t-Minmaxwang distortion introduced in Section 4.3.3. The key distinction of this approach is the inclusion of an additional parameter, denoted as t[44]. This parameter serves as a metric for selecting the appropriate distortion in each instance. This is a useful feature because, as highlighted in Section 4.3.3, often the choice of the distortion function to be used is somewhat arbitrary. And, therefore, having a simple tool allowing to discriminate between the possible choices available can often come in handy. Through the analysis of the results we find that the majority of points in the volatility surface exhibit a

43 We remark here that, in this framework, as we are not dealing with analytical formulae, it does not make a difference, from an implementation angle, whether we use one distortion function or another. On the contrary, in the case of the conic SABR model (see Section 4.3.1), the choice of the Wang transform (instead of another distortion) allows to still remain in a framework where (approximated) analytical formulae could be used.

44 We use t in this context due to the interpretation of this distortion function provided in Section 4.3.3.

(close to) zero value for t, indicating a predominance of the Wang transform in terms of performance accuracy. Based on this observation, we can empirically assess that, at least for the cases considered in the examples available in the paper at hand, the Wang transform exhibits superior performances compared to the Minmaxvar distortion. We present in Table 5 the (average) calibration errors computed using the different distortion functions mentioned above (i.e., Wang, Minmaxvar, t-Minmaxwang, Wang-Minmaxvar and Minmaxvar-Wang). From the results obtained, and as expected by construction, we observe that the t-Minmaxwang provides the best fit (due to the extra degree of freedom introduced).

Err. type	Minmaxvar	Wang	t-Minmaxwang	Wang-Minmaxvar	Minmaxvar-Wang
Bid Err.	0.000599	0.000585	0.000575	0.000591	0.000595
Ask Err.	0.000516	0.000508	0.000501	0.000511	0.000513

Table 5: Average absolute implied volatility error per distortion function given the dataset considered.

4.8 CONCLUSION

In the article at hand we have investigated new ways of combining the usage of neural networks in quantitative finance and, in particular, in the context of derivatives valuation in a two-price economy according to the paradigm of Cherny and Madan (2010). Notably, we have examined two different applications of neural networks in conic financial markets. The first explores the possibility of replicating bid and ask pricing by the use of (vector-valued) multilayer perceptrons, while the second aims to combine the latter technology with LSV modeling. Specifically, (i) we provide a fully neural network based architecture for bid and ask derivative valuations. Moreover, this methodology (ii) allows to replicate theoretical bid and ask prices with a high level of accuracy (iii) in a model agnostic manner. Furthermore, this novel approach (iv) allows for extraordinary computational performance improvements that can reach orders of magnitude compared to simulation methods, and also provides a fast and efficient risk-neutral model calibration as additional advantage. This technique (v) has also been extended for calculating sensitivities. Furthermore, we also (vi) provide novel developments in LSV modeling by extending the work of Cuchiero et al. (2020) to bid-ask pricing by means of conic Monte Carlo techniques. On the side, (vii) we additionally introduce the conic version of the SABR Hagan et al. (2014) model and, lastly, (viii) we provide simple techniques to generate arbitrary families of hybrid distortion functions that can be used as decision making tools for conic modeling purposes. All the work outlined in the document at

hand is novel and, to the best of our knowledge, it provides the first examples of neural network based bid-ask pricing under conic settings.

4.A REPLICATION OF THE CONIC SABR MODEL BY MEANS OF A CNN

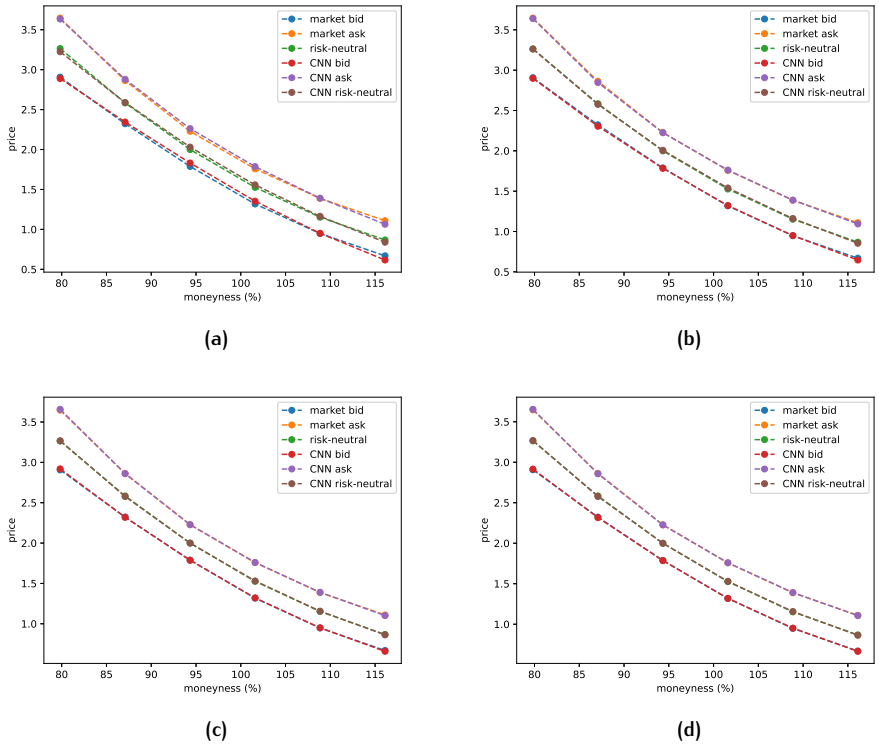


Figure 32: Replication of quoted bid and ask option prices, as well as their risk-neutral counterparts, using the conic SABR model and its CNN extension. Risk-neutral prices have been calculated by considering $\lambda = 0$. The number of points used in the CNN equals 5 in panel (a), 10 in panel (b), 15 in panel (c), and 20 in panel (d). The number of points should be interpreted per training parameter.

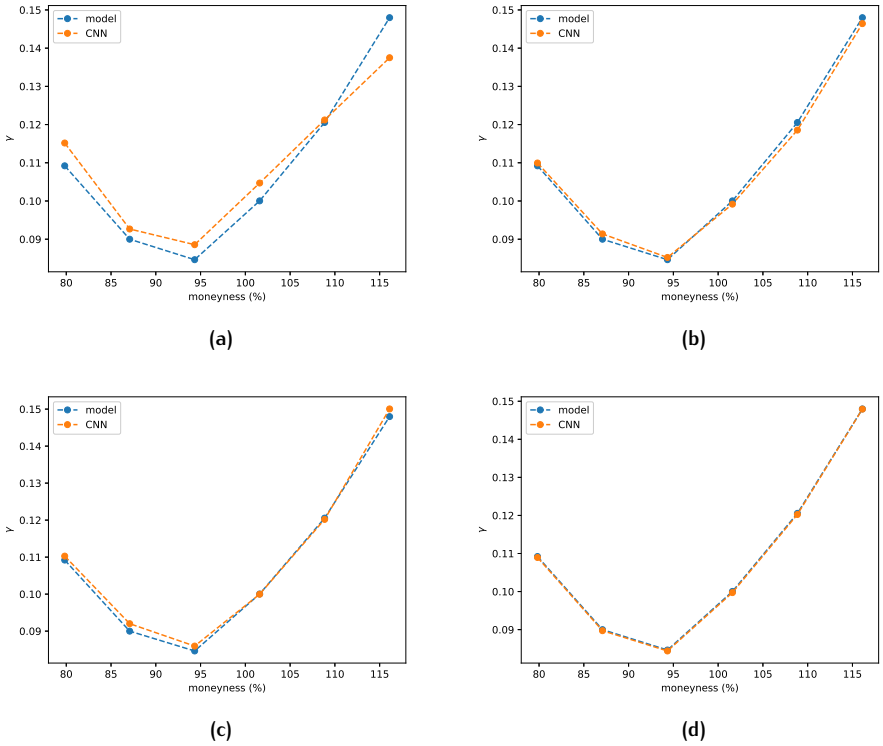


Figure 33: Replication of implied liquidity amounts computed using the conic SABR model by means of its CNN extension. The number of points used in the CNN equals 5 in panel (a), 10 in panel (b), 15 in panel (c), and 20 in panel (d). The number of points should be interpreted per training parameter.

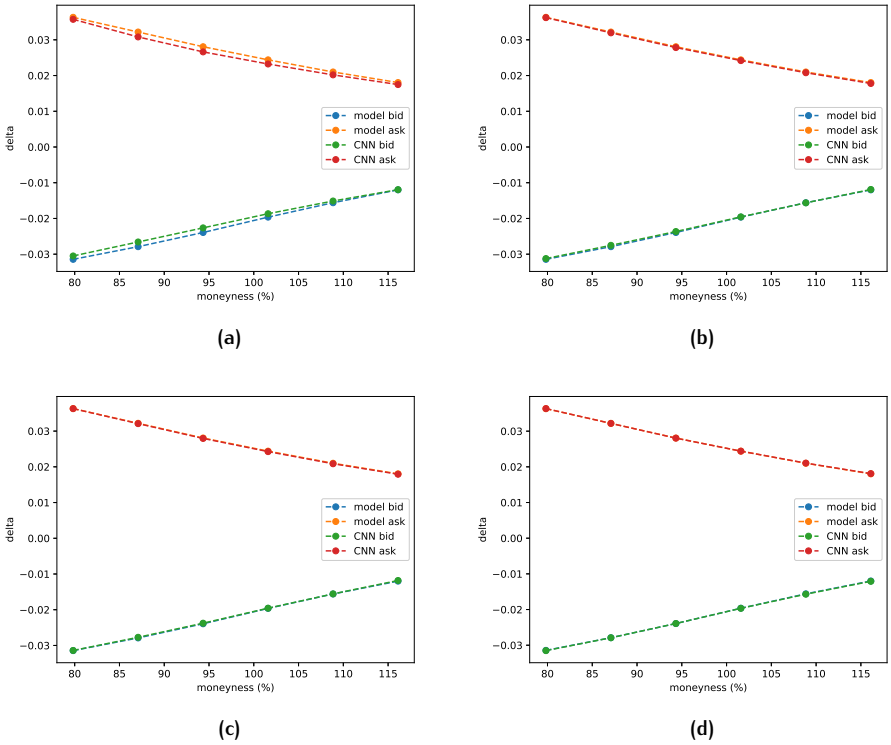


Figure 34: Replication of implied liquidity delta amounts computed using the conic SABR model by means of its CNN extension. The number of points used in the CNN equals 5 in panel (a), 10 in panel (b), 15 in panel (c), and 20 in panel (d). The number of points should be interpreted per training parameter. Liquidity delta amounts have been multiplied by 1%, in line with what is often done in risk monitoring, e.g., for vega calculations.

Part II

WASSERSTEIN DISTANCES AND PROXY CREDIT CURVES

5 PROXYING CREDIT CURVES VIA WASSERSTEIN DISTANCES

5.1 ABSTRACT

Credit risk plays a key role in financial modeling, and financial institutions are required to incorporate it in their pricing, as well as in capital requirement calculations. A common manner to extract credit worthiness information for existing and potential counterparties is based on the credit default swap (CDS) market. Nonetheless, not all counterparties of a financial institution have (liquid) CDSs traded in the market. In this case, financial institutions shall employ a proxy methodology to estimate the default probabilities of these counterparties. Starting from the intersection methodology for credit curves, in this article we investigate whether it is possible to construct proxy credit curves from CDS quotes by means of (weighted) Wasserstein barycenters. We show how, under simple and common assumptions, this revised methodology leads to elementary and intuitive formulae to calculate distances between CDS-implied default probability distributions. Further, we illustrate how to use this information to construct proxy CDS quotes.

5.2 INTRODUCTION

In this article we investigate an alternative approach to construct proxy credit curves starting from credit default swaps (CDSs) market data. In particular, we revise the *intersection methodology* by calculating the (weighted) *Wasserstein barycenter* of the implied CDS credit curves considered as inputs. Under simple assumptions our new methodology results in analytical formulae for the proxy CDS hazard rates with an intuitive interpretation.

Banks are required, for instance in valuation adjustment, risk limits and capital calculations, to account for credit risk as accurately as possible. Nonetheless, not for all counterparties credit-related information can be (easily) extracted from financial instruments traded in the market. For instance, simple instruments used to strip default probabilities of given counterparties are CDSs. However, not all counterparties have liquid CDSs trading in the market, and many have no CDSs trading at all. Therefore, proxy methodologies should be employed, as for instance the intersection methodology described in EBA (2013) (see also

Chourdakis et al. (2013a) and Sourabh et al. (2018) for further details). In this approach the proxy spread for an illiquid entity with a given rating, region, sector, etc. is calculated as the average of all the available liquid spreads of the entities with the same characteristics, i.e., rating, region, sector, etc.[1] In this article we investigate whether the Wasserstein square distance can be considered as a tool to substitute the aforementioned average in the intersection model. In particular, we show how using the Wasserstein distance allows, given some simple assumptions, to obtain proxy CDS curves in a simple way.

The modeling of the default time of a counterparty within the risk-neutral framework can be easily embedded in pricing and hedging equations; see Jarrow and Turnbull (1995) and (Ammann, 1999, Ch. 5), amongst others. However, this is based on the implicit assumption that faithful credit data is available. Unfortunately, this is not necessarily guaranteed in practice. In fact, see (Green, 2015, Ch. 4.6), the number of counterparties worldwide for which CDS quotes are available is in the order of a few thousand and, further, not all the available quotes are liquid enough to be considered trustworthy. Therefore, it does not come as a surprise that, for the most of the counterparties of financial institutions, no (liquid) CDSs are available to imply the required default probabilities. From this, the need of using a CDS proxy methodology to compensate for the lack of available (or reliable) data arises.

Despite the fact that financial institution can decide, provided approval is granted by validators, auditors and regulators, for the methodology to be used to proxy CDS curves, the literature is quite scarce. For instance, financial institutions can map the missing or unreliable CDS data needed for given counterparties to that of other counterparties for which liquid single-name CDS data is available, based on selection criteria such as rating, region, sector, etc.; see (Green, 2015, Ch. 4.6.1). This has the advantage that, if the mapping selection is done carefully, then the proxy curves would reflect the same rating-, region- and sector-specific risks of the company whose CDS curve needs to be proxied. From a theoretical perspective, this would allow to hedge the credit risk by taking positions in the mapped CDS. However, this would not guarantee synchronicity of default or credit downgrades/upgrades (if any) between the illiquid counterparty and its proxy. This might result in unexpected gains or losses, as well as in potential re-hedging costs. Alternatively, see (Green, 2015, Ch. 4.6.2), missing single-name CDS data can be mapped to given CDS indices. Notwithstanding, this can lead to similar situations as the one highlighted in the single-name mapping case. A way to construct proxy CDS curves that does not involve simply mapping missing CDS data to that of other companies or indices is the so-called intersection methodology EBA (2013). This approach suggests to bucket liquid CDSs according to their rating, region and sector. By averaging the CDS quotes

1 In principle, more granular bucketing can take place, but in this article we do not deal with the choice of the bucketing methodology, and we consider it as given.

in each bucket, one can then define proxy CDS quotes per bucket and maturity (see also Chourdakis et al. (2013a) and Sourabh et al. (2018)). This approach is very intuitive and easy to implement. Also, should one need to make the approach more granular, then this could be easily achieved by considering a finer bucketing methodology.[2] However, the intersection methodology has the disadvantage that, in some cases, not many CDS names are available with a certain rating, region and sector, which might cause some buckets to only contain a handful of quotes (or even to be empty). We also recall the so-called *cross-sectional methodology* developed in Chourdakis et al. (2013a) (see also Chourdakis et al. (2013b) and (Green, 2015, Ch. 4.6.3)). This methodology assumes that each CDS spread can be decomposed in a number of factors, which are then calibrated to the available CDS spreads by means of least-square regression. This approach has the advantage, as its name suggests, of taking into account the cross-sectional information of CDS spreads arising from different ratings, regions and sectors. The cross-sectional methodology of Chourdakis et al. (2013a) has been extended in Sourabh et al. (2018) in order to potentially include equity returns in the regression to further enhance the reliability and stability of the results.

In this article we consider the intersection methodology as a starting point, and we look at it from another angle. Assuming the available CDS quotes have been allocated according to a given bucketing criterion we investigate whether, instead of averaging the quotes, it is possible to create CDS curves in a different manner. In particular, we attempt to do so by means of calculating Wasserstein square distances between the CDS-implied default probability curves corresponding to the instruments allocated in the same bucket. We show that, under some well-known approximations, this is possible. This leads to analytic formulae for the hazard rates of the proxy CDSs, which are also intuitive from the interpretation angle.

This article is organized as follows. Section 5.3 provides the essential concepts on Wasserstein distances relevant for the methodology developed in this article, and it also outlines the basic notions related to CDSs. In Section 5.4 the methodology to proxy CDS curves using Wasserstein square distances is available. In Section 5.5 we comment on the possible use of some alternative metrics between probability distributions for the purpose of proxying CDS curves and further motivate our choice. In Section 5.6 two extensive examples of proxy CDS quotes construction are provided. Section 5.7 concludes.

2 Therefore, from here onwards we will often not refer explicitly to the choice of the buckets as, from the implementation angle, this does not matter. We will thus often use more general expressions such as, for instance, "bucketing methodology", for brevity.

5.3 BACKGROUND INFORMATION

In this section we outline the basic notions which are needed throughout this article. In particular, Section 5.3.1 provides the essential background concerning Wasserstein distances, while Section 5.3.2 that on CDSs. In both Sections 5.3.1 and 5.3.2 the interested reader will find relevant references should more detailed and general information be needed.

5.3.1 A brief overview of Wasserstein distances

In this section, we provide the essential notions concerning Wasserstein distances as necessary in the rest of the article. For an in-depth treatment of Wasserstein distances see (Villani, 2003, Ch. 7), (Villani, 2009, Ch. 6), or (Panaretos and Zemel, 2020, Ch. 1 and 2), where Wasserstein distances are treated from a more general angle covering aspects that go beyond our needs in this article.

The concept of Wasserstein distance finds its origin in the area of optimal transport. In particular, it aims to measure the minimal work that is required to rearrange the probability masses of one distribution in order to reconstruct a second distribution. This transportation plan should be performed in order to minimize the cumulative "distance" that the various probability masses have to go through.

Given any two probability measures on $\mathbb{R}^d$ denoted with $\mathbb{P}$ and $\mathbb{Q}$, we indicate with $\mathcal{J}(\mathbb{P},\mathbb{Q})$ the set containing all the joint distributions which have $\mathbb{P}$ and $\mathbb{Q}$ as marginals. For $p \geqslant 1$, the p-Wasserstein distance between $\mathbb{P}$ and $\mathbb{Q}$ is defined as

$$W_p(\mathbb{P},\mathbb{Q}) := \left(\inf_{J \in \mathcal{J}(\mathbb{P},\mathbb{Q})} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\|^p \, J(dx \times dy) \right)^{\frac{1}{p}}, \tag{94}$$

where $\|\cdot\|$ denotes the Euclidean norm. The Wasserstein distance (94) is finite when the distributions have finite moments of order p; see Panaretos and Zemel (2019). When $p = 2$, (94) is referred to as *Wasserstein square distance*, which is the analogous in a probabilistic sense of the Euclidean space distance. In the case of unidimensional random variables (94) can be easily calculated via the relationship

$$W_p(\mathbb{P},\mathbb{Q}) = \left(\int_0^1 |F_{\mathbb{P}}^{-1}(u) - F_{\mathbb{Q}}^{-1}(u)|^p \, du \right)^{\frac{1}{p}}, \tag{95}$$

where $F_{\mathbb{P}}^{-1}(\cdot)$ ($F_{\mathbb{Q}}^{-1}(\cdot)$) denotes the (pseudo-)inverse of the cumulative distribution function of $\mathbb{P}$ ($\mathbb{Q}$); see Panaretos and Zemel (2019). In the present article, we will be interested in the unidimensional case only and, in particular, in the case when $p = 2$.

