DEUTSCHE
BUNDESBANK
EUROSYSTEM

Discussion Paper

Deutsche Bundesbank
No 43/2014

Updating the option implied probability of default methodology

Johannes Vilsmeier

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet http://www.bundesbank.de

ISBN 978–3–95729–106–6 (Printversion)
ISBN 978–3–95729–107–3 (Internetversion)

Non-technical summary

Research Question

We propose some technical modifications to the option implied probability of default (option iPoD) methodology suggested in Capuano (2008). The framework allows to estimate a probability distribution for a firm's/bank's future value of assets in such a way that an implicit probability of default for the firm/bank can be derived. For the estimation process solely the information from equity option prices is needed. Despite its general attractiveness, the application of the originally suggested form of the option iPoD framework is difficult to apply due to computational problems that arise in the estimation of the asset distribution and the PoD. This paper largely solves these problems.

Contribution

We stabilize the estimation process of the asset distribution by deriving an alternative objective function which can be solved for arbitrary applications. For the new objective function we can solve the involved integrals analytically which further stabilizes the estimation procedure. Moreover, we propose a new algorithm to determine the PoD after showing that the algorithm used in Capuano (2008) gives rise to arbitrary results. Finally, we carry out comprehensive numerical evaluations for the updated framework.

Results

Numerical evaluations for the suggested framework - using a large number of hypothetical PoDs and asset distributions - show the accuracy of the algorithms. In addition, we provide an illustrative empirical application of the updated option iPoD framework by applying it to real world option data and show that the methodology was able to anticipate the downgrading of Bank of America by Moody's in 2011.

Nicht-technische Zusammenfassung

Forschungsfrage

In diesem Papier werden technische Modifikationen für die in Capuano (2008) vorgestellte statistische Methodik zur Schätzung von Ausfallwahrscheinlichkeiten aus Optionspreisen vorgeschlagen. Diese Methodik erlaubt es, eine Wahrscheinlichkeitsverteilung für den zukünftigen Vermögenswert einer Firma/Bank zu schätzen, wobei hieraus implizit eine Ausfallwahrscheinlichkeit für die jeweilige Firma/Bank abgeleitet werden kann. Als Daten für die Schätzung der Verteilung werden ausschließlich Aktienoptionen zu verschiedenen Ausübungspreisen benötigt. Trotz der generellen Attraktivität des Ansatzes ist die Anwendbarkeit der Methodik – in der Form wie sie in Capuano (2008) vorgeschlagen wurde – stark eingeschränkt, da es technische Probleme bei der Schätzung der optimalen Vermögensverteilung sowie der Ausfallwahrscheinlichkeit gibt. Dieses Papier löst diese Probleme weitestgehend.

Beitrag

Wir stabilisieren die Schätzungen durch die Herleitung einer neuen Zielfunktion, welche Lösungen für beliebige Anwendungen liefert. Die Integrale der neuen Zielfunktion können analytisch gelöst werden, was die Schätzungen zusätzlich stabilisiert und die Präzision der Ergebnisse erhöht. Weiterhin wird ein alternativer Algorithmus zur Bestimmung der Ausfallwahrscheinlichkeit aus der Vermögensverteilung entwickelt, da der ursprünglich von Capuano (2008) vorgeschlagene Algorithmus die Generierung weitgehend beliebiger Ergebnisse begünstigt.

Ergebnisse

Numerische Evaluationen – unter Verwendung einer Vielzahl von hypothetischen Ausfallwahrscheinlichkeiten und Wahrscheinlichkeitsverteilungen für den Vermögenswert – zeigen, dass die aktualisierte Methodik gute Schätzungen für die Ausfallwahrscheinlichkeit ermöglicht. Zudem wenden wir den Ansatz auf reale Optionsdaten an und geben dadurch ein illustratives Beispiel für die praktische Relevanz des Ansatzes. Es zeigt sich, dass die ermittelten Ausfallwahrscheinlichkeiten frühzeitig eine klare Indikation für den durch Moody's im Jahr 2011 vorgenommenen Downgrade der Bank of America geben konnten.

Bundesbank Discussion Paper No 43/2014

Updating the Option Implied Probability of Default Methodology [‡]

Johannes Vilsmeier[§]
Deutsche Bundesbank

Abstract

In this paper we 'update' the option implied probability of default (option iPoD) approach recently suggested in the literature. First, a numerically more stable objective function for the estimation of the risk neutral density is derived whose integrals can be solved analytically. Second, it is reasoned that the originally proposed approach for the estimation of the PoD produces arbitrary results and hence an alternative procedure is suggested that is based on the Lagrange multipliers. Based on numerical evaluations and an illustrative empirical application we conclude that the framework provides very promising results.

Keywords: Option Implied Probability of Default, Risk Neutral Density, Cross Entropy.

JEL classification: C51, C52, C61, G12, G24, G32.

[‡]This paper will be published in the Journal of Computational Finance in the course of 2015.
The author thanks the FAZIT-STIFTUNG Gemeinnützige Verlagsgesellschaft mbH and the Bavarian Graduates Program in Economics (BGPE) for financial support during the writing process. He also thanks Professor Rolf Tschernig for insightful comments and valuable support. Any errors, misrepresentations, and omissions are my own. The views expressed in this paper are those of the authors and do not necessarily reflect the opinions of the Deutsche Bundesbank.
[§]Deutsche Bundesbank, D-60431 Frankfurt am Main, Germany, johannes.vilsmeier@bundesbank.de.

1 Introduction

This paper deals with methodological issues concerning the so-called option implied probability of default (option iPoD) approach suggested in Capuano (2008). In the option iPoD approach option implied risk neutral densities (RNDs) are estimated in such a way that they reveal information about the probability of default (PoD) of the issuer of the underlying. Default in this framework corresponds to the event that the stock price of the firm falls to zero during the time to maturity of the options. As shown in this paper, the approach derives the PoDs in a purely statistical way, and neither accounting data nor recovery rate assumptions are needed. This contrasts the methodology to traditional concepts of PoD estimation, which in general are either based on the so called 'structural approach' of Merton (1974) (see eg Crosbie and Kocagil (2003), J.P.Morgan (1997)) or use debt based financial instruments, like Credit Default Swaps (CDS) or bonds (see eg Chan-Lau (2006)).

In structural approaches the PoDs are measured as the probability that the asset value of a firm falls below the value of its debt (default barrier). The main drawback of structural approaches is that they are based on historical data. While the firm's asset distribution is derived from historical stock prices, the default barrier is calculated from accounting data. On the one hand this leads to properties of the asset distribution that are mostly based on investors' expectations in the past, and on the other hand to simplified definitions of default points based on the book value of liabilities. Especially for firms with liquid assets, volatile leverages and highly complex and opaque capital structures (including eg off balance liabilities) the structural approach is too restrictive and hence hard to apply. Particularly critical in this context are applications to financial institutions, as eg pointed out in Crosbie and Kocagil (2003). In theory more appealing is to use information in prices of debt based market instruments (credit spreads), like for CDS or bonds. It is generally agreed on that it is extremely difficult to permanently 'beat the market', meaning that the degree to which information is processed by the market is hardly outperformed by any single model. The problem, though, is that in order to derive the PoDs from credit spreads, it is necessary to make assumptions about the recovery rate in the case of default. To assign reliable values to the recovery rate is difficult in practice since especially in the case of financial institutions the volatility of asset values and liabilities makes it hard to predict firm values in the event of default. While this difficulty in practice is often neglected by assuming constant recovery rates across firms and over time, the uncertainty about the recovery rates has severe impact on the information content of extracted PoDs since we do not know whether credit spreads differ/change due to differences/changes in recovery rates or due to the PoDs.

The option iPoD approach overcomes these shortcomings, because it does neither require balance sheet nor recovery rate information, and derives asset distributions using exclusively up-to-date information implied by the daily observed set of option prices. While the framework is applicable to any type of firm for which equity options are available, the framework is particularly well suited for applications to financial institutions due to the problems pointed out above. Capuano (2008) applies the approach to financial firms facing the early stage of the US subprime crisis and presents promising results regarding the signalling power of the framework. Hence, the methodology seems apt to derive financial distress indicators. Moreover, it provides an optimal basis for modelling multivariate distributions of financial systems using copulas, since the entire (marginal) asset distributions of the firms are estimated.*

The idea of the option iPoD methodology is i) to determine for a firm the risk-neutral density of its future asset value from option prices and ii) to specify a sub-domain of the risk-neutral density on which the future asset prices imply a future stock price of zero. The probability assigned to the sub-domain then defines the PoD. In order to implement this idea, a highly flexible RND estimation approach is required. First, it has to allow for an interaction between the shape of the RND and the level of the PoD, such that the combination that best fits the observed prices can be identified. Second, as shown in section 4, it is essential for the PoD algorithm that the RND can take arbitrary shapes at the left tail of the distribution, such that for any length of the sub-domain the PoD can be roughly approximated.

*For copula approaches marginals and the dependence structure between the marginals are needed. The dependence structure could be calculated on basis of estimated time series of marginals.

A straightforward implementation is possible with the so-called cross-entropy approach as suggested in Capuano (2008).[†] The cross entropy approach is a semi-parametric estimation procedure, which uses moment constraints and the entropy concept (Shannon (1948)) to derive densities under limited information. Redefining the RND domain transforms the basic entropy based RND estimation set-up, introduced in Buchen and Kelly (1996),[‡] into the option iPoD framework. The entropy based approach has the appealing property that in the estimation process no more information (in the entropic sense) is imposed as what is known from the data (see eg Jaynes (1957a)).

