TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

DISSERTATION

Dependent Credit Rating Transitions and the Generalization of the Dybvig–Ingersoll–Ross Theorem

ausgeführt zum Zwecke der Erlangung des akademischen Grades einer Doktorin der Naturwissenschaften unter der Leitung von

Univ.-Prof. Dr. Uwe Schmock

E 105-1
Institut für Wirtschaftsmathematik
Forschungsgruppe Finanz- und Versicherungsmathematik

eingereicht an der Technischen Universität Wien
bei der Fakultät für Mathematik und Geoinformation

von

Dipl.-Math. Verena Goldammer

Matrikelnummer: 0627943
Geologengasse 8/14, 1030 Wien

Wien, am 20. Mai 2010

Kurzfassung der Dissertation

Diese Dissertation befasst sich mit einem neuen Ansatz für die Modellierung von Kreditrisiken und einer Erweiterung des bekannten Theorems von Dybvig, Ingersoll und Ross in der Zinstheorie. Im ersten Teil der Arbeit beschäftigen wir uns mit der Modellierung von abhängigen Kreditratingübergängen. Kreditratings beschreiben die Kreditwürdigkeit von Schuldnern und klassifizieren die Schuldner nach ihrer Wahrscheinlichkeit auszufallen. Dadurch hängt der Wert eines Kreditportfolios von den einzelnen Kreditratings der Schuldner ab, an die diese Kredite vergeben wurden. Ohne die Abhängigkeitsstruktur der Schuldner mit einzubeziehen, unterschätzen wir in der Regel das Risiko von gemeinsamen Ausfällen, die hohe Verluste in einem Kreditportfolio erzeugen.

Um diese Abhängigkeit zu modellieren, wollen wir den Effekt von Schocks auf zeitstetige Kreditratings internalisieren. Manche Ereignisse verursachen eine gemeinsame Veränderung der Kreditwürdigkeit von Schuldnern. Um diese Auswirkung zu modellieren, verwenden wir einen markierten Punktprozess. Dieser modelliert die zufälligen Zeiten der Änderungen und eine zufällige Marke gibt die möglichen Änderungen der Kreditratings an. In den meisten Ansätzen, die bisher zur Modellierung von abhängigen, stetigen Kreditratings angewandt werden, wird die Abhängigkeit nur dadurch eingeführt, dass die Intensität der Ratingübergänge der einzelnen Schuldner von dem Gesamtzustand der Ökonomie abhängt. Gleichzeitige Ratingänderungen von mehreren Schuldnern sind hier nicht möglich. Im Unterschied zu früheren Arbeiten erlauben wir in dieser Dissertation auch Ratingübergänge von mehreren Schuldnern zu derselben Zeit.

Dazu studieren wir ein spezielles Modell in unserem allgemeinen Rahmen ausführlicher. Die Ratingübergänge werden durch einen zeitlich homogenen Markovprozess beschrieben und alle Schuldner mit demselben Rating dürfen nur zu derselben Klasse wechseln. Wir nehmen an, dass es einen Poissonprozess mit Intensität $\lambda > 0$ gibt. Jedes Mal, wenn der Poissonprozess springt, wählen wir eine Funktion $s \in S^S = \{s : S \to S\}$, wobei S die Menge der Ratingklassen ist. Die Wahrscheinlichkeit die Funktion s zu wählen, wird durch eine Verteilung $P \in \mathcal{M}_1(S^S)$ gegeben. Nun können alle Schuldner mit Rating 1 entweder zum Rating $s(1)$ wechseln, oder in Klasse 1 bleiben, alle Schuldner mit Rating 2 zu $s(2)$ wechseln, oder in 2 bleiben, usw. Ein Übergang eines Schuldners mit Rating $x \in S$ findet tatsächlich mit Wahrscheinlichkeit $p_x \in [0, 1]$ statt, wobei jeder Schuldner unabhängig von den anderen die Klasse wechselt.

In unserem allgemeinen Modell gibt es zwei Möglichkeiten, um Abhängigkeit einzuführen. Einerseits dürfen alle Schuldner ihr Rating gleichzeitig ändern gemäß der Wahrscheinlichkeitsverteilung P. Dadurch können die Übergänge so verknüpft werden, dass zum Beispiel die Kreditwürdigkeit aller Schuldner gleichzeitig erhöht bzw. erniedrigt wird. Andererseits werden die Schuldner in derselben Klasse durch den Abhängigkeitsvektor $(p_x)_{x \in S}$ gekoppelt. Durch diese Struktur kann man gleichzeitige Ausfälle erklären und verschiedene Wahrscheinlichkeitsverteilungen von Verlustfunktionen erzeugen.

Unser Modell besitzt die nützliche Eigenschaft, dass die einzelnen Schuldner die glei-

chen Übergangsintensitäten für ihre Kreditratings haben. Dadurch ist es möglich, sich unabhängig bewegende Schuldner mit Schuldnern in unserem Modell zu vergleichen, zwischen denen es Abhängigkeit gibt. In einem einfachen Beispiel zeigen wir, dass die Kovarianz der Ausfälle von zwei Schuldnern in unserem Modell höher ist, und wie sie sich bei verschiedener Wahl der Verteilung P ändert. Die numerische Simulation eines Kreditportfolios zeigt in einem realistischeren Rahmen, wie die Wahl von P die Verteilung der Verlustfunktion eines Portfolios beeinflusst. Durch das Einführen von simultanen Ratingänderungen hat unser Modell mehr Freiheitsgrade und wir können verschiedene Arten von Abhängigkeitsstrukturen zwischen den Ratingklassen darstellen.

Im restlichen Kreditratingteil kalibrieren wir unser Modell an historische Ratingänderungen mithilfe des Maximum-Likelihood-Verfahrens. Für unser allgemeines Modell stellen wir die Likelihood-Funktion auf. Da die Berechnung des Maximums im Allgemeinen keine geschlossene analytische Form liefert, spezialisieren wir unser allgemeines Modell und betrachten die stark gekoppelte Irrfahrt, welche von Spitzer (1981) vorgestellt wurde. Wenn der wahre Parameter p_x in dem Intervall $(0, 1)$ liegt, dann ist der Maximum-Likelihood-Schätzer in diesem Fall eindeutig und wird durch die Nullstelle eines Polynoms gegeben, dessen Grad höchstens der Anzahl der Schuldner entspricht. Um die Genauigkeit des Schätzers zu beurteilen, zeigen wir, dass der Schätzer konsistent und asymptotisch normalverteilt ist. Das bedeutet, dass mit wachsender Zahl an Beobachtungen der Schätzer gegen den wahren Parameter konvergiert und der skalierte Fehler asymptotisch normalverteilt ist, was uns asymptotische Konfidenzintervalle für unsere Schätzung liefert.

Der zweite Teil der Dissertation widmet sich dem Verhalten von langfristigen Investitionsrenditen. Für das Bepreisen langfristiger Verträge modellieren Praktiker die Preise von Nullkuponanleihen mit langen Fälligkeiten, um sinnvolle Diskontierungssätze zu finden. Empirische Untersuchungen dieser Preise sind schwierig, da nur Nullkuponanleihen mit einer Fälligkeit von bis zu 30 Jahren gehandelt werden, aber z. B. für eine Leibrente die Zinssätze für Fälligkeiten von bis zu 100 Jahren benötigt werden. Um sinnvolle Modelle zu konstruieren, müssen wir daher wissen, wie sich langfristige Zinsen verhalten.

Dybvig, Ingersoll und Ross (1996) zeigen, dass die langfristigen Investitionsrenditen in einem arbitragefreien Marktmodell niemals fallen können. Falls also diese Renditen in einem Modell fallen, erlaubt das Modell Arbitrage. Sie setzen dabei voraus, dass der Grenzwert der Investitionsrenditen existiert. Es ist allerdings auch in bekannten Zinsmodellen wie dem Vašíček-Modell, dem Cox-Ingersoll-Ross-Modell oder dem Heath-Jarrow-Morton-Modell möglich, dass dieser Grenzwert nicht existiert, wie wir anhand von Beispielen in Kapitel 9 zeigen. In diesem Fall können wir das Theorem von Dybvig, Ingersoll und Ross nicht nutzen, um das Verhalten von langfristigen Investitionsrenditen zu erklären und zu entscheiden, ob das Modell Arbitrage zulässt. Deshalb verallgemeinern wir in dieser Doktorarbeit das Theorem auf Modelle, in denen der Grenzwert nicht existiert. Wir zeigen, dass der *Limes superior der Investitionsrenditen und der Terminzinssätze niemals fällt*, was wir als asymptotische Monotonie bezeichnen. Aus Sicht eines Investors ist der Limes superior die natürliche Erweiterung, da er langfristige Investitionen bevorzugt, deren Nullkuponanleihen einen hohen Ertrag liefern. Die Verallgemeinerung beweisen wir sowohl unter einer etwas schwächeren Annahme, als der Existenz eines risikoneutralen Wahrscheinlichkeitsmaßes, als auch unter der Annahme, dass es keine Arbitragemöglichkeit mit verschwindendem Risiko im Grenzwert gibt.

Neben dem Hauptsatz geben wir Bedingungen für asymptotische Minimalität des Limes superiors der Investitionsrenditen an. Das bedeutet, dass der Limes superior der Investitionsrenditen die größte Zufallsvariable ist, welche zu diesem Zeitpunkt bekannt ist und von dem zukünftigen Limes superior der Invesititonsrenditen dominiert wird.

Summary

This thesis addresses a new approach in the modeling of credit risk and an extension of the well-known Dybvig–Ingersoll–Ross theorem in interest rate theory. In the first part we deal with the modeling of dependent credit rating transitions. Credit ratings describe the credit-worthiness of obligors and classify the obligors depending on their probability to default. Therefore the value of credit portfolios depends on the rating of the underlying obligors. Without incorporation of dependencies between the ratings of the obligors into the modeling we underestimate in general the risk of joint defaults, which may induce high losses in the portfolio.

For modeling these dependent credit rating transitions we want to internalize the effect of shocks on the continuous-time rating. Some events can cause a simultaneous change of the credit quality of different obligors. To model this effect we use a marked point process. This process models random event times, and a random mark specifies the possible simultaneous credit rating transitions. In most of the models, so far used for the modeling of continuous-time credit rating transitions, the dependence is introduced via the dependence of the individual intensities on the current ratings of all the obligors. In this case no simultaneous up- or downgrades are possible. In contrast to the previous work, we allow the obligors to change their credit ratings simultaneously.

In more detail we study a special model within the general framework, that is a Markov jump process and all obligors with the same rating are only allowed to change to the same rating class at the same time. In our general model we use a Poisson process with intensity $\lambda > 0$. Each time, at which this process jumps, we choose a map s in $S^S = \{s : S \to S\}$, where S contains the credit rating classes. The probability for choosing the credit rating function $s \in S^S$ is given by a probability distribution $P \in \mathcal{M}_1(S^S)$. Then all obligors with rating 1 either remain in this class or change their rating to $s(1)$ with a certain probability, all obligors with rating 2 remain in this class or change to $s(2)$ and so on. More precisely, we independently throw a coin for each obligor with probability $p_x \in [0, 1]$ of heads, where $x \in S$ is the rating of the obligor. If the coin shows head, then the obligor changes the rating to $s(x)$, otherwise it remains in the rating class x.

Our general model provides two possibilities to introduce dependence. On the one hand, all obligors may change simultaneously according to the distribution P on the credit rating classes. For example, we can link the obligors such that either the credit quality of the obligors is upgraded or downgraded. On the other hand, obligors within the same rating class are coupled by the dependence vector $(p_x)_{x \in S}$ and may change their credit rating simultaneously. In this framework clustering of defaults is possible and we can generate different shapes of loss distributions.

The model has the useful property that the single obligors have the same individual transition intensities for their credit rating. Therefore, we can compare the independent case, where all obligors change their rating independently of each other, with our model,

where we have dependence between the obligors. A toy example shows that the covariance of the default of two obligors in our model is higher than in the independent case, and how the choice of the distribution P influences the covariance. Simulation of a credit portfolio illustrates the influence of the choice of P and p on the distribution of the losses for more realistic behavior of the obligors. By introducing simultaneous rating transitions we have more degrees of freedom in the modeling and are capable of reproducing different kind of dependence structures between the rating classes of the obligors.

In the remaining of the credit rating part we calibrate our model to historical rating transitions using maximum likelihood techniques. For the general model we state the likelihood function but the computation of the maximum is not tractable in general. In case of the extended strongly coupled random walk, which is a specialization of our general model, we are able to give an analytic expression for the maximum likelihood estimator. If the true dependence parameter p_x is in the interval $(0, 1)$, then the maximum likelihood estimator is unique and given by the root of a polynomial whose degree is bounded by the number of obligors and the coefficients contain the number of historical rating changes and the historical waiting times. To evaluate the accuracy of the estimators we also show consistency and asymptotic normality. Therefore with increasing number of observations the estimator converges to the true parameter. The scaled approximation errors converge in distribution to a normal distribution, which provides confidence intervals for our estimation.

The second part of the thesis is devoted to the behavior of long-term zero-coupon rates. To price long-term contracts, like life insurance policies, practitioners model zero-coupon bond prices with long-term maturities to find reasonable discount factors. Empirical investigations of these prices are difficult, since there are only zero-coupon bonds traded with maturity of up to 30 years, and for a life annuity, for example, discount factors for up to 100 years are needed. To construct reasonable models, we need to know how the long-term zero-coupon rates behave.

Dybvig, Ingersoll and Ross (1996) showed that long-term zero-coupon rates can never fall in an arbitrage-free market model under the assumption that the limit of the zero-coupon rates exists. Therefore, if the rates in a model decrease, it is not arbitrage-free. However, even in well-known interest rate models like the Vašíček model, the Cox–Ingersoll–Ross model or the Gaussian Heath–Jarrow–Morton model, it is possible that the limit of the zero-coupon rates does not exist, as we show with examples in Chapter 9. In this case we cannot use the Dybvig–Ingersoll–Ross theorem to explain the behavior of the long-term zero-coupon rates and decide if the model admits arbitrage opportunities. To assess also models where the limit does not exist we generalize the Dybvig–Ingersoll–Ross theorem in this thesis. We prove that the *limit superior of the zero-coupon rates and the forward rates never fall*, which is called asymptotic monotonicity. From the investor's point of view, the limit superior is the natural extension, because he prefers for long-term investments those zero-coupon bonds which give a high investment return. We prove this generalization either under a slightly weaker condition than assuming the existence of a forward risk neutral probability measure, or the assumption that there is no arbitrage opportunity in the limit with vanishing risk.

Besides the main theorem, we state conditions for asymptotic minimality of the limit superior of the zero-coupon rates. That means, the limit superior of the zero-coupon rates is the largest random variable, which is known at this time and dominated by the future limit superior of the long-term zero-coupon rates.

Acknowledgements

This work would have not been fulfilled without the contribution, either direct or indirect, of many people and best friends. At this point I want to take the opportunity to thank them. First, I want to thank my supervisor Prof. Uwe Schmock. He introduced me to the subject of my work and with his help I published my first paper in a recognized journal. Furthermore, he always encouraged me to present my work on international conferences that gave me the opportunity to improve my presentation skills and to see the world. I am also deeply grateful to my co-examiner Prof. Wolfgang Runggaldier for devoting a lot of time on a thorough revision of the thesis and his many helpful comments that helped improving the presentation. Furthermore, my thanks go out to all my colleagues at FAM for the nice and friendly research atmosphere.

I gratefully acknowledge financial support from UniCredit Bank Austria and the Christian Doppler Research Association (CDG) via the PRisMa laboratory. Especially, I want to thank Christoph Konvicka and Matthias Baron, my colleagues from Bank Austria, for the good collaboration and giving me insights in the practical side of financial mathematics. Further, I thank the WWTF for financial support under grant MA04-013 and grant MA09-003.

I owe a lot to the support by my friends, my family and especially my boyfriend, but this private part of the acknowledgements is done separately.

Contents

Part I

Modeling and Estimation of Dependent Credit Rating Transitions

Chapter 1

Introduction

On 15. September 2008 the investment bank Lehmann Brothers failed, which was a shock for the financial world. The bankruptcy caused huge losses for banks, which had invested in Lehmann and immediately affected the credit quality of these banks. Simultaneous changes of the credit quality can also be observed in other economical sectors. In Iceland, e. g., the volcano Eyjafjallajökull erupted and on 15. April 2010 the volcanic ash forced the aviation authorities to close most of the European airports. Most of the flights were cancelled and the European airlines suffered huge losses. The unexpected eruption of the volcano caused a joint deterioration of the credit quality of the airlines. These two examples show that special events can influence the credit rating of firms immediately and cause a simultaneous credit rating change. Without consideration of such joint changes in credit quality, we underestimate the risk of a portfolio with such firms as underlyings.

In this part of the thesis we address such simultaneous changes of the credit quality. To this end, we internalize shocks and given there is a shock, the credit ratings of the firms[1] are simultaneously affected. We use a marked point process to model the credit ratings of the firms. At each event time the credit ratings change according to a random mark. This is a random function that maps the rating of each firm and an idiosyncratic component to another rating class. Therefore all firms can change their rating simultaneously.

Credit rating transitions are already discussed in the literature. An overview of credit rating models is given in the books of e. g. Bielecki and Rutkowski (2002), Duffie and Singleton (2003), or Lando (2004). A usual assumption is that the credit ratings follow a Markov process. Jarrow, Lando and Turnbull (1997) consider a time-homogeneous Markov process for the credit rating of one firm and assume that the transitions of different firms are independent. Since there is empirical evidence that the probability of credit rating transitions varies over time, several authors suggest to model the credit rating of a single firm by a time-inhomogeneous Markov process, e. g. Couderc and Renault (2005), Duffie and Singleton (1999), Kavvathas (2000), Lando (1998), or Lando and Skødeberg (2002). The basic assumption in these models is that the intensities of the credit rating transitions of one firm are time-dependent and depend on macroeconomic factors, like e. g. the business cycle. Therefore the credit rating process of different firms depends on the same macroeconomic factors, but conditionally on these factors the credit ratings are independent.

Jarrow and Yu (2001) show that the dependence through macroeconomic factors is apparently insufficient to explain the clustering of defaults during a recession. To capture the microeconomic structure between two firms they add a counterparty specific jump

[1]In the following we tacitly assume that all obligors are firms. If the obligors are private persons or states, we can use the same mathematical framework to model dependent credit rating transitions.

term to the intensity of the credit ratings. Yu (2007) generalizes this idea and provides a construction for the distribution of the dependent defaults of more than two firms, where the intensities depend on the observed defaults and a common macroeconomic factor. In the model of Jarrow et al. the contagion is based on a direct link between the different firms. Another explanation for credit contagion are information effects, i. e., the default intensity of the firms depends on unobservable macroeconomic variables (often called *frailty*), and the default of one firm gives information on the default risk of the other firms (see e. g. Collin-Dufresne et al. (2003), Duffie et al. (2006), or Schönbucher (2003)). Credit contagion in the discrete-time setting is also considered by Davis and Lo (2001). They assume that the default of a firm directly causes simultaneous defaults of further firms, similar to a spread of an infection. Horst (2007) models a cascade effect of defaults with a mean-field approach in which the rating transitions depend only on the average rating of the other firms. After an economic shock some firms default immediately. This causes additional downgrades through the deterioration of the average rating and triggers a chain reaction. For description of the direct interaction between the firms, Egloff, Leippold and Vanini (2007) use a directed graph representation. In the continuous-time setting the common way to model contagion is that the individual credit rating intensity depends on the credit rating of the other firms. In this case simultaneous credit rating transitions are not possible. To model this effect, interacting particle systems have become popular because these models provide a convenient description of the direct interaction between the firms. Giesecke and Weber (2003, 2006) apply the voter model to construct the dependence structure between the defaults. The voter model is a spin system, i. e., each coordinate has only two possible states, interpreted as default of a special firm or not. The default of one firm increases the default intensity of neighbors of the firm. Bielecki, Crèpey, Jeanblanc and Rutkowski (2007) introduce a Markov model modulated by a Lévy process to define the credit rating transitions. The dependence is introduced via the dependence of the individual intensities on the current ratings of all other firms in the economy. Bielecki, Vidozzi and Vidozzi (2006) calibrate this model at market data and price several credit derivatives. Frey and Backhaus (2007) and Dai Pra, Runggaldier, Sartori and Tolotti (2009) use a mean-field interaction model. Here, the firms are divided into several groups and only the number of defaulted firms in the different groups influences the individual default intensities.

Another main approach to model dependence of the defaults are copula models. An analysis in this context is given, e. g., by Laurant and Gregory (2005), Li (2000), or Schönbucher and Schubert (2001).

In contrast to the previous work, we want to model *simultaneous* credit rating transitions in the continuous-time setting. Therefore it is not sufficient to model the intensities for credit rating changes for the single firms, like in the contagion models, we have to model the intensity of the credit rating changes of the whole economy. In the Markov setup; Avellaneda and Wu (2001) model the credit ratings of the firms with a general Markov process without specifying the dependence structure between the credit ratings. Therefore they estimate the intensity matrix of the credit rating transitions for the whole system of firms. For an economy with many firms this leads to an extremely complex calibration problem. We also model the credit rating process of all firms but we specialize the dependence structure.

This part of the thesis is organized as follows. In Chapter 2 we provide a general framework where we can embed other model classes by defining the parameters such that the possible transitions are restricted. The credit ratings of the firms are specified by a marked point process with random event times and random marks. The random marks specify the possible rating class changes, which take place at the random event times. This modeling framework is versatile, e. g., the event times can be given by a renewal process or the random marks can depend on each other. Especially, the process is not Markovian in general. Since the calibration is too complex, we study a model within the general framework in more detail, which is a Markov jump process and all firms with the same rating are only allowed to change to the same rating class at the same time. In our general model we use a Poisson process with intensity $\lambda > 0$. Each time, at which this process jumps, we choose a map s in $S^S = \{s : S \to S\}$, where S contains the credit rating classes. The probability for choosing the credit rating function $s \in S^S$ is given by the probability distribution $P \in \mathcal{M}_1(S^S)$. Then all firms with rating 1 either remain in this class or change their rating to $s(1)$ with a certain probability, all firms with rating 2 remain in this class or change to $s(2)$ and so on. More precisely, we independently throw a coin for each firm with probability $p_x \in [0, 1]$ of heads, where $x \in S$ is the rating of the firm. If the coin shows head, then the firm changes the rating to $s(x)$, otherwise it remains in the rating class x.

Our general model provides two parameters to introduce dependence. On the one hand, all firms may change simultaneously according to the distribution P on the credit rating classes. For example, we can link the firms such that either the firms are all upgraded or all firms are downgraded. On the other hand, firms within the same rating class are coupled by the vector $p = (p_x)_{x\in S}$ and may change their credit rating simultaneously.

To illustrate the possible dependence structures of the general model we introduce two possible credit rating processes by specifying the parameters of the general model, namely the extended strongly coupled random walk process and the scheme model. In the so-called strongly coupled random walk process, developed by Spitzer (1981) for an infinite state space, there are no simultaneous rating transitions of firms in different credit rating classes, only firms with the same rating may change their rating at the same rating. The scheme model additionally allows that firms with different credit ratings change their rating simultaneously. Chapter 3 illustrates the loss distribution of a simulated credit portfolio where the firms follow these two processes. Stronger dependence between the firms increases the probability of high losses. If we assume that the transition intensities of the individual firms and p is the same in both models, then we observe that there are more joint defaults in the scheme model than in the extended strongly coupled random walk because of the additional dependence through the link between firms in different credit rating classes.

To calibrate our model to historical rating transitions, we consider the maximum likelihood function in Chapter 4. Maximizing this function gives the maximum likelihood estimators, which we calculate for the extended strongly coupled random walk. If the true parameter p_x in the interval $(0, 1)$, then the estimator is unique and given by the root of a polynomial whose degree is bounded by the number of firms and the coefficients contain the number of historical rating changes and the historical waiting times. A detailed derivation for the estimators is given in Section 4.2.3. In Chapter 5 we show consistency and asymptotic normality of the estimators. That means with increasing number of observations the estimator converges to the true parameter. The scaled approximation errors converge in distribution to a normal distribution, which provides confidence intervals for our estimation.

Chapter 2

The credit rating model

We call $F = \{1, \ldots, n\}$ the set of the firms and $S = \{1, \ldots, K\}$ symbolizes the rating classes, where rating K means that the firm is in default and 1 is the best rating class. In the following we model the evolution of the credit ratings of the firms. We assume that the transitions of the firms from one rating class to another are given by a process $(X_t)_{t\geq 0}$ with state space S^n, i. e., X_t is a vector with the rating of the firms at time $t \geq 0$. In the following we define a general framework where all firms may change the rating simultaneously. We model with a marked point process the time of the transitions of the credit ratings and the type of transition, which takes place.

2.1 The general framework

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(I, \mathcal{I})$ be a measurable space. Define the set E of functions by

$$E = \big\{ r : S \times I \to S \big| \; r \text{ is } \big(\mathcal{P}(S) \otimes \mathcal{I}\big)\text{-}\mathcal{P}(S) \text{ measurable} \big\}, \tag{2.1}$$

where $\mathcal{P}(S)$ is the power set of S. Let $\mathcal{E}$ be a σ-algebra on E.

Definition 2.2. *Let $(\tau_i)_{i\in\mathbb{N}}$ be $(0, \infty]$-valued random variables with $\tau_i < \tau_{i+1}$ on the set $\{\tau_i < \infty\}$ and $\tau_i = \tau_{i+1} = \infty$ on $\{\tau_i = \infty\}$. Let ρ_i be a random variable with $\rho_i \in E$ on the set $\{\tau_i < \infty\}$ and $\rho_i := \rho_\infty$ on the set $\{\tau_i = \infty\}$ for each $i \in \mathbb{N}$, where ρ_∞ is an external point of E. Then we call $\big((\tau_i, \rho_i)\big)_{i\in\mathbb{N}}$* a marked point process *(see [41, Chapter 1]).*

The time of the i-th credit rating transition is given by τ_i for $i \in \mathbb{N}$ and the random function ρ_i specifies the possible rating transition. More precisely, ρ_i maps the current rating of the firm and an idiosyncratic component to a rating class, namely the rating of the firm after the transition. The idiosyncratic component is for each $i \in \mathbb{N}$ given by a collection $\{U_i(j) : j \in F\}$ of I-valued random variables. Altogether, the dynamics of the rating transitions are determined by the collection $\{\rho_i, \tau_i, U_i(j) : i \in \mathbb{N}, j \in F\}$ of random variables. In reality we cannot observe these random variables directly. We are only able to watch the rating class transitions of the firms, i. e., we observe the credit rating process $X = (X_t)_{t\geq 0}$ given by the following definition.

Definition 2.3. *Let $\big((\tau_i, \rho_i)\big)_{i\in\mathbb{N}}$ be a marked point process as in Definition 2.2. Let $U_i(j)$ be an I-valued random variable for each $i \in \mathbb{N}$ and $j \in F$. We say that the process $X = (X_t)_{t\geq 0}$ with state space S^n* follows the general framework*, if $X = \big((X_t(1), \ldots, X_t(n))\big)_{t\geq 0}$ satisfies the following:*

(i) $X_t = X_0$ *for* $t \in [0, \tau_1)$,

(ii) For each $i \in \mathbb{N}$ *and firm* $j \in F$

$$X_t(j) = \rho_i\big(X_{\tau_i -}(j), U_i(j)\big) \quad \textit{for } t \in [\tau_i, \tau_{i+1}).$$

In the next lemma we consider the probability of a rating transition from $z \in S^n$ to $\tilde{z} \in S^n$ at the i-th jump time τ_i. Assuming that the collection $\{U_i(j) : j \in F\}$ is identically distributed and independent, then each firm changes the credit rating according to a random stochastic transition matrices $M^{(i)}$ conditional on the σ-algebra $\sigma(\rho_i)$, generated by ρ_i, and the jump time τ_i.

Lemma 2.4. *Additionally to the assumption of Definition 2.3, suppose for each* $i \in \mathbb{N}$ *the collection* $\{U_i(j) : j \in F\}$ *is identically distributed and conditionally independent given* ρ_i*. Define for each* $i \in \mathbb{N}$ *the random stochastic transition matrix* $M^{(i)} = (M^{(i)}_{xy})_{x,y \in S}$ *by*

$$M^{(i)}_{xy} = \mathbb{P}\big[\rho_i(x, U_i(j)) = y \,\big|\, \sigma(\rho_i)\big], \quad \textit{for any } j \in F.$$

Then for each $z, \tilde{z} \in S^n$ *with* $z = \big(z(j)\big)_{j \in F}$ *and* $\tilde{z} = \big(\tilde{z}(j)\big)_{j \in F}$ *it follows that*

$$\mathbb{P}\big[\tilde{z}(j) = \rho_i(z(j), U_i(j)) \textit{ for all } j \in F\big] = \mathbb{E}\Big[\prod_{j \in F} M^{(i)}_{z(j), \tilde{z}(j)}\Big].$$

Proof. Since for each $i \in \mathbb{N}$ the collection $\big(\rho_i(x, U_i(j))\big)_{j \in F}$ of random variables is conditionally independent given $\sigma(\rho_i)$, we obtain

$$\begin{aligned}
\mathbb{P}\big[\tilde{z}(j) = \rho_i(z(j), U_i(j)) \text{ for all } j \in F\big] &= \mathbb{E}\big[\mathbb{P}\big[\tilde{z}(j) = \rho_i(z(j), U_i(j)) \text{ for all } j \in F \,\big|\, \sigma(\rho_i)\big]\big] \\
&= \mathbb{E}\Big[\prod_{j \in F} \mathbb{P}\big[\tilde{z}(j) = \rho_i(z(j), U_i(j)) \,\big|\, \sigma(\rho_i)\big]\Big] = \mathbb{E}\Big[\prod_{j \in F} M^{(i)}_{z(j), \tilde{z}(j)}\Big].
\end{aligned}$$

□

Processes following the general framework are versatile. The jump times can be modelled as the jump times of a renewal process, for example. Then all the transitions occur according to i. i. d. waiting times but they do not have to be exponentially distributed.

2.1.1 Markov jump process in the general framework

A process within the general framework is not Markovian in general. To obtain a Markov process, we define the time points of the transitions by the jump times of a Poisson process, and assume that the random marks are independent for different transitions.

Definition 2.5. *Assume* X_0 *is a* S^n*-valued random variable. Let* $(N_t)_{t \geq 0}$ *be a Poisson process with intensity* $\lambda > 0$*, independent of* X_0*. Define the random time* τ_i *for* $i \in \mathbb{N}$ *by*

$$\tau_i = \inf\big\{t \geq 0 : N_t = i\big\}.$$

Let $(\rho_i)_{i \in \mathbb{N}}$ *be an i. i. d. sequence of* $(E, \mathcal{E})$*-valued random variables, independent of the Poisson process and* X_0*. Let* $\{U_i(j) : i \in \mathbb{N}, j \in F\}$ *be an i. i. d. collection of* I*-valued random variables, and assume the collection is independent of the Poisson process, the random variable* X_0 *and the sequence* $(\rho_i)_{i \in \mathbb{N}}$*. We call the process* $(X_t)_{t \geq 0}$ *a* Markov jump process in the general framework, *if* $X_t = X_0$ *for* $t \in [0, \tau_1)$ *and* $X_t(j) = \rho_i(X_{\tau_i -}(j), U_i(j))$ *for* $j \in F$ *and* $t \in [\tau_i, \tau_{i+1})$.

Remark 2.6. Additionally to the assumptions for a general process in our framework we assume that the type of transitions and the individual components are independent for different jump times, i. e., the random functions ρ_i that determine the type of the transition, and the random collection $\{U_i(j) : j \in F\}$ are independent for different $i \in \mathbb{N}$. Furthermore we suppose that the jump times are given by a Poisson process and the waiting times in between are exponentially distributed.

The above defined credit rating process is indeed a Markov process, which is shown by the following theorem.

Theorem 2.7. *Let $(N_t)_{t\geq 0}$ be a Poisson process with intensity $\lambda > 0$. Let the collection $\{\rho_i, \tau_i, U_i(j) : i \in \mathbb{N}, j \in F\}$ of random variables be defined as in Definition 2.5 and let $(X_t)_{t\geq 0}$ be the appropriate process given by Definition 2.5. Let ρ be an $(E, \mathcal{E})$-valued random variable with the same distribution as the ρ_i. Furthermore assume U is an I-valued random variable, independent of ρ, with the same distribution as $U_i(j)$ for any $i \in \mathbb{N}$ and $j \in F$. Define the random matrix $M = (M_{xy})_{x,y\in S}$ by*

$$M_{xy} = \mathbb{P}\big[\rho(x, U) = y \,\big|\, \sigma(\rho)\big]. \tag{2.8}$$

Then the process $(X_t)_{t\geq 0}$ is a Markov process*. Its transition probability $p_t(\cdot,\cdot) : [0,\infty) \times S^n \times S^n \to [0, 1]$ is given by*

$$p_t(z, \tilde{z}) = \mathbb{E}\big[P_Y^{N_t}(z, \tilde{z})\big] = e^{-\lambda t} \sum_{i=0}^{\infty} \frac{\lambda^i t^i}{i!} P_Y^i(z, \tilde{z}),$$

where P_Y^i is the i-th power of the stochastic transition matrix defined by

$$P_Y(z, \tilde{z}) = \mathbb{E}\Big[\prod_{j\in F} M_{z(j),\tilde{z}(j)}\Big] \quad \textit{for } z = \big(z(j)\big)_{j\in F} \textit{ and } \tilde{z} = \big(\tilde{z}(j)\big)_{j\in F} \textit{ in } S^n.$$

Proof. Define the discrete-time process $(Y_i)_{i\in\mathbb{N}_0}$ with state space S^n by $Y_0 = X_0$ and for $i \geq 1$

$$Y_i(j) = \rho_i\big(Y_{i-1}(j), U_i(j)\big), \quad \text{for each } j \in F.$$

Let $i \in \mathbb{N}$. Since ρ_i as well as $\big\{U_i(j) : j \in F\big\}$ are independent of $\sigma(Y_m : m \leq i-1)$, we obtain for $k \in \{0, i-1\}$

$$\begin{aligned}
&\mathbb{P}[Y_i = z_i \,|\, Y_k = z_k, Y_{k+1} = z_{k+1}, \ldots, Y_{i-1} = z_{i-1}] \\
&\quad = \mathbb{P}\big[\rho_i(Y_{i-1}(j), U_i(j)) = z_i(j) \text{ for all } j \in F \,|\, Y_k = z_k, Y_{k+1} = z_{k+1}, \ldots, Y_{i-1} = z_{i-1}\big] \\
&\quad = \mathbb{P}\big[\rho_i(z_{i-1}(j), U_i(j)) = z_i(j) \text{ for all } j \in F\big] \\
&\quad = \mathbb{E}\Big[\mathbb{P}\big[\rho(z_{i-1}(j), U(j)) = z_i(j) \text{ for all } j \in F\big] \,\Big|\, \sigma(\rho)\Big] \\
&\quad = \mathbb{E}\Big[\prod_{j\in F} M_{z_{i-1}(j),z_i(j)}\Big] = P_Y(z_i, z_{i-1}),
\end{aligned}$$

for all $z_k, \ldots, z_i \in S^n$ with $\mathbb{P}[Y_k = z_k, Y_{k+1} = z_{k+1}, \ldots, Y_{i-1} = z_{i-1}] > 0$. Hence, $(Y_i)_{i\in\mathbb{N}_0}$ is a Markov chain with transition probability matrix P_Y.

Since $X_t = Y_{N_t}$ for all $t \geq 0$, the process $(X_t)_{t\geq 0}$ is a Markov process (see [21, Chapter 4.2, p. 163]). For the transition probability matrix of X let z be in S^n and assume $X_0 = z$. Since $(Y_i)_{i\in\mathbb{N}}$ and $(N_t)_{t\geq 0}$ are independent, we obtain for each $\tilde{z} \in S^n$ and $t \geq 0$

$$\mathbb{P}[X_t = \tilde{z}] = \mathbb{P}\big[Y_{N_t} = \tilde{z}\big] = \sum_{i=0}^{\infty} \mathbb{P}[N_t = i]\, \mathbb{P}[Y_i = \tilde{z}] = \sum_{i=0}^{\infty} \mathbb{P}[N_t = i]\, P_Y^i(z, \tilde{z}),$$

which concludes the proof. □

If we simulate the random collection $\{\rho_i, \tau_i, U_i(j) : i \in \mathbb{N}, j \in F\}$ we can construct the credit rating process X by Definition 2.5. Therefore the construction provides a nice way to simulate the credit rating process. If the state space of a Markov process is finite, we can identify the distribution of the process also by its corresponding Q-matrix. If we are interested in properties of the distribution of the process we work with the Q-matrix instead of the explicit construction of X. Below, we recall the definition of the Q-matrix of a general Markov process with finite state space (see also [37, Chapter 17.3]). Then we deduce the Q-matrix corresponding to the Markov jump process X given by Definition 2.5.