We recall now the concept of *weighted Wasserstein barycenter*, as it will be useful for an analogy we will provide in Section 5.4. Given N probability measures

$\mathbb{P}_1, \ldots, \mathbb{P}_N$ with the characteristics as above and positive weights $w_1, \ldots, w_N$ such that $\sum_{i=1}^{N} w_i = 1$, a Wasserstein barycenter, see Agueh and Carlier (2011), is any probability distribution $\mathbb{P}^*$ solution of

$$\inf \left\{ \sum_{i=1}^{N} w_i W_2^2(\mathbb{P}_i, \mathbb{P}) : \mathbb{P} \text{ has finite second moment} \right\}. \tag{96}$$

For practical applications, we recall that if the probabilities $\mathbb{P}_i$'s are absolutely continuous, then both existence and uniqueness of the Wasserstein barycenter are guaranteed; see Agueh and Carlier (2011).[3] Often, the definition of weighted Wasserstein barycenter is provided with uniform weights, i.e., with $w_i := \frac{1}{N}$ for every i. In that case, the dependence on $w_1, \ldots, w_N$ can be then simply omitted.[4]

The popularity of Wasserstein distances has increased considerably in the last decade, as well as the number of applications of Wasserstein distances in different fields. This is due to some desirable properties Wasserstein distances exhibit, including, but not limited to, their behavior along geodesics (see (Ambrosio et al., 2005, Ch. 7)), and to Wasserstein distances incorporating the geometry of the ground space (see Panaretos and Zemel (2019)). We also recall the use of Wasserstein distances in statistical learning Karimi et al. (2020), auto-encoders Zhang et al. (2021), martingale optimal transport Guo and Obłój (2019), and image labelling Hühnerbein et al. (2018), to name but a few. For computational challenges concerning Wasserstein barycenters see Altschuler and Boix-Adserà (2022); Altschuler and Boix-Adserà (2021), while for an overview of optimal transport theory from a computational angle with applications in machine learning see Peyré and Cuturi (2019).

5.3.2 A brief overiew of CDSs

A CDS is a credit derivative contract involving two counterparties: a protection buyer and a protection seller. In particular, by entering the transaction the protection buyer obtains protection against a contract-specific credit event (e.g., a credit rating downgrade or a default) concerning a given reference entity. In exchange for holding the credit risk, the protection seller expects the counterparty to pay a recurrent protection fee until either the contract expires or a credit event occurs. CDSs are nowadays traded given an upfront premium and a fixed coupon. The CDS market provides an important source of data that can be used to estimate the default probabilities of different entities. These default proba-

3 Note that more general conditions are sufficient for the existence and uniqueness of a Wasserstein barycenter; see Agueh and Carlier (2011). However, for our purpose of constructing proxy CDS curves we will be working with absolutely continuous probability measures only.

4 For simplicity, from here onward we will refer to (96) simply as Wasserstein barycenter.

bilities can be then employed in various areas related to credit risk modeling, including valuation adjustment calculations.

The process of extracting default probabilities from CDS quotes can be performed in different manners. A simple and effective way to do so is by means of *reduced-form* models, which are based on the idea of assigning a given functional form for the default probability function and to calibrate its parameters in order to reprice the available CDS contracts in the market (for a detailed overview of several approaches concerning modelling default probabilities see Bielecki and Rutkowski (2010)). In particular, a simple manner to model the default probability of an entity is that of assuming its (risk-neutral) default probability function to be defined via the relationship

$$\mathbb{Q}(\tau \leqslant t) \coloneqq 1 - e^{-\int_0^t \lambda(s)\,ds}. \tag{97}$$

In (97), $\mathbb{Q}$ denotes the risk-neutral measure, τ the (random) default time of the entity, and the (deterministic) function $\lambda : (0, +\infty) \to (0, +\infty)$ is named *hazard rate* (or *default intensity*).[5] Different specifications for the hazard rate function can be chosen: common ones are flat, piecewise-constant, piecewise-linear and cubic spline.

We now recall some simple and well-known approximations for hazard rates, known as *credit triangles*, which are often used in practice (see, amongst others, Berd (2005), Berd (2011), White (2013), Gambetti et al. (2018) and (De Spiegeleer, Hulle, et al., 2014, Ch. 10.4) for further details). The CDS *par spread* is defined as the coupon of the fixed leg that would make the CDS contract trade at par. This used to be the standard way of quoting CDSs. Having denoted the recovery rate of a given CDS as R[6] and its par spread as s, assuming the hazard rate function in equation (97) is flat, one can then approximate the constant hazard rate in (97) via

$$\lambda \approx \frac{s}{1 - R}.^{7} \tag{98}$$

This means that the distribution of the default time τ is modeled by means of an *exponential distribution* with *rate parameter* λ. Note that, as a consequence of the CDS standardization process that took place after the credit crisis (see, amongst others, Markit (2009a) and Markit (2009b)), nowadays it is market practice to quote CDSs is by means of a standard coupon and an upfront payment. In this case, the credit triangle relationship would read $\lambda \approx (c + \mathrm{PUF}/T)/(1 - R)$, where c denotes the standard coupon of the CDS, PUF the *points up-front*, i.e., the upfront payment quoted as percentage of the notional, while T the time to maturity of the contract considered.[8] The simplifying assumption of a constant

5 Note that we have excluded the case where $\lambda(t) = 0$ as this corresponds to the (instantaneous) default-free case.

6 The recovery rate R is assumed exogenously given. Further details can be found in Das and Hanouna (2009).

7 For a derivation of the credit triangle relationship see, for instance, White (2013).

8 Also in this case, the derivation of the credit triangle relationship can be found in White (2013).

hazard rate is often used in practice to convert from one quoting convention (i.e., points upfront and standard coupon) to the other (i.e., par spread); see White (2013).

5.4 CDS PROXY CURVES USING WASSERSTEIN SQUARE DISTANCES

Assume we have grouped liquid CDS quotes given some bucketing criteria (e.g., their sector, region, etc.), and that one bucket for which we want to extract a proxy CDS curve has been selected. For each of the CDSs in the group we have a set of maturities. We choose a maturity amongst the ones available, as the process can be performed for each maturity separately. We assume that $N \geqslant 2$ par spreads are available (see Section 5.3.2), and we denote them as $s_1, \ldots, s_N$ (we could follow a similar approach if we decided to start from upfront premia instead of from par spreads; again, see Section 5.3.2). Thus, for each of the CDS quotes in this group (recall the maturity is fixed) we have a par spread. For each CDS we now compute the *flat* hazard rate via the credit triangle relationship (98). The approximation of having a flat hazard rate per CDS basically says that, once a CDS is considered, this CDS belongs to an economy composed by the risk-free bond and the CDS itself. Thus, the easiest way to calculate a hazard rate is by assuming it flat, as done in Section 5.3.2. This assumption is a simplifying one, as it does not take into account the term structure of the hazard rates for different CDS maturities. However, it is a common approximation used in practice, as for instance credit triangles are based on this assumption; see (98). For our group of N CDSs, we use the assumption of a constant hazard rate, and thus for each par spread s_i we can calculate a flat hazard rate λ_i. This approximation based on a constant hazard rate assumption is the same used to perform the conversion between upfront payments and related par spreads; see White (2013). This approximation is also used, for instance, in Madan (2014) in the context of measuring risk acceptability within the CDS market. The reason why we propose to use this simplifying assumption is that it allows to exactly calculate proxy hazard rates. More precisely, this is because having a constant hazard rate per maturity allows to analytically calculate Wasserstein distances between CDS-implied distributions, as shown in Theorem 3. The distribution function of the default time implied from the i^{th} CDS is denoted as $F_{Q_i}(t) \coloneqq \mathbb{Q}(\tau_i \leqslant t)$, with τ_i indicating the default time of the i^{th} reference entity. For the i^{th} CDS, we can calculate the inverse of the default probability distribution explicitly, i.e., $F^{-1}_{Q_i}(u) = -\frac{\ln(1-u)}{\lambda_i}$ by inverting (97). We can consider an additional CDS (i.e., the *proxy* CDS) which, under the same assumptions of constant hazard rate, defines the quantile function $F^{-1}_{Q^*}(u) = -\frac{\ln(1-u)}{\lambda}$ for a given constant λ to be

determined. We can then calculate the *weighted square error* in Wasserstein square distance between the distribution of the proxy CDS $\mathbb{Q}^*$ and those of the other N CDSs $\mathbb{Q}_1, \dots, \mathbb{Q}_N$, i.e.,

$$wse(\lambda) \coloneqq \sum_{i=1}^{N} w_i W_2^2(\mathbb{Q}_i, \mathbb{Q}^*), \tag{99}$$

where $\sum_{i=1}^{N} w_i = 1$ and with $w_i > 0$ for every i.[9] By minimizing (99) with respect to λ we can then obtain the optimal hazard rate. Note the similarity of (99) with the definition of Wasserstein barycenter (96): it is clear how (99) represents the Wasserstein barycenter between the probabilities $\mathbb{Q}_i$'s after a functional form for the barycenter itself has been chosen. Therefore, the probability measure $\mathbb{Q}^*$ in (99) can be interpreted as a *pseudo* Wasserstein barycenter for the default distributions we have considered.

We now provide an analytical expression for the optimal hazard rate in Theorem 3, which illustrates how the optimal hazard rate is actually the *weighted harmonic mean* of the hazard rates λ_i's. Note that, from here onward, the index i will be always assumed to range in $\{1, \dots, N\}$, for ease of exposition.

Theorem 3. *Let $\mathbb{Q}_1, \dots, \mathbb{Q}_N$ be exponential distributions where, for every i, $\mathbb{Q}_i$ has rate parameter λ_i. Further, let $\mathbb{Q}^*$ be another exponential distribution with rate parameter*

$$\lambda^* \coloneqq \frac{1}{\sum_i \frac{w_i}{\lambda_i}}. \tag{100}$$

Then, λ^* is the unique minimizer of $wse(\,\cdot\,)$.

Proof. Recalling that the quantile function of an exponential distribution $\mathbb{Q}$ with rate parameter λ is given by $F_{\mathbb{Q}}^{-1}(u) = -\frac{\ln(1-u)}{\lambda}$, it results that

$$W_2^2(\mathbb{Q}_i, \mathbb{Q}) = \int_0^1 \left(\frac{\ln(1-u)}{\lambda_i} - \frac{\ln(1-u)}{\lambda} \right)^2 du = 2 \left(\frac{1}{\lambda_i} - \frac{1}{\lambda} \right)^2. \tag{101}$$

Let $x \coloneqq \frac{1}{\lambda}$. It is sufficient to consider the function

$$f(x) \coloneqq \sum_i w_i \left(\frac{1}{\lambda_i} - x \right)^2$$

9 Note that, in the case of uniform weights, the weights would be redundant. For completeness, however, we explicitly consider the weights in (99). This is because, in practice, if a bucket only contains a handful of names, then the modeler might want to add to the bucket some other CDSs (for instance with the same sector and rating but belonging to a different region) and assign them a small weight to have more stable results. However, as mentioned in the beginning of the section, in this article a bucketing methodology is assumed as given. We provide the formulation (99) (as well as that of weighted Wasserstein barycenter (96)) as they are slightly more general than the uniform case, allowing thus for more flexibility.

and to observe that $\frac{d}{dx}f(x) = 0$ when $x = \frac{1}{\lambda^*}$, while $\frac{d^2}{dx^2}f(x) \equiv 2$ due to the weights w_i's summing up to one. □

Note that the mean of $\mathbb{Q}_i$ equals $\frac{1}{\lambda_i}$. Thus, due to Theorem 3 the reciprocal of the proxy hazard, i.e., $\frac{1}{\lambda^*}$ can be interpreted as the weighted average expectation of the N default probability distributions considered. This means that the weighted harmonic mean of the λ_i's is actually the correct way of averaging them, as they can be interpreted as rates.

We highlight that using flat hazard rates is the key simplifying hypothesis that guarantees the simple formula of the optimal hazard rate (100) to hold. If this assumption were to be dropped and a more elaborate functional form for the hazard rate term structure were to be chosen, then a closed form for computing the optimal (weighted sum of) Wasserstein distance(s) (99) would not be available. Instead, it would require to solve a minimization problem in several unknowns (i.e., one per CDS maturity available), making the whole procedure more complicated and computationally expensive. More precisely, given N CDS names available per maturity, let M denote the total number of quoted CDS maturities (M, in many case would be equal to 11, as ideally CDS quotes are available for the 6M, 1Y, 2Y, 3Y, 4Y, 5Y, 7Y, 10Y, 15Y, 20Y and 30Y maturities). Assume that default probabilities were to be modeled by means of (97) in a more elaborate manner such as piecewise-constant or piecewise-linear.[10] Denote with λ_i^j the hazard rate corresponding to the i^{th} CDS name and to the j^{th} maturity. Also, denote with $F^{-1}_{\mathbb{Q}_i}(\lambda_1^i, \ldots, \lambda_M^i; \cdot)$ the quantile function implied from the i^{th} CDS, and with $F^{-1}_{\mathbb{Q}^*}(\lambda_1, \ldots, \lambda_M; \cdot)$ the quantile function generated by the (unknown) proxy CDS yet to be estimated. One would then need to find a solution, via (95), of the constrained optimization problem

$$\min_{\lambda_1,\ldots,\lambda_M} \sum_i \int_0^1 \left(F^{-1}_{\mathbb{Q}_i}(\lambda_1^i, \ldots, \lambda_M^i; u) - F^{-1}_{\mathbb{Q}^*}(\lambda_1, \ldots, \lambda_M; u) \right)^2 du, \tag{102}$$

where each λ_i is required to be positive. First of all, it is not guaranteed that, for arbitrary choices of the hazard rate function, a unique solution exists. Furthermore, the integrand in (102) cannot be, in general, integrated analytically, and the relatively high number of degrees of freedom (i.e., M), together with the potentially high number of CDSs in the bucket considered, would make the problem very expensive from the computational angle. Thus, these observations tip the balance in favor of a simpler approach, as the one we have adopted involving a constant hazard rate.

We now provide a second intuitive result highlighting some relationship between the observed hazard rates (implied default distributions) and its optimal counterpart. Its proof is trivial, and it directly follows from the interpretation of

10 Note that, in principle, also other parametrizations of the hazard rate functions are possible, such as cubic spline.

the optimal hazard rate λ^* as an harmonic mean and from the (strict) monotonicity of the distribution function of an exponential random variable with respect to the rate parameter. For ease of readability, we denote with $\lambda_{\min} \coloneqq \min_i \lambda_i$ and with $\lambda_{\max} \coloneqq \max_i \lambda_i$, while with $i_{\min}$ and $i_{\max}$ the corresponding indices, i.e., such that $\lambda_{i_{\min}} = \lambda_{\min}$ and $\lambda_{i_{\max}} = \lambda_{\max}$.

Corollary 1. *Under the assumptions of Theorem 3 it holds that*

$$\lambda_{\min} \leqslant \lambda^* \leqslant \lambda_{\max}. \tag{103}$$

Furthermore, for every $t \in [0, +\infty)$ *it results that*

$$F_{Q_{i_{\min}}}(t) \leqslant F_{Q^*}(t) \leqslant F_{Q_{i_{\max}}}(t). \tag{104}$$

So far, we have defined the probability distribution Q^* as the pseudo-barycenter generated by considering $Q_1, \ldots, Q_N$. We now show that Q^* represents the actual Wasserstein barycenter of the probability distributions considered.

Theorem 4. *Under the assumptions of Theorem 3 it holds that*

$$\sum_i w_i W_2^2(Q_i, Q^*) = \min\left\{\sum_i w_i W_2^2(Q_i, Q) : Q \text{ \textit{has finite second moment}}\right\}. \tag{105}$$

Proof. We denote with $F_Q^{-1}(\,\cdot\,)$ the quantile function of a given probability distribution Q with finite second moment, and define

$$\Delta_Q \coloneqq \sum_i w_i W_2^2(Q_i, Q) = \sum_i w_i \int_0^1 \left(F_{Q_i}^{-1}(u) - F_Q^{-1}(u)\right)^2 \, du. \tag{106}$$

By adding and subtracting $F_{Q^*}^{-1}(\,\cdot\,)$ inside the brackets in the third member of (106) it results that

$$\begin{aligned}\Delta_Q = &\sum_i w_i \int_0^1 \left(F_{Q_i}^{-1}(u) - F_{Q^*}^{-1}(u)\right)^2 \, du \\ &+ \int_0^1 \left(F_{Q^*}^{-1}(u) - F_Q^{-1}(u)\right)^2 \, du \\ &+ 2\sum_i w_i \int_0^1 \left(F_{Q_i}^{-1}(u) - F_{Q^*}^{-1}(u)\right)\left(F_{Q^*}^{-1}(u) - F_Q^{-1}(u)\right) \, du.\end{aligned}$$

By definition of the optimal hazard rate λ^*, the quantity $\sum_i w_i (F_{Q_i}^{-1}(u) - F_{Q^*}^{-1}(u))$ equals zero. From this it follows that

$$\Delta_Q = \sum_i w_i W_2^2(Q_i, Q^*) + \int_0^1 \left(F_{Q^*}^{-1}(u) - F_Q^{-1}(u)\right)^2 \, du$$

$$\geqslant \sum_i w_i W_2^2(\mathbb{Q}_i, \mathbb{Q}^*).$$

From the arbitrariness of $\mathbb{Q}$ and from Theorem 3, equality (105) follows. □

For a given group of CDSs one can calculate, for each maturity, the optimal hazard rate λ^* minimizing the squared error in the Wasserstein square distance (99). The corresponding "optimal" par credit spreads, per maturity, can then be implied. This can be done, for instance, by again employing the credit triangle relationship (98), and by assuming a recovery rate equal either to a prespecified value or calculated as the average (e.g., arithmetic or harmonic) of the recovery rates of the CDSs belonging to the same bucket.

Assume now, for simplicity, that for all the CDSs in a given bucket the recovery rate is constant.[11] If we define the proxy recovery rate as the average (arithmetic or harmonic) of the observed recovery rates and compute the proxy par spread by inverting (98), it then still follows that this CDS par spread equals the harmonic mean of the CDS spreads of the single contracts in the bucket. We recall that the harmonic mean of a set of (non-zero) numbers is at most equal to its arithmetic counterpart due to Jensen's inequality. While the arithmetic mean equally averages all the data points, which makes it give less importance to observations with very small magnitude, by considering the harmonic mean higher weights are given to those observations. Hence, it then follows that, under the constant recovery rate assumption, the methodology proposed here is less conservative than simply averaging the observed CDS quotes, as well as less (more) sensitive to large (small) outliers.

From a computational perspective, the methodology highlighted in the present article is simple. As Theorem 3 shows, calculating the optimal hazard rates only requires the computation of a weighted harmonic mean. This also allows to interpret the proxy CDS curves in terms of Wasserstein barycenters in virtue of Theorem 4. However, we provide the following remark concerning the interpretation of the technique we have used in this article. Calculating the proxy spread by simply averaging the relevant CDS quotes is an elementary methodology, which very simple to apply and very intuitive. However, calculating the average of the observed spreads is a procedure that is only based on the magnitude of the observations and that does not take into account how the implicit default distributions implied from the different observations interact with each other. On the other hand, the approach we have proposed here bases the construction of the proxy quotes on the implicit distributions arising from the different quotes. This is done by means of finding a new distribution that is as close as possible to the observed ones from an optimal transportation perspective. This gives a fresh view to credit curve proxy methodologies by means of attempting to tackle the

11 In principle, recovery rates can vary. However, in many cases fixed values for recovery rates are used; see Das and Hanouna (2009).

problem not from a data perspective but, instead, by involving metrics between probability distributions.

5.5 SUITABILITY OF OTHER METRICS BETWEEN PROBABILITY DISTRIBUTIONS TO PROXY CDS CURVES

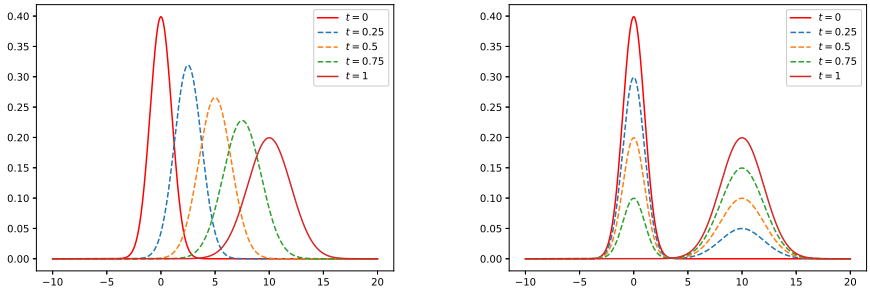


Figure 35: Example of geodesics between probability density functions. In particular, panel (a) represents the geodesic in Wasserstein square distance between a normal density function with mean parameter $\mu_0 = 0$ and standard deviation $\sigma_0 = 1$ (denoted with $n_0(\cdot)$) and a normal density function with mean parameter $\mu_1 = 10$ and standard deviation $\sigma_1 = 2$ (denoted with $n_1(\cdot)$). For each $t \in [0,1]$ the geodesic can be explicitly calculated (see Malagó et al. (2018)), and its density (denoted with $n_t(\cdot)$) corresponds to that of a normal distribution with mean $\mu_t := (1-t)\mu_0 + t\mu_1$ and standard deviation $\sigma_t := (1-t)\sigma_0 + t\sigma_1$. Note how the geodesic is shape-preserving. On the other hand, the geodesic in the Euclidean space is illustrated in panel (b) by means of calculating, for $t \in [0,1]$, the density function $n_t(\cdot) := (1-t)n_0(\cdot) + tn_1(\cdot)$ (i.e., the segment joining $n_0(\cdot)$ and $n_1(\cdot)$), which is clearly not shape-preserving.