Despite the depicted general attractiveness of the method proposed by Capuano (2008), it suffers from two methodological problems: First, the optimization process in the RND estimation is numerically very unstable and second, the PoD determination algorithm is rather arbitrary.

The contributions of this paper are: We suggest technical modifications to the original framework which solve the methodological problems to a large degree and hence considerably improve the general applicability of the option iPoD approach. The first modification concerns the optimization procedure for the cross entropy distribution as the approach applied in Capuano (2008) is generally known to be highly unstable due to singularity problems in wide ranges of the relevant parameter space. Following Alhassid, Agmon, and Levine (1978), the search for the roots of a highly non-linear system of equations is transformed into a stable and computationally efficient minimization problem for a strictly convex scalar function. Further, we provide an analytical solution to the integrals of the objective function, such that no numerical methods are necessary.

The second modification concerns the determination of the correct PoD. By analysing the underlying statistical mechanisms and carrying out numerical evaluations we show that the

[†]The idea could in theory also be implemented with modifications of more frequently used RND estimation frameworks, like truncated mixtures of normals, but there would be the additional problem to decide how many mixtures are required in each specific case.

[‡]Other applications of the concept to option pricing are eg Avellaneda (1998), Choe and Jeong (2008), Guo (2001), Neri and Schneider (2012), Stutzer (1996), Stutzer (2000), Rompolis (2010).

procedure suggested by Capuano (2008) for the estimation of the PoD is unreliable and crucially depends on discretionary model parameters. In Capuano (2008) these model parameters were taken from balance sheet information as the framework there is interpreted as a structural approach based model with an endogenous distress barrier. In contrast, we stress that the approach is of a purely statistical nature and that the use of balance sheet information leads to arbitrary results. For this reason we suggest an alternative procedure which is based on the evolution of the Lagrange multipliers when estimating the RND for different default barriers. Despite of the ad hoc nature of our approach, the numerical evaluations show clearly its accuracy. We illustrate the practical use of the updated framework with an empirical application to real world option data.

The remainder of the paper is organized as follows. In section 2 we present the option iPoD framework and contrast our purely statistical point of view to the interpretation given in Capuano (2008). Section 3 presents the basic estimation set-up used in Capuano (2008) and then 'updates‘ the methodology for the estimation of the cross-entropy density. The update comprises a detailed derivation of a new objective function and an analytical solution to the involved integrals. Section 4 provides an analysis of the mechanism that allows for the estimation of the option iPoD followed by an discussion on how to determine the optimal PoD. In this context an ad-hoc procedure is suggested whose accuracy is comprehensively evaluated in section 5. The evaluation comprises numerical examples as well as an illustrative real world application. Section 6 offers some conclusions and prospects on future research.

2 The Option iPoD Framework

We start by presenting the basic idea of the option iPoD framework from a purely statistical point of view, which contrasts the 'structural approach' interpretation given in Capuano (2008). The aim of the option iPoD is to modify traditional RND estimation approaches such that it is possible to estimate a 'mass point' in the RND that indicates the probability that the underlying of a stock option will have value zero at time of maturity of the option. A RND, $f(S_T)$, is a density function that describes the investors' expectations regarding the value of the underlying stock S at time of maturity T implied by the observed option prices for different strikes K. The mass point can be interpreted as a PoD if we assume that a stock price of zero implies a firm's default. In order to be able to estimate the potential 'jump' in the RND (if a PoD exists) at a stock price of zero within a continuous estimation framework a 'trick' has to be applied. The trick is to extend the domain of the RND for the stock price such that all realisations within this additional interval of values imply a future stock price of zero. The integral over the density assigned to this interval corresponds to our PoD estimate. The additional interval of values is obtained by shifting the domain of S_T upwards by some constant D, and estimating $f(V_T)$, with $V_T = S_T + D$. The pay-off for a call option with strike K_i in T using the new domain is now defined by: $C_T^{K_i} = \max(V_T - D - K_i; 0)$, and consequently, there will be no pay-off for asset values within the interval $[0; D]$. In the RND estimation, a density will be assigned to this interval such that an optimal fit to the observed option prices is achieved and the PoD is then given by:

$$PoD(D) = \int_0^D f(V_T) dV_T. \tag{1}$$

As shown in section 4, any PoD level can be approximated if we choose the length of the interval D correctly.

In Capuano (2008) the above described mechanism is interpreted as a 'structural approach' based model (Merton (1974)) with an endogenous distress barrier. In the structural approach a firm's value of assets is given by the value of its debt plus the value of its equity. The firm defaults if the value of assets does not cover the value of debt. Hence, in the PoD mechanism V can be interpreted as value of assets, S as the value of equity and D as the value of debt. Capuano (2008) uses balance sheet information about a firm's debt and value of assets to define bounds on the domain for the RND which, as we will see in section 4, critically influences the search for the optimal D and leads to arbitrary effects on the estimated PoD. Henceforward, we will adopt the 'structural approach' interpretation solely to give a theoretical meaning to the variables involved in the PoD estimation procedure. It is important to note, though, that the assumptions of the structural approach have no implications for our PoD estimation.

To estimate the option implied RND, $f(V_T)$, we will use moment constraints given by the theory of risk neutral pricing (Cox and Ross (1976)) and observed call option prices.[§] The theory of risk neutral pricing postulates that for a given strike K_i the expectation over all possible pay-offs of an option in T (measured in years), discounted with the (annual) risk free rate r, should be equal to the current option price observed at the market. Hence, our moment constraints read as follows:

$$C_0^{K_i} = e^{-rT} \int\limits_{V_T = D + K_i}^{\infty} (V_T - D - K_i) f(V_T) dV_T, \quad i = 1 \dots B, \tag{2}$$

with B denoting the number of observable option prices $C_0^{K_i}$, whereat the current stock price S_0 is included as an option with strike $K_1 = 0$.

One intends to solve the system of equations (2) with respect to the unknown density for given option prices $C_0^{K_i}$ at different strikes K_i. One faces an under-determined estimation problem, as we do not have an infinite set of strikes. Different statistical approaches are possible to determine a unique density $f(V_T)$ out of the infinite many that are compatible with the observed prices (see eg Jackwerth (2004)).

We apply the so called cross-entropy function, originally introduced by Kullback and Leibler (1951), to find a unique density. The cross-entropy-function $CE[f(V_T), f^0(V_T)]$ is defined as

$$CE[f(V_T), f^0(V_T)] = \int\limits_{0}^{\infty} f(V_T) \log \frac{f(V_T)}{f^0(V_T)} dV_T \tag{3}$$

and for a given density $f^0(V_T)$, we find the optimal $f(V_T)$ by minimizing (3) under the moment constraints given by the system of equations (2).

The cross entropy function is based on the concept of entropy which - as shown by Shan-

[§]Whether put or calls are used does not matter as they are deterministically linked by the put-call parity.

non (1948) - can be interpreted as a measure of the average uncertainty in a random variable. Using a weak law of large numbers and Stirling's approximation, the entropy function $H[f(x)] = -\int_0^\infty f(x) \log f(x) dx$ can be directly derived from the multinomial coefficient in which the relative frequencies of the different outcomes are replaced by probabilities (see eg Jaynes (1968)). Each vector of probabilities (assignable to a given domain) entails a certain number of possible outcomes, whereat the degree of uncertainty in a random variable increases with the number of possible outcomes. The maximum entropy distribution for a range of possible outcomes (domain) will hence be that distribution that provides the most uncertainty regarding a future outcome and it is therefore the least informative distribution. On a closed interval this will be the uniform distribution, on a unbounded positive real valued domain (for a given mean) the exponential distribution and on a unbounded real valued interval (given a mean and a variance) the normal distribution.

The related cross entropy function (3), that we minimize in our estimations, can be interpreted as an entropic measure of the discrepancy between the two probability distributions $f(x)$ and $f^0(x)$. In the following the latter can be thought of as a prior distribution. As suggested by Jaynes (1957a), both, the entropy and the cross entropy function, can be used for the estimation of probability distributions when there is only partial information available. Minimizing the cross-entropy function and maximizing the entropy function yield the same optimal solution if the prior distribution $f^0(V_T)$ in (3) is chosen to be of maximal entropy on the defined domain. In both cases the principle of maximum entropy (Jaynes (1957a), Jaynes (1957b)) holds, which states that given the information from the data (the moment constraints) the distribution which best describes our current state of knowledge is the one that maximizes the entropy.[¶] The reason why we use the cross entropy instead of the maximum entropy concept, is that it is a more general framework for which (if available) additional prior information can be used in the estimation.

As shown in the next section, the minimum cross-entropy density will be in the family of

[¶]In our case this means that we minimize the additional assumed knowledge about the future realisation of the stock price beyond what is known from the data (the option prices).

exponential distributions where the number of parameters equals the number of available option prices. This implies that the more data is available the more flexible shapes we can approximate. A great advantage over eg the use of (truncated) mixtures of normals is that we do not have to decide how much flexibility we allow for modelling the density (by choosing the number of used mixtures) as this will be automatically determined by the amount of data available.[‖]

3 Estimation of the RND

3.1 Basic Setup

Using the cross-entropy principle, the option implied RND is obtained by minimizing equation (3) under the moment constraints given by the equations (2).