Definition 2.9. *Let $(X_t)_{t\geq 0}$ be a Markov process with finite state space Ψ and transition probability matrix $p_t(z, \tilde{z})$ for $z, \tilde{z} \in \Psi$ and $t \geq 0$. For each $z, \tilde{z} \in \Psi$ with $z \neq \tilde{z}$ define*

$$Q(z, \tilde{z}) = \lim_{t \downarrow 0} \frac{1}{t} p_t(z, \tilde{z}),$$

and assume this limit exists. For each $z \in \Psi$ define

$$Q(z, z) = -\sum_{\substack{\tilde{z} \in \Psi \\ z \neq \tilde{z}}} Q(z, \tilde{z}). \tag{2.10}$$

Then we call Q the Q-matrix *of the Markov process X.*

Theorem 2.11. *Let Ψ be a finite space. Let $Q : \Psi \times \Psi \to \mathbb{R}$ be a function with $Q(z, \tilde{z}) \geq 0$ for all $z, \tilde{z} \in \Psi$ with $z \neq \tilde{z}$ and $Q(z, z)$ as in (2.10). Then Q is the Q-matrix of a Markov process X, where Q and the initial distribution of X uniquely determine the finite dimensional distributions of X.*

Proof. See [37, Theorem 17.25] or [21, Chapter 4, Prop. 1.6 and Section 2]. □

Lemma 2.12. *Assume X_0 is a random variable on S^n. Let the collection $\{\rho, \rho_i, U, U_i(j), \tau_i : i \in \mathbb{N}, j \in \mathbb{F}\}$ of random variables be as in Theorem 2.7. Define the random matrix $M = (M_{xy})_{x,y \in S}$ by (2.8). Define the matrix $Q_n : S^n \times S^n \to \mathbb{R}$ by*

$$Q_n(z, \tilde{z}) = \begin{cases} \lambda\, \mathbb{E}\Big[\prod\limits_{j \in F} M_{z(j), \tilde{z}(j)}\Big], & \text{if } z \neq \tilde{z}, \\ -\sum\limits_{z' \in S^n \setminus \{z\}} Q_n(z, z'), & \text{if } z = \tilde{z}, \end{cases}$$

where $z = \big(z(j)\big)_{j \in F}$ and $\tilde{z} = \big(\tilde{z}(j)\big)_{j \in F}$. Then Q_n is the Q-matrix of the Markov process X with transition probability matrix $p_t(z, \tilde{z})$, defined in Theorem 2.7.

Proof. Let $z, \tilde{z}$ be in S^n with $z \neq \tilde{z}$, then

$$\lim_{t \downarrow 0} \frac{1}{t} p_t(z, \tilde{z}) = \lambda\, P_Y(z, \tilde{z}) = Q_n(z, \tilde{z}).$$

□

In our general framework the dynamics of the firms are not directly depending on each other. The dependence is introduced by the random function ρ and the fact that all firms change the rating simultaneously. Given ρ and the current rating, the transitions of the individual firms are independent. Therefore if we consider a model with $n \in \mathbb{N}$ firms and are interested in the distribution of the rating transitions of the first $m < n$ firms, the distribution of the projected process is the same as if we consider the transitions in a model with m firms. The dynamics do not depend on the number of rated firms. Especially, the distribution of the transitions of a single firm is the same for all firms starting with the same credit rating.

Theorem 2.13. *Fix $m, n \in \mathbb{N}$ with $m < n$. Let Q-matrices Q_m and Q_n be defined as in Lemma 2.12 for the state space S^m and S^n, respectively. Define the projection $\pi : S^n \to S^m$ by $\pi(z) = z|_{S^m}$. Suppose $(X_t)_{t\geq 0}$ is a Markov jump process with state space S^n and Q-matrix Q_n. Then $Y_t = \pi(X_t)$ for $t \geq 0$ is a Markov jump process with state space S^m, generated by Q_m.*

Proof. The process X has the same distribution as the credit rating process defined as in Definition 2.5 and the theorem follows immediately from the construction of that process. □

Remark 2.14. If we do not have the construction of a Markov process generated by the Q-matrix Q_n, resp. Q_m, then we can prove the theorem more generally by using the Q-matrices directly. To this end see Lemma 6.3 in the appendix of this part.

The theorem implies, that the estimated parameters of a larger model have to be the same as if we only consider a smaller number of firms. Furthermore, if we are only interested in the behavior of a smaller number of firms, it is sufficient to simulate the model with less firms.

The general framework includes many different processes. Since the structure is rather general the calibration of this general framework is complex. That is the reason why we restrict our further studies to the case where all firms with the same rating may change only to the same rating class simultaneously. Other simplifications are also possible and postponed for future research.

2.2 The general model

We say that the credit rating process $X = (X_t)_{t\geq 0}$ *follows the general model* (as opposed to 'framework'), if X is a time-homogeneous Markov jump process with state space S^n and if X has the following dynamics. A Poisson process with intensity $\lambda \in (0, \infty)$ is given. Each time, at which this process jumps, we choose a map s in

$$S^S = \{s : S \to S\}.$$

The probability for choosing the credit rating function $s \in S^S$ is given by the probability distribution $P \in \mathcal{M}_1(S^S)$. Then all firms with rating 1 either remain in this class or change their rating to $s(1)$ with a certain probability, all firms with rating 2 remain in this class or change to $s(2)$ and so on. More precisely, we independently throw a coin for each firm with probability $p_x \in [0, 1]$ of heads, where $x \in S$ is the rating of the firm. If the coin shows head, then the firm changes the rating to $s(x)$, otherwise it remains in the rating class x.

This credit rating process is a special process within the general framework. All firms with the same rating may only change to the same rating class. The following definition formalizes this general model.

Definition 2.15. *Let $p = (p_x)_{x\in S}$ be in $[0, 1]^S$ and P be a probability distribution on S^S. Let $(E, \mathcal{E})$ be defined as in (2.1), where $I = [0, 1]$ and $\mathcal{I} = \mathcal{B}\big([0, 1]\big)$. Define for each $s \in S^S$*

the function $r_s \in E$ *by*

$$r_s(x,u) = \begin{cases} s(x), & \text{if } u \in [0,p_x], \\ x, & \text{if } u \in [p_x,1]. \end{cases} \tag{2.16}$$

Let the collection $\{\rho_i, \tau_i, U_i(j) : i \in \mathbb{N}, j \in F\}$ *of random variables be defined as in Definition 2.5, where* τ_i *are the jump times of a Poisson process with intensity* $\lambda > 0$*. Additionally assume the distribution of every* ρ_i *for* $i \in \mathbb{N}$ *is given by* $\mathbb{P}[\rho_i = r_s] = P(s)$ *for each* $s \in S^S$*. Furthermore suppose* $U_i(j)$ *are uniformly distributed on* $[0,1]$ *for each* $i \in \mathbb{N}$ *and* $j \in F$*.*

We say that the Markov jump process $X = (X_t)_{t\geq 0}$ follows the general model with parameters (λ, P, p)*, if* $X_t = X_0$ *for* $t \in [0, \tau_1)$ *and* $X_t(j) = \rho_i(X_{\tau_i-}(j), U_i(j))$ *for* $j \in F$ *and* $t \in [\tau_i, \tau_{i+1})$*.*

Since each transition has the special form $r_s \in E$, at each jump time, the firms with the same rating may either change all to the same rating class, or remain in their original rating class. Whether the j-th firm actually changes the rating at the i-th transition, depends on the idiosyncratic component $U_i(j)$, which represents the coin.

Remark 2.17. The distribution of Markov jump processes in our general framework is uniquely determined by the intensity λ of the Poisson process, the distribution of the random variables ρ_i and $U_i(j)$ and the initial distribution of the process, see Theorem 2.7. In our general model $U_i(j)$ are uniformly distributed and the distribution of ρ_i only depends on the vector $p = (p_x)_{x\in S} \in [0,1]^S$ and the probability distribution $P \in \mathcal{M}_1(S^S)$. Therefore the distribution of the Markov jump process X in the general model is uniquely determined by the parameters (λ, P, p).

Example 2.18. Let $S = \{1,2,3\}$ be the rating classes. Consider the S^n-valued credit rating $(X_t)_{t\geq 0}$ following our general model, given by Definition 2.15 above. Let $p_x = 1$ for all $x \in S$. Then for each $s \in S^S$ the function r_s, defined by (2.16), equals $r_s(x,u) = s(x)$ for all $x \in S$ and $u \in [0,1]$. For $i \in \mathbb{N}$ and $s \in S^S$ consider the set $A = \{\omega \in \Omega : \rho_i(\omega) = r_s\}$. Then $\mathbb{P}(A) = P(s)$ and for all $\omega \in A$ and $j \in F$

$$X_t(j)(\omega) = \rho_i\big(X_{\tau_i-}(j)(\omega), U_i(j)(\omega)\big)(\omega) = s\big(X_{\tau_i-}(j)(\omega)\big), \quad \text{for } t \in \big[\tau_i(\omega), \tau_{i+1}(\omega)\big),$$

i. e., at each jump time i of the process all firms actually change their rating according to some $s \in S^S$, since the coin shows head with probability 1. Define three rating transition functions $s_1, s_2, s_3 \in S^S$ by

- $s_1(x) = 3$ for all $x \in S$,
- $s_2(1) = 2$, $s_2(2) = 3$ and $s_2(3) = 3$,
- $s_3(1) = 1$, $s_3(2) = 1$ and $s_3(3) = 3$.

Define the probability distribution $P \in \mathcal{M}_1(S^S)$ by $P(s_k) > 0$ for $k \in \{1,2,3\}$ and $P(s) = 0$ otherwise. Therefore at the i-th jump of the Poisson process the random function $\rho_i \in \{r_{s_1}, r_{s_2}, r_{s_3}\}$ a. s., i. e., if $\rho_i = r_{s_1}$, then all firms change the rating to 3, if $\rho_i = r_{s_2}$, then all firms with rating 1 are downgraded to 2 and all firms with rating 2 are downgraded to 3, and if $\rho_i = r_{s_3}$, then all firms with rating 2 are upgraded to rating 1. ◊

Q-matrix of the general model

We introduce some notation to carry out the computation of the Q-matrix given by Lemma 2.12 for a process X following the general model.

Definition 2.19. *For each $x \in S$ and $z = \big(z(j)\big)_{j \in F} \in S^n$ denote by $a_x(z)$ the number of firms in z with rating x, i. e., $a_x(z) = \#\{j \in F : z(j) = x\}$. Given another vector $\tilde{z} \in S^n$ the number $b_x(z, \tilde{z})$ of firms, that change their rating $x \in S$, is given by*

$$b_x(z, \tilde{z}) = \#\{j \in F : z(j) = x \text{ and } \tilde{z}(j) \neq x\}, \quad \text{for } x \in S \text{ and } z, \tilde{z} \in S^n.$$

If firms with the same credit rating change to different credit ratings, then the probability for this transition is zero. Such a rating transition is not feasible for the process X in an infinitesimally small time step.

Definition 2.20. *We call a transition from $z \in S^n$ to $\tilde{z} \in S^n$* weakly feasible, *if $z \neq \tilde{z}$ and for all $j, k \in F$ with $j \neq k$, such that $z(j) = z(k)$ and $\tilde{z}(j) \neq z(j) \neq \tilde{z}(k)$, it follows that $\tilde{z}(j) = \tilde{z}(k)$, i. e., for each $x \in S$ all firms with rating x may only change to the same rating class.*

Definition 2.21. *For each weakly feasible transition from $z \in S^n$ to $\tilde{z} \in S^n$ denote the set of* possible rating class changes *by*

$$\begin{aligned}\mathcal{S}^{\mathrm{c}}(z, \tilde{z}) := \big\{ s \in S^S :\ s(x) = \tilde{z}(j) \ &\text{for each } x \in S \text{ s. t. there exists } j \in F \\ &\text{with } z(j) = x \text{ and } \tilde{z}(j) \neq x \big\}.\end{aligned}$$

For each $s \in \mathcal{S}^{\mathrm{c}}(z, \tilde{z})$ the entry $s(x)$ is determined, if there is a transition of a firm with rating $x \in S$. Using the notation and the convention $0^0 = 1$ we state the Q-matrix of the process X in the next theorem.

Theorem 2.22. *Let $(p_x)_{x \in S}$ be in $[0,1]^S$ and P be a probability distribution on S^S. Let the collection $\{\rho_i, \tau_i, U_i(j) : i \in \mathbb{N}, j \in F\}$ of random variables be defined as in Definition 2.15, where τ_i are the jump times of a Poisson process with intensity $\lambda > 0$. Define the matrix $Q_n^{\mathrm{g}} : S^n \times S^n \to \mathbb{R}$ by*

$$Q_n^{\mathrm{g}}(z, \tilde{z}) = \begin{cases} \lambda \sum\limits_{s \in \mathcal{S}^{\mathrm{c}}(z,\tilde{z})} P(s) \prod\limits_{\substack{x \in S \\ s(x) \neq x}} p_x^{b_x(z,\tilde{z})} (1 - p_x)^{a_x(z) - b_x(z,\tilde{z})}, & \\ \qquad\qquad \text{if the transition } z \to \tilde{z} \text{ is weakly feasible,} & \\ -\lambda \sum\limits_{\substack{s \in S^S \\ s \neq \mathrm{id}}} P(s) \Bigg(1 - \prod\limits_{\substack{x \in S \\ s(x) \neq x}} (1 - p_x)^{a_x(z)} \Bigg), & \\ \qquad\qquad \text{if } z = \tilde{z}, & \\ 0, \qquad\quad \text{otherwise,} & \end{cases}$$

where $\mathrm{id} : S \to S$ is the identity, i. e. $\mathrm{id}(x) = x$ for all $x \in S$.

Then Q_n^{g} is the Q-matrix of the time-homogeneous Markov jump process $X = (X_t)_{t \geq 0}$ following the general model with parameters (λ, P, p), given by Definition 2.15.

Proof. Let ρ be a random E-valued function with the same distribution as ρ_i for any $i \in \mathbb{N}$. Let U be a random variable uniformly distributed on $[0,1]$ independent of ρ. If X follows the general model, X especially follows the general framework. Therefore the Q-matrix Q_n

of X is given by Lemma 2.12, i. e. for each $z, \tilde{z} \in S^n$ with $z \neq \tilde{z}$[1]

$$
\begin{aligned}
Q_n(z,\tilde{z}) &= \lambda\,\mathbb{E}\Big[\prod_{j\in F}\mathbb{P}\big[\rho(z(j),U)=\tilde{z}(j)\,\big|\,\sigma(\rho)\big]\Big] = \lambda\sum_{s\in S^S}P(s)\prod_{j\in F}\mathbb{P}\big[r_s(z(j),U)=\tilde{z}(j)\big] \\
&= \lambda\sum_{s\in S^S}P(s)\prod_{j\in F}\Big(\mathbb{P}\Big[s(z(j))=\tilde{z}(j),U\le p_{z(j)}\Big]+\mathbb{P}\Big[z(j)=\tilde{z}(j),U>p_{z(j)}\Big]\Big), \\
&= \lambda\sum_{s\in S^S}P(s)\prod_{j\in F}\Big(p_{z(j)}\,\mathbb{1}_{\{s(z(j))=\tilde{z}(j)\}}+(1-p_{z(j)})\,\mathbb{1}_{\{z(j)=\tilde{z}(j)\}}\Big).
\end{aligned}
\tag{2.23}
$$

It is sufficient to prove that $Q_n(z,\tilde{z}) = Q_n^{\mathrm{g}}(z,\tilde{z})$ for all $z, \tilde{z} \in S^n$.

If the transition from z to $\tilde{z}$ is not weakly feasible, then there exists $i, j \in F$ such that $z(i) = z(j)$ but $\tilde{z}(i) \neq \tilde{z}(j)$ and $\tilde{z}(i), \tilde{z}(j) \neq z(i)$. It follows

$$
\mathbb{1}_{\{s(z(i))=\tilde{z}(i)\}}\,\mathbb{1}_{\{s(z(j))=\tilde{z}(j)\}} = 0 \quad \text{for all } s \in S^S.
$$

By (2.23) we obtain $Q_n(z,\tilde{z}) = 0$. So, if the transition is not weakly feasible, then $Q_n(z,\tilde{z}) = Q_n^{\mathrm{g}}(z,\tilde{z})$.

Assume the transition from z to $\tilde{z}$ is weakly feasible. If the rating function $s \in S^S$ is not in the set $\mathcal{S}^{\mathrm{c}}(z,\tilde{z})$ of possible rating class changes, then there exists $j \in F$ such that $z(j) \neq \tilde{z}(j)$ and $s(z(j)) \neq \tilde{z}(j)$. It follows

$$
p_{z(j)}\,\mathbb{1}_{\{s(z(j))=\tilde{z}(j)\}}+(1-p_{z(j)})\,\mathbb{1}_{\{z(j)=\tilde{z}(j)\}} = 0.
$$

Therefore (2.23) simplifies to

$$
Q_n(z,\tilde{z}) = \lambda\sum_{s\in\mathcal{S}^{\mathrm{c}}(z,\tilde{z})}P(s)\prod_{\substack{x\in S\\ s(x)\neq x}}p_x^{b_x(z,\tilde{z})}(1-p_x)^{a_x(z)-b_x(z,\tilde{z})} = Q_n^{\mathrm{g}}(z,\tilde{z}).
$$

The diagonal entries of the Q-matrix Q_n are the negative sum over the non-diagonal entries of the same row, i. e. by (2.23)

$$
\begin{aligned}
Q_n(z,z) &= -\sum_{\substack{\tilde{z}\in S^n\\ z\neq\tilde{z}}}Q_n(z,\tilde{z}) \\
&= -\lambda\sum_{s\in S^S}P(s)\sum_{\substack{\tilde{z}\in S^n\\ z\neq\tilde{z}}}\prod_{j\in F}\Big(p_{z(j)}\,\mathbb{1}_{\{s(z(j))=\tilde{z}(j)\}}+(1-p_{z(j)})\,\mathbb{1}_{\{z(j)=\tilde{z}(j)\}}\Big).
\end{aligned}
\tag{2.24}
$$

For simplification of the last sum consider for all $s \in S^S$

$$
\begin{aligned}
&\sum_{\tilde{z}\in S^n}\prod_{j\in F}\Big(p_{z(j)}\,\mathbb{1}_{\{s(z(j))=\tilde{z}(j)\}}+(1-p_{z(j)})\,\mathbb{1}_{\{z(j)=\tilde{z}(j)\}}\Big) \\
&\qquad = \prod_{j\in F}\sum_{\tilde{z}(j)\in S}\Big(p_{z(j)}\,\mathbb{1}_{\{s(z(j))=\tilde{z}(j)\}}+(1-p_{z(j)})\,\mathbb{1}_{\{z(j)=\tilde{z}(j)\}}\Big) = 1.
\end{aligned}
\tag{2.25}
$$

[1]The function $\mathbb{1}_{\{\text{condition}\}} = 1$, if the condition is satisfied, and zero otherwise.

Putting (2.25) into (2.24) we obtain

$$\begin{aligned}
Q_n(z,z) &= -\lambda \sum_{s\in S^S} P(s)\Big(1 - \prod_{j\in F}\big(p_{z(j)}\,\mathbb{1}_{\{s(z(j))=z(j)\}} + (1-p_{z(j)})\,\mathbb{1}_{\{z(j)=z(j)\}}\big)\Big),\\
&= -\lambda \sum_{s\in S^S} P(s)\Big(1 - \prod_{j\in F}\big(1 - p_{z(j)}\,\mathbb{1}_{\{s(z(j))\neq z(j)\}}\big)\Big)\\
&= -\lambda \sum_{\substack{s\in S^S\\ s\neq \mathrm{id}}} P(s)\Big(1 - \prod_{\substack{x\in S\\ s(x)\neq x}} (1-p_x)^{a_x(z)}\Big) = Q_n^{\mathrm{g}}(z,z).
\end{aligned}$$

Altogether $Q_n^{\mathrm{g}} = Q_n$ and Q_n^{g} is the Q-matrix of X. □

The Q-matrix Q_n^{g} of the process X depends only on the parameters $(\lambda, P, p) \in (0,\infty)\times\mathcal{M}_1(S^S)\times[0,1]^S$ and therefore the distribution of the process X depends on these parameters and the initial distribution, see also Remark 2.17. For different parameters we obtain in general different types of dependence. The parameter p is the probability that a firm actually changes the rating class. The distribution P determines to which rating classes the firms may change at the transitions. In the following we specify two possible distributions P. First, only firms with the same credit rating may change at the same time, which is the so-called strongly coupled random walk process. In this model there are no simultaneous rating transitions of firms with different credit ratings. The dependence in this model is induced only by the vector p between firms with the same rating. Secondly, firms with different ratings may change according to a specialized transition function. Here we have higher dependence as in the strongly coupled random walk since there is additionally dependence between the credit rating classes.

2.3 Extended strongly coupled random walk process

We embed the strongly coupled random walk process in the general model by specifying the parameters appropriately. The strongly coupled random walk process was introduced by Spitzer (1981) and has the following dynamics. For each credit rating class $x\in S$ we have an independent Poisson process with intensity $\lambda_x\in(0,\infty)$. When the Poisson process jumps, then the firms may change to the rating class $y\in S$ according to the stochastic transition function $P^{\mathrm{c}} : S\times S\to[0,1]$, i. e. $\sum_{y\in S} P^{\mathrm{c}}(x,y) = 1$ for all $x\in S$. The firms actually change the rating independently of one another with probability $p_x\in[0,1]$.

The strongly coupled random walk allows simultaneous transitions only from exactly one rating class to the same other rating class. The dynamics of the strongly coupled random walk within our general model are given by the following definition.

Definition 2.26. *Let $\lambda_x\in(0,\infty)$ for each $x\in S$ and let $P^{\mathrm{c}} : S\times S\to[0,1]$ be a stochastic transition function. Let $p=(p_x)_{x\in S}$ be a probability vector in $[0,1]^S$. Define $\lambda=\sum_{x\in S}\lambda_x$. Define the rating class distribution P on S^S by*

$$P(s) = \begin{cases} \frac{\lambda_x}{\lambda} P^{\mathrm{c}}(x,y), & \text{if there exist } x,y\in S \text{ with } x\neq y \text{ and} \\ & \quad s(x)=y \text{ and } s(u)=u \text{ for all } u\in S\setminus\{x\},\\ \sum\limits_{x\in S}\frac{\lambda_x}{\lambda}P^{\mathrm{c}}(x,x), & \text{if } s(x)=x \text{ for all } x\in S,\\ 0, & \text{otherwise.}\end{cases} \tag{2.27}$$

We say that the time-homogeneous Markov jump process X is a strongly coupled random walk process *with parameters* $\big((\lambda_x)_{x\in S}, P^{\mathrm{c}}, p\big)$*, if X follows the general model with parameters (λ, P, p), given by Definition 2.15.*

Remark 2.28. The above defined P is obviously a distribution, since

$$\sum_{s\in S^S} P(s) = \sum_{\substack{x,y\in S\\ x\neq y}} \frac{\lambda_x}{\lambda} P^{\mathrm{c}}(x,y) + \sum_{x\in S} \frac{\lambda_x}{\lambda} P^{\mathrm{c}}(x,x) = \sum_{x\in S} \frac{\lambda_x}{\lambda}\big(1 - P^{\mathrm{c}}(x,x) + P^{\mathrm{c}}(x,x)\big) = 1.$$

In the general model the firms with the same rating may only change *to one* other rating class or remain in the original class. For the strongly coupled random walk, we have additionally that changes are only possible *from one* rating class at the same time, since P is only non-zero, if $s(x) \neq x$ for exactly one $x \in S$. The probability that the firms actually change is still given by p. So, in the strongly coupled random walk less transitions are feasible and we now strengthen the definition of feasible.

Definition 2.29. *We call a transition from $z \in S^n$ to $\tilde{z} \in S^n$* strongly feasible*, if $z \neq \tilde{z}$ and there exists exactly one pair $(x, y) \in S^2$ with $x \neq y$ such that for all $j \in F$ with $z(j) \neq \tilde{z}(j)$ it follows that $z(j) = x$ and $\tilde{z}(j) = y$, i. e., the firms may only change from x to y. Obviously, strongly feasible implies weakly feasible.*

For estimation we extend the strongly coupled random walk and reduce the number of parameters. Instead of $(\lambda_x)_{x\in S}$ and P^{c} we use the Q-matrix Q_1^{g} of the strongly coupled random walk with one firm, which is defined by the following corollary of Theorem 2.22.

Corollary 2.30. *Let $\lambda_x \in (0,\infty)$ for each $x \in S$ and let $P^{\mathrm{c}} : S \times S \to [0,1]$ be a stochastic transition function. Let $p = (p_x)_{x\in S}$ be a probability vector in $[0,1]^S$. Then the Q-matrix $Q_1^{\mathrm{g}} : S \times S \to \mathbb{R}$ of the strongly coupled random walk with state space S and parameters $\big((\lambda_x)_{x\in S}, P^{\mathrm{c}}, p\big)$ is given by*

$$Q_1^{\mathrm{g}}(x,y) = \lambda_x P^{\mathrm{c}}(x,y) p_x, \quad \textit{for } x, y \in S \textit{ with } x \neq y. \tag{2.31}$$

Proof. Let X be a strongly coupled random walk with state space S and parameters $\{(\lambda_x)_{x\in S}, P^{\mathrm{c}}, p\}$. Then X follows the general model with parameters (λ, P, p), given by Definition 2.26, and the Q-matrix Q_1^{g} of the credit rating process X is given by Theorem 2.22. Using (2.27) we obtain for each $x, y \in S$ with $x \neq y$

$$Q_1^{\mathrm{g}}(x,y) = \lambda \sum_{s\in \mathcal{S}^{\mathrm{c}}(x,y)} P(s) p_x = \lambda_x P^{\mathrm{c}}(x,y) p_x,$$

since for each $x, y \in S$ with $x \neq y$ the set $\mathcal{S}^{\mathrm{c}}(x,y)$ of possible rating class changes as in Definition 2.21 equals $\mathcal{S}^{\mathrm{c}}(x,y) = \{s : S \to S : s(x) = y\}$. □

Remark 2.32. The embedding property in Theorem 2.13 implies that the intensity for the transitions considering a single firm in a credit rating process with n firms, i. e. with state space S^n, is the same as the intensities of the transitions in a credit rating process with just one firm, i. e. with state space S, if we use the same parameters for both processes. Therefore Q_1^{g} corresponds to the intensity of the transitions of each individual firm in the strongly coupled random walk process with n firms.

Depending on the Q-matrix Q_1^{g} we state the Q-matrix of the strongly coupled random walk process.

Lemma 2.33. *Let $\mu : S \times S \to \mathbb{R}$ be a Q-matrix and let $p = (p_u)_{u\in S}$ be a probability vector in $[0,1]^S$. Let $z, \tilde{z}$ be in S^n. For each rating class $u \in S$ define $a_u(z)$ and $b_u(z,\tilde{z})$ by Definition 2.19. Let $\mathcal{S}^{\mathrm{c}}(z,\tilde{z})$ be as in Definition 2.21. For each strongly feasible transition from $z \in S^n$ to $\tilde{z} \in S^n$ define $x, y \in S$ with $x \neq y$ such that $z(j) = x$ and $\tilde{z}(j) = y$ for some $j \in F$. Define the matrix $Q_n^{\mathrm{c}} : S^n \times S^n \to \mathbb{R}$ by*

$$Q_n^{\mathrm{c}}(z,\tilde{z}) = \begin{cases} \mu_{xy} p_x^{b_x(z,\tilde{z})-1}(1-p_x)^{a_x(z)-b_x(z,\tilde{z})}, & \text{if } z \to \tilde{z} \text{ is strongly feasible,} \\ -\sum\limits_{u\in S} \mu_u\, q\big(a_u(z), p_u\big), & \text{if } z = \tilde{z}, \\ 0, & \text{otherwise,} \end{cases} \tag{2.34}$$

where the function $q : \mathbb{N}_0 \times [0,1] \to [0,\infty)$ is given by

$$q(k,p) = \sum_{j=0}^{k-1} (1-p)^j = \begin{cases} \frac{1-(1-p)^k}{p}, & \text{if } p \in (0,1], \\ k, & \text{if } p = 0, \end{cases} \tag{2.35}$$

and for all $u \in S$

$$\mu_u = \sum_{v \in S\setminus\{u\}} \mu_{uv}. \tag{2.36}$$

Then the following holds:

(i) The matrix Q_n^{c} is a Q-matrix.

(ii) Let $\lambda_u \in (0,\infty)$ for each $u \in S$ and let $P^{\mathrm{c}} : S \times S \to [0,1]$ be a stochastic transition function. Assume $\mu = (\mu_{uv})_{u,v\in S}$ is the Q-matrix Q_1^{g} of the strongly coupled random walk with state space S and parameters $\big((\lambda_u)_{u\in S}, P^{\mathrm{c}}, p\big)$, given by Corollary 2.30. Then Q_n^{c} is the Q-matrix of the strongly coupled random walk process with state space S^n and parameters $\big((\lambda_u)_{u\in S}, P^{\mathrm{c}}, p\big)$.

Proof. (i) Obviously the non-diagonal entries of Q_n^{c} are non-negative. We parametrize the strongly feasible transitions by $x, y \in S$ with $x \neq y$ and $b \in \{1, \ldots, a_x(z)\}$, where b firms with rating x change their rating to y in the transition. Using this parametrization the sum over the non-diagonal entries of Q_n^{c} in row $z \in S^n$ equals

$$\sum_{\tilde{z}\in S^n\setminus\{z\}} Q_n^{\mathrm{c}}(z,\tilde{z}) = \sum_{\substack{\tilde{z}\in S^n \\ z\to\tilde{z} \\ \text{str. feas.}}} Q_n^{\mathrm{c}}(z,\tilde{z}) = \sum_{\substack{x,y\in S \\ x\neq y}} \sum_{b=1}^{a_x(z)} \binom{a_x(z)}{b} \mu_{xy} p_x^{b-1}(1-p_x)^{a_x(z)-b} \tag{2.37}$$

$$= \sum_{x\in S} \Big(\sum_{\substack{y\in S \\ x\neq y}} \mu_{xy} \Big) \sum_{b=1}^{a_x(z)} \binom{a_x(z)}{b} p_x^{b-1}(1-p_x)^{a_x(z)-b}. \tag{2.38}$$

Since the binomial theorem says for $p \in [0,1]$ and $k \in \mathbb{N}_0$

$$\sum_{j=0}^{k} \binom{k}{j} p^j (1-p)^{k-j} = \big(p + (1-p)\big)^k = 1, \tag{2.39}$$

Equation (2.38) simplifies to

$$\begin{aligned} \sum_{\tilde{z}\in S^n\setminus\{z\}} Q_n^{\mathrm{c}}(z,\tilde{z}) &= \sum_{x\in S} \mu_x \left(\frac{1-(1-p_x)^{a_x(z)}}{p_x} \mathbb{1}_{\{p_x\in(0,1]\}} + a_x(z)\, \mathbb{1}_{\{p_x=0\}} \right) \\ &= \sum_{x\in S} \mu_x\, q\big(a_x(z), p_x\big) = Q_n^{\mathrm{c}}(z,\tilde{z}). \end{aligned}$$

Hereby the negative sum over the non-diagonal entries in row z is the corresponding diagonal entry of Q_n^{c} and it is therefore a Q-matrix by Theorem 2.11, since the state space is finite.

(ii) Let X be a strongly coupled random walk with state space S^n and parameters $\big((\lambda_x)_{x\in S}, P^{\mathrm{c}}, p\big)$. Then X follows the general model with parameters (λ, P, p), given by Definition 2.26, and the Q-matrix Q_n^{g} of X is given by Theorem 2.22. To prove that Q_n^{c} is the Q-matrix of X, we show $Q_n^{\mathrm{c}}(z,\tilde z) = Q_n^{\mathrm{g}}(z,\tilde z)$ for all $z, \tilde z \in S^n$ with $z \neq \tilde z$.

Let $z = (z(j))_{j\in F}, \tilde z = (\tilde z(j))_{j\in F} \in S^n$ with $z \neq \tilde z$. Assume the transition from z to $\tilde z$ is not strongly feasible. Then at least two firms with different credit rating change their rating, i. e., there exist $j, k \in F$ with $z(j) \neq z(k)$, $z(j) \neq \tilde z(j)$ and $z(k) \neq \tilde z(k)$. Then $s(z(j)) \neq z(j)$ and $s(z(k)) \neq z(k)$ for each $s \in \mathcal{S}^{\mathrm{c}}(z,\tilde z)$ and $P(s) = 0$ by (2.27). Therefore $Q_n^{\mathrm{g}}(z,\tilde z) = 0$ and $Q_n^{\mathrm{c}}(z,\tilde z) = Q_n^{\mathrm{g}}(z,\tilde z)$ for $z, \tilde z \in S^n$ not strongly feasible.

Assume z to $\tilde z$ is a strongly feasible transition. Then there exist $x, y \in S$ with $x \neq y$ such that for all $j \in F$ with $z(j) \neq \tilde z(j)$, it follows $z(j) = x$ and $\tilde z(j) = y$. Therefore $s(x) = y$ for all $s \in \mathcal{S}^{\mathrm{c}}(z,\tilde z)$. Define $\tilde s : S \to S$ by $\tilde s(x) = y$ and $\tilde s(u) = u$ for all $u \neq x$. Using the definition of the probability distribution P in (2.27) for the strongly coupled random walk,

$$P(\tilde s) = \frac{\lambda_x P^{\mathrm{c}}(x,y)}{\lambda}, \quad \text{and } P(s) = 0 \text{ for all } s \in \mathcal{S}^{\mathrm{c}}(z,\tilde z) \setminus \{\tilde s\}.$$

Since μ is the Q-matrix of the strongly coupled random walk with state space S and parameters $\big((\lambda_x)_{x\in S}, P^{\mathrm{c}}, p\big)$, Corollary 2.30 implies

$$\mu_{xy} = \lambda_x\, P^{\mathrm{c}}(x,y) p_x, \quad \text{for all } x, y \in S \text{ with } x \neq y.$$

Altogether we obtain

$$\begin{aligned} Q_n^{\mathrm{g}}(z,\tilde z) &= \lambda \sum_{s\in\mathcal{S}^{\mathrm{c}}(z,\tilde z)} P(s) \prod_{\substack{x'\in S\\ s(x')\neq s}} p_{x'}^{b_{x'}(z,\tilde z)} (1-p_{x'})^{a_{x'}(z)-b_{x'}(z,\tilde z)} \\ &= \lambda\, P(\tilde s) p_x^{b_x(z,\tilde z)} (1-p_x)^{a_x(z)-b_x(z,\tilde z)} \\ &= \mu_{xy}\, p_x^{b_x(z,\tilde z)-1} (1-p_x)^{a_x(z)-b_x(z,\tilde z)} = Q_n^{\mathrm{c}}(z,\tilde z), \end{aligned}$$

which concludes the proof. □

For all parameters $\lambda_x \in (0,\infty)$ for $x \in S$, all stochastic transition functions $P^{\mathrm{c}} : S \times S \to [0,1]$ and all $p = (p_x)_{x\in S} \in [0,1]^S$, the strongly coupled random walk has the Q-matrix given by (2.34) for $\mu : S \times S \to \mathbb{R}$ given by (2.31). Therefore all strongly coupled random walks are in the class of Markov processes, where the Q-matrix is defined by (2.34) for parameters μ and p. On the other hand, not all Markov processes within this class are strongly coupled random walks. Consider $p = (p_u)_{u\in S}$ with $p_x = 0$ for any $x \in S$ and the Q-matrix $\mu = (\mu_{uv})_{u,v\in S}$ with $\mu_{xy} > 0$ for any $y \in S$ with $x \neq y$. Then there is no $\lambda_x \in (0,\infty)$ and $P^{\mathrm{c}}(x,y) \in [0,1]$ such that $\mu_{xy} = \lambda_x\, P^{\mathrm{c}}(x,y) p_x$. Therefore the Markov process with Q-matrix Q_n^{c} with parameters μ and p is not a strongly coupled random walk.

In this case we obtain $Q_n^{\mathrm{c}}(z,\tilde z) = \mu_{xy}$ for $p_x = 0$, if there is only one single firm changing the credit rating class, i. e. $b_x(z,\tilde z) = 1$. All other non-diagonal entries are zero. This represents a Markov jump process where the firms with rating x change their credit rating independently of the other firms according to Q-matrix μ. Therefore the broader class of Markov processes with Q-matrix Q_n^{c} includes independent transitions of the firms for $p_x = 0$ and with $p_x > 0$ we obtain dependence since the firms can change their rating simultaneously. In the following we consider strongly coupled walks extended by the independent case.

Definition 2.40. *Let $\mu = (\mu_{xy})_{xy \in S}$ be a Q-matrix and $p = (p_x)_{x \in S}$ be a vector in $[0,1]^S$. We say the Markov process $X = (X_t)_{t \geq 0}$ is an* extended strongly coupled random walk process (esc–process) *with state space S^n and parameters (μ, p), if the Q-matrix of X is given by $Q_n^{\mathrm{c}} : S^n \times S^n \to \mathbb{R}$ depending on μ and p, given by* (2.34).

For each strongly coupled random walk the embedding property in Theorem 2.13 holds, since the process follows the general framework. Since the class of extended strongly coupled random walks is larger, we show that the embedding property still holds. Then the parameter μ corresponds to the transition intensity of a single firm in the process, i. e., if X is an esc–process with parameter (μ, p) and state space S^n, the embedding property implies that the credit rating $(X_t(j))_{t \geq 0}$ of the j-th firm is a Markov jump process with Q-matrix Q_1^{c}, where $Q_1^{\mathrm{c}} = \mu$ by definition.

