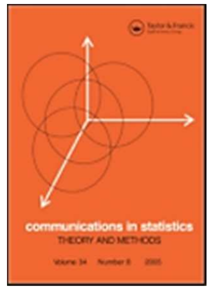

Parameter estimation in credit models under incomplete information

Journal:	*Communications in Statistics – Theory and Methods*
Manuscript ID:	LSTA-2012-0909
Manuscript Type:	Special Issue - MSMPRF2011
Date Submitted by the Author:	25-Oct-2012
Complete List of Authors:	HERBERTSSON, ALEXANDER; University of Gothenburg, Department of economics/Centre for Finance Frey, Rüdiger; Institute for Statistics and Mathematics, Vienna University of Economics and Business
Keywords:	Credit risk, incomplete information, CDS index spread
Abstract:	We consider the filtering model of Frey & Schmidt (2012) stated under the real probability measure and develop a method for estimating the parameters in this framework by using time-series data of CDS index spreads and classical maximum-likelihood algorithms. The estimation-approach incorporates the Kushner-Stratonovich SDE for the dynamics of the filtering probabilities. The convenient formula for the survival probability is a prerequisite for our estimation algorithm. We apply the developed maximum-likelihood algorithms on market data for historical CDS index spreads (iTraxx Europe Main Series) in order to estimate the parameters in the nonlinear filtering model for an exchangeable credit portfolio.

PARAMETER ESTIMATION IN CREDIT MODELS UNDER INCOMPLETE INFORMATION

ALEXANDER HERBERTSSON AND RÜDIGER FREY

Abstract. We consider the filtering model of Frey & Schmidt (2012) stated under the real probability measure and develop a method for estimating the parameters in this framework by using time-series data of CDS index spreads and classical maximum-likelihood algorithms. The estimation-approach incorporates the Kushner-Stratonovich SDE for the dynamics of the filtering probabilities. The convenient formula for the survival probability is a prerequisite for our estimation algorithm. We apply the developed maximum-likelihood algorithms on market data for historical CDS index spreads (iTraxx Europe Main Series) in order to estimate the parameters in the nonlinear filtering model for an exchangeable credit portfolio. Several such estimations are performed as well as accompanying statistical and numerical computations.

1. Introduction

Dynamic modelling of portfolio credit risk is a big and important area in applied mathematical finance. Developing credit portfolio models that conveniently include stochastic dynamics as well as default dependence is challenging. In Frey & Schmidt (2012) the authors developed a full dynamic information-based approach to credit risk modelling where the prices of traded credit derivatives are given by the solution of a nonlinear filtering problem. Frey & Schmidt (2012) solve this problem using the innovations approach to nonlinear filtering and derive in particular the Kushner-Stratonovich SDE describing the dynamics of the filtering probabilities. Moreover, they give interesting theoretical results on the dynamics of the credit spreads and on risk minimizing hedging strategies.

In this paper consider the filtering model of Frey & Schmidt (2012) stated under the real probability measure and outline an algorithm for estimating the parameters in this framework by using time-series data of index CDS spreads and classical maximum-likelihood algorithms. In particular, the estimation approach incorporates the Kushner-Stratonovich SDE for the dynamics of the filtering probabilities. The practical expression for the survival probability is a prerequisite for our estimation algorithm since this formula will be used to back out the filtering probabilities under the market filtration (i.e. the noisy information) specified in the framework of Frey & Schmidt (2012). These filtering probabilities satisfy the Kushner-Stratonovich SDE at each time point, which in turn makes it possible to state

Date: October 26, 2012.

The research of Alexander Herbertsson was supported by AMAMEF travel grants 2206 and 2678, by the Jan Wallanders and Tom Hedelius Foundation and by Vinnova.

The research of Rüdiger Frey was supported by Deutsche Forschungsgemeinschaft.

a likelhood function. We apply our developed maximum-likelihood algorithms on market data for historical CDS index spreads in order to estimate the parameters in the nonlinear filtering model for an exchangeable (i.e. homogeneous) credit portfolio.

Our paper makes several references to Herbertsson & Frey (2012) in which the authors derive tractable formulas for the conditional survival distribution and the conditional number of defaults under the market filtration (i.e. the noisy information) in terms of the filtering probabilities specified in the framework of Frey & Schmidt (2012).

The idea of using maximum likelihood (MLE) techniques in credit risk modelling is not new. For example, Hurd & Zhou (2011) develops a structural credit risk model with noisy information. In order to estimate the parameters Hurd & Zhou (2011) utilities MLE techniques together with CDS spread time series data.

The rest of the paper is organized as follows. In Section 2 we describe the model used in this paper, originally presented in Frey & Schmidt (2012). Section 3 gives a short recapitulation of the the Kushner-Stratonovich SDE describing the dynamics of the filtering probabilities in the model where we in particular focus on a homogeneous portfolio. Furthermore, in Section 3 we also restate explicit formulas for the conditional survival distribution originally derived in Herbertsson & Frey (2012). Continuing, in Section 4 we present an algorithm for estimating the parameters in the filtering models by using time-series data of index CDS spreads and classical maximum-likelihood algorithms. The calibration-approach incorporates the Kushner-Stratonovich SDE for the dynamics of the filtering probabilities. Our estimation-procedure heavily relies on the convenient expression for the conditional survival distribution since this formula will be used to back out the conditional distribution of the state space given the noisy information each time point, which also follows the Kushner-Stratonovich SDE. The transformation of the market index CDS spread to a survival probability under the real measure is done by first considering the market spread as a (transformed) proxy measure for the corresponding survival probability under the risk neutral measure given the noisy information. After this we use another proxy approximation between the survival probabilities under the risk-neutral and historical measure

Finally, in Section 5 we apply the maximum-likelihood algorithm on market data for historical CDS index spreads in order to estimate the parameters in the nonlinear filtering model for an homogeneous credit portfolio. Furthermore, other numerical studies are are performed as well as accompanying statistical testes.

2. The model

In this section we shortly recapitulate the model of Frey & Schmidt (2012). Thus, we will consider a reduced-form model driven by an unobservable background factor process X modelling the "true" state of the economy. For tractability reasons X is modelled as finite-state Markov chain. The factor process X is not directly observable. Instead prices of liquidly traded securities are given as conditional expectation with respect to the so called market filtration $\mathbb{F}^M = (\mathcal{F}_t^M)_{t\geq 0}$. The filtration $\mathbb{F}^M$ is generated by the factor process X plus noise, which will be specified in detail below. Intuitively speaking, this means that market investors observe the securities given the "noisy" history of the state of the economy. Furthermore, in the model of Frey & Schmidt (2012) the default times of all obligors are conditionally independent given the information of the factor process X. This setup is close to the one found in e.g. Graziano & Rogers (2009).

Frey & Schmidt (2012) treat the case with stochastic recoveries in a general theoretical setting. In this paper we will take a simplified approach and only consider deterministic recoveries, which up to the recent credit crises has been considered as standard in the credit literature.

2.1. **The factor process.** In this section we introduce the model that we will consider under the full information.

Let X_t be a finite state continuous time Markov chain on the state space $S^X = \{1, 2, \dots, K\}$ with generator $\boldsymbol{Q}$. Let $\mathcal{F}_t^X = \sigma(X_s; s \leq t)$ be the filtration generated by the factor process X. Consider m obligors with default times $\tau_1, \tau_2 \dots, \tau_m$ and let the mappings $\lambda_1, \lambda_2 \dots, \lambda_m$ be the corresponding $\mathcal{F}_t^X$ default intensities, where $\lambda_i : S^X \mapsto \mathbb{R}^+$ for each obligor i. This means that the default time τ_i are modelled as the first jump of a Cox-process, with intensity $\lambda_i(X_t)$. It is well known (see e.g. Lando (1998)) that given an i.i.d sequence $\{E_i\}$ where E_i is exponentially distributed with parameter one, such that all $\{E_i\}$ are independent of $\mathcal{F}_\infty^X$, then

$$\tau_i = \inf\left\{ t > 0 : \int_0^t \lambda_i(X_s) ds \geq E_i \right\}. \tag{2.1.1}$$

Hence, for any $T \geq t$ we have

$$\mathbb{P}\left[\tau_i > t \,|\, \mathcal{F}_T^X \right] = \exp\left(- \int_0^t \lambda_i(X_s) ds \right) \tag{2.1.2}$$

and thus

$$\mathbb{P}\left[\tau_i > t \right] = \mathbb{E}\left[\exp\left(- \int_0^t \lambda_i(X_s) ds \right) \right]. \tag{2.1.3}$$

Note that the default times are conditionally independent, given $\mathcal{F}_\infty^X$.

The states in $S^X = \{1, 2, \dots, K\}$ are ordered so that state 1 represents the best state and K represents the worst state of the economy. Consequently, the mappings $\lambda_i(\cdot)$ are chosen to be strictly increasing in $k \in \{1, 2, \dots, K\}$, that is $\lambda_i(k) < \lambda_i(k+1)$ for all $k \in \{1, 2, \dots, K-1\}$ and for every obligor in the portfolio.

2.2. **The market filtration and full information.** In this subsection we formally introduce the market filtration, that is the information observed by the market participation. Recall that the prices of all securities are given as conditional expectations with respect to this filtration. We also shortly discuss the full information $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$, which is the biggest filtration containing all other filtrations, where $(\Omega, \mathcal{G}, \mathbb{Q})$ with $\mathcal{G} = \mathcal{F}_\infty$ will be the underlying probability space assumed in the rest of this paper.

Let $Y_{t,i}$ denote the random variable $Y_{t,i} = 1_{\{\tau_i \leq t\}}$ and Y_t be the vector $Y_t = (Y_{t,1}, \dots, Y_{t,m})$. The filtration $\mathcal{F}_t^Y = \sigma(Y_s; s \leq t)$ represents the default portfolio information at time t, generated by the process $(Y_s)_{s \geq 0}$. Furthermore, let B_t be a one-dimensional Brownian motion independent of $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ and let $a(\cdot)$ be a function from $\{1, 2, \ldots, K\}$ to $\mathbb{R}$. Next, define the process Z_t as

$$Z_t = \int_0^t a(X_s)ds + B_t. \tag{2.2.1}$$

We here remark that Frey & Schmidt (2012) allows for multivariate Brownian motion B_t in (2.2.1) as well as a vector valued mapping $\boldsymbol{a}(\cdot)$ with same dimension as B_t. In this paper however, we restrict ourselves to only one source of randomness in the noise representation (2.2.1). Intuitively Z_t represents the noisy history of X_t and the functional form of Z_t given by (2.2.1) is a representation that are standard in the nonlinear filtering theory, see e.g. Davis & Marcus (1981). Following Frey & Schmidt (2012), we define the market filtration $\mathbb{F}^M = (\mathcal{F}_t^M)_{t\geq 0}$ as

$$\mathcal{F}_t^M = \mathcal{F}_t^Y \vee \mathcal{F}_t^Z. \tag{2.2.2}$$

We set the full information $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ to be the biggest filtration containing all other filtrations with $\mathcal{G} = \mathcal{F}_\infty$. We can for example let $\mathcal{F}_t$ be given by

$$\mathcal{F}_t = \mathcal{F}_t^X \vee \mathcal{F}_t^Y \vee \mathcal{F}_t^Z \vee \mathcal{F}_t^B \tag{2.2.3}$$

where $(\mathcal{F}_t^B)_{t\geq 0}$ is the filtration generated by the Brownian motion B_t. Note that $\mathcal{F}_t^X$ is not a subfiltration of $\mathcal{F}_t^Z$, and similarly, $\mathcal{F}_t^B$ is not contained in $\mathcal{F}_t^Z$.

3. The Kushner-Stratonovic SDE in the modell

In this section we study the Kushner-Stratonovic SDE in our filtering model. We use the same notation as in Frey & Schmidt (2012). First, define π_t^k as the conditional probability of the event $\{X_t = k\}$ given the market information $\mathcal{F}_t^M$ at time t, that is

$$\pi_t^k = \mathbb{P}\left[X_t = k \,|\, \mathcal{F}_t^M\right] \tag{3.1}$$

and let $\boldsymbol{\pi}_t \in \mathbb{R}^K$ be a row-vector such that $\boldsymbol{\pi}_t = \left(\pi_t^1, \ldots, \pi_t^K\right)$. In the sequel, for any $\mathcal{F}_t$-adapted process U_t we let $\widehat{U}_t$ denote the optional projection of $\widehat{U}_t$ onto the filtration $\mathcal{F}_t^M$, that is $\widehat{U}_t = \mathbb{E}\left[U_t \,|\, \mathcal{F}_t^M\right]$. To this end, we have for example

$$\begin{aligned}
\widehat{\lambda_i(X_t)} &= \mathbb{E}\left[\lambda(X_t) \,|\, \mathcal{F}_t^M\right] = \sum_{k=1}^K \lambda_i(k)\pi_t^k \\
\widehat{a(X_t)} &= \mathbb{E}\left[a(X_t) \,|\, \mathcal{F}_t^M\right] = \sum_{k=1}^K a(k)\pi_t^k.
\end{aligned}$$

Next, define $M_{t,i}$ and μ_t as

$$\begin{aligned}
M_{t,i} &= Y_{t,i} - \int_0^{t\wedge\tau_i} \widehat{\lambda_i(X_{s-})}ds \qquad \text{for } i = 1, \ldots, m \\
\mu_t &= Z_t - \int_0^t \widehat{a(X_s)}\,ds
\end{aligned}$$

In Frey & Schmidt (2012) it is shown that $M_{t,i}$ is an $\mathcal{F}_t^M$-martingale, for $i = 1, 2, \ldots, m$ and that μ_t is a Brownian motion with respect to the filtration $\mathcal{F}_t^M$. Thus, the vector $M_t = (M_{t,1}, \ldots, M_{t,m})$ is an $\mathcal{F}_t^M$-martingale.

Furthermore, Frey & Schmidt (2012) also proof the following proposition, which is a version of the Kushner- Stratonovic equations, adopted to the filtering models presented in this paper (originally developed in Frey & Schmidt (2012)).