Applying the Lagrange multiplier technique and taking into account the additivity constraint $\int_0^\infty f(V_T)dV_T = 1$ to ensure that the density integrates to one, the optimization problem of Capuano (2008) reads as:

$$L = \int\limits_{V_T=0}^{\infty} f(V_T)\left[\log\frac{f(V_T)}{f^0(V_T)}\right]dV_T + \lambda_0\left[1 - \int\limits_{V_T=0}^{\infty} f(V_T)dV_T\right] + \sum_{i=1}^{B}\lambda_i\left[C_0^{K_i} - e^{-rT}\int\limits_{V_T=D+K_i}^{\infty}(V_T - D - K_i)f(V_T)dV_T\right] \tag{4}$$

where $f^0(V_T)$ is the distribution of maximum entropy on the defined domain and $\lambda_0, \ldots, \lambda_B$ are the Lagrange multipliers. To obtain the first-order conditions for $f(V_T)$, one needs the derivative of the Lagrange function with respect to the density (see eg, Cover and Thomas (2006)), which then yields:

$$f^*(V_T) = f^0(V_T)\exp\left[\lambda_0 - 1 + \sum_{i=1}^{B}\lambda_i e^{-rT}\mathbf{1}_{V_T>D+K_i}(V_T - D - K_i)\right] \tag{5}$$

‖ In addition, as eg shown in Bahra (1997) and Cooper (1999), already the use of a small number of log-normals leads to quite irregular RND estimates with 'spikes'.

where $f^*(V_T)$ is an infinite-dimensional vector and $\mathbf{1}$ is an indicator function that is one if the condition is true and zero otherwise.

By inserting equation (5) in the additivity constraint, $\exp[\lambda_0 - 1]$ can be expressed as function of the remaining λ_i such that $f^*(V_T)$ can be rewritten as:

$$f^*(V_T) = \frac{1}{\mu(\lambda)} f^0(V_T) \exp\left[\sum_{i=1}^{B} \lambda_i e^{-rT} \mathbf{1}_{V_T > D + K_i}(V_T - D - K_i)\right] \tag{6}$$

with

$$\mu(\lambda) = \exp(1 - \lambda_0) = \exp(-\lambda_0') = \int\limits_{V_T=0}^{\infty} f^0(V_T) \exp\left[\sum_{i=1}^{B} \lambda_i e^{-rT} \mathbf{1}_{V_T > D + K_i}(V_T - D - K_i)\right] dV_T \tag{7}$$

It turns out that for a given value of debt D the optimization for $f(V_T)$ results in the necessity to determine the optimal set of λ_i's in (6). In Capuano (2008) this is achieved by inserting equation (6) in equation (4), deriving the resulting function with respect to the remaining λ_i's and setting the latter to zero. One obtains the following non-linear system of equations consisting of the partial derivatives:

$$\frac{\partial L}{\partial \lambda_i} = e^{-rt} \int\limits_{V_T=0}^{\infty} \mathbf{1}_{V_T > D + K_i}(V_T - D - K_i) f^*(V_T) dV_T - C_0^{K_i} \stackrel{!}{=} 0 \quad i = 1 \dots B \tag{8}$$

The optimal set is calculated in Capuano (2008) by solving (8) with a multivariate Newton-Raphson algorithm (see eg, Zellner and Highfield (1988)), that is by linearising the system with a first order Taylor approximation. Unfortunately, the search for the roots of the system is infeasible in many applications for various reasons. First, the Jacobi matrix resulting from the Taylor approximation has near singularities in large regions of the λ-space which makes the required inversion of the Jacobi matrix numerically impossible in most cases. Matters are further complicated by the fact that the iterative procedure used by the Newton-Raphson algorithm is very vulnerable to inaccuracies in the numerical solution of the integrals involved in equation (8). As a result, the search for the roots is unstable and converges only for a small number of constraints and when the initial values for λ are set near the final solution (see eg, Ormoneit and White (1999), Maasoumi (1993)). To overcome these problems we suggest a robust and computationally efficient algorithm to calculate the optimal set of λ in equation (6) by applying an approach introduced by Alhassid et al. (1978).

3.2 Derivation of the New Objective Function

Alhassid et al. (1978) showed that a function can be defined such that for any trial set of parameters $\lambda_1^T \dots \lambda_B^T$ it provides a theoretical upper bound to the entropy of the maximum entropy density that satisfies the imposed moment conditions. Equivalently, we will derive a lower bound to the cross entropy of the corresponding minimum cross-entropy.
To derive the objective function, we start by denoting every density that satisfies the moment constraints given by equations (2) with $f(V_T)$ and the particular $f(V_T)$ that is of minimum cross entropy with $f^*(V_T)$. Further we define $f^{Tr}(V_T)$ as any (trial) distribution of minimum cross entropy, that is a distribution of form (6) with parameters $\lambda_1^{Tr} \dots \lambda_B^{Tr}$. Subsequently we will show that a strictly convex function W of $\lambda_1^{Tr} \dots \lambda_B^{Tr}$ exists which has a minimum at that set of $\lambda_i^{Tr} = \lambda_i^*$ that satisfies the system of equations (8) and therefore provides us with $f^*(V_T)$.

In order to obtain W we use the non-negativity characteristic (Cover and Thomas (2006), p. 28) of the cross entropy function, that is

$$CE[f(V_T), f^{Tr}(V_T)] = \int\limits_{V_T=0}^{\infty} f(V_T) \log \frac{f(V_T)}{f^{Tr}(V_T)} dV_T \geq 0. \tag{9}$$

Adding and subtracting $\int_{V_T=0}^{\infty} f(V_T) \log f^0(V_T) dV_T$ on the LHS of (9) and rearranging terms yields:

$$CE[f(V_T), f^0(V_T)] = \int\limits_{V_T=0}^{\infty} f(V_T) \log \frac{f(V_T)}{f^0(V_T)} dV_T \geq \int\limits_{V_T=0}^{\infty} f(V_T) \log \frac{f^{Tr}(V_T)}{f^0(V_T)} dV_T \tag{10}$$

with equality if and only if $f(V_T) = f^{Tr}(V_T)$. Next we insert equation (6) for $f^{Tr}(V_T)$ and get for the RHS of (10):

$$\int_{V_T=0}^{\infty} f(V_T) \left[\lambda_0^{Tr'} + \sum_{i=1}^{B} \lambda_i^{Tr} e^{-rT} \mathbf{1}_{V_T > D+K_i} (V_T - D - K_i) \right] dV_T \tag{11}$$

where $\lambda_0^{Tr'} = (\lambda_0^{Tr} - 1)$.

As it holds that $\int_{V_T=0}^{\infty} f(V_T)\lambda_0^{Tr'} dV_T = \lambda_0^{Tr'}$ and $\int_{V_T=0}^{\infty} f(V_T)e^{-rT}\mathbf{1}_{V_T>D+K_i}(V_T-D-K_i)dV_T = C_0^{K_i}$, one finally obtains:

$$CE[f(V_T), f^0(V_T)] \geq CE[f^*(V_T), f^0(V_T)] \geq \lambda_0^{Tr'} + \sum_{i=1}^{B} \lambda_i^{Tr'} C_0^{K_i}, \tag{12}$$

whereat the first inequality holds because $f^*(V_T)$ is just a particular $f(V_T)$ such that the RHS also applies to $CE[f^*(V_T), f^0(V_T)]$. Therefore equation (12) provides a lower bound on the cross-entropy of the distribution of minimum cross entropy, with equality if and only if $f^{Tr}(V_T) = f^*(V_T)$, implying $\lambda_1^{Tr} \ldots \lambda_B^{Tr} = \lambda_1^* \ldots \lambda_B^*$.

Rewriting (12) we get our working function W:

$$W = \left(CE[f^*(V_T), f^0(V_T)] - (\lambda_0^{Tr'} + \sum_{i=1}^{B} \lambda_i^{Tr} C_0^{K_i}) \right) \geq 0 \tag{13}$$

which can be interpreted as a 'goodness of fit' measure of $f^{Tr}(V_T)$ regarding $f^*(V_T)$ and is therefore minimized.

The FOCs for W are given by the conditions:

$$\partial W / \partial \lambda_i^{Tr} \stackrel{!}{=} 0 \quad \text{or} \quad -\partial \lambda_0^{Tr'} / \partial \lambda_i^{Tr} \stackrel{!}{=} C_0^{K_i} \quad i = 1 \ldots B \tag{14}$$

In the Appendix it is shown that W is a strictly convex function for any set of λ_i^{Tr} implying a unique minimum. Consequently one faces a simple minimization problem for a scalar function in B variables which, given that there is a solution**, will yield convergence for arbitrary starting values $\lambda_{i,0}^{Tr}$.

In practice we will minimize the function $F = -\lambda_0^{Tr'} - \sum_{i=1}^{B} \lambda_i C_0^{K_i}$ rather than W as the two functions differ only by the constant $CE[f^*(V_T), f^0(V_T)]$ and hence F is also strictly convex and has a unique minimum. Further, following Agmon, Alhassid, and Levine (1979), we can calculate F in a computationally more efficient way by multiplying equation (5) with $\exp(\sum_{i=1}^{B} \lambda_i C_0^{K_i} - \sum_{i=1}^{B} \lambda_i C_0^{K_i})$, yielding:

$$\exp(-\lambda_0^{Tr''}) = \exp(-\lambda_0^{Tr'} + \sum_{i=1}^{B} \lambda_i^{Tr} C_0^{K_i}) \tag{15}$$

and

$$F = -\lambda_0^{Tr''} = \log \left\{ \int_{V_T=0}^{\infty} f^0(V_T) \exp \left[\sum_{i=1}^{B} \lambda_i^{Tr} (e^{-rT} \mathbf{1}_{V_T > D+K_i} (V_T - D - K_i) - C_0^{K_i}) \right] dV_T \right\} \tag{16}$$

which is the function that we minimize in our applications. Subsequently we show that we can carry out the integration implied by (16) analytically such that no numerical quadrature methods are necessary.