In Section 5.3.1 we have recalled how Wasserstein distances have become popular due to some intuitive properties they guarantee. In particular, geodesics defined by means of Wasserstein distances are shape-preserving, as Figure 35 illustrates.

Geodesics are shape-preserving due to the fact that, in some sense, the Wasserstein distance accounts for the form of the underlying distribution. For instance, if one considers three continuous uniform distributions $\mathbb{U}_1$, $\mathbb{U}_2$ and $\mathbb{U}_3$ on $[0,1]$, $[1,2]$ and $[100,101]$, respectively, then by (95) it results that $W_2(\mathbb{U}_1,\mathbb{U}_2) = 1$ while $W_2(\mathbb{U}_1,\mathbb{U}_3) = 10$. Therefore, distributions with "similar" densities are "close", according to the (square) Wasserstein distance, while this is not the case if densities are "far" from each other. Note that other metrics between prob-

ability distributions would not be sensitive to "geometrical" similarities. For instance, we recall that the *total variation distance* (see (Villani, 2009, Ch. 6)) between two (absolutely continuous) probability measures $\mathbb{P}$ and $\mathbb{Q}$ with densities $f_{\mathbb{P}}(\cdot)$ and $f_{\mathbb{Q}}(\cdot)$, respectively, denoted as $TV(\mathbb{P},\mathbb{Q})$, can be computed as $\frac{1}{2}\int_{-\infty}^{\infty}|f_{\mathbb{P}}(x)-f_{\mathbb{Q}}(x)|\,dx$. Additionally, the *(square) Hellinger distance* (see (Le Cam, 1986, Ch. 4)) between $\mathbb{P}$ and $\mathbb{Q}$, denoted as $H_2(\mathbb{P},\mathbb{Q})$, can be calculated as $\int_{-\infty}^{\infty}(\sqrt{f_{\mathbb{P}}(x)}-\sqrt{f_{\mathbb{Q}}(x)})^2\,dx$. It then results that $TV(\mathbb{U}_1,\mathbb{U}_2)=TV(\mathbb{U}_1,\mathbb{U}_3)=1$ and, similarly, that $H_2(\mathbb{U}_1,\mathbb{U}_2)=H_2(\mathbb{U}_1,\mathbb{U}_3)=1$. The intuitive fact that Wasserstein distances take into account the shape of the probability distributions makes them attractive from the geometrical perspective as well.

As far as exponential distributions are concerned, which are the ones used in this article, we observe that given two exponential distributions with parameters λ and μ (assume in this case $\lambda<\mu$), their total variation distance equals $\left(\frac{\mu}{\lambda}\right)^{-\frac{\mu}{\mu-\lambda}}-\left(\frac{\mu}{\lambda}\right)^{-\frac{\lambda}{\mu-\lambda}}$. Further, their (square) Hellinger distance equals $\frac{(\sqrt{\mu}-\sqrt{\lambda})^2}{\lambda+\mu}$. Thus, the expressions just provided would make the problem of finding the optimal hazard rate with respect to the total variation and (square) Hellinger distances more complicated. We also recall that the *Kullback-Leibler divergence* Kullback and Leibler (1951) between two (absolutely continuous) probability distributions $\mathbb{P}$ and $\mathbb{Q}$ with densities $f_{\mathbb{P}}(\cdot)$ and $f_{\mathbb{Q}}(\cdot)$, respectively, is defined as $KL(\mathbb{P},\mathbb{Q})=\int_{-\infty}^{\infty}f_{\mathbb{P}}(x)\ln\left(\frac{f_{\mathbb{P}}(x)}{f_{\mathbb{Q}}(x)}\right)dx$. Thus, the Kullback-Leibler divergence is not symmetric. Given $\mathbb{P}$ and $\mathbb{Q}$ exponentially distributed as before (i.e., with rate parameters equal to λ and μ, respectively), we obtain that $KL(\mathbb{P},\mathbb{Q})=-\log\frac{\mu}{\lambda}-1+\frac{\mu}{\lambda}$. If in (99) $W_2^2(\mathbb{Q}_i,\mathbb{Q}^*)$ were to be substituted with $KL(\mathbb{Q}_i,\mathbb{Q}^*)$ for every i, then the optimal hazard rate would still coincide with the weighted harmonic average (100). On the other hand, if $W_2^2(\mathbb{Q}_i,\mathbb{Q}^*)$ were to be substituted with $KL(\mathbb{Q}^*,\mathbb{Q}_i)$, then the optimal hazard rate would coincide with the weighted arithmetic average of the λ_i's. As the λ_i's can be interpreted as rates, then the natural way of averaging them is that of considering their (weighted) harmonic mean rather than their (weighted) arithmetic one.

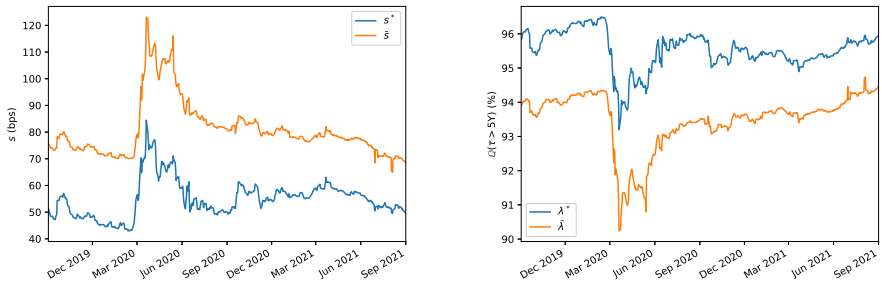
The examples and considerations provided above illustrate how Wasserstein distances are very suitable tools to measure, in an intuitive manner, distances between probability distributions. This makes them suitable tools to be employed for the purpose of computing distances between default probability distributions as proposed in this article, as opposed to some other well-known metrics.

5.6 EXAMPLES

In this section we apply the methodology described in this article to two different datasets. We compare the results to those obtained by applying the intersection methodology when the proxy CDS par spread for a given bucket is

defined as the arithmetic average of the CDS spreads therein allocated. For the sake of clarity and brevity, in this section we will often refer to the methodology outlined in this article as the *Wasserstein methodology*, while we will refer to the intuitive procedure of arithmetically averaging CDS spreads as the *average methodology*. In both examples considered here we take into account the time period spanning from September 2019 to September 2021. As far as the tenor of the CDS contracts is concerned, we focus our analysis on the 5Y tenor, which is in general the most liquid.

As a first example we take into account the bucket where the region is "North America", the sector "financials", the rating "A", and the seniority "senior unsecured debt".

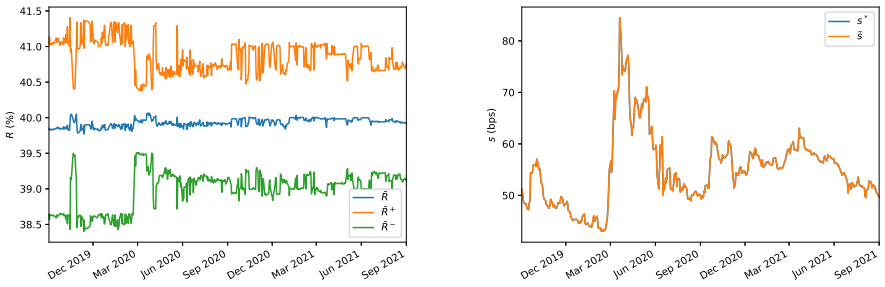


(a) (b)

Figure 36: 5Y CDS proxy par spreads and survival probabilities for the bucket with region "North America", sector "financials", rating "A" and seniority "senior unsecured debt"; see panels (a) and (b), respectively. The proxy CDS par spreads depicted in panel (a) labelled with s^* ($\bar{s}$) have been computed using the Wasserstein (average) methodology. Similarly, the 5Y proxy survival probabilities of panel (b) labelled with λ^* ($\bar{\lambda}$) are based on the hazard rate calculated with the Wasserstein (average) methodology.

In this bucket, for the time period and the tenor considered, the number of entities oscillates between a minimum of 55 and a maximum of 62. Figure 36a displays the 5Y proxy CDS spreads computed using the Wasserstein methodology and those computed using the average one. It is clear that calculating proxy CDS spreads by means of Wasserstein barycentres produces results that are consistently below their average counterparts. In particular, we observe differences of at least 15-20bps and, during one of the peaks of the Covid-19 pandemic (i.e., in the neighbourhood of the second quarter of 2020), the magnitude of the discrepancies observed reaches values higher than 40bps. The differences observed here are therefore reflected in the associated implied survival probabilities. In fact, as Figure 36b shows, arithmetically averaging the CDS spreads in the bucket largely underestimates (overestimates) the proxy survival (default) probabilities.

In particular, Figure 36b highlights how, given a 5Y horizon, survival probabilities implied from the averaged CDS quotes differ by at least 1.5%-2% from those implied using the Wasserstein methodology, with peaks between 3% and 3.5% during the Covid-19 period mentioned above.



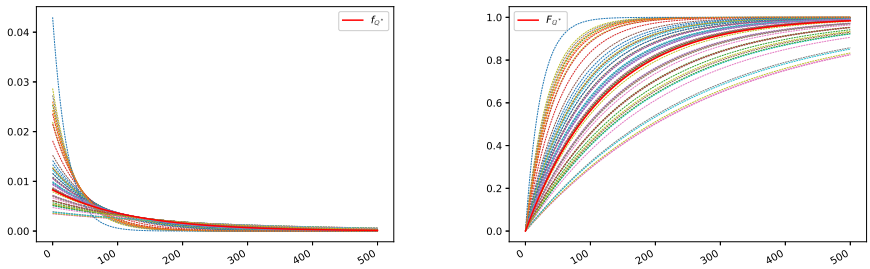
(a) (b)

Figure 37: Panel (a) illustrates the 5Y CDS average recovery rates for the bucket with region "North America", sector "financials", rating "A" and seniority "senior unsecured debt". In particular, $\bar{R}$ denotes the arithmetic average of the recovery rates of the CDSs in the chosen bucket, while $\bar{R}^+$ ($\bar{R}^-$) denotes the arithmetic average of the recovery rates of the CDSs in the chosen bucket plus (minus) their standard deviation. Panel (b) displays the 5Y proxy survival probabilities for this bucket calculated with the Wasserstein methodology. The label s* ($\tilde{s}$) indicates that the proxy recovery rate has been calculated as the arithmetic (harmonic) average of the recovery rates of the CDSs in the bucket considered.

The reasons why simply averaging CDS spreads overestimates their proxy counterparts (and therefore, overestimates default probabilities) is due to the fact that, often, recovery rates for CDSs are set equal (or close) to some standard values (see Section 5.4 and Das and Hanouna (2009)). Therefore, CDSs with similar characteristics in terms of sector, region, rating, etc. often share similar recovery rates. In particular, frequently recovery rates are set equal to (or neighbouring) 40%.

This is illustrated by Figure 37a, which shows that the average recovery rate for this bucket often lies between 39.5% and 40.5%, and that it has a small standard deviation. In Section 5.4 we have observed that, if the recovery rates in a given bucket are all equal, then by inverting (98) the resulting proxy CDS par spread equals the harmonic mean of the CDS spreads in the bucket (if all recovery rates are equal and the proxy recovery rate equals their average). Due to the small standard deviation of the recovery rates observed, it results that all the recovery rates for the CDSs in the bucket are roughly equal to each other. Therefore, by inverting (98), the proxy CDS spread for this bucket is approximately equal to

the harmonic average of the observed CDS spreads, which is smaller than its arithmetic counterpart. Note that, given the dataset considered, if we compute the proxy recovery rate as the harmonic mean of the observed recovery rates, then results do not change significantly as recovery rates are stable through time; see Figure 37b. This shows that the methodology is robust with respect to the definition of the proxy recovery rate.



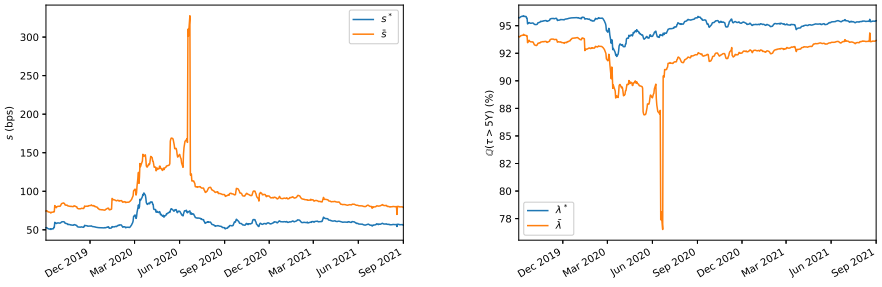
(a) (b)

Figure 38: Density and distribution functions implied from the 5Y CDSs with region "North America", sector "financials", rating "A" and seniority "senior unsecured debt", as well as their proxy counterparts (thick red lines) computed using the Wasserstein methodology as of 1-Sep-2021; see panels (a) and (b), respectively.

For completeness, we also illustrate how the density and distribution functions for the default times as of 1-Sep-2021 look like in this specific case, and compare them with their proxy counterparts computed using the Wasserstein methodology; see Figures 38a and 38b, respectively.

We now consider a second example where the selected CDS bucket is identified by the "Europe" region, "industrials" sector, "BBB" rating and "senior unsecured debt" seniority. The number of constituents for this bucket varies, throughout the period taken into account, between a minimum of 27 and a maximum of 38. We have chosen this bucket as, during the time period considered for the analysis, the dataset shows some outliers. Therefore, this allows to test the robustness of the proposed methodology in these circumstances compared to the standard intersection methodology. As Figure 39a illustrates, the proxy CDS quotes computed by means of Wasserstein barycentres are, as in the previous example, consistently below their counterparts computed using arithmetic averages. Differences are, for the majority of the dates considered, of the order of magnitude between 20 and 25bps. This is clearly reflected by the related implied survival probabilities, where those implied from arithmetically averaged CDS spreads are underestimated; see Figure 39b. We observe how, between June and September 2020, the dataset considered displays a peak in the proxy

spread (and survival probabilities) in the case proxy CDS spreads are calculated using the average methodology due to some outliers (some other minor peaks are also present in the time frame we consider in this example, but with smaller magnitude).

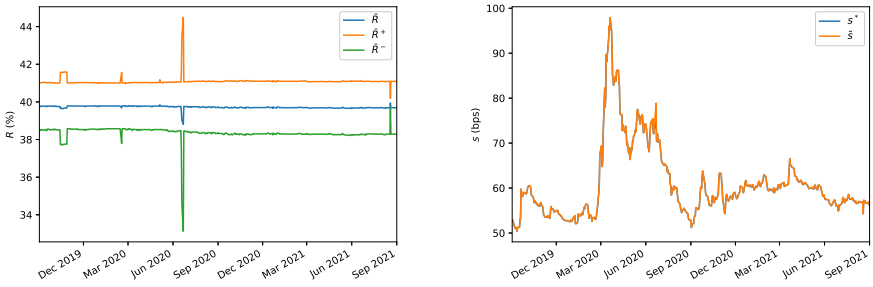


(a) (b)

Figure 39: 5Y CDS proxy par spreads and survival probabilities for the bucket with region "Europe", sector "industrials", rating "BBB" and seniority "senior unsecured debt"; see panels (a) and (b), respectively. The proxy CDS par spreads depicted in panel (a) labelled with s^* ($\bar{s}$) have been computed using the Wasserstein (average) methodology. Similarly, the 5Y proxy survival probabilities of panel (b) labelled with λ^* ($\bar{\lambda}$) are based on the hazard rate calculated with the Wasserstein (average) methodology.

We clearly see, as Figures 39a and 39b illustrate, that computing proxy CDS spreads by means of Wasserstein barycentres, results are much more stable. The peak observed in the case the average methodology is used results in a difference in the proxy CDS spread (5Y survival probability) of the order of more than 250bps (15%) compared to the Wasserstein methodology case.

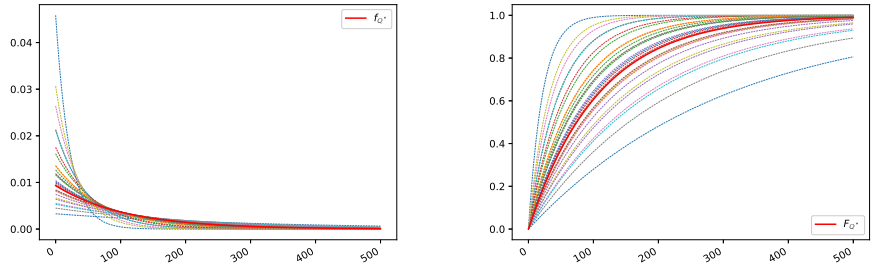
Also in this case we observe that, as expected, for the CDSs in the chosen bucket recovery rates do not vary much, with a few exceptions; see Figure 40a. Therefore, also in these circumstances the same rationale as outlined in the former example can be followed to explain why proxy CDS spreads calculated with the average methodology (and corresponding default probabilities) overestimate their counterparties computed following the Wasserstein methodology. We also show, see Figure 40b, that the Wasserstein methodology is robust with respect to the averaging type used to compute the proxy recovery rate (i.e., arithmetic or harmonic).



(a) (b)

Figure 40: Panel (a) illustrates the 5Y CDS average recovery rates for the bucket with region "Europe", sector "industrials", rating "BBB" and seniority "senior unsecured debt". In particular, $\bar{R}$ denotes the arithmetic average of the recovery rates of the CDSs in the chosen bucket, while $\bar{R}^+$ ($\bar{R}^-$) denotes the arithmetic average of the recovery rates of the CDSs in the chosen bucket plus (minus) their standard deviation. Panel (b) displays the 5Y proxy survival probabilities for this bucket calculated with the Wasserstein methodology. The label s* ($\tilde{s}$) indicates that the proxy recovery rate has been calculated as the arithmetic (harmonic) average of the recovery rates of the CDSs in the bucket considered.

Again, for completeness we also illustrate how the density and distribution functions for the default times as of 1-Sep-2021 look like in this second example as well, and how they compare with respect to their proxy counterparts computed by means of the Wasserstein methodology; see Figures 41a and 41b, respectively.



(a) (b)

Figure 41: Density and distribution functions implied from the 5Y CDSs with region "Europe", sector "industrials", rating "BBB" and seniority "senior unsecured debt", as well as their proxy counterparts computed using the Wasserstein methodology as of 1-Sep-2021 (thick red lines); see panels (a) and (b), respectively.

We have provided a comparison between a common way to compute proxy CDS spreads and related credit curves (i.e., intersection approach with average methodology) and those computed following the approach proposed based on Wasserstein distances. We have shown how, by considering real market scenarios, the differences in default probabilities produced using the two methodologies are non-marginal. In particular, given the examples taken into account, we have shown how proxying CDS spreads by means of the average methodology consistently overestimates the resulting implied survival probabilities. This means that simply arithmetically averaging CDS spreads can produce unjustifiably conservative survival probabilities. This is due to the fact that, despite simple and intuitive, this approach lacks of adequate theoretical foundations. As proxy survival probabilities are often used by financial institution to compute Credit Valuation Adjustment (CVA), using the intersection approach with the average methodology given the datasets considered here would result in excessively conservative CVA charges compared to those calculated using proxy survival probabilities computed using the Wasserstein methodology. Thus, under market circumstances for which the examples taken into account here can be considered representative, the Wasserstein methodology would result in more aggressive CVA charges than those computed when the average methodology is used.