Proposition 3.1. *With notation as above, the processes π_t^k satisfies the following K-dimensional system of SDE-s,*

$$d\pi_t^k = \sum_{\ell=1}^{K} \boldsymbol{Q}_{\ell,k} \pi_t^\ell dt + (\boldsymbol{\gamma}^k(\boldsymbol{\pi}_{t-}))^\top dM_t + \alpha^k(\boldsymbol{\pi}_t)\, d\mu_t\,, \tag{3.2}$$

where $(\boldsymbol{\gamma}^k)^\top = \left(\gamma_1^k(\boldsymbol{\pi}_t), \ldots, \gamma_1^k(\boldsymbol{\pi}_t)\right)$ *and the coefficients are given by*

$$\gamma_i^k(\boldsymbol{\pi}_t) = \pi_t^k \left(\frac{\lambda_i(k)}{\sum_{n=1}^{K} \lambda_i(n) \pi_t^n} - 1 \right), \quad 1 \le i \le m \tag{3.3}$$

and

$$\alpha^k(\boldsymbol{\pi}_t) = \pi_t^k \left(a(k) - \sum_{n=1}^{K} \pi_t^n a(n) \right), \quad 1 \le k \le K. \tag{3.4}$$

The K-dimensional SDE-system partly uses the vector notation for the M_t vector. However, as will be seen below, it will be beneficial to rewrite this SDE on component form, especially when we consider homogeneous credit portfolios. Thus, let us rewrite (3.2) on component form, so that

$$d\pi_t^k = \sum_{\ell=1}^{K} \boldsymbol{Q}_{\ell,k} \pi_t^\ell dt + \sum_{i=1}^{m} \gamma_i^k(\boldsymbol{\pi}_t) dM_{t,i} + \alpha^k(\boldsymbol{\pi}_t) d\mu_t. \tag{3.5}$$

Next, let us consider a homogeneous credit portfolio, that is, all obligors are exchangeable so that $\lambda_i(X_t) = \lambda(X_t)$ and $\gamma_i^k(\boldsymbol{\pi}_t) = \gamma^k(\boldsymbol{\pi}_t)$ for each obligor i and define N_t as

$$N_t = \sum_{i=1}^{m} Y_{t,i} = \sum_{i=1}^{m} 1_{\{\tau_i \le t\}}. \tag{3.6}$$

Furthermore, define $\boldsymbol{\lambda}$ as $\boldsymbol{\lambda} = (\lambda(1), \ldots, \lambda(K))$ and let $\boldsymbol{e}_k \in \mathbb{R}^m$ be a row vector where the entry at position k is 1 and the other entries are zero. For a homogeneous portfolio the results of Proposition 3.1 can be simplified to the following corollary, proved in Herbertsson & Frey (2012).

Corollary 3.2. *Consider a homogeneous credit portfolio with m obligors. Then, with notation as above, the processes π_t^k satisfies the following K-dimensional system of SDE-s,*

$$d\pi_t^k = \gamma^k(\boldsymbol{\pi}_t) dN_t + \boldsymbol{\pi}_t \left(\boldsymbol{Q} \boldsymbol{e}_k^\top - \gamma^k(\boldsymbol{\pi}_t) \boldsymbol{\lambda}^\top (m - N_t) \right) dt + \alpha^k(\boldsymbol{\pi}_t) d\mu_t \tag{3.7}$$

where $\gamma^k(\boldsymbol{\pi}_t)$ *and* $\alpha^k(\boldsymbol{\pi}_t)$ *are given by*

$$\gamma^k(\boldsymbol{\pi}_t) = \pi_t^k\left(\frac{\lambda(k)}{\boldsymbol{\pi}_t\boldsymbol{\lambda}^\top} - 1\right) \qquad \textit{and} \qquad \alpha^k(\boldsymbol{\pi}_t) = \pi_t^k\Big(a(k) - \sum_{n=1}^{K}\pi_t^n a(n)\Big). \tag{3.8}$$

From the SDE (3.7) in Corollary 3.2 we clearly see that the dynamics of the conditional probabilities π_t^k contains a drift part, a diffusion part and a jump part. The diffusion part is due to the $d\mu_{t,j}$ components and the jump part is due to the defaults in the portfolio, given by the differential dN_t.

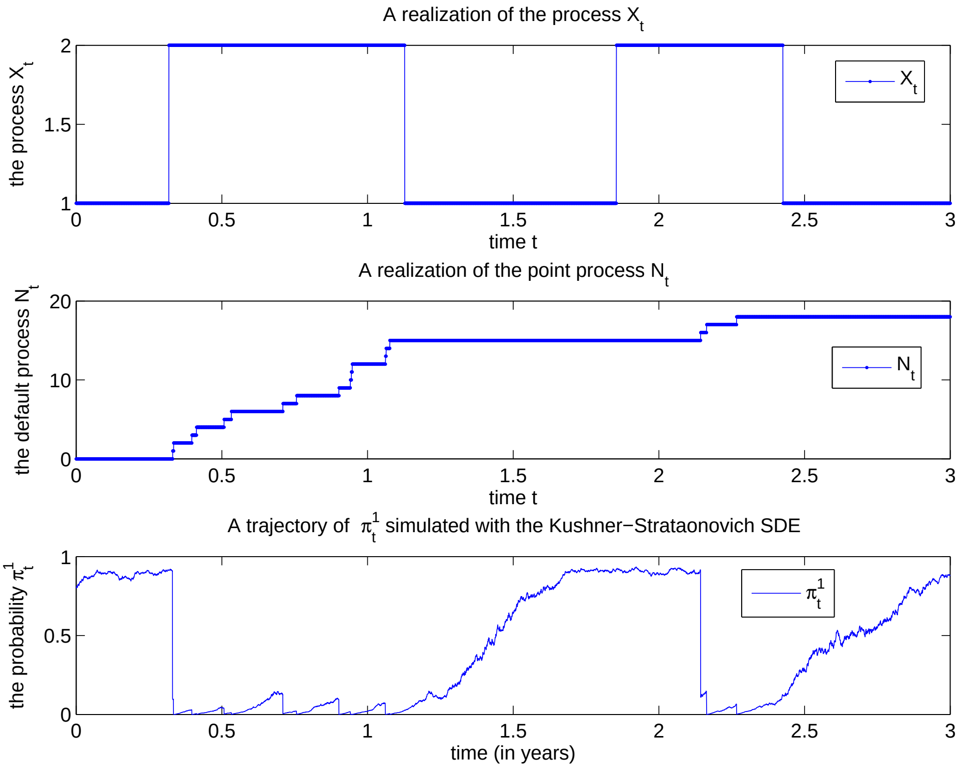

Figure 1. A simulated trajectory of X_t, N_t and π_t^1 where $K = 2$, $l = 1$ and $m = 125$.

Figure 1 visualizes a simulated path of π_t^1 given by (3.7) in Corollary 3.2 in an example where $K = 2$ and $m = 125$, using fictive parameters for $\boldsymbol{Q}$ and $\boldsymbol{\lambda}$ assuming $a(k) = c\cdot\ln\lambda(k)$ for a constant c. From the third Figure 1 we clearly see that π_t^1 has jump, drift and diffusion parts. The first and second subfigures in Figure 1 shows the corresponding trajectories for X_t and N_t. Note how the defaults presented by N_t clusters as X_t switches to state 2, representing the more worse economic state among $\{1, 2\}$.

Next, we briefly describe the conditional survival distribution given the noisy market information, that is, we state expressions for the quantities $\mathbb{P}\left[\tau_i > T \,\middle|\, \mathcal{F}_t^M\right]$ where $T > t$

and $\mathcal{F}_t^M$ described in Equation (2.2.2). First, we need to introduce some notation. If X_t is a finite state Markov jump process on $S^X = \{1, 2, \ldots, K\}$ with generator $\boldsymbol{Q}$, then, for a function $\lambda(x) : S^X \mapsto \mathbb{R}$ we denote the matrix $\boldsymbol{Q}_\lambda = \boldsymbol{Q} - \boldsymbol{I}_\lambda$ where $\boldsymbol{I}_\lambda$ is a diagonal-matrix such that $(\boldsymbol{I}_\lambda)_{k,k} = \lambda(k)$. Furthermore, we let $\mathbf{1}$ be a column vector in $\mathbb{R}^K$ where all entries are 1. The following theorem is a perquisite for all other results in this paper and is therefore a core result. A detailed proof is given in Herbertsson & Frey (2012).

Theorem 3.3. *Consider a homogeneous credit portfolio and let $\lambda(X_t)$ be the $\mathcal{F}_t^X$-intensity for obligor i. If $T \geq t$ then, with notation as above*

$$\mathbb{P}\left[\tau_i > T \,\middle|\, \mathcal{F}_t^M\right] = 1_{\{\tau_i > t\}} \boldsymbol{\pi}_t e^{\boldsymbol{Q}_\lambda (T-t)} \mathbf{1} \tag{3.9}$$

where the matrix $\boldsymbol{Q}_\lambda = \boldsymbol{Q} - \boldsymbol{I}_\lambda$ is defined as above.

The convenient expressions for the conditional survival distribution stated in (3.9) will be used in our estimation algorithm, as will be seen in the next section.

4. Parameter estimation in the filtering model

In this section we outline an algorithm for estimating the parameters in the filtering model by using time-series data of index CDS spreads and classical maximum-likelihood algorithms. The estimation-approach incorporates the Kushner-Stratonovich SDE for the dynamics of the filtering probabilities. Our estimation-procedure heavily relies on the convenient expression for the survival probability formula in Theorem 3.3 since this formula will be used to back out the distribution of the state space under the noisy information. We then use that these conditional probabilities must follow the Kushner-Stratonovich SDE at each time point.

The main ideas of our estimation-approach are first given in Subsection 4.1, and then the details of the algorithm is described in three steps presented through Subsection 4.2 to Subsection 4.4.

Finally, in Subsection 4.5 we give a brief discussion of related maximum-likelihood estimations both in filtering credit models as well as other intensity based frameworks described in recent academic papers

4.1. The main ideas and assumptions.

In this subsection we outline the main ideas in the algorithm for calibrating the parameters in our model using time-series data on CDS index spreads, the survival probability formula under the physical probability measure $\mathbb{P}$ and the Kushner-Stratonovic SDE.

Our task is to estimate the parameters $\boldsymbol{\theta} = (\boldsymbol{Q}, \boldsymbol{\lambda})$ from market data. Let $\{S_M(t, t+5)\}_{t \in \boldsymbol{t}^{(s)}}$ be a historical time-series trajectory of the 5-year market CDS index spreads observed at $N^{(s)}$ sample time points $\boldsymbol{t}^{(s)} = \{t_1^{(s)}, \ldots, t_{N^{(s)}}^{(s)}\}$. Recall that CDS spreads are priced under the risk neutral measure $\mathbb{Q}$ while we study a model under the real probability measure $\mathbb{P}$. We will use the time series data of the iTraxx Europe index CDS market spread as follows. At each time point t we consider the market spread $S_M(t, t+5)$ as a (transformed) proxy measure for the 5-year survival probability under the risk neutral measure $\mathbb{Q}$ given the noisy information. In order to relate $S_M(t, t+5)$ with a number under the physical probability measure $\mathbb{P}$ we make some crude assumptions (typically done in financial engineering) as follows. By letting $q_t^{(5)}$ be the noisy 5-year survival probability under $\mathbb{Q}$ and $p_t^{(5)}$ be the corresponding noisy 5-year survival probability under $\mathbb{P}$ we assume that

$$q_t^{(5)} = e^{-\gamma_t^Q 5} + \text{noise}_t \quad \text{and} \quad p_t^{(5)} = e^{-\gamma_t^P 5} + \text{noise}_t \tag{4.1.1}$$

where γ_t^Q and γ_t^P are stochastic non-negative functions. Inspired by the results in e.g. Berndt, Douglas, Duffie, Ferguson & Schranz (2008) we assume that γ_t^Q and γ_t^P are related as follows

$$\frac{\gamma_t^Q}{\gamma_t^P} \approx \beta \tag{4.1.2}$$

where β is a constant which can be fixed before the estimation alternatively be included with the other parameters in the estimation procedure. In Berndt et al. (2008) the authors determine the parameter β in (4.1.2) by regression over KMV Moody data and market CDS spreads, see also in Subsection 9.3 in McNeil, Frey & Embrechts (2005).

In view of (4.1.1) we will for a fixed outcome $\omega \in \Omega$ consider $\{S_M(t, t+5)\}_{t\in \boldsymbol{t}^{(s)}}$ as a observed trajectory of the 5-year market CDS index spreads sampled at the time points $\boldsymbol{t}^{(s)} = \{t_1^{(s)}, \ldots, t_{N^{(s)}}^{(s)}\}$ where $\{S_M(t, t+5)\}_{t\in \boldsymbol{t}^{(s)}}$ relates to the trajectory $\{\gamma_t^Q(\omega)\}_{t\in \boldsymbol{t}^{(s)}}$ as follows. For each time point $t \in \boldsymbol{t}^{(s)}$ we assume that the observed outcome $S_M(t, t+5)$ is related to the outcome of $\gamma_t^Q(\omega)$ by letting

$$\gamma_t^Q(\omega) + \text{noise}_t \approx \frac{S_M(t, t+5)}{(1-\phi)} \tag{4.1.3}$$

where ϕ is a constant recovery retrieved at default by a obligor in the homogeneous portfolio representing the CDS index. Note that the right hand side in (4.1.3) is "noisy" via the observed market CDS index spread $S_M(t, t+5)$ which fluctuates on a daily basis. Furthermore, the relation (4.1.3) is simply the standard credit triangle frequently used among market practitioners assuming a "flat" CDS term structure, i.e., assuming that γ_t^Q will be constant for all time points after t, see also Equation (9.11) on p.404 in McNeil et al. (2005). So in view of (4.1.2) and (4.1.3) we can for a fixed a outcome $\omega \in \Omega$ relate $\gamma_t^P(\omega)$ to $S_M(t, t+5)$ as

$$\gamma_t^P(\omega) + \text{noise}_t \approx \frac{S_M(t, t+5)}{(1-\phi)\beta} \tag{4.1.4}$$

and using the formula for $p_t^{(5)}$ in (4.1.1) then renders that

$$p_t^{(5)} \approx e^{\frac{-S_M(t,t+5)}{(1-\phi)\beta}5} + \text{noise}_t. \tag{4.1.5}$$

Finally, recalling that the quantity $p_t^{(5)}$ is the 5-year survival probability under the real probability measure $\mathbb{P}$ for an obligor under the noise information we have that

$$p_t^{(5)} = \mathbb{P}\left[\tau_i > t+5 \,\middle|\, \mathcal{F}_t^M\right]. \tag{4.1.6}$$