3.3 Analytical Solution of the Integrals

In order to derive an analytical solution for the integration we assume a finite domain for V_T with a generally defined lower bound $V_{min} \in [0; D]$ (until now we assumed: $V_{min} = 0$) and an upper bound V_{max}. Further we define an uniform prior, ie $f^0(V_T) = \frac{1}{V_{max} - V_{min}}$. Then we split up the integral in (16) such that we can rewrite the objective function F without the

**This requires the derivative to change sign from positive to negative as the set λ_i^{Tr} varies from $-\infty$ to $+\infty$ (see Alhassid et al. (1978)).

indicator function:

$$F = \log\left(\frac{1}{V_{max} - V_{min}}\right) + \log\left\{\int_{V_{min}}^{D} \exp\left(-\sum_{i=1}^{B} \lambda_i C_0^{K_i}\right) dV_T\right.$$
$$+ \sum_{i=1}^{B-1} \int_{D+K_i}^{D+K_{i+1}} \exp\left(\sum_{j=1}^{i} \lambda_j (e^{-rT}(V_T - D - K_j) - C_0^{K_j}) - \sum_{k=i+1}^{B} \lambda_k C_0^{K_k}\right) dV_T$$
$$\left. + \int_{D+K_B}^{V_{max}} \exp\left(\sum_{j=1}^{B} \lambda_j (e^{-rT}(V_T - D - K_j) - C_0^{K_j})\right) dV_T \right\} \quad (17)$$

For this form of F the implied integrals can be solved in a straightforward way, leading to:

$$F = \log\left(\frac{1}{V_{max} - V_{min}}\right) + \log\left\{\exp\left(-\sum_{i=1}^{B} \lambda_i C_0^{K_i}\right)(D - V_{min})\right.$$
$$- \sum_{i=1}^{B-1}\left[\frac{\exp\left(\sum_{j=1}^{i} \lambda_j (e^{-rT}(K_i - K_j) - C_0^{K_j}) - \sum_{k=i+1}^{B} \lambda_k C_0^{K_k}\right)}{e^{-rT}(\sum_{j=1}^{i} \lambda_j)}\right.$$
$$\left. - \frac{\exp\left(\sum_{j=1}^{i} \lambda_j (e^{-rT}(K_{i+1} - K_j) - C_0^{K_j}) - \sum_{k=i+1}^{B} \lambda_k C_0^{K_k}\right)}{e^{-rT}(\sum_{j=1}^{i} \lambda_j)}\right]$$
$$\left. - \left[\frac{\exp\left(\sum_{j=1}^{B} \lambda_j (e^{-rT}(K_B - K_j) - C_0^{K_j})\right) - \exp\left(\sum_{j=1}^{B} \lambda_j (e^{-rT}(V_{max} - D - K_j) - C_0^{K_j})\right)}{e^{-rT}(\sum_{j=1}^{B} \lambda_j)}\right]\right\} \quad (18)$$

4 Estimation of the Option iPoD

So far we focused on the estimation of the optimal set of λ where we had to assume that the default barrier D is known. In this section we turn to the estimation of the optimal D and the related determination of the PoD. We first analyse the statistical mechanism which allows to estimate the option iPoD, such that we can show that i) a book value based definition of the RND domain bounds, as in Capuano (2008), is unnecessary and under some circumstances will severely bias the results (section 4.1), and ii) the cross-entropy based D

estimation approach of Capuano (2008) leads to arbitrary results (section 4.2). After describing the drawbacks of the originally suggested approach, we propose in section 4.2 a new algorithm to estimate the PoD which incorporates the evolution of the Lagrange multipliers if the RND is estimated for different D.

To allow for comparisons with the Capuano (2008) framework, we will use the generally defined lower bound $V_{min} \in [0; D]$ for our domain throughout this section. In Capuano (2008) the lower bound V_{min} is defined by a firm's book value of debt (per share).[††] Similarly the upper bound V_{max} is based on the book value of assets.[‡‡]

Further, we introduce the following definitions: We refer to the intervals $[V_{min}, D]$ and $[D, V_{max}]$ as the PoD domain and the pricing domain respectively. The density that we assume to have priced the observable options we denote as the True Pricing Density (TPD) with mass point $f(S_T = 0) = PoD_{TPD}$.

4.1 The Model Parameters

Choice of the Domain Bounds

To obtain an option implied PoD we allow in the RND estimation for a mass point at the value of zero for the future stock price. In order to estimate that mass point we define an uniform prior $f^0(V_T) = \frac{1}{(V_{max}-V_{min})}$ for an interval $[V_{min}, V_{max}]$, with $V_{min} \leq D < V_{max}$. Hence, V_{min} and V_{max} are the domain bounds for our V_T. These bounds influence the estimation results solely through the domain length $(V_{max} - V_{min})$ that they define, but not by their actual values. This is because the future pay-offs $\max(V_T - D - K_i; 0)$ are for $V_T > D$ (and a given K) the same for arbitrary pricing domains $[D, V_{max}]$ with constant length. Equally, the pay-offs for $V_T \leq D$ are the same for any PoD domain $[V_{min}, D]$ with constant length, as all values within this domain imply a future pay-off of zero. It follows that book value

[††]Note that in Capuano (2008) V_{min} is denoted by D_0, the initial guess for the debt value of the firm. We find the term V_{min} more apt as the prior and hence the domain is defined from V_{min} to V_{max}.

[‡‡]More precisely, it is derived from the book value of the assets, the growth rate of the value of assets over the last year and its standard deviation.

based domain bounds, as in Capuano (2008), do not have a real life meaning since any domain with the same length implies identical results. In contrast, a book value based domain length will influence the results in a bad way if the implied domain is too short as we see next.

From a statistical point of view we should choose the domain large enough to have enough possible pay-offs to price the options correctly. The PoD domain must be large enough to approximate the mass point and the pricing domain must be large enough to price the options with the residual probability $1 - PoD$. If the pricing domain is larger than needed (ie the implied pay-offs of the true pricing density (TPD) are a subset of the implied pay-offs of the empirical domain), a density of ('practically') zero will be assigned to the additional pay-offs on the upper part of the domain as the options are optimally priced using solely the TPD domain. As the spare domain implies large pay-offs, assigning density to these values would significantly influence the price of the option and hence 'destroy' the optimal fit. For this reason, the pricing domain can be chosen very large without (significantly) influencing the results. Put another way, there won't be significant damage if we choose the domain (compared to the TPD) too large, but there will be an severe impact on the results if we choose it too small.

Influence of the Distress Barrier D on the PoD Estimation

For given domain bounds we are left with the choice of the optimal D and hence the length of the PoD domain. As will be seen, the decision about D crucially influences the size of our estimated PoD. The estimation procedure uses the moment constraints (2) to modify the prior to the posterior density where for any $V_T \leq D$ all restrictions are zero except of the additivity constraint. The additivity constraint assigns constant density according to $f^*(V_T) = \frac{1}{(V_{max}-V_{min})} \frac{\exp(-\sum_{i=1}^{B} C_0^{K_i} \lambda_i)}{\exp(F)}$ to all V_T in the PoD domain. As the pay-offs are zero for all $V_T \leq D$ the V_T in the PoD domain do not influence the option prices directly. But because the pricing domain density does not integrate up to one if density is assigned to the PoD domain there is an indirect effect of these values on the prices.

Of course, as the density assigned to the spare pay-offs in the upper part of the domain is never exactly zero, an 'unreasonably' large domain will still influence the results to some degree. In practice we set the domain length five times the stock price.

Knowing that the integral over the PoD domain corresponds to our PoD we can define the following PoD function:

$$PoD(\lambda, D) = \frac{1}{(V_{max} - V_{min})} \frac{\exp(-\sum_{i=1}^{B} C_0^{K_i}\lambda_i)}{\exp(F(\lambda, D))}(D - V_{min}), \tag{19}$$

which provides further insights into the mechanism. First, the PoD function shows that the size of the density assigned to the PoD domain is a non-linear function of all RND shape parameters λ_i, $i = 1...B$, which implies that the PoD and the shape of the pricing domain density interact. Hence, for a given D the combination of PoD and RND shape that best fits the prices is identified as optimal. Second, equation (19) reveals that the PoD depends crucially on the length of the PoD domain, ie for a given V_{min} on the choice of the 'optimal' $D = D^*$. If we rewrite the PoD-function as $PoD(\lambda, D) = f^*(V_T = D)(D - V_{min})$, we immediately see that in order to obtain the true PoD (PoD_{TPD}) it must hold: $D^* = V_{min} + \frac{PoD_{TPD}}{f^*(V_T = D^*)}$. This simply means that for D^* the mass point of the TPD must be perfectly approximable as the integral over the density in the PoD domain. As a consequence also the TPD for $S_T > 0$ is only perfectly approximable by $f^*(V_T > D)$ if $D = D^*$.

If $D \neq D^*$ then the TPD will not be estimateable and instead the TPD will be 'modified' such that the prices can be met given the 'wrong' PoD domain. The more strikes we have, the more flexible shapes are possible (as one has more λ_i). The parameter λ_1 belonging to the moment constraint given by S_0 will provide flexibility at $V_T \approx D$ which is necessary to roughly approximate PoD_{TPD} also for 'wrong' PoD domains. If $D < D^*$, then a good fit is achieved if, compared to the TPD, $f^*(V_T = D)$ is larger than $f^*(V_T = D^*)$ and/or higher density is assigned to very small pay-offs (and consequently less density to high pay-offs). For $D > D^*$, $f^*(V_T = D)$ will be be smaller than $f^*(V_T = D^*)$ and/or less density is assigned to very small pay-offs. For too large Ds the PoD will in general be larger than PoD_{TPD} due to the fact, as will be seen in the numerical evaluations, that the estimation framework will

Note that for our method it is not required to have information from out of the money options to provide PoD estimates. Any data set of options might be priced more consistently if the RND exhibits a PoD. This flexibility is provided because any shape parameter λ_i influences the PoD function.

assign as much density as possible to the PoD domain given that the prices are still met. The reason is that by assigning density to the PoD domain the density as a whole gets more entropic, ie more uniform.