Lemma 2.41. *Let $\mu = (\mu_{xy})_{xy \in S}$ be a Q-matrix and $p = (p_x)_{x \in S}$ be a vector in $[0,1]^S$. Fix $m, n \in \mathbb{N}$ with $m < n$. Define the Q-matrices Q_m^{c} and Q_n^{c} by* (2.34) *for the state space S^m and S^n, respectively. Define the projection $\pi : S^n \to S^m$ by $\pi(z) = z|_{S^m}$. Suppose $(X_t)_{t \geq 0}$ is a Markov jump process with state space S^n and Q-matrix Q_n^{c}. Then $Y_t = \pi(X_t)$ for $t \geq 0$ is a Markov jump process with state space S^m, generated by Q_m^{c}.*

Proof. By Lemma 6.3 in the appendix it is sufficient to show for all $z_m, \tilde{z}_m \in S^m$

$$Q_m^{\mathrm{c}}(z_m, \tilde{z}_m) = \sum_{\tilde{z}_n \in \pi^{-1}(\tilde{z}_m)} Q_n^{\mathrm{c}}(z_n, \tilde{z}_n), \qquad \text{for } z_n \in \pi^{-1}(z_m), \tag{2.42}$$

since the projection π is surjective.

Assume $z_m = \big(z_m(j)\big)_{j \in F} \in S^m$ and $\tilde{z}_m = \big(\tilde{z}_m(j)\big)_{j \in F} \in S^m$ with $z_m \neq \tilde{z}_m$. If the transition from z_m to $\tilde{z}_m$ is not strongly feasible, in the sense of Definition 2.29, then there exist $i, j \in \{1, \ldots, m\}$ such that $z_m(i) \neq \tilde{z}_m(i)$ and $z_m(j) \neq \tilde{z}_m(j)$ but $z_m(i) \neq z_m(j)$ or $\tilde{z}_m(i) \neq \tilde{z}_m(j)$. Therefore for each $z_n \in \pi^{-1}(z_m)$ and $\tilde{z}_n \in \pi^{-1}(\tilde{z}_m)$ the transition from z_n to $\tilde{z}_n$ is not strongly feasible as well. By definition of Q_m^{c} and Q_n^{c} we obtain $Q_m^{\mathrm{c}}(z_m, \tilde{z}_m) = 0 = Q_n^{\mathrm{c}}(z_n, \tilde{z}_n)$ and (2.42) holds. Assume that the transition from z_m to $\tilde{z}_m$ is strongly feasible, i. e., there exists a pair $x, y \in S$ with $x \neq y$ such that for all $j \in \{1, \ldots, m\}$ with $z_m(j) \neq \tilde{z}_m(j)$ follows that $z_m(j) = x$ and $\tilde{z}_m(j) = y$. Define the number of firms with rating x in z_m by

$$a_m = \#\{j \in \{1, \ldots, m\} : z_m(j) = x\},$$

and the number of firms that change the rating x to y in the transition from z_m to $\tilde{z}_m$ by

$$b_m = \#\{j \in \{1, \ldots, m\} : \, z_m(j) = x, \, \tilde{z}_m(j) = y\}.$$

Let $z_n \in \pi^{-1}(z_m)$. Define the number a_n of firms with rating x in z_n by

$$a_n = \#\{j \in \{1, \ldots, n\} : z_n(j) = x\}.$$

We parametrize the strongly feasible transitions from z_n to $\tilde{z}_n \in \pi^{-1}(\tilde{z}_m)$ by the number $b \in \{0, \ldots, a_n - a_m\}$ of firms in $j \in \{m+1, \ldots, n\}$ that change the rating from x to y. By definition of Q_n^{c} we obtain

$$\begin{aligned}
\sum_{\tilde{z}_n \in \pi^{-1}(\tilde{z}_m)} Q_n^{\mathrm{c}}(z_n, \tilde{z}_n) &= \sum_{b=0}^{a_n - a_m} \binom{a_n - a_m}{b} \mu_{xy} \, p_x^{(b_m + b) - 1} (1 - p_x)^{a_n - (b_m + b)} \\
&= \mu_{xy} \, p_x^{b_m - 1} (1 - p_x)^{a_m - b_m} \sum_{b=0}^{a_n - a_m} \binom{a_n - a_m}{b} p_x^b (1 - p_x)^{(a_n - a_m) - b} \\
&= \mu_{xy} \, p_x^{b_m - 1} (1 - p_x)^{a_m - b_m} = Q_m^{\mathrm{c}}(z_m, \tilde{z}_m),
\end{aligned}$$

using the binomial theorem in (2.39). Therefore (2.42) holds for all $z_m, \tilde{z}_m \in S^m$ with $z_m \neq \tilde{z}_m$.

Since Q^{c}_m and Q^{c}_n are Q-matrices and (2.42) holds for all $z_m, \tilde{z}_m \in S^m$ with $z_m \neq \tilde{z}_m$, we obtain for $z_m \in S^m$ and $z_n \in \pi^{-1}(z_m)$

$$\begin{aligned} Q^{\mathrm{c}}_m(z_m, z_m) &= -\sum_{\substack{\tilde{z}_m \in S^m \\ z_m \neq \tilde{z}_m}} Q^{\mathrm{c}}_m(z_m, \tilde{z}_m) \\ &= -\sum_{\substack{\tilde{z}_m \in S^m \\ z_m \neq \tilde{z}_m}} \sum_{\tilde{z}_n \in \pi^{-1}(\tilde{z}_m)} Q^{\mathrm{c}}_n(z_n, \tilde{z}_n) = \sum_{\tilde{z}_n \in \pi^{-1}(z_m)} Q^{\mathrm{c}}_n(z_n, \tilde{z}_n). \end{aligned}$$

□

The esc–process incorporates dependencies between firms with the same credit rating. This process shows the influence of the vector p. Since the transitions of a single firm in the model follows the Q-matrix $Q^{\mathrm{c}}_1 = \mu$ by the embedding property, we are able to compare this model with a model, where all firms move independently of one another according to the Q-matrix μ. The following example illustrates how the vector p couples firms with the same rating and introduces dependence. Therefore with the esc–process we are capable to reproduce different kinds of dependencies.

Example 2.43. Consider the rating classes $S = \{1, 2, 3\}$ where 3 means default and is an absorbing state. Assume there are $n = 2$ firms and the single firms change their credit rating according to Q-matrix

$$\mu = \begin{pmatrix} -0.5 & 0.25 & 0.25 \\ 0.5 & -1 & 0.5 \\ 0 & 0 & 0 \end{pmatrix}. \tag{2.44}$$

In the model where both firms change their rating $(X^{(i)}_t)_{t \geq 0}$ for $i = 1, 2$ independently of each other, the probability that both firms default within one year is

$$\mathbb{P}_{(1,2)}\big[X^{(1)}_1 = 3, X^{(2)}_1 = 3\big] = \exp(\mu)_{1,3} \exp(\mu)_{2,3} \approx 0.086,$$

where $X^{(1)}_0 = 1$ and $X^{(2)}_0 = 2$ almost surely.

Let $(X^c_t)_{t \geq 0}$ be the credit rating process of two firms following the esc–process with parameters μ and $p_x = 1$ for all $x \in S$. Then the probability that both firms default within one year is

$$\mathbb{P}_{(1,2)}[X^c_1 = (3, 3)] = \exp(Q^{\mathrm{c}}_2)_{(1,2),(3,3)} \approx 0.141,$$

where $X^c_0 = (1, 2)$ almost surely. If we couple the firms with the same rating then the probability increases that both firms default compared to the independent case. Once both firms have the same credit rating, they change their rating always in the same way at the same time, since $p_x = 1$ for all $x \in S$. This increases the probability that both firms default.

The probability of joint defaults also influences the loss in a portfolio of two credits where $(X^c_t)_{t \geq 0}$ is the credit rating process of the obligors with $X^c_0 = (1, 2)$ a. s. Suppose the maturity of the credits is $T \geq 0$ and the credit amount is 1. Assuming that the recovery rate and the default free interest rate are zero, the loss at time $t \in (0, T]$, induced by the credit of firm $j \in \{1,2\}$, is given by $L_t^{(j)} = \mathbb{1}_{\{X_t^c(j)=K\}}$. The covariance of the losses after one year is

$$\begin{aligned} \operatorname{Cov}\big(L_1^{(1)}, L_1^{(2)}\big) &= \mathbb{E}[L_1^{(1)} L_1^{(2)}] - \mathbb{E}[L_1^{(1)}]\mathbb{E}[L_1^{(2)}] \\ &= \mathbb{P}_{(1,2)}\big[X_1^c = (3,3)\big] - \exp(\tilde{\mu})_{1,3} \exp(\tilde{\mu})_{2,3} \approx 0.055. \end{aligned} \tag{2.45}$$

The linear correlation coefficient equals

$$\rho\big(L_1^{(1)}, L_1^{(2)}\big) = \frac{\operatorname{Cov}\big(L_1^{(1)}, L_1^{(2)}\big)}{\sqrt{\operatorname{Var}\big(L_1^{(1)}\big)}\sqrt{\operatorname{Var}\big(L_1^{(2)}\big)}} \approx 0.27.$$

There is positive correlation between the loss of the two credits and a portfolio with these two credits has higher risk than a portfolio, where the credit rating of the firms is independent.

◊

2.4 The scheme model

In the scheme model we specify the distribution P of the rating class functions $s : S \to S$ in our general model, such that rating transitions of firms with different ratings have positive probability. Thereby, we have dependence also between the credit rating classes not only between firms with the same rating because firms with different ratings may change simultaneously. The probability, that firms with rating $x \in S$ actually change the rating class to $s(x)$, is still given by $p_x \in [0,1]$.

Let $(p_{xy})_{x,y\in S} \in [0,1]$ be a stochastic transition function, i. e. $\sum_{y\in S} p_{xy} = 1$ for each $x \in S$. Furthermore let V be a random variable, uniformly distributed on the interval $[0,1]$. Define the S^S-valued random function $\tilde{s}$ by

$$\tilde{s}(x) = \max\Bigg\{ y \in S : \sum_{k=1}^{y-1} p_{xk} \le V \Bigg\}, \quad \text{for } x \in S. \tag{2.46}$$

A graphical illustration of realizations of $\tilde{s}$ is given in Figure 2.1. For each rating class $x \in S$ the interval $[0,1]$ is divided into K subintervals with length p_{xy} for the y-th subinterval. The subinterval, where the random variable V lies in, represents the rating class, to which the firm may change. In the figure the random variable V is represented by the black line and for $x = 1$ it lies in the first subinterval, i. e. $s(1) = 1$, for $x = 2$ also in the first, i. e. $s(2) = 1$, and so on.

Lemma 2.47. *Let $(p_{xy})_{x,y\in S} \in [0,1]$ be a stochastic transition function and V be a random variable, uniformly distributed on $[0,1]$. Define the S^S-valued random function $\tilde{s}$ by* (2.46)*. Then the measure P^{s} on the power set $\mathcal{P}(S^S)$, given by*

$$P^{\mathrm{s}}(s) = \max\Bigg\{ \min_{x\in S} \sum_{k=1}^{s(x)} p_{xk} - \max_{x\in S} \sum_{k=1}^{s(x)-1} p_{xk}, 0 \Bigg\}, \tag{2.48}$$

is the distribution of $\tilde{s}$.

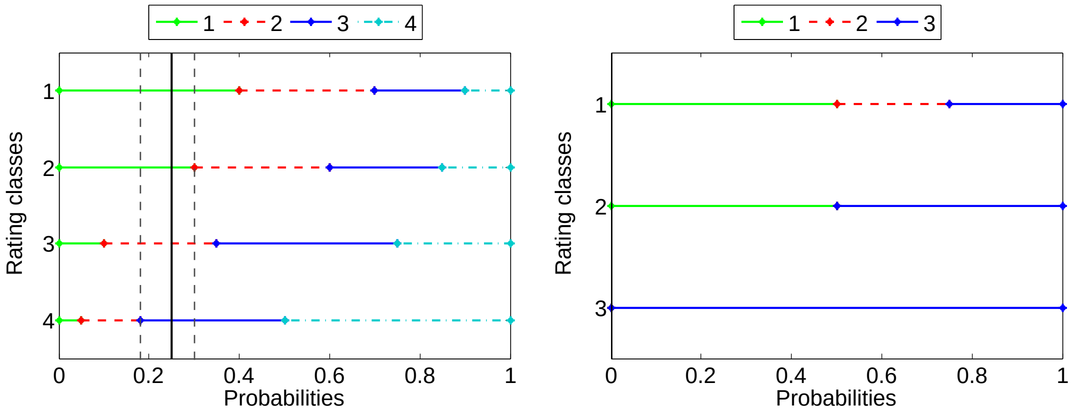

Figure 2.1: Schemes of rating class distributions. Left: Black line is value of V. Interval length between dotted lines is the probability of $\tilde{s} = (1, 1, 2, 3)$. Right: Scheme of Example 2.59.

Proof. For each $s \in S^S$ we obtain

$$\mathbb{P}\left[\tilde{s}(x) = s(x), \text{ for all } x \in S\right] = \mathbb{P}\left[\max\left\{y \in S : \sum_{k=1}^{y-1} p_{xk} \leq V\right\} = s(x), \text{ for all } x \in S\right]$$
$$= \mathbb{P}\left[\sum_{k=1}^{s(x)-1} p_{xk} \leq V < \sum_{k=1}^{s(x)} p_{xk}, \text{ for all } x \in S\right] = \mathbb{P}\left[\max_{x \in S} \sum_{k=1}^{s(x)-1} p_{xk} \leq V < \min_{x \in S} \sum_{k=1}^{s(x)} p_{xk}\right]$$
$$= P^{\mathrm{s}}(s).$$

□

In the scheme model we assume that our random rating class functions have the same distribution as $\tilde{s}$, i. e. $P = P^{\mathrm{s}}$.

Definition 2.49. *Let $(p_{xy})_{x,y \in S} \in [0, 1]$ be a stochastic transition function. Define the probability distribution P^{s} of the rating configurations $s \in S^S$ by (2.48). Let $\lambda \in (0, \infty)$ and let $p = (p_x)_{x \in S}$ be a vector in $[0, 1]^S$.*

We say that the time-homogeneous Markov jump process $(X_t^s)_{t \geq 0}$ follows the scheme model *with parameters $\big(\lambda, (p_{xy})_{x,y \in S}, p\big)$, if X follows the general model with parameters $(\lambda, P^{\mathrm{s}}, p)$, given by Definition 2.15.*

The scheme model allows simultaneous transitions of firms with different credit ratings. By choosing an appropriate scheme $(p_{xy})_{xy \in S}$ we can link different credit rating classes $x, y \in S$ such that for each transition all firms with rating x and y are either all up- or downgraded, i. e. for all $i \in \mathbb{N}$

$$\mathbb{P}\Big[\rho_i \in \big\{r_s \in E : s(x) \leq x \text{ and } s(y) \leq y, \text{ or } s(x) \geq x \text{ and } s(y) \geq y\big\}\Big]$$
$$= P^{\mathrm{s}}\Big(\big\{s \in S^S : s(x) \leq x \text{ and } s(y) \leq y, \text{ or } s(x) \geq x \text{ and } s(y) \geq y\big\}\Big) = 1,$$

where the random mark ρ_i and the possible rating class changes r_s are as in Definition 2.15. The next lemma shows the assumptions on the scheme such that this is satisfied.

Lemma 2.50. *Let $(p_{xy})_{x,y\in S} \in [0,1]$ be a stochastic transition function. Define the probability distribution P^{s} on S^S by (2.48). If for $x, y \in S$ with $x \neq y$*

$$\max\Big\{\sum_{k=1}^{x-1} p_{xk}, \sum_{k=1}^{y-1} p_{xy}\Big\} \le \min\Big\{\sum_{k=1}^{x} p_{xk}, \sum_{k=1}^{y} p_{xy}\Big\}, \tag{2.51}$$

then

$$P^{\mathrm{s}}\Big(\big\{s \in S^S : s(x) \le x \text{ and } s(y) \le y, \text{ or } s(x) \ge x \text{ and } s(y) \ge y\big\}\Big) = 1.$$

Proof. Let V be a random variable, uniformly distributed on $[0,1]$, and define the S^S-valued random function $\tilde{s}$ by (2.46). Then $\tilde{s}$ has distribution P^{s} by Lemma 2.47. Using the Definition (2.46) of $\tilde{s}$ we obtain for each $x, y \in S$

$$\begin{aligned}
&P^{\mathrm{s}}\Big(\big\{s \in S^S : s(x) \le x \text{ and } s(y) \le y, \text{ or } s(x) \ge x \text{ and } s(y) \ge y\big\}\Big) \\
&\quad = \mathbb{P}\Big[\big\{\tilde{s}(x) \le x \text{ and } \tilde{s}(y) \le y\big\} \cup \big\{\tilde{s}(x) \ge x \text{ and } \tilde{s}(y) \ge y\big\}\Big] \\
&\quad = \mathbb{P}\Big[\Big\{V < \min\Big\{\sum_{k=1}^{x} p_{xk}, \sum_{k=1}^{y} p_{yk}\Big\}\Big\} \cup \Big\{\max\Big\{\sum_{k=1}^{x-1} p_{xk}, \sum_{k=1}^{y-1} p_{yk}\Big\} \le V\Big\}\Big].
\end{aligned}$$

This equals 1, if for x and y the equation (2.51) is satisfied. □

To compare this model with the esc–process, we again use as parameters the vector $p = (p_x)_{x\in S}$ and the Q-matrix Q_1^{g} of an individual moving firm, which is given by the next Lemma for processes following the scheme model.

Lemma 2.52. *Let $\lambda \in (0,\infty)$ and let $(p_{xy})_{x,y\in S} \in [0,1]$ be a stochastic transition function. Let $p = (p_x)_{x\in S}$ be a probability vector in $[0,1]^S$. Then the Q-matrix $Q_1^{\mathrm{g}} : S \times S \to \mathbb{R}$ of the scheme model with state space S and parameters $\big(\lambda, (p_{xy})_{x,y\in S}, p\big)$ is given by*

$$Q_1^{\mathrm{g}}(x,y) = \lambda p_{xy} p_x, \quad \text{for } x, y \in S \text{ with } x \neq y. \tag{2.53}$$

Proof. Let X follow the scheme model with state space S and parameters $\big(\lambda, (p_{xy})_{x,y\in S}, p\big)$. By definition, X follows the general model with parameters $(\lambda, P^{\mathrm{s}}, p)$, where the distribution P^{s} is given by (2.48). Using Theorem 2.22 the Q-matrix Q_1^{g} of X is given by

$$Q_1^{\mathrm{g}}(x,y) = \lambda\, p_x \sum_{\substack{s\in S^S \\ s(x)=y}} P^{\mathrm{s}}(s), \quad \text{for each } x, y \in S \text{ with } x \neq y. \tag{2.54}$$

Let V be a random variable, uniformly distributed on $[0,1]$, and define the S^S-valued random function $\tilde{s}$ by (2.46). Then $\tilde{s}$ has distribution P^{s} by Lemma 2.47. Therefore we obtain

$$\sum_{\substack{s\in S^S \\ s(x)=y}} P^{\mathrm{s}}(s) = \mathbb{P}[\tilde{s}(x) = y] = \mathbb{P}\Big[\sum_{k=1}^{y-1} p_{xk} < V \le \sum_{k=1}^{y} p_{xk}\Big] = p_{xy}.$$

□

Assuming $p_x > 0$ for all $x \in S$, we can parametrize the processes X following the scheme model also by parameters (μ, p), where $\mu : S \times S \to \mathbb{R}$ is the Q-matrix of the single firms and $p = (p_x)_{x\in S}$.

Definition 2.55. *Let $\mu : S \times S \to \mathbb{R}$ be a Q-matrix and $p = (p_x)_{x\in S}$ be a vector in $[0,1]^S$, where $p_x > 0$ for all $x \in S$. We say the time-homogeneous Markov jump process X follows the scheme model with parameters (μ, p), if it follows the scheme model with parameters $\big(\lambda, (p_{xy})_{x,y\in S}), p\big)$, where*

$$\lambda = \max_{x\in S} \frac{\mu_x}{p_x}, \quad \text{and} \quad p_{xy} = \frac{\mu_{xy}}{\lambda p_x}, \quad \text{for each } x, y \in S \text{ with } x \neq y. \tag{2.56}$$

Lemma 2.57. *Let $\mu : S \times S \to \mathbb{R}$ be a Q-matrix and $p = (p_x)_{x\in S}$ be a vector in $[0,1]^S$, where $p_x > 0$ for all $x \in S$. Let X follow the scheme model with state space S^n and parameters (μ, p). Then μ corresponds to the transition intensity of each single firm in the process X, i. e., for each firm $j \in F$ the credit rating process $\big(X_t(j)\big)_{t\geq 0}$ has Q-matrix μ.*

Proof. By definition, X follows the scheme model with parameters $\big(\lambda, (p_{xy})_{x,y\in S}, p\big)$, where λ and $(p_{xy})_{x,y\in S}$ are defined by (2.56). By Lemma 2.52 the Q-matrix Q_1^{g} of a Markov process following the scheme model with state space S and parameters $\big(\lambda, (p_{xy})_{x,y\in S}, p\big)$ is given by

$$Q_1^{\mathrm{g}}(x,y) = \lambda p_{xy} p_x, \quad \text{for each } x, y \in S \text{ with } x \neq y.$$

Since X follows the scheme model, it follows also the general model and the embedding property in Theorem 2.13 holds. Therefore the process $\big(X_t(j)\big)_{t\geq 0}$ has Q-matrix Q_1^{g}, which equals μ with (2.56). □

The new parametrization by μ and p generates the same class of Markov processes as the parametrization by $(\lambda, (p_{xy})_{x,y\in S}, p)$ assuming $p_x > 0$ for all $x \in S$. This is shown by the following lemma.

Lemma 2.58. *Let $\lambda \in (0, \infty)$, and let $(p_{xy})_{x,y\in S} \in [0,1]$ be a stochastic transition function. Let $p = (p_x)_{x\in S}$ be a probability vector in $[0,1]^S$, where $p_x > 0$ for all $x \in S$.*

Then the Q-matrix Q of the scheme model X with parameters $(\lambda, (p_{xy})_{x,y\in S}, p)$ equals the Q-matrix $\tilde{Q}$ of the process $\tilde{X}$ following the scheme model with parameters (μ, p), where the Q-matrix $\mu : S \times S \to \mathbb{R}$ is given by (2.53).

Proof. Define

$$\tilde{\lambda} = \max_{x\in S} \frac{\mu_x}{p_x}, \quad \text{and} \quad \tilde{p}_{xy} = \frac{\mu_{xy}}{\tilde{\lambda} p_x}, \quad \text{for each } x, y \in S \text{ with } x \neq y.$$

By Definition 2.55 the process $\tilde{X}$ follows the scheme model with parameters $(\tilde{\lambda}, (\tilde{p}_{xy})_{x,y\in S}, p)$. The process X, resp. $\tilde{X}$, follows the general model with parameters $(\lambda, P^{\mathrm{s}}, p)$, resp. $(\tilde{\lambda}, \tilde{P}^{\mathrm{s}}, p)$, by Definition 2.49 of the scheme model, where P^{s}, resp. $\tilde{P}^{\mathrm{s}}$, is defined by (2.48). Using (2.53) for μ and the definition of $\tilde{p}_{xy}$ we obtain for each $s \in S^S$

$$\begin{aligned} P^{\mathrm{s}}(s) &= \max\left\{ \min_{x\in S} \sum_{k=1}^{s(x)} p_{xk} - \max_{x\in S} \sum_{k=1}^{s(x)-1} p_{xk}, 0 \right\} = \max\left\{ \min_{x\in S} \sum_{k=1}^{s(x)} \frac{\mu_{xk}}{\lambda p_x} - \max_{x\in S} \sum_{k=1}^{s(x)-1} \frac{\mu_{xk}}{\lambda p_x}, 0 \right\} \\ &= \max\left\{ \min_{x\in S} \sum_{k=1}^{s(x)} \tilde{p}_{xk} \frac{\tilde{\lambda}}{\lambda} - \max_{x\in S} \sum_{k=1}^{s(x)-1} \tilde{p}_{xk} \frac{\tilde{\lambda}}{\lambda}, 0 \right\} = \frac{\tilde{\lambda}}{\lambda} \tilde{P}^{\mathrm{s}}(s). \end{aligned}$$

The definition of the Q-matrix in Theorem 2.22 for the general model implies $Q = \tilde{Q}$. □

Example 2.59 (Continuation of Example 2.43)**.** Consider the rating classes $S = \{1, 2, 3\}$, where 3 means default. Let $(X_t^s)_{t\geq 0}$ be the credit rating process of two firms following the scheme model with parameters μ and $p_x = 1$ for all $x \in S$, where μ is defined by (2.44). Then the intensity λ and the stochastic transition function p_{xy} for $x, y \in S$ are by Definition 2.55

$$\lambda = \max_{x\in S} \frac{\mu_x}{p_x} = 1, \quad (p_{xy})_{x,y\in S} = \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0 & 0 & 1 \end{pmatrix},$$

see also the graph on the right-hand side in Figure 2.1. We obtain for the distribution on the credit configurations $s \in S^S$, given by (2.48),

$$P^{\mathrm{s}}(s_1) = 0.25,\ P^{\mathrm{s}}(s_2) = 0.25,\ P^{\mathrm{s}}(s_3) = 0.5,$$

where $s_1(x) = 3$ for each $x \in S$, $s_2(1) = 2$, $s_2(2) = 3 = s_2(3)$ and $s_3(1) = 1$, $s_3(2) = 1$ and $s_3(3) = 3$. The probability that both firms default within one year is

$$\mathbb{P}_{(1,2)}\big[X_1^s = (3,3)\big] = \exp(Q_2^{\mathrm{s}})_{(1,2),(3,3)} \approx 0.239,$$

where $X_0^s = (1, 2)$ almost surely, where Q_2^{s} is the Q-matrix of the process X^s, given by Theorem 2.22. In the scheme model the dependence increases the probability that both firms default compared to the esc–process in Example 2.43 since we additionally introduced dependence between the credit rating classes. Here, it is possible that the firms with rating 1 and 2 default at the same time. Therefore the probability for the default of both firms is higher in this model than in the extended strongly coupled random walk model.

In addition to the probability for a joint default we also consider the losses of two credits, where the credit rating of the two obligors follows the credit rating process $(X_t^s)_{t\geq 0}$ with $X_0^s = (1, 2)$ a. s. Assume the maturity of the credits is $T \geq 0$ and the credit amount is 1. Furthermore we suppose that the recovery rate and the default free interest rate are zero. Then the loss at time $t \in (0, T]$, induced by the default of firm $j \in \{1, 2\}$, is given by $L_t^{(j)} = \mathbb{1}_{\{X_t^s(j)=K\}}$. The covariance of the losses after one year is

$$\mathrm{Cov}\big(L_1^{(1)}, L_1^{(2)}\big) = \mathbb{P}_{(1,2)}\big[X_1 = (3,3)\big] - \exp(\tilde{\mu})_{1,3}\, \exp(\tilde{\mu})_{2,3} \approx 0.153. \tag{2.60}$$

The linear correlation coefficient equals $\rho\big(L_1^{(1)}, L_1^{(2)}\big) \approx 0.75$. The dependence between the losses is stronger than in the extended strongly coupled random walk model. ♢

In the next chapter we simulate the esc–process and the scheme model for specialized parameters p and μ and compare the losses in a portfolio of credits. We expect that there is more dependence in the scheme model, since we have additional dependence between the credit rating classes. Therefore extremal losses/gains should be more likely.

Chapter 3

Simulation of the model

The following chapter illustrates how the dependence in our model influences the value of a portfolio of credits. We assume that it is not possible for the bank to sell the risk of the credits. This is reasonable for small clients/firms. If the firm defaults and cannot meet its obligations, then the bank has a loss. If there exists positive correlation between the defaults of the firms within a portfolio, then the probability of high losses increases compared to the independent case.

3.1 Loss of a credit portfolio

Assume we have a portfolio of $n \in \mathbb{N}$ credits with credit amount $C_i^{\mathrm{a}} \in [0, \infty)$ and maturity $T_i \in (0, \infty)$ for the i-th credit. The obligors change their credit rating according to the credit rating process X. The credit rating classes are given by $S = \{1, \ldots, K\}$ and K is the default state which we assume that is absorbing. A loss occurs if a firm defaults, i. e., the firm is rated with K. Let $(\mathcal{F}_t)_{t \geq 0}$ be the natural filtration of X. Furthermore we assume that there is no interest rate, and the recovery rate of the i-th credit is constant $\delta_i \in [0, 1]$ for $i \in \{1, \ldots, n\}$. Since K is an absorbing state, the loss L of the portfolio is given by

$$L(t) = \sum_{i=1}^{n} C_i^{\mathrm{a}} (1 - \delta_i) \mathbb{1}_{\{X_{t \wedge T_i}(i) = K\}}, \quad \text{for all } t \geq 0.$$

In the following we consider the distribution of the loss assuming X is an esc–process or follows the scheme model with parameters (μ, p), where $\mu = (\mu_{xy})_{x,y \in S} \in \mathbb{R}^{K \times K}$ is the Q-matrix of the transitions of an individual firm and $p = (p_x)_{x \in S}$ is a probability vector in $[0, 1]^S$. Furthermore we suppose the process starts in $X_0 = z_0 \in S^n$. By Lemma 2.41, resp. Theorem 2.13, the embedding property holds for the process X. Therefore the projected process $(\pi_i(X_t))_{t \geq 0}$, where $\pi_i : S^n \to S$ is defined by $\pi_i(x_1, \ldots, x_n) = x_i$ for each $i \in \{1, \ldots, n\}$, is a Markov jump process with respect to $(\mathcal{F}_t)_{t \geq 0}$ generated by the Q-matrix μ and starting in $z_0(i)$. Hence we obtain for all $t \geq 0$

$$\mathbb{E}[L(t)] = \sum_{i=1}^{n} C_i^{\mathrm{a}} (1 - \delta_i) \mathbb{P}_{z_0} \left[X_{t \wedge T_i}(i) = K \right] = \sum_{i=1}^{n} C_i^{\mathrm{a}} (1 - \delta_i) \exp\left\{ \mu (t \wedge T_i) \right\}_{z_0(i), K}.$$

The expected loss depends only on the Q-matrix μ of the transitions of an individual firm and does not depend on the parameter p. The distribution of the loss, however, depends also on p. Since we are interested in the probability that we loose more than certain values, we consider the excess distribution of the loss at time $t \geq 0$, i. e.

$$\mathbb{P}[L(t) > x] = \mathbb{P}\Big[\sum_{i=1}^{n} C_i^{\mathrm{a}}(1-\delta_i)\mathbb{1}_{\{X_{t\wedge T_i}(i)=K\}} > x\Big], \quad \text{for } x \geq 0.$$

Assuming the recovery rate $\delta_i = \delta \in [0,1]$, the credit amount $C_i^{\mathrm{a}} = C^{\mathrm{a}} \in (0,\infty)$ and the maturity $T_i = T \in (0,\infty)$ for all $i \in \{1,\dots,n\}$ we obtain for $x \in [0,(1-\delta)C^{\mathrm{a}}n]$

$$\mathbb{P}[L(t) > x] = \sum_{j=\left\lceil \frac{x}{(1-\delta)C^{\mathrm{a}}} \right\rceil}^{n} \mathbb{P}\left[\sum_{i=1}^{n} \mathbb{1}_{\{X_{t\wedge T}(i)=K\}} = j\right] = \sum_{j=\left\lceil \frac{x}{(1-\delta)C^{\mathrm{a}}} \right\rceil}^{n} \sum_{z\in\mathcal{Z}} \exp\big\{Q(t\wedge T)\big\}_{z_0,z},$$

where

$$\mathcal{Z} := \Big\{z \in S^n : \#\big\{i \in \{1,\dots,n\} : z(i) = K\big\} = j\Big\},$$

and $Q : S^n \times S^n \to \mathbb{R}$ is the Q-matrix of X, i. e. $Q = Q_n^{\mathrm{c}}$ given by (2.34) for the esc–process, resp. $Q = Q_n^{\mathrm{g}}$ given by Theorem 2.22 for the scheme model. For example for $K = 8$ credit rating classes and $n = 100$ firms the matrix Q has $4 \cdot 10^{180}$ entries and the numerical computation of the exact excess loss distribution is impossible. Therefore, we simulate the credit rating process X and compute the loss for each simulation. To show the influence of dependence we continue with a numerical example.

3.2 Numerical example of the loss of a credit portfolio

To model realistic behavior of transitions of an individual firm we estimate the Q-matrix μ using the average one-year transition probability matrix P_S from 1981 to 2006 in the annual European corporate default report of Standard & Poor's [56, Table 10]. In this case we have $K = 8$ credit rating classes. To find a Q-matrix μ corresponding to P_S, we apply the method of Israel, Rosenthal and Wei (2001). If all diagonal entries of P_S are greater than 1/2, the series

$$\tilde{\mu} = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(P_S - I)^k}{k},$$

converges, where I is the $K \times K$ identity matrix, and $P_S = \exp\tilde{\mu}$ (cf. [31]). Therefore, $\tilde{\mu}$ is a Q-matrix of P_S, if all non-diagonal entries of $\tilde{\mu}$ are non-negative.

Computing the series $\tilde{\mu}$ for the transition matrix of Standard & Poor's we obtain negative non-diagonal entries. Hence, $\tilde{\mu}$ is not a valid Q-matrix. To solve this problem, we set the negative non-diagonal entries equal to zero and add the absolute value to the diagonal entry in the same row (cf. [31]). Thus, it approximately holds $P_S \approx \exp(\mu)$, where the Q-matrix

$$\mu = \begin{pmatrix} -0.106 & 0.103 & 0.002 & 0.002 & 0 & 0 & 0 & 0 \\ 0.003 & -0.107 & 0.101 & 0.003 & 0 & 0 & 0 & 0 \\ 0 & 0.033 & -0.088 & 0.054 & 0.001 & 0.001 & 0 & 0 \\ 0 & 0.002 & 0.056 & -0.098 & 0.033 & 0.004 & 0.002 & 0.001 \\ 0 & 0 & 0 & 0.043 & -0.171 & 0.122 & 0.003 & 0.004 \\ 0 & 0 & 0.003 & 0.004 & 0.098 & -0.233 & 0.097 & 0.031 \\ 0 & 0 & 0 & 0 & 0 & 0.225 & -1.008 & 0.783 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

is computed in the way described above.

We consider a portfolio of $n = 100$ credits maturing at $T = 15$ with credit amount $C^{\mathrm{a}} = 1$. Suppose the credit rating process X of the underlying firms follows either the extended strongly coupled random walk model or the scheme model with the vector $(p_x)_{x\in S}$ and the Q-matrix μ. The process starts with 16 firms in the best rating class 1 and 14 firms in each of the classes 2 to 7. For simplification we choose the same dependence parameter $p_x = p$ for all rating classes $x \in S$. Assume $\delta = 0.4$ is the recovery rate δ_i for all firms $i \in \{1, \ldots, 100\}$.

Figure 3.1 and 3.2 show the empirical distribution functions of the excess loss for the esc–process and for the scheme model at times $t = 1$ and $t = 5$ for different parameters p. The estimation is based on 5 000 simulations. We see an increase of the probability for high losses, if we have a stronger coupling between the firms with the same rating, i. e., p is closer to one. For example, after five years the probability for a loss greater than 15 for $p = 1$ equals approximately 31.7%, for $p = 0.5$ only 13%, and for $p = 0$ only 0.1% in the extended strongly coupled random walk model. Therefore the dependence within the rating classes allows more possible shapes of loss distributions.

Simulation of both models leads to a similar shape of the excess loss distribution. The influence of the coupling by p dominates the induced dependence between the credit rating classes by P^{s} of the scheme model. To illustrate the differences, Figure 3.3 shows the histograms of the simulated losses for both models at time $t = 5$ for the parameter $p = 0.5$. We simulate the paths of the processes 10 000 times. We observe that the number of high losses is higher in the scheme model than in the extended strongly coupled random walk model. Since the transition intensities of an individual firm are the same in both models this indicates stronger positive correlation in the scheme model. In the esc–process each rating class has an independent Poisson process. If this process jumps, then we choose another rating class. Therefore, downgrades, resp. upgrades, of different credit rating classes are independent. In the scheme model the downgrades of firms in different rating classes are linked by the uniformly distributed random variable V. If V is close to one, resp. zero, then all firms are simultaneously downgraded, resp. upgraded, or keep their credit rating. The transitions of the firms with different ratings are correlated and the probability of high losses increases compared to the esc–process.

Another remarkable observation is a peak at 8.4 in both histograms. This corresponds to 14 defaults within 5 years. A firm with credit rating 7 defaults in average after 1.3 years. Therefore if there is high dependence between the firms, then it is likely that all 14 firms starting with rating 7 are defaulted after 5 years. On the other hand, firms in rating class 6 default in average after 12 years. Therefore the number of simulations, where more than 14 firms default, is significantly less than the number where 14 firms default. If we consider the paths after a longer time period, then this effect is smoothing out.

To measure the riskiness of the credit portfolio we estimate the Value-at-Risk (VaR) and the expected shortfall (ES) at the confidence level $\alpha = 0.95$, resp. $\alpha = 0.99$. In Table 3.1 the firms follow the esc–process and in Table 3.2 the scheme model. Varying p shows that the risk increases with higher dependence. Underestimation of the dependence leads to unexpected high losses, as our model shows. The incorporation of the dependence allows more flexibility in the model. Furthermore, the risk of a credit portfolio following the scheme model is higher than following the esc–process.