Thus, using (4.1.5) and (4.1.6) together with the explicit expression of $\mathbb{P}\left[\tau_i > t+5 \,\middle|\, \mathcal{F}_t^M\right]$ given by (3.9) in Theorem 3.3 and dropping the noise term in the right hand side of (4.1.5) we can for a fixed outcome $\omega \in \Omega$, relate the vector $\boldsymbol{\pi}_t(\omega)$ to the "observation" $S_M(t, t+5)$ as follows

$$e^{-\frac{S_M(t,t+5)}{(1-\phi)\beta}5} = \boldsymbol{\pi}_t(\omega)e^{\boldsymbol{Q}_\lambda 5}\mathbf{1} \tag{4.1.7}$$

where we used that $1_{\{\tau_i > t\}} = 1$, i.e. the average obligor is still alive at time t, that is, not all 125 obligors have defaulted up to time t. Also recall that the probability vector $\boldsymbol{\pi}_t(\omega)$ is "noisy by construction" which motivates the dropping of the noise term in the right hand side of (4.1.5) when using Equation (4.1.7)

We here remark that the formula (4.1.7) is a somewhat rough "back of the envelope" relation between the market 5-year CDS spread and the "noisy" 5-year survival probability under the physical probability measure $\mathbb{P}$ in the framework of Section 2. The main point in this section is not to develop a full-fledged theory for the relationship between the risk neutral and physical probability measures in the model by Frey & Schmidt (2012) but rather to give one example of how to estimate the parameters in this model via the Kushner-Stratonovic SDE given by Equation (3.7) in Corollary 3.2. The latter method is independent of how we extract the filtering probabilities $\boldsymbol{\pi}_t(\omega)$ from e.g. observed market CDS index spreads. Thus, going back to the the formula (4.1.7), we acknowledge that other methods describing the relationships between $S_M(t, t+5)$ and the probability vector $\boldsymbol{\pi}_t(\omega)$ (stated under the real measure $\mathbb{P}$) can be used, in particular when considering several maturities simultaneously. For example, if one wants to simultaneously use market CDS index spreads for e.g. four maturities, that is $S_M(t, t+3), S_M(t, t+5), S_M(t, t+7), S_M(t, t+10)$, when establishing a relationship between these spreads and the survival probability (expressed in terms of the filtering probabilities $\boldsymbol{\pi}_t(\omega)$) under the physical probability measure, then the "flat CDS term structure" formula (4.1.3) have to be replaced with another relationship. One such example is to let $\beta = \beta(t)$ be piecewise constant and replacing the proxy (4.1.3) with a more sophisticated relation such as a CDS pricer with piecewise constant intensity, allowing γ_t^Q to be bootstrapped from these spreads. After this, a similar version of (4.1.2) can be utilized with piecewise constant $\beta(t)$, in order to extract the corresponding γ_t^P on each time intervall. Finally, extending (4.1.7) is then trivial with the left hand side giving the formula for the survival probability in the piecewise constant framework, while the right hand side is the same but with different maturities $3, 5, 7$ and 10. Let us formalize the above ideas a bit more. Let J be an integer denoting how many market CDS spreads we want to include in our transformation, and let $\{T_1, \dots, T_J\}$ be a subset of $\{3, 5, 7, 10\}$. If e.g. $J = 4$ then $T_1 = 3, \dots, T_4 = 10$ and if $J = 1$ there are several options for T_1, but typically $T_1 = 5$. Thus, for a fixed outcome $\omega \in \Omega$, letting $\boldsymbol{x}_t$ denote the observed vector $\boldsymbol{x}_t = (S_M(t, t+T_1), \dots, S_M(t, t+T_j))$ we will for each $j = 1, \dots, J$ be able to find a mapping $F_{M,j}(\boldsymbol{x}_t, \boldsymbol{\beta})$ which relates the J CDS market spreads $\boldsymbol{x}_t$ to the filtering probability vector $\boldsymbol{\pi}_t(\omega)$ under the historical measure $\mathbb{P}$ as follows

$$F_{M,j}(\boldsymbol{x}_t, \boldsymbol{\beta}) = \boldsymbol{\pi}_t(\omega)e^{\boldsymbol{Q}_\lambda T_j}\mathbf{1} \quad \text{for} \quad j = 1, \dots, J \tag{4.1.8}$$

where $\boldsymbol{\beta}$ is a vector of parameters describing the relations $F_{M,j}$. Although the mappings $F_{M,j}(\boldsymbol{x}_t, \boldsymbol{\beta})$ may not be explicit in terms of analytical functions as in (4.1.7), we will always be able to numerically evaluate $F_{M,j}(\boldsymbol{x}_t, \boldsymbol{\beta}, T)$ given the arguments $\boldsymbol{x}_t, \boldsymbol{\beta}$ and $\{T_1, \dots, T_J\}$.

In the case $J = 1$ where we only use one market spread, say $S_M(t, t+5)$ so that $\boldsymbol{x}_t = S_M(t, t+5)$, $\boldsymbol{\beta} = (\beta, \phi)$, $T_1 = 5$ we can inspired by for example Equation (4.1.7) let $F_{M,1} = F_M$ be given by

$$F_M\left(\boldsymbol{x}_t, \boldsymbol{\beta}\right) = e^{-\frac{S_M(t,t+5)}{(1-\phi)\beta}5}. \tag{4.1.9}$$

This relation together with (4.1.8) for $T = 5$ is (4.1.7) restated.

Next, going back to the general case with $J \geq 1$, and for a fixed outcome $\omega \in \Omega$, we note that using $\boldsymbol{\pi}_t(\omega)\mathbf{1} = 1$ together with (4.1.7) will yield a linear equation system for $\boldsymbol{\pi}_t(\omega)$ as follows

$$\boldsymbol{A}\boldsymbol{\pi}_t^{\top}(\omega) = \boldsymbol{b}_t \tag{4.1.10}$$

where the matrix $\boldsymbol{A} = \boldsymbol{A}\left(\boldsymbol{\theta}\right) \in \mathbb{R}^{2\times K}$ is given by

$$\boldsymbol{A}\left(\boldsymbol{\theta}\right) = \begin{pmatrix} \boldsymbol{e}_1 e^{\boldsymbol{Q}_\lambda T_1}\mathbf{1} & ,\ldots\ldots, & \boldsymbol{e}_K e^{\boldsymbol{Q}_\lambda T_1}\mathbf{1} \\ \vdots & & \\ \boldsymbol{e}_1 e^{\boldsymbol{Q}_\lambda T_J}\mathbf{1} & ,\ldots\ldots, & \boldsymbol{e}_K e^{\boldsymbol{Q}_\lambda T_J}\mathbf{1} \\ 1 & ,\ldots\ldots, & 1 \end{pmatrix}. \tag{4.1.11}$$

Furthermore, the column vector $\boldsymbol{b}_t = \boldsymbol{b}\left(\boldsymbol{x}_t, \boldsymbol{\beta}\right) \in \mathbb{R}^J$ in (4.1.10) is given by

$$\boldsymbol{b}\left(\boldsymbol{x}_t, \boldsymbol{\beta}\right) = \begin{pmatrix} F_{M,1}\left(\boldsymbol{x}_t, \boldsymbol{\beta}\right) \\ \vdots \\ F_{M,J}\left(\boldsymbol{x}_t, \boldsymbol{\beta}\right) \\ 1 \end{pmatrix}. \tag{4.1.12}$$

Hence, given market spread vector $\boldsymbol{x}_t = (S_M(t, t+T_1), \ldots, S_M(t, t+T_J))$ and vector $\boldsymbol{\beta}$, we can for a fixed $\boldsymbol{\theta} = (\boldsymbol{Q}, \boldsymbol{\lambda})$ compute the matrix $\boldsymbol{A}$. Furthermore, if we assume that $\boldsymbol{A}^{-1}$ exists then $\boldsymbol{\pi}_t(\omega) = \left(\pi_t^1(\omega), \pi_t^2(\omega), \ldots, \pi_t^K(\omega)\right)$ is obtained by solving (4.1.10). Of course, we also have to make sure that the entries in $\boldsymbol{\pi}_t(\omega)$ are probabilities, that is $\pi_t^k(\omega) \in [0, 1]$ for every state k.

Our main idea is to estimate $\boldsymbol{\theta} = (\boldsymbol{Q}, \boldsymbol{\lambda})$ with maximum likelihood techniques by using observed market time-series data $\{\boldsymbol{x}_t\}_{t\in\boldsymbol{t}^{(s)}}$, the equation system for $\boldsymbol{\pi}_t(\omega)$ in (4.1.10) and the Kushner-Stratonovic SDE given by Equation (3.7) in Corollary 3.2. This will be done assuming that the vector $\boldsymbol{\beta}$ is exogenously given.

A necessary (but not sufficient) condition for $\boldsymbol{A}^{-1}$ to exists is to choose the number of states for X to be as many as the rows in $\boldsymbol{A}$, that is letting $K = J + 1$. As already mentioned, in this article we will for tractability reasons only consider the CDS index for one maturity (i.e. $J = 1$) where $T_1 = T = 5$ in the calibration of $\boldsymbol{\theta}$ which implies that that we set $K = 2$ for the Markov chain X. Note however that all ideas presented in this section can easily be carried over to an arbitrary number of states K, combined with $J = K - 1$ relations each given by (4.1.8). Since we in this paper only consider the case $J = 1$ with $T = 5$ then the vector $\boldsymbol{\beta}$ is reduced to a scalar β and the first row in $\boldsymbol{b}_t$ is given by Equation (4.1.9).

In the rest of this paper we define the function $a(X_t)$ used in construction (2.2.1) for the noisy observation Z_t, as

$$a(X_t) = c \ln \lambda(X_t) \quad \text{that is} \quad a_k = c \ln \lambda_k \quad \text{for each state } k \tag{4.1.13}$$

where c is a constant. The parametrization (4.1.13) have previously also been used in Frey & Schmidt (2012) and Frey & Schmidt (2011), see e.g. Subsection 5.3 in Frey & Schmidt (2012) and Example 7.6.1 in Frey & Schmidt (2011). Given (4.1.13) the unknown parameters to be estimated are then $\boldsymbol{\theta} = (\boldsymbol{Q}, \boldsymbol{\lambda}, c)$. Furthermore, we assume that the scalar β is exogenously given as well as the recovery rate ϕ.

In the following subsections we will describe the maximum likelihood algorithm for the parameters $\boldsymbol{\theta}$ in three steps. The first step (Subsection 4.2) is to extract $\boldsymbol{\pi}_t(\omega)$ by solving the linear system $\boldsymbol{A}\boldsymbol{\pi}_t^\top(\omega) = \boldsymbol{b}_t$ and to find conditions guarantying that all entries in $\boldsymbol{\pi}_t(\omega)$ are probabilities. The second step (Subsection 4.3) is to simplify and study the discrete version of the Kushner-Stratonovic SDE in (3.7). Finally, in the third step (Subsection 4.4) we use the previous two steps to derive a computationally tractable expression for the likelihood function of the parameters $\boldsymbol{\theta} = (\boldsymbol{Q}, \boldsymbol{\lambda}, c)$, as well as an analytical first-order condition for the variable c. We remark that the functional relationship (4.1.13) is only used in the final step when deriving the expression for the likelihood function, while the first two steps holds for any function $a(X_t)$ used in (2.2.1) for the noisy observation Z_t.

4.2. **Step 1: Extracting and constraining the filtering probabilities.** In this subsection we derive necessary and sufficient conditions so that the filtering probabilities satisfies $\pi_t^k(\omega) \in [0,1]$ for $k = 1, 2$, when $\boldsymbol{\pi}_t(\omega) = (\pi_t^1(\omega), \pi_t^2(\omega))$ is obtained by solving $\boldsymbol{A}\boldsymbol{\pi}_t^\top(\omega) = \boldsymbol{b}_t$ in (4.1.10). We also shortly discuss an extended method that can be used when the linear equation system method in fails (4.1.10), for example when $\boldsymbol{A}$ no longer is quadratic. The extended method will hold for arbitrary number states K.

For $K = 2$ we can easily solve $\boldsymbol{\pi}_t(\omega) = (\pi_t^1(\omega), \pi_t^2(\omega))$ in (4.1.10). From (4.1.11) with $J = 1$ and $K = 2$ we conclude that the second constraint in (4.1.10) trivially implies that $\pi_t^2(\omega) = 1 - \pi_t^1(\omega)$ for all $t \geq 0$ so for $K = 2$ it is enough to study $\pi_t^1(\omega)$. This fact inserted in the first constraint of (4.1.10) in the case $J = 1, K = 2$ and using (4.1.9) in $\boldsymbol{b}_t$ then immediately renders the probability $\pi_t^1(\omega) = \pi_t^1(\omega; \boldsymbol{\theta}, S_M(t,T))$ as

$$\pi_t^1(\omega) = \frac{e^{-\frac{S_M(t,t+5)}{(1-\phi)\beta}5} - \boldsymbol{e}_2 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1}}{\boldsymbol{e}_1 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1} - \boldsymbol{e}_2 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1}}. \tag{4.2.1}$$

We write $\pi_t^1(\omega) = \pi_t^1(\omega; \boldsymbol{\theta}, S_M(t, t+5))$ or alternatively $\pi_t^1(\omega) = \pi_t^1(\omega; \boldsymbol{\theta}, \beta, \phi, S_M(t, t+5))$ to emphasize that $\pi_t^1(\omega)$ is extracted from the market spread $S_M(t, t+5)$, given the fixed parameters $\boldsymbol{\theta}$, β and recovery ϕ. However, from now on we will for notational convenience just write $\pi_t^1(\omega)$ suppressing the dependence of $S_M(t, t+5)$, $\boldsymbol{\theta}$, β and ϕ.