The Figures 1.(a)-1.(d) as well as the Tables 1 and 2 illustrate the described effects using numerical evaluation examples for our estimation procedure.

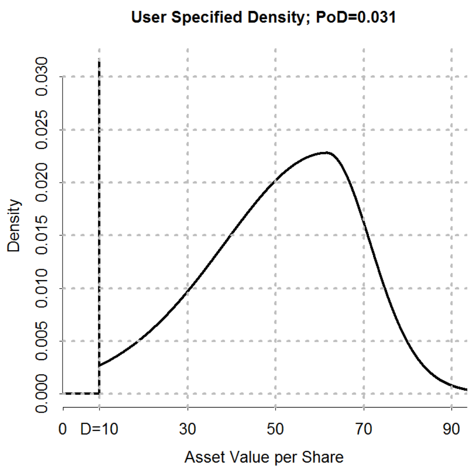

Figure 1.(a): Specified TPD (PoD= 0.032).

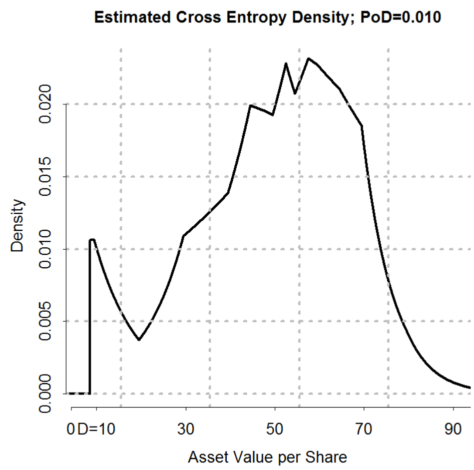

Figure 1.(b): Cross-Entropy density ($V_{min} = 9$, $D = 10$, Est. PoD=0.010).

Assigning density to the pricing domain always requires some 'structure' as this density influences the prices directly. Hence, assignment to the PoD domain is always 'the more entropic' alternative.

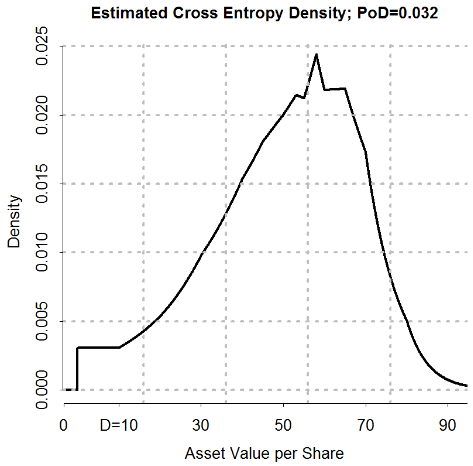

Figure 1.(c): Cross-Entropy density ($V_{min} = 4$, $D = 10$, Est. PoD=0.32).

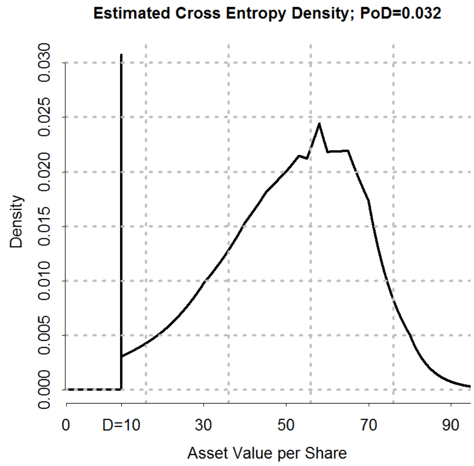

Figure 1.(d): Cross-Entropy density ($V_{min} = 4$, $D = 10$). PoD assigned to $V_T = D$.

For the numerical evaluation, artificial TPDs were created (see Figure 1.(a)) and the theoretically implied option prices for different strike prices were calculated from these densities. The theoretical option prices were then used as input data in the cross-entropy estimations for different lengths of the PoD domain (see Figures 1.(b)-1.(d)). To obtain a realistic impression of the reliability of the estimates in practice, we used about ten equally spaced different strike prices within a range of $[0.7 \times S_0, 1.3 \times S_0]$ in order to estimate the respective densities. Such a set of strikes corresponds roughly to the available strikes of 3 month option contracts for major financial institutions, which, as stressed in the introduction, are the type of firms the iPoD approach seems most appropriate for.

The Figures 1.(b) and 1.(c) illustrate the effect of differing lengths of the PoD domain on the estimated densities. The option iPoD procedure provides correct PoDs and smooth density estimates if the length of the PoD domain allows to approximate the mass point of the TPD. The smoothness is provided as the cross-entropy function acts as a type of smoothness criterion that 'dislikes' deviations from the uniform prior distribution. As the PoD domain in Figure1.(b) is too short more spiky densities are required to fit the option prices. The fit is provided by assigning large density to small pay-offs and removing density for larger pay-offs.

Note, if we would have used strike prices closer to zero in the estimation then the sharp increase in the

The Tables 1 and 2 show the PoD estimates for differing lengths of the PoD domain ($D - V_{min}$) for two different TPDs. As expected the estimated PoDs increase with the length of the intervals, but notably the procedure approximates the magnitude of the PoDs even for too long intervals quite well. A further notable fact is, as will be illustrated in section 4.2, that the approach clearly discriminates between TPDs exhibiting PoDs and TPDs that do not.

$D - V_{min}$	1	2	3	4	5	**6**	10	15	20
PoD	0.010	0.017	0.022	0.027	0.027	0.032	0.037	0.042	0.045

Table 1: Estimated PoDs for different lengths of the PoD domain (Specified PoD: 0.031).

$D - V_{min}$	1	2	3	4	5	**7**	10	15	20
PoD	0.0004	0.0009	0.0012	0.0016	0.0020	0.00279	0.0037	0.0055	0.0065

Table 2: Estimated PoDs for different lengths of the PoD domain (Specified PoD: 0.0028).

4.2 Approximation of the Correct PoD

A closer look at the results in Table 1 and 2 reveals that we should aim for more than just rough estimates for the PoD. The framework provides very accurate estimates for any PoD level and RND form if we are able to identify the correct interval length.

Analysis of the Approach Suggested in Capuano (2008)

The approach suggested in Capuano (2008) evaluates the objective function (18) for different D. We can derive a formula for the optimal D for a given set of λ from the objective function in (18). Solving $\frac{\partial F}{\partial D} = 0$ for D yields

$$D^* = V_{max} - \frac{\sum_{j=1}^{B} \lambda_j K_j}{\sum_{j=1}^{B} \lambda_j}. \tag{20}$$

density for small pay-offs would be closer to D. The increase would be larger and the PoD (the resulting $f^*(V_T = D)$) would be higher.

Inserting this formula for D in the objective function and optimizing will give us the optimal interval length.

However, the flexibility of our approach is a curse here. To show this, a loss function is defined which measures the quadratic distance between the observed prices and the prices implied by the respective RND estimates for differing lengths of the PoD domain (keeping V_{min} fixed). In Figure 2.(a) the loss function for the numerical example of Table 1 is displayed. One finds that for arbitrary interval lengths a good fit to the data can be obtained. Hence, there are no moment restrictions to identify the optimal D as many D provide a good fit, and only the cross-entropy is minimized in the optimization of (18). Consequently the density that is closest to the uniform distribution will be identified as optimal.

As discussed before, density is uniformly assigned to the PoD domain which implies that the 'uniform nature' of the posterior will increase with increasing D. Hence, the D closest to V_{max} that still provides a good fit to the data will be identified as optimal. The fit decreases if the pricing domain becomes too short to cover the pay-offs of the TPD. Hence, given enough pay-offs in the pricing domain, the length of the PoD domain will be maximized.

Eventually, the estimated PoD will depend on the length of the RND domain. Since we can not infer the domain length of the TPD from any theoretical considerations, the suggested PoD mechanism will produce arbitrary results for the PoD. This is also holds for the book value based definition of the domain bounds in Capuano (2008). Figure 2.(b) illustrates the described problem of the Capuano (2008) approach by showing the value of the cross entropy function for estimated densities with different D (and keeping V_{min} fixed) using the same numerical example as in Table 1.

We would have to know how long the true PoD domain is. But the PoD domain is a purely statistical concept solely used to be able to approximate the mass point $f(S_T = 0)$.

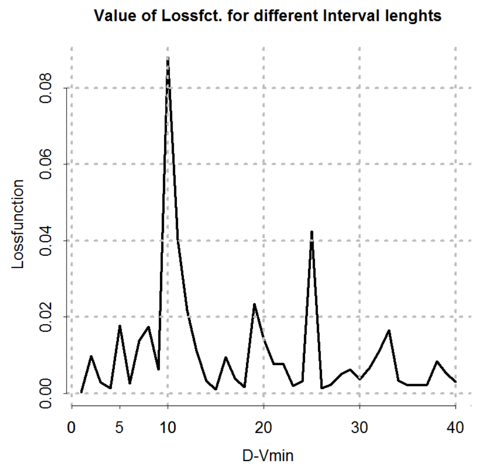

Figure 2.(a): Quadratic loss function for estimates with different lengths of the PoD domain (keeping V_{min} fixed).