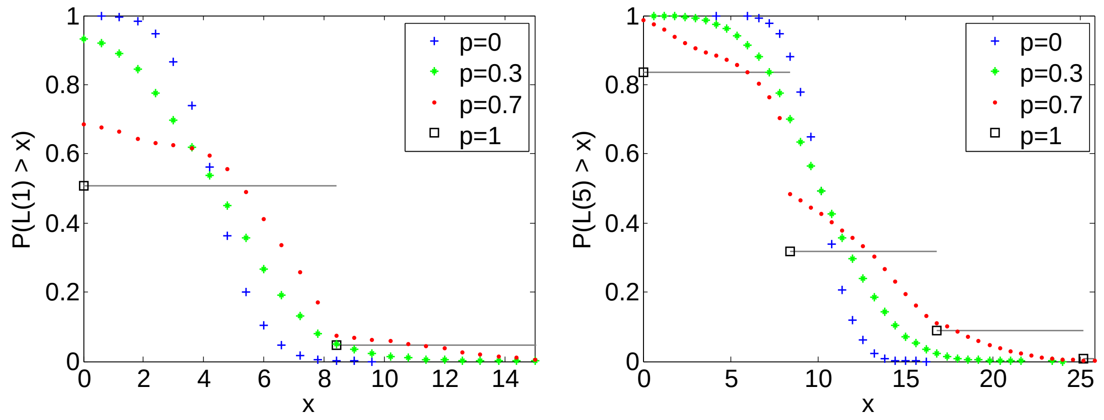

Figure 3.1: Empirical excess loss distribution of the credit portfolio for the extended strongly coupled random walk at times $t = 1$ and $t = 5$.

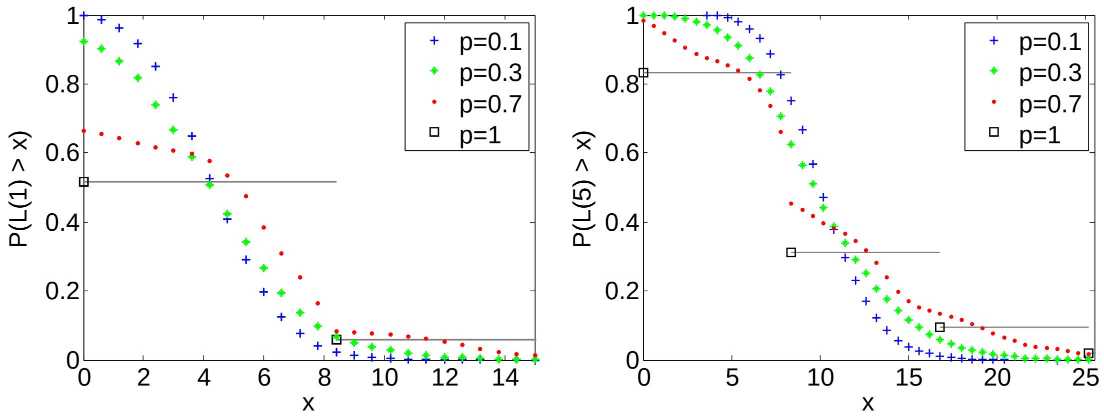

Figure 3.2: Empirical excess loss distribution of the credit portfolio for the scheme model at times $t = 1$ and $t = 5$.

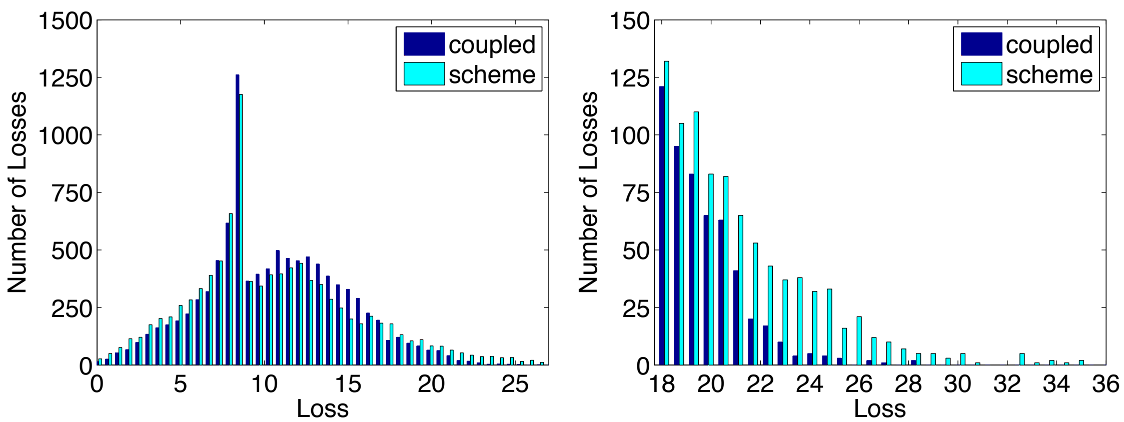

Figure 3.3: Histogram of the simulated losses for the extended strongly coupled random walk (coupled) and the scheme model (scheme) where $p = 0.5$ and $t = 5$, based on 10 000 simulations.

	$p = 0$	$p = 0.3$	$p = 0.5$	$p = 0.7$	$p = 1$
$t = 1$	6.6	8.4	10.2	11.4	8.4
$t = 2$	9.0	11.4	13.2	14.4	16.8
$t = 5$	13.2	16.2	18.0	19.8	25.2
$t = 10$	16.8	19.8	22.2	24.0	25.2
$t = 15$	19.8	23.4	25.2	27.6	33.6

$\mathrm{VaR}_{0.95}$ of the credit portfolio for time points t and probability p.

	$p = 0$	$p = 0.3$	$p = 0.5$	$p = 0.7$	$p = 1$
$t = 1$	7.43	10.14	11.88	13.43	16.50
$t = 2$	9.87	13.00	14.71	16.56	19.05
$t = 5$	13.61	17.37	19.69	22.16	26.52
$t = 10$	17.74	21.51	24.47	26.86	33.44
$t = 15$	20.69	24.91	27.77	31.20	39.91

$\mathrm{ES}_{0.95}$ of the credit portfolio for time points t and probability p.

	$p = 0$	$p = 0.3$	$p = 0.5$	$p = 0.7$	$p = 1$
$t = 1$	7.8	10.8	13.2	14.4	16.8
$t = 2$	10.2	13.8	15.6	18.0	25.2
$t = 5$	13.8	18.0	21.0	23.4	25.2
$t = 10$	18.0	22.8	25.8	28.8	33.6
$t = 15$	21.0	25.8	29.4	33.6	43.2

$\mathrm{VaR}_{0.99}$ of the credit portfolio for time points t and probability p.

	$p = 0$	$p = 0.3$	$p = 0.5$	$p = 0.7$	$p = 1$
$t = 1$	8.18	12.17	13.99	16.33	18.81
$t = 2$	10.63	14.86	17.33	20.17	25.87
$t = 5$	14.53	19.43	22.25	25.51	31.78
$t = 10$	18.88	23.81	28.06	32.14	40.70
$t = 15$	21.83	27.22	31.42	36.92	52.68

$\mathrm{ES}_{0.99}$ of the credit portfolio for time points t and probability p.

Table 3.1: Value-at-Risk (VaR) and expected shortfall (ES) of the credit portfolio with 100 firms and constant recovery of 40%. Firms follow the esc–process with probability parameter p.

	$p = 0.1$	$p = 0.3$	$p = 0.5$	$p = 0.7$	$p = 1$
$t = 1$	7.8	9.6	10.8	12.6	16.8
$t = 2$	10.8	12.6	13.8	15.0	16.8
$t = 5$	15.0	17.4	19.8	21.6	25.2
$t = 10$	19.8	22.2	24.6	26.4	33.6
$t = 15$	22.8	26.4	28.8	30.0	33.6

$\mathrm{VaR}_{0.95}$ of the credit portfolio for time points t and probability p.

	$p = 0.1$	$p = 0.3$	$p = 0.5$	$p = 0.7$	$p = 1$
$t = 1$	8.95	11.15	13.28	15.17	18.28
$t = 2$	11.86	14.55	17.11	18.97	21.24
$t = 5$	16.30	19.69	23.17	25.42	29.00
$t = 10$	21.04	25.11	28.60	31.18	36.28
$t = 15$	24.23	29.48	32.96	35.54	41.91

$\mathrm{ES}_{0.95}$ of the credit portfolio for time points t and probability p.

	$p = 0.1$	$p = 0.3$	$p = 0.5$	$p = 0.7$	$p = 1$
$t = 1$	9.6	12.0	15.0	16.2	16.8
$t = 2$	12.6	16.2	19.2	21.0	25.2
$t = 5$	17.4	21.0	24.6	27.6	33.6
$t = 10$	22.2	27.0	31.2	34.8	42.0
$t = 15$	25.2	31.8	36.0	39.0	50.4

$\mathrm{VaR}_{0.99}$ of the credit portfolio for time points t and probability p.

	$p = 0.1$	$p = 0.3$	$p = 0.5$	$p = 0.7$	$p = 1$
$t = 1$	10.76	13.69	17.56	20.76	24.19
$t = 2$	13.92	17.39	21.55	24.98	29.06
$t = 5$	18.41	22.86	28.15	31.76	36.98
$t = 10$	23.21	29.28	34.62	39.07	45.48
$t = 15$	26.42	33.96	38.58	43.85	54.82

$\mathrm{ES}_{0.99}$ of the credit portfolio for time points t and probability p.

Table 3.2: Value-at-Risk (VaR) and expected shortfall (ES) of the credit portfolio with 100 firms and constant recovery of 40%. Firms follow the scheme model with probability parameter p.

Chapter 4

Maximum likelihood estimation

In the following chapter we compute the maximum likelihood estimators for the parameters in the general model and in the esc–process. Using these estimated parameters, we are able to simulate the process X and predict the rating transitions of the firms. The observations for the estimation are realizations of the sample paths of the Markov process. In reality the credit rating of firms within different industry sectors may be independent of each other and follow different credit rating processes. The number of firms in the process is given by the size of the industry. Furthermore, the time length of the observations may depend on the sector, since some industries have shorter rating history than others.

For the estimation we want to use the observations with different length of the paths and different number of firms. We assume that each Markov process follows the general model, resp. the esc–process, with the same parameters and constant number of firms. The credit rating transitions of different Markov processes are independent of each other. By this, we can use more observations for our estimation.

4.1 Likelihood function for the general model

The parameters in our general model are the intensity $\lambda \in (0, \infty)$ of the Poisson process, the probability distribution $P \in \mathcal{M}_1(S^S)$ of the rating configurations and the probabilities $p_x \in [0, 1]$ for each $x \in S$ that the firms actually change the rating class, i. e., $\Theta = (0, \infty) \times \mathcal{M}_1(S^S) \times [0, 1]^K$ is the set of the parameters. In the following we estimate the true parameter $\theta_0 \in \Theta$ via the maximum likelihood technique.

For each parameter $\theta \in \Theta$ and each number $n \in \mathbb{N}$ of firms let $(X_{t,n})_{t\geq 0}$ follow the general model with state space S^n and parameter θ, given by Definition 2.15. We observe independent sample paths of the processes $(X_{t,n})_{t\geq 0}$ with $n \in \mathbb{N}$ firms of length $T \in (0, \infty)$. We assume that the probability that we observe a Markov process with n firms is given by $P_N(n) \in [0, 1]$ and ξ is the probability distribution on $\mathcal{B}([(0, \infty))$ of the observed length of the paths. If for some particular $n \in \mathbb{N}$ and $T \in (0, \infty)$ the probability $P_N(n) = 1$ and $\xi(T) = 1$, then we always observe paths of the Markov process $(X_{t,n})_{t\geq 0}$ with length T.

The maximum likelihood estimator gives the parameter that maximizes the density of the observed sample paths of the process. Therefore we start with the construction of the space of the sample paths of $(X_{t,n})_{t\geq 0}$ for $n \in \mathbb{N}$ and the density of the sample paths in the time interval $[0, T)$. This is inspired by the work of Albert (1962), where he describes the density on the space of the sample paths, generated by a Markov jump process with a finite state space and a given Q-matrix, in general.

4.1.1 The space of the sample paths

Every (time-homogeneous) Markov jump process with finite state space has a modification, where the sample paths are piecewise constant functions with a finite number of jumps in the time interval $[0, T)$ with $T > 0$, see e. g. [21, Chapter 4.2]. We assume in the following, that the sample paths ω of the process $(X_{t,n})_{t\in[0,T)}$ for all $n \in \mathbb{N}$ are of the form $\omega = ((z_0, t_0), \dots, (z_{l-1}, t_{l-1}), z_l)$, where $l \in \mathbb{N}_0$ is the number of jumps, $z_i \in S^n$ for $i = 0, \dots, l$, $z_i \neq z_{i+1}$ and $t_i > 0$ for $i = 0, \dots, l-1$. The process starts in the state z_0 and remains there for the time t_0. Then the process jumps to the state z_1 and remains there for the time t_1 and so on. Finally it reaches z_l and stays there until T.

So, for all $T > 0$ every sample path in a model with $n \in \mathbb{N}$ firms, which jumps $l \in \mathbb{N}$ times in $[0, T)$, is an element of $\mathcal{S}_{n,l} = (S^n \times \mathbb{R})^l \times S^n$, and the path is in $\mathcal{S}_{n,0} = S^n$, if there are no jumps. Therefore a sample path of $(X_{t,n})_{t\in[0,T)}$ with n firms is in

$$\mathcal{S}_n = \bigcup_{l=0}^{\infty} \mathcal{S}_{n,l}, \quad \text{for each } n \in \mathbb{N}.$$

Sample paths with an arbitrary number $n \in \mathbb{N}$ of firms lie in

$$\mathcal{S} = \bigcup_{n=1}^{\infty} \mathcal{S}_n \,. \tag{4.1}$$

To define a measure on the space of the sample paths, we define $\mathcal{C}$ as the smallest σ-algebra containing all subsets of $\mathcal{S}$ whose intersection with $\mathcal{S}_{n,l}$ is a Borel set for each pair $(n, l) \in \mathbb{N} \times \mathbb{N}_0$. Furthermore we define for each $n \in \mathbb{N}$ the smallest σ-algebra $\mathcal{C}_n$, which contains all subsets of $\mathcal{S}_n$ whose intersection with $\mathcal{S}_{n,l}$ is a Borel set for each pair $(n, l) \in \mathbb{N}\times\mathbb{N}_0$. Let $\mathcal{V}$ be the Lebesgue–Borel measure on $\mathbb{R}$ and $\mathcal{N}_n$ be the counting measure on S^n. Define $\sigma_{n,l}$ as the product measure on $\mathcal{S}_{n,l}$ by

$$\sigma_{n,l} = (\mathcal{N}_n \otimes \mathcal{V})^{\otimes l} \otimes \mathcal{N}_n, \quad \text{for each } n \in \mathbb{N} \text{ and } l \in \mathbb{N}_0,$$

and for the sets $C_n \in \mathcal{C}_n$ the measure

$$\sigma_n(C_n) = \sum_{l=0}^{\infty} \sigma_{n,l}(C_n \cap \mathcal{S}_{n,l}), \quad \text{for each } n \in \mathbb{N}.$$

For each $C \in \mathcal{C}$ the intersection $(C \cap \mathcal{S}_n) \cap \mathcal{S}_{n,l} = C \cap \mathcal{S}_{n,l}$ is a Borel set for each $l \in \mathbb{N}_0$. Therefore $(C \cap \mathcal{S}_n) \in \mathcal{C}_n$ and we define the measure

$$\sigma(C) = \sum_{n=1}^{\infty} \sigma_n(C \cap \mathcal{S}_n). \tag{4.2}$$

Denote by $\mathcal{O}$ the *space of the observations*, given by

$$\mathcal{O} = \mathbb{N} \times (0, \infty) \times \mathcal{S} = \mathbb{N} \times (0, \infty) \times \bigcup_{n=1}^{\infty} \mathcal{S}_n. \tag{4.3}$$

The first entry is the number of firms in the observed path. The second entry is the observation time, and the third are the credit ratings of the firms in the observed time.

4.1.2 Density of the sample paths

For the description of a density of the observed sample paths we need the definition of a weakly feasible path. Since we later restrict the model to the extended strongly coupled random walk, we also introduce the definition of a strongly feasible path.

Definition 4.4. *Let $(\omega, T) \in \mathcal{S} \times (0, \infty)$ be a path until time T , where*

$$\omega = \big((z_0, t_0), \ldots, (z_{l-1}, t_{l-1}), z_l\big), \quad \text{with } l \in \mathbb{N}_0,\ t_j > 0 \text{ and } z_i \in S^n \text{ for } n \in \mathbb{N}.$$

(i) We call (ω, T) a weakly feasible path until time T*, if $\sum_{j=0}^{l-1} t_j < T$ and there are only weakly feasible transitions, i. e., the transitions between z_i to z_{i+1} are weakly feasible for $i = 0, \ldots, l-1$ in the sense of Definition 2.20.*

(ii) We call (ω, T) a strongly feasible path until time T*, if $\sum_{j=0}^{l-1} t_j < T$ and there are only strongly feasible transitions in the sense of Definition 2.29.*

Definition 4.5. *Let $(\omega, T) \in \mathcal{S}_n \times (0, \infty)$ be a weakly feasible path until time T with*

$$\omega = ((z_0, t_0), \ldots, (z_{l-1}, t_{l-1}), z_l), \quad \text{with } l \in \mathbb{N}_0,\ t_j > 0 \text{ and } z_i \in S^n.$$

Let $\bar{a} : S \to \mathbb{N}_0$ and $\bar{b} : S \to \mathbb{N}_0$ be functions that map each rating class to a number of firms. Define the total time that the path spends *with $\bar{a}(x)$ firms in the rating class $x \in S$ for all $x \in S$, by*

$$T(\omega, T, \bar{a}) = \sum_{j=0}^{l-1} t_j \prod_{x \in S} \mathbb{1}_{\{\bar{a}(x) = a_x(z_j)\}} + \Big(T - \sum_{j=0}^{l-1} t_j\Big) \prod_{x \in S} \mathbb{1}_{\{\bar{a}(x) = a_x(z_l)\}}, \tag{4.6}$$

where $a_x(z)$ for $x \in S$ and $z \in S^n$ is given by Definition 2.19. Define the number of transitions, where $\bar{b}(x)$ firms of $\bar{a}(x)$ firms with rating $x \in S$ change the rating class to $s(x)$ *for all $x \in S$, according to any rating class function $s \in S^{\mathrm{c}} \subseteq S^S$, by*

$$\begin{aligned} N(\omega, S^{\mathrm{c}}, \bar{a}, \bar{b}) = \#\big\{ j \in \{0, \ldots, l-1\} : \mathcal{S}^{\mathrm{c}}(z_j, z_{j+1}) = S^{\mathrm{c}}, \\ a_x(z_j) = \bar{a}(x),\ b_x(z_j, z_{j+1}) = \bar{b}(x),\ \text{for all } x \in S \big\}, \end{aligned} \tag{4.7}$$

where $b_x(z, \tilde{z})$ for $x \in S$ and $z, \tilde{z} \in S^n$ is given by Definition 2.19 and $\mathcal{S}^{\mathrm{c}}$ by Definition 2.21.

Theorem 4.8. *Let $\theta = (\lambda, P, p) \in \Theta$ be a parameter, where $p = (p_x)_{x \in S}$. For each $n \in \mathbb{N}$ let the Markov jump process $(X_{t,n})_{t \geq 0}$ follow the general model with state space S^n and parameter θ, given by Definition 2.15. Let ν_n be the distribution of the initial state $X_{0,n}$ and assume it is the same for all $\theta \in \Theta$. For each $n \in \mathbb{N}$ define the function $f_n : \mathcal{S} \times (0, \infty) \to [0, \infty)$ by*

$$\begin{aligned} f_n(\omega, T; \theta) = \nu_n\big(\{z_0\}\big) \exp\Big\{ -\lambda \sum_{\bar{a} \in \mathbb{N}_0^K} T(\omega, T, \bar{a}) \sum_{\substack{s \in S^S \\ s \neq \mathrm{id}}} P(s) \big(1 - \prod_{\substack{x \in S \\ s(x) \neq x}} (1 - p_x)^{\bar{a}_x}\big) \Big\} \\ \times \prod_{S^{\mathrm{c}} \subseteq S^S} \prod_{\bar{a} \in \mathbb{N}_0^K} \prod_{\substack{\bar{b} \in \mathbb{N}_0^K \\ \bar{a} \geq \bar{b}}} \Big(\lambda \sum_{s \in S^{\mathrm{c}}} P(s) \prod_{\substack{x \in S \\ s(x) \neq x}} p_x^{\bar{b}_x} (1 - p_x)^{\bar{a}_x - \bar{b}_x} \Big)^{N(\omega, S^{\mathrm{c}}, \bar{a}, \bar{b})}, \end{aligned}$$

if $\omega \in \mathcal{S}_n$, where $z_0 \in S^n$ is the first entry of ω, and (ω, T) is a weakly feasible path until time T, or $f_n(\omega, T; \theta) = 0$ otherwise.

Then for each $T > 0$ and $n \in \mathbb{N}$ the function $f_n(\cdot, T)$ is a density of the sample paths of the process $(X_{t,n})_{t \geq 0}$ until time T with respect to the measure σ, which is given by (4.2).

Remark 4.9. The function f_n is $\mathcal{C} \otimes \mathcal{B}((0,\infty))$-$\mathcal{B}(\mathbb{R})$ measurable, since f_n is continuous on the open set

$$\{(\omega, T) \in \mathcal{S}_n \times (0,\infty) : (\omega, T) \text{ is a weakly feasible path}\}.$$

Proof. Let $n \in \mathbb{N}$ and $T \in (0,\infty)$. Since $X = (X_{t,n})_{t\geq 0}$ follows the general model, the Q-matrix Q_n^{g} is given by Theorem 2.22. Albert proved that for each $T > 0$ a density $g : \mathcal{S}_n \to \mathbb{R}$ of the paths of a finite-state Markov process $(X_{t,n})_{t\in[0,T)}$ following the Q-matrix Q_n^{g} with respect to σ_n is given by

$$g(\omega) = \begin{cases} \mathbb{P}(X_{0,n} = z_0) e^{-Q_n^{\mathrm{g}}(z_0)T}, & \text{if } \omega = (z_0), \\ \mathbb{P}(X_{0,n} = z_0) e^{-Q_n^{\mathrm{g}}(z_l)T} \prod\limits_{j=0}^{l-1} Q_n'(z_j, z_{j+1}) e^{-(Q_n^{\mathrm{g}}(z_j) - Q_n^{\mathrm{g}}(z_l))t_j}, & \\ \qquad \text{if } \omega = ((z_0, t_0), \ldots, (z_{l-1}, t_{l-1}), z_l), & \\ \qquad \text{with } l \geq 1,\ z_j \in S^n, t_j > 0, \text{ and } \sum\limits_{j=0}^{l-1} t_j < T, & \\ 0, & \text{otherwise,} \end{cases} \tag{4.10}$$

where $Q_n^{\mathrm{g}}(z) = -Q_n^{\mathrm{g}}(z,z)$ and for $z, \tilde{z} \in S^n$

$$Q_n'(z, \tilde{z}) = \begin{cases} 0, & \text{if } z = \tilde{z}, \\ Q_n^{\mathrm{g}}(z, \tilde{z}), & \text{otherwise.} \end{cases}$$

If $\omega \in \mathcal{S}_n$ has an infeasible transition z_j to z_{j+1}, it can easily be seen that the density $g(\omega) = 0$ for each $T > 0$, since $Q_n'(z_j, z_{j+1}) = 0$. Therefore $g(\omega) = f_n(\omega, T)$. Let $(\omega, T) \in \mathcal{S}_n \times (0,\infty)$ be a weakly feasible path until time T with

$$\omega = ((z_0, t_0), \ldots, (z_{l-1}, t_{l-1}), z_l), \quad \text{with } l \in \mathbb{N}_0,\ t_j > 0 \text{ and } z_i \in S^n \text{ for } n \in \mathbb{N}.$$

Applying the definition of Q_n^{g} in Theorem 2.22 to the density of Albert in (4.10) we obtain for each $T > 0$

$$\begin{aligned} g(\omega) = \nu_n\big(\{z_0\}\big) \prod_{j=0}^{l-1} \lambda \Bigg(\sum_{s \in \mathcal{S}^{\mathrm{c}}(z_j, z_{j+1})} P(s) \prod_{\substack{x \in S \\ s(x) \neq x}} p_x^{b_x(z_j, z_{j+1})} (1 - p_x)^{a_x(z_j) - b_x(z_j, z_{j+1})} \Bigg) \\ \times \exp\Bigg\{ -\lambda \sum_{\substack{s \in S^S \\ s \neq \mathrm{id}}} P(s) \sum_{j=0}^{l-1} t_j \Big(1 - \prod_{\substack{x \in S \\ s(x) \neq x}} (1 - p_x)^{a_x(z_j)} \Big) \Bigg\} \\ \times \exp\Bigg\{ -\lambda \sum_{\substack{s \in S^S \\ s \neq \mathrm{id}}} P(s) \Big(T - \sum_{j=0}^{l-1} t_j \Big) \Big(1 - \prod_{\substack{x \in S \\ s(x) \neq x}} (1 - p_x)^{a_x(z_l)} \Big) \Bigg\}. \end{aligned}$$

Using the definition of $N(\omega, S^{\mathrm{c}}, \bar{a}, \bar{b})$ and $T(\omega, T, \bar{a})$ we obtain $g(\omega) = f_n(\omega, T)$ for all $\omega \in \mathcal{S}_n$. Since $f_n(\omega, T) = 0$ for $\omega \notin \mathcal{S}_n$ we finally have for all $C \in \mathcal{C}$

$$\begin{aligned} \mathbb{P}\big[(X_{t,n})_{t\in[0,T)} \in C\big] &= \mathbb{P}\big[(X_{t,n})_{t\in[0,T)} \in C \cap \mathcal{S}_n\big] \\ &= \int_{C \cap \mathcal{S}_n} g(\omega)\, \sigma_n(d\omega) = \int_C f_n(\omega, T)\, \sigma(d\omega). \end{aligned}$$

□

The observed paths are sample paths of $(X_{t,n})_{t\geq 0}$ where the number $n \in \mathbb{N}$ of firms has probability $P_N(n)$ and the time length of the paths has distribution ξ. The density of the sample paths of $(X_{t,n})_{t\geq 0}$ until time $T \in (0,\infty)$ is the function $f_n(\cdot, T)$ for each $n \in \mathbb{N}$.

Definition 4.11. *Let $\mathcal{O}$ be the space of observations defined in* (4.3). *Let ξ be a probability measure on the Borel σ-algebra $\mathcal{B}((0,\infty))$ and P_N be a probability measure on the power set $\mathcal{P}(\mathbb{N})$. For each parameter $\theta \in \Theta$ define the function $h_\theta : \mathcal{O} \to [0,\infty)$ by*

$$h_\theta(n, T, \omega) = f_n(\omega, T; \theta), \tag{4.12}$$

where $f_n(\omega, T; \theta)$, given by Theorem 4.8, is a density of the sample path of $(X_{t,n})_{t\in[0,T)}$ until time $T > 0$ with respect to measure σ, given by (4.2).

(i) Define the stochastic kernel $\kappa : \big(\mathbb{N} \times (0,\infty)\big) \times \mathcal{C} \to [0,\infty]$ by

$$\kappa\big((n,T), C\big) = \int_C f_n(\omega, T)\sigma(d\omega).$$

By Theorem 4.8 for each $n \in \mathbb{N}$ and $T > 0$ the measure $C \mapsto \kappa((n,T), C)$ is the distribution of the sample path of $(X_{t,n})_{t\in[0,T)}$.

(ii) We call η a random observation *in $\mathcal{O}$ of the* general model, *if the distribution of η is the measure $P_N \otimes \xi \otimes \kappa$ on $\big(\mathcal{O}, \mathcal{P}(\mathbb{N}) \otimes \mathcal{B}((0,\infty)) \otimes \mathcal{C}\big)$, i. e. for each $N \in \mathcal{P}(\mathbb{N})$, $\mathcal{T} \in \mathcal{B}((0,\infty))$ and $C \in \mathcal{C}$*

$$\begin{aligned}\mathbb{P}\big[\eta \in N \times \mathcal{T} \times C\big] &= \sum_{n\in N} P_N(n) \int_{\mathcal{T}} \kappa((n,T), C)\,\xi(dT)\\ &= \sum_{n\in N} P_N(n) \int_{\mathcal{T}} \int_C h_\theta(n, T, \omega)\sigma(d\omega)\xi(dT).\end{aligned}$$

Therefore h_θ is a density of η with respect to the measure $P_N \otimes \xi \otimes \sigma$.

Remark 4.13. If $P_N(n) = 1$ for some particular $n \in \mathbb{N}$ and $\xi(T) = 1$ for some particular $T \in (0,\infty)$, then η has density $f_n(\omega, T)$ with respect to σ. The path has the distribution of a random sample path of $(X_{t,n})_{t\geq 0}$ until time T.

Remark 4.14. The probability $P_N(n)$ is the probability that we observe Markov processes with n firms. We assume that the economy consists of industry sectors with size $n \in \mathbb{N}$ and the credit rating of these sectors follow general models with the same parameter θ. However, the credit ratings of firms in different sectors are independent. The measure ξ is the probability distribution of the observation time of the credit ratings. In different sectors and different observations the rating history may vary. The length of the observed paths are independent of the credit rating dynamics in the paths. We assume that the observed firms in the different paths change their rating independently of firms in other paths.

4.1.3 Likelihood function for the general model

Let $\theta_0 \in \Theta$ be the true parameter. Let $(n_k, T_k, \omega_k) \in \mathcal{O}$ for $k \in \{1, \ldots, m\}$ with $m \in \mathbb{N}$ be observed paths, i. e., (n_k, T_k, ω_k) are independent realizations of the paths η, where the paths have density h_{θ_0}, and assume $h_{\theta_0}(n_k, T_k, \omega_k) > 0$. In particular, (ω_k, T_k) is weakly feasible. To estimate the true parameter θ_0 with the maximum likelihood technique, we maximize the likelihood function $\mathcal{L} : \Theta \to [0, \infty)$, which is given by

$$\mathcal{L}(\theta) = \prod_{k=1}^{m} h_\theta(n_k, T_k, \omega_k).$$

For the general model we cannot expect a tractable solution for the maximization problem. However, in the case of the extended strongly coupled random walk we are able to deduce the maxima.

4.2 Likelihood estimator for the esc–process

The parameters in the extended strongly coupled random walk model are the probabilities $p_x \in [0, 1]$ for each $x \in S$ that the firms actually change the rating class and the Q-matrix $\mu = (\mu_{xy})_{x,y \in S}$ of an independently moving firm.

Definition 4.15. *Let $\Theta_c = \big([0, \infty)^{K-1} \times [0, 1]\big)^K$ be the set of parameters. For the parameter $\theta \in \Theta_c$ set $\theta = (\theta_x)_{x \in S}$, where $\theta_x = (\mu_{x,1}, \dots, \mu_{x,x-1}, \mu_{x,x+1}, \dots, \mu_{x,K}, p_x)$. The vector $p = (p_x)_{x \in S}$ is a probability vector and $\mu = (\mu_{xy})_{x,y \in S}$ is the Q-matrix of the individually moving firms, where $\mu_{xx} = -\mu_x$ is given by* (2.36)*.*

4.2.1 Density of the sample paths

To obtain the density of the sample path we do the same steps as for the general model. We start with the density for the paths of the Markov process $(X^{c}_{t,n})_{t \geq 0}$ w. r. t. the measure σ until time $T > 0$ for each $n \in \mathbb{N}$. If we weight the sample paths of the different Markov processes by the probability measure P_N, which is the probability to observe a path with n firms, and the distribution of the length of the paths, then we obtain the density of our observed sample paths.

Definition 4.16. *Let $(\omega, T) \in \mathcal{S} \times (0, \infty)$ be a path until time T, where*

$$\omega = \big((z_0, t_0), \dots, (z_{l-1}, t_{l-1}), z_l\big), \quad \text{with } l \in \mathbb{N}_0,\ t_j > 0 \text{ and } z_i \in S^n \text{ for } n \in \mathbb{N}.$$

Denote by $a_x(z)$ the number of firms in state $z \in S^n$ with rating $x \in S$ by

$$a_x(z) = \#\big\{j \in \{1, \dots, n\} : z(j) = x\big\}.$$

(i) For each $x \in S$ and $a \in \mathbb{N}$ define the total time $T^{(\omega,T)}_{x,a}$ that the path spends in rating class x with a firms by

$$T^{(\omega,T)}_{x,a} = \sum_{i=0}^{l-1} t_i \mathbb{1}_{\{a_x(z_i)=a\}} + (T - \sum_{i=0}^{l-1} t_i) \mathbb{1}_{\{a_x(z_l)=a\}}, \tag{4.17}$$

if (ω, T) is strongly feasible and by $T^{(\omega,T)}_{x,a} = 0$ otherwise.

(ii) For each $x, y \in S$ with $x \neq y$ and $a, b \in \mathbb{N}$ with $a \geq b$ call $N^{(\omega)}_{x,y,a,b}$ the number of transitions from rating class x to y with a firms having class x before and $a - b$ firms are in class x after the rating change, i. e.

$$\begin{aligned} N^{(\omega)}_{x,y,a,b} = \#\Big\{ &i \in \{0, \dots, l-1\} : a_x(z_i) = a, \\ &b = \#\big\{j \in \{1, \dots, n\} : z_i(j) = x, z_{i+1}(j) = y\big\}\Big\}, \end{aligned} \tag{4.18}$$

if ω has only strongly feasible transitions, and $N^{(\omega)}_{x,y,a,b} = 0$ otherwise.

For the likelihood function we need the density of the sample paths of $(X^{\mathrm{c}}_{t,n})_{t\geq 0}$ with respect to the measure σ.

Lemma 4.19. *Let $\theta = (\theta_x)_{x\in S} \in \Theta_c$ be a parameter, where*

$$\theta_x = (\mu_{x,1}, \ldots, \mu_{x,x-1}, \mu_{x,x+1}, \ldots, \mu_{x,K}, p_x), \quad \text{for each } x \in S.$$

For each $n \in \mathbb{N}$ let $(X^{\mathrm{c}}_{t,n})_{t\geq 0}$ be an esc–process with state space S^n and parameter θ, given by Definition 2.40. Let ν^{c}_n be the distribution of the initial state $X^{\mathrm{c}}_{0,n}$ and assume it is the same for all $\theta \in \Theta_c$. For each $n \in \mathbb{N}$ define the function $f^{\mathrm{c}}_n : \mathcal{S} \times (0,\infty) \to [0,\infty)$ by

$$f^{\mathrm{c}}_n(\omega, T; \theta) = \nu^{\mathrm{c}}_n\big(\{z_0(\omega)\}\big) \exp\Big\{ -\sum_{x\in S} \mu_x \sum_{a=1}^{n} T^{(\omega,T)}_{x,a}\, q(a, p_x)\Big\} \times \prod_{\substack{x,y\in S\\ x\neq y}} \prod_{\substack{a,b=1\\ a\geq b}}^{n} \Big(\mu_{xy} p_x^{b-1}(1-p_x)^{a-b}\Big)^{N^{(\omega)}_{x,y,a,b}},$$

for a strongly feasible path $(\omega, T) \in \mathcal{S}_n \times (0,\infty)$, or $f^{\mathrm{c}}_n(\omega, T; \theta) = 0$ otherwise, where $z_0(\omega)$ is the first component of ω and q is defined in (2.35), and μ_x in (2.36).

Then for each $T > 0$ the function $f^{\mathrm{c}}_n(\cdot, T)$ is a density of the sample paths of the process $(X^{\mathrm{c}}_{t,n})_{t\geq 0}$ until time T with respect to the measure σ, which is given by (4.2).

Proof. For each $T > 0$ and $n \in \mathbb{N}$ the density of the sample paths of the process $(X^{\mathrm{c}}_{t,n})_{t\in[0,T)}$ is given by $g : \mathcal{S}_n \to [0,\infty)$ defined by (4.10) for the Q-matrix Q^{c}_n. If a transition from $z \in S^n$ to $\tilde{z} \in S^n$ is not strongly feasible, we obtain $Q^{\mathrm{c}}_n(z, \tilde{z}) = 0$. Therefore $g(\omega) = 0$ for $(\omega, T) \in \mathcal{S}_n \times (0,\infty)$ not strongly feasible.