Before we state conditions implying that $0 \leq \pi_t^1(\omega) \leq 1$, we need the following assumption. Given $\lambda(1) < \lambda(2)$ we assume that $\boldsymbol{Q}$ and $\lambda(1) < \lambda(2)$ are chosen so that

$$\mathbb{E}\left[e^{-\int_0^t \lambda(X_s)ds} \,\middle|\, X_0 = 1 \right] > \mathbb{E}\left[e^{-\int_0^t \lambda(X_s)ds} \,\middle|\, X_0 = 2 \right] \tag{4.2.2}$$

for all $0 \leq t \leq 5$. The condition (4.2.2) is intuitively clear for "small" t and also independent of the matrix $\boldsymbol{Q}$ and the size K of the state space as long as $\lambda(1) < \lambda(2) < \ldots, \lambda(K)$. To see this, note that in view of Theorem A.1, Appendix A in Herbertsson & Frey (2012) we have

$$\mathbb{E}\left[e^{-\int_0^t \lambda(X_s)ds} \,\middle|\, X_0 = k\right] = \boldsymbol{e}_k e^{\boldsymbol{Q}_\lambda t}\mathbf{1} \tag{4.2.3}$$

and recall that a first order approximation of the matrix $e^{\boldsymbol{Q}_\lambda t}$ is given by $\boldsymbol{I} + \boldsymbol{Q}_\lambda t$. Hence, for "small" t this implies

$$\boldsymbol{e}_k e^{\boldsymbol{Q}_\lambda t}\mathbf{1} \approx \boldsymbol{e}_k\left(\boldsymbol{I} + \boldsymbol{Q}_\lambda t\right)\mathbf{1} = 1 - \lambda(k)t \tag{4.2.4}$$

because $\boldsymbol{e}_k\mathbf{1} = 1$ for every k and

$$\boldsymbol{e}_k\boldsymbol{Q}_\lambda\mathbf{1} = \sum_{j=1}^{K}(\boldsymbol{Q}_\lambda)_{k,j} = \sum_{j=1}^{K}\boldsymbol{Q}_{k,j} - (\boldsymbol{I}_\lambda)_{k,k} = -\lambda(k) \quad \text{since} \quad \sum_{j=1}^{K}\boldsymbol{Q}_{k,j} = 0$$

and $\boldsymbol{Q}_\lambda = \boldsymbol{Q} - \boldsymbol{I}_\lambda$ where $\boldsymbol{I}_\lambda$ is a diagonal-matrix with $(\boldsymbol{I}_\lambda)_{k,k} = \lambda(k)$. Hence, assuming that t is "small" enough and inserting (4.2.4) into (4.2.3) yields

$$\mathbb{E}\left[e^{-\int_0^t \lambda(X_s)ds} \,\middle|\, X_0 = k\right] \approx 1 - \lambda(k)t. \tag{4.2.5}$$

Thus, if $\lambda(1) < \lambda(2) < \ldots, \lambda(K)$ then (4.2.5) will for "small" t imply that

$$\mathbb{E}\left[e^{-\int_0^t \lambda(X_s)ds} \,\middle|\, X_0 = 1\right] > \ldots > \mathbb{E}\left[e^{-\int_0^t \lambda(X_s)ds} \,\middle|\, X_0 = 1\right] \tag{4.2.6}$$

which motivates assumption (4.2.2) for any matrix $\boldsymbol{Q}$ and any size K of the state space, as long as $\lambda(1) < \lambda(2) < \ldots, \lambda(K)$. However, as t becomes larger the relationship (4.2.6) is no longer obvious since the first order approximation $\boldsymbol{I} + \boldsymbol{Q}_\lambda t$ is not accurate enough. Clearly, the region $[0, 5]$ where (4.2.6) holds for $0 \leq t \leq 5$ obviously depends on the matrix $\boldsymbol{Q}$. Hence, rather than stating explicit conditions for $\boldsymbol{Q}$ implying (4.2.2) for $0 \leq t \leq 5$, we therefore simply state that $\boldsymbol{Q}$ is chosen so that assumption (4.2.2) is true for all $0 \leq t \leq 5$. In the case $K = 2$ one can in fact prove that assumption (4.2.2) is true all $t > 0$ by using so called coupling arguments, see e.g in Lindvall (2002). In Subsection 5.2 we graphically display the difference $\Delta\mathbb{E}_t^{1,2}$ given by

$$\Delta\mathbb{E}_t^{1,2} = \mathbb{E}\left[e^{-\int_0^t \lambda(X_s)ds} \,\middle|\, X_0 = 1\right] - \mathbb{E}\left[e^{-\int_0^t \lambda(X_s)ds} \,\middle|\, X_0 = 2\right] \tag{4.2.7}$$

for $0 \leq t \leq 10$ with different parameters $\boldsymbol{Q}$ and $\boldsymbol{\lambda}$ obtained in different calibrations. Note that condition (4.2.2) is equivalent with the condition that the quantity in (4.2.7) is positive.

We can now give conditions implying that $0 \leq \pi_t^1(\omega) \leq 1$ when $K = 2$ as stated in the following lemma.

Lemma 4.1. *Under assumption* (4.2.2) *and with notation as above, we have that*

$$0 \leq \pi_t^1(\omega) \leq 1 \;\; \textit{if and only if} \;\; -\ln\left(\boldsymbol{e}_1 e^{\boldsymbol{Q}_\lambda 5}\mathbf{1}\right) \leq \frac{S_M(t, t+5)5}{(1-\phi)\beta} \leq -\ln\left(\boldsymbol{e}_2 e^{\boldsymbol{Q}_\lambda 5}\mathbf{1}\right). \tag{4.2.8}$$

Furthermore, (4.2.8) *will hold for all* $t \in \boldsymbol{t}^{(s)}$ *if*

$$-\ln\left(\boldsymbol{e}_1 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1}\right) \le \min_{t\in\boldsymbol{t}^{(s)}} \frac{S_M(t,t+5)5}{(1-\phi)\beta} \quad \textit{and} \quad \max_{t\in\boldsymbol{t}^{(s)}} \frac{S_M(t,t+5)5}{(1-\phi)\beta} \le -\ln\left(\boldsymbol{e}_2 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1}\right). \tag{4.2.9}$$

Proof. First, recall from (4.2.1) that

$$\pi_t^1(\omega) = \frac{e^{-\frac{S_M(t,t+5)}{(1-\phi)\beta}5} - \boldsymbol{e}_2 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1}}{\boldsymbol{e}_1 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1} - \boldsymbol{e}_2 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1}}. \tag{4.2.10}$$

Since $\lambda(1) < \lambda(2)$ by construction of the filtering model, then assumption (4.2.2) implies that the denominator in (4.2.10) is positive. Hence, $\pi_t^1(\omega) \ge 0$ if and only if the nominator in (4.2.10) is positive, that is if

$$e^{-\frac{S_M(t,t+5)}{(1-\phi)\beta}5} \ge \boldsymbol{e}_2 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1} \quad \text{or} \quad -\ln\left(\boldsymbol{e}_2 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1}\right)(1-\phi)\beta \ge S_M(t,t+5)5 \tag{4.2.11}$$

and this proves the upper equality in (4.2.8) since $\beta > 0$ and $0 < \phi < 1$ Next, from (4.2.10) we see that $\pi_t^1(\omega) \le 1$ if and only if

$$e^{-\frac{S_M(t,t+5)}{(1-\phi)\beta}5} - \boldsymbol{e}_2 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1} \le \boldsymbol{e}_1 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1} - \boldsymbol{e}_2 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1}$$

that is,

$$e^{-\frac{S_M(t,t+5)}{(1-\phi)\beta}5} \le \boldsymbol{e}_1 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1} \quad \text{or} \quad -\ln\left(\boldsymbol{e}_1 e^{\boldsymbol{Q}_\lambda t}\boldsymbol{1}\right)(1-\phi)\beta \le S_M(t,t+5)5 \tag{4.2.12}$$

and this proves the lower equality in (4.2.8). Finally, since the equalities in (4.2.11) and (4.2.12) holds for all time points, they will also hold for max and min over the sample time point set $\boldsymbol{t}^{(s)}$, which proves that (4.2.9) implies (4.2.8). □

It is interesting to note that for fixed parameters $\boldsymbol{\theta} = (\boldsymbol{Q}, \boldsymbol{\lambda})$, ϕ and given a historical time-series trajectory $\{S_M(t,t+5)\}_{t\in\boldsymbol{t}^{(s)}}$ of the 5-year market CDS index spreads observed at $N^{(s)}$ sample time points $\boldsymbol{t}^{(s)} = \{t_1^{(s)}, \ldots, t_{N^{(s)}}^{(s)}\}$, then (4.2.9) implies upper and lower bounds for the parameter β given by

$$\max_{t\in\boldsymbol{t}^{(s)}} \frac{S_M(t,t+5)5}{(\phi-1)\ln\left(\boldsymbol{e}_2 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1}\right)} \le \beta \le \min_{t\in\boldsymbol{t}^{(s)}} \frac{S_M(t,t+5)5}{(\phi-1)\ln\left(\boldsymbol{e}_1 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1}\right)} \tag{4.2.13}$$

where we remind the reader that $\ln\left(\boldsymbol{e}_2 e^{\boldsymbol{Q}_\lambda 5}\boldsymbol{1}\right) < 0$ and $0 < \phi < 1$.

In Subsection 5.2 we present some numerical examples with real market data that illustrates the use of Lemma 4.1 and condition (4.2.2) via (4.2.7) as well as other related issues.

The constraints for the filtering probabilities in Lemma 4.1 was presented for the case with two states. When there are more than two states we can still find constraints guaranteing that $\pi_t^k(\omega) \in [0,1]$ for every state $k = 1, \ldots, K$ where $K > 2$. In particular, when $K > 2$ we need to use $K-1$ other market CDS index observations (i.e. more maturities on the CDS index) so that $\boldsymbol{A}$ is quadratic making the linear system $\boldsymbol{A}\boldsymbol{\pi}_t^\top(\omega) = \boldsymbol{b}_t$ in (4.1.10) solvable. However, it is likely that the conditions for $\pi_t^k(\omega) \in [0,1]$ will be much more involved compared with the case $K = 2$. Furthermore, for larger K we may not find enough market instruments and thus $\boldsymbol{A}$ may no longer be quadratic. An alternative approach to find $\boldsymbol{\pi}_t(\omega)$ which simultaneously makes $\pi_t^k(\omega) \in [0,1]$ for all k and which works for any size K of the state space, is to solve the following quadratic programming problem

$$\begin{aligned} &\min_{\boldsymbol{\pi}_t} \left\| \boldsymbol{A}\boldsymbol{\pi}_t^\top - \boldsymbol{b}_t \right\|^2 \\ &\text{subject to} \\ &\boldsymbol{\pi}_t \mathbf{1} = 1 \\ &0 \leq \pi_t^k \leq 1 \quad \text{for} \quad k = 1, \dots, K \end{aligned} \tag{4.2.14}$$

where $\|\cdot\|$ is the standard Euclidian norm in $\mathbb{R}^K$. Thus, (4.2.14) can then be reformulated on the equivalently form

$$\begin{aligned} &\min_{\boldsymbol{\pi}_t} \left(\boldsymbol{\pi}_t \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{\pi}_t^\top - 2\boldsymbol{b}_t^\top \boldsymbol{A}\boldsymbol{\pi}_t^\top \right) \\ &\text{subject to} \\ &\boldsymbol{\pi}_t \mathbf{1} = 1 \\ &0 \leq \pi_t^k \leq 1 \quad \text{for} \quad k = 1, \dots, K. \end{aligned} \tag{4.2.15}$$

Solvers for quadratic programming problems (QP-problems) such as (4.2.15) are implemented in most standard math-software packages. Furthermore, the QP-problem (4.2.15) have a unique solution for $\boldsymbol{\pi}_t$ and the probabilistic constraints are automatically fulfilled.

4.3. **Step 2: Discretizing the Kushner-Stratonovic SDE.** In this subsection we consider the discrete version of the Kushner-Stratonovic SDE. Our outline is done for a general formula of the noise mapping $a(\cdot)$ and is thus not restricted to the form in Equation (4.1.13).

From (4.2.1) in Subsection 4.2 we conclude that for a fixed t, the quantity π_t^1 can be seen as a function of $S_M(t, t+5)$, β, ϕ and $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{Q})$. However, $\pi_t^1(\omega)$ must also satisfy the Kushner-Stratonovic equation stated in (3.7), that is, the following SDE

$$d\pi_t^1 = \gamma^1(\boldsymbol{\pi}_t)dN_t + \boldsymbol{\pi}_t \left(\boldsymbol{Q}\boldsymbol{e}_1^\top - \gamma^1(\boldsymbol{\pi}_t)\boldsymbol{\lambda}^\top (m - N_t) \right) dt + \alpha^1(\boldsymbol{\pi}_t)d\mu_t \tag{4.3.1}$$

where μ_t is a Brownian motion with respect to the filtration $\mathcal{F}_t^M$ and where we have assumed only one source of randomness in the noise that creates the noisy signal Z_t, given by (2.2.1). Furthermore, $\gamma^1(\boldsymbol{\pi}_t)$ and $\alpha^1(\boldsymbol{\pi}_t)$ are given in (3.8), that is

$$\gamma^1(\boldsymbol{\pi}_t) = \pi_t^1 \left(\frac{\lambda_1}{\pi_t^1\lambda_1 + (1-\pi_t^1)\lambda_2} - 1 \right) = \frac{\pi_t^1(\pi_t^1 - 1)(\lambda_1 - \lambda_2)}{\pi_t^1\lambda_1 + (1-\pi_t^1)\lambda_2} \tag{4.3.2}$$

and

$$\alpha^1(\boldsymbol{\pi}_t) = \pi_t^1 \left(a_1 - \pi_t^1 a_1 - (1-\pi_t^1)a_2 \right) = \pi_t^1 \left(1 - \pi_t^1 \right)(a_1 - a_2) \tag{4.3.3}$$

where we have used the fact that $\pi_t^2 = 1 - \pi_t^1$. Further, we let $a_k = a(k)$ and $\lambda_k = \lambda(k)$ for $k = 1, 2$.

We now consider a discrete version of the Kushner-Stratonovic SDE (4.3.1) on the mesh $t_1^{(s)}, t_2^{(s)}, \dots, t_{N^{(s)}}^{(s)}$ where, $t_n^{(s)} = n\Delta t$ which give raise to the following discrete SDE

$$\Delta\tilde{\pi}^1_{t_n^{(s)}} = \gamma^1(\tilde{\boldsymbol{\pi}}_{t_n^{(s)}})\Delta N_{t_n^{(s)}} + \tilde{\boldsymbol{\pi}}_{t_n^{(s)}} \left(\boldsymbol{Q}\boldsymbol{e}_1^\top - \gamma^1(\tilde{\boldsymbol{\pi}}_{t_n^{(s)}})\boldsymbol{\lambda}^\top (m - N_t)\right)\Delta t + \alpha^1(\tilde{\boldsymbol{\pi}}_{t_n^{(s)}})\Delta\mu_{t_n^{(s)}} \quad (4.3.4)$$

where $\tilde{\pi}^1_t$ is the solution to (4.3.4) and $\Delta\tilde{\pi}^1_{t_n^{(s)}}, \Delta N_{t_n^{(s)}}, \Delta\mu_{t_n^{(s)}}$ are given by

$$\Delta\tilde{\pi}^1_{t_n^{(s)}} = \tilde{\pi}^1_{t_{n+1}^{(s)}} - \tilde{\pi}^1_{t_n^{(s)}}, \quad \Delta N_{t_n^{(s)}} = N_{t_{n+1}^{(s)}} - N_{t_n^{(s)}}, \quad \text{and} \quad \Delta\mu_{t_n^{(s)}} = \mu_{t_{n+1}^{(s)}} - \mu_{t_n^{(s)}}$$

and $\tilde{\boldsymbol{\pi}}_t = (\tilde{\pi}^1_t, \tilde{\pi}^2_t)$. Next we use "financial engineering" and approximate the solution $\tilde{\pi}^1_t$ in (4.3.4) with the solution π^1_t to the SDE (4.3.1). Thus, replacing $\tilde{\pi}^1_t$ with π^1_t in (4.3.4) yields

$$\Delta\pi^1_{t_n^{(s)}} = \gamma^1(\boldsymbol{\pi}_{t_n^{(s)}})\Delta N_{t_n^{(s)}} + \boldsymbol{\pi}_{t_n^{(s)}} \left(\boldsymbol{Q}\boldsymbol{e}_1^\top - \gamma^1(\boldsymbol{\pi}_{t_n^{(s)}})\boldsymbol{\lambda}^\top (m - N_t)\right)\Delta t + \alpha^1(\boldsymbol{\pi}_{t_n^{(s)}})\Delta\mu_{t_n^{(s)}} \quad (4.3.5)$$

where $\Delta\pi^1_{t_n^{(s)}} = \pi^1_{t_{n+1}^{(s)}} - \pi^1_{t_n^{(s)}}$.