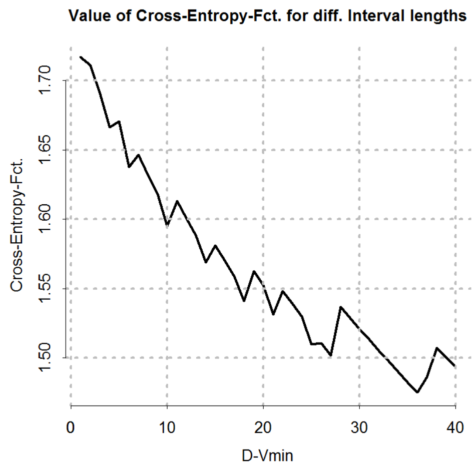

Figure 2.(b): Value of cross-entropy function for different lengths of the PoD domain (keeping V_{min} fixed).

The 'Averaging Approach' to Estimate the PoD

Noting the arbitrariness of the PoD estimates in the Capuano (2008) framework, we suggest an alternative procedure. It is important to note that the D estimation must not depend on the domain length $[V_{min}, V_{max}]$, since we can not infer the TPD domain by any means. Our approach is still quite ad hoc and requires further research but, as will be seen in section 5, provides very promising results. It is based on the evolution of the PoD function (19) and of the Lagrange multipliers when estimating the optimal density for different D (and setting $V_{min} = 0$). We start by looking at a numerical example in which we specify a TPD with no PoD and estimate it for different D. Figure 3.(a) shows the evolution of the PoD and Figure 3.(b) the 'aggregated' evolution of the estimated Lagrange multipliers.

It can be seen that the estimated Lagrange multipliers stay the same for all chosen D and consequently the PoD function increases linearly with growing interval length (which is implied by the PoD function (19) for constant λ_i). The reason is that there is no PoD to assign to the PoD domain and hence no shape modifications have to be carried out in the pricing domain (would be displayed in changing λs) for changing D.

In contrast, we now look at the evolution of the PoD and the Lagrange multipliers if we define the same RND as for Table 1 (see Figure 4.(a) and 4.(b)). The evolution of the Lagrange multipliers illustrates the shape adjustments that are necessary for increasing D in order to get a good fit to the data in each case. The PoD function displays roughly a concave form and hence the slope of the function decreases with growing D. Empirically we found the PoD to be more concave if the PoD is high. Looking at the evolution of the λs one detects strong fluctuations. These fluctuations will be the stronger the higher the PoD is. But equal to the PoD function (which is governed by the λ) the λ-function is flattening with increasing D. This characteristic would clearly be more striking if the function would be smoothed.

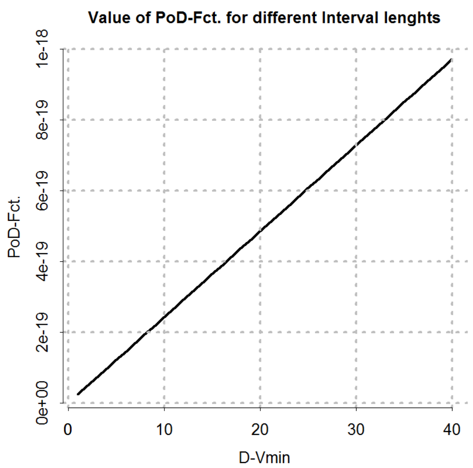

Figure 3.(a): Evolution of the PoD for different D ($V_{min} = 0$); Specified PoD=0.

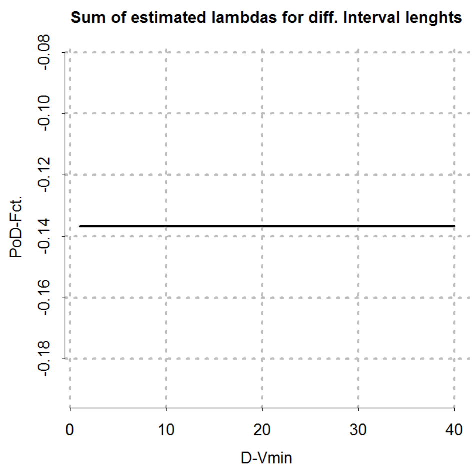

Figure 3.(b): Sum of estimated Lagrange multipliers ($\sum_{i=1}^{B} \lambda_i$); Specified PoD=0.

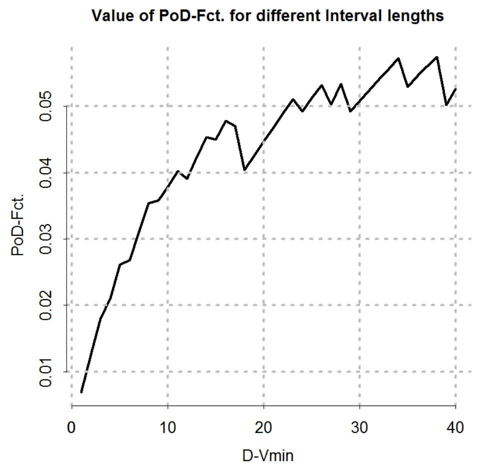

Figure 4.(a): Evolution of the PoD for different D ($V_{min} = 0$); Specified PoD=0.031

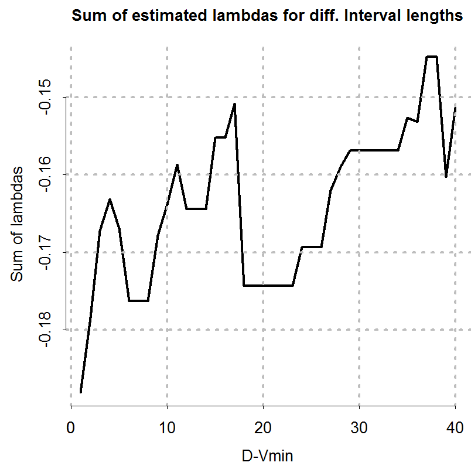

Figure 4.(b): Sum of estimated Lagrange multipliers ($\sum_{i=1}^{B} \lambda_i$); Specified PoD=0.031

The exact nature of the evolution of the PoD and the λ-Function are matter of current research. It seems that the exact determination of the optimal interval length has to be based on the second derivative of the functions as the slope of the functions is the steepest before reaching the optimal interval length of six (see Table 1) and is flattening afterwards. This fast convergence to the true PoD was found in all of our empirical evaluations. So far though, it is not clear what the exact decision rule should be as the degree of flattening depends on the level of the PoD.

Until now we suggest the following ad hoc procedure which led to convincing results in our numerical experiments (see section 5). Since we do not know the exact decision rule, we decide to average estimated PoDs over several lengths of the PoD domain ((setting $V_{min} = 0$)) after having defined an upper bound to the length, ie we average over the PoD estimates for $D = 1 \ldots D_{max}$. Empirically we found that choosing 20 as D_{max} provides accurate results for arbitrary PoD levels and RND forms. The decision for $D_{max} = 20$ is backed by the finding that the PoD function is quite flat for this value of D for any PoD level that we specified in our numerical experiments. This strongly indicates that in practical applications the 'true PoD' can be very well approximated with an maximal interval length of 20, and averaging over estimates for lengths close to the optimal one will provide good results. To identify the optimal D and RND one chooses that length for the PoD domain that provides the PoD estimate closest to the 'average PoD'. Finally, it is important to note, that the reason for the mechanism to work is the flexibility of the entropy based estimation approach which guarantees that the true PoD can be roughly approximated for any length of the PoD domain.

5 Evaluation

As addressed in the previous section, we evaluate the suggested estimation procedure by defining different densities (TPDs) from which we generate our option data.

Figure 5.(a) and 6.(a) show two TPDs, Table 3 their respective statistical characteristics. The density in 5.(a) exhibits the typical shape of RNDs often found in empirical studies (negative skewness and positive excess kurtosis ('fat tails')) except that the entire density below the default barrier is assigned to D leading in this case to a negative (excess) kurtosis. In contrast, in Figure 6.(a) a density of quite unusual form is specified such that the great flexibility of the estimation procedure can be demonstrated. In Figures 5.(b), 6.(b) and Table 3 the respective results of the estimation are shown.

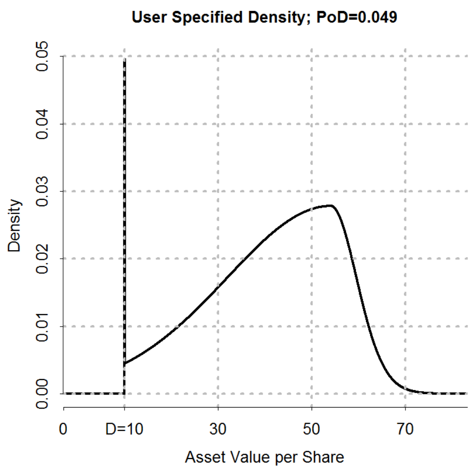

Figure 5.(a): TPD with PoD= 0.049.

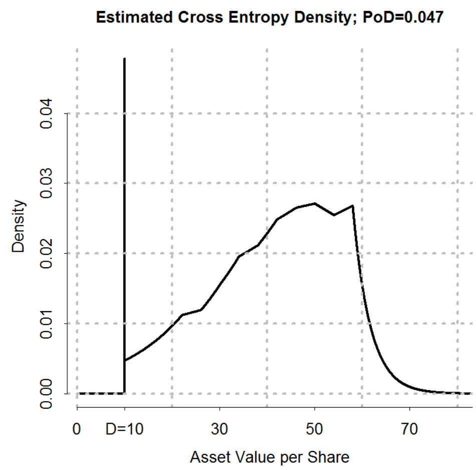

Figure 5.(b): Cross-Entropy density for the density in Figure 5.(a), with vector of strikes $K = (0, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39)$.