5.7 CONCLUSION

In this article we have investigated an alternative way of estimating proxy CDS curves starting from the intersection methodology. In particular, instead of simply averaging the different CDS quotes belonging to the given buckets, we have constructed, under some simple assumptions, a proxy CDS curve by means of an optimal transportation problem. That is, after having specified a given functional form for the CDS proxy curve to be estimated, we have provided simple formulae that allow to calculate the proxy hazard rates (and, thus, proxy par spreads) by simply minimizing the Wasserstein square distances between the proxy curve and the other curves belonging to the chosen bucket. The approach we have adopted is simple and relies on basic assumptions concerning CDSs that are often employed in practice. It also provides a simple interpretation of the proxy CDS default distribution as the actual Wasserstein barycenter of the observed default probability distributions.

6 ON WASSERSTEIN DISTANCES, BARYCENTERS, AND THE CROSS-SECTION METHODOLOGY FOR PROXY CREDIT CURVES

6.1 ABSTRACT

The credit default swap (CDS) market plays an important role for financial institutions. This is not only for their trading activities, but also as it provides a source of information to extract default probabilities to be used for (counterparty) credit risk purposes, as for instance in credit valuation adjustment calculations. Nonetheless, the number of entities for which liquid single-name CDSs are traded is of the order of a few thousands. This requires financial institutions to employ proxy methodologies to estimate the credit risk they face when trading with counterparties for which no (liquid) CDSs are available in the market. In this article we propose and compare different approaches to take into account counterparty-specific information in terms of rating, region, sector, etc. at cross-sectional level to strip risk-neutral default probabilities from CDSs. This is achieved by taking into account the intrinsic probabilistic information characterizing each CDS by means of suitably-defined Wasserstein distances and barycenters. The results suggest that default probabilities are likely to be overestimated if the construction of the proxy credit curves overlooks the probability structure underlying the CDS market, potentially resulting in a too conservative counterparty credit risk pricing framework.

6.2 INTRODUCTION

In the article at hand we develop a novel approach for constructing proxy credit curves from credit default swaps (CDSs) starting from the *cross-section methodology*. In particular, we investigate how to embed the concepts of *Wasserstein distance* and *Wasserstein barycenter* between implied CDS probability distributions in a cross-sectional framework.

Credit risk and counterparty credit risk are key aspects financial institutions. Quantifying (and, thus, risk-managing) these risks involves procedures aiming to estimate the credit worthiness of given counterparties, which necessitate the

calculation of their default probabilities within the relevant portfolio-specific horizons. For this reason techniques to compute the likelihood of a default given the observable quantities available in the financial markets need to be in place. A common procedure to estimate risk-neutral default probabilities leverages on CDS quotes: from these observations it is possible to compute default probabilities in a simple and nowadays practically standardized manner White (2013). These probabilities are particularly important, amongst other areas, for valuation adjustments calculations; see for instance Brigo and Pallavicini (2014) and Abbas-Turki et al. (2018), amongst others. However, the liquidity of the CDS market is often below-par compared to that of other markets Junge and Trolle (2015), which makes the estimation of default risk starting from these instruments not a straightforward task. This is mainly due to a couple of key factors. On the one hand, at global level the number of counterparties for which CDS quotes are available is somewhat limited, i.e., in the order of a couple of thousands; see (Green, 2015, Ch. 4.6) and (Ruiz, 2015, Ch. 7.4). On the other hand, an important distinction needs to be made between liquid and illiquid CDS names: in the former case the procedure to compute default risk for counterparties with liquid CDSs quoted in the market is straightforward, while in the latter case it might happen that either quoted CDSs for the given counterparties are highly illiquid (and, as a consequence, not reliable) or, even worse, no CDSs for the relevant names are traded in the market at all. This is a very common situation faced by several financial institutions (especially banks) which do business with small retail companies and enterprises, for which in the vast majority of cases no CDSs are traded. Therefore, under these circumstances, procedures and methodologies need to be set up to indirectly calculate the default probabilities for these companies.

There are procedures in place to extract the aforementioned default probabilities in the case the relevant CDS market data is unreliable or not directly observable. Other than mapping the illiquid counterparties to specific single-name or index CDSs (see (Green, 2015, Ch. 4.6.1) and (Green, 2015, Ch. 4.6.2), respectively), the easiest way of doing so is by means of the so-called *intersection methodology* EBA (2013) (see also Chourdakis et al. (2013a) and Sourabh et al. (2018)). This involves proxying the unobservable quotes for a counterparty with the average of the available CDS quotes of companies with the same characteristics (in terms of rating, region, sector, etc.). This approach has been extended in Michielon et al. (2022b), where the intersection methodology has been enhanced by computing the Wasserstein barycenter of the probability distributions implied from the CDSs belonging to the same bucket (i.e., with the same rating, region, sector, etc.). Nonetheless, the intersection methodology has two drawback. That is, in some cases the number of liquid CDS quotes in a given bucket can be low or even zero, which might make some proxy curves unreliable or, in the worse-case scenario, impossible to compute. Further, the intersection methodology often seems to fail to guarantee, given buckets with different ratings but

the same region, sector, etc., that the higher the rating the lower the likelihood of default is. To remedy these shortcomings, Chourdakis et al. (2013a) propose a new approach, also known as *cross-section methodology*, which assumes that a given CDS spread can be split as the product of multiple factors (either global or depending on rating, region, sector, etc.). These factors can be then optimally calibrated in a least-squares sense by means of a linear regression procedure with categorical variables. From a data perspective, the approach of Chourdakis et al. (2013a) has been further extended in Sourabh et al. (2018), where it is illustrated how to include additional information (i.e., equity data such as returns and volatilities) in the regression equations to further enhance the explicative power of the technique.

In this article we propose a hybrid calibration procedure to proxy CDS spreads. We propose different variations for this methodology and highlight their advantages and drawbacks by comparing them to the cross-section approach. We show how, overall, the best amongst the approaches proposed in this article shares the main qualitative features of the cross-section methodology and, at the same time, tends to produce lower default probabilities. This is in line with the results outlined in Michielon et al. (2022b) as far as the intersection methodology is concerned: ignoring the structure of the probability distributions implied from CDSs might result in overly-conservative counterparty credit risk pricing.

The main contribution of this paper is that of developing a new framework to construct proxy credit curves starting from CDS quotes. In particular, the methodology proposed allows to build credit curves by means of the Wasserstein barycenters of the (probability distributions implied from the) available CDS buckets, as well as the cross-sectional information between them. The resulting approach allows to explicitly take into account, in a sensible and intuitive manner, the implicit probabilistic information hidden in CDS quotes by means of Wasserstein distances, and to easily "interpolate" this information using a simple regression step based on categorical variables. The article at hand clearly illustrates how Wasserstein distances can be considered suitable tools for modelling credit curves and, to the best of our knowledge, it is one of the few exploring the possibility of modelling credit risk using distances defined on probability spaces.

The paper at hand is organized as follows. Section 6.3 illustrates the methodologies currently available in the literature for proxying CDS spreads from liquid market quotes. In Section 6.4 different novel approaches to incorporate Wasserstein distances and barycenters within the cross-sectional framework are proposed. Further, Section 6.5 compares the different methodologies given a real market data set, while Section 6.6 concludes. In Appendix 6.A a remark concerning the possibility of applying the cross-section methodology at hazard rate level is provided.

6.3 LITERATURE REVIEW

In this section, we provide an overview of the main methodologies currently available in the literature to build proxy credit curves from CDS data. This is necessary to make the article self-contained, as the methodologies proposed in Section 6.4.1 are based on the main concepts that are outlined in the current section. We highlight that, at the time of writing, the number of articles available in the literature concerning proxying credit curves from CDSs is quite limited, as only two main approaches are available, i.e., the intersection and the cross-section methodologies. Section 6.3.1 focuses on the former, while Section 6.3.2 concerns with the latter.[1]

6.3.1 The intersection methodology

A way to construct proxy CDS curves that does not involve simply mapping missing CDS data to that of other companies or indices is given by the so-called intersection methodology EBA (2013). This approach suggests to bucket liquid CDSs according to their rating, region, sector, etc. By averaging the CDS quotes in each bucket, one can then define proxy CDS data per bucket and maturity; see also Chourdakis et al. (2013a) and Sourabh et al. (2018). This methodology is very intuitive and easy to implement. Also, should one need to make the approach more granular, then this could be easily achieved by considering a finer bucketing procedure, despite this might results in more buckets with few or no observations. Despite the simplicity of the approach, the intersection methodology exhibits two main drawbacks, as highlighted in Section 6.4.1.

The intersection methodology has been investigated from a different angle in Michielon et al. (2022b). In particular, therein the technique is revised from a probabilistic perspective. That is, instead of simply arithmetically-averaging the market quotes within a given bucket, proxy CDS quotes for a given group are constructed by finding the (unique) hazard rate which defines the Wasserstein barycenter of the probability distributions implied from the considered CDSs. In particular, in Michielon et al. (2022b) it is assumed that, for a fixed maturity, default distributions implied from CDSs are of the form

$$Q(\tau \leqslant t) \coloneqq 1 - e^{-\lambda t}, \tag{107}$$

1 For completeness, we highlight that, in the sections that follows, default probabilities are always calculated under risk-neutral settings and within the framework of structural credit models, to which the specifications available in White (2013) apply. For methodologies to compare actual and risk-neutral default probabilities the interested reader can refer, e.g., to and Heynderickx et al. (2016) and Zou and Li (2022), while for alternative approaches to calculate hazard rates to Lee and Kuo (2015), amongst others.

where τ denotes the (random) default time of a given entity, and where $\lambda \in (0, +\infty)$ denotes the (constant) hazard rate. A simple manner to estimate λ is given by the well-known *credit triangle* relationship White (2013), which reads

$$\lambda \approx \frac{s}{1-R}, \tag{108}$$

where s refers to the par spread of the CDS taken into account, while R to its recovery rate.[2]

We denote with $W_2(\mathbb{P}, \mathbb{Q})$ the square Wasserstein distance between the two absolutely continuous probability distributions on $\mathbb{R}$, i.e.,

$$W_2(\mathbb{P}, \mathbb{Q}) := \left(\inf_{J \in \mathcal{J}(\mathbb{P}, \mathbb{Q})} \int_{\mathbb{R}^2} \|x - y\|^2 \, J(dx \times dy) \right)^{\frac{1}{2}},$$

with $\mathcal{J}(\mathbb{P}, \mathbb{Q})$ denoting the set containing all the joint distributions which have $\mathbb{P}$ and $\mathbb{Q}$ as marginals, and where $\|\cdot\|$ stands for the Euclidean norm. For more details concerning Wasserstein distances refer, amongst others, to (Villani, 2003, Ch. 7), (Villani, 2009, Ch. 6), or to (Panaretos and Zemel, 2020, Ch. 1 and 2).

Assume now that N CDSs with the same maturity and with spreads denoted as $s_1, \ldots, s_N$ are bucketed together. The Wasserstein barycenter of the N CDS-implied exponential distributions $\mathbb{Q}_1, \ldots, \mathbb{Q}_N$ is any probability distribution $\tilde{\mathbb{Q}}$ solution of

$$\inf \left\{ \sum_{i=1}^{N} W_2^2(\mathbb{Q}_i, \mathbb{Q}) : \mathbb{Q} \text{ has finite second moment} \right\}. \tag{109}$$

It can be proven, see Michielon et al. (2022b), that the solution of (109) is unique. In particular, it results that the unique solution of (109) is still of the form (107), and that its hazard rate can be computed analytically by finding the minimum of

$$\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} - \frac{1}{\tilde{\lambda}} \right)^2 \tag{110}$$

with respect to $\tilde{\lambda}$. This hazard rate coincides with the harmonic mean of the implied hazard rates $\lambda_1, \ldots, \lambda_N$ associated with $\mathbb{Q}_1, \ldots, \mathbb{Q}_N$; see Michielon et al. (2022b). For completeness, observe that

$$W_2^2(\mathbb{Q}_i, \tilde{\mathbb{Q}}) = 2 \left(\frac{1}{\lambda_i} - \frac{1}{\tilde{\lambda}} \right)^2. \tag{111}$$

2 Observe that one could alternatively imply hazard rates numerically, but the credit triangle relationship (108) provides an easy approximation formula that is very useful for comparison purposes and for the explanation of the results obtained. Therefore, this approach is followed throughout this paper (see also Michielon et al. (2022b)).

Assume that recovery rates within the same bucket are close to each other (usually this is the case as CDSs belonging to the same bucket share the same characteristics in terms of rating, region, sector, etc., and, thus, also similar recovery rates), and that the recovery rate for the proxy CDS is defined as their (harmonic) average. Then, again by means of (108), it results that the proxy CDS spread corresponding to $\tilde{\lambda}$ is close to the harmonic mean of the CDS spreads $s_1, \dots, s_N$. This leads to CDS proxy spreads that are smaller than their counterparts calculated by simply averaging the available CDS quotes within the bucket considered, as illustrated in Michielon et al. (2022b), as the harmonic mean of $s_1, \dots, s_N$ is bounded from above by their arithmetic counterpart. As default probabilities are strictly increasing with respect to CDS spreads, the intersection methodology with Wasserstein barycenter produces in general lower default probabilities than its standard counterpart.

We highlight that the idea outlined in Michielon et al. (2022b) was that of investigating how proxy credit curves (therein within the intersection approach) can take into account the distance between the probability distributions implied by the different CDSs. This is because just averaging CDS quotes does not explicitly accounts for the probabilistic information between default times. In particular, in order to do so, it is advisable to define a metric between probability distributions. One convenient (and nowadays very popular) choice due to its properties is the Wasserstein distance as, e.g., it produces geodesics which are shape preserving and, further, it takes into account the "shape" of the probability distributions considered. These properties are not in general satisfied by other metrics, and in Michielon et al. (2022b), see Section 4 therein, these concepts are described and examples are provided to endorse the choice made.

Further, we now provide some considerations to concerning the intuitive interpretation of the Wasserstein barycenter as highlighted in Michielon et al. (2022b). To do so, we assume to have a default time τ exponentially distributed under $\mathbb{Q}$ with rate parameter λ. It results, given a time interval δt and assuming λ "small", that

$$\frac{\mathbb{Q}(t \leqslant \tau \leqslant t + \delta t)}{\delta t} \left(= \frac{\mathbb{Q}(\tau \leqslant t + \delta t) - \mathbb{Q}(\tau \leqslant t)}{\delta t} \right) = \frac{\int_t^{t+\delta t} \lambda e^{-\lambda s} \, ds}{\delta t} \approx \lambda. \quad (112)$$

Note that the numerator in the first member of (112) measures by how much the default probability changes during the time interval δt. Therefore, (112) measures a change in probability during the time frame δt, i.e., λ can be intuitively interpreted as the "default velocity".

We take into account a very simple framework where we have only two CDSs, denoted as CDS_1 and CDS_2, respectively. We also assume that both CDS_1 and CDS_2 have similar specifications, i.e., that they are linked to entities with the same rating, region, sector, etc., so that it makes sense considering CDS_1 and CDS_2 both in the same bucket. We denote with s_1 the CDS spread for CDS_1, and with s_2 the CDS spread for CDS_2 (we assume $s_1 \neq s_2$ to avoid trivial cases).

We assume that both CDS_1 and CDS_2 have the same recovery rate R. Further, we approximate the hazard rates corresponding to CDS_1 and CDS_2, denoted with λ_1 and λ_2, respectively, by means of the credit triangle relationship (108). That is, $\lambda_1 \approx \frac{s_1}{1-R}$, and $\lambda_2 \approx \frac{s_2}{1-R}$. We denote with τ_1 and τ_2 the default times linked to CDS_1 and CDS_2, respectively.

We consider a fixed time $\bar{t}$ and a sufficiently small amount $\delta Q \in (0,1)$ such that $Q(\tau_1 \leqslant \bar{t} + \delta t_1) = Q(\tau_1 \leqslant \bar{t}) + \delta Q$ and $Q(\tau_2 \leqslant \bar{t} + \delta t_2) = Q(\tau_2 \leqslant \bar{t}) + \delta Q$, for $\delta t_1 > 0$ and $\delta t_2 > 0$. That is, during the time intervals δt_1 and δt_2 the probabilities of default of τ_1 and τ_2 both increase by the same amount δQ, i.e., δQ represents the change in probability for each default time during the two time frames considered. Therefore, due to (112), it results that $\delta t_1 \approx \frac{\delta Q}{\lambda_1}$ and $\delta t_2 \approx \frac{\delta Q}{\lambda_2}$. Thus, the average default velocity of τ_1 and τ_2 results being equal to $\frac{\text{total decrease in probability}}{\text{total time}} = \frac{2\delta Q}{\frac{\delta Q}{\lambda_1} + \frac{\delta Q}{\lambda_2}} = \frac{2\lambda_1\lambda_2}{\lambda_1+\lambda_2} =: \hat{h}$, which corresponds to the harmonic mean of the default velocities λ_1 and λ_2. Therefore, the proxy hazard rate relevant for the bucket, which is calculated as the harmonic mean of the relevant hazard rates, can be naturally interpreted as "average default velocity". We also note, see Theorems 1 and 2 in Michielon et al. (2022b), that the Wasserstein barycenter between exponential distributions is still exponentially distributed, with rate parameter equal to the harmonic mean of the relevant rate parameters. This makes Wasserstein distances also convenient from a computational perspective.

6.3.2 The cross-section methodology

The cross-section methodology, introduced in Chourdakis et al. (2013a), is based on the idea that the logarithm of a CDS spread can be decomposed as the product of one global and different bucket-dependent factors. These factors can then be estimated in a least-squares sense by taking into account all the CDS spreads available in the market simultaneously; see also (Gregory, 2020, Ch. 12.4.4). By doing so, the approach by Chourdakis et al. (2013a) allows to determine CDS spreads also for buckets for which no data is directly observable in the market. The cross-section technique has been enhanced in Sourabh et al. (2018), where equity data is used simultaneously to credit data to take into account even more information in the least-squares optimization procedure. We now describe the cross-section methodology more in detail. For consistency with the rest of the paper, see Section 6.4, in our exposition we follow a similar approach to that outlined in Sourabh et al. (2018), but without taking into account equity data.

Assume that, given a CDS dataset for a fixed maturity, we are able to distinguish between N_{Rating} possible ratings, N_{Region} possible regions, N_{Sector} possible sectors, and $N_{\text{Seniority}}$ possible seniorities. This identifies a total of $N_{\text{Bucket}} :=$

$N_{\text{Rating}} \cdot N_{\text{Region}} \cdot N_{\text{Sector}} \cdot N_{\text{Seniority}}$ possible rating-region-sector-seniority combinations. Having defined an enumeration for the ratings, one for the regions, one for the sectors and one for the seniorities, we fix a rating, a region, a sector and a seniority to be used as a baseline. We are thus left with $N_{\text{Bucket}} - 1$ possible combinations different from the base one. Given a total of K CDS quotes available for a fixed maturity, denoted as $s_1, \dots, s_K$, it is assumed that, for every $i \in \{1, \dots, K\}$, the following relationship holds:

$$\begin{aligned} \ln(s_i) = x_0^{\text{Base}} + \sum_{j=1}^{N_{\text{Rating}}-1} x_j^{\text{Rating}} \mathbb{1}_j(i) + \sum_{k=1}^{N_{\text{Region}}-1} x_k^{\text{Region}} \mathbb{1}_k(i) \\ + \sum_{l=1}^{N_{\text{Sector}}-1} x_l^{\text{Sector}} \mathbb{1}_l(i) + \sum_{m=1}^{N_{\text{Seniority}}-1} x_m^{\text{Seniority}} \mathbb{1}_m(i). \end{aligned} \tag{113}$$

In (113), $\mathbb{1}_j(i)$ denotes the indicator function that equals one if the rating of the i^{th} CDS coincides with the j^{th} rating, while similar conventions apply for $\mathbb{1}_k(i)$, $\mathbb{1}_l(i)$ and $\mathbb{1}_m(i)$ as far as region, sector and seniority are concerned. Therefore, this means that (113) relates to a linear regression problem involving categorical variables. For this reason, in the first of the four summations in (113), j ranges from one to $N_{\text{Rating}} - 1$ rather than to N_{Rating}, and similar considerations hold for k, l, and m, respectively. From an interpretation angle, see Sourabh et al. (2018), x_0^{Base} can be understood as the logarithm of the proxy spread of the base bucket, with the other $N_{\text{Total}} - 4$ unknowns representing the changes in the logarithm when migrating from the base bucket to a different one. Estimating the unknowns in (113) can be done by means of a simple least-squares minimization. As we use the right-hand-side of (113) also in the rest of this article, for conciseness and readability we define it as $\xi_i(\mathbf{x})$, where the first component of the vector $\mathbf{x}$ coincides with x_0^{Base}, followed by x_1^{Rating}, and so on until $x_{N_{\text{Seniority}}-1}^{\text{Seniority}}$.