Assume $(\omega, T) \in \mathcal{S}_n \times (0,\infty)$ is a strongly feasible path until time T given by

$$\omega = \big((z_0, t_0), \ldots, (z_{l-1}, t_{l-1}), z_l\big), \quad \text{with } l \in \mathbb{N}_0, t_j > 0 \text{ and } z_i \in S^n.$$

For each $i \in \{0, \ldots, l-1\}$ define by $x_i \in S$ the original rating of firms which are changing the rating to $y_i \in S$ in the transition z_i to z_{i+1}, i. e., for all $j \in \{1, \ldots, n\}$ with $z_i(j) \neq z_{i+1}(j)$ we have $z_i(j) = x_i$ and $z_{i+1}(j) = y_i$, which is well-defined since z_i to z_{i+1} is a strongly feasible transition. Substituting Q^{c}_n, given by Lemma 2.33, into the density in (4.10), we obtain

$$g(\omega) = \nu^{\mathrm{c}}_n\big(\{z_0(\omega)\}\big) \prod_{i=1}^{l-1} \mu_{x_i y_i} p_{x_i}^{b_{x_i}(z_i, z_{ji-1})-1} (1-p_{x_i})^{a_{x_i}(z_i) - b_{x_i}(z_i, z_{i+1})} \times \exp\Big\{ -\sum_{x\in S} \mu_x \Big[\sum_{i=0}^{l-1} t_i\, q(a_x(z_i), p_x) + \Big(T - \sum_{i=0}^{l-1} t_i\Big) q(a_x(z_l), p_x)\Big]\Big\}.$$

The definition of $N^{(\omega)}_{x,y,a,b}$ and $T^{(\omega,T)}_{x,a}$ in (4.18) and (4.17) yield $g(\omega) = f^{\mathrm{c}}_n(\omega, T)$ for all $\omega \in \mathcal{S}_n$. Since $f^{\mathrm{c}}_n(\omega, T) = 0$ for $\omega \notin \mathcal{S}_n$ we finally have for all $C \in \mathcal{C}$

$$\begin{aligned}\mathbb{P}\big[(X^{\mathrm{c}}_{t,n})_{t\in[0,T)} \in C\big] &= \mathbb{P}\big[(X^{\mathrm{c}}_{t,n})_{t\in[0,T)} \in C \cap \mathcal{S}_n\big] \\ &= \int_{C\cap\mathcal{S}_n} g(\omega)\, \sigma_n(d\omega) = \int_C f^{\mathrm{c}}_n(\omega, T)\, \sigma(d\omega).\end{aligned}$$

$\square$

Analogously to the definition of the observations in the general model we define the density of observation of the esc–processes. Again, we use as observations sample paths of different Markov processes, where the distribution of the paths conditional on the firm number $n \in \mathbb{N}$ and the observation length $T \in (0, \infty)$ is given by the density $f_n^{\mathrm{c}}(\cdot, T)$.

Definition 4.20. *Let $\mathcal{O}$ be the space of observations defined in* (4.3)*. Let ξ be a probability measure on the Borel σ-algebra $\mathcal{B}((0, \infty))$ and P_N be a probability measure on the power set $\mathcal{P}(\mathbb{N})$. For each parameter $\theta \in \Theta$ define the function $h_\theta^{\mathrm{c}} : \mathcal{O} \to [0, \infty)$ by*

$$h_\theta^{\mathrm{c}}(n, T, \omega) = f_n^{\mathrm{c}}(\omega, T; \theta), \tag{4.21}$$

where $f_n^{\mathrm{c}}(\omega, T; \theta)$, given by Lemma 4.19, is a density of the sample path of $(X_{t,n}^{\mathrm{c}})_{t\in[0,T)}$ until time $T > 0$ with respect to measure σ, given by (4.2)*.*

(i) Define the stochastic kernel $\kappa^c : \big(\mathbb{N} \times (0, \infty)\big) \times \mathcal{C} \to [0, \infty]$ by

$$\kappa^c\big((n, T), C\big) = \int_C f_n^{\mathrm{c}}(\omega, T)\sigma(d\omega).$$

By Lemma 4.19 for each $n \in \mathbb{N}$ and $T > 0$ the measure $C \mapsto \kappa^c((n, T), C)$ is the distribution of the sample path of $(X_{t,n}^{\mathrm{c}})_{t\in[0,T)}$.

(ii) We call η a random observation *in $\mathcal{O}$ of the* esc–process*, if the distribution of η is the measure $P_N \otimes \xi \otimes \kappa^c$ on $\big(\mathcal{O}, \mathcal{P}(\mathbb{N}) \otimes \mathcal{B}((0, \infty)) \otimes \mathcal{C}\big)$, i. e. for each $N \in \mathcal{P}(\mathbb{N})$, $\mathcal{T} \in \mathcal{B}((0, \infty))$ and $C \in \mathcal{C}$*

$$\begin{aligned}\mathbb{P}\big[\eta \in N \times \mathcal{T} \times C\big] &= \sum_{n\in N} P_N(n) \int_{\mathcal{T}} \kappa^c((n, T), C)\, \xi(dT)\\ &= \sum_{n\in N} P_N(n) \int_{\mathcal{T}} \int_C h_\theta^{\mathrm{c}}(n, T, \omega)\sigma(d\omega)\xi(dT).\end{aligned}$$

Therefore h_θ^{c} is a density of η with respect to the measure $P_N \otimes \xi \otimes \sigma$.

4.2.2 Maximum likelihood estimators

With the density of the observations we can state the likelihood function. To reduce notation we define the number of rating changes and the time, which the firms spend in the configurations.

Definition 4.22. *Suppose $(\omega_k, T_k) \in \mathcal{S} \times (0, \infty)$ are strongly feasible sample paths until time T_k for $k \in \{1, \dots, m\}$ with $m \in \mathbb{N}$ in the sense of Definition 4.4. We define the* number of rating changes of $b \in \mathbb{N}$ firms from $x \in S$ to $y \in S$, *where $a \geq b$ firms had rating x originally, by*

$$N_{x,y,a,b} = \sum_{k=1}^{m} N_{x,y,a,b}^{(\omega_k)}, \quad \text{for } x, y \in S \text{ with } x \neq y \text{ and } a, b \in \mathbb{N} \text{ with } a \geq b, \tag{4.23}$$

where $N_{x,y,a,b}^{(\omega_k)}$ is defined in (4.18)*. The* total number of rating changes of b firms from x, *where a firms had originally rating x, is given by*

$$\tilde{N}_{x,a,b} = \sum_{y\in S\setminus\{x\}} N_{x,y,a,b}, \quad \text{for } x \in S \text{ and } a, b \in \mathbb{N} \text{ with } a \geq b. \tag{4.24}$$

Using $T_{x,a}^{(\omega_k,T_k)}$ *in* (4.17)*, we define the* total time of a firms in x *by*

$$T_{x,a} = \sum_{k=1}^{m} T_{x,a}^{(\omega_k,T_k)}, \quad \textit{for } x \in S \textit{ and } a \in \mathbb{N}. \tag{4.25}$$

Let $\theta_0 \in \Theta_c$ be the true parameter. Suppose $(n_k, T_k, \omega_k) \in \mathcal{O}$ for $k \in \{1, \dots, m\}$ with $m \in \mathbb{N}$ are observed paths until time T_k, i. e., (n_k, T_k, ω_k) are independent realizations of the random path η of the esc–process, where the path has density $h^{\mathrm{c}}_{\theta_0}$, and assume $h^{\mathrm{c}}_{\theta_0}(n_k, T_k, \omega_k) > 0$. In particular, (ω_k, T_k) is strongly feasible. Let $n := \max_{k\in\{1,\dots,m\}} n_k$ be the maximal number of firms in the paths. Applying the notation in 4.22 and the definition of the density of the observed paths in 4.20, the likelihood function of the esc–process is given by

$$\begin{aligned} \mathcal{L}^{\mathrm{c}}(\theta) = \prod_{k=1}^{m} h^{\mathrm{c}}_{\theta}(n_k, T_k, \omega_k) &= \bigg(\prod_{k=1}^{m} \nu^{\mathrm{c}}_{n_k}\big(\{z_0(\omega_k)\}\big)\bigg) \exp\bigg\{ - \sum_{x\in S} \mu_x \sum_{a=1}^{n} T_{x,a}\, q(a, p_x)\bigg\} \\ &\quad \times \prod_{\substack{x,y\in S\\ x\neq y}} \prod_{\substack{a,b=1\\ a\geq b}}^{n} \Big(\mu_{xy} p_x^{b-1} (1-p_x)^{a-b}\Big)^{N_{x,y,a,b}}, \end{aligned} \tag{4.26}$$

for each $\theta = (\theta_x)_{x\in S} \in \Theta_c$ with $\theta_x = (\mu_{x,1}, \dots, \mu_{x,x-1}, \mu_{x,x+1}, \dots, \mu_{x,K}, p_x)$, where function q is defined in (2.35) and μ_x in (2.36). The maximum $\hat{\theta}$ of the function $\mathcal{L}^{\mathrm{c}}$ is a maximum likelihood estimator for the true parameter θ_0.

After preparation we state the likelihood estimators of the esc–process in the next theorem and prove it in Section 4.2.3.

Theorem 4.27. *Let* $(n_k, T_k, \omega_k) \in \mathbb{N} \times (0, \infty) \times \mathcal{S}$ *for* $k \in \{1, \dots, m\}$ *with* $m \in \mathbb{N}$ *be independent realizations of the observed paths in the esc–process, i. e., the paths have density* $h^{\mathrm{c}}_{\theta_0}$ *with* $\theta_0 \in \Theta_c$*, given by* (4.20)*. Suppose* $h^{\mathrm{c}}_{\theta_0}(n_k, T_k, \omega_k) > 0$ *for each* $k \in \{1, \dots, m\}$*. Let* $n := \max_{k\in\{1,\dots,m\}} n_k$ *be the maximal number of firms in the paths. Let* $F = \{1, \dots, n\}$ *be the set of observed firms. Furthermore, for every* $x \in S$ *define the polynomial* $P_x : [0, 1] \to \mathbb{R}$ *with at most degree* n *by*

$$P_x(p) = \sum_{j=0}^{n} c_j(x)\, p^j, \tag{4.28}$$

with the coefficients

$$c_0(x) = \sum_{\substack{a,b=1\\ a\geq b}}^{n} (b-1)\tilde{N}_{x,a,b} \sum_{k=1}^{n} k\, T_{x,k}$$

and for $j \in \{1, \dots, n\}$

$$c_j(x) = (-1)^j \sum_{\substack{a,b=1\\ a\geq b}}^{n} \tilde{N}_{x,a,b} \sum_{k=j}^{n} \binom{k}{j} \Big(\frac{k-j}{j+1}\, b + a - k\Big)\, T_{x,k}.$$

Define the index $l_x \in \{0, \dots, n\}$ *by*

$$l_x = \max\big\{ j \in \{0, \dots, n\} : \ c_i(x) = 0, \textit{ for } 0 \leq i \leq j-1 \big\}.$$

Then the parameters in the set $\hat{\Theta} \subset \Theta_c$ *are exactly the maximum likelihood estimators, where for* $\hat{\theta} = (\hat{\theta}_x)_{x\in S} \in \hat{\Theta}$ *with* $\hat{\theta}_x = (\hat{\mu}_{x,1}, \dots, \hat{\mu}_{x,x-1}, \hat{\mu}_{x,x+1}, \dots, \hat{\mu}_{x,K}, \hat{p}_x)$ *holds:*

(i) *For each* $x \in S$ *with* $T_{x,a} = 0$ *for all* $a \in F$*, the entry* $\hat{\theta}_x \in [0,\infty)^{K-1} \times [0,1]$.

(ii) *For each* $x \in S$ *the entry* $\hat{p}_x$ *is the unique root in* $(0,1)$ *of the polynomial* P_x*, if there exists* $a, b \in F$ *with* $a > b$ *such that* $\tilde{N}_{x,a,b} > 0$ *and either there exists* $a, b \in F$ *with* $a \geq b \geq 2$ *such that* $\tilde{N}_{x,a,b} > 0$ *or* $c_{l_x}(x) > 0$.

(iii) *For* $x \in S$ *suppose* $\tilde{N}_{x,a,b} = 0$ *for all* $a, b \in F$ *with* $a \geq b \geq 2$ *and* $N_{x,a,1} > 0$ *for any* $a \in \{2, \ldots, n\}$*. Furthermore assume* $c_{l_x}(x) < 0$*. Then the entry* $\hat{p}_x = 0$.

(iv) *For* $x \in S$ *assume* $\tilde{N}_{x,a,b} = 0$ *for all* $a, b \in F$ *with* $a > b$*. If there exists* $a \geq 2$ *with* $\tilde{N}_{x,a,a} > 0$ *or* $\tilde{N}_{x,1,1} > 0$ *and* $T_{x,a} > 0$ *for any* $a \in \{2, \ldots, n\}$*, then the maximum likelihood entry* $\hat{p}_x = 1$.

(v) *The entry* $\hat{p}_x \in [0,1]$ *for each* $x \in S$*, where* $\tilde{N}_{x,a,b} = 0$ *for all* $a, b \in F$ *with* $a \geq 2$ *and either* $\tilde{N}_{x,1,1} = 0$ *or* $T_{x,a} = 0$ *for all* $a \in \{2, \ldots, n\}$.

(vi) *For each* $x \in S$*, such that there exists* $a \in F$ *with* $T_{x,a} > 0$*, the entries* $\hat{\mu}_{xy}$ *equal*

$$\hat{\mu}_{xy} = \sum_{\substack{a,b=1 \\ a \geq b}}^{n} N_{x,y,a,b} \Big(\sum_{k=1}^{n} T_{x,k}\, q(k, \hat{p}_x) \Big)^{-1}, \quad \text{for each } y \in S \text{ with } x \neq y, \tag{4.29}$$

where the function q *is defined in* (2.35).

Remark 4.30. Suppose $\tilde{N}_{x,a,b} = 0$ for all $a, b \in F$ with $a \geq b \geq 2$ and $N_{x,a,1} > 0$ for any $a \in \{2, \ldots, n\}$ for $x \in S$. Then the polynomial P_x is not constant zero or equivalently, $c_{l_x}(x) \neq 0$, which is proved in the next section.

Remark 4.31. The maximum likelihood estimator is unique, if each rating class is attained by the observed paths, and if there is at least one rating change during the observation for all rating classes. Additionally, if for one rating class this rating transition always takes place, when just one firm has this credit rating, then it is necessary for a unique maximum likelihood estimator, that at least two firms have this rating at the same time, anytime during the observation.

Remark 4.32. If the maximal number of the observed firms is $n = 1$, then the number of rating transitions $N_{x,y,a,b} = 0$ for all $a \geq 2$ and $b \in \{1, \ldots, a\}$. Therefore each $\hat{p}_x \in [0,1]$ is a maximum likelihood estimator of p_x for all $x \in S$. We cannot estimate p_x uniquely, since there is only one firm in the system and the transitions of the firm are according to μ independently of p. Since $q(1,p) = 1$ for all $p \in [0,1]$, the estimator for μ is given by

$$\hat{\mu}_{xy} = \frac{N_{x,y,1,1}}{T_{x,1}}, \quad \text{for all } x, y \in S \text{ with } x \neq y \text{ and } T_{x,1} > 0,$$

which is the usual maximum likelihood estimator for a Q-matrix, where the transitions of the underlying process are according to μ.

4.2.3 Proof of the maximum likelihood estimators

Throughout this subsection $(n_k, T_k, \omega_k) \in \mathbb{N} \times (0,\infty) \times \mathcal{S}$ are supposed to be independent realizations of observed paths of the extended strongly coupled random walk process for $k \in \{1, \ldots, m\}$ with $m \in \mathbb{N}$, where $\theta_0 \in \Theta_c = \big([0,\infty)^{K-1} \times [0,1]\big)^K$ is the true parameter, and the paths have density $h^{\mathrm{c}}_{\theta_0}$. Furthermore, we assume $h^{\mathrm{c}}_{\theta_0}(n_k, T_k, \omega_k) > 0$ for all $k \in$

$\{1, \ldots, m\}$. In particular, all paths (ω_k, T_k) are strongly feasible and for each $k \in \{1, \ldots, m\}$ the probability is strictly greater than zero, that the process $(X^{\mathrm{c}}_{t,n_k})_{t\geq 0}$ starts in $z_0(\omega_k)$, which is the first component of ω_k, i. e. $\nu^{\mathrm{c}}_{n_k}\big(\{z_0(\omega_k)\}\big) > 0$.

Let $n := \max_{j\in\{1,\ldots,m\}} n_k$ be the maximal number of firms in the paths. Suppose $F = \{1, \ldots, n\}$ is the set of observed firms. In sense of Definition 4.22 let $N_{x,y,a,b}$ be the number of rating changes, where $a \in \mathbb{N}$ firms have rating $x \in S$ initially and $b \leq a$ firms change the rating to the class $y \in S$, and let $\tilde{N}_{x,a,b}$ be the total number of rating changes. Let $T_{x,a}$ be the total time of $a \in \mathbb{N}$ firms with rating $x \in S$ given by Definition 4.22.

To prove Theorem 4.27, we maximize the likelihood function $\mathcal{L}^{\mathrm{c}}$, which is given by (4.26). Using the conventions $\log(0) = -\infty$ and $0 \cdot (-\infty) = 0$, the log-likelihood function $\log \mathcal{L}^{\mathrm{c}} : \Theta_c \to \mathbb{R} \cup \{-\infty\}$ is given by

$$\log \mathcal{L}^{\mathrm{c}}(\theta) = \sum_{k=1}^{m} \log \nu^{\mathrm{c}}_{n_k}\big(\{z_0(\omega_k)\}\big) + \sum_{x\in S} L_x(\theta_x), \tag{4.33}$$

where $\theta = (\theta_x)_{x\in S}$ with $\theta_x = (\mu_{x,1}, \ldots, \mu_{x,x-1}, \mu_{x,x+1}, \ldots, \mu_{x,K}, p_x)$ and $L_x : [0,\infty)^{K-1} \times [0,1] \to \mathbb{R} \cup \{-\infty\}$ is defined by

$$\begin{aligned} L_x(\theta_x) = &\sum_{\substack{y\in S\\ x\neq y}} \sum_{\substack{a,b=1\\ a\geq b}}^{n} N_{x,y,a,b}\big[\log \mu_{xy} + (b-1)\log p_x + (a-b)\log(1-p_x)\big] \\ &- \mu_x \sum_{a=1}^{n} T_{x,a}\, q(a, p_x). \end{aligned} \tag{4.34}$$

Therefore we reduce the maximization problem to the maximization of L_x with respect to θ_x for each $x \in S$ separately and show that the maximum of L_x is finite.

Remark 4.35. To show asymptotic properties of the estimator in Chapter 5, we have to solve an analogous maximization problem. Therefore we maximize here a more general function L_x, where the variables $N_{x,y,a,b} \in [0,\infty)$ and $T_{x,a} \in [0,\infty)$ for all $x, y \in S$ and $a, b \in \mathbb{N}$ with $a \geq b$. Furthermore, we assume, that if for $x \in S$ and $a \in \mathbb{N}$ there exists $b \in \mathbb{N}$ with $a \geq b$, such that $\tilde{N}_{x,a,b} > 0$, then $T_{x,a} > 0$ which holds in our case.

Suppose total time $T_{x,a} = 0$ for all $a \in F$. Then there is no rating change from x observable in a strongly feasible path and $N_{x,y,a,b} = 0$ for all $y \in S$ with $x \neq y$ and $a, b \in F$ with $a \geq b$. This implies that the function $L_x(\theta_x) = 0$ for all $\theta_x \in [0,\infty)^{K-1} \times [0,1]$ and each θ_x maximizes the function L_x. Therefore, Part (i) of Theorem 4.27 holds. In the following we assume for $x \in S$ that there exists $a \in F$ such that $T_{x,a} > 0$.

First, we maximize L_x with respect to μ_{xy} for all $y \in S$ with $x \neq y$ for constant $p_x = p \in [0,1]$.

Lemma 4.36. *For each $x \in S$ with $T_{x,a} > 0$ for any $a \in F$ and $p \in [0,1]$ the constrained function $L_x : [0,\infty)^{K-1} \times \{p\} \to \mathbb{R} \cup \{-\infty\}$, defined by (4.34), is maximized at*

$$\hat{\theta}_x = (\varphi_{x,1}(p), \ldots, \varphi_{x,x-1}(p), \varphi_{x,x+1}(p), \ldots, \varphi_{x,K}, p),$$

where for each $y \in S$ with $x \neq y$ the function $\varphi_{xy} : [0,1] \to [0,\infty)$ is given by

$$\varphi_{xy}(p) := \sum_{\substack{a,b=1\\ a\geq b}}^{n} N_{x,y,a,b}\Big(\sum_{k=1}^{n} T_{x,k}\, q(k,p)\Big)^{-1}, \tag{4.37}$$

and the function q is defined by (2.35). *If $L_x(\hat{\theta}_x) > -\infty$, then the maximal point $\hat{\theta}_x$ of L_x is unique.*

Proof. Fix $x \in S$ with $T_{x,a} > 0$ for any $a \in F$ and $p \in [0,1]$. The function L_x equals

$$L_x(\theta_x) = \sum_{\substack{y \in S \\ x \neq y}} L_{x,y}(\mu_{xy}) + \sum_{\substack{a,b=1 \\ a \geq b}}^{n} \tilde{N}_{x,a,b}\big[(b-1)\log p + (a-b)\log(1-p)\big],$$

where for each $y \in S$ with $x \neq y$ the function $L_{x,y} : [0,\infty) \to \mathbb{R} \cup \{-\infty\}$ is given by

$$L_{x,y}(\mu_{xy}) = \sum_{\substack{a,b=1 \\ a \geq b}}^{n} N_{x,y,a,b} \log \mu_{xy} - \mu_{xy} \sum_{a=1}^{n} T_{x,a}\, q(a,p).$$

To find a maximal point of L_x it is sufficient to maximize $L_{x,y}$ for each μ_{xy} with $y \in S$ with $x \neq y$ separately. Fix $y \in S$ with $x \neq y$. If $N_{x,y,a,b} = 0$ for all $a, b \in F$ with $a \geq b$, then $\varphi_{xy}(p) = 0$. Furthermore $\mu_{xy} = 0$ maximizes $L_{x,y}$ since $L_{x,y}$ is strictly decreasing with respect to μ_{xy}. Assume that there exists a pair $a, b \in F$ with $a \geq b$ with $N_{x,y,a,b} > 0$. Setting the derivative of $L_{x,y}$ equal to zero, we obtain the critical point

$$\mu_{xy} = \sum_{\substack{a,b=1 \\ a \geq b}}^{n} N_{x,y,a,b} \Big(\sum_{k=1}^{n} T_{x,k}\, q(k,p)\Big)^{-1},$$

which is well defined since $q(k,p) \geq 1$ for all $k \in \{1,\dots,n\}$ and there exist $k \in F$ with $T_{x,k} > 0$. Since the second derivative of $L_{x,y}$ is negative, $L_{x,y}$ has a unique maximum at the critical point. Since $L_{x,y}(\varphi_{xy}(p)) > -\infty$ for each $y \in S$ with $y \neq x$, the maximal point $\hat{\theta}_x$ of L_x is unique, if

$$\sum_{\substack{a,b=1 \\ a \geq b}}^{n} \tilde{N}_{x,a,b}\big[(b-1)\log p + (a-b)\log(1-p)\big] > -\infty,$$

i. e., the maximal point is unique, if $L_x(\hat{\theta}_x) > -\infty$. □

Using Lemma 4.36 it is sufficient to maximize L_x with respect to $p \in [0,1]$ assuming $\mu_{xy} = \varphi_{xy}(p)$ and to show that the maximum of L_x is finite. In the following we maximize the function $\Phi_x : [0,1] \to \mathbb{R} \cup \{-\infty\}$ given by

$$\begin{aligned}\Phi_x(p) :&= L_x\left(\varphi_{x,1}(p),\dots,\varphi_{x,x-1}(p),\varphi_{x,x+1}(p),\dots,\varphi_{x,K}(p),p\right)\\ &= C_x - \sum_{\substack{a,b=1 \\ a \geq b}}^{n} \tilde{N}_{x,a,b}\, \log\Big(\sum_{k=1}^{n} T_{x,k}\, q(k,p)\Big)\\ &\quad + \sum_{\substack{a,b=1 \\ a \geq b}}^{n} \tilde{N}_{x,a,b}\big((b-1)\log p + (a-b)\log(1-p)\big),\end{aligned} \tag{4.38}$$

where C_x equals

$$C_x = \sum_{\substack{y \in S \\ x \neq y}} \sum_{\substack{a,b=1 \\ a \geq b}}^{n} N_{x,y,a,b}\, \Big(\log\Big(\sum_{\substack{j,k=1 \\ j \geq k}}^{n} N_{x,y,j,k}\Big) - 1\Big). \tag{4.39}$$

Lemma 4.40. *Suppose $x \in S$ with $T_{x,a} > 0$ for any $a \in F$. Then Φ_x is constant and $\Phi_x > -\infty$, if and only if $\tilde{N}_{x,a,b} = 0$ for all $a, b \in F$ with $a \geq b$ or $\tilde{N}_{x,a,b} = 0$ for all $a, b \in F$ with $a \geq 2$ and $\tilde{N}_{x,1,1} > 0$ and $T_{x,k} = 0$ for all $k \in \{2, \dots, n\}$.*

Proof. Assume $\tilde{N}_{x,a,b} = 0$ for all $a, b \in F$ with $a \geq b$. Then $\Phi_x(p) = 0$ for all $p \in [0,1]$. If $\tilde{N}_{x,a,b} = 0$ for all $a, b \in F$ with $a \geq 2$ and $\tilde{N}_{x,1,1} > 0$ and $T_{x,k} = 0$ for all $k \in \{2, \dots, n\}$, then we obtain for all $p \in [0,1]$

$$\Phi_x(p) = -\sum_{y \in S \setminus \{x\}} N_{x,y,1,1}(1 + \log T_{x,1} - \log N_{x,y,1,1}) > -\infty,$$

which is also constant.

Define the constant C_x by (4.39). Then $|C_x| < \infty$. Evaluating Φ_x at zero yields

$$\Phi_x(0) = C_x + \sum_{\substack{a,b=1 \\ a \geq b}}^{n} \tilde{N}_{x,a,b}\Big((b-1)\log(0) - \log\Big(\sum_{k=1}^{n} T_{x,k}\, k\Big)\Big).$$

If there exists a pair $a, b \in \{2, \dots, n\}$ with $a \geq b$ such that $\tilde{N}_{x,a,b} > 0$, then $-\infty = \Phi_x(0) < \Phi_x(1/2)$. Therefore $\tilde{N}_{x,a,b} = 0$ for all $a, b \in \{2, \dots, n\}$ with $a \geq b$ is necessary for constant Φ_x and we assume that this holds in the following. Therefore, Φ_x simplifies to

$$\Phi_x(p) = C_x + \sum_{a=1}^{n} \tilde{N}_{x,a,1}\Big((a-1)\log(1-p) - \log\Big(\sum_{k=1}^{n} T_{x,k}\, q(k,p)\Big)\Big).$$

If there exists $a \in \{2, \dots, n\}$ with $\tilde{N}_{x,a,1} > 0$, then $-\infty = \Phi_x(1) < \Phi_x(1/2)$ and Φ_x is not constant. Furthermore, if $\tilde{N}_{x,1,1} > 0$, it is necessary for $\Phi_x(0) = \Phi_x(1)$, that $T_{x,k} = 0$ for all $k \in \{2, \dots, n\}$. □

For proving that all critical points of Φ_x are maxima in Lemma 4.42 below, we need the following.

Lemma 4.41. *For $n \in \mathbb{N}$ and constants $a_0, a_1, \dots, a_n > 0$ define the function $g : [0, \infty) \to (0, \infty)$ by*

$$g(x) = \sum_{j=0}^{n} a_j\, x^j.$$

Then it holds

$$x\left(g'(x)\right)^2 < g(x)g'(x) + x\, g(x)g''(x), \quad \text{for all } x \in [0, \infty).$$

Proof. We get for the left-hand side of the inequality

$$x\left(g'(x)\right)^2 = \sum_{i,j=0}^{n} ij\, a_i a_j\, x^{i+j-1} < \sum_{i,j=0}^{n} \frac{i^2 + j^2}{2} a_i a_j x^{i+j-1} = \sum_{i,j=0}^{n} i^2 a_i a_j x^{i+j-1},$$

which is a strict inequality, since $a_0 a_1 > 0$, if $a_j > 0$ for all $j \in \{0, \dots, n\}$ and $n \geq 1$. The right-hand side equals

$$g(x)g'(x) + x\, g(x)g''(x) = \sum_{i,j=0}^{n} i\, a_i a_j x^{i+j-1} + \sum_{i,j=0}^{n} i(i-1) a_i a_j x^{i+j-1}.$$

Therefore the inequality holds. □

Lemma 4.42. *The critical points of the function Φ_x, defined by* (4.38)*, for $x \in S$ with $T_{x,a} > 0$ for any $a \in F$ are the roots of the polynomial P_x, defined in* (4.28)*, that belong to the interval* $(0,1)$*. Furthermore each critical point is a maximum.*

Proof. The function Φ_x is continuous for $p \in [0,1]$ and differentiable for $p \in (0,1)$. The first derivative of Φ_x equals for $p \in (0,1)$

$$\Phi_x'(p) = \sum_{\substack{a,b=1 \\ a \geq b}}^{n} \tilde{N}_{x,a,b} \left(\frac{b-1}{p} - \frac{a-b}{1-p} - \frac{\sum_{k=1}^{n} T_{x,k} \frac{\partial q}{\partial p}(k,p)}{\sum_{j=1}^{n} T_{x,j}\, q(j,p)} \right). \tag{4.43}$$

Since $q(k,p) = \frac{1-(1-p)^k}{p}$ for $p > 0$ and for all $k \in \{1, \dots, n\}$, we obtain

$$\begin{aligned}\Phi_x'(p) = \Big(\sum_{k=1}^{n} T_{x,k}(1-(1-p)^k) \Big)^{-1} \Bigg[&\sum_{\substack{a,b=1 \\ a \geq b}}^{n} \tilde{N}_{x,a,b} \sum_{k=1}^{n} T_{x,k}(1-(1-p)^k) \left(\frac{b-ap}{p(1-p)} - \frac{1}{p} \right) \\ &- \sum_{\substack{a,b=1 \\ a \geq b}}^{n} \tilde{N}_{x,a,b} \sum_{k=1}^{n} T_{x,k} \left(k(1-p)^{k-1} - \frac{1-(1-p)^k}{p} \right) \Bigg].\end{aligned}$$

Rearrangement leads to

$$\begin{aligned}\Phi_x'(p) = &\Big(p(1-p) \sum_{k=1}^{n} T_{x,k}(1-(1-p)^k) \Big)^{-1} \\ &\times \sum_{\substack{a,b=1 \\ a \geq b}}^{n} \tilde{N}_{x,a,b} \sum_{k=1}^{n} T_{x,k} \Big[b - ap - b(1-p)^k + ap(1-p)^k - k\,p(1-p)^k \Big].\end{aligned}$$

Since the binomial theorem says for $p \in (0,1)$ and $k \in \mathbb{N}_0$

$$(1-p)^k = 1 + \sum_{j=1}^{k} \binom{k}{j} (-p)^j,$$

the derivative is

$$\begin{aligned}\Phi_x'(p) = &\Big(p(1-p) \sum_{k=1}^{n} T_{x,k}(1-(1-p)^k) \Big)^{-1} \\ &\times \sum_{\substack{a,b=1 \\ a \geq b}}^{n} \tilde{N}_{x,a,b} \sum_{k=1}^{n} T_{x,k} \Big(-kp + (ap-b-kp) \sum_{j=1}^{k} \binom{k}{j} (-p)^j \Big).\end{aligned} \tag{4.44}$$

Define the polynomial $\tilde{P}$ with at most degree n by

$$\begin{aligned}\tilde{P}(p) := \sum_{\substack{a,b=1 \\ a \geq b}}^{n} \tilde{N}_{x,a,b} \Bigg[&(b-1) \sum_{k=1}^{n} k\, T_{x,k} \\ &+ \sum_{k=1}^{n} T_{x,k} \Big(b \sum_{j=1}^{k-1} \binom{k}{j+1} (-p)^j + (a-k) \sum_{j=1}^{k} \binom{k}{j} (-p)^j \Big) \Bigg].\end{aligned}$$

Using (4.44) and the definition of $\tilde{P}$, we get

$$\Phi'_x(p) = \Big(p(1-p)\sum_{k=1}^{n} T_{x,k}\, q(k,p))\Big)^{-1} \tilde{P}(p). \tag{4.45}$$

So, all roots of polynomial $\tilde{P}$ in $(0,1)$ are the critical values of Φ_x. Using the fact

$$\binom{k}{j+1} = \binom{k}{j}\frac{k-j}{j+1},$$

and changing the order of summation, we obtain for the polynomial

$$\begin{aligned}\tilde{P}(p) = \sum_{\substack{a,b=1\\ a\geq b}}^{n} \tilde{N}_{x,a,b}\Bigg[&(b-1)\sum_{k=1}^{n} k\,T_{x,k}\\ &+(-p)^n\,T_{x,n}(a-n) + \sum_{j=1}^{n-1} p^j(-1)^j \sum_{k=j}^{n} T_{x,k}\binom{k}{j}\left(\frac{k-j}{j+1}b + a - k\right)\Bigg].\end{aligned}$$

Therefore the coefficients of $\tilde{P}$ equal the coefficients of P_x in (4.28). Thus the critical values of Φ_x are the roots of the polynomial P_x, that belong to the interval $(0,1)$.

If $\tilde{N}_{x,a,b} = 0$ for all $a, b \in F$ with $a \geq b$, or if only $\tilde{N}_{x,1,1} > 0$ and $T_{x,a} = 0$ for all $a \in \{2,\dots,n\}$, then the function Φ_x is constant. In this case each point is critical and a maximum. In the following, we assume the existence of a pair $a, b \in F$ with $a \geq b$, such that $\tilde{N}_{x,a,b} > 0$. If $a \geq 2$ then there exists $k \in \{2,\dots,n\}$, such that $T_{x,k} > 0$. If $a = 1$, we assume existence of such k, since otherwise Φ is constant. Next, we derive the second order derivative of Φ_x for $p \in (0,1)$

$$\begin{aligned}\Phi''_x(p) = -\sum_{\substack{a,b=1\\ a\geq b}}^{n} \tilde{N}_{x,a,b}\Bigg[&\frac{b-1}{p^2} + \frac{a-b}{(1-p)^2} + \sum_{k=1}^{n} T_{x,k}\frac{\partial^2 q}{\partial p^2}(k,p)\Big(\sum_{j=1}^{n} T_{x,j}\, q(j,p)\Big)^{-1}\\ &-\Big(\sum_{k=1}^{n} T_{x,k}\frac{\partial q}{\partial p}(k,p)\Big)^2\Big(\sum_{j=1}^{n} T_{x,j}\, q(j,p)\Big)^{-2}\Bigg].\end{aligned} \tag{4.46}$$

To apply Lemma 4.41, define for $j \in \{0,\dots,n-1\}$ the sequence

$$a_j := \sum_{k=j+1}^{n} T_{x,k}.$$

It is obvious that $a_0 \geq a_1 \geq a_2 \geq \dots \geq a_{n-1} \geq 0$. Since we assumed the existence of $k \in \{2,\dots,n\}$ with $T_{x,k} > 0$, we obtain $a_0 a_1 > 0$. Hence

$$\tilde{n} = \max\big\{j \in \{1,\dots,n-1\} : a_j > 0\big\}$$

is well-defined and $a_j = 0$ for all $j \in \{\tilde{n}+1,\dots,n-1\}$. Define the function $g : [0,1] \to (0,\infty)$ by

$$g(y) = \sum_{k=1}^{n} T_{x,k}\sum_{j=0}^{k-1} y^j = \sum_{j=0}^{\tilde{n}} a_j\, y^j. \tag{4.47}$$

Using the definition of the function q in (2.35), we obtain

$$g(1-p) = \sum_{k=1}^{n} T_{x,k}\, q(k,p). \tag{4.48}$$

Therefore we get for the second order derivative in (4.46)

$$\Phi_x''(p) = -\sum_{\substack{a,b=1\\ a\geq b}}^{n} \tilde{N}_{x,a,b}\Big(\frac{b-1}{p^2} + \frac{a-b}{(1-p)^2} + \frac{g''(1-p)}{g(1-p)} - \Big(\frac{g'(1-p)}{g(1-p)}\Big)^2\Big).$$

Using that the first order derivative of Φ_x in (4.43) equals zero at the critical points $p_c \in (0,1)$, we obtain

$$\begin{aligned}\Phi_x''(p_c) = &-\sum_{\substack{a,b=1\\ a\geq b}}^{n} \tilde{N}_{x,a,b}\Big(\frac{b-1}{p_c^2} + \frac{b-1}{p_c(1-p_c)}\Big)\\ &-\sum_{\substack{a,b=1\\ a\geq b}}^{n} \tilde{N}_{x,a,b}\left(\frac{1}{1-p_c}\frac{g'(1-p_c)}{g(1-p_c)} + \frac{g''(1-p_c)}{g(1-p_c)} - \left(\frac{g'(1-p_c)}{g(1-p_c)}\right)^2\right).\end{aligned} \tag{4.49}$$

The first summand is smaller than zero, since $\tilde{N}_{x,a,b} \geq 0$ for all $a,b \in F$ with $a \geq b$. Applying Lemma 4.41 and since we assumed that there exists $a,b \in F$ with $\tilde{N}_{x,a,b} > 0$, we see that the second summand is strictly smaller than zero. Therefore each critical point is a maximum of Φ_x. □

If Φ_x has no critical point, the maximum is attained at the boundary. Depending on the observed rating changes and the total time the next lemma shows, which p maximizes the function Φ_x.