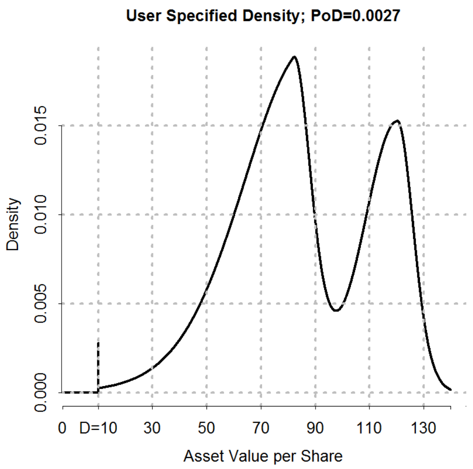

Figure 6.(a): TPD with PoD= 0.0027.

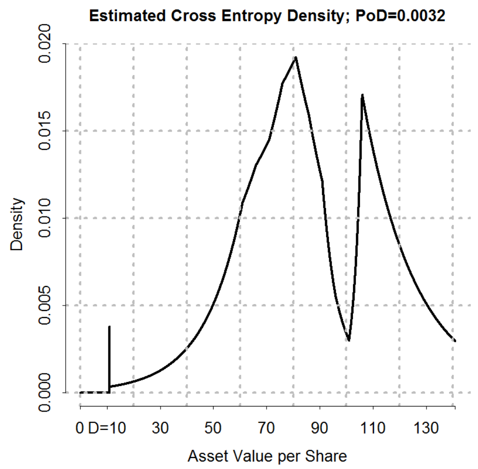

Figure 6.(b): Cross-Entropy density for the density in Figure 6.(a), with vector of strikes $K = (0, 55, 60, 65, 70, 75, 80, 85, 90, 95)$.

One can easily verify the accuracy of the estimates regarding shape and PoD of the RNDs. In each case the optimal PoD was determined as the mean of the PoDs estimated for differing interval lengths ($D_{max} = 20$). The optimal density was then identified as the density that exhibits a PoD that is the closest to the average PoD.

	PoD	Mean	Variance	Skewness	Kurtosis
Figure 5.(a)	0.0496	40.6024	213.3029	-0.4614	-0.6215
Figure 5.(b)	0.0478	40.6205	214.1831	-0.4392	-0.6058
Figure 6.(a)	0.0027	85.8212	680.5172	-0.1146	-0.6574
Figure 6.(b)	0.0032	86.2269	676.8333	-0.0584	-0.4193

Table 3: PoDs and moments of the TPDs and their corresponding cross-entropy density estimates.

	PoD	Mean	Variance	Skewness	Kurtosis
Specified	0.1977	45.8692	179.3605	0.5490	-0.0811
Estimated	0.2044	45.0393	185.9304	0.7241	0.5073
Specified	0.0838	70.6151	9.5013	0.6569	4.5840
Estimated	0.0873	70.6192	8.8162	0.3001	1.4496
Specified	0.0159	23.0837	23.1187	-0.5768	-0.0459
Estimated	0.0188	23.1054	23.5051	-0.5481	0.0833
Specified	0.0010	70.0067	8.8962	-0.6949	0.4190
Estimated	0.0040	70.0634	8.9593	-0.7501	0.8354
Specified	0.000078	90.6685	51.5272	-3.6873	24.6349
Estimated	0.000121	90.6784	51.3571	-3.6397	19.8385
Specified	–	155.1997	24.9505	-1.0545	2.4086
Estimated	10^{-23}	155.0426	25.4788	-0.9470	1.9518

Table 4: PoDs and Moments of specified TPDs and their corresponding (optimal) Cross-Entropy density estimates.

In Table 4 one can see the results of density estimates for further TPD specifications with PoDs ranging from very high ($\approx 20\%$) to very low (0.0078%) and differing statistical moments. Also, in these numerical experiments the framework provided very reliable results. Especially remarkable is that our simple ad hoc procedure regarding the determination of the PoD is able to obtain accurate estimates for any levels of the true PoD. A clear feature of the framework seems to be that lower PoDs can be estimated more accurately than high PoDs. This is due to the strong shape modifications that have to be carried out when estimating RNDs with high PoDs for different interval lengths. Consequently the PoD estimates for different D vary more for high PoDs than for low PoDs and the averaging approach is less accurate.

Finally, we want to provide a short empirical illustration of our framework with real option prices of banks. It shows that the proposed framework also produces plausible and promising results if applied to real world data. To evaluate the results of the approach we contrast the estimation results with real events, since the true RND shapes and PoDs of the banks are of course not known.

On 9/21/2011 the Bank of America (BoA) was downgraded two levels by Moody's. This implies that BoA should clearly exhibit a higher level of PoD than eg, JP Morgan Chase & Co. (JPM), whose top rating stayed unchanged. We now want to test if our framework is able to identify this elevated risk for BoA relative to JPM months before the actual downgrade did happen. We look at two data sets, each available at Yahoo!Finance (http://finance.yahoo.com). One for 4/25/2011, and one for 8/30/2011. The used options are 3 month contracts with maturity at 07/15/2011 and 11/19/2011 respectively.

We clearly see by comparing the results in the Tables 5 and 6 that the framework indicates an elevated PoD for BoA relative to JPM for both dates. In addition, the PoD for BoA increases sharply getting closer to the date of the downgrading, which seems highly plausible. Finally, there is an increasing variance for both banks, JPM and BoA, which might indicate a general increase in risk perception in the market regarding financial institutions.

4/25/2011	PoD	Mean	Variance	Skewness	Kurtosis
JP Morgan Chase & Co. (JPM)	5.8×10^{-9}	54.8529	19.9191	0.0382	0.6768
Bank of America (BoA)	0.00152	22.4115	3.5336	-0.3701	6.6342

Table 5: PoD and moments of RNDs for JP Morgan Chase & Co. and Bank of America based on Options from 4/25/2011.

8/30/2011	PoD	Mean	Variance	Skewness	Kurtosis
JP Morgan Chase & Co. (JPM)	2.4×10^{-14}	47.6295	42.8599	1.3031	0.9717
Bank of America (BoA)	0.04568	17.4347	8.2326	-0.2795	2.6762

Table 6: PoD and moments of RNDs for JP Morgan Chase & Co. and Bank of America based on Options from 8/30/2011.

We want to stress that this is not meant to be an comprehensive empirical evaluation of the methodology, for which we refer to Matros and Vilsmeier (2012). There, the proposed method is used to estimate time series of PoDs for a sample of major US financial institutions. It is found that the informational content in the time series of option iPoDs is superior to the one implied by time series of CDS spreads during the 2007/2008 US sub-prime crisis.

Note that both data sets have the same time to maturity such that the densities for both dates can be compared, as the maturity dependence is equal in both cases.

For the applications time series of PoDs were estimated. This can be achieved by using alternately call

6 Conclusion

In this paper we presented some technical modifications to the framework proposed in Capuano (2008) to derive an option implied probability of default (option iPoD). The first modification concerns the optimization algorithm to calculate the cross-entropy density associated with the option prices observed at the market. We derived an objective function whose minimization is stable and yields unique solutions for the Lagrange parameters which determine the optimal density. Further, we show how the integrals of the objective function can be solved analytically. Another modification was proposed regarding the determination of the optimal PoD. After reasoning that a pure entropic approach to determine the optimal default barrier leads to arbitrary results, we suggested an easy to implement algorithm for the calculation of the optimal PoD based on the characteristics of the PoD function. Both modifications to the framework increase the general practical applicability of the option iPoD framework.

In section 5 we comprehensively tested our approach by applying it to option prices generated from user-specified RNDs. The results are very convincing as the estimation procedure was shown to be highly accurate regarding the estimation of the moments and the PoD of the true density. Especially remarkable is the ability of the framework to estimate densities with very low probability of defaults. This is essential for practical applications where one will mostly deal with low-PoD densities. In an illustrative application to real option data the framework is able to anticipate the downgrading decision taken by Moody's regarding Bank of America.

Further, throughout the paper we showed the purely statistical nature of the option iPoD framework which implies that no theoretical default model or exogenous book values are needed to derive the PoDs. This contrasts the interpretation of the methodology as a 'structural approach' based model with endogenous distress barrier provided in Capuano (2008).

---

options with a fixed maturity cycle as eg 5-,6-, and 7-months. The inherent maturity dependence can be removed by estimating (non-linearly) the maturity effect for the respective time to maturities on basis of pooled PoDs over a large period of time and several banks/firms. By doing so in Matros and Vilsmeier (2012), iPoDs exhibiting a theoretical evaluation horizon of 6-month were derived.

We conclude that the suggested framework provides a promising tool for the estimation of (risk neutral) PoDs for firms with exchange traded options. Compared to competing approaches, as structural approach based models or the derivation of PoDs from debt based market instruments, the methodology has the great advantage that neither historical balance sheet data nor recovery rate assumptions are required. Particularly attractive seem applications to financial institutions, which due to volatile assets and leverage as well as their complex capital structures are highly sensible to the assumptions in the competing approaches. In addition, the fact, that we derive the PoDs jointly with the corresponding risk neutral densities, provides an optimal basis for multivariate - copula based - density modelling. Hence, the estimation of an option implied multivariate asset distribution for a proxy portfolio of the financial system, based on time series of option iPoDs/RNDs, would be an obvious empirical application of the framework.

Further research is required regarding the exact nature of the PoD function and the evolution of the Lagrange multipliers when the length of the PoD domain is changed. The objective should be to derive an exact decision rule for the determination of the optimal option iPoD.

References

Agmon, N., Y. Alhassid, and R. Levine (1979). An algorithm for finding the distribution of maximal entropy. *Journal of Computational Physics 30*, 250–258.