The cross-section methodology provides a framework to proxy credit curves that is very intuitive and computationally inexpensive. Therefore, for this reason the aim of the current article is that of extending this approach by also taking into account the underlying probability distributions of the considered CDSs by means of Wasserstein barycenters as per Michielon et al. (2022b).

6.4 WASSERSTEIN DISTANCES AND THE CROSS-SECTION METHODOLOGY

Now that the general aspects of the cross-section methodology have been outlined in Section 6.3.2, we describe alternative approaches allowing to embed Wasserstein distances within this framework. In particular, we will outline three

different possibilities before providing a comparison between them in Section 6.5. For the ease of readability and clarity, in the sections that follow we will adopt the following convention: quantities with the tilde superscript always indicate that a bucketing procedure has already taken place, while the absence of the tilde superscript is used in the opposite situation.

6.4.1 Wasserstein cross-intersection methodology on spreads

Assume that, given the CDSs observed in the market with the same maturity, it is possible to bucket them in $N_{\text{Bucket}}^{\text{Market}}$ distinct rating-region-sector-seniority buckets (in general, it holds that $N_{\text{Bucket}}^{\text{Market}} \leqslant N_{\text{Bucket}}$, as CDS quotes are not usually available for every possible rating-region-sector-seniority combination). In Michielon et al. (2022b), see also Section 6.3.1, it has been described how, for each bucket, it is possible to define the unique optimal CDS quote corresponding to the Wasserstein barycenter. This is done by reversing the credit triangle relationship (108), where the recovery rate therein is estimated as the (harmonic[3]) average of the recovery rates of the CDSs in the bucket (later in this section an alternative approach to estimate the recovery rate for a given bucket is illustrated). Assume that the $N_{\text{Bucket}}^{\text{Market}}$ proxy CDS quotes $\tilde{s}_1, \dots, \tilde{s}_{N_{\text{Bucket}}^{\text{Market}}}$ have been defined for each of the market-observable buckets by means of Wasserstein barycenters following the framework outlined in Section 6.3.1. It is then assumed that, for every $i \in \{1, \dots, N_{\text{Bucket}}\}$, the following relationship, like in (113), holds:

$$\ln(\tilde{s}_i) = \xi_i(\boldsymbol{x}). \tag{114}$$

We point out that also appropriate recovery rates need to be calculated. Otherwise, unless default values are assigned to them, it would be not possible to strip credit curves solely from the spreads. Therefore, in order to do so, observe that given a recovery rate $R \in (0,1)$, then $-\ln(R) \in (0, +\infty)$.[4] Hence, a similar approach to that just outlined can be followed to proxy recovery rates by substituting the opposite of the logarithm of the i^{th} (harmonically) averaged recovery rate $\tilde{R}_i$ in an expression as (114). This time, however, one needs to be aware that the regressed variables still need to be positive, otherwise negative recovery rates could be calculated. For this reason, it is needed to either perform a constrained regression where the components of $\boldsymbol{x}$ have to be positive or, otherwise, to perform an unconstrained regression and to eventually floor the negative recovery rates (if any) with a reasonable positive constant. Note that,

3 In Michielon et al. (2022b) it is shown that considering the arithmetic average rather than the harmonic one has a negligible impact on the magnitude of the proxy CDS quotes constructed. Notwithstanding, as here we are dealing with the averaging of (recovery) rates, the choice of the harmonic average seems the natural one to consider.

4 By considering $R \in (0,1)$ we assume that it is not possible to recover all the notional given a default and, further, that always at least a (very) small fraction of the invested notional, which corresponds to a small R, can be recovered if a default event is triggered.

as the step to perform the regression on the recovery rates should be performed before the regression on spreads takes place, the regression on recovery rates can potentially also be performed without taking into account any bucketing.[5]

6.4.2 Wasserstein cross-intersection methodology on hazard rates

We have assumed, throughout this paper, that the default probability distribution implied from a CDS is of the form (107). Therefore, the survival probability implied by means of Wasserstein barycenters for the i^{th} bucket, see Section 6.3.1, is given by

$$\mathbb{Q}(\tilde{\tau}_i > t) = e^{-\tilde{\lambda}_i t}.$$

We can assume that, for every $i \in \{1, \dots, N_{\text{Bucket}}\}$, the following relationship holds

$$\tilde{\lambda}_i = \xi_i(\mathbf{x}).$$

Also in this case, a procedure as that outlined in Section 6.4.1 should be followed to compute proxy recovery rates for each bucket. Note that, as hazard rates need to be positive, then either a constrained regression should be performed, or the computed hazard rates should be properly floored in the case of negative regressed quantities.

6.4.3 Global Wasserstein distance optimization

We now look at the cross-section methodology from a different angle compared to Sections 6.4.1 and 6.4.2. That is, we consider all the K CDS quotes available for the chosen maturity simultaneously. Here, no preliminary bucketing takes place. By considering the flat hazard rate hypothesis (107), we assume that the default distribution implied from the i^{th} CDS can be written as

$$\mathbb{Q}(\tau_i \leqslant t) := 1 - e^{-(\lambda_{\text{Rating}(i)} + \lambda_{\text{Region}(i)} + \lambda_{\text{Sector}(i)} + \lambda_{\text{Seniority}(i)})t}, \tag{115}$$

where the subscript $\text{Rating}(i)$ denotes the rating of the i^{th} CDS, $\text{Region}(i)$ its region, and so forth. Note that if we consider four independent exponentially distributed random variable with rate parameters $\lambda_{\text{Rating}(i)}$, $\lambda_{\text{Region}(i)}$, $\lambda_{\text{Sector}(i)}$, and $\lambda_{\text{Seniority}(i)}$, respectively, then their minimum is still exponentially distributed with rate parameter given by the sum $\lambda_{\text{Rating}(i)} + \lambda_{\text{Region}(i)} + \lambda_{\text{Sector}(i)} + \lambda_{\text{Seniority}(i)}$. Therefore, (115) can be interpreted as the probability distribution that corresponds to the "first-to-default" among the four aforementioned independent

5 Observe that, in principle, one can also perform an unconstrained regression after having transformed the (potentially harmonically averaged) hazard rates by means of the map such that $x \mapsto -\ln\left(\frac{1}{x} - 1\right)$, or by any other function that maps $(0, 1)$ to $(-\infty, +\infty)$ in a smooth and strictly increasing manner, as for instance via the trigonometric functions such that $x \mapsto \cot(\pi x)$ or $x \mapsto \tan\left(\pi\left(x - \frac{1}{2}\right)\right)$, to name but a few.

exponential random variables. Given the implied probability distribution $\mathbb{Q}_i$ of the i^{th} CDS calculated from the market data assuming a flat hazard rate, we denote with $\mathbb{Q}_i^*$ its counterpart defined via (115). For every quoted CDS with distribution $\mathbb{Q}_i$ and flat hazard rate λ_i we can then calculate, via (111) (see also Michielon et al. (2022b) for further details), the amount

$$W_2^2(\mathbb{Q}_i, \mathbb{Q}_i^*) = 2\left(\frac{1}{\lambda_i} - \frac{1}{\lambda_{\text{Rating}(i)} + \lambda_{\text{Region}(i)} + \lambda_{\text{Sector}(i)} + \lambda_{\text{Seniority}(i)}}\right)^2. \quad (116)$$

Thereafter, one can attempt to minimize the aggregate amount

$$\frac{1}{2}\sum_{i=1}^{K} W_2^2(\mathbb{Q}_i, \mathbb{Q}_i^*) \quad (117)$$

with respect to the $N_{\text{Total}} \coloneqq N_{\text{Rating}} + N_{\text{Region}} + N_{\text{Sector}} + N_{\text{Seniority}}$ hazard rates (see (110)). Note, however, that (117) is not strictly convex, as the function $f(x) \coloneqq \left(\frac{1}{\lambda} - \frac{1}{x}\right)^2$ has zero second derivative at $\frac{3}{2}\lambda$. Therefore, the existence of a unique global minimum for (117) is in principle not necessarily guaranteed. Also, note that for common datasets there will be several CDS which share similar features (e.g., they might have the rating and different region, sector and seniority, etc.). This obviously means that equations of the form (116) are not, irrespectively on i, independent from each other. Thus, this implies that finding the minimum (or minima) of (117) analytically is not possible in general and, therefore, numerical techniques should be employed in this respect. As the uniqueness of a global minimum for (117) is not guaranteed, one would need to properly assess which area of the domain of the function to be minimized should be taken into account for the optimization procedure. This suggests, therefore, the use of stochastic optimization algorithms as Basin-hopping Wales and Doye (1997) or Dual Annelling Xiang et al. (2013), at least to determine a sensible estimate for the initial point for the minimization of (117). As a consequence, given the high number of parameters involved in the optimization, we now propose a modified version of the optimization problem (117) to guarantee strict convexity.

Given $\epsilon > 0$ small and fixed a-priori, define the function

$$g_\epsilon^\lambda(x) \coloneqq \begin{cases} f(x) \text{ if } x \leqslant \frac{3}{2}\lambda - \epsilon \\ f\left(\frac{3}{2}\lambda - \epsilon\right) + f'\left(\frac{3}{2}\lambda - \epsilon\right)\left(x - \frac{3}{2}\lambda + \epsilon\right) \\ \qquad + f''\left(\frac{3}{2}\lambda - \epsilon\right)\left(x - \frac{3}{2}\lambda + \epsilon\right)^2 \text{ if } x > \frac{3}{2}\lambda - \epsilon \end{cases}, \quad (118)$$

i.e., $g_\epsilon^\lambda(\cdot)$ coincides with $f(\cdot)$ until just before its inflection point, and afterwards it coincides with its second order Taylor approximation. By construction, this function is twice continuously differentiable and strictly convex on the positive

real line. We can therefore consider, once a fixed value for ϵ has been chosen, instead of (117) its approximated counterpart given by

$$\sum_{i=1}^{K} g_{\epsilon}^{\lambda_i}\left(\frac{1}{\lambda_{\text{Rating}(i)}+\lambda_{\text{Region}(i)}+\lambda_{\text{Sector}(i)}+\lambda_{\text{Seniority}(i)}}\right). \tag{119}$$

We now show that the alternative function to be minimized, i.e., (119), is strictly convex on $(0,+\infty)^{N_{\text{Total}}}$.

Theorem 5. *The function* (119) *is strictly convex on* $(0,+\infty)^{N_{\text{Total}}}$.

Proof. Fix an index i and observe that the map $\pi_i : (0,+\infty)^{N_{\text{Total}}} \to (0,+\infty)$ such that $\pi_i(\boldsymbol{\lambda}) \coloneqq \lambda_{\text{Rating}(i)}+\lambda_{\text{Region}(i)}+\lambda_{\text{Sector}(i)}+\lambda_{\text{Seniority}(i)}$ is linear. Given $\theta \in [0,1]$ and $\boldsymbol{\lambda}_*$ and $\boldsymbol{\lambda}^*$ in $(0,+\infty)^{N_{\text{Total}}}$, it results that

$$\begin{aligned} g_{\epsilon}^{\lambda} \circ \pi_i(\theta\boldsymbol{\lambda}_* + (1-\theta)\boldsymbol{\lambda}^*) &= g_{\epsilon}^{\lambda}(\pi_i(\theta\boldsymbol{\lambda}_* + (1-\theta)\boldsymbol{\lambda}^*) \\ &= g_{\epsilon}^{\lambda}(\theta\pi_i(\boldsymbol{\lambda}_*) + (1-\theta)\pi_i(\boldsymbol{\lambda}^*)) \qquad (120) \\ &< \theta g_{\epsilon}^{\lambda}(\pi_i(\boldsymbol{\lambda}_*)) + (1-\theta) g_{\epsilon}^{\lambda}(\pi_i(\boldsymbol{\lambda}^*)) \qquad (121) \\ &= \theta g_{\epsilon}^{\lambda} \circ \pi_i(\boldsymbol{\lambda}_*) + (1-\theta) g_{\epsilon}^{\lambda} \circ \pi_i(\boldsymbol{\lambda}^*), \end{aligned}$$

where equality (120) comes from the linearity of $\pi_i(\,\cdot\,)$, while the strict inequality in (121) is a consequence of the strict convexity of $g_{\epsilon}^{\lambda}(\,\cdot\,)$. Recalling that the sum of strictly convex functions is still strictly convex, if follows that (119) is strictly convex on its (convex) domain $(0,+\infty)^{N_{\text{Total}}}$. □

We highlight that, from a practical perspective, it is not necessary to perform a minimization for each variable in (119) on the whole positive line, as setting a sufficiently large upper bound suffices. This is because, despite in principle hazard rates can reach extremely large values in $(0,+\infty)$, in practice this is usually not the case. For example, on 12 September 2008 for Lehman Brothers, just before its bankruptcy was announced on 14 September 2008, the implied hazard rate for its one-year CDS was of the order of 230,000bps while the other hazard rates corresponding to longer-dated maturities were all below 10,000bps; see Brigo et al. (2010). Also note that, for close-to-default names, often in practice no proxy CDS curves are calculated, and more conservative approaches are used, instead. Therefore, minimizing (119) can be performed on a suitable compact set of $(0,+\infty)^{N_{\text{Total}}}$, which guarantees existence and unicity of a global minimum in virtue of Theorem 5. The fact that it is possible to approximate the minimum of (117) by minimizing (119) by means of a twice continuously differentiable function guarantees that quasi-Newton methods such that the Broyden–Fletcher–Goldfarb–Shanno (Press et al., 2007, Ch. 10.9), often used in financial practical applications, can be employed directly. Note that there are other minimization techniques that can be used as, for instance, the Powell's method

(Press et al., 2007, Ch. 10.7). For more details about optimization methods used in different financial applications the interested reader can refer to Cornuejols et al. (2018).[6]

6.5 COMPARISON OF THE DIFFERENT METHODOLOGIES

In this section we provide a comparison between the different methodologies proposed in this article to assess how they perform and compare with each other under real market circumstances. Before doing so in Section 6.5.2, we describe the data preprocessing steps we have followed in Section 6.5.1.

6.5.1 Data preprocessing

The data preprocessing steps we have followed are quite similar to those considered in Sourabh et al. (2018). As CDS quotes that are related to the "CCC" or "D" rating are often not reliable or simply not available, we exclude those ratings from the dataset. Further, we only consider, as regions, "Asia", "Europe" and "North America", and label the remaining ones as "Other". For "Europe" we consider only CDSs in EUR currency, while for all the other regions we only consider USD CDSs, as they are more liquid. As far as the sector is concerned, we remove all the "Government" quotes, since there seems to be evidence that those should be analysed separately; see Longstaff et al. (2011) and Pan and Singleton (2008). We also label the "Basic materials", "Consumer services", "Energy", "Industry", "Technology" and "Telecommunication services" sectors as "Cyclical", while "Consumer goods", "Healthcare", "Utilities" as "Non-cyclical".[7] The "Financials" sector is left unchanged. The choices and mappings for the ratings, regions, sectors and seniorities have been summarized in Table 6, for convenience. The mapping performed allows to define 6 possible ratings, 4 possible regions, 3 possible sectors and 2 possible seniorities, leading to a total of 144 buckets. We also keep a distinction between senior and subordinated credit, and we remove all the CDS quotes with spread larger than 1,000bps. Further, we only take into account CDS names for which data was provided by at least two market participants.

6 Observe that many optimization algorithms are nowadays available in standard libraries for various programming languages as, e.g., QuantLib (`www.quantlib.org`), or SciPy (`www.scipy.org`).

7 We recall that cyclical and non-cyclical sectors refer, in this case, to how correlated the credit quality of a given entity is with the relevant macroeconomic factors: cyclical sectors are closely related to the current economic trend, while the opposite holds for the non-cyclical ones.

	From	To
Rating	AAA AA A BBB BB B	AAA AA A BBB BB B
	CCC D	–
Region	Asia	Asia
	Europe	Europe
	North America	North America
	Africa Eastern Europe India Latin America Middle East Oceania Offshore	Other
Sector	Basic materials Consumer services Energy Industrials Technology Telecommunications services	Cyclical
	Consumer goods Healthcare Utilities	Non-cyclical
	Financials	Financials
	Government	–
Seniority	Senior	Senior
	Subordinated	Subordinated

Table 6: The mapping performed in terms of rating, region, sector and seniority for the datasets considered in this article. The dash symbol "–" means that all the CDSs linked to that category have been removed from the dataset: the left column collects the raw labels, while the right one contains the preprocessed ones.

6.5.2 Results

Before providing a comparison between the different methodologies outlined in this article, we introduce some abbreviations for the ease of clarity and readability. In particular, we refer to the cross-section methodology described in Section 6.3.2 as CSM, to the Wasserstein cross-intersection methodology on spreads of Section 6.4.1 as WCIMS, to the Wasserstein cross-intersection methodology on hazard rates of Section 6.4.2 as WCIMHR and, further, to the global Wasserstein distance optimization methodology of Section 6.4.3 as GWDOM. Further, again for comprehensibility purposes, if the approach outlined in Section 6.4.1 is applied to the arithmetically-averaged CDS quotes for each observable bucket (i.e., the standard intersection methodology is applied before the regression takes place), then we name this variation as average cross-intersection methodology on spreads, and refer to it as ACIMS. This last approach will be solely used in this section for benchmarking purposes with the WCIMS.

As a first step in our analysis, we start by considering the CSM given the dataset taken into account and the preprocessing of the data outlined in Section 6.5.1. In particular, for comparison purposes we consider the spreads computed for the base bucket using the CSM during the yearly period considered in this article. Moreover, keeping region, sector and seniority unchanged, we consider all the possible ratings, and depict the results in Figure 42. As expected, Figure 42 illustrates how the CSM properly differentiates amongst ratings, in the sense that to lower (higher) CDS spreads correspond higher (lower) ratings.

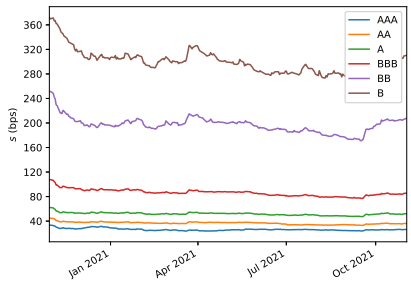


Figure 42: Proxy CDS spreads computed using the CSM, given different ratings, for the base bucket during the time period spanning from 1 November 2020 to 1 November 2021.

As a second step, as this will be useful for benchmarking purposes, we implement the ACIMS. We do this as, by the results available in Michielon et al. (2022b), see also Section 6.3.1, the proxy spreads generated by means of Wasserstein barycenters should be lower than their counterparts computed by averag-

ing the spreads in each bucket. Therefore, we would overall expect a similar behaviour also in the cases considered here when the regression takes place. For the base bucket, and keeping its region, sector and seniority fixed while varying the rating, we obtain the results illustrated in Figure 43.

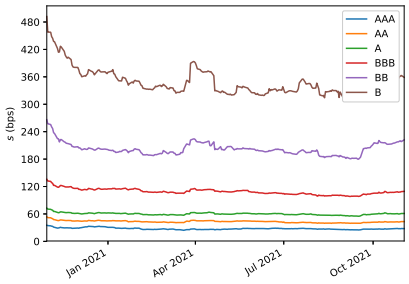


Figure 43: Proxy CDS spreads computed using the ACIMS, given different ratings, for the base bucket during the time period spanning from 1 November 2020 to 1 November 2021.

Also in this case, in line with what observed in Figure 42, we see a clear distinction between ratings, with the expected monotonicity property being accurately reflected in the graphs, that is, to higher (lower) ratings correspond lower (higher) credit spreads. We now analyse the approach proposed in Section 6.4.1, i.e, the WCIMS. The results are available in Figure 44.

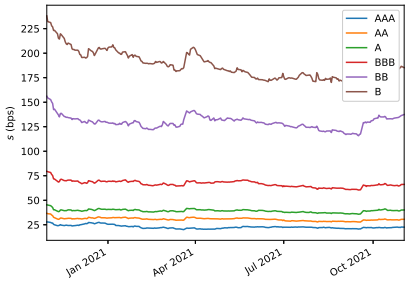


Figure 44: Proxy CDS spreads computed using the WCIMS, given different ratings, for the base bucket during the time period spanning from 1 November 2020 to 1 November 2021.

As expected, in this case we obtain proxy spreads which are below their counterparts depicted in Figure 43. We also recall that, once proxy spreads are computed with the WCIMS, then in order to imply proxy credit curves, proxy recovery rates are also necessary. For this reason, as outlined in Section 6.4.1, we have proposed different approaches (which share the same base rationale) for doing so.