If there has been no defaults in the portfolio that constitute the market times series data during the sampling period $[t_1^{(s)}, t_{N^{(s)}}^{(s)}]$ then $N_t = 0$ for all $t \in [t_1^{(s)}, t_{N^{(s)}}^{(s)}]$ and consequently $\Delta N_t = 0$ for all $t \in [t_1^{(s)}, t_{N^{(s)}}^{(s)}]$ where we start our counting-process N_t at $t_1^{(s)}$. Up to the writing moment there has been no defaults of a entity in a on-the-run series of iTraxx Europe, which is the series that we consider in this paper. Hence, we will therefore set $\Delta N_{t_n^{(s)}} = 0$ and $N_{t_n^{(s)}} = 0$ for all $t_n^{(s)}$ in Equation (4.3.5). Thus, instead of (4.3.5), we will from now on consider the following discrete SDE

$$\Delta\pi^1_{t_n^{(s)}} = \boldsymbol{\pi}_{t_n^{(s)}} \left(\boldsymbol{Q}\boldsymbol{e}_1^\top - \gamma^1(\boldsymbol{\pi}_{t_n^{(s)}})\boldsymbol{\lambda}^\top m\right)\Delta t + \alpha^1(\boldsymbol{\pi}_{t_n^{(s)}})\Delta\mu_{t_n^{(s)}}. \quad (4.3.6)$$

Before deriving the likelihood function for the parameters $\boldsymbol{\theta}$, we will for notational convenience introduce some further notation. First, the CDS index spreads $S_M(t_n^{(s)}, t_n^{(s)} + 5)$ sampled at $t_n^{(s)}$ will be denoted by x_n, that is,

$$x_n = S_M(t_n^{(s)}, t_n^{(s)} + 5). \quad (4.3.7)$$

Further, since $\pi^1_{t_n^{(s)}}$ is extracted from the market spread x_n via (4.2.1) using the fixed parameters $\boldsymbol{\theta} = (\boldsymbol{Q}, \boldsymbol{\lambda})$, the quantities $\pi^1_{t_n^{(s)}}$, $\boldsymbol{\pi}_{t_n^{(s)}}$ and $\Delta\pi^1_{t_n^{(s)}}$ can be seen as functions of the "pair" $(\boldsymbol{\theta}, x_n)$ and will therefore for notational be denoted by $\pi^1(\boldsymbol{\theta}, x_n)$, $\boldsymbol{\pi}(\boldsymbol{\theta}, x_n)$ and $\Delta\pi_{n,1}(\boldsymbol{\theta}, x_n)$, that is

$$\pi^1(\boldsymbol{\theta}, x_n) = \pi^1_{t_n^{(s)}}, \quad \boldsymbol{\pi}(\boldsymbol{\theta}, x_n) = \boldsymbol{\pi}_{t_n^{(s)}} \quad \text{and} \quad \Delta\pi_{n,1}(\boldsymbol{\theta}, x_n) = \Delta\pi^1_{t_n^{(s)}} \quad (4.3.8)$$

where $\Delta\pi_{n,1}(\boldsymbol{\theta}, x_n) = \pi^1(\boldsymbol{\theta}, x_{n+1}) - \pi^1(\boldsymbol{\theta}, x_n)$. In the same spirit we let $g(\boldsymbol{\theta}, x_n)$ and $\alpha(\boldsymbol{\theta}, x_n)$ denote

$$g(\boldsymbol{\theta}, x_n) = \boldsymbol{\pi}_{t_n^{(s)}} \left(\boldsymbol{Q}\boldsymbol{e}_1^\top - \gamma^1(\boldsymbol{\pi}_{t_n^{(s)}})\boldsymbol{\lambda}^\top m\right) \quad \text{and} \quad \alpha(\boldsymbol{\theta}, x_n) = \alpha^1(\boldsymbol{\pi}_{t_n^{(s)}}) \quad (4.3.9)$$

where $\alpha^1(\boldsymbol{\pi}_t)$ is given by (4.3.3), that is $\alpha^1(\boldsymbol{\pi}_t) = \pi_t^1\left(1-\pi_t^1\right)(a_1-a_2)$. Hence, (4.3.7), (4.3.8) and (4.3.9) allows us to rewrite the discrete SDE (4.3.6) as

$$\Delta\pi_{n,1}\left(\boldsymbol{\theta}, x_n\right) = g\left(\boldsymbol{\theta}, x_n\right)\Delta t + \alpha\left(\boldsymbol{\theta}, x_n\right)\Delta\mu_{t_n^{(s)}} \tag{4.3.10}$$

where for a fixed $\boldsymbol{\theta}$ the sequence $\{\pi^1\left(\boldsymbol{\theta}, x_n\right)\}_{n=1}^{N^{(s)}}$ is extracted via (4.2.1) using the observed trajectory of market spreads $\{x_n\}_{n=1}^{N^{(s)}}$ of the CDS index, sampled at the time points $\{t_n^{(s)}\}_{n=1}^{N^{(s)}}$.

4.4. **Step 3: Deriving the likelihood function and first order conditions.** In this subsection we use the results from Subsection 4.2 and 4.3 to derive a computationally tractable expression for the likelihood function $\mathcal{L}$ of the parameters $\boldsymbol{\theta} = (\boldsymbol{Q}, \boldsymbol{\lambda}, c)$, as well as an analytical first-order condition for the variable c.

First, by construction recall that for a small time interval $[t, t+\Delta t]$ the incremental noise $\Delta\mu_t$ defined as $\Delta\mu_t = \mu_{t+\Delta t} - \mu_t$ will be normally distributed with zero mean and variance Δt, that is $\Delta\mu_t \sim \mathcal{N}(0, \Delta t)$. Inspired by this, and the fact that the mesh points $\{t_n^{(s)}\}_{n=1}^{N^{(s)}}$ are given by $t_n^{(s)} = n\Delta t$ where Δt is chosen to be small enough (one trading day), we then conclude that the right hand side in (4.3.10) is approximately normally distributed with mean $g\left(\boldsymbol{\theta}, x_n\right)\Delta t$ and variance $\alpha\left(\boldsymbol{\theta}, x_n\right)^2\Delta t$, that is

$$g\left(\boldsymbol{\theta}, x_n\right)\Delta t + \alpha\left(\boldsymbol{\theta}, x_n\right)\Delta\mu_{t_n^{(s)}} \sim \mathcal{N}\left(g\left(\boldsymbol{\theta}, x_n\right)\Delta t, \alpha\left(\boldsymbol{\theta}, x_n\right)^2\Delta t\right). \tag{4.4.1}$$

Furthermore, the random increments $\{g\left(\boldsymbol{\theta}, x_n\right)\Delta t + \alpha\left(\boldsymbol{\theta}, x_n\right)\Delta\mu_{t_n^{(s)}}\}_{n=1}^{N^{(s)}}$ are independent since μ_t is a Brownian motion. Given the extracted sequences $\{\Delta\pi_{n,1}\left(\boldsymbol{\theta}, x_n\right)\}_{n=1}^{N^{(s)}}$ which thus is a realization of the random variables defined in (4.4.1), we can then write our likelihood function $\mathcal{L}\left(\boldsymbol{\theta}|x_1, x_2, \ldots, x_{N^{(s)}}\right)$ as

$$\mathcal{L}\left(\boldsymbol{\theta}|x_1, \ldots, x_{N^{(s)}}\right) = \prod_{n=1}^{N^{(s)}} \frac{1}{\sqrt{2\pi\alpha\left(\boldsymbol{\theta}, x_n\right)^2\Delta t}} \exp\left(-\frac{\left(\Delta\pi_{n,1}\left(\boldsymbol{\theta}, x_n\right) - g\left(\boldsymbol{\theta}, x_n\right)\Delta t\right)^2}{2\alpha\left(\boldsymbol{\theta}, x_n\right)^2\Delta t}\right). \tag{4.4.2}$$

The optimal estimation $\hat{\boldsymbol{\theta}}$ obtained from (4.4.2) is preserved when considering the log-likelihood function $\ell\left(\boldsymbol{\theta}|x_1, \ldots, x_{N^{(s)}}\right)$ defined as $\ell\left(\boldsymbol{\theta}|x_1, \ldots, x_{N^{(s)}}\right) = \ln\mathcal{L}\left(\boldsymbol{\theta}|x_1, \ldots, x_{N^{(s)}}\right)$. By taking the logarithm of (4.4.2) we get

$$\ell\left(\boldsymbol{\theta}|x_1, \ldots, x_{N^{(s)}}\right) = -N^{(s)}\frac{\ln(2\pi\Delta t)}{2} - \sum_{n=1}^{N^{(s)}}\left(\ln\left|\alpha\left(\boldsymbol{\theta}, x_n\right)\right| + \frac{\left(\Delta\pi_{n,1}\left(\boldsymbol{\theta}, x_n\right) - g\left(\boldsymbol{\theta}, x_n\right)\Delta t\right)^2}{2\alpha\left(\boldsymbol{\theta}, x_n\right)^2\Delta t}\right). \tag{4.4.3}$$

Furthermore, recalling the notation of $\alpha\left(\boldsymbol{\theta}, x_n\right)$ in (4.3.9) together with (4.3.3) then implies that $\alpha^1(\boldsymbol{\pi}_t) = \pi^1\left(\boldsymbol{\theta}, x_n\right)\left(1-\pi^1\left(\boldsymbol{\theta}, x_n\right)\right)\left(a_1-a_2\right)$ which in (4.4.3) renders

$$\begin{aligned}\ell\left(\boldsymbol{\theta}|x_1,\ldots,x_{N^{(s)}}\right) = &-N^{(s)}\frac{\ln(2\pi\Delta t)}{2} - \sum_{n=1}^{N^{(s)}}\ln\left|\pi^1\left(\boldsymbol{\theta}, x_n\right)\left(1-\pi^1\left(\boldsymbol{\theta}, x_n\right)\right)\right| - N^{(s)}\ln\left|a_1-a_2\right| \\ &-\frac{1}{\left(a_1-a_2\right)^2}\sum_{n=1}^{N^{(s)}}\frac{\left(\Delta\pi_{n,1}\left(\boldsymbol{\theta}, x_n\right)-g\left(\boldsymbol{\theta}, x_n\right)\Delta t\right)^2}{2\left[\pi^1\left(\boldsymbol{\theta}, x_n\right)\left(1-\pi^1\left(\boldsymbol{\theta}, x_n\right)\right)\right]^2\Delta t}.\end{aligned} \tag{4.4.4}$$

So our maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ will then be given by $\operatorname{argmax}_{\boldsymbol{\theta}}\,\ell\left(\boldsymbol{\theta}|x_1,\ldots,x_{N^{(s)}}\right)$ where $\ell\left(\boldsymbol{\theta}|x_1,\ldots,x_{N^{(s)}}\right)$ is defined by (4.4.4). Observe that this is a maximization problem, but we can without loss of generally instead consider the minimization problem

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \operatorname*{argmin}_{\boldsymbol{\theta}}\,\hat{\ell}\left(\boldsymbol{\theta}|x_1,\ldots,x_{N^{(s)}}\right) \tag{4.4.5}$$

where $\hat{\ell}\left(\boldsymbol{\theta}|x_1,\ldots,x_{N^{(s)}}\right)$ is given by

$$\begin{aligned}\hat{\ell}\left(\boldsymbol{\theta}|x_1,\ldots,x_{N^{(s)}}\right) = &\,N^{(s)}\ln\left|a_1-a_2\right| + \sum_{n=1}^{N^{(s)}}\ln\left|\pi^1\left(\boldsymbol{\theta}, x_n\right)\left(1-\pi^1\left(\boldsymbol{\theta}, x_n\right)\right)\right| \\ &+\frac{1}{\left(a_1-a_2\right)^2}\sum_{n=1}^{N^{(s)}}\frac{\left(\Delta\pi_{n,1}\left(\boldsymbol{\theta}, x_n\right)-g\left(\boldsymbol{\theta}, x_n\right)\Delta t\right)^2}{2\left[\pi^1\left(\boldsymbol{\theta}, x_n\right)\left(1-\pi^1\left(\boldsymbol{\theta}, x_n\right)\right)\right]^2\Delta t}.\end{aligned} \tag{4.4.6}$$

Note that the term $N^{(s)}\frac{\ln(2\pi\Delta t)}{2}$ have been ignored in (4.4.6) since it do not affect the optimization routine when finding the parameters $\hat{\boldsymbol{\theta}}_{\text{MLE}}$. Before we continue, we remind the reader that in view of (4.1.13) we use the parameterizations $a_k = c\ln\lambda_k$ for $k=1,2$. This parametrization implies that we can rewrite $\hat{\ell}\left(\boldsymbol{\theta}|x_1,\ldots,x_{N^{(s)}}\right)$ as

$$\begin{aligned}\hat{\ell}\left(\boldsymbol{\theta}|x_1,\ldots,x_{N^{(s)}}\right) = &\,N^{(s)}\left(\ln c + \ln\left|\ln\frac{\lambda_1}{\lambda_2}\right|\right) + \sum_{n=1}^{N^{(s)}}\ln\left|\pi^1\left(\boldsymbol{\theta}, x_n\right)\left(1-\pi^1\left(\boldsymbol{\theta}, x_n\right)\right)\right| \\ &+\frac{1}{c^2\left(\ln\frac{\lambda_1}{\lambda_2}\right)^2}\sum_{n=1}^{N^{(s)}}\frac{\left(\Delta\pi_{n,1}\left(\boldsymbol{\theta}, x_n\right)-g\left(\boldsymbol{\theta}, x_n\right)\Delta t\right)^2}{2\left[\pi^1\left(\boldsymbol{\theta}, x_n\right)\left(1-\pi^1\left(\boldsymbol{\theta}, x_n\right)\right)\right]^2\Delta t}\end{aligned} \tag{4.4.7}$$