Lemma 4.50. *Fix $x \in S$ with $T_{x,a} > 0$ for any $a \in F$. Define the function Φ_x by (4.38) and the polynomial P_x with at most degree n with coefficients c_j for $j \in \{1,\dots,n\}$ by (4.28). Suppose $l \in \{0,\dots,n\}$ is given by $l = \max\big\{j \in \{0,\dots,n\} : c_i = 0 \text{ for all } 0 \leq i \leq j-1\big\}$. Then the following holds:*

- *(i) The root of P_x in $(0,1)$ maximizes the function Φ_x uniquely, if there exists $a,b \in F$ with $a > b$ such that $\tilde{N}_{x,a,b} > 0$ and either there exists $a,b \in F$ with $a \geq b \geq 2$ such that $\tilde{N}_{x,a,b} > 0$ or $c_l > 0$.*
- *(ii) Suppose $\tilde{N}_{x,a,b} = 0$ for all $a,b \in F$ with $a \geq b \geq 2$ and $N_{x,a,1} > 0$ for any $a \in \{2,\dots,n\}$. Furthermore assume $c_l < 0$. Then the unique maximum of Φ_x is attained at $p = 0$.*
- *(iii) Assume $\tilde{N}_{x,a,b} = 0$ for all $a,b \in F$ with $a > b$. If there exists $a \geq 2$ with $\tilde{N}_{x,a,a} > 0$ or $\tilde{N}_{x,1,1} > 0$ and $T_{x,a} > 0$ for any $a \in \{2,\dots,n\}$, then the unique maximum of Φ_x is attained at $p = 1$.*

Proof. (i) Assume there exists $a,b \in F$ with $a > b$ such that $\tilde{N}_{x,a,b} > 0$. Lemma 4.40 implies Φ_x is not constant. Since each critical point is a maximum and Φ_x is not constant, there can be at most one critical point. Therefore if a root of P_x exists in $(0,1)$, then this root is unique and it is the unique maximum of Φ_x with Lemma 4.42. Furthermore the polynomial P_x is not constant zero which means $c_l \neq 0$. To show the existence of a root of P_x, we see

$$\begin{aligned} P_x(1) = \sum_{\substack{a,b=1\\ a\geq b}}^{n} \tilde{N}_{x,a,b} \sum_{k=1}^{n} T_{x,k} \Bigg[b\Bigg(k - \sum_{j=1}^{k-1} (-1)^{j+1} \binom{k}{j+1}\Bigg)\Bigg) \\ + a \sum_{j=1}^{k} (-1)^j \binom{k}{j} - k - k \sum_{j=1}^{k} (-1)^j \binom{k}{j} \Bigg]. \end{aligned}$$

Since the binomial theorem implies

$$\sum_{j=0}^{k} (-1)^j \binom{k}{j} = 0,$$

the polynomial, evaluated at one, simplifies to

$$P_x(1) = \sum_{\substack{a,b=1\\ a\geq b}}^{n} (b-a) \tilde{N}_{x,a,b} \sum_{k=1}^{n} T_{x,k}. \tag{4.51}$$

Since there exists a pair $a, b \in F$ with $a > b$, where $\tilde{N}_{x,a,b} > 0$, it is $P_x(1) < 0$. Furthermore, the value of P_x at zero is

$$P_x(0) = c_0 = \sum_{\substack{a,b=1\\ a\geq b}}^{n} (b-1) \tilde{N}_{x,a,b} \sum_{k=1}^{n} k\, T_{x,k}. \tag{4.52}$$

If there exists $a, b \in \{2, \dots, n\}$ with $a \geq b$ such that $\tilde{N}_{x,a,b} > 0$, then this is strictly greater than zero. The continuity of P_x and the change of sign in $(0, 1)$ imply that the polynomial has a root in $(0, 1)$.

If $\tilde{N}_{x,a,b} = 0$ for all $a, b \in \{2, \dots, n\}$, then $P_x^{(j)}(0) = j!\, c_j = 0$ for all $j \in \{0, \dots, l-1\}$, where $l \geq 1$, because $c_0 = 0$. Using the Taylor formula it exists $\delta > 0$, such that $P_x(\varepsilon)$ has the same sign as $P_x^{(l)}(0)$ for all $\varepsilon \in (0, \delta)$. Therefore if $c_l > 0$, then $P_x(\varepsilon) > 0$. It follows, P_x has a root in $(0, 1)$, since P_x is continuous and $P_x(1) < 0$.

(ii) Suppose $\tilde{N}_{x,a,b} = 0$ for all $a, b \in F$ with $a \geq b \geq 2$ and $\tilde{N}_{x,a,1} > 0$ for any $a \geq 2$. Then analogously to Part (i) we obtain $P_x(1) < 0$, $P_x(0) = 0$ and there exists $\delta > 0$ such that $P_x(\varepsilon) < 0$ for all $\varepsilon \in (0, \delta)$. Assume there exists a root $p_c \in (0, 1)$ of P_x. Then p_c is a maximal point of Φ_x with Lemma 4.42. Differentiation of Φ'_x in (4.45) and evaluating at p_c yields

$$\Phi''_x(p_c) = P'_x(p_c) \Big(p_c(1-p_c) \sum_{k=1}^{n} T_{x,k} q(k, p_c) \Big)^{-1},$$

since $P_x(p_c) = 0$. Therefore $P'_x(p_c) < 0$, since $\Phi''_x(p_c) < 0$. Thus there exists $\tilde{\delta} > 0$ such that $P_x(p_c - \tilde{\varepsilon}) > 0$ for all $\tilde{\varepsilon} \in (0, \tilde{\delta})$. The sign of P_x changes in $(\delta, p_c - \tilde{\delta})$ and P_x has another root in $(0, 1)$. This is a contradiction, since there is at most one root of the polynomial P_x in $(0, 1)$. Hence, there is no root of the polynomial in $(0, 1)$ and Φ_x is maximized for $p \in \{0, 1\}$. The definition of Φ_x in (4.38) implies $\Phi_x(1) = -\infty$, but $\Phi_x(0) > -\infty$, if $\tilde{N}_{x,a,b} = 0$ for all $a, b \in F$ with $a \geq b \geq 2$ and $\tilde{N}_{x,a,1} > 0$ for any $a \geq 2$. Altogether Φ_x is maximized at zero.

(iii) Suppose $\tilde{N}_{x,a,b} = 0$ for all $a, b \in F$ with $a > b$. Using (4.43), the derivative of Φ_x is for all $p \in (0,1)$ given by

$$\Phi_x'(p) = \sum_{a=1}^{n} \tilde{N}_{x,a,a} \left(\frac{a-1}{p} - \frac{\sum_{k=1}^{n} T_{x,k} \frac{\partial q}{\partial p}(k,p)}{\sum_{j=1}^{n} T_{x,j}\, q(j,p)} \right),$$

With definition of q in (2.35) we get

$$\sum_{k=1}^{n} T_{x,k} \frac{\partial q}{\partial p}(k,p) = -\sum_{k=2}^{n} T_{x,k} \sum_{j=1}^{k-1} j(1-p)^{j-1} \leq 0. \tag{4.53}$$

If there exists $a \geq 2$, such that $\tilde{N}_{x,a,a} > 0$, then $\Phi_x'(p) > 0$ for $p \in (0,1)$ and $p = 1$ maximizes the continuous function Φ_x.

Suppose $\tilde{N}_{x,a,b} = 0$ for all $a, b \in F$ with $a \geq 2$. Using (4.43) and (4.53) it follows for all $p \in (0,1)$

$$\Phi_x'(p) = -\tilde{N}_{x,1,1} \frac{\sum_{k=1}^{n} T_{x,k} \frac{\partial q}{\partial p}(k,p)}{\sum_{j=1}^{n} T_{x,j}\, q(j,p)} > 0.$$

If there exists $k \geq 2$ with $T_{x,k} > 0$ and $\tilde{N}_{x,1,1} > 0$, then $p = 1$ maximizes Φ_x. □

Altogether, we sum up the results of this subsection to prove Theorem 4.27.

Proof of Theorem 4.27. The maximum likelihood estimators are the maxima of the function $\log \mathcal{L}^{c} : \Theta_c \to \mathbb{R} \cup \{-\infty\}$ given by

$$\log \mathcal{L}^{c}(\theta) = \sum_{k=1}^{m} \log \nu_{n_k}^{c}\big(\{z_0(\omega_k)\}\big) + \sum_{x \in S} L_x(\theta_x),$$

where $\theta = (\theta_x)_{x \in S}$ with $\theta_x = (\mu_{x,1}, \dots, \mu_{x,x-1}, \mu_{x,x+1}, \dots, \mu_{x,K}, p_x)$ and $L_x : [0,\infty)^{K-1} \times [0,1] \to \mathbb{R} \cup \{-\infty\}$ is given by (4.34).

If $T_{x,a} = 0$ for all $a \in F$, then $N_{x,y,a,b} = 0$ for all $y \in S \setminus \{x\}$ and $a, b \in F$ with $a \geq b$. Therefore $L_x \equiv 0$ and each $\theta_x \in [0,\infty)^{K-1} \times [0,1]$ maximizes L_x, that shows Part (i) of Theorem 4.27.

Assume there exists $a \in F$ such that $T_{x,a} > 0$. By Lemma (4.36) the maxima $\hat{\theta}_x$ of L_x are in the set

$$\hat{\theta}_x \in \Big\{ \big(\varphi_{x,1}(p), \dots, \varphi_{x,x-1}(p), \varphi_{x,x+1}(p), \dots, \varphi_{x,K}(p), p\big) : p \in [0,1] \Big\},$$

where $\varphi_{x,y}$ is given by (4.37). Therefore it is sufficient to maximize the function $\Phi_x : [0,1] \to \mathbb{R} \cup \{-\infty\}$ given by

$$\Phi_x(p) := L_x\left(\varphi_{x,1}(p), \dots, \varphi_{x,x-1}(p), \varphi_{x,x+1}(p), \dots, \varphi_{x,K}(p), p\right),$$

to find the maxima of L_x. By Lemma 4.40 and Lemma 4.50 the maxima of Φ are given by Part (ii) - (v) of Theorem 4.27. Furthermore, for each $x \in S$ the function L_x is finite, since Φ is either uniquely maximized by Lemma 4.50, or constant by Lemma 4.40, but greater than $-\infty$. Therefore if $\hat{\theta} = (\hat{\theta}_x)_{x \in S}$ maximizes $\log \mathcal{L}^{c}$, then $\hat{\theta}_x$ maximizes L_x for each $x \in S$, which concludes the proof. □

Chapter 5

Asymptotic properties of the estimator

In Chapter 4 we proved the maximum likelihood estimators for the extended strongly coupled random walk process. In the following we show consistency and asymptotic normality of the estimator. For proving consistency we maximize a function which is similar to the log-likelihood function in (4.33). Hence, we state the maximization result in the following theorem for slightly more general functions.

Theorem 5.1. *Let $n \in \mathbb{N}$ be a natural number and define the set $F = \{1, \ldots, n\}$. For each $a, b \in F$ with $a \geq b$ and $x, y \in S$ with $x \neq y$ define constants $C^{\mathrm{N}}_{x,y,a,b} \in [0, \infty)$ and $C^{\mathrm{T}}_{x,a} \in [0, \infty)$. Suppose $\tilde{C}^{\mathrm{N}}_{x,a,b} = \sum_{y \in S\setminus\{x\}} C^{\mathrm{N}}_{x,y,a,b}$ for each $a, b \in F$ with $a \geq b$ and $x \in S$. Furthermore we assume that if there exists $b \in F$ with $a \geq b$ such that $\tilde{C}^{\mathrm{N}}_{x,a,b} > 0$, then $C^{\mathrm{T}}_{x,a} > 0$ for $a \in F$ and $x \in S$. For each $x \in S$ we define the constants*

$$c_0(x) = \sum_{\substack{a,b=1 \\ a \geq b}}^{n} (b-1)\tilde{C}^{\mathrm{N}}_{x,a,b} \sum_{k=1}^{n} k\, C^{\mathrm{T}}_{x,k}$$

and for $j \in \{1, \ldots, n\}$

$$c_j(x) = (-1)^j \sum_{\substack{a,b=1 \\ a \geq b}}^{n} \tilde{C}^{\mathrm{N}}_{x,a,b} \sum_{k=j}^{n} \binom{k}{j} \Big(\frac{k-j}{j+1}\, b + a - k \Big)\, C^{\mathrm{T}}_{x,k}.$$

Define the index $l_x \in \{0, \ldots, n\}$ by $l_x = \max\big\{ j \in \{0, \ldots, n\} :\ c_i(x) = 0, \text{ for } 0 \leq i \leq j-1 \big\}$. For each $x \in S$ let $g_x : [0, \infty)^{K-1} \times [0, 1] \to \mathbb{R} \cup \{-\infty\}$ be a function with

$$\begin{aligned} g_x(\theta_x) = &\sum_{\substack{y \in S \\ x \neq y}} \sum_{\substack{a,b=1 \\ a \geq b}}^{n} C^{\mathrm{N}}_{x,y,a,b} \big[\log \mu_{xy} + (b-1)\log p_x + (a-b)\log(1-p_x) \big] \\ &- \mu_x \sum_{a=1}^{n} C^{\mathrm{T}}_{x,a}\, q(a, p_x), \end{aligned}$$

where $\theta_x = (\mu_{x,1}, \ldots, \mu_{x,x-1}, \mu_{x,x+1}, \ldots, \mu_{x,K}, p_x)$. If there exists $a \in F$ with $C^{T}_{x,a} > 0$, then define the function $\hat{\theta}_x : [0,1] \to [0, \infty)^{K-1} \times [0,1]$ by

$$\hat{\theta}_x(p) = (\varphi_{x,1}(p), \ldots, \varphi_{x,x-1}(p), \varphi_{x,x+1}(p), \ldots, \varphi_{x,K}(p), p),$$

where for each $y \in S$ with $x \neq y$ the function $\varphi_{xy} : [0,1] \to [0,\infty)$ is given by

$$\varphi_{xy}(p) := \sum_{\substack{a,b=1 \\ a \geq b}}^{n} C^{\mathrm{N}}_{x,y,a,b} \Big(\sum_{k=1}^{n} C^{\mathrm{T}}_{x,k}\, q(k,p) \Big)^{-1}.$$

Then the following holds:

(i) The function g_x is finite at the maximal points.

(ii) Suppose there exists $a, b \in F$ with $a > b$ such that $\tilde{C}^{\mathrm{N}}_{x,a,b} > 0$ and $a, b \in F$ with $a \geq b \geq 2$ such that $\tilde{C}^{\mathrm{N}}_{x,a,b} > 0$. Then $g_x(\hat{\theta}_x(\cdot)) : [0,1] \to \mathbb{R}$ has a unique critical point $\hat{p}$ in $(0,1)$, which is the unique maximum. Furthermore $\hat{\theta}_x(\hat{p})$ maximizes the function g_x uniquely.

(iii) Suppose there exists $a \in F$ with $a \geq 2$ such that $\tilde{C}^{\mathrm{N}}_{x,a,1} > 0$ and $\tilde{C}^{\mathrm{N}}_{x,a,b} = 0$ for all $a, b \in F$ with $a \geq b \geq 2$. Then g_x is uniquely maximized at $\hat{\theta}_x(\hat{p})$ for $\hat{p} \in [0,1)$. If $c_{l_x} < 0$, then $\hat{p} = 0$.

(iv) Suppose there exists $a \in F$ with $a \geq 2$ such that $\tilde{C}^{\mathrm{N}}_{x,a,a} > 0$ and $\tilde{C}^{\mathrm{N}}_{x,a,b} = 0$ for all $a, b \in F$ with $a > b$. Then the unique maximum of g_x is at $\hat{\theta}_x(1)$.

Proof. (i) Obviously. (ii)-(iv) Apply Remark 4.35, Lemma 4.42 and Lemma 4.50. □

5.1 Consistency of the maximum likelihood estimator

In this section we show the consistency of the likelihood estimators. We assume that the observed paths, given by Definition 4.20, follow a true distribution $\mathcal{P}_{\theta_0}$, which is unknown and estimate the true parameter $\theta_0 \in \Theta_c = \big([0,\infty)^{K-1} \times [0,1]\big)^K$. Then the estimated parameters converge almost surely to θ_0, if the number of observed sample paths increases.

For each $\theta \in \Theta_c$ let η be a random observed path in the esc–process, where the space of the sample path is given by $\mathcal{O} = \mathbb{N} \times (0,\infty) \times \mathcal{S}$, defined in (4.1), and density h^{c}_{θ}, given by Definition 4.20. For each $n \in \mathbb{N}$ and $T \in (0,\infty)$ let $\omega_{n,T}$ be a random sample path in $\mathcal{S}_n$ with density $f^{\mathrm{c}}_n(\cdot, T)$. By Lemma 4.19 the random path $\omega_{n,T}$ has the distribution of a sample path of the Markov process $(X^{\mathrm{c}}_{t,n})_{t \geq 0}$ until time T. Denote by $\mathbb{E}_{\theta}[\cdot]$ the expectation, where the paths η, resp. $\omega_{n,T}$, have density h^{c}_{θ}, resp. $f^{\mathrm{c}}_n(\cdot, T; \theta)$ for each $n \in \mathbb{N}$ and $T > 0$.

In this chapter we always assume the following.

Assumption 5.2. *We assume the probability measure P_N is concentrated on $\{1, \ldots, \tilde{N}\}$ with $\tilde{N} \in \mathbb{N}$, i. e. $P_N(n) = 0$ for $n > \tilde{N}$. Furthermore the probability measure ξ satisfies*

$$\int_0^{\infty} T\, \xi(dT) < \infty. \tag{5.3}$$

Remark 5.4. For real observations this assumption is a weak restriction. We assume that the size of the industries is bounded by $\tilde{N}$. Furthermore a random variable with distribution ξ has a finite first moment.

Lemma 5.5. *Let $F : \mathbb{N} \times (0,T) \times \mathcal{S} \to \mathbb{R}$ be a $\mathcal{P}(\mathbb{N}) \otimes \mathcal{B}\big((0,\infty)\big) \otimes \mathcal{C}$-measurable function and be non-negative or $F(\eta)$ is integrable. Then for all $\theta \in \Theta_c$ holds*

$$\mathbb{E}_{\theta}\big[F(\eta)\big] = \sum_{n=1}^{\infty} P_N(n) \int_0^{\infty} \mathbb{E}_{\theta}\big[F(n, T, \omega_{n,T})\big]\, \xi(dT).$$

Proof. Using Fubini's Theorem and the definition of the density h_θ^{c} in Definition 4.20 we obtain

$$\begin{aligned}\mathbb{E}_\theta\big[F(\eta)\big] &= \sum_{n=1}^{\infty} P_N(n) \int_0^\infty \int_{\mathcal{S}} F(n,T,\omega) f_n^{\mathrm{c}}(\omega,T;\theta)\,\sigma(d\omega)\,\xi(dT)\\ &= \sum_{n=1}^{\infty} P_N(n) \int_0^\infty \int_{\mathcal{S}_n} F(n,T,\omega) f_n^{\mathrm{c}}(\omega,T;\theta)\,\sigma_n(d\omega)\,\xi(dT).\end{aligned}$$

□

Next, we show a connection between the expected number of transitions and the expected waiting time, and afterwards, the finiteness of these two.

Lemma 5.6. *Let θ be in the parameter set Θ_c, $n \in \mathbb{N}$ be the number of firms and $T > 0$ be the length of the path. Then for each $x, y \in S$ with $x \neq y$ and $a, b \in \mathbb{N}$ with $a \geq b$*

$$\mathbb{E}_\theta\Big[N_{x,y,a,b}^{(\omega_{n,T})}\Big] = \binom{a}{b} \mu_{xy} p_x^{b-1}(1-p_x)^{a-b}\,\mathbb{E}_\theta\Big[T_{x,a}^{(\omega_{n,T},T)}\Big]. \tag{5.7}$$

For each $x_i, y_i \in S$ with $x_i \neq y_i$ and $a_i, b_i \in \mathbb{N}$ with $a_i \geq b_i$ for $i = 1, 2$ define the function $S_i : \mathcal{S} \times (0,\infty) \to \mathbb{R}$ by

$$S_i(\omega,T) = N_{x_i,y_i,a_i,b_i}^{(\omega)} - \binom{a_i}{b_i} \mu_{x_i y_i} p_{x_i}^{b_i-1}(1-p_{x_i})^{a_i-b_i}\, T_{x_i,a_i}^{(\omega,T)}, \tag{5.8}$$

for (ω,T) strongly feasible, defined in 4.4, and $S_i(\omega,T) = 0$ otherwise. Then we obtain

$$\mathbb{E}_\theta\big[S_1(\omega_{n,T},T)S_2(\omega_{n,T},T)\big] = \begin{cases} \mathbb{E}_\theta\Big[N_{x_1,y_1,a_1,b_1}^{(\omega_{n,T})}\Big], & \text{if } x_1 = x_2,\ y_1 = y_2,\\ & \quad a_1 = a_2 \text{ and } b_1 = b_2,\\ 0, & \text{otherwise.}\end{cases}$$

Proof. Suppose $x_i, y_i \in S$ with $x_i \neq y_i$ and $a_i, b_i \in \mathbb{N}$ with $a_i \geq b_i$ for $i = 1, 2$. Let

$$\omega = ((z_0,t_0),\ldots,(z_{l-1},t_{l-1}),z_l)$$

be an arbitrary path in $\mathcal{S}_n$, where $l \in \mathbb{N}_0$, $z_i \in S^n$ for all $i \in \{0,\ldots,l\}$ and $t_i \in (0,T)$ for all $i \in \{0,\ldots,l-1\}$, such that (ω,T) is strongly feasible. Define the number of transitions $N_{z\tilde{z}}(\omega)$ from $z \in S^n$ to $\tilde{z} \in S^n$ with $z \neq \tilde{z}$ by

$$N_{z\tilde{z}}(\omega) = \#\{0 \leq k \leq l-1 : z_k = z,\ z_{k+1} = \tilde{z}\}$$

and the total time $T_z(\omega,T)$ of ω spending in $z \in S^n$ until T by

$$T_z(\omega,T) = \sum_{k=0}^{l-1} t_k \mathbb{1}_{\{z_k = z\}} + (T - \sum_{k=0}^{l-1} t_k)\mathbb{1}_{\{z_l = z\}}. \tag{5.9}$$

Albert shows in Theorem 5.1, that for the Markov jump process X_n^{c} with Q-matrix $Q_n^{\mathrm{c},\theta}$ and random sample path $\omega_{n,T}$

$$\mathbb{E}_\theta\big[N_{z\tilde{z}}(\omega_{n,T})\big] = Q_n^{\mathrm{c},\theta}(z,\tilde{z})\,\mathbb{E}_\theta\big[T_z(\omega_{n,T},T)\big], \quad \text{for each } z,\tilde{z} \in S^n \text{ with } z \neq \tilde{z}. \tag{5.10}$$

Define the set $Z_i \subset S^n$ of states, where a_i firms have rating x_i, by

$$Z_i = \{z \in S^n : a_{x_i}(z) = a_i\}, \qquad \text{for } i = 1, 2.$$

For each $z \in Z_i$ we define the set Z'_i, where b_i firms change the rating from x_i to y_i, by

$$Z'_i(z) = \Big\{\tilde{z} \in S^n : z \to \tilde{z} \text{ strongly feasible}, \\ b_i = \#\big\{j \in \{1, \dots, n\} : z(j) = x_i,\, \tilde{z}(j) = y_i\big\}\Big\}, \qquad \text{for } i = 1, 2.$$

Using this definition and (5.10), we obtain

$$\mathbb{E}_\theta[N^{(\omega_{n,T})}_{x_1,y_1,a_1,b_1}] = \sum_{z \in Z_1} \sum_{\tilde{z} \in Z'_1(z)} \mathbb{E}_\theta[N_{z\tilde{z}}(\omega_{n,T})] = \sum_{z \in Z_1} \sum_{\tilde{z} \in Z'_1(z)} Q^{c,\theta}_n(z, \tilde{z}) \mathbb{E}_\theta[T_z(\omega_{n,T}, T)]. \quad (5.11)$$

The definition of $Q^{c,\theta}_n$ in (2.34) implies

$$\mathbb{E}_\theta[N^{(\omega_{n,T})}_{x_1,y_1,a_1,b_1}] = \binom{a_1}{b_1} \mu_{x_1 y_1} p_{x_1}^{b_1 - 1} (1 - p_{x_1})^{a_1 - b_1} \sum_{z \in Z_1} \mathbb{E}_\theta[T_z(\omega_{n,T}, T)].$$

With the definition of $T^{(\omega,T)}_{x_1,a_1}$ in (4.17) for strongly feasible $(\omega, T) \in \mathcal{S} \times (0, \infty)$, it follows

$$T^{(\omega,T)}_{x_1,a_1} = \sum_{z \in Z_1} T_z(\omega, T), \quad (5.12)$$

which concludes the proof of (5.7).

To show the second equation, Theorem 5.1 of Albert says for each $z_i \in Z_i$ and $\tilde{z}_i \in Z'_i(z_i)$ for $i = 1, 2$

$$\mathbb{E}_\theta\Big[\big(N_{z_1\tilde{z}_1}(\omega_{n,T}) - Q^{c,\theta}_n(z_1, \tilde{z}_1) T_{z_1}(\omega_{n,T}, T)\big)\big(N_{z_2\tilde{z}_2}(\omega_{n,T}) - Q^{c,\theta}_n(z_2, \tilde{z}_2) T_{z_2}(\omega_{n,T}, T)\big)\Big] \\ = \begin{cases} Q^{c,\theta}_n(z_1, \tilde{z}_1) \mathbb{E}[T_{z_1}(\omega_{n,T}, T)], & \text{if } z_1 = z_2,\ \tilde{z}_1 = \tilde{z}_2, \\ 0, & \text{otherwise.} \end{cases} \quad (5.13)$$

Therefore $\mathbb{E}_\theta[S_1 S_2]$ is only non-zero, if $x_1 = x_2$, $y_1 = y_2$, $a_1 = a_2$ and $b_1 = b_2$, since otherwise $z_1 \neq z_2$ or $\tilde{z}_1 \neq \tilde{z}_2$ for all $z_i \in Z_i$ and $\tilde{z}_i \in Z'_i(z_i)$. Using (5.12) and (5.13), we obtain

$$\mathbb{E}_\theta\big[S_1^2(\omega_{n,T}, T)\big] = \sum_{z_1 \in Z_1} \sum_{\tilde{z}_1 \in Z'_1(z_1)} \mathbb{E}_\theta\Big[\big(N_{z_1,\tilde{z}_1}(\omega_{n,T}) - Q^{c,\theta}_n(z_1, \tilde{z}_1) T_{z_1}(\omega_{n,T}, T)\big)^2\Big] \\ = \binom{a_1}{b_1} \mu_{x_1 y_1} p_{x_1}^{b_1 - 1} (1 - p_{x_1})^{a_1 - b_1} \mathbb{E}_\theta\Big[T^{(\omega_{n,T},T)}_{x_1,a_1}\Big].$$

□

Corollary 5.14. *Let θ be in the parameter set Θ_c. Then for each $x, y \in S$ with $x \neq y$ and $a, b \in \mathbb{N}$ with $a \geq b$*

$$\mathbb{E}_\theta\Big[N^{(\eta)}_{x,y,a,b}\Big] = \binom{a}{b} \mu_{xy} p_x^{b-1} (1 - p_x)^{a-b}\, \mathbb{E}_\theta\Big[T^{(\eta)}_{x,a}\Big]. \quad (5.15)$$

For each $x_i, y_i \in S$ with $x_i \neq y_i$ and $a_i, b_i \in \mathbb{N}$ with $a_i \geq b_i$ for $i = 1, 2$ we obtain

$$\mathbb{E}_\theta\big[S_1(\eta)S_2(\eta)\big] = \begin{cases} \mathbb{E}_\theta\Big[N^{(\eta)}_{x_1,y_1,a_1,b_1}\Big], & \text{if } x_1 = x_2,\ y_1 = y_2, \\ & \quad a_1 = a_2 \text{ and } b_1 = b_2, \\ 0, & \text{otherwise,} \end{cases} \tag{5.16}$$

where S_i is defined in (5.8).

Proof. Apply Lemma 5.5 and 5.6. □

Remark 5.17. We use e. g. $\mathbb{E}_\theta\Big[N^{(\eta)}_{x,y,a,b}\Big]$ as abbreviation of $\mathbb{E}_\theta\Big[N^{\pi(\eta)}_{x,y,a,b}\Big]$, where

$$\pi : \mathbb{N} \times (0,\infty) \times \mathcal{S} \to \mathcal{S}, \text{ with } \pi(n,T,\omega) = \omega$$

is the projection in the third coordinate.

Lemma 5.18. *Suppose the assumptions in 5.2 holds. Let θ_0 be in the parameter set Θ_c. Then we obtain for the random sample path η with density $h^{\mathrm{c}}_{\theta_0}$ the following.*

(i) $\mathbb{E}_{\theta_0}\big[T^{(\eta)}_{x,a}\big] < \infty$, *for all $x \in S$ and $a \in \mathbb{N}$,*

(ii) $\mathbb{E}_{\theta_0}[N^{(\eta)}_{x,y,a,b}] < \infty$, *for all $x, y \in S$ with $x \neq y$ and $a, b \in \mathbb{N}$ with $a \geq b$.*

(iii) $\mathbb{E}_{\theta_0}\big[\,|\log h^{\mathrm{c}}_\theta(\eta)|\,\big] < \infty$, *for all $\theta \in \Theta_c$.*

Proof. (i) The waiting time $T^{(\omega,T)}_{x,a}$ is for all $(\omega, T) \in \mathcal{S} \times (0,\infty)$ bounded by the total observation time T. With Lemma 5.5 we see

$$\mathbb{E}_{\theta_0}\big[T^{(\eta)}_{x,a}\big] = \sum_{n=1}^{\infty} P_N(n) \int_0^\infty \mathbb{E}_{\theta_0}\big[T^{(\omega_{n,T},T)}_{x,a}\big]\,\xi(dT) \leq \int_0^\infty T\,\xi(dT) < \infty.$$

(ii) Apply Corollary 5.14 and Part (i).

(iii) Since the sets $\mathcal{S}_n$ are pairwise disjoint for $n \in \mathbb{N}$, we estimate with the definition of the density h^{c}_θ in Definition 4.20 and the definition of f^{c}_n in Lemma 4.19

$$\begin{aligned}
\mathbb{E}_{\theta_0}\big[\,|\log h^{\mathrm{c}}_\theta(\eta)|\,\big] &= \sum_{n=1}^{\tilde N} P_N(n) \int_0^\infty \int_{\mathcal{S}_n} |\log\left(f^{\mathrm{c}}_n(\omega,T;\theta)\right)|\, f^{\mathrm{c}}_n(\omega,T;\theta_0)\,\sigma_n(d\omega)\,\xi(dT) \\
&\leq \sum_{n=1}^{\tilde N} P_N(n) \int_0^\infty \int_{\mathcal{S}_n} \big|\log \nu^{\mathrm{c}}_n\big(\{z_0(\omega)\}\big)\big|\, f^{\mathrm{c}}_n(\omega,T;\theta_0)\,\sigma_n(d\omega)\,\xi(dT) \\
&\quad + \sum_{x\in S} \mu_x \sum_{a=1}^{\tilde N} q(a,p_x)\mathbb{E}_{\theta_0}\big[T^{(\eta)}_{x,a}\big] + \sum_{\substack{x,y\in S\\ x\neq y}} \sum_{\substack{a,b=1\\ a\geq b}}^{\tilde N} \Big|\log\Big(\mu_{xy} p_x^{b-1}(1-p_x)^{a-b}\Big)\Big|\,\mathbb{E}_{\theta_0}\big[N^{(\eta)}_{x,y,a,b}\big].
\end{aligned}$$

The second and third summands are finite, since $\mathbb{E}_{\theta_0}[T^{(\eta)}_{x,a}]$ and $\mathbb{E}_{\theta_0}[N^{(\eta)}_{x,y,a,b}]$ are finite. For the first summand we show that the integrand is bounded. Define for each $n \in \mathbb{N}$ the set $\tilde S := \{z \in S^n : \nu^{\mathrm{c}}_n(\{z\}) > 0\}$. Then for each $\omega \in \mathcal{S}_n$ with $f^{\mathrm{c}}_n(\omega,T;\theta_0) > 0$, the first entry $z_0(\omega)$ of ω is in $\tilde S$. Therefore there exists a constant $C_n \in [0,\infty)$ for each $n \in \mathbb{N}$, such that for all $T > 0$ and $\omega \in \mathcal{S}_n$ with $f^{\mathrm{c}}_n(\omega,T;\theta_0) > 0$

$$\big|\log \nu^{\mathrm{c}}_n(\{z_0(\omega)\})\big| \leq \max_{z\in\tilde S} \big|\log \nu^{\mathrm{c}}_n(\{z\})\big| =: C_n < \infty.$$

□

To show asymptotic consistency, we first show, which parameters $\theta \in \Theta_c$ identify the law of the sample paths.

Lemma 5.19. *Suppose the assumption in 5.2 holds. Define the subset $\Theta_{id} \subset \Theta_c$ of the parameters by*

$$\Theta_{id} = \Big\{ \theta = (\theta_x)_{x \in S} \in \Theta_c : \sum_{a=2}^{\tilde{N}} \mathbb{E}_\theta[T^{(\eta)}_{x,a}] > 0, \text{ and } \mu_x = \sum_{y \in S \setminus \{x\}} \mu_{xy} > 0, \text{ for all } x \in S,$$
$$\text{where } \theta_x = (\mu_{x,1}, \dots, \mu_{x,x-1}, \mu_{x,x+1}, \dots, \mu_{x,K}, p_x) \Big\}. \tag{5.20}$$

Then the parameters in Θ_{id} are identifiable, i. e., if $\mathcal{P}_\theta = \mathcal{P}_{\tilde{\theta}}$ for $\theta, \tilde{\theta} \in \Theta_{id}$, then $\theta = \tilde{\theta}$, where $\mathcal{P}_\theta$, resp. $\mathcal{P}_{\tilde{\theta}}$, is the distribution of the random sample path η with density h^{c}_θ, resp. $h^{\mathrm{c}}_{\tilde{\theta}}$.

Proof. Suppose $\theta = (\theta_x)_{x \in S} \in \Theta_{id}$ with $\theta_x = (\mu_{x,1}, \dots, \mu_{x,x-1}, \mu_{x,x+1}, \dots, \mu_{x,K}, p_x)$ and $\tilde{\theta} = (\tilde{\theta}_x)_{x \in S} \in \Theta_{id}$ with $\tilde{\theta}_x = (\tilde{\mu}_{x,1}, \dots, \tilde{\mu}_{x,x-1}, \tilde{\mu}_{x,x+1}, \dots, \tilde{\mu}_{x,K}, \tilde{p}_x)$, such that $\mathcal{P}_\theta = \mathcal{P}_{\tilde{\theta}}$. Let x be in S arbitrary. Since $\theta \in \Theta_{id}$ and $T^{(\eta)}_{x,a} \geq 0$ for all $a \in \mathbb{N}$, there exists an $a \in \{2, \dots, \tilde{N}\}$ with $\mathbb{E}_\theta[T^{(\eta)}_{x,a}] > 0$. Since $\mathcal{P}_\theta = \mathcal{P}_{\tilde{\theta}}$, we have

$$\mathbb{E}_\theta\big[T^{(\eta)}_{x,a}\big] = \mathbb{E}_{\tilde{\theta}}\big[T^{(\eta)}_{x,a}\big] \quad \text{and} \quad \mathbb{E}_\theta\big[N^{(\eta)}_{x,y,a,b}\big] = \mathbb{E}_{\tilde{\theta}}\big[N^{(\eta)}_{x,y,a,b}\big],$$

for all $y \in S$ with $x \neq y$ and $b \leq a$. These equations and Corollary 5.14 imply for $b = 1$

$$\mu_{xy}(1 - p_x)^{a-1} = \tilde{\mu}_{xy}(1 - \tilde{p}_x)^{a-1}, \quad \text{for all } y \in S \text{ with } x \neq y, \tag{5.21}$$

and for $b = a$

$$\mu_{xy} p_x^{a-1} = \tilde{\mu}_{xy} \tilde{p}_x^{a-1}, \quad \text{for all } y \in S \text{ with } x \neq y. \tag{5.22}$$

First, suppose $p_x = 1$. With (5.21) and using the fact $\tilde{\mu}_x > 0$, we obtain $(1 - \tilde{p}_x)^{a-1} = 0$ and therefore $\tilde{p}_x = 1$. It follows immediately $\mu_{xy} = \tilde{\mu}_{xy}$ for all $y \in S$ using (5.22). Now, suppose $p_x \in [0, 1)$. Rearrangement of (5.21) yields

$$\mu_{xy} = \tilde{\mu}_{xy} \Big(\frac{1 - \tilde{p}_x}{1 - p_x} \Big)^{a-1}, \quad \text{for all } y \in S \text{ with } x \neq y.$$

Substituting this into equation (5.22), implies $p_x = \tilde{p}_x$, since $\tilde{\mu}_x > 0$. With (5.21) it follows $\mu_{xy} = \tilde{\mu}_{xy}$ for all $y \in S$. $\square$

Remark 5.23. If $P_N(1) = 1$, then the identifiable set Θ_{id} is empty. Since there is almost surely only a single firm observable, the law of the sample paths does not depend on the choice of p_x for all $x \in S$. Therefore the estimator for p_x can arbitrarily be chosen and does not converge to the true estimator, if we increase the number of sample path. In this case consistency and asymptotic normality does not hold.