We provide in Figures 45a and 45b two examples. As expected, the regressed recovery rates all have an order of magnitude neighbouring 40%. Also, as it can be seen for instance in Varna et al. (2003), in practice there is not always monotonicity in recovery rates with respect to ratings (i.e., higher credit ratings are expected to have higher recovery rates), which is what we can see in Figures 45a and 45b (despite regressing without harmonically averaging the recovery rates first, see Figure 45b, produces results which better satisfy the monotonicity with respect to ratings compared to Figure 45a). For additional details concerning the link between ratings and recovery rates refer, e.g., to Altman et al. (2005).

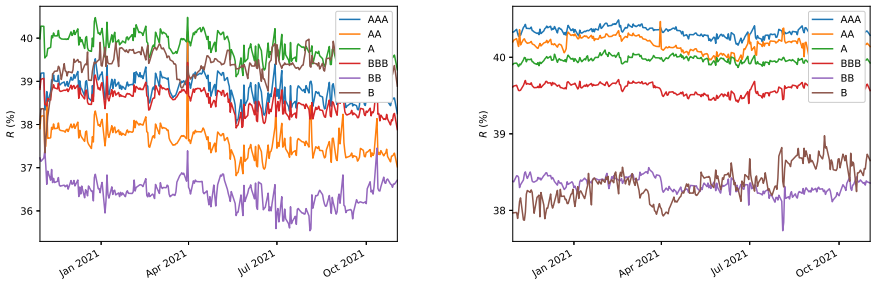


Figure 45: Proxy recovery rates computed using the cross-intersection methodology for recovery rates available in Section 6.4.1 (based on regressing minus the logarithms of the recovery rates) with and without harmonically averaging the recovery rates before performing the regression, given different ratings, for the base bucket during the time period spanning from 1 November 2020 to 1 November 2021; see panels (a) and (b), respectively.

The results of Figures 45a and 45b indicate that, ball-park, just always using a recovery rate of 40% would be, in practice, an acceptable choice. In order to assess the impact on proxy default probabilities of a 40% recovery rate rather than any of the other values observed in Figures 45a and 45b, we consider a CDS belonging to the base bucket. For each of the dates in the period taken into account, we calculate the default probability for the relevant entity within a 5-year horizon. The results are available in Figure 46.

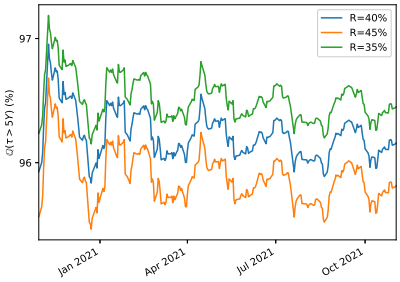


Figure 46: 5-year default probabilities from an entity belonging to the base bucket during the time period spanning from 1 November 2020 to 1 November 2021 calculated using different recovery rates.

It is a well-known fact that default probabilities are not very sensitive to the value chosen for the recovery rate, and this is confirmed by Figure 46. Figure 45a indicates that the uncertainty in recovery rate values is of the order of 4%, while Figure 45b seems to indicate a slightly lower uncertainty, of the order of a couple of percentage points. Taking into account recovery rates which differ from the somewhat standard value of 40% as done in Figure 46 results in default probabilities which do not differ much more than 1%. Given that recovery rates are not quoted, simply using the (often) commonly-used value of 40% seems a justified choice from a practitioner's angle in this respect.

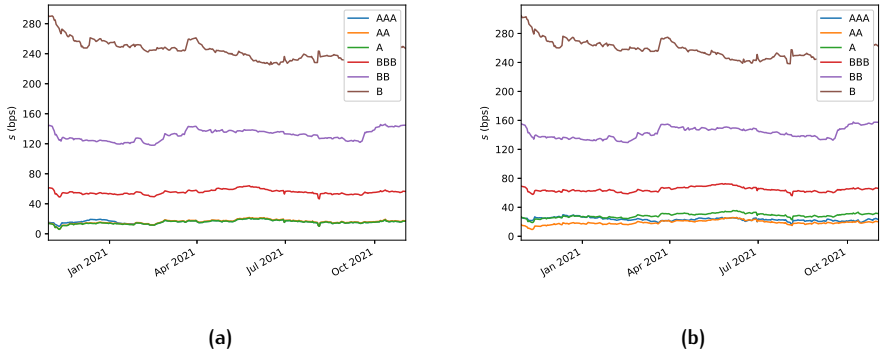


Figure 47: Proxy CDS spreads computed using the WCIMHR with and without constraints to guarantee positiveness, given different ratings, for the base bucket during the time period spanning from 1 November 2020 to 1 November 2021; see panels (a) and (b), respectively.

We now show how the proxy CDS spreads for the base bucket look like if they are proxied by means of the WCIMHR. The results have been provided in Figures 47a and 47b, which display a similar behavior.

We observe that the WCIMHR seems to perform less well than the WCIMS. In particular, the WCIMHR fails to properly distinguish between the top three investment-grade ratings. For the remaining three ratings, the results seem slightly more conservative than their counterparts in Figure 44, and the monotonicity property with respect to the ratings remains satisfied in this case. Note that, when constraints to guarantee the positiveness of the hazard rates are imposed as in Figure 47a (otherwise, this might in principle produce negative hazard rates), having the regression coefficients less degrees of freedom causes the results to be less accurate in the area where the spreads are small (i.e., for entities with high credit worthiness), as in this case we are close to the boundary with zero. This is confirmed by the fact that, for lower credit ratings, the distinction between credit spreads linked to different ratings is crystal clear and, ballpark, in line with what observed in Figure 44. If the constraints are removed, see Figure 47b, the distinction between the three top ratings is clearer than in Figure 47a, but the desirable monotonicity property is not always guaranteed. We highlight that recovery rates are not market-observable quantities, and that the same applies to hazard rates. Therefore, as it is also suggested by Figures 47a and 47b compared to Figure 44, the choice of the WCIMHR would be suboptimal compared to the WCIMS.

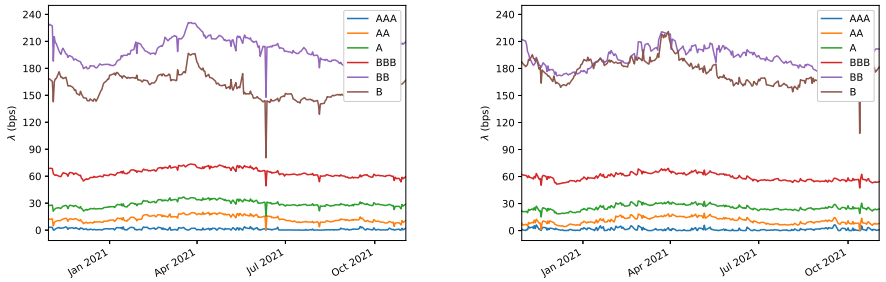


Figure 48: Proxy CDS spreads computed using the GWDOM in its original form (117) and via its approximation (119), given different ratings, for the base bucket during the time period spanning from 1 November 2020 to 1 November 2021; see panels (a) and (b), respectively.

We now look at the third option analysed in this article, i.e., the GWDOM. In particular, we look at this approach from two angles, i.e., by means of perform-

ing the minimization of (117) directly, but also by considering its approximated counterpart (119). The results, in terms of hazard rates, have been depicted in Figures 48a and 48b, respectively.

First of all, we notice how, for investment-grade ratings, approaching the optimization problem in its original form (117) or via its approximation (119) produce similar results. This means that setting up the approximated error function (118) (here, we have chosen ϵ equal to 1bp) produces results in line with expectations. We notice in the results outlined in Figures 48a and 48b that the hazard rates are monotonic with respect to the rating, expect for the "B" and "BB" cases. Here, the approximated approach (119) seems to perform slightly better than its counterpart (117). Nonetheless, we note that in both cases, and independently on the credit worthiness level, the trend of the hazard rates exhibits some evident sawtooth patterns. This is due to the fact that both optimization problems require several parameters to be estimated and, therefore, seem to be quite sensitive to the market data used as input. We observe this especially for the worst two ratings where, due to the data containing CDS quotes of badly-rated entities, the results show even less smooth features. Therefore, despite sound from a theoretical perspective, the GWDOM seem to underperform compared to both the WCIMS and the WCIMHR. We also empathise that the GWDOM is computationally much more intensive than the other two.

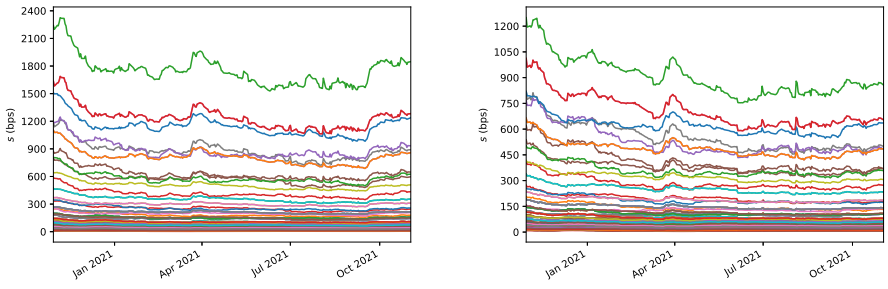


Figure 49: Proxy CDS spreads computed using the WCIMHR with and without constraints to guarantee positiveness, given different ratings, for the base bucket during the time period spanning from 1 November 2020 to 1 November 2021; see panels (a) and (b), respectively.

From the analysis we have carried out in this section we can conclude that, amongst the three approaches analysed, the WCIMS clearly outperforms the other two. Therefore, we make a final comparison between the WCIMS and the CSM. In particular, for all the 144 possible buckets that can be defined given the

choices and mappings described in Table 6, we plot the proxy spreads for the time period considered. The results have been provided in Figures 49a and 49b, respectively.

We first highlight how both methodologies provide stable results for all the buckets with the same qualitative features (i.e., curve shapes). This also illustrates the robustness of the methodology originally proposed in Chourdakis et al. (2013a) and further enhanced in Sourabh et al. (2018). We observe that the WCIMS produces, for rating, region, sector and seniority where CDS spreads are high, hazard rates that are below those of their counterparts calculated using the CSM. For combinations where, on the other hand, CDS are low, the two approaches tend to be closer to each other, or the other relationship might hold. This is because, as far as the intercept of the model is concerned, the CSM is consistent with the average of the (log) CDS spreads observed, while the WCIMS with the mean of the (log) harmonic averages of the bucketed spreads. This means that, in the WCIMS, lower weights (higher) are given to high (low) credit spreads. During the time frame taken into consideration, on average approximately 77% of the curves produced by the WCIMS are below their counterparts constructed by means of the CSM, with a standard deviation of roughly 3%.

As an example, we consider the red lines in Figures 49a and 49b. In the former case, at the beginning of the period, the CDS spread is approximately 1,600bps, while in the second case it is roughly 1,000bps. In the first case, by using the credit triangle relationship (108) and considering a recovery rate of 40%, it results that the hazard rate approximately equals 2,777bps, while in the second case it roughly equals 1,667bps. These hazard rates lead to 5-year default probabilities of approximately 74% and 57%. However, for lower CDS spreads the magnitude of the differences between hazard rates would be milder, if not slightly lower in the case the CSM is used, in some cases.

6.6 CONCLUSION

The main contribution available in this paper is the development of a novel methodology for constructing proxy credit curves starting from observable CDS market quotes. In particular, we have investigated three possible approaches to take into account the default probability structure of CDSs within the cross-sectional framework of Chourdakis et al. (2013a) and Sourabh et al. (2018), with the addition of explicitly including probability-related considerations by means of Wasserstein distances. Amongst the three possibilities envisaged, we have assessed how regressing proxy CDS spreads computed by means of Wasserstein barycenters for the observable buckets, in order to interpolate and extrapolate the available information to the empty buckets as well, seems to be the best strategy to follow. We have compared this methodology to the cross-section approach

and observed that, in line with Michielon et al. (2022b), neglecting the default probability distribution of the underlying CDSs might lead to overestimating credit risk. The analysis performed on the chosen methodology also confirms how stable, robust and flexible the cross-section methodology of Chourdakis et al. (2013a) and Sourabh et al. (2018) is. By means of the hybrid curve construction framework outlined in the article at hand we have shown how Wasserstein distances can be employed to model default risk, and this hopefully paves the way to further developments towards this direction.

6.A THE CROSS-SECTION METHODOLOGY ON HAZARD RATES

In Section 6.3.2 the basic features of the cross-section methodology Chourdakis et al. (2013a) and Sourabh et al. (2018) have been highlighted. Despite this article deals with the cross-section methodology from a different angle (i.e., by employing Wasserstein distances and barycenters), we want to highlight that, even if the regression equations in (113) are based on the logarithms of the CDS spreads, one could look at this approach from an hazard rate perspective. That is, one could reinterpret (113) by considering, for every $i \in \{1, \dots, K\}$, a relationship of the form

$$\lambda_i = \xi_i(\mathbf{x}), \tag{122}$$

where each hazard rate is computed by means of the credit triangle (108).[8]

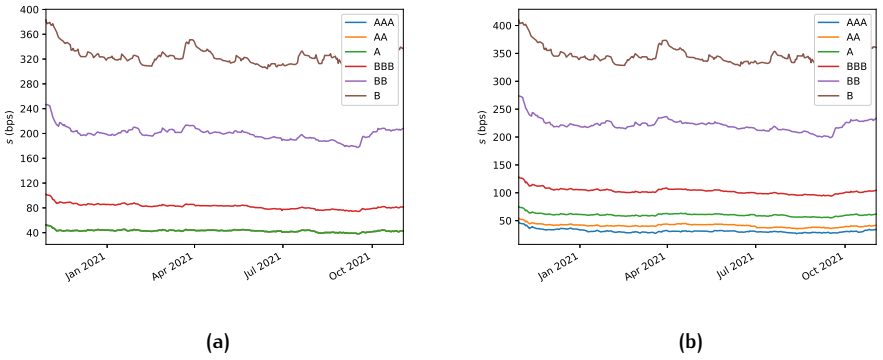


Figure 50: Proxy CDS spreads computed using the WCIMHR with and without constraints to guarantee positiveness, given different ratings, for the base bucket during the time period spanning from 1 November 2020 to 1 November 2021; see panels (a) and (b), respectively.

Also in this case, due to the fact that the regressed hazard rates are expected to be positive, one could either impose the regression coefficients to be positive a priori, or floor the results, if needed, by a sufficiently small positive constant. By also regressing for every $i \in \{1, \dots, K\}$ the recovery rates in a similar manner as Section 6.4.1 outlines, one could then employ the credit triangle (108) again to back out the proxy CDS spreads. For completeness, in Figures 50a and 50b we illustrate how the proxy CDS spreads will look like in this case.

As Figure 50a shows, restraining the regression coefficients to be positive imposes too strict conditions on the regression results, leading to the ratings equal

8 As an additional variation, one could also perform the regression on the logarithms of the hazard rates.

to "A" or better to be almost indistinguishable from each other (a similar behaviour has been observed in Figure 47a). On the other hand, not imposing the constrains (in this case it is not flooring of the regressed variables is needed) allows to clearly distinguish amongst all the ratings and, in addition, to guarantee the desired monotonicity property with respect to the credit worthiness. We also observe that regressing at hazard rate level results in proxy CDS spreads that are more conservative than their counterparts calculated using the cross-section methodology depicted in Figure 42. Also in this case, we highlight that recovery rates are derived quantities and, therefore, they are not quoted (and the same applies to hazard rates). So, despite in this case the regression methodology applied to hazard rates provides stable results which are monotonic with respect to credit ratings, regressing the spreads directly (which are observable) should still be the preferred choice.

7 SUMMARY

The research material available in this practice-oriented PhD thesis delves into two areas of financial mathematics, namely *bid-ask pricing* and *credit risk*.

Firstly, in Part i we investigate different applications of the *conic finance* theory, which aims to model the dynamics of financial markets in the presence of frictions and, in particular, of bid-ask spreads. These applications include estimating risk-neutral model parameters (and market data) starting from bid and ask security prices directly, as well as the deployment of machine learning techniques to price in markets with direction-dependent prices.

Secondly, in Part ii we present applied work pertaining to the area of credit risk. In particular, the research conducted in this respect relates to estimating default probabilities of counterparties in the presence of insufficient, (partly) missing and/or untrustworthy *credit default swap* (CDS) market data. For this purpose, the concepts of *Wasserstein distance* and *Wasserstein barycenter* play a pivotal role.

Part i concerns applications of the conic finance theory of Cherny and Madan (2010) in different areas. More precisely, in Chapter 2 (based on Michielon et al. (2021a)), the relationship between bid-ask market quotes of CDSs and their corresponding risk-neutral default probabilities is investigated. That is, in Chapter 2 it is illustrated how it is possible to extract, under mild and intuitive assumptions, risk-neutral default probabilities from bid and ask CDS quotes directly. This is achieved assuming that the dynamics of the default time of a given CDS reference entity follow the specifications of a Poisson process with piecewise-constant hazard rate term structure. The methodology outlined allows not to rely on any mid-quote approximation for the CDS market data used as model input. From a numerical perspective, the results available in Chapter 2 only requires to solve a non-linear system in two equations and two unknowns, easily achievable by means of standard computational methods.

The research work available in Chapter 2 is then extended to a more general framework in Chapter 3 (based on Michielon et al. (2021b)). Specifically, in Chapter 3 the problem of computing risk-neutral implied volatilities from bid and ask (European) option quotes directly is considered and investigated, providing insights on the differences between risk-neutral implied volatility smiles and their counterparts, computed starting from mid-quote option prices for a variety of underlying assets. The technique used in Chapter 3 (and, similarly, in Chapter 2), can also be used to compute model-implied parameters from bid and ask

security prices directly if their risk-neutral counterparts are strictly increasing with respect to the former.

Chapter 4 (based on Michielon et al. (2023)) provides applications of *neural networks* (and, specifically, of *multilayer perceptrons*) to model liquidity in financial markets in two different frameworks. That is, we first show how to replicate bid and ask model prices, within the conic finance paradigm, by means of (vector-valued) multilayer perceptrons. This approach allows for accurate price replication (as well as sensitivities calculations), and at the same time provides remarkable computational speed-ups compared to, e.g., (conic) Monte Carlo simulations. It is worth noting that the methodology proposed here for this purpose is agnostic of the underlying valuation model, and that it can be trivially adapted to all pricing techniques compatible with the conic finance paradigm. Furthermore, in Chapter 4 we adapt the work available in Cuchiero et al. (2020), which concerns modeling the local volatility component of a *local stochastic volatility model* by a (combination of) neural network(s), to markets exhibiting bid-ask spreads. On a side note, we additionally illustrate how it is possible to generate *hybrid families of distortion functions* allowing for extra flexibility during bid-ask model calibration routines, and we also introduce the *conic SABR model*, which still allows for (pseudo-)analytical formulae for bid and ask (European) option prices.

Part ii relates to the area of credit risk. In particular, the research work available therein aims to provide alternative methodologies to compute proxy credit curves when unreliable and/or incomplete CDS market data is available. The methodologies developed for this purpose are based on the notions of Wasserstein distance and Wasserstein barycenter (refer, e.g., to Agueh and Carlier (2011), (Panaretos and Zemel, 2020, Ch. 1 and 2), (Villani, 2003, Ch. 7) and (Villani, 2009, Ch. 6)), which make it possible to tackle the underlying problem using robust probabilistic tools.

More in detail, Chapter 5 (based on Michielon et al. (2022b)) aims to look at the *intersection methodology* EBA (2013) for credit curves from a probabilistic perspective. Concretely, therein we show, under simple and standard assumptions, how this revised methodology leads to elementary and intuitive formulae to calculate distances between CDS-implied default probability distributions. Further, we illustrate how to use this information to construct proxy CDS quotes by means of Wasserstein barycenters.

On the other hand, in Chapter 6 (based on Michielon et al. (2022a)) we propose and compare different approaches to take into account counterparty-specific information in terms of rating, region, sector, etc. at *cross-sectional level*, as in Chourdakis et al. (2013b) and Sourabh et al. (2018), with the aim of calculating risk-neutral default probabilities from CDSs. The methodologies proposed in Chapter 6 attempt to provide improvements to the intersection methodology EBA (2013), again aiming to take into account the implicit probabilistic informa-

tion embedded in CDS markets, which is in general overlooked by the classical and mainly data-driven approaches for counterparty risk.