The MLE-estimation in (4.4.5) is a minimization problem, and to this setup we can therefore impose the necessary first order conditions that must hold for $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{\text{MLE}}$, viz.

$$\frac{\partial\hat{\ell}\left(\boldsymbol{\theta}|\boldsymbol{x}\right)}{\partial\theta_i} = 0 \quad \text{for all } \theta_i \in \boldsymbol{\theta} \tag{4.4.8}$$

where we for notational convenience let $\boldsymbol{x}$ denote $\boldsymbol{x} = (x_1,\ldots,x_{N^{(s)}})$. We want to find these first order conditions explicitly for some of our parameters $\boldsymbol{\theta} = (\boldsymbol{Q}, \boldsymbol{\lambda}, c)$. Let us consider the MLE-estimation $\hat{c}$ of c. First note that for fixed $\boldsymbol{Q}$ and $\boldsymbol{\lambda}$ the parameter

c does not enter in the expression (4.2.1) which (with the notation in (4.3.7)- (4.3.8)) is used to extract the the sequence $\{\pi^1(\boldsymbol{\theta}, x_n)\}_{n=1}^{N^{(s)}}$ from the observed trajectory of market spreads $\{x_n\}_{n=1}^{N^{(s)}}$ of the CDS index sampled at the time points $\{t_n^{(s)}\}_{n=1}^{N^{(s)}}$. Neither does c emerge in the expression of $g(\boldsymbol{\theta}, x_n)$ in (4.3.9). Hence, using these observations in (4.4.7) we immediately get

$$\frac{\partial \hat{\ell}(\boldsymbol{\theta}|\boldsymbol{x})}{\partial c} = N^{(s)}\frac{1}{c} - \frac{2}{c^3\left(\ln\frac{\lambda_1}{\lambda_2}\right)^2}\sum_{n=1}^{N^{(s)}}\frac{\left(\Delta\pi_{n,1}(\boldsymbol{\theta}, x_n) - g(\boldsymbol{\theta}, x_n)\Delta t\right)^2}{2\left[\pi^1(\boldsymbol{\theta}, x_n)\left(1-\pi^1(\boldsymbol{\theta}, x_n)\right)\right]^2\Delta t} \tag{4.4.9}$$

which together with condition (4.4.8), some rearranging and replacing c with $\hat{c}$ implies that the MLE-estimation $\hat{c}$ in our method must satisfy

$$\hat{c} = \pm\frac{1}{\left|\ln\frac{\lambda_1}{\lambda_2}\right|}\sqrt{\frac{1}{N^{(s)}}\sum_{n=1}^{N^{(s)}}\frac{\left(\Delta\pi_{n,1}(\boldsymbol{\theta}, x_n) - g(\boldsymbol{\theta}, x_n)\Delta t\right)^2}{\left[\pi^1(\boldsymbol{\theta}, x_n)\left(1-\pi^1(\boldsymbol{\theta}, x_n)\right)\right]^2\Delta t}}. \tag{4.4.10}$$

Since we only consider positive values of $\hat{c}$, the negative version in (4.4.10) can be disregarded. Note that the $\hat{c}$ in (4.4.10) can be seen as a function of the parameters $\boldsymbol{Q}$ and $\boldsymbol{\lambda}$, that is $\hat{c} = \hat{c}(\boldsymbol{Q}, \boldsymbol{\lambda})$ where the mapping $\hat{c}(\boldsymbol{Q}, \boldsymbol{\lambda})$ thus is positive version of the right hand side of (4.4.10). The expression in (4.4.10) is useful if we would like to estimate c in our model for fixed values of $\boldsymbol{Q}$ and $\boldsymbol{\lambda}$ since no numerical optimization routines is needed. To see this, note that $\hat{\ell}(c|\boldsymbol{x})$ is strictly convex in c if and only if

$$\frac{\partial^2\hat{\ell}(c|\boldsymbol{x})}{\partial^2 c} > 0 \tag{4.4.11}$$

so differentiation the right hand side in (4.4.9) with respect to c and inserting into (4.4.11) renders that $\hat{\ell}(c|\boldsymbol{x})$ is strictly convex in c if and only if

$$c < \frac{\sqrt{3}}{\left|\ln\frac{\lambda_1}{\lambda_2}\right|}\sqrt{\frac{1}{N^{(s)}}\sum_{n=1}^{N^{(s)}}\frac{\left(\Delta\pi_{n,1}(\boldsymbol{\theta}, x_n) - g(\boldsymbol{\theta}, x_n)\Delta t\right)^2}{\left[\pi^1(\boldsymbol{\theta}, x_n)\left(1-\pi^1(\boldsymbol{\theta}, x_n)\right)\right]^2\Delta t}} \tag{4.4.12}$$

where we have assumed a positive c. Hence, since $\sqrt{3} > 1$, we see that choosing a positive $\hat{c}$ in (4.4.10) will always yield the optimal maximum likelihood estimation $\hat{c}$ of the parameter c. Above all, no numerical optimization routine is needed to find the MLE-calibrated $\hat{c}$ in the case when $\boldsymbol{Q}$ and $\boldsymbol{\lambda}$ are given. In Section 5 we will perform an MLE estimation of the parameters $\boldsymbol{\theta} = (\boldsymbol{Q}, \boldsymbol{\lambda}, c)$ when $K = 2$ by numerically solving the minimization routine in (4.4.5) with $\hat{\ell}(\boldsymbol{\theta}|x_1, \ldots, x_{N^{(s)}})$ given by (4.4.7). In this approach we thus obtain the MLE-estimation $\hat{c}$ of c without using the explicit formula (4.4.10) but rather with a numerical minimization method jointly with the parameters of $\hat{\boldsymbol{Q}}$ and $\hat{\boldsymbol{\lambda}}$. However, after inserting the MLE estimators $\hat{\boldsymbol{Q}}$ and $\hat{\boldsymbol{\lambda}}$ into the formula in the right hand side of (4.4.10) we observe that the obtained estimate $\hat{c}_{\mathrm{fc}}$ (with the subscript "fc" denoting the first order condition) of c coincides with the corresponding MLE-estimate $\hat{c}$ of c obtained via the numerical minimization method of (4.4.7). For more on this see in Subsection 5.1.

The likelihood expression $\hat{\ell}(\boldsymbol{\theta}|\boldsymbol{x})$ in (4.4.7) is a highly nonlinear function of the parameters $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ and q_{12}, q_{21} (recall that q_{12}, q_{21} describes the generator $\boldsymbol{Q}$). Thus, it seems difficult to use the first order conditions in order to find analytical closed-form solutions for MLE-estimators of the parameters $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ and q_{12}, q_{21}. Instead we resort to numerical optimization routines when solving the minimization problem (4.4.5) for MLE-estimators of these parameters. For more on this see in Subsection 5.1.

4.5. **Other estimation methods.** In this subsection we give a brief discussion of related maximum-likelihood estimations for nonlinear filtering credit models and other credit models described in some recent academic papers.

One can estimate the parameters $\boldsymbol{\theta} = (\boldsymbol{Q}, \boldsymbol{\lambda}, c)$ in several different ways. In Frey & Schmidt (2012) the authors calibrate the parameter c by using the quadratic variation of the diffusion part of the index spread dynamics. Just as in our method, Frey & Schmidt (2012) observe that there were no defaults within the iTraxx Europe in their observation period for the time series, so the empirical quadratic variation of the index spreads is an estimate of the continuous part of the quadratic variation on the same index spread. They find a value of $c = 0.42$ for the 2009 data and $c = 0.71$ for 2006 data, see p.127 in Frey & Schmidt (2012). As will be seen in in Section 5, our MLE-estimation $\hat{c}$ obtained with the algorithm in Section 4 is in the same order as the corresponding estimation of c in the 2009 data case computed in Frey & Schmidt (2012).

The idea of using maximum likelihood (MLE) techniques under the risk neutral measure in credit risk modelling is not new. For example, in Hurd & Zhou (2011) the authors utilities MLE techniques together with CDS market time series data but in a structural credit risk model with noisy information. Furthermore, in Azizpour, Giesecke & Kim (2011) the authors uses MLE techniques with CDS index market data to find parameters in a self-exciting intensity based model under the risk neutral measure.

5. Numerical studies

In this section we perform the estimation and other numerical studies. First, in Subsection 5.1 we estimate the parameters for a homogeneous credit portfolio representing a CDS index by using real market data from the iTraxx Europe series sampled in the period November 2007 to June 2012. In this process we utilize the calibration algorithm presented in Section 4. Then, in Subsection 5.2 we perform some statistical and numerical observations in our estimated model.

5.1. **Parameter estimation.** In this subsection we estimate the parameters for a homogeneous credit portfolio representing an index CDS by using real market data from the iTraxx Europe series. This is done with the method presented in Section 4.

Our data set consist of the iTraxx Europe on-the-run series sampled in the period November 2007 to June 2012 (1169 sample points), collected from Reuters CreditViews. In the data set the observations are sampled daily so $\Delta t = 1/250$. The number of obligors representing the portfolio are $m = 125$ and we set the individual recovery rate to $\phi = 40\%$. Furthermore, we also assume that the parameter β is exogenously given by $\beta = 3$.

Our time series did not contain any defaults. With this time series we perform three different estimations of the parameters, using three different periods; these are November 2007 to February 2010 (596 sample points), March 2010 to June 2012 (573 sample points), and finally for the whole period November 2007 to June 2012 (1169 sample points). The period November 2007 to February 2010 (see upper plot in Figure 2) contains the turbulent credit crises emerged from the US subprime crises, then propagating through the Bear Sterns bailout in March 2008, the default of Lehman Brother in September 2008 as well as the following period of turmoil on the financial markets late 2008 and spring 2009 in US, Europe and the rest of the world. In late 2009/early 2010 this financial crises started to spill-over into a sovereign debt crises for several European countries (the PIIGS countries) belonging to the EURO-area. This motivates the separate estimation for the period March 2010 to June 2012 (see lower plot in Figure 2). Finally, we also perform an estimation of the parameters on the whole period November 2007 to June 2012.

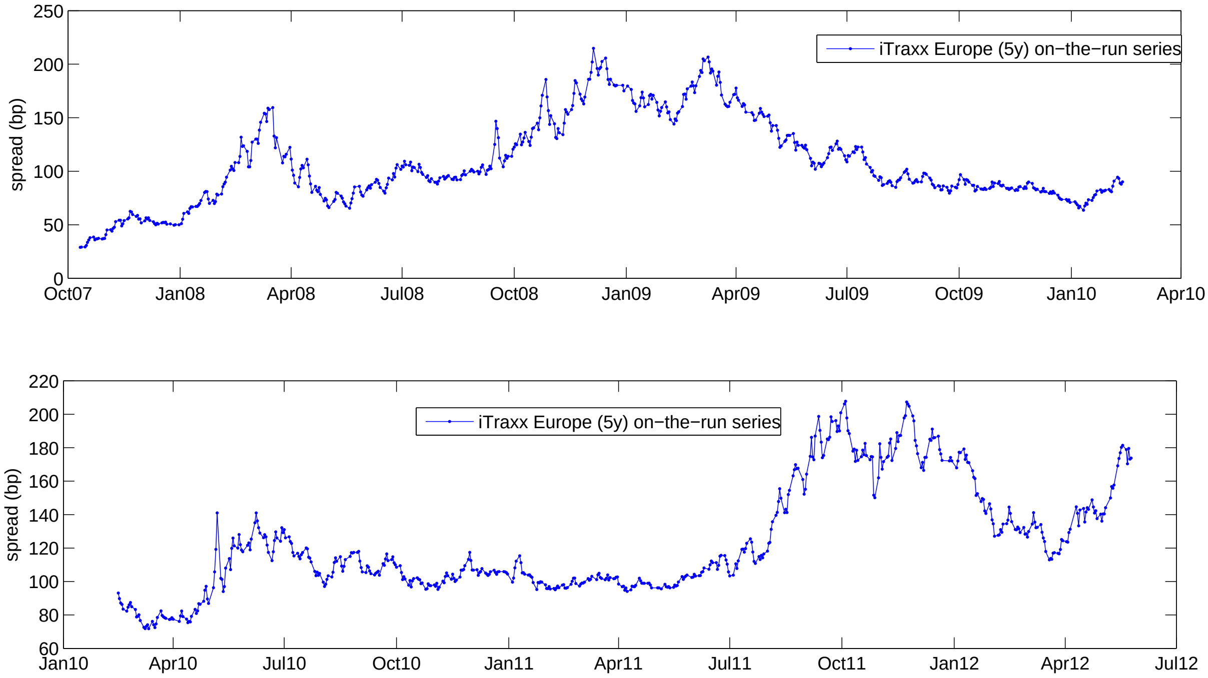

Figure 2. The five-year iTraxx Europe on the run series for November 2007 to February 2010 (top) and March 2010 to June 2012 (bottom).

Recall that we estimate the parameters $\boldsymbol{\theta} = (\boldsymbol{Q}, \boldsymbol{\lambda}, c)$ assuming the underlying factor process has two states, i.e. $K = 2$. Thus, we have $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ and q_{12}, q_{21} are the two transition intensities describing the generator $\boldsymbol{Q}$ for $K = 2$, so we can rewrite $\boldsymbol{\theta}$ as $\boldsymbol{\theta} = (q_{12}, q_{21}, \lambda_1, \lambda_2, c)$. The minimization problem (4.4.5) with $\hat{\ell}\left(\boldsymbol{\theta}|x_1, \ldots, x_{N^{(s)}}\right)$ given by

(4.4.7) is accompanied with constraints for the parameters $\boldsymbol{\theta} = (q_{12}, q_{21}, \lambda_1, \lambda_2, c)$. Obviously all parameters in $\boldsymbol{\theta}$ should be nonnegative. Furthermore, in order to make sure that assumption (4.2.2) holds so Lemma 4.1 can be invoked guaranteing that $0 \le \pi_t^1(\omega) \le 1$ for all time points t in our sample set, we have to find both upper and lower bounds for the parameters $\boldsymbol{\theta} = (q_{12}, q_{21}, \lambda_1, \lambda_2, c)$ given the fixed parameter $\beta = 3$. This is something which have to be checked "manually" for the given data set at hand and in our case we use the following upper bounds for λ_1, λ_2; $\lambda_1 < 0.004$ and $\lambda_2 < 0.09$. Similarly, we put lower bounds for q_{12}, q_{21} given by $q_{12} > 0.002$ and $q_{21} > 0.004$. Equipped with these constraints we solve (4.4.5) using the matlab-routine `fmincon` and the results are displayed in Table 1.