The parameter is identifiable, if the expected time is positive, that more than one firm has rating x, for all $x \in S$. To ensure that, the following lemma shows an intuitive condition on the parameters.

Definition 5.24. *Let x and y be two rating classes in S. We call y* accessible *from x, if there exist $m \in \mathbb{N}$ and $x_0, \dots, x_m \in S$ with $x_0 = x$ and $x_m = y$, such that*

$$\prod_{i=1}^{m} \mu_{x_{i-1}, x_i} > 0.$$

Lemma 5.25. *Let x be an arbitrary rating class in S. Assume $\theta = (\theta_x)_{x \in S} \in \Theta_c$ with $\theta_x = (\mu_{x,1}, \dots, \mu_{x,x-1}, \mu_{x,x+1}, \dots, \mu_{x,K}, p_x)$. Furthermore suppose $P_N(n) > 0$ for $n \in \mathbb{N}$. If there exists $z_0 \in S^n$ with $\nu_n^{\mathrm{c}}(\{z_0\}) > 0$, such that there are two different firms $j_1, j_2 \in \{1, \dots, n\}$, where either $z_0(j_i) = x$ or x is accessible from $z_0(j_i)$ for $i = 1, 2$, then $\sum_{a=2}^{\infty} \mathbb{E}_\theta[T_{x,a}^{(\eta)}] > 0$.*

Proof. Based on the assumptions there exists a sequence of rating classes $y_0, \dots, y_m \in S$ for $m \geq 0$ with $y_0 = z_0(j_1)$ and $y_m = x$, such that

$$\prod_{i=1}^{m} \mu_{y_{i-1} y_i} > 0.$$

Define $z_i \in S^n$ for $i \in \{1, \dots, m\}$ inductively. Set $z_i(j_1) = y_i$ and $z_i(k) = z_{i-1}(k)$ for $k \neq j_1$, if $p_{y_{i-1}} \in [0, 1)$. Otherwise define $z_i(k) = y_i$ for all $k \in \{1, \dots, n\}$ with $z_{i-1}(k) = y_{i-1}$ and $z_i(k) = z_{i-1}(k)$ otherwise.

If $z_m(j_2) = x$, set $m' = m$. Otherwise, there exist $y_m, \dots, y_{m'} \in S$ for $m' > m$ with $y_m = z_m(j_2)$, $y_i \neq x$ for $i \in \{m+1, \dots, m'-1\}$ and $y_{m'} = x$, such that

$$\prod_{i=m+1}^{m'} \mu_{y_{i-1} y_i} > 0.$$

This is possible, since either $z_m(j_2) = z_0(j_2)$ or $z_m(j_2) = y_i$ with $i \in \{1, \dots, m\}$. Define $z_i \in S^n$ for $i \in \{m+1, \dots, m'\}$ inductively, as above. Then $z_{m'}(k) = x$ for at least two $k \in \{1, \dots, n\}$, i.e. $a_x(z_{m'}) \geq 2$.

For each $T > 0$ define the set $C(T) \subset \mathcal{S}_n$ of paths by

$$C(T) = \Big\{ \omega = \big((z_0, t_0), \dots, (z_{m'-1}, t_{m'-1}), z_{m'} \big) : t_i \in (0, T) \text{ and } \sum_{i=0}^{m'-1} t_i < T \Big\}.$$

Then the measure $\sigma_n\big(C(T)\big) > 0$ and the paths (ω, T) with $\omega \in C(T)$ are strongly feasible. Let $b_i = b_{y_i}(z_i, z_{i+1})$ be the number of firms, which change the rating, in the transition z_i to z_{i+1}, and $a_i = a_{y_i}(z_i)$ the number of firms with rating y_i in state z_i for $i \in \{0, \dots, m'\}$. Then the density of the path $\omega_{n,T}$ of the process $(X_{t,n}^{\mathrm{c}})_{t \in [0,T)}$ equals for each $\omega \in C(T)$

$$\begin{aligned} f_n^{\mathrm{c}}(\omega, T; \theta) = \nu_n^{\mathrm{c}}(\{z_0(\omega)\}) \exp \Big\{ -\sum_{\tilde{x} \in S} \mu_{\tilde{x}} \sum_{k=1}^{n} T_{\tilde{x},k}^{(\omega,T)}\, q(k, p_{\tilde{x}}) \Big\} \\ \times \prod_{i=0}^{m'-1} \mu_{y_i y_{i+1}} p_{y_i}^{b_i - 1} (1 - p_{y_i})^{a_i - b_i} > 0, \end{aligned}$$

since $b_i = 1$, if $p_{y_i} = 0$, and $b_i = a_i$, if $p_{y_i} = 1$, for all $i \in \{0, \dots, m'-1\}$. The construction of the set $C(T)$ implies

$$T_{x, a_x(z_{m'})}^{(\omega,T)} \geq T - \sum_{i=0}^{m'-1} t_i > 0, \qquad \text{for all } \omega \in C(T).$$

See Lemma 5.5 to conclude

$$\mathbb{E}_\theta[T_{x, a_x(z_{m'})}^{(\eta)}] \geq P_N(n) \int_0^\infty \int_{C(T)} T_{x, a_x(z_{m'})}^{(\omega,T)}\, f_n^{\mathrm{c}}(\omega, T; \theta)\, \sigma_n(d\omega)\, \xi(dT) > 0.$$

$\square$

To show that the maximum likelihood estimators converge to the true parameter almost surely, we use a modification of [54, Chapter 5, Lemma 5.10], where we replace the convergence in probability by almost sure convergence.

Lemma 5.26. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $I \subset \mathbb{R}$ be a real interval. Let $\psi_n : I \times \Omega \to \mathbb{R}$ be measurable functions for each $n \in \mathbb{N}$ and $\psi : I \to \mathbb{R}$ be a deterministic function, such that $\psi_n(x)$ converges almost surely to $\psi(x)$ for $n \to \infty$ for each $x \in I$. Assume $x \mapsto \psi_n(x, \omega)$ is almost surely continuous and has exactly one root at $x_n(\omega)$ for almost every $\omega \in \Omega$. Let $x_0 \in I$ be a point such that there exists $\delta > 0$ with $x_0 - \delta \in I$ and $x_0 + \delta \in I$ and $\psi(x_0 - \varepsilon) < 0 < \psi(x_0 + \varepsilon)$ for all $\varepsilon \in (0, \delta]$.*

Then $x_n \to x_0$ almost surely for $n \to \infty$.

Proof. Since $x \mapsto \psi_n(x, \omega)$ is almost surely continuous and the root $x_n(\omega)$ is unique, $x_n : \Omega \to I$ is well-defined and measurable. Suppose $\varepsilon \in (0, \delta]$ arbitrary. Since ψ_n converges almost surely to ψ and $\psi(x_0 - \varepsilon) < 0 < \psi(x_0 + \varepsilon)$, there exists $n_0 \in \mathbb{N}$ with $\psi_n(x_0 - \varepsilon, \omega) < 0 < \psi_n(x_0 + \varepsilon, \omega)$ for all $n \geq n_0$ and almost every $\omega \in \Omega$. Using the continuity of $\psi_n(x, \omega)$ in x, it follows for the root x_n that $x_0 - \varepsilon < x_n(\omega) < x_0 + \varepsilon$ for almost every $\omega \in \Omega$. □

Lemma 5.27. *Let $\theta_0 = (\theta_x^0)_{x \in S} \in \Theta_{id}$ with*

$$\theta_x^0 = (\mu_{x,1}^0, \dots, \mu_{x,x-1}^0, \mu_{x,x+1}^0, \dots, \mu_{x,K}^0, p_x^0) \quad \text{for each } x \in S,$$

be the true parameter. Suppose $(n_j, T_j, \omega_j) \in \mathbb{N} \times (0, \infty) \times \mathcal{S}$ for $j \in \mathbb{N}$ are independent realizations of the sample path η with density $h_{\theta_0}^{\mathrm{c}}$ and assume $h_{\theta_0}^{\mathrm{c}}(n_j, T_j, \omega_j) > 0$. Define for each $x \in S$ and $a, b \in \mathbb{N}$ with $a \geq b$

$$M_0(x,a,b) = \sum_{\substack{y \in S \\ x \neq y}} \mathbb{E}_{\theta_0}\big[N_{x,y,a,b}^{(\eta)}\big], \quad \text{and} \quad M_m(x,a,b) = \sum_{\substack{y \in S \\ x \neq y}} \sum_{j=1}^{m} N_{x,y,a,b}^{(\omega_j)}.$$

Then the following holds:

(i) Suppose $p_x^0 \in (0,1)$ for $x \in S$. Then there exists $a \in \{2, \dots, \tilde{N}\}$ and $m_0 \in \mathbb{N}$ such that $M_m(x,a,b) > 0$ a. s. for all $m \geq m_0$ and $M_0(x,a,b) > 0$ for all $b \in \{1, \dots, a\}$.

(ii) Suppose $p_x^0 = 0$ for $x \in S$. Then $M_0(x,a,b) = 0$ and $M_m(x,a,b) = 0$ a. s. for all $a, b \in \mathbb{N}$ with $a \geq b \geq 2$ and $m \in \mathbb{N}$. Furthermore there exists $a \in \{2, \dots, \tilde{N}\}$ and $m_0 \in \mathbb{N}$ such that $M_m(x,a,1) > 0$ a. s. for all $m \geq m_0$ and $M_0(x,a,1) > 0$.

(iii) Suppose $p_x^0 = 1$ for $x \in S$. Then $M_0(x,a,b) = 0$ and $M_m(x,a,b) = 0$ a. s. for all $a, b \in \mathbb{N}$ with $a > b$ and $m \in \mathbb{N}$. Furthermore there exists $a \in \{2, \dots, \tilde{N}\}$ and $m_0 \in \mathbb{N}$ such that $M_m(x,a,a) > 0$ a. s. for all $m \geq m_0$ and $M_0(x,a,a) > 0$.

Proof. Using the connection between the expected number of rating transitions and the waiting times in Corollary 5.14 we obtain for all $x \in S$ and $a, b \in \mathbb{N}$ with $a \geq b$

$$M_0(x,a,b) = \binom{a}{b} \mu_x^0 (p_x^0)^{b-1} (1 - p_x^0)^{a-b} \mathbb{E}_{\theta_0}\big[T_{x,a}^{(\eta)}\big].$$

Suppose $p_x^0 = 0$. Therefore $M_0(x,a,b) = 0$ and $M_m(x,a,b) = 0$ a. s. for all $a, b \in \mathbb{N}$ with $a \geq b \geq 2$ and for all $m \in \mathbb{N}$ using the strong law of large numbers and the non-negativity of $N_{x,y,a,b}^{(\omega)}$. Since $\theta_0 \in \Theta_{id}$, there exists $\tilde{a} \in \mathbb{N}$ with $\tilde{a} \geq 2$ such that $\mathbb{E}_{\theta_0}\big[T_{x,\tilde{a}}^{(\eta)}\big] > 0$. It follows $M_0(x, \tilde{a}, 1) > 0$ and there exists $m_0 \in \mathbb{N}$ such that $M_m(x, \tilde{a}, 1) > 0$ a. s. for all $m \geq m_0$. The remaining follows in an analogous way. □

After the preparing lemmata we show the strong consistency of the maximum likelihood estimators.

Theorem 5.28. *Assume the assumption in 5.2 holds. Suppose the true parameter θ_0 is in the identifiable parameter set Θ_{id}, which is given by* (5.20)*. Let $(n_j, T_j, \omega_j) \in \mathbb{N} \times (0, \infty) \times \mathcal{S}$ for $j \in \mathbb{N}$, be independent realizations of the sample path η with density $h^{\mathrm{c}}_{\theta_0}$, given by Definition 4.20, and assume $h^{\mathrm{c}}_{\theta_0}(n_j, T_j, \omega_j) > 0$. For each $m \in \mathbb{N}$ let $\theta^m = (\theta^m_x)_{x \in S}$ with*

$$\theta^m_x = (\mu^m_{x,1}, \dots, \mu^m_{x,x-1}, \mu^m_{x,x+1}, \dots, \mu^m_{x,K}, p^m_x) \quad \textit{for each } x \in S,$$

be a maximum likelihood estimator for the paths (n_j, T_j, ω_j) for $j \in \{1, \dots, m\}$, i. e., θ^m_x maximizes L^m_x given by (4.34) *for each $x \in S$.*

Then the estimators θ^m are strongly asymptotically consistent, i. e.

$$\theta^m \to \theta_0, \quad \textit{a. s. for } m \to \infty.$$

Proof. Define the entries of $\theta_0 = (\theta^0_x)_{x \in S}$ by

$$\theta^0_x = (\mu^0_{x,1}, \dots, \mu^0_{x,x-1}, \mu^0_{x,x+1}, \dots, \mu^0_{x,K}, p^0_x) \quad \text{for each } x \in S.$$

Furthermore define $M_0(x, a, b)$ and $M_m(x, a, b)$ for each $x \in S$, $a, b \in \mathbb{N}$ with $a \geq b$ and $m \in \mathbb{N}$ as in Lemma 5.27. For asymptotical consistency, we show for each $x, y \in S$ with $x \neq y$, that μ^m_{xy} converges a. s. to μ^0_{xy} for m tending to infinity and p^m_x converges a. s. to p^0_x. Van der Vaart proves in [54, Chapter 5, Lemma 5.35], that the true parameter maximizes the function $L^0 : \Theta_{id} \to \mathbb{R}$ defined by

$$L^0(\theta) := \mathbb{E}_{\theta_0}\big[\log h^{\mathrm{c}}_\theta(\eta)\big], \tag{5.29}$$

which is well-defined with Lemma 5.18. Analogously to the proof of Lemma 5.18, Part (iii), we obtain for each $\theta = (\theta_x)_{x \in S}$ with $\theta_x = (\mu_{x,1}, \dots, \mu_{x,x-1}, \mu_{x,x+1}, \dots, \mu_{x,K}, p_x)$

$$L^0(\theta) = C + \sum_{x \in S} L^0_x(\theta_x),$$

where $|C| < \infty$ is independent of θ and for each $x \in S$ the function $L^0_x : [0, \infty)^{K-1} \times [0, 1] \to \mathbb{R}$ is given by

$$\begin{aligned} L^0_x(\theta_x) = &\sum_{y \in S \setminus \{x\}} \sum_{\substack{a,b=1 \\ a \geq b}}^{\tilde{N}} \mathbb{E}_{\theta_0}\big[N^{(\eta)}_{x,y,a,b}\big] \left(\log \mu_{xy} + (b-1)\log p_x + (a-b)\log(1-p_x)\right) \\ &- \mu_x \sum_{a=1}^{\tilde{N}} \mathbb{E}_{\theta_0}\big[T^{(\eta)}_{x,a}\big]\, q(a, p_x). \end{aligned} \tag{5.30}$$

If $\sum_{y \in S \setminus \{x\}} \mathbb{E}_{\theta_0}[N^{(\eta)}_{x,y,a,b}] > 0$ for $a, b \in \{1 \dots, \tilde{N}\}$ and $x \in S$, then $\mathbb{E}_\theta[T^{(\eta)}_{x,a}] > 0$ by Corollary 5.14. Theorem 5.1 implies L^0_x is finite at the maximal points for each $x \in S$. Therefore if L^0_x is uniquely maximized, then the maximum is at θ^0_x by [54, Lemma 5.35]. Analogous argumentation yields that the maximum of L^m_x is a. s. at θ^m_x if the maximal point is unique. In the following we show the almost sure convergence of the maximal points of L^m_x to the maximal points of L^0_x for each $x \in S$.

Fix $x \in S$. Since $\theta_0 \in \Theta_{id}$ there exists $a^T \in \mathbb{N}$ with $a^T \geq 2$ such that $\mathbb{E}_{\theta_0}\big[T^{(\eta)}_{x,a^T}\big] > 0$ and there exists $m_0 \in \mathbb{N}$ such that $\sum_{j=1}^m T^{(\omega_j,T_j)}_{x,a^T} > 0$ a.s. for all $m \geq m_0$. Define for each $x, y \in S$ with $x \neq y$ the function $\varphi^0_{xy} : [0,1] \to [0,\infty)$ by

$$\varphi^0_{xy}(p) = \Big(\sum_{\substack{a,b=1\\ a\geq b}}^{\tilde{N}} \mathbb{E}_{\theta_0}\big[N^{(\eta)}_{x,y,a,b}\big]\Big)\Big(\sum_{a=1}^{\tilde{N}} \mathbb{E}_{\theta_0}\big[T^{(\eta)}_{x,a}\big]\, q(a,p)\Big)^{-1}, \tag{5.31}$$

and the vector $\hat{\theta}^0_x(p) = (\varphi^0_{x,1}(p), \dots, \varphi^0_{x,x-1}(p), \varphi^0_{x,x+1}(p), \dots, \varphi^0_{x,K}(p), p)$. For each $m \in \mathbb{N}$ and $x, y \in S$ with $x \neq y$ define

$$\varphi^m_{xy}(p) = \Big(\sum_{\substack{a,b=1\\ a\geq b}}^{\tilde{N}} \sum_{j=1}^{m} N^{(\omega_j)}_{x,y,a,b}\Big)\Big(\sum_{a=1}^{\tilde{N}}\sum_{j=1}^{m} T^{(\omega_j,T_j)}_{x,a}\, q(a,p)\Big)^{-1},$$

and the vector $\hat{\theta}^m_x(p) = (\varphi^m_{x,1}(p), \dots, \varphi^m_{x,x-1}(p), \varphi^m_{x,x+1}(p), \dots, \varphi^m_{x,K}(p), p)$.

By Lemma 5.27 and Theorem 5.1 L^0_x is uniquely maximized at $\hat{\theta}^0_x(p^0_x)$ which is θ^0_x. Furthermore there exists $m_0 \in \mathbb{N}$ such that L^m_x is uniquely maximized at $\hat{\theta}^m_x(p^m_x) = \theta^m_x$ a.s. for each $m \geq m_0$. If p^m_x converges to p^0_x a.s., then $\mu^m_{xy} = \varphi^m_{xy}(p^m_x)$ converges a.s. to $\varphi^0_{xy}(p^0_x) = \mu^0_{xy}$ for m tending to infinity for each $y \in S$ with $x \neq y$ using the strong law of large numbers.

Assume $p^0_x \in (0,1)$. To show the almost sure convergence of p^m_x to p^0_x, we verify the assumptions of Lemma 5.26. Define the deterministic function $\Phi : [0,1] \to \mathbb{R}$ by

$$\Phi(p) = \frac{\partial L^0_x(\hat{\theta}^0_x(p))}{\partial p} = \sum_{\substack{a,b=1\\ a\geq b}}^{\tilde{N}} \sum_{\substack{x,y\in S\\ x\neq y}} \mathbb{E}_{\theta_0}\big[N^{(\eta)}_{x,y,a,b}\big]\Big(\frac{b-1}{p} - \frac{a-b}{1-p} - \frac{\sum_{k=1}^{\tilde{N}} \mathbb{E}_{\theta_0}\big[T^{(\eta)}_{x,k}\big]\frac{\partial}{\partial p} q(k,p)}{\sum_{j=1}^{\tilde{N}} \mathbb{E}_{\theta_0}[T^{(\eta)}_{x,j}] q(j,p)}\Big).$$

Since $p^0_x \in (0,1)$, the continuous differentiable function $L^0_x(\hat{\theta}_x(p))$ has a unique maximum at p^0_x by Lemma 5.27, Theorem 5.1 and [54, Lemma 5.35]. That implies for Φ, that there exists $\delta > 0$, such that

$$\Phi(p^0_x + \varepsilon) < 0 < \Phi(p^0_x - \varepsilon), \qquad \text{for all } \varepsilon \in (0, \delta].$$

Analogously define for each $m \in \mathbb{N}$ the stochastic function $\Phi_m : [0,1] \to \mathbb{R}$ by

$$\begin{aligned}\Phi_m(p) &= \frac{\partial L^m_x(\hat{\theta}^m_x(p))}{\partial p}\\ &= \sum_{\substack{a,b=1\\ a\geq b}}^{\tilde{N}} \sum_{\substack{x,y\in S\\ x\neq y}} \sum_{i=1}^{m} N^{(\omega_i)}_{x,y,a,b}\Big(\frac{b-1}{p} - \frac{a-b}{1-p} - \frac{\sum_{k=1}^{\tilde{N}}\sum_{j=1}^m T^{(\omega_j,T_j)}_{x,k}\ \frac{\partial}{\partial p} q(k,p)}{\sum_{l=1}^{\tilde{N}}\sum_{j=1}^m T^{(\omega_j,T_j)}_{x,l}\ q(l,p)}\Big).\end{aligned}$$

It follows $\Phi_m(p)/m \to \Phi(p)$ a.s. for $m \to \infty$ for each $p \in (0,1)$. Since $p^0_x \in (0,1)$, there exists $m_0 \in \mathbb{N}$ such that $M_m(x, a^T, b) > 0$ a.s. for each $b \leq a^T$ and $m \geq m_0$ by Lemma 5.27. Using Theorem 5.1 the continuous differentiable function $L^m_x(\hat{\theta}^m_x(p))$ has a unique critical point which is the maximum. Therefore Φ_m has the unique root at p^m_x. Altogether by Lemma 5.26, we obtain $p^m_x \to p^0_x$ almost surely for $m \to \infty$.

Assume $p_x^0 = 0$. For each $m \in \mathbb{N}$ define the constant

$$c_0^m = \sum_{\substack{a,b=1 \\ a \geq b}}^{\tilde{N}} (b-1) \sum_{j=1}^{m} \sum_{\substack{y \in S \\ x \neq y}} N_{x,y,a,b}^{(\omega_j)} \sum_{k=1}^{\tilde{N}} k \sum_{l=1}^{m} T_{x,k}^{(\omega_l, T_l)}.$$

By Lemma 5.27 $M_m(x,a,b) = 0$ a. s for all $a, b \in \mathbb{N}$ with $a \geq b \geq 2$ and $m \in \mathbb{N}$. Therefore we obtain $c_0^m = 0$ a. s. for all $m \in \mathbb{N}$. For each $m \in \mathbb{N}$ define the constant

$$c_1^m = -\frac{1}{2} \sum_{a=1}^{\tilde{N}} \sum_{j=1}^{m} \sum_{\substack{y \in S \\ x \neq y}} N_{x,y,a,1}^{(\omega_j)} \sum_{k=1}^{\tilde{N}} k(2a-k-1) \sum_{l=1}^{m} T_{x,k}^{(\omega_l, T_l)}.$$

Using the strong law of large numbers we obtain

$$\frac{c_1^m}{m^2} \to \frac{1}{2} \sum_{a=1}^{\tilde{N}} \sum_{\substack{y \in S \\ x \neq y}} \mathbb{E}_{\theta_0}\big[N_{x,y,a,1}^{(\eta)}\big] \sum_{k=1}^{\tilde{N}} k(2a-k-1) \mathbb{E}_{\theta_0}\big[T_{x,k}^{(\eta)}\big] =: c_1^0, \quad \text{a. s. for } m \to \infty.$$

By Corollary 5.14 we obtain

$$\sum_{\substack{y \in S \\ x \neq y}} \mathbb{E}_{\theta_0}[N_{x,y,a,1}^{(\eta)}] = a\mu_x^0 \, \mathbb{E}_{\theta_0}[T_{x,a}^{(\eta)}], \quad \text{for each } a \in \mathbb{N}.$$

Putting this into c_1^0 we get

$$c_1^0 = -\frac{\mu_x^0}{2} \sum_{a=1}^{\tilde{N}} \sum_{k=1}^{\tilde{N}} ak(2a-k-1) \mathbb{E}_{\theta_0}\big[T_{x,a}^{(\eta)}\big] \, \mathbb{E}_{\theta_0}\big[T_{x,k}^{(\eta)}\big].$$

Now we split the second sum and change the order of summation, i. e.

$$\begin{aligned} c_1^0 = &-\frac{\mu_x^0}{2} \sum_{a=2}^{\tilde{N}} \sum_{k=1}^{2a-1 \wedge \tilde{N}} ak(2a-k-1) \mathbb{E}_{\theta_0}\big[T_{x,a}^{(\eta)}\big] \, \mathbb{E}_{\theta_0}\big[T_{x,k}^{(\eta)}\big] \\ &-\frac{\mu_x^0}{2} \sum_{k=2}^{\tilde{N}} \sum_{a=1}^{\lfloor \frac{k}{2} \rfloor} ak(2a-k-1) \mathbb{E}_{\theta_0}\big[T_{x,a}^{(\eta)}\big] \, \mathbb{E}_{\theta_0}\big[T_{x,k}^{(\eta)}\big]. \end{aligned}$$

Changing the summation index in the second sum and adding the first and the second sum yields

$$\begin{aligned} c_1^0 = &-\frac{\mu_x^0}{2} \sum_{a=2}^{\tilde{N}} \sum_{k=1}^{\lfloor \frac{a}{2} \rfloor} ak(a+k-2) \mathbb{E}_{\theta_0}\big[T_{x,a}^{(\eta)}\big] \, \mathbb{E}_{\theta_0}\big[T_{x,k}^{(\eta)}\big] \\ &-\frac{\mu_x^0}{2} \sum_{a=2}^{\tilde{N}} \sum_{k=\lfloor \frac{a}{2} \rfloor + 1}^{2a-1 \wedge \tilde{N}} ak(2a-k-1) \mathbb{E}_{\theta_0}\big[T_{x,a}^{(\eta)}\big] \, \mathbb{E}_{\theta_0}\big[T_{x,k}^{(\eta)}\big] < 0. \end{aligned}$$

Therefore there exists $m_0 \in \mathbb{N}$ such that $c_1^m < 0$ a. s. for all $m \geq m_0$. By Lemma 5.27 and Theorem 5.1 there exists $m_1 \geq m_0$ such that $p_x^m = 0$ a. s. for all $m \geq m_1$. Altogether p_x^m converges a. s. to $p_x^0 = 0$.

Assume $p_x^0 = 1$. Lemma 5.27 and Theorem 5.1 imply the existence of $m_0 \in \mathbb{N}$ such that $p_x^m = 1$ a. s. for all $m \geq m_0$. □

5.2 Asymptotical normality

Let Θ_{id} be the set of identifiable parameters, given by (5.20). For proving asymptotic normality we assume that the true parameter is in the interior of Θ_{id}. Define the open set

$$\dot{\Theta} = \Big\{\theta = (\theta_x)_{x\in S} \in \Theta_c : p_x \in (0,1) \text{ and } \mu_{xy} \in (0,\infty) \text{ for each } x, y \in S \text{ with } x \neq y, \\ \text{where } \theta_x = (\mu_{x,1}, \dots, \mu_{x,x-1}, \mu_{x,x+1}, \dots, p_x)\Big\}.$$

Lemma 5.32. *Assume $P_N(1) < 1$. Then $\dot{\Theta}$ is the interior of Θ_{id}.*

Proof. For $\theta \in \dot{\Theta}$ each rating class is accessible from all other rating classes. Therefore by Lemma 5.25 we have $\dot{\Theta} \subset \Theta_{id}$ since there exists $n \in \mathbb{N}$ with $n \geq 2$ such that $P_N(n) > 0$ and each rating class is accessible from all other rating classes. Furthermore each parameter $\theta \in \Theta_{id}$, which is not contained in $\dot{\Theta}$, is a boundary point of Θ_{id}. □

Definition 5.33. *Define the set*

$$\mathcal{S}_{feas} := \big\{(n, T, \omega) \in \mathbb{N} \times (0,\infty) \times \mathcal{S} :\ h^{\mathrm{c}}_\theta(n, T, \omega) > 0, \text{ for any } \theta \in \dot{\Theta}\big\}.$$

The definition of h^{c}_θ in Definition 4.20 implies for each path $(n, T, \omega) \in \mathcal{S}_{feas}$, that $h^{\mathrm{c}}_\theta(n, T, \omega) > 0$ for all $\theta \in \dot{\Theta}$. Furthermore (ω, T) is strongly feasible in the sense of Definition 4.4 and

$$\nu^{\mathrm{c}}_n\big(\{z_0(\omega)\}\big) > 0,$$

where $z_0(\omega)$ is the first component of ω.

For asymptotic normality we prove some properties of the derivatives of $\log h^{\mathrm{c}}_\theta$ first.

Lemma 5.34. *Let $\theta_0 \in \Theta_c$ be the true parameter. Suppose $\eta \in \mathcal{O}$ is a random sample path with density $h^{\mathrm{c}}_{\theta_0}$ given by Definition 4.20. Then for all $(n, T, \omega) \in \mathcal{S}_{feas}$ the function $\dot{\Theta} \ni \theta \mapsto \log h^{\mathrm{c}}_\theta(n, T, \omega)$ is three times continuously differentiable and the first, second and third order partial derivatives with respect to $\theta \in \dot{\Theta}$ of $\log h^{\mathrm{c}}_\theta(\eta)$ have finite expectation.*

Proof. Let (n, T, ω) be in $\mathcal{S}_{feas}$. For every $\theta = (\theta_x)_{x\in S} \in \Theta_c$ with

$$\theta_x = (\mu_{x,1}, \dots, \mu_{x,x-1}, \mu_{x,x+1}, \dots, \mu_{x,K}, p_x) \quad \text{for } x \in S,$$

we obtain

$$\log h^{\mathrm{c}}_\theta(n, T, \omega) = \log \nu^{\mathrm{c}}_n\big(\{z_0(\omega)\}\big) - \sum_{x\in S} \mu_x \sum_{a=1}^{n} T^{(\omega,T)}_{x,a} q(a, p_x) \\ + \sum_{\substack{x,y\in S\\ x\neq y}} \sum_{\substack{a,b=1\\ a\geq b}}^{n} N^{(\omega)}_{x,y,a,b}\big(\log \mu_{xy} + (b-1)\log p_x + (a-b)\log(1-p_x)\big).$$

Because $(n, T, \omega) \in \mathcal{S}_{feas}$, it follows $|\log \nu^{\mathrm{c}}_n\big(\{z_0(\omega)\}\big)| < \infty$. For $\theta \in \dot{\Theta}$ the parameter entries $\mu_{xy} > 0$ and $p_x \in (0,1)$ for all $x, y \in S$. So, $\log h^{\mathrm{c}}_\theta(n, T, \omega)$ is three times continuously differentiable in $\dot{\Theta}$.

The first order partial derivative of $\log h^{\mathrm{c}}_\theta(n, T, \omega)$ with respect to μ_{xy} for each $x, y \in S$ with $x \neq y$ is equivalent to

$$\frac{\partial}{\partial \mu_{xy}} \log h^{\mathrm{c}}_\theta(n, T, \omega) = \sum_{\substack{a,b=1\\ a\geq b}}^{n} \frac{N^{(\omega)}_{x,y,a,b}}{\mu_{xy}} - \sum_{a=1}^{n} T^{(\omega,T)}_{x,a}\, q(a, p_x), \quad \text{for every } (n, T, \omega) \in \mathcal{S}_{feas}.$$

Using the integrability of $N^{(\eta)}_{x,y,a,b}$ and $T^{(\eta)}_{x,a}$ in Lemma 5.18 it follows

$$\mathbb{E}_{\theta_0}\Big[\Big|\frac{\partial}{\partial \mu_{xy}} \log h^{\mathrm{c}}_\theta(\eta)\Big|\Big] = \sum_{\substack{a,b=1\\ a\geq b}}^{\tilde{N}} \frac{\mathbb{E}_{\theta_0}\big[N^{(\eta)}_{x,y,a,b}\big]}{\mu_{xy}} + \sum_{a=1}^{\tilde{N}} \mathbb{E}_{\theta_0}\big[T^{(\eta)}_{x,a}\big] q(a,p_x) < \infty.$$

Differentiation with respect to p_x yields also an integrable function and analogous argumentation shows finite expectation of the second and third order partial derivatives. $\square$

Lemma 5.35. *Let $\theta \in \dot{\Theta}$ be an arbitrary parameter. Then we obtain for the random sample path $\eta \in \mathcal{O}$ with density h^{c}_θ, given by Definition 4.20, for each $i,j \in \{1,\dots,K^2\}$*

$$\mathbb{E}_\theta\Big[\frac{\partial}{\partial \theta_i} \log h^{\mathrm{c}}_\theta(\eta) \frac{\partial}{\partial \theta_j} \log h^{\mathrm{c}}_\theta(\eta)\Big] = -\mathbb{E}_\theta\Big[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log h^{\mathrm{c}}_\theta(\eta)\Big]. \tag{5.36}$$

Proof. Using the binomial theorem we obtain for each $k \in \mathbb{N}$ and $p \in (0,1]$

$$\sum_{j=1}^{k} \binom{k}{j} p^{j-1}(1-p)^{k-j} = \frac{1}{p}\sum_{j=1}^{k} \binom{k}{j} p^{j}(1-p)^{k-j} = \frac{1-(1-p)^k}{p} = q(k,p),$$

by definition of the function q in (2.35). For $p=0$ we obtain for each $k \in \mathbb{N}$

$$\sum_{j=1}^{k} \binom{k}{j} p^{j-1}(1-p)^{k-j} = k = q(k,0),$$

by definition of q. It follows for each $k \in \mathbb{N}$ and $p \in [0,1]$

$$q(k,p) = \sum_{j=1}^{k} \binom{k}{j} p^{j-1}(1-p)^{k-j}. \tag{5.37}$$

The derivative of q equals

$$\frac{\partial}{\partial p} q(k,p) = \sum_{j=1}^{k} \binom{k}{j} p^{j-1}(1-p)^{k-j}\Big(\frac{j-1}{p} - \frac{k-j}{1-p}\Big). \tag{5.38}$$

By Lemma 5.34 the function $\dot{\Theta} \ni \theta \mapsto \log h^{\mathrm{c}}_\theta(n,T,\omega)$ is twice continuously differentiable for $(n,T,\omega) \in \mathcal{S}_{feas}$. Fix $(n,T,\omega) \in \mathcal{S}_{feas}$. Using (5.37) the first derivative of $\log h^{\mathrm{c}}_\theta(n,T,\omega)$ with respect to μ_{xy} with $x,y \in S$, $x \neq y$, equals

$$\frac{\partial}{\partial \mu_{xy}} \log h^{\mathrm{c}}_\theta(n,T,\omega) = \sum_{\substack{a,b=1\\ a\geq b}}^{n} \Big(\frac{N^{(\omega)}_{x,y,a,b}}{\mu_{xy}} - \binom{a}{b} p_x^{b-1}(1-p_x)^{a-b} T^{(\omega,T)}_{x,a}\Big), \tag{5.39}$$

where $\theta = (\theta_x)_{x\in S} \in \dot{\Theta}$ with $\theta_x = (\mu_{x,1},\dots,\mu_{x,x-1},\mu_{x,x+1},\dots,\mu_{x,k},p_x)$. The first derivative with respect to p_x for $x \in S$ is given by

$$\begin{aligned}\frac{\partial}{\partial p_x} \log h^{\mathrm{c}}_\theta(n,T,\omega) = &\sum_{\substack{a,b=1\\ a\geq b}}^{n} \sum_{\substack{y\in S\\ x\neq y}} \Big(\frac{b-1}{p_x} - \frac{a-b}{1-p_x}\Big)\\ &\times \Big(N^{(\omega)}_{x,y,a,b} - \mu_{xy}\binom{a}{b} p_x^{b-1}(1-p_x)^{a-b} T^{(\omega,T)}_{x,a}\Big).\end{aligned} \tag{5.40}$$

Define $S_i(\omega, T)$ as in (5.8) for $x_i, y_i \in S$ with $x_i \neq y_i$ and $a_i, b_i \in \mathbb{N}$ with $a_i \geq b_i$ for $i = 1, 2$. Hence, we obtain

$$\begin{aligned}
&\mathbb{E}_\theta\Big[\frac{\partial}{\partial \mu_{x_1,y_1}} \log h^{\mathrm{c}}_\theta(\eta) \frac{\partial}{\partial p_{x_2}} \log h^{\mathrm{c}}_\theta(\eta)\Big] \\
&\quad = \sum_{\substack{a_1,b_1=1\\ a_1\geq b_1}}^{\tilde N} \sum_{\substack{a_2,b_2=1\\ a_2\geq b_2}}^{\tilde N} \sum_{\substack{y_2\in S\\ x_2\neq y_2}} \frac{1}{\mu_{x_1,y_1}} \Big(\frac{b_2-1}{p_{x_2}} - \frac{a_2-b_2}{1-p_{x_2}}\Big) \mathbb{E}_\theta[S_1(\eta)S_2(\eta)].
\end{aligned}$$

If $x_1 \neq x_2$, then this equals zero by (5.16). Differentiation of the first derivative with respect to p_{x_2} in (5.40) with respect to μ_{x_1,y_1} is also zero. If $x_1 = x = x_2$, then the equation simplifies to

$$\begin{aligned}
&\mathbb{E}_\theta\Big[\frac{\partial}{\partial \mu_{xy}} \log h^{\mathrm{c}}_\theta(\eta) \frac{\partial}{\partial p_x} \log h^{\mathrm{c}}_\theta(\eta)\Big] \\
&\quad = \sum_{\substack{a,b=1\\ a\geq b}}^{\tilde N} \frac{1}{\mu_{xy}} \Big(\frac{b-1}{p_x} - \frac{a-b}{1-p_x}\Big) \mu_{xy} \binom{a}{b} p_x^{b-1}(1-p_x)^{a-b} \mathbb{E}_\theta\big[T^{(\eta)}_{x,a}\big],
\end{aligned}$$

using (5.15) and (5.16). Since the derivative of q with respect to p_x is given by (5.38), this is equivalent to

$$\begin{aligned}
&\mathbb{E}_\theta\Big[\frac{\partial}{\partial \mu_{xy}} \log h^{\mathrm{c}}_\theta(\eta) \frac{\partial}{\partial p_x} \log h^{\mathrm{c}}_\theta(\eta)\Big] \\
&\qquad\quad = \sum_{a=1}^{\tilde N} \frac{\partial}{\partial p_x} q(a, p_x)\, \mathbb{E}_\theta\big[T^{(\eta)}_{x,a}\big] = -\mathbb{E}_\theta\Big[\frac{\partial^2}{\partial \mu_{xy} \partial p_x} \log h^{\mathrm{c}}_\theta(\eta)\Big],
\end{aligned}$$

which is obtained by differentiating the derivative in (5.40) with respect to μ_{xy}. Analogous argumentation yields the result for the other second derivatives. $\square$

Lemma 5.41. *Suppose $P_N(1) < 1$ and $\theta \in \dot\Theta$. Let $I(\theta) = (I(\theta)_{ij})_{i,j\in\{1,\ldots,K^2\}}$ be the Fisher matrix for the density h^{c}_θ, i. e.*

$$I(\theta)_{ij} = -\mathbb{E}_\theta\Big[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log h^{\mathrm{c}}_\theta(\eta)\Big] \quad \text{for each } i, j \in \{1, \ldots, K^2\}. \tag{5.42}$$

Then the entries of the Fisher matrix are finite and the Fisher matrix is positive definite.