To conclude, the research work carried out in the dissertation at hand proposes and investigates promising techniques in financial mathematics, aiming to offer enhancements, as well as diverse perspectives, within the practical domains of bid-ask pricing and credit risk.

8 SAMENVATTING (IN DUTCH)

Het onderzoek dat gedocumenteerd is in dit praktijkgerichte proefschrift gaat dieper in op twee gebieden van de financiële wiskunde, namelijk *bid-ask pricing* en *kredietrisico*.

Ten eerste onderzoeken we in Deel i verschillende toepassingen van de conische financiële theorie (*conic finance*), die tot doel heeft de dynamiek van financiële markten te modelleren in aanwezigheid van fricties en, in het bijzonder, van bid-ask spreads. Deze toepassingen omvatten het schatten van risiconeutrale modelparameters, uitgaande van directe bid- en askprijzen, evenals het gebruik van machine learning-technieken om prijzen te bepalen op markten met richtingsafhankelijke prijzen.

Ten tweede presenteren we in Deel ii toegepast werk op het gebied van kredietrisico. Het onderzoek dat in dit verband wordt uitgevoerd heeft in het bijzonder betrekking op het schatten van de kans op *default* (in gebreke blijven) van tegenpartijen in geval van onvoldoende, (deels) ontbrekende en/of onbetrouwbare marktgegevens van een *credit default swap* (CDS). Voor dit doel spelen de concepten *Wassersteinafstand* en *Wassersteinbarycentrum* een centrale rol.

In Deel i behandelen we toepassingen van de conische financieringstheorie (*conic finance*) uit Cherny en Madan (2010) op verschillende gebieden. Preciezer gezegd, in Hoofdstuk 2 (gebaseerd op Michielon e.a. (2021a)) wordt de relatie tussen bid-askmarktkoersen van CDS's en hun overeenkomstige risiconeutrale defaultkansen onderzocht. Dat wil zeggen, in Hoofdstuk 2 laten we zien hoe het mogelijk is om, onder milde en intuïtieve aannames, risiconeutrale defaultkansen rechtstreeks uit bid- en ask-CDS-koersen te extraheren. Dit wordt bereikt onder de veronderstelling dat de tijd tot default van een gegeven CDS-referentie-entiteit de specificaties volgt van een Poissonproces met een stuksgewijs constante intensiteit. De geschetste methodologie leidt tot een alternatieve aanpak voor de gebruikelijke benadering met de middenkoersen voor de CDS-marktgegevens als modelinvoer. Vanuit numeriek perspectief vereisen de resultaten die verkregen zijn in Hoofdstuk 2 alleen het oplossen van een niet-lineair systeem in twee vergelijkingen en twee onbekenden, wat gemakkelijk te realiseren is met behulp van standaard computermethoden.

Het onderzoek waarover gerapporteerd is in Hoofdstuk 2 wordt vervolgens uitgebreid naar een algemener raamwerk in Hoofdstuk 3 (gebaseerd op Michielon e.a. (2021b)). Specifiek wordt in Hoofdstuk 3 het probleem van het rechtstreeks berekenen van risiconeutrale impliciete volatiliteiten uit bid- en askkoer-

sen van (Europese) opties beschouwd en onderzocht. Hierdoor worden nieuwe inzichten verkregen in de verschillen tussen risiconeutrale impliciete volatiliteitssmiles en hun tegenhangers die berekend worden uitgaande van middenkoersoptieprijzen. De techniek die wordt gebruikt in Hoofdstuk 3 (en op soortgelijke wijze in Hoofdstuk 2), kan ook worden gebruikt om model-geïmpliceerde parameters direct te berekenen uit bid- en askeffectenprijzen, als hun risiconeutrale tegenhangers strikt stijgend afhankelijk zijn van eerstgenoemde.

Hoofdstuk 4 (gebaseerd op Michielon e.a. (2023)) bevat toepassingen van *neurale netwerken* (en specifiek van *meerlaagse perceptrons*) om de liquiditeit in financiële markten in twee verschillende raamwerken te modelleren. We laten eerst zien hoe we bid- en askmodelprijzen kunnen repliceren, binnen het conische financiële paradigma, door middel van (vectorwaardige) meerlaagse perceptrons. Deze aanpak maakt nauwkeurige prijsreplicatie mogelijk (evenals het berekenen van gevoeligheden), en verschaft tegelijkertijd opmerkelijke verbeteringen van rekensnelheden vergeleken met bijvoorbeeld (conische) Monte Carlosimulaties. Het is vermeldenswaard dat de hier voor dit doel voorgestelde methodologie agnostisch is ten opzichte van het onderliggende waarderingsmodel, en dat deze op triviale wijze kan worden aangepast aan alle prijstechnieken die compatibel zijn met het conische financiële paradigma. Verder passen we in dit hoofdstuk resultaten toe uit Cuchiero e.a. (2020), dat betrekking heeft op het modelleren van de lokale volatiliteitscomponent van een *lokaal stochastisch volatiliteitsmodel* door een combinatie van neurale netwerken, in modellen voor markten met bid-asksprеads. Als een bijproduct laten we bovendien zien hoe het mogelijk is om *hybride families van distortiefuncties* te genereren, wat extra flexibiliteit genereert tijdens kalibratieroutines van bid-askmodellen. We introduceren ook het *conische SABR-model*, waarvoor we nog steeds in staat zijn (pseudo-)analytische formules voor bid- en askprijzen van (Europese) opties af te leiden.

Deel ii heeft betrekking op het gebied van kredietrisico. Het daarin gepresenteerde onderzoek is met name gericht op het verschaffen van alternatieve methodologieën voor het berekenen van proxy-kredietcurves wanneer alleen onbetrouwbare en/of onvolledige CDS-marktgegevens beschikbaar zijn. De voor dit doel ontwikkelde methodologieën zijn gebaseerd op de begrippen Wassersteinafstand en Wassersteinbarycentrum (zie bijvoorbeeld Agueh en Carlier (2011), (Panaretos en Zemel, 2020, Hfdst. 1 en 2), (Villani, 2003, Hfdst. 7) en (Villani, 2009, Hfdst. 6)), die het mogelijk maken het onderliggende probleem aan te pakken met behulp van robuuste probabilistische technieken.

Meer in detail beoogt Hoofdstuk 5 (gebaseerd op Michielon e.a. (2022b)) de *intersectiemethodologie* EBA (2013) voor kredietcurves vanuit een probabilistisch perspectief te bekijken. Concreet laten we daarin, onder eenvoudige en standaardaannames, zien hoe deze herziene methodologie leidt tot elementaire en intuïtieve formules om de Wassersteinafstanden tussen CDS-geïmpliceerde kansverdelingen voor tijdstippen van defaults te berekenen. Verder illustreren we hoe

deze informatie kan worden gebruikt om proxy-CDS-koersen te construeren met behulp van Wassersteinbarycentra.

In aanvulling hierop introduceren en vergelijken we in Hoofdstuk 6 (gebaseerd op Michielon e.a. (2022a)) verschillende benaderingen die rekening te houden met tegenpartijspecifieke informatie in termen van rating, regio, sector, enz. op *cross-sectioneel niveau*, zoals in Chourdakis e.a. (2013b) en Sourabh e.a. (2018), met als doel de risiconeutrale defaultkansen van CDS's te berekenen. Met de in dit hoofdstuk voorgestelde methodologieën brengen we verbeteringen aan in de intersectiemethodologie uit EBA (2013). Hierbij houden we opnieuw rekening met de impliciete probabilistische informatie die is ingebed in CDS-markten, informatie die voorheen in het algemeen over het hoofd werd gezien bij de klassieke en voornamelijk datagestuurde benaderingen voor tegenpartijrisico.

Concluderend kan worden gesteld dat het onderzoek dat in dit proefschrift wordt beschreven, veelbelovende technieken in de financiële wiskunde introduceert en onderzoekt, met als doel zowel verbeteringen als verschillende perspectieven te bieden binnen de praktische toepassingsdomeinen van bid-askprijzen en kredietrisico.

9 AUTHOR CONTRIBUTIONS

Chapter 2, which resulted in Michielon et al. (2021b), is based on an idea of Matteo Michielon. Both implementation and writing were also done by Matteo Michielon. The research work of this chapter was conducted under the supervision of both Peter Spreij and Asma Khedher.

The work related to Chapter 3, which resulted in Michielon et al. (2021b), is also based on an idea of Matteo Michielon, who also performed the implementation and did the writing. The idea of studying the invariance property of the Wang transform when applying a non-decreasing and left-continuous function to a (log)normal random variable is due to Peter Spreij. The proof of the related result (available in Appendix 3.C) is due to Matteo Michielon. The research work of this chapter was conducted under the supervision of both Peter Spreij and Asma Khedher.

As far as Chapter 4 is concerned, the idea of investigating possible applications of machine learning techniques in the area of bid-ask pricing is due to Asma Khedher. The modeling ideas of Chapter 4 are due to Matteo Michielon, including those related to the conic SABR model (see Section 4.3.1) and to hybrid distortion functions (see Section 4.3.3). From an implementation perspective, a prototype of the neural-network-based methodology available in Chapter 4.6 was done by Alessandro Gentile, who also suggested the idea of using vector-valued multilayer perceptrons as possible choice for two-dimensional price approximations. Further, the implementation of the methodology available in Section 4.7 was performed by Diogo Franquinho. The rest of the work, including writing, was done by Matteo Michielon. The research work of this chapter was conducted under the supervision of both Peter Spreij and Asma Khedher.

Chapter 5, which resulted in Michielon et al. (2022b), is based on an idea of Matteo Michielon, who also performed the implementation and did the writing. Theorem 3 is due to Peter Spreij. The research work of this chapter was conducted under the supervision of both Peter Spreij and Asma Khedher.

Chapter 6, leading to Michielon et al. (2022a), originates from an idea of Matteo Michielon. He also undertook both the implementation and the writing tasks associated with this chapter. The research work of this chapter was conducted under the supervision of both Peter Spreij and Asma Khedher.

BIBLIOGRAPHY

Abbas-Turki, L.A., S. Crépey, and B. Diallo (2018). "XVA principles, nested Monte Carlo strategies, and GPU optimization". In: *International Journal of Theoretical and Applied Finance* 21.6.

Abbott, S. (2015). *Understanding Analysis*. New York: Springer.

Agueh, M. and G. Carlier (2011). "Barycenters in the Wasserstein space". In: *SIAM Journal on Mathematical Analysis* 43.2, pp. 904–924.

Albanese, C., S. Crépey, R. Hoskinson, and B. Saadeddine (2020). "XVA analysis from the balance sheet". In: *Quantitative Finance* 21.1, pp. 99–123.

Albrecher, H., F. Guillaume, and W. Schoutens (2013). "Implied liquidity: Model sensitivity". In: *Journal of Empirical Finance* 23, pp. 48–67.

Altman, E.I., B. Brooks, A. Resti, and A. Sironi (2005). "The link between default and recovery rates: Theory, empirical evidence and implications". In: *The Journal of Business* 78.6, pp. 2203–2228.

Altschuler, J.M. and E. Boix-Adserà (2021). "Wasserstein barycenters can be computed in polynomial time in fixed dimension". In: *Journal of Machine Learning Research* 22.44, pp. 1–19.

– (2022). "Wasserstein barycenters are NP-hard to compute". In: *SIAM Journal on Mathematics of Data Science* 4.1, pp. 179–203.

Ambrosio, L., N. Gigli, and G. Savaré (2005). *Gradient flows in metric spaces and in the space of probability measures*. Basel: Birkhäuser Verlag.

Ammann, M. (1999). *Pricing derivative credit risk*. Berlin: Springer-Verlag.

An, B.-J., A. Ang, T.G. Bali, and N. Cakici (2014). "The joint cross section of stocks and options". In: *The Journal of Finance* 69.5, pp. 2279–2337.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). "Coherent measures of risk". In: *Mathematical Finance* 9.3, pp. 203–228.

Awoyemi, J.O., A.O. Adetunmbi, and S.A. Oluwadare (2017). "Credit card fraud detection using machine learning techniques: A comparative analysis". In: *2017 International Conference on Computing Networking and Informatics (ICCNI)*, pp. 1–9.

Bayer, C., P. Friz, and J. Gatheral (2016). "Pricing under rough volatility". In: *Quantitative Finance* 16.6, pp. 887–904.

Bennedsen, M., A. Lunde, and M.S. Pakkanen (2017). "Hybrid scheme for Brownian semistationary processes". In: *Finance and Stochastics* 21, pp. 931–965.

Berd, A.M. (2005). "Recovery swaps". In: *Journal of Credit Risk* 1.3, pp. 61–70.

– (2011). "A guide to modeling credit term structures". In: *The Oxford handbook of credit derivatives*. Ed. by A. Lipton and A. Rennie. Oxford: Oxford University Press.

Biagini, S. and J. Bion-Nadal (2014). "Dynamic quasi concave performance measures". In: *Journal of Mathematical Economics* 55, pp. 143–153.

Bielecki, T.R., I. Cialenco, and T. Chen (2015). "Dynamic conic finance via backward stochastic difference equations". In: *SIAM Journal of Financia Mathematics* 6.1, pp. 1068–1122.

Bielecki, T.R., I. Cialenco, I. Iyigunler, and R. Rodriguez (2013). "Dynamic conic finance: Pricing and hedging in market models with transaction costs via dynamic coherent acceptability indices". In: *International Journal of Theoretical and Applied Finance* 16.1.

Bielecki, T.R., I. Cialenco, and M. Pitera (2017). "A survey of time consistency of dynamic risk measures and dynamic performance measures in discrete time: LM-measure perspective". In: *Probability, Uncertainty and Quantitative Risk* 2.3.

– (2018). "A unified approach to time consistency of dynamic risk measures and dynamic performance measures in discrete time". In: *Mathematics of Operations Research* 43.1, pp. 204–221.

Bielecki, T.R., I. Cialenco, and Z. Zhang (2014). "Dynamic coherent acceptability indices and their applications to finance". In: *Mathematical Finance* 24.3, pp. 411–441.

Bielecki, T.R. and M. Rutkowski (2010). *Credit risk: Modeling, valuation and hedging*. Berlin: Springer-Verlag.

Bion-Nadal, J. (2009). "Bid-ask dynamic pricing in financial markets with transaction costs and liquidity risk". In: *Journal of Mathematical Economics* 45.11, pp. 738–750.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Singapore: Springer.

Black, F. and J.C. Cox (1976). "Valuing corporate securities: Some effects of bond indenture provisions". In: *The Journal of Finance* 31.2, pp. 351–367.

Borchani, H., G. Varando, C. Bielza, and P. Larrañaga (2015). "A survey on multi-output regression". In: *WIREs Data Mining and Knowledge Discovery* 5.5, pp. 216–233.

Bottou, L. and O. Bousquet (2007). "The tradeoffs of large scale learning". In: *Advances in Neural Information Processing Systems*. Ed. by J. Platt, D. Koller, Y. Singer, and S. Roweis. Vol. 20.

Brigo, D., M. Morini, and M. Tarenghi (2010). "Credit calibration with structural models and equity return swap valuation under counterparty risk". In: *Credit risk frontiers: Subprime crisis, pricing and hedging, CVA, MBS, ratings and liquidity*. Ed. by T. Bielecki, D. Brigo, and F. Patras. Bloomberg Press.

Brigo, D. and A. Pallavicini (2014). "Nonlinear consistent valuation of CCP cleared or CSA bilateral trades with initial margins under credit, funding and wrong-way risks". In: *Journal of Financial Engineering* 1.1.

Buehler, H. (2010). "Volatility and dividends – Volatility modelling with cash dividends and simple credit risk". SSRN Working paper.

Buehler, H., L. Gonon, J. Teichmann, and B. Wood (2019). "Deep hedging". In: *Quantitative Finance* 19.8, pp. 1271–1291.

Cao, C., T. Simin, and H. Xiao (2020). "Predicting the equity premium with the implied volatility spread". In: *Journal of Financial Markets* 51.

Carmona, R. and S. Nadtochiy (2008). "Local volatility dynamic models". In: *Finance and Stochastics* 13, pp. 1–48.

Carr, P., D.B. Madan, M. Melamed, and W. Schoutens (2016). "Hedging insurance books". In: *Insurance: Mathematics and Economics* 70, pp. 364–372.

Chataigner, M., A. Cousin, S. Crépey, M. Dixon, and D. Gueye (2022). "Beyond surrogate modeling: Learning the local volatility via shape constraints". ArXiv working paper.

Chen, Z., T. Chen, and M. Davison (2005). "Choquet expectation and Peng's g-expectation". In: *The Annals of Probability* 33.3, pp. 1179–1199.

Chen, Z., K. He, and R. Kulperger (2013). "Risk measures and nonlinear expectations". In: *Journal of Mathematical Finance* 3.3, pp. 383–391.

Chen, Z., L.D. Van Khoa, E.N. Teoh, A. Nazir, E.K. Karuppiah, and K.S. Lam (2018). "Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: A review". In: *Knowledge and Information Systems* 57, pp. 245–285.

Cherny, A. and D.B. Madan (2009). "New measures for performance evaluation". In: *The Review of Financial Studies* 22.7, pp. 2571–2606.

– (2010). "Markets as a counterparty: An introduction to conic finance". In: *International Journal of Theoretical and Applied Finance* 13.8, pp. 1149–1177.

Chopra, S. (2020). "Reinforcment learning methods for conic finance". PhD thesis. University of Maryland.

Choquet, G. (1954). "Theory of capacities". In: *Annales de l'Institut Fourier* 5, pp. 131–295.

Chourdakis, K., E. Epperlein, M. Jeannin, and J. McEwen (2013a). "A cross-section across CVA". In: *Nomura*.

– (2013b). "A cross-section for CVA". In: *Risk* 26.3, pp. 20–21.

Corcuera, J.M., F. Guillaume, D.B. Madan, and W. Schoutens (2012). "Implied liquidity: Towards stochastic liquidity modeling and liquidity trading". In: *International Journal of Portfolio Analysis and Management* 1.1, pp. 80–91.

Cornuejols, G., J. Peña, and R. Tütüncü (2018). *Optimization Methods in Finance*. Cambridge: Cambridge University Press.

Crépey, S. and M. Dixon (2019). "Gaussian process regression for derivative portfolio modeling and application to CVA computations". ArXiv Working paper.

Crépey, S., R. Hoskinson, and B. Saadeddine (2019). "Balance sheet XVA by deep learning and GPU". Working paper.

Cuchiero, C., W. Khosrawi, and J. Teichmann (2020). "A generative adversarial network approach to calibration of local stochastic volatility models". In: *Risks* 8.4.

Das, S.R. and P. Hanouna (2009). "Implied recovery". In: *Journal of Economic Dynamics and Control* 33.11, pp. 1837–1857.

Davis, J., L. Devos, S. Reyners, and W. Schoutens (2021). "Gradient boosting for quantitative finance". In: *Jounal of Computational Finance* 24.4, pp. 1–40.

De Spiegeleer, J., C. Van Hulle, and W. Schoutens (2014). *The handbook of hybrid securities: Convertible bonds, CoCo bonds and bail-in*. Chichester: Wiley.

De Spiegeleer, J., D.B. Madan, S. Reyners, and W. Schoutens (2018). "Machine learning for quantitative finance: Fast derivative pricing, hedging and fitting". In: *Quantitative Finance* 18.10, pp. 1635–1643.

Delbaen, F. (2002). "Coherent risk measures on general probability spaces". In: *Advances in Finance and Stochastics*. Ed. by K. Sandmann and P.J. Schönbucher. Berlin: Springer, pp. 1–37.

Denneberg, D. (1994). *Non-additive measure and integral*. Dordrecht: Kluver Academic Publishers.

Dixon, M.F., I. Halperin, and P. Bilokon (2020). *Machine learning in finance. From theory to practice*. Cham: Springer.

Domashova, J. and N. Mikhailina (2021). "Usage of machine learning methods for early detection of money laundering schemes". In: *Procedia Computer Science* 190, pp. 184–192.

Doran, J.S., A. Fodor, and D. Jiang (2013). "Call-put implied volatility spreads and option returns". In: *The Review of Asset Pricing Studies* 3.2, pp. 258–290.

Driouchi, T., L. Trigeorgis, and Y. Gao (2015). "Choquet-based European option pricing with stochastic (and fixed) strikes". In: *OR Spectrum* 37, pp. 787–802.

Duchi, J., E. Hazan, and Y. Singer (2011). "Adaptive subgradient methods for online learning and stochastic optimization". In: *Journal of Machine Learning Research* 12, pp. 2121–2159.