Table 1. MLE estimators $\boldsymbol{\theta}_{\text{MLE}}$, that is $\hat{q}_{12}, \hat{q}_{21}, \hat{\lambda}_1, \hat{\lambda}_2, \hat{c}$ for three different data sets. The last column is the estimate of c using (4.4.10) with input parameters $\hat{q}_{12}, \hat{q}_{21}, \hat{\lambda}_1, \hat{\lambda}_2$.

Time period	$\hat{q}_{12}$	$\hat{q}_{21}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{c}$	$\hat{c}_{\text{fc}}$
Nov 2007 - Feb 2010	0.002	0.004	0.001	0.09	0.3095	0.3095
Mar 2010 - Jun 2012	0.0118	0.004	0.001	0.09	0.32	0.32
Nov 2007 - Jun 2012	0.002	0.004	0.001	0.09	0.2753	0.2753

As can bee seen in Table 1, the two MLE-estimates $\hat{q}_{12}, \hat{q}_{21}$ hit their lower bounds except in one case, and by increasing these bounds the same thing happens again. It is interesting to note that when we solving (4.4.5) with e.g. λ_1 fixed, the obtained $\hat{q}_{12}, \hat{q}_{21}$ will no longer hit their lower bounds. We have not been able to find a good explanation of this phenomena. Furthermore, λ_2 reaches the upper bound for all three data sets.

After the calibration we use the estimated parameters $\hat{q}_{12}, \hat{q}_{21}, \hat{\lambda}_1, \hat{\lambda}_2$ to compute an alternative estimate of c via Equation (4.4.10) and this value $\hat{c}_{\text{fc}}$ coincide with the MLE estimate $\hat{c}$ for all three data sets, see in the last column in Table 1. The fact that $\hat{c} = \hat{c}_{\text{fc}}$ lends some confidence to the correctness of our implementation of the algorithm presented in Section 4 and Equation (4.4.5). However this observation is of course not a proof that our numerical implementation is correct.

5.2. **Some statistical and numerical observations.** In this subsection we perform some statistical observations on the data used in the previous subsection.

After finding the MLE-estimation $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ of the parameters $\boldsymbol{\theta} = (q_{12}, q_{21}, \lambda_1, \lambda_2, c)$ we can check the "quality" of our calibration, as follows. Let G_n be the right hand side in the discretized KS-SDE given by (4.4.1), i.e.

$$G_n = g\left(\boldsymbol{\theta}, x_n\right) \Delta t + \alpha_1\left(\boldsymbol{\theta}, x_n\right) \Delta \mu_{t_n^{(s)}}$$

and define H_n as

$$H_n = \frac{G_n - g\left(\boldsymbol{\theta}, x_n\right) \Delta t}{\alpha_1\left(\boldsymbol{\theta}, x_n\right) \sqrt{\Delta t}}. \tag{5.2.1}$$

Hence, given the observed trajectory of market spreads $\{x_n\}_{n=1}^{N^{(s)}}$ of the CDS index, sampled at the time points $\{t_n^{(s)}\}_{n=1}^{N^{(s)}}$, then if our model is describing the market dynamics realistically and if the parameters $\boldsymbol{\theta}$ are properly estimated, the random vector $\{H_n\}_{n=1}^{N^{(s)}}$ should be an i.i.d sequence where H_n are standard normal random variables, that is $H_n \sim \mathcal{N}(0,1)$.

Thus, replacing $\boldsymbol{\theta}$ with $\boldsymbol{\theta}_{\mathrm{MLE}}$ in (5.2.1) as well as changing G_n to the corresponding observation $\Delta\pi_{n,1}(\boldsymbol{\theta}_{\mathrm{MLE}}, x_n)$ in (5.2.1), we can then define the "calibrated implied noise" as $h_n(\boldsymbol{\theta}_{\mathrm{MLE}}, x_n)$

$$h_n(\boldsymbol{\theta}_{\mathrm{MLE}}, x_n) = \frac{\Delta\pi_{n,1}(\boldsymbol{\theta}_{\mathrm{MLE}}, x_n) - g(\boldsymbol{\theta}_{\mathrm{MLE}}, x_n)\Delta t}{\alpha_1(\boldsymbol{\theta}_{\mathrm{MLE}}, x_n)\sqrt{\Delta t}}. \tag{5.2.2}$$

Figure 3 displays the "estimated implied noise" $h_n(\boldsymbol{\theta}_{\mathrm{MLE}}, x_n)$ for the two periods November 2007 to February 2010 (top) and March 2010 to June 2012 (bottom) using the corresponding $\boldsymbol{\theta}_{\mathrm{MLE}}$ parameters in Table 1. If our model is describing the market dynamics realistically the sequence $\{h_n(\boldsymbol{\theta}_{\mathrm{MLE}}, x_n)\}_{n=1}^{N^{(s)}}$ should be a realization of the i.i.d sequence $\{H_n\}_{n=1}^{N^{(s)}}$ that is, an i.i.d sequence of standard normal random variables with mean $\hat{\eta} = 0$ and standard deviation $\hat{\sigma} = 0$.

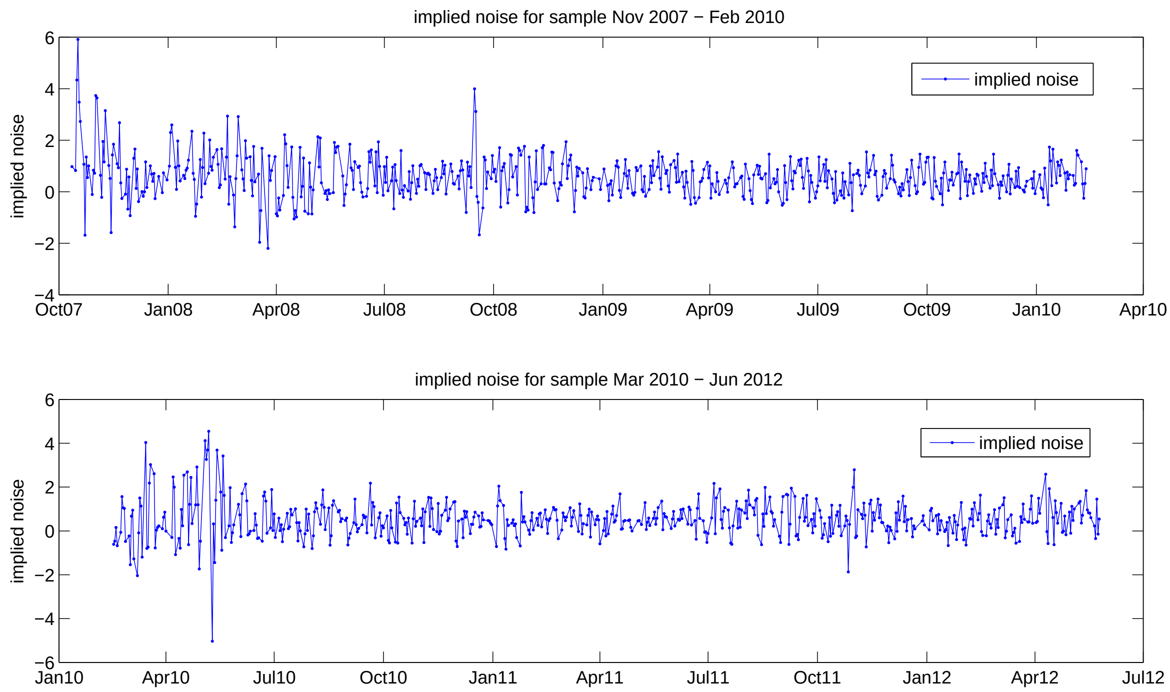

Figure 3. The calibrated implied noise computed with (5.2.2) for November 2007 to February 2010 (top) and March 2010 to June 2012 (bottom).

Table 2. Point estimators $\hat{\eta}$ and $\hat{\sigma}$ for the sample $\{h_n(\boldsymbol{\theta}_{\text{MLE}}, x_n)\}_{n=1}^{N(s)}$ as well as the corresponding 95% confidence intervals.

Time period	$\hat{\eta}$	$\hat{\sigma}$
Nov 2007 - Feb 2010	0.5822 [0.5167,0.6477]	0.8137 [0.77,0.8628]
Mar 2010 - Jun 2012	0.5298 [0.4601,0.5996]	0.8488 [0.8023,0.9011]
Nov 2007 - Jun 2012	0.6271 [0.5824,0.6719]	0.7792 [0.7489,0.8122]

In view of this observation we can now compute both a point estimate and a confidence interval estimate of the sample mean $\hat{\eta}_{\text{MLE}}$ and sample standard deviation $\hat{\sigma}_{\text{MLE}}$ for the sample $\{h_n(\boldsymbol{\theta}_{\text{MLE}}, x_n)\}_{n=1}^{N(s)}$. Recall that this data set is obtained via (5.2.2) after we have performed our MLE-estimation of $\boldsymbol{\theta}_{\text{MLE}}$. We use the matlab function `normfit` to compute the sample mean $\hat{\eta}$ and sample standard deviation $\hat{\sigma}$ the data set $\{h_n(\boldsymbol{\theta}_{\text{MLE}}, x_n)\}_{n=1}^{N(s)}$ as well as the accompanying 95% confidence intervals. If our model assumptions are correct, the estimate $\hat{\eta}$ should be approximately zero and $\hat{\sigma}$ should be around one, but as seen in Table 2 there is a deviation from these values.

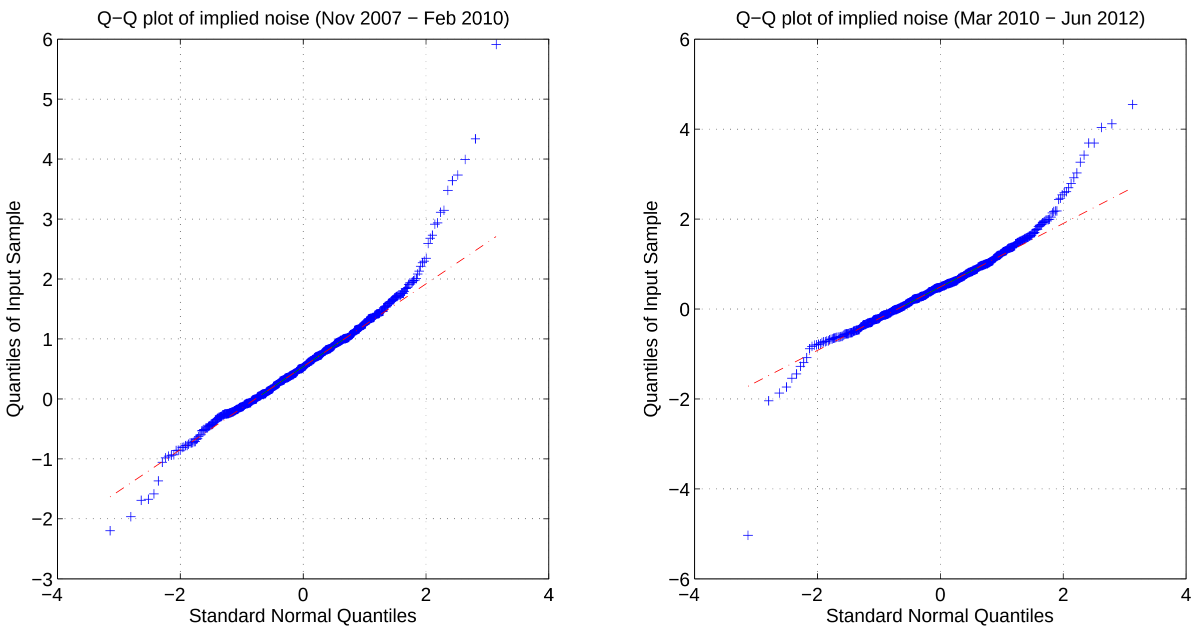

Figure 4. Q-Q plots for the "calibrated implied" noise computed with (5.2.2) for November 2007 to February 2010 (left) and March 2010 to June 2012 (right).

The Q-Q plots in Figure 4 for the two data set in Figure 3 shows that the lines formed by the empirical quantiles do not cross through origo $(0, 0)$ and do not have a slope of rate 1. Furthermore, as seen in Figure 5, a plot of the empirical cumulative distribution function (CDF) displays a deviation from the CDF of a standard normal random variable for both data sets. It is difficult to point out the exact reasons for the deviations in Table 2, Figure 4 and Figure 5, and this will be an issue for future research.

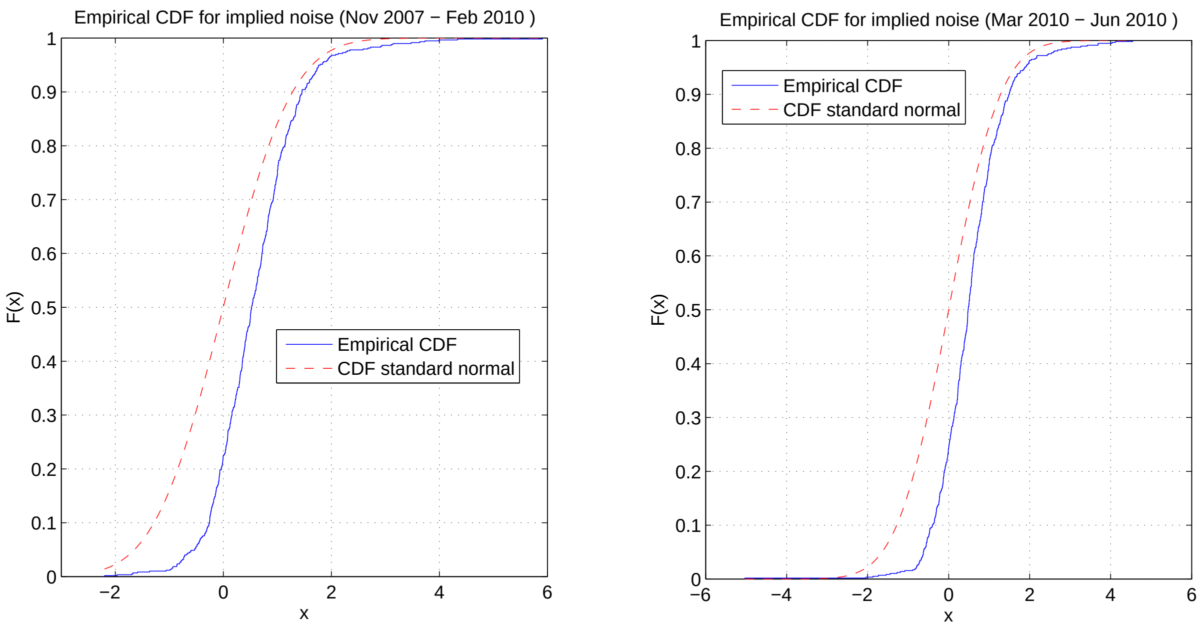

Figure 5. Empirical CDF plots for the "calibrated implied" for data set of November 2007 to February 2010 (left) and March 2010 to June 2012 (right). The dashed curve is the CDF of a standard normal.