Proof. Finiteness of the entries of the Fisher matrix follows from Lemma 5.34. Define the entries of $\theta = (\theta_x)_{x\in S}$ by $\theta_x = (\mu_{x,1}, \ldots, \mu_{x,x-1}, \mu_{x,x+1}, \ldots, \mu_{x,K}, p_x)$. Computation of the second derivatives of $\log h^{\mathrm{c}}_\theta$ yields the following structure of the Fisher matrix

$$I(\theta) = \begin{pmatrix} I^{(1)} & 0 & \cdots & \cdots & 0 \\ 0 & I^{(2)} & \ddots & & \vdots \\ \vdots & \ddots & I^{(3)} & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & I^{(K)} \end{pmatrix}$$

with the block matrices $I^{(x)} = (I^{(x)}_{ij})_{i,j\in\{1,\dots,K\}}$ for $x \in S$ defined by

$$I^{(x)} = \begin{pmatrix} D^{\mu}_{x1} & 0 & \cdots & \cdots & \cdots & \cdots & 0 & D^{\mu,p}_{x} \\ 0 & D^{\mu}_{x2} & \ddots & & & & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & & \vdots & \vdots \\ \vdots & & \ddots & D^{\mu}_{x,x-1} & \ddots & & \vdots & \vdots \\ \vdots & & & \ddots & D^{\mu}_{x,x+1} & \ddots & \vdots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & D^{\mu}_{xK} & D^{\mu,p}_{x} \\ D^{\mu,p}_{x} & \cdots & \cdots & \cdots & \cdots & \cdots & D^{\mu,p}_{x} & D^{p}_{x} \end{pmatrix},$$

where the function D^{μ} is defined by

$$D^{\mu}_{xy} = -\mathbb{E}_\theta\Big[\frac{\partial^2 \log h^{c}_\theta(\eta)}{\partial \mu^2_{xy}}\Big] = \sum_{\substack{a,b=1\\ a\geq b}}^{\tilde N} \frac{\mathbb{E}_\theta\big[N^{(\eta)}_{x,y,a,b}\big]}{\mu^2_{xy}}, \quad \text{for each } x,y\in S, \text{ with } x\neq y.$$

Furthermore $D^{\mu,p}_{x}$ is given by

$$D^{\mu,p}_{x} = -\mathbb{E}_\theta\Big[\frac{\partial^2 \log h^{c}_\theta(\eta)}{\partial p_x \partial \mu_{xy}}\Big] = \sum_{a=1}^{\tilde N} \frac{\partial}{\partial p_x} q(a,p_x)\mathbb{E}_\theta\big[T^{(\eta)}_{x,a}\big], \text{ for } x,y\in S, \text{ with } x\neq y,$$

where q is defined in (2.35). The function D^{p}_{x} equals for each $x \in S$

$$\begin{aligned} D^{p}_{x} &= -\mathbb{E}_\theta\Big[\frac{\partial^2 \log h^{c}_\theta(\eta)}{\partial p^2_x}\Big] \\ &= \sum_{\substack{a,b=1\\ a\geq b}}^{\tilde N} \sum_{\substack{y\in S\\ x\neq y}} \mathbb{E}_\theta\big[N^{(\eta)}_{x,y,a,b}\big]\Big(\frac{b-1}{p^2_x} + \frac{a-b}{(1-p_x)^2}\Big) + \mu_x \sum_{a=1}^{\tilde N} \frac{\partial^2}{\partial p^2_x} q(a,p_x)\mathbb{E}_\theta\big[T^{(\eta)}_{x,a}\big]. \end{aligned}$$

The Fisher matrix $I(\theta)$ is positive definite, if all principal minors are positive, i. e., for each $x \in S$ the determinant of the matrix $I^{(x)}_m = (I^{(x)}_{ij})_{i,j=1,\dots,m}$ is positive for each $m \in \{1,\dots,K\}$. Since $P_N(1) < 1$, the set $\dot\Theta \subset \Theta_{id}$ by Lemma 5.32. Using the definition of the identifiable set Θ_{id} in (5.20) and Corollary 5.14 there exists some $a \in \{2,\dots,\tilde N\}$ for each $x \in S$, such that

$$\mathbb{E}_\theta\big[N^{(\eta)}_{x,y,a,b}\big] > 0 \quad \text{for each } b\in\{1,\dots,a\} \text{ and each } y \in S \text{ with } x\neq y. \tag{5.43}$$

Therefore $D^{\mu}_{xy} > 0$ and the determinant of $I^{(x)}_m$ is strictly positive for $m \in \{1,\dots,K-1\}$ and all $x \in S$. By adding an appropriate factor of the k-th column to the last column for all $k \in \{1,\dots,K-1\}$, we transform $I^{(x)}_K$ in triangular form. Then the determinant of $I^{(x)}_K$ equals

$$\det(I^{(x)}_K) = \det(I^{(x)}_{K-1})\Big(D^{p}_{x} - \sum_{y\in S\setminus\{x\}} \frac{(D^{\mu,p}_{x})^2}{D^{\mu}_{xy}}\Big).$$

Define the function $g : [0,1] \to (0,\infty)$ by

$$g(1-p) = \sum_{k=1}^{\tilde{N}} \sum_{j=0}^{k-1} \mathbb{E}_\theta\big[T_{x,k}^{(\eta)}\big](1-p)^j = \sum_{j=0}^{\tilde{N}-1} \sum_{k=j+1}^{\tilde{N}} \mathbb{E}_\theta\big[T_{x,k}^{(\eta)}\big](1-p)^j.$$

Lemma 4.41 says for all $p \in [0,1]$

$$(1-p)g'(1-p)^2 < g(1-p)g'(1-p) + (1-p)g(1-p)g''(1-p). \tag{5.44}$$

Corollary 5.14 and the definition of q in (5.37) imply for each $x, y \in S$ with $x \neq y$

$$\mu_{xy} = \sum_{\substack{a,b=1\\a\geq b}}^{\tilde{N}} \mathbb{E}_\theta\big[N_{x,y,a,b}^{(\eta)}\big]\Big(\sum_{k=1}^{\tilde{N}} q(k,p)\mathbb{E}_\theta\big[T_{x,k}^{(\eta)}\big]\Big)^{-1} = \sum_{\substack{a,b=1\\a\geq b}}^{\tilde{N}} \mathbb{E}_\theta\big[N_{x,y,a,b}^{(\eta)}\big]\big(g(1-p_x)\big)^{-1}.$$

Substituting that into D_{xy}^μ, $D_x^{\mu,p}$ and D_x^p yields

$$\begin{aligned} D_x^p - \sum_{\substack{y\in S\\x\neq y}} \frac{(D_x^{\mu,p})^2}{D_{xy}^\mu} = \sum_{\substack{a,b=1\\a\geq b}}^{\tilde{N}} \sum_{\substack{y\in S\\x\neq y}} \mathbb{E}_\theta\big[N_{x,y,a,b}^{(\eta)}\big]\Big(\frac{b-1}{p_x^2} + \frac{a-b}{(1-p_x)^2} + g''(1-p_x)(g(1-p_x))^{-1} \\ - (g'(1-p_x))^2(g(1-p_x))^{-2}\Big). \end{aligned} \tag{5.45}$$

Since $\mathbb{E}_\theta[(\partial/\partial p_x)\log h_\theta^{\mathrm{c}}(\eta)] = 0$ by (5.40) and Corollary 5.14, we obtain

$$\sum_{\substack{a,b=1\\a\geq b}}^{\tilde{N}} \mathbb{E}_\theta\big[\tilde{N}_{x,a,b}^{(\eta)}\big]\frac{a-b}{(1-p_x)^2} = \sum_{\substack{a,b=1\\a\geq b}}^{\tilde{N}} \sum_{\substack{y\in S\\x\neq y}} \mathbb{E}_\theta\big[N_{x,y,a,b}^{(\eta)}\big]\Big(\frac{b-1}{p_x(1-p_x)} + \frac{g'(1-p_x)}{(1-p_x)g(1-p_x)}\Big).$$

Substituting this into (5.45), and using (5.43) and inequality (5.44), we proved the positive definiteness of $I(\theta)$. $\square$

Lemma 5.46. *Let $\theta_0 \in \dot{\Theta}$ be the true parameter. Suppose η is a random sample path with density $h_{\theta_0}^{\mathrm{c}}$ given by Definition 4.20. Furthermore, let $B_r(\theta_0)$ be a closed ball with center θ_0 and radius $r > 0$, such that $B_r(\theta_0) \subset \dot{\Theta}$. Then there exists a function $M_{ikl} : \mathcal{O} \to [0,\infty)$ for each $i, k, l \in \{1, \ldots, K^2\}$, such that $\mathbb{E}_{\theta_0}[M_{ikl}(\eta)] < \infty$ and for each $(n, T, \omega) \in \mathcal{S}_{feas}$*

$$\left|\frac{\partial^3}{\partial\theta_i\partial\theta_k\partial\theta_j} \log h_\theta^{\mathrm{c}}(n,T,\omega)\right| \leq M_{ikl}(n,T,\omega), \quad \textit{for all } \theta \in B_r(\theta_0),$$

where $\mathcal{S}_{feas}$ is given by Definition 5.33.

Proof. Since $B_r(\theta_0)$ is closed, there exists $\mu_{xy}^u < \infty$ for each $x, y \in S$ with $x \neq y$, such that

$$0 < \mu_{xy} \leq \mu_{xy}^u, \quad \text{for all } \theta = (\theta_x)_{x\in S} \in B_r(\theta_0),$$

where $\theta_x = (\mu_{x,1}, \ldots, \mu_{x,x-1}, \mu_{x,x+1}, \ldots, \mu_{x,K}, p_x)$. Furthermore there exist $p_x^l \in (0,1)$ and $p_x^u \in (0,1)$ for each $x \in S$, such that

$$p_x^l \leq p_x \leq p_x^u, \quad \text{for all } \theta \in B_r(\theta_0).$$

For each $x \in S$ define the function $M_{xK,xK,xK} : \mathcal{O} \to [0, \infty)$ by

$$M_{xK,xK,xK}(n,T,\omega) = \mu_x^u \sum_{a=1}^{\tilde{N}} T_{x,a}^{(\omega,T)} \sum_{j=3}^{a-1} j^3 + 2 \sum_{\substack{y \in S \\ x \neq y}} \sum_{\substack{a,b=1 \\ a \geq b}}^{\tilde{N}} N_{x,y,a,b}^{(\omega)} \Big(\frac{b-1}{(p_x^l)^3} + \frac{a-b}{(1-p_x^u)^3} \Big),$$

for $(n,T,\omega) \in \mathcal{S}_{feas}$ and set $M_{xK,xK,xK}(n,T,\omega) = 0$ otherwise. For each $(n,T,\omega) \in \mathcal{S}_{feas}$ we have

$$\left| \frac{\partial^3}{\partial p_x^3} \log h_\theta^{\mathrm{c}}(n,T,\omega) \right| \leq M_{xK,xK,xK}(n,T,\omega),$$

for all $\theta \in B_r(\theta_0)$. Furthermore, using Lemma 5.18

$$\begin{aligned} \mathbb{E}_{\theta_0}[M_{xK,xK,xK}(\eta)] = 2 \sum_{\substack{y \in S \\ x \neq y}} \sum_{\substack{a,b=1 \\ a \geq b}}^{\tilde{N}} \mathbb{E}_{\theta_0}[N_{x,y,a,b}^{(\eta)}] \Big(\frac{b-1}{(p_x^l)^3} + \frac{a-b}{(1-p_x^u)^3} \Big) \\ + \mu_x^u \sum_{a=1}^{\tilde{N}} \mathbb{E}_{\theta_0}[T_{x,a}^{(\eta)}] \sum_{j=3}^{a-1} j^3 < \infty. \end{aligned}$$

The bounds for the remaining partial derivatives can be obtained in an analogous way. $\square$

After the preparing lemmata we prove asymptotic normality.

Theorem 5.47. *Suppose the assumption in 5.2 holds and additionally $P_N(1) < 1$. Let the true parameter θ_0 be in the open set $\dot{\Theta}$. Let $(n_j, T_j, \omega_j) \in \mathcal{O}$ for $j \in \mathbb{N}$ be independent realizations of the random sample path η with density $h_{\theta_0}^{\mathrm{c}}$, given by Definition 4.20. Assume $h_{\theta_0}^{\mathrm{c}}(n_j, T_j, \omega_j) > 0$ for all $j \in \mathbb{N}$. For each $m \in \mathbb{N}$ let θ^m be a maximum likelihood estimator for the paths (n_j, T_j, ω_j) for $j \in \{1, \dots, m\}$, i. e., θ_m maximizes the log-likelihood function $\log \mathcal{L}_m^{\mathrm{c}}$, given by*

$$\log \mathcal{L}_m^{\mathrm{c}}(\theta) = \sum_{j=1}^{m} \log h_\theta^{\mathrm{c}}(n_j, T_j, \omega_j).$$

Then $\sqrt{m}\,(\theta^m - \theta_0)$ is asymptotically normal with mean zero and covariance matrix $I(\theta_0)^{-1}$, which is the inverse Fisher matrix given by (5.42).

Proof. The proof is based on the proof of Lehmann and Casella in [43, Chapter 6, Theorem 5.1]. Let $r > 0$ be a radius, such that the closed ball $B_r(\theta_0) \subset \dot{\Theta}$. Since $P_N(1) < 1$, the true parameter $\theta_0 \in \Theta_{id}$ by Lemma 5.32. By Theorem 5.28 the estimator θ^m converges a. s. to θ_0 and there exists $m_0 \in \mathbb{N}$, such that $\theta^m \in B_r(\theta_0)$ a. s. for all $m \geq m_0$. By Lemma 5.34 the function $\log \mathcal{L}_m^{\mathrm{c}}$ admits all third derivatives with respect to $\theta \in \dot{\Theta}$. For each $i \in \{1, \dots, K^2\}$ and $m \geq m_0$ a second-order Taylor series expansion of $\partial / \partial \theta_i \log \mathcal{L}_m^{\mathrm{c}}(\theta^m)$ about θ_0 yields

$$\begin{aligned} \frac{\partial}{\partial \theta_i} \log \mathcal{L}_m^{\mathrm{c}}(\theta^m) = \frac{\partial}{\partial \theta_i} \log \mathcal{L}_m^{\mathrm{c}}(\theta_0) + \sum_{k=1}^{K^2} (\theta_k^m - \theta_{0k}) \frac{\partial^2}{\partial \theta_i \partial \theta_k} \log \mathcal{L}_m^{\mathrm{c}}(\theta_0) \\ + \frac{1}{2} \sum_{k,l=1}^{K^2} (\theta_k^m - \theta_{0k})(\theta_l^m - \theta_{0l}) \frac{\partial^3}{\partial \theta_i \partial \theta_k \partial \theta_l} \log \mathcal{L}_m^{\mathrm{c}}(\tilde{\theta}^{\mathrm{m}}), \end{aligned} \tag{5.48}$$

where $\tilde{\theta}^{\mathrm{m}} = t\,\theta^m + (1-t)\theta_0$ for $t \in [0,1]$ and therefore $\tilde{\theta}^{\mathrm{m}} \in B_r(\theta_0)$ for all $m \geq m_0$ a. s. Since θ^m is the maximum of the log-likelihood function, the first partial derivatives of $\log \mathcal{L}^{\mathrm{c}}_m$ are zero at θ^m. Hence, the left-hand side of (5.48) is zero.

For each $m \geq m_0$ define the vector $y^m = (y^m_1, \ldots, y^m_{K^2}) \in \mathbb{R}^{K^2}$ by

$$y^m_k = \sqrt{m}(\theta^m_k - \theta_{0k}), \qquad \text{for } k \in \{1, \ldots, K^2\}.$$

Furthermore define the matrix $A^m = (A^m_{ik})_{i,k \in \{1,\ldots,K^2\}}$ by

$$A^m_{ik} = \frac{1}{m} \frac{\partial^2}{\partial\theta_i \partial\theta_k} \log \mathcal{L}^{\mathrm{c}}_m(\theta_0) + \frac{1}{2m} \sum_{l=1}^{K^2} (\theta^m_l - \theta_{0l}) \frac{\partial^3}{\partial\theta_i \partial\theta_k \partial\theta_l} \log \mathcal{L}^{\mathrm{c}}_m(\tilde{\theta}^{\mathrm{m}}),$$

and the vector $b^m \in \mathbb{R}^{K^2}$ by

$$b^m_i = -\frac{1}{\sqrt{m}} \frac{\partial}{\partial\theta_i} \log \mathcal{L}^{\mathrm{c}}_m(\theta_0), \qquad \text{for } i \in \{1, \ldots, K^2\}.$$

Then equation (5.48) is equivalent to $A^m y^m = b^m$. If A^m_{ik} converges to $-I(\theta_0)_{ik}$ in probability for m tending to infinity for each $i, k \in \{1, \ldots, K^2\}$ and b^m converges weakly to a normally distributed random vector b with mean zero and covariance matrix $I(\theta_0)$, then the solutions y^m tend in probability to the solution y of the linear equation

$$I(\theta_0) y = b,$$

see [43, Chapter 6, Lemma 5.2]. Due to the positive definiteness of $I(\theta_0)$, proved in Lemma 5.41, the vector y is normally distributed with mean zero and covariance matrix $I(\theta_0)^{-1}$ and y^m is asymptotically normal.

The strong law of large numbers and the definition of the Fisher matrix in Lemma 5.41 implies for each $i, k \in \{1, \ldots, K^2\}$

$$\begin{aligned} \frac{1}{m} \frac{\partial^2}{\partial\theta_i \partial\theta_k} \log \mathcal{L}^{\mathrm{c}}_m(\theta_0) &= \frac{1}{m} \sum_{j=1}^{m} \frac{\partial^2}{\partial\theta_i \partial\theta_k} \log h^{\mathrm{c}}_{\theta_0}(n_j, T_j, \omega_j) \\ &\to \mathbb{E}_{\theta_0}\Big[\frac{\partial^2}{\partial\theta_i \partial\theta_k} \log h^{\mathrm{c}}_{\theta_0}(\eta)\Big] = -I(\theta_0)_{ik}, \quad \text{a. s. for } m \to \infty. \end{aligned}$$

Since $\tilde{\theta}^{\mathrm{m}} \in B_r(\theta_0)$ for all $m \geq m_0$, we obtain by Lemma 5.46 for each $i, k, l \in \{1, \ldots, K^2\}$

$$\frac{1}{m} \sum_{j=1}^{m} \frac{\partial^3}{\partial\theta_i \partial\theta_k \partial\theta_l} \log h^{\mathrm{c}}_{\tilde{\theta}^{\mathrm{m}}}(n_j, T_j, \omega_j) \leq \frac{1}{m} \sum_{j=1}^{m} M_{i,k,l}(n_j, T_j, \omega_j), \quad \text{for all } m \geq m_0.$$

Using the strong law of large numbers and Lemma 5.46, we see

$$\frac{1}{m} \sum_{j=1}^{m} M_{i,k,l}(n_j, T_j, \omega_j) \to \mathbb{E}_{\theta_0}[M_{i,k,l}(\eta)] < \infty, \quad \text{a. s. for } m \to \infty.$$

Due to the strong consistency of θ^m, the difference $(\theta^m_l - \theta_{0l})$ tends almost surely to zero for m tending to infinity for each $l \in \{1, \ldots, K^2\}$ and altogether A^m_{ik} tends a. s. to $-I(\theta_0)_{ik}$ for each $i, k \in \{1, \ldots, K^2\}$.

The first partial derivatives of $\log h^{\mathrm{c}}_\theta$ in (5.39) and (5.40), together by Corollary 5.14 yield

$$\mathbb{E}_{\theta_0}[\partial/\partial\theta_i \log h^{\mathrm{c}}_{\theta_0}(\eta)] = 0, \quad \text{for all } i \in \{1, \ldots, K^2\}.$$

Furthermore Lemma 5.35 shows for each $i, k \in \{1, \ldots, K^2\}$

$$\mathbb{E}_{\theta_0}\Big[\frac{\partial}{\partial\theta_i}\log h^{\mathrm{c}}_{\theta_0}(\eta)\frac{\partial}{\partial\theta_k}\log h^{\mathrm{c}}_{\theta_0}(\eta)\Big] = I(\theta_0)_{ik}.$$

Applying the central limit theorem the vector b^m tends weakly to a normally distributed random vector b with mean zero and covariance matrix $I(\theta_0)$ for m tending to infinity. $\square$

Chapter 6

Appendix

In the appendix we show, for which kind of functions f and Q-matrices, in general, $f(X)$ is again a Markov jump process if X is a Markov process with countable state space. The following definition is in the way as [58, Chapter 2.3].

Definition 6.1. *Let Ψ be a countable state space. We call a Borel measurable function $Q : [0, \infty) \times \Psi^2 \to \mathbb{R}$ a time-dependent Q-matrix, if Q is bounded and $Q_t(z, z') \geq 0$ for all $z, z' \in \Psi$ with $z \neq z'$ and $t \geq 0$ and*

$$Q_t(z, z) = -\sum_{z' \in \Psi \setminus \{z\}} Q_t(z, z'), \quad \text{for all } z \in \Psi \text{ and } t \geq 0.$$

Furthermore we define for the time-dependent Q-matrix the generator $\mathcal{Q}$ for each bounded, Borel measurable function $f : \Psi \to \mathbb{R}$ and $t \geq 0$ by

$$\mathcal{Q}(t)f(z) = \sum_{z' \in \Psi} Q_t(z, z')f(z'). \tag{6.2}$$

Ethier and Kurtz prove in [21, Chapter 4, Lemma 7.2] that $\mathcal{Q}$ generates a unique transition function for a Markov jump process and that there is a one-to-one correspondence between the generator and the Markov jump process.

Lemma 6.3. *Let S_1 and S_2 be countable state spaces and $\Phi : S_1 \to S_2$ be a Borel measurable function. Let $Q : [0, \infty) \times S_1^2 \to \mathbb{R}$ be a time-dependent Q-matrix , where for each $\eta, \eta' \in S_2$ and $t \geq 0$*

$$\sum_{z' \in \Phi^{-1}(\eta')} Q_t(z_1, z') = \sum_{z' \in \Phi^{-1}(\eta')} Q_t(z_2, z'), \quad \text{for all } z_1, z_2 \in \Phi^{-1}(\eta). \tag{6.4}$$

Define the generator $\mathcal{Q}$ for the time-dependent Q-matrix Q by (6.2) and suppose $(X_t)_{t\geq 0}$ is the Markov jump process with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$, generated by $\mathcal{Q}$, with state space S_1. We define the function $\tilde{Q} : [0, \infty) \times S_2^2 \to \mathbb{R}$ by

$$\tilde{Q}_t(\eta, \eta') = \sum_{z' \in \Phi^{-1}(\eta')} Q_t(z, z'), \quad \text{for } z \in \Phi^{-1}(\eta), \tag{6.5}$$

if $\Phi^{-1}(\eta) \neq \emptyset$, and by $\tilde{Q}_t(\eta, \eta') = 0$ otherwise.

Then the function $\tilde{Q}$ is well defined and a time-dependent Q-matrix. Furthermore $(\Phi(X_t))_{t\geq 0}$ is a Markov jump process with respect to $(\mathcal{F}_t)_{t\geq 0}$ with state space S_2, corresponding to the time-dependent Q-matrix $\tilde{Q}$.

Proof. The Borel measurability of $\tilde{Q}$ follows from (6.5) and the measurability of Q. Since the non-diagonal entries of $\tilde{Q}$ are the sum over non-diagonal entries of Q or zero, they are non-negative. The definition of $\tilde{Q}$ in (6.5) implies for all $\eta \in S_2$ with $\Phi^{-1}(\eta) \neq \emptyset$ and $t \geq 0$

$$\tilde{Q}_t(\eta,\eta) = \sum_{z' \in \Phi^{-1}(\eta)} Q_t(z,z') = -\sum_{\eta' \in S_2 \setminus \{\eta\}} \sum_{z' \in \Phi^{-1}(\eta')} Q_t(z,z') = -\sum_{\eta' \in S_2 \setminus \{\eta\}} \tilde{Q}_t(\eta,\eta'),$$

where $z \in \Phi^{-1}(\eta)$ arbitrary. For $\eta \in S_2$ with $\Phi^{-1}(\eta) = \emptyset$ this equation is obviously true. Since Q is bounded, there exists a constant $C \geq 0$ such that $|Q_t(z,z')| \leq C$ for all $z, z' \in S_1$ and $t \geq 0$. This implies for each $\eta, \eta' \in S_2$ with $\Phi^{-1}(\eta) \neq \emptyset$ and $t \geq 0$

$$|\tilde{Q}_t(\eta,\eta')| = \left| \sum_{z' \in \Phi^{-1}(\eta')} Q_t(z,z') \right| \leq |Q_t(z,z)| \leq C,$$

and therefore $\tilde{Q}$ is bounded as well. Altogether $\tilde{Q}$ is a time-dependent Q-matrix. Define the bounded generator $\tilde{\mathcal{Q}}$ for the time-dependent Q-matrix $\tilde{Q}$ by (6.2). Let f be a Borel measurable, bounded, real-valued function. Then for each $t \geq 0$

$$\begin{aligned}
\tilde{\mathcal{Q}}(t) f(\Phi(X_t)) &= \sum_{\eta \in S_2} \tilde{Q}_t(\Phi(X_t),\eta) f(\eta) = \sum_{\eta \in S_2} \sum_{z' \in \Phi^{-1}(\eta)} Q_t(X_t,z') f(\Phi(z')) \\
&= \sum_{z' \in S_1} Q_t(X_t,z') f(\Phi(z')) = \mathcal{Q}(t)(f \circ \Phi)(X_t),
\end{aligned}$$

since the preimages of all $\eta \in S_2$ under Φ divide S_1 in disjoint subsets. Hence, we obtain

$$f(\Phi(X_t)) - \int_0^t \tilde{\mathcal{Q}}(s) f(\Phi(X_s))\, ds = f \circ \Phi(X_t) - \int_0^t \mathcal{Q}(s)(f \circ \Phi)(X_s)\, ds.$$

Since the Markov jump process X is generated by $\mathcal{Q}$ and $f \circ \Phi$ is bounded, this is an $\mathcal{F}_t$–martingale, see [21, Chapter 4, Proposition 1.7].

Furthermore, the finite-dimensional distributions of $\Phi(X)$ are uniquely determined and the process $\Phi(X)$ is a Markov process with respect to $(\mathcal{F}_t)_{t \geq 0}$ generated by $\tilde{\mathcal{Q}}$, see [21, Chapter 4, Theorem 7.3 and Theorem 4.2]. □

Part II

Generalization of the Dybvig–Ingersoll–Ross Theorem

Chapter 7

Introduction

To price long-term contracts, like life insurance policies, practitioners model zero-coupon bond prices with long-term maturities to find reasonable discount factors. Empirical investigations of these prices are difficult, since there are only zero-coupon bonds traded with maturity of up to 30 years, and for a life annuity, for example, discount factors for up to 100 years are needed, see e. g. Carriere (1999). To construct reasonable models, we need to know how the long-term zero-coupon rates behave.

Dybvig, Ingersoll and Ross (1996) showed that long-term zero-coupon rates can never fall in an arbitrage-free market under the assumption that the limit of the zero-coupon rates exists. Therefore, if the rates in a model decrease, it is not arbitrage-free. This fundamental theorem is part of textbooks, see e. g. Cairns (2004), and can be used to constrain the parameters of factor models to avoid arbitrage. Yao (1999) and El Karoui, Frachot and Geman (1998) discussed the long-term rates for several well-known models and used the theorem in this context.

Even in well-known interest rate models like the Vašíček model, the Cox–Ingersoll–Ross model or the Gaussian Heath–Jarrow–Morton model, it is possible that the limit of the zero-coupon rates does not exist, as we show with examples in Chapter 9. In this case we cannot use the Dybvig–Ingersoll–Ross theorem to explain the behavior of the long-term zero-coupon rates and decide if the model admits arbitrage opportunities. To assess also models where the limit does not exist we generalize the Dybvig–Ingersoll–Ross theorem in this part of the thesis. We prove that the *limit superior of the zero-coupon rates and the forward rates never fall*, which is called asymptotic monotonicity. From the investor's point of view, the limit superior is the natural extension, because he prefers for long-term investments those zero-coupon bonds which give a high investment return.

In the literature there are two approaches to prove the Dybvig–Ingersoll–Ross theorem. The first approach constructs an arbitrage strategy, if long-term rates fall. Dybvig et al. provide an arbitrage strategy for a general infinite state space in the appendix of their paper. In the case of finitely many states they construct a second arbitrage strategy, which was made rigorous by McCulloch (2000). Recently, Schulze (2007) showed a further arbitrage strategy using another definition of arbitrage than Dybvig et al. The second approach to prove the Dybvig–Ingersoll–Ross theorem is to assume the existence of an equivalent martingale measure. Hubalek, Klein and Teichmann (2002) give a general proof in this setting. We also prove the generalization for the limit superior of the zero-coupon rates and forward rates in these two different ways. For the first approach, we assume a slightly weaker condition than assuming the existence of an equivalent martingale measure. This proof is inspired by the proof of Hubalek et al. For the second approach, we assume that there is no arbitrage opportunity in the limit with vanishing risk, and show again that asymptotic monotonicity holds.

Besides the main theorem, Dybvig et al. showed that the long-term zero-coupon rate equals its minimum future value, if the state space is finite. Using a stricter definition of no-arbitrage, Schulze extended this result to infinite state spaces. Again, the authors assume the existence of the long-term limit. Here we state conditions for asymptotic minimality of the limit superior of the zero-coupon rates. That means, the limit superior of the long-term limit of the zero-coupon rates is the largest random variable, which is known at this time and dominated by the future limit superior of the long-term limit. Examples in Chapter 9 show, that either no arbitrage opportunity in the limit nor the existence of an equivalent martingale measure are sufficient for asymptotic minimality. In both cases we state additional conditions, such that asymptotic minimality holds. No arbitrage opportunity in the limit or the existence of an equivalent martingale measure are not even necessary which is also shown in Chapter 9.

Kardaras and Platen also show the Dybvig–Ingersoll–Ross theorem without assuming the existence of the long-term limit. In contrast to this thesis they use another definition for the limit superior and concentrate on the maximal order that long-term rates at earlier dates can dominate long-term rates at later dates.

The outline of this part is the following. In Chapter 8 we give the general notation, state the main theorem about asymptotic monotonicity, and justify the use of the limit superior from the investor's point of view. Furthermore, we specify conditions for asymptotic minimality and define two notions of an *arbitrage opportunity in the limit.* In Chapter 9, we provide several interest rate models, where the long-term limit of the zero-coupon rates does not exist, to show that our generalization of asymptotic monotonicity is useful. Further examples illustrate the conditions for asymptotic minimality. Chapter 10 contains the proofs of our results. After proving two auxiliary lemmas in Section 10.1, we give in Section 10.2 the proof for asymptotic monotonicity and minimality using an equivalent martingale measure. The proofs using arbitrage arguments are given in Section 10.3.

Chapter 8

Statement of the generalized Dybvig–Ingersoll–Ross theorem

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$ a filtration of $\mathcal{F}$ with a discrete-time parameter $t \in \mathbb{N}_0$ or a continuous-time parameter $t \in [0, \infty)$. For every maturity $T \in \mathbb{N}$ or $T \in (0, \infty)$, respectively, we assume that the corresponding zero-coupon bond price process $P(t, T)$ with $t \in \{0, 1, \dots, T\}$ or $t \in [0, T]$, respectively, is strictly positive and $\mathbb{F}$-adapted with normalization $P(T, T) = 1$.

Define the zero-coupon rate for maturity $T > 0$ in the discrete-time case by

$$R(t,T) := P(t,T)^{-1/(T-t)} - 1, \qquad t \in \{0, 1, \dots, T-1\}, \tag{8.1}$$

and in the continuous-time case by

$$R(t,T) := -\frac{\log P(t,T)}{T-t}\,, \qquad t \in [0,T). \tag{8.2}$$

The arbitrage-free forward rate $F(s, t, T)$ for a loan over the future time period $[t, T]$, contracted at time s, is in the discrete-time case defined by

$$F(s,t,T) := \left(\frac{P(s,t)}{P(s,T)}\right)^{1/(T-t)} - 1, \qquad s,t \in \{0,1,\dots,T-1\},\ s \leq t, \tag{8.3}$$

and in the continuous-time case by

$$F(s,t,T) := \frac{1}{T-t}\,\log\frac{P(s,t)}{P(s,T)}\,, \qquad s,t \in [0,T),\ s \leq t. \tag{8.4}$$

For both time scales we define the long-term spot rate process by

$$l(t) := \limsup_{T\to\infty} R(t,T) = \lim_{n\to\infty}\ \operatorname*{ess\,sup}_{T>n\vee t} R(t,T), \qquad t \geq 0, \tag{8.5}$$

and the long-term forward rate process by

$$l_F(s,t) := \limsup_{T\to\infty} F(s,t,T) = \lim_{n\to\infty}\ \operatorname*{ess\,sup}_{T>n\vee t} F(s,t,T), \qquad 0 \leq s \leq t, \tag{8.6}$$

Remark 8.7. For clarity we want to point out that for each $t \geq 0$ the limit superior $l(t)$ of the zero-coupon rates is the pointwise infimum of $\{R^*_n(t)\}_{n\in\mathbb{N}}$, where each $R^*_n(t)$ denotes the essential supremum of $\{R(t,T)\}_{T>n\vee t}$. The essential supremum is the smallest

$\mathcal{F}_t$-measurable upper bound. That means, $R_n^*(t)$ is an $\mathcal{F}_t$-measurable random variable, $\mathbb{P}(R_n^*(t) \geq R(t,T)) = 1$ for all $T > n \vee t$, and every other random variable X dominating a. s. these zero-coupon rates satisfies $\mathbb{P}(X \geq R_n^*(t)) = 1$. In particular, the essential supremum is uniquely determined up to a set of $\mathbb{P}$-measure zero. The existence of the essential supremum for a collection of random variables is proved, for example, in [22, Appendix A.5]. Note that $\mathbb{P}(R_m^*(t) \geq R_n^*(t)) = 1$ for all $m \leq n$, hence the infimum of $\{R_n^*(t)\}_{n\in\mathbb{N}}$ is $\mathbb{P}$-almost surely equal to the almost surely existing pointwise limit. The limit superior of the forward rates is to be understood in an analogue manner.

In comparison to Dybvig et al. and Hubalek et al., we do not assume that the long-term limits of the zero-coupon rates or the forward rates exist. In Section 9.1 we present (extensions of) popular interest rate models, which need this generalization.

From the investor's point of view, the limit superior of the zero-coupon rates is the natural definition, because he/she prefers for the long-term investment those zero-coupon bonds, which give a high investment return based on the information at time t. The following lemmas (proved in Section 10.1) show that the long-term spot rate $l(t)$ can indeed be approximated by investing in a zero-coupon bond with a suitable maturity, which is chosen based on the information available at time t. Furthermore, $l(t)$ agrees with the long-term forward rates, so it suffices to investigate the behaviour of the long-term spot rates.

Lemma 8.8. *Given $t \geq 0$, there exists a sequence of $\mathcal{F}_t$-measurable random maturities*[1] *$T_n \colon \Omega \to (n \vee t, \infty)$, each one taking only a finite number of values, such that*

$$l(t) \overset{a.s.}{=} \lim_{n\to\infty} R(t, T_n).$$

Lemma 8.9. *The long-term forward and spot rates are almost surely equal, meaning that $l_F(s,t) \overset{a.s.}{=} l(s)$ for all $0 \leq s \leq t$