Duffie, D. and K.J. Singleton (1999). "Modeling the term structure of defaultable bonds". In: *The Review of Financial Studies* 12.4, pp. 687–720.

Dupire, B. (1994). "Pricing with a smile". In: *Risk* 7.1, pp. 18–20.

EBA (2013). *EBA final draft regulatory technical standards on credit valuation adjustment risk for the determination of a proxy spread and the specification of a limited number of smaller portfolios under article 383(7) of regulation (EU) no 575/2013 (capital requirements regulation – CRR)*.

– (2015). *EBA final draft regulatory technical standards on prudent valuation under article 105(14) of regulation (EU) no 575/2013 (capital requirements regulation – CRR)*.

Eberlein, E., T. Gehrig, and D.B. Madan (2012). "Pricing to acceptability: With applications to valuation of one's own credit risk". In: *Journal of Risk* 15.1, pp. 91–120.

Ferguson, R. and A. Green (2018). "Deeply learning derivatives". ArXiv Working paper.

Föllmer, H. and A. Schied (2016). *Stochastic finance. An introduction in discrete time*. 4th edition. Berlin: De Gruyter.

Fritsch, F.N. and R.E. Carlson (1980). "Monotone piecewise cubic interpolation". In: *SIAM Journal on Numerical Analysis* 17.2, pp. 238–246.

Fu, X., Y.E. Arisoy, M.B. Shackleton, and M. Umutlu (2016). "Option-implied volatility measures and stock return predictability". In: *The Journal of Derivatives* 24.1, pp. 58–78.

Fukushima, K. (1975). "Cognitron: A self-organizing multilayered neural network". In: *Biological Cybernetics* 20, pp. 121–136.

Gambetti, P., G. Gauthier, and F. Vrins (2018). "Stochastic recovery rate: Impact of pricing measure's choice and financial consequences on single-name products". In: *New methods in fixed income modeling. Contributions to management science*. Ed. by M. Mili, R. Samaniego Medina, and F. di Pietro. Cham: Springer.

Gatheral, J. (2006). *The volatility surface: A practitioner's guide*. Hoboken: John Wiley & Sons.

Gatheral, J., T. Jaisson, and M. Rosenbaum (2018). "Volatility is rough". In: *Quantitative Finance* 18.6, pp. 933–949.

Glorot, X., A. Bordes, and Y. Bengio (2011). "Deep sparse rectifier neural networks". In: *Proceedings of Machine Learning Research*. Vol. 15. JMLR Workshop and Conference Proceedings, pp. 315–323.

Grabisch, M. and M. Ridaoui (2016). "Choquet integral calculus on a continuous support and its applications". In: *Operations Research and Decisions* 26.1, pp. 73–93.

Green, A. (2015). *XVA: Credit, funding and capital valuation adjustments*. Chichester: Wiley.

Gregory, J. (2020). *The xVA Challenge: Counterparty risk, funding, collateral, capital and initial margin*. Chichester: Wiley.

Guillaume, F. (2015). "The LIX: A model-independent liquidity index". In: *Journal of Banking & Finance* 58, pp. 214–231.

Guillaume, F. and W. Schoutens (2013). "Bid-ask spread for exotic options under conic finance". In: *Innovations in Quantitative Risk Management*. Ed. by K. Glau, M. Scherer, and R. Zagst. New York: Springer, pp. 59–74.

Guo, G. and J. Obłój (2019). "Computational methods for martingale optimal transport problems". In: *Annals of Applied Probability* 29.6, pp. 3311–3347.

Guyon, J. and P. Henry-Labordère (2011). "The smile calibration problem solved". In: *Risk Magazine*. SSRN working paper.

Hagan, P.S., D. Kumar, A. Lesniewski, and D.E. Woodward (2014). "Arbitrage-free SABR". In: *Wilmott* 2014.69, pp. 60–75.

Hagan, P.S., D. Kumar, Andrew S. Lesniewski, and D.E. Woodward (2002). "Managing smile risk". In: *Wilmott* 1, pp. 84–108.

Hasan, I., M. Marra, T.Y. To, E. Wu, and G. Zhang (2023). "COVID-19 pandemic and global corporate CDS spreads". In: *Journal of Banking & Finance* 147. Special Issue: The Impact of Global Pandemic on Financial Markets and Institutions.

Haug, E.G. (2007). *The complete guide to option pricing formulas*. New York: McGraw-Hill.

Henrard, M. (2017). *Algorithmic differentiation in finance explained*. Cham: Springer.

Hernandez, A. (2017a). "Model calibration with neutal networks". In: *Risk*.
– (2017b). "Model calibration: Global optimizer vs. neural network". SSRN working paper.
Heston, S.L. (1993). "A closed-form solution for options with stochastic volatility with applications to bond and currency options". In: *The Review of Financial Studies* 6.2, pp. 327–343.
Heynderickx, W., J. Cariboni, W. Schoutens, and B. Smits (2016). "The relationship between risk-neutral and actual default probabilities: The credit risk premium". In: *Applied Economics* 48.42, pp. 4066–4081.
Huang, J., J. Chai, and S. Cho (2020). "Deep learning in finance and banking: A literature review and classification". In: *Frontiers of Business Research in China* 12.13.
Hühnerbein, R., F. Savarino, F. Åström, and C. Schnörr (2018). "Image labeling based on graphical models using Wasserstein messages and geometric assignment". In: *SIAM Journal on Imaging Sciences* 11.2, pp. 1317–1362.
ISDA (1998). *Confirmation of OTC credit swap transaction single reference entity – Non-sovereign*.
– (2003). *2003 ISDA credit derivatives definitions*.
– (2014). *2014 ISDA credit derivatives definitions*.
– (2015). *Frequently asked questions. Amending when single name CDS roll to new on-the-run contracts: December 20, 2015 go-live*.
Jarrow, R.A. and S.M. Turnbull (1995). "Pricing derivatives on financial securities subject to credit risk". In: *The Journal of Finance* 50.1, pp. 53–85.
Johri, P., J.K. Verma, and S. Paul (2020). *Applications of machine learning*. Singapore: Springer.
Junge, B. and A.B. Trolle (2015). "Liquidity risk in credit default swap markets". In: *Swiss Finance Institute Research Paper* 13–65.
Kaelo, P. and M.M. Ali (2007). "Integrated crossover rules in real coded genetic algorithms". In: *European Journal of Operational Research* 176.1, pp. 60–76.
Karimi, A., L. Ripani, and T.T. Georgiou (2020). "Statistical learning in Wasserstein space". In: *IEEE Control Systems Letters* 5.3, pp. 899–904.
Kast, R., A. Lapied, and D. Roubaud (2014). "Modelling under ambiguity with dynamically consistent Choquet random walks and Choquet-Brownian motions". In: *Economic Modelling* 38, pp. 495–503.
Kingma, D.P. and J. Ba (2014). "Adam: A method for stochastic optimization". ArXiv working paper.
Knight, F. (1921). *Risk, uncertainty and profit*. New York: Harper.
Kullback, S. and R.A. Leibler (1951). "On information and sufficiency". In: *The Annals of Mathematical Statistics* 22.1, pp. 79–86.
Le Cam, L. (1986). *Asymptotic methods in statistical decision theory*. New York: Springer-Verlag.

Lee, C.-W. and C.-K. Kuo (2015). "Combining hazard rates with the CreditGrades model: A hybrid method to value CDS contracts". In: *International Journal of Financial Engineering* 2.4.

Liu, S., A. Borovykh, L.A. Grzelak, and C.W. Oosterlee (2019). "A neural network-based framework for financial model calibration". In: *Journal of Mathematics in Industry* 9.9.

Liu, S., C.W. Oosterlee, and S.M. Bohte (2019). "Pricing options and computing implied volatilities using neural networks". In: *Risks* 7.1.

Longstaff, F.A., J. Pan, L.H. Pedersen, and K.J. Singleton (2011). "How sovereign is sovereign credit risk?" In: *American Economic Journal: Macroeconomics* 3.2, pp. 75–103.

Madan, D.B. (2009). "Capital requirements, acceptable risks and profit". In: *Quantitative Finance* 9.7, pp. 767–773.

– (2012). "From credit valuation adjusemtnts to credit capital commitments". In: *Quantitative Finance* 12.6, pp. 839–845.

– (2014). "Modeling and monitoring risk acceptability in markets: The case of the credit default swap market". In: *Journal of Banking & Finance* 47, pp. 63–73.

– (2016). "Conic portfolio theory". In: *International Journal of Theoretical and Applied Finance* 19.3.

Madan, D.B., S. Reyners, and W. Schoutens (2019). "Advanced model calibration on Bitcoin options". In: *Digital Finance* 1, pp. 117–137.

Madan, D.B. and W Schoutens (2017). "Conic Option Pricing". In: *The Journal of Derivatives* 25 (1), pp. 10–36.

– (2011a). "Conic coconuts: The pricing of contingent capital notes using conic finance". In: *Mathematics and Financial Economics* 4, pp. 87–106.

– (2011b). "Conic finance and the corporate balance sheet". In: *International Journal Theoretical and Applied Finance* 14.5, pp. 587–610.

– (2012). "Structured products equilibria in conic two price markets". In: *Mathematics and Financial Economics* 6, pp. 37–57.

– (2016a). *Applied conic finance*. Cambridge: Cambridge University Press.

– (2016b). "Conic CVA and DVA". In: *Wilmott Magazine* 82, pp. 45–50.

Madan, D.B. and Y.M. Sharaiha (2020). "Machine trading: Theory, advances, and applications". In: *The Journal of Financial Data Science* 2.3, pp. 8–24.

Malagó, L., L. Montrucchio, and G. Pistone (2018). "Wasserstein Riemannian geometry of Gaussian densities". In: *Information Geometry* 1, pp. 137–179.

Mandelbrot, B.B. and J.W. van Ness (1968). "Fractional Brownian motions, fractional noises and applications". In: *SIAM Review* 10.4, pp. 422–437.

Markit (2009a). *CDS small bang: Understanding the global contract & European convention changes*.

– (2009b). *The CDS big bang: Understanding the changes to the global CDS contract and North American conventions*.

Marquet, I. and W. Schoutens (2018). "Conic CPPIs". In: *International Journal of Theoretical and Applied Finance* 21.2.

Mazzon, A. and A. Pascucci (2017). "The forward smile in local–stochastic volatility models". In: *The Journal of Computational Finance* 20.3, pp. 1–29.

McCrickerd, R. and M.S. Pakkanen (2018). "Turbocharging Monte Carlo pricing for the rough Bergomi model". In: *Quantitative Finance* 18.11, pp. 1877–1886.

Merton, R.C. (1974). "On the pricing of corporate debt: The risk structure of interest rates". In: *The Journal of Finance* 29.2, pp. 449–470.

Mesiar, R., J. Li, and E. Pap (2010). "The Choquet integral as Lebesgue integral and related inequalities". In: *Kybernetika* 46.6, pp. 1098–1107.

Michielon, M., D. Franquinho, A. Gentile, A. Khedher, and P.J.C. Spreij (2023). "Neural network empowered liquidity pricing in a two-price economy under conic finance settings using multilayer perceptrons". Revised version to appear in *Quantitative Finance*.

Michielon, M., A. Khedher, and P.J.C. Spreij (2021a). "From bid-ask credit default swap quotes to risk-neutral default probabilities using distorted expectations". In: *International Journal of Theoretical and Applied Finance* 24.3.

– (2021b). "Liquidity-free implied volatilities: An approach using conic finance". In: *International Journal of Financial Engineering* 8.4.

– (2022a). "On Wasserstein distances, barycenters, and the cross-section methodology for proxy credit curves".

– (2022b). "Proxying credit curves via Wasserstein distances". In: *Annals of Operations Research*.

Morelli, M.J., G. Montagna, O. Nicrosini, M. Treccani, M. Farina, and P. Amato (2004). "Pricing financial derivatives with neural networks". In: *Physica A: Statistical Mechanics and its Applications* 338.1, pp. 160–165.

Moussa, K. (2018). "Arbitrage-based filtering of option price data". SSRN working paper.

Musiela, M. and M. Rutkowski (2005). *Martingale methods in financial modelling*. Berlin: Springer-Verlag.

Ni, H., X. Dong, J. Zheng, and G. Yu (2021). *An intoduction to machine learning in quantitative finance*. London: World Scientific Publishing Europe Ltd.

O'Kane, D. and S. Turnbull (2003). "Valuation of credit default swaps". In: *Lehman Brothers Quantitative Credit Research Quarterly* Q1-Q2.

Oh, K.-S. and K. Jung (2004). "GPU implementation of neural networks". In: *Pattern Recognition* 37.6, pp. 1311–1314.

Overhaus, M., A. Bermúdez, H. Buehler, A. Ferraris, C. Jordison, and A. Lamnouar (2007). *Equity hybrid derivatives*. Hoboken: John Wiley & Sons.

Pan, J. and K.J. Singleton (2008). "Default and recovery implicit in the term structure of sovereign CDS spreads". In: *The Journal of Finance* 63.5, pp. 2345–2384.

Panaretos, V.M. and Y. Zemel (2019). "Statistical aspects of Wasserstein distances". In: *Annual Review of Statistics and Its Applications* 6, pp. 405–431.

– (2020). *An Invitation to Statistics in Wasserstein Space*. Cham: Springer.

Peng, S.G. (1997). "Backward SDE and related g-expectation". In: *Pitman Research Notes in Mathematics* 364, pp. 141–159.

Peyré, G. and M. Cuturi (2019). "Computational optimal transport: With applications to data science". In: *Foundations and trends in machine learning* 11.5–6, pp. 355–607.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery (2007). *Numerical recipes. The art of scientific computing*. Cambridge: Cambridge University Press.

Rosazza Gianin, E. and C. Sgarra (2013). "Acceptability indexes via g-expectations: An application to liquidity risk". In: *Mathematics and Financial Economics* 7.4, pp. 457–475.

Ruiz, I. (2015). *XVA desks – A new era for risk management*. New York: Palgrave.

Rundo, F., F. Trenta, A.L. di Stallo, and S. Battiato (2019). "Machine learning for quantitative finance applications: A survey". In: *Applied Sciences* 9.24.

Sadgali, I., N. Sael, and F. Benabbou (2019). "Performance of machine learning techniques in the detection of financial frauds". In: *Procedia Computer Science* 148, pp. 45–54.

Schmeidler, D. (1986). "Integral representation without additivity". In: *Proceedings of the American Mathematical Society* 97.2, pp. 255–261.

– (1989). "Subjective probability and expected utility without additivity". In: *Econometrica* 57.3, pp. 571–587.

Schoutens, W., E. Simons, and J. Tistaert (2003). "A perfect calibration! Now what?" In: *The Best of Wilmott* 281.

Sornette, D., P. Cauwels, and G. Smilyanov (2018). "Can we use volatility to diagnose financial bubbles? Lessons from 40 historical bubbles". In: *Quantitative Finance and Economics* 2.1, pp. 486–594.

Sourabh, S., M. Hofer, and D. Kandhai (2018). "Liquidity risk in derivatives valuation: An improved credit proxy method". In: *Quantitative Finance* 18.3, pp. 467–481.

Szakmary, A., E. Ors, J.K. Kim, and W.N. Davidson (2003). "The predictive power of implied volatility: Evidence from 35 futures markets". In: *Journal of Banking & Finance* 27.11, pp. 2151–2175.

Szandała, T. (2021). "Review and comparison of commonly used activation functions for deep neural networks". In: *Bio-inspired Neurocomputing*. Ed. by A.K. Bhoi, P.K. Mallick, C.-M. Liu, and V.E. Balas. Singapore: Springer Singapore, pp. 203–224.

Ulrich, M. and S. Walther (2020). "Option-implied information: What's the vol surface got to do with it?" In: *Review of Derivatives Research* 23, pp. 323–355.

van Bakel, S., S. Borovkova, and M. Michielon (2020). "Conic CVA and DVA for option portfolios". In: *International Journal of Theoretical and Applied Finance* 23.5.

van der Stoep, A.W., L.A. Grzelak, and C.W. Oosterlee (2015). "The time-dependent FX-SABR model: Efficient calibration based on effective parameters". In: *International Journal of Theoretical and Applied Finance* 18.6.

Varna, P., R. Cantor, and D. Hamilton (2003). "Recovery rates on defaulted corporate bonds and preferred stocks". In: *Moody's Investors Service*.

Villani, C. (2003). *Topics in optimal transportation*. Providence: American Mathematical Society.
– (2009). *Optimal transport: Old and new*. Berlin: Springer-Verlag.
Wales, D.J. and J.P.K. Doye (1997). "Global optimization by Basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms". In: *The Journal of Physical Chemistry A* 101.28, pp. 5111–5116.
Wang, S.S. (2000). "A class of distortion operators for pricing financial and insurance risks". In: *Journal of Risk and Insurance* 67.1, pp. 15–36.
Wang, Y.E., G.-Y. Wei, and D. Brooks (2019). "Benchmarking TPU, GPU, and CPU platforms for deep learning". ArXiv working paper.
Wang, Z. and G.J. Klir (2009). *Generalized measure theory*. New York: Springer.
Wang, Z., N. Privault, and C. Guet (2022). "Deep self-consistent learning of local volatility". ArXiv woring paper.
White, R. (2013). "The pricing and risk management of credit default swaps, with a focus on the ISDA model". In: *OpenGamma Quantitative Research* 16.
Xiang, Y., S. Gubian, B. Suomela, and J. Hoeng (2013). "Generalized simulated annealing for global optimization: The GenSA Package". In: *The R Journal* 5.1, pp. 13–29.
Yang, S.Y., Q. Qiao, P.A. Beling, W.T. Scherer, and A.A. Kirilenko (2015). "Gaussian process-based algorithmic trading strategy identification". In: *Quantitative Finance* 15.10, pp. 1683–1703.
Ye, T. and L. Zhang (2019). "Derivatives pricing via machine learning". In: *Journal of Mathematical Finance* 9.3, pp. 561–589.
Zeier, M.D. (2015). "ADADELTA: An adaptive learning rate method". ArXiv working paper.
Zhang, X., J. Zhong, and K. Liu (2021). "Wasserstein autoencoders for collaborative filtering". In: *Neural Computing and Applications* 33, pp. 2793–2802.
Zou, L. and W. Li (2022). "Estimating actual probability of default from structural models". In: *International Journal of Financial Engineering* 9.1.

10 ACKNOWLEDGEMENTS

This is a list of the people I would like to thank, though it may not be exhaustive. I apologize in advance if I have overlooked anyone, I hope I haven not missed too many.

First of all, I would like to express my gratitude to my supervisors, Peter and Asma, for their guidance and support, and for allowing me the flexibility to choose my research topics.

I owe a special thanks to my parents, Guido and Rita, for their unwavering encouragement and support, and to my wonderful sisters Beatrice and Elisabetta, for always being there when I need them. Speaking of family, thank you also to Gualtiero, Pietro, Susi, and Tobia.

To my paranymphs, Alessandro and Karim, I feel incredibly fortunate to count on you as friends. I also want to thank my dear and long-time friends Antonio, Carlo, Davide, Laura, Luca, and Maddalena for their constant support. Thank you to Ezra, Felix, Fredrik, Georgios, Hanae, Jaap, Jiayi, Loizos, Nicoletta, Norina, Sander, Sonja, Vil, Xiang, and Yang Lu for your friendship. I consider myself very lucky to have you all in my life. I also appreciate Diogo for his great collaboration on my last project, Riccardo for his wise suggestions, and Gergely, Imre, and Laszlo.

My sincere thanks go to my manager, Leo, for his support over the past few years. I also want to thank my former managers Geert and Tim, and my current and former colleagues Andrei, Arnout, Chris, Frits, Nancy, Patrick, and Vincent, with special thanks to Ana, André, Jelle, Jovanka, and Raoul.

I am grateful to Diederik, Jacqueline, Marieke, Roger, and all the others from TopQuants, with whom I have collaborated over the years.

A special mention goes to Pierandrea Bon for sparking my curiosity in the world of mathematics, to Bruno Girotto for introducing me to financial mathematics, and to Silvano Holzer for simply being the best teacher I could have ever hoped for.

Finally, I would like to thank Camillo Colleluori and Raphael Saini for being exceptional drum teachers, as well as my all-time idol Pete "Commando" Sandoval: after all these years, you continue to dominate; ageless, still you are.

COLOPHON

This document has been typeset with LaTeX using ArsClassica, a reworking of the ClassicThesis style designed by André Miede, developed by Lorenzo Pantieri.