Finally, before we end this sections we present some numerical examples with real market data that illustrates the correctness of assumption (4.2.2) and validity of Lemma 4.1 for our estimated model.

Recall from Equation (4.2.13) following Lemma 4.1 that under assumption (4.2.2) then $0 \leq \pi_t^1(\omega) \leq 1$ if and only if

$$\max_{t \in \boldsymbol{t}^{(s)}} \frac{S_M(t,t+5)5}{(\phi-1)\ln\left(\boldsymbol{e}_2 e^{\boldsymbol{Q}_\lambda 5}\mathbf{1}\right)} \leq \beta \leq \min_{t \in \boldsymbol{t}^{(s)}} \frac{S_M(t,t+5)5}{(\phi-1)\ln\left(\boldsymbol{e}_1 e^{\boldsymbol{Q}_\lambda 5}\mathbf{1}\right)} \qquad (5.2.3)$$

Table 3 displays the upper and lower bounds for β in (5.2.3) using the three data sets November 2007 to February 2010, March 2010 to June 2012 and November 2007 to June 2012. From Table 3 we clearly see that $\beta = 3$ lies in the corresponding intervals for all three data sets.

Table 3. The upper and lower bounds for the parameter β computed with (5.2.3) using the parameters $\boldsymbol{\theta}_{\mathrm{MLE}}$ in Table 1

Time period	lower bound	upper bound
Nov 2007 - Feb 2010	0.4024	3.488
Mar 2010 - Jun 2012	0.3891	3.698
Nov 2007 - Jun 2012	0.4024	3.488

Furthermore, Figure 6 and 7 demonstrate that $0 \leq \pi_t^1(\omega) \leq 1$ is true with the calibrated parameters $\boldsymbol{\theta}_{\mathrm{MLE}}$ in Table 1 for the two data sets November 2007 to February 2010 and March 2010 to June 2012. Figure 6 and 7 also displays the calibrated 5-year default probability computed with Equation (3.9) in Theorem 3.3.

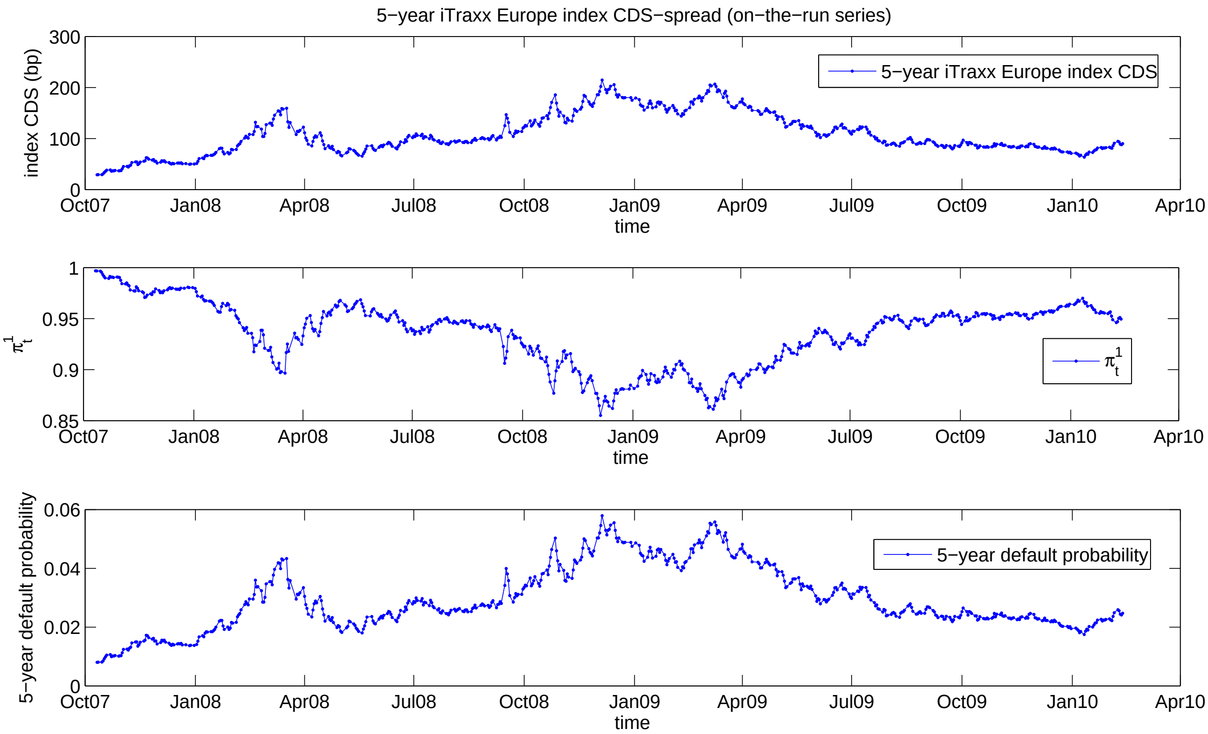

Figure 6. The five-year iTraxx Europe on the run series in November 2007 to February 2010 (top) and the corresponding filtering probability π_t^1 (middle) and the five-year default probability $\mathbb{P}\left[\tau_i < t+5 \,|\, \mathcal{F}_t^M\right]$ (bottom) both computed with the parameters $\boldsymbol{\theta}_{\mathrm{MLE}}$ in Table 1

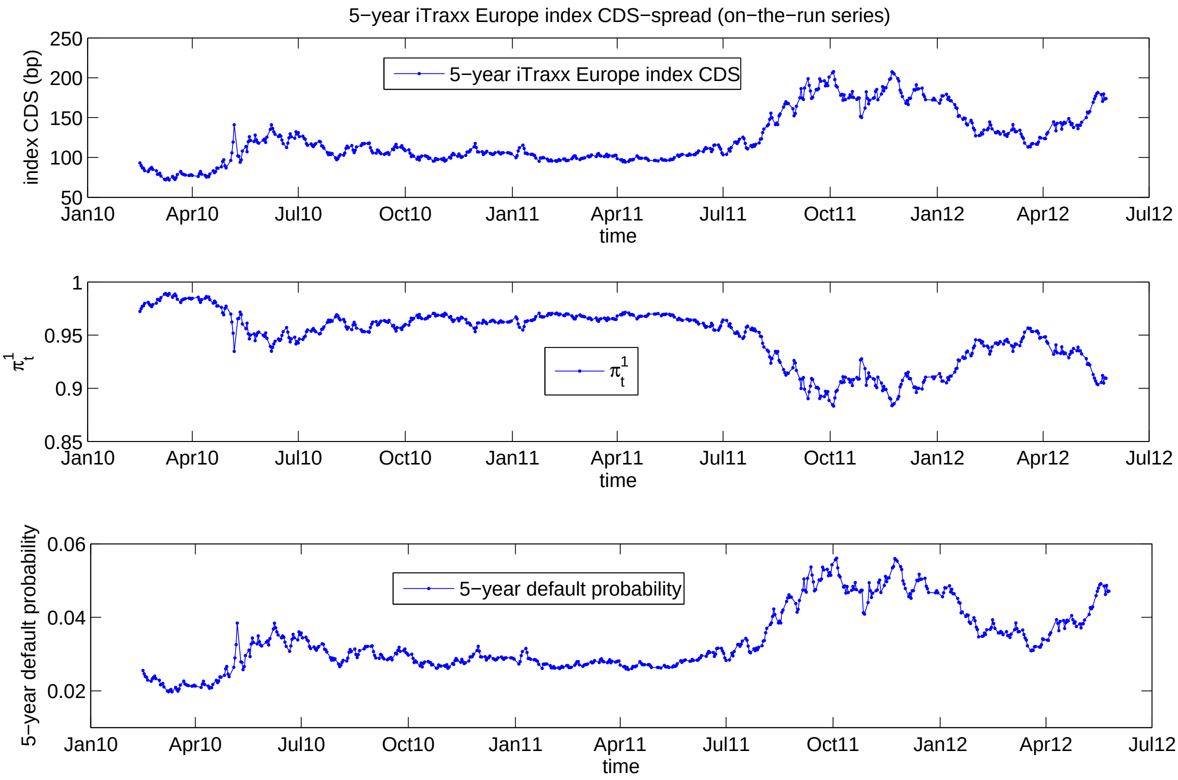

Figure 7. The five-year iTraxx Europe on the run series in March 2010 to June 2012 (top) and the corresponding filtering probability π_t^1 (middle) and the five-year default probability $\mathbb{P}\left[\tau_i < t+5 \,|\, \mathcal{F}_t^M\right]$ (bottom) both computed with the parameters $\boldsymbol{\theta}_{\text{MLE}}$ in Table 1

Finally, using the same parameters $\boldsymbol{\theta}_{\text{MLE}}$, in Figure 8 we also plot the quantity (4.2.7), that is

$$\Delta\mathbb{E}_t^{1,2} = \mathbb{E}\left[e^{-\int_0^t \lambda(X_s)ds} \,\middle|\, X_0 = 1\right] - \mathbb{E}\left[e^{-\int_0^t \lambda(X_s)ds} \,\middle|\, X_0 = 2\right]$$

for $0 \leq t \leq 10$ and see that this quantity positive in the period $0 \leq t \leq 10$. Consequently, assumption (4.2.2) will therefore also be true in the period $0 \leq t \leq 10$ which was necessary condition for Lemma 4.1 to be true. In fact, since our index CDS data had maturity 5 year it is enough for (4.2.7) to hold in the interval $0 \leq t \leq 5$

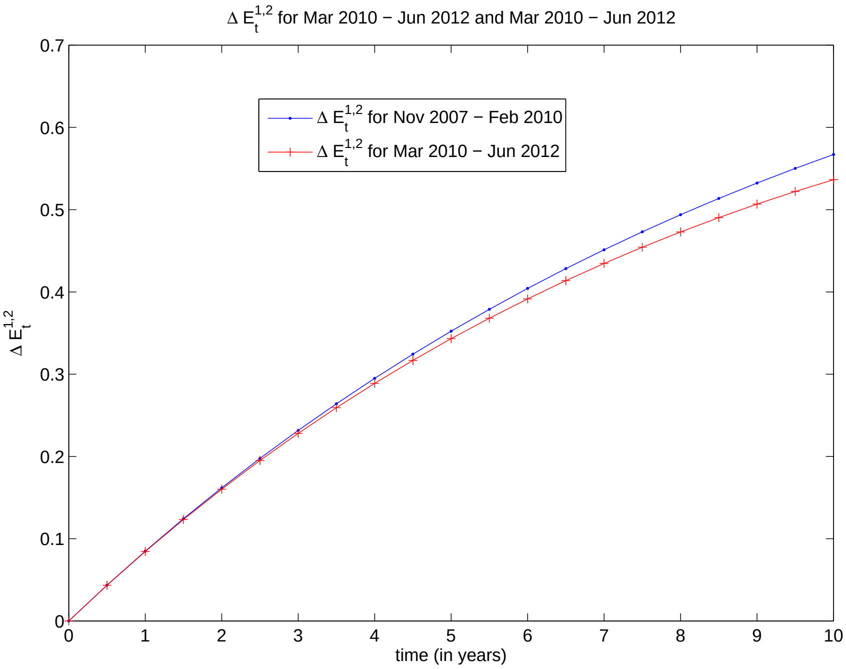

Figure 8. The quantity (4.2.7), computed in the period $0 \leq t \leq 10$ with the parameters $\boldsymbol{\theta}_{\mathrm{MLE}}$ in Table 1 for the two data sets November 2007 to February 2010 and March 2010 to June 2012.

References

Azizpour, S., Giesecke, K. & Kim, B. (2011), 'Premia for correlated default risk', *Journal of Economic Dynamics and Control* **35**(8), 1340–1357.

Berndt, A., Douglas, R., Duffie, D., Ferguson, M. & Schranz, D. (2008), Measuring default risk premia from default swap rates and edfs. Working paper.

Davis, M. H. A. & Marcus, S. I. (1981), An introduction to nonlinear filtering. M. Hazewinkel and J.C. Willems, eds, "Stochastic Systems: The Mathematics of Filtering and Identifications and Applications", Reidel Publishing Company, pp.53-75.

Frey, R. & Schmidt, T. (2011), Filtering and incomplete information in credit risk. Chapter 7 in "Recent Advancements in the Theory and Practice of Credit Derivatives", Brigo, D., Bielecki, T., and Patras, F. (eds), Wiley, New Jersey.

Frey, R. & Schmidt, T. (2012), 'Pricing and hedging of credit derivatives via the innovations approach to nonlinear filtering', *Finance and Stochastics* **16**(1), 105–133.

Graziano, G. D. & Rogers, L. C. G. (2009), 'A dynamic approach to the modeling of correlation credit derivatives using Markov chains', *International Journal of Theoretical and Applied Finance* **12**(1), 45–62.

Herbertsson, A. & Frey, R. (2012), Pricing CDS index options in a nonlinear filtering model. Working paper, available at www.economics.gu.se/herbertsson.

Hurd, T. R. & Zhou, Z. (2011), Structural credit risk using time-changed brownian motions: a tale of two models. Qorking paper. Department of Mathematics and Statistics, McMaster University, September 2011.

Lando, D. (1998), 'On Cox processes and credit risky securities.', *Review of Derivatives Research* **2**, 99–120.

Lindvall, T. (2002), *Lectures on the Coupling Method*, Dover Publications, New York.

McNeil, A. J., Frey, R. & Embrechts, P. (2005), *Quantitative Risk Management*, Princeton University Press, Oxford.

(Alexander Herbertsson), Centre For Finance, Department of Economics, School of Business, Economics and Law, University of Gothenburg, P.O Box 640, SE-405 30 Göteborg, Sweden

E-mail address: Alexander.Herbertsson@economics.gu.se

(Rüdiger Frey), Institute for Statistics and Mathematics, Vienna University of Economics and Business, Augasse 2-6, A-1090 Vienna, Austria

E-mail address: ruediger.frey@wu.ac.at