Alhassid, Y., N. Agmon, and R. D. Levine (1978). An upper bound for the entropy and its applications to the maximal entropy problem. *Chemical Physics Letters 53*, 22–26.

Avellaneda, M. (1998). Minimum entropy calibration of asset pricing models. *International Journal of Theoretical and Applied Finance*.

Bahra, B. (1997). Probability distributions of future asset prices implied by option prices. *Bank of England Working Paper*.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods* (2nd ed.). Springer.

Buchen, P. W. and M. Kelly (1996). The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis 1*, 143–159.

Capuano, C. (2008). The option-ipod. the probability of default implied by option prices based on entropy. *IMF Working Paper 08*(194).

Chan-Lau, J. A. (2006). Market-based estimation of default probabilities and its application to financial market surveillance. *IMF Working Paper* (104).

Choe, G. H. and M. G. Jeong (2008). Estimation of the asset price distribution using the maximum entropy principle. *Working Paper*.

Cooper, N. (1999). Testing techniques for estimating implied rnds from the prices of european-style options. In *Proceedings of the BIS workshop on implied PDFs.*

Cover, T. M. and J. A. Thomas (2006). *Elements of Information Theory* (2nd ed.). John Wiley & Sons.

Cox, J. and S. Ross (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics 3*, 145–166.

Crosbie, P. and A. Kocagil (2003). Modeling default risk. *Moody's KMV Technical Paper*.

Guo, W. (2001). Maximum entropy in option pricing: A convex-spline smoothing method. *The Journal of Future Markets 21*(9), 819–832.

Jackwerth, J.-C. (2004). *Oprion implied Risk Neutral Distributions and Risk Aversion*. Research Foundation of AIMR (CFA Institute).

Jaynes, E. T. (1957a). Information theory and statistical mechanics. *Physical Review 106*, 620–630.

Jaynes, E. T. (1957b). Information theory and statistical mechanics. ii. *The Physical Review 108*(2), 171–190.

Jaynes, E. T. (1968). Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics 4*(3), 227–241.

J.P.Morgan (1997). Creditmetrics. *Technical Document*.

Kullback, S. and R. Leibler (1951). On information and sufficiency. *Annals of Mathematical Statistics 22*, 79–86.

Maasoumi, E. (1993). A compendium to information theory in economics and econometrics. *Econometric Reviews 12*, 137–182.

Matros, P. and J. Vilsmeier (2012). Measuring option implied degree of distress in the us financial sector using the entropy principle. *Deutsche Bundesbank Discussion Paper 30.*

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance 29*, 449–470.

Neri, C. and L. Schneider (2012). Maximum entropy distributions inferred from option portfolios on an asset. *Finance and Stochastics 16*, 293–318.

Ormoneit, D. and H. White (1999). An efficient algorithm to compute maximum entropy densities. *Econometric Reviews 18*(2), 127–140.

Rompolis, L. S. (2010). A new method of employing the principle of maximum entropy to retrieve the risk neutral density. *Journal of Empirical Finance 17*, 918–937.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal 27*, 379–423.

Stutzer, M. J. (1996). A simple nonparametric approach to derivative security valuation. *Journal of Finance 51*, 1633–1652.

Stutzer, M. J. (2000). Simple entropic derivation of a generalized black-scholes option pricing model. *Entropy 2*, 70–77.

Zellner, A. and R. A. Highfield (1988). Calculation of maximum entropy distributions and approximation of marginal posterior distributions. *Journal of Econometrics 37*, 195–209.

Appendix

Proof A: The Strict Convexity of W

The proof follows strongly Alhassid et al. (1978) and it is shown that the Hessian matrix for the function W is positive definite. We start by deriving the negative definiteness and hence strict concavity for the function $F'' = -F = \lambda_0^{Tr'} + \sum_{i=1}^{B} \lambda_i^{Tr} C_0^{K_i}$.

From our definition for W in equation (13) it follows that F'' is smaller than $CE[f^*(V_T), f^0(V_T)]$ for arbitrary $\lambda_1^{Tr} \dots \lambda_B^{Tr}$, except if $\lambda_1^{Tr} \dots \lambda_B^{Tr} = \lambda_1^* \dots \lambda_B^*$ for which both terms are equal. Therefore, F'' is a concave function with a unique maximum at $F'' = CE[f^*(V_T), f^0(V_T)]$ if its Hessian matrix is negative definite for arbitrary $\lambda_1^{Tr} \dots \lambda_B^{Tr}$.

We define the following shorthand notations:

$$\phi_i(V_T) = e^{-rT} \mathbf{1}_{V_T > D + K_i} (V_T - D - K_i) \tag{21}$$

and

$$
\begin{aligned}
C_0^{K_i,Tr} &= \int\limits_0^{V_{max}} \phi_i(V_T) f^{Tr}(V_T) dV_T \\
&= \int\limits_0^{V_{max}} \phi_i(V_T) \frac{1}{\mu(\lambda)} f^0(V_T) \exp\left[\sum_{i=1}^{B} \lambda_i^{Tr} \phi_i(V_T)\right] dV_T = -\partial \lambda_0^{Tr'} / \partial \lambda_i^{Tr} \qquad (22)
\end{aligned}
$$

Hence, we get for the first derivative of F'' with respect to λ_i^{Tr}:

$$
\partial F''/\partial \lambda_i^{Tr} = \partial \lambda_0^{Tr'}/\partial \lambda_i^{Tr} + C_0^{K_i} = C_0^{K_i} - C_0^{K_i,Tr} \qquad (23)
$$

and the Hessian of F'' is given by:

$$
\begin{aligned}
\partial F''^2/\partial \lambda_i^{Tr} \partial \lambda_j^{Tr} &= -\partial C_0^{K_i,Tr}/\partial \lambda_j^{Tr} = -\partial C_0^{K_j,Tr}/\partial \lambda_i^{Tr} \\
&= (C_0^{K_i,Tr} C_0^{K_j,Tr}) - \int\limits_0^{V_{max}} \phi_i(V_T)\phi_j(V_T) f^{Tr}(V_T) dV_T \\
&= -\int\limits_0^{V_{max}} f^{Tr}(V_T) \left[\phi_i(V_T) - C_0^{K_i,Tr}\right] \left[\phi_j(V_T) - C_0^{K_j,Tr}\right] dV_T \qquad (24)
\end{aligned}
$$

where we use the expansion $\pm C_0^{K_i,Tr} C_0^{K_j,Tr}$ and the fact that $C_0^{K_i,Tr} C_0^{K_j,Tr}$ can be rewritten as $C_0^{K_i,Tr} \int_0^{Vmax} f^{Tr}(V_T)\phi_j(V_T) dV_T$ to get from the second to the third line.

The Hessian can be interpreted as a covariance matrix of $\phi_i(V_T)$ and $\phi_j(V_T)$ where $C_0^{K_i,Tr}$ and $C_0^{K_j,Tr}$ are the respective expected values. To formally show that the Hessian is positive definite we have to show that for linearly independent constraints the matrix M with

$$
M_{i,j} = \partial F''^2/\partial \lambda_i^{Tr} \partial \lambda_j^{Tr} \qquad (25)
$$

satisfies for any column vector x ($x \neq 0$), $x^t M x > 0$. To do so we note that we can write $M_{i,j}$ as a (weighted) scalar-product of the two constraints $B_i(V_T) = \phi_i(V_T) - C_0^{K_i,Tr}$ and $B_j(V_T) = \phi_j(V_T) - C_0^{K_j,Tr}$ (see eg, Brockwell and Davis (1991) for the axioms that define a scalar product), ie,

$$M_{i,j} = -\int_{0}^{V_{max}} f^{Tr}(V_T)B_i(V_T)B_j(V_T)dV_T = - < B_i, B_j >^{f^{Tr}} \tag{26}$$

where $f^{Tr}(V_T)$ is a strictly positive weighting function. Using the properties of a scalar product we can further write $x^t Mx$ as:

$$x^t Mx = -\sum_{i,j=1}^{B} x_i < B_i, B_j >^{f^{Tr}} x_j = - < \sum_{i=1}^{B} x_i B_i, \sum_{j=1}^{B} x_j B_j >^{f^{Tr}} = - < C, C >^{f^{Tr}} \tag{27}$$

An alternative way to write $< C, C >$ is:

$$< C, C >= (x_1 B_1(V_1) + \ldots + x_B B_B(V_1))^2 + \ldots + (x_1 B_1(V_{V_{max}}) + \ldots + x_B B_B(V_{V_{max}}))^2 \tag{28}$$

so that that it holds that $x^t Mx < 0$ and hence M is negative definite for arbitrary sets of λ_i^{Tr} if $C \neq 0$. One obtains $C = 0$ if and only if for every V_T holds:

$$\sum_{i=1}^{B} x_i B_i = \sum_{i=1}^{B} x_i \phi_i(V_T) - \sum_{i=1}^{B} x_i C_0^{K_i,Tr} = 0 \tag{29}$$

or

$$\sum_{i=0}^{B} x_i \phi_i(V_T) = 0 \tag{30}$$

with $\phi_0 = 1$ and x_0 equal to the second term in (29). Equation (30) will only be satisfied if one or more constraints are linearly dependent. If this is the case the cross entropy distribution will be still unique but the magnitude of the λ_i^{Tr} are not identified uniquely. In practice one can always eliminate one or more constraints in order to obtain a linearly independent set of moment conditions (constraints).

The derivation of the positive definiteness for our working function W is straightforward as we defined $F'' = -F$ and $W = CE[f^*(V_T), f^0(V_T)] + F$ where $CE[f^*(V_T), f^0(V_T)]$ is just a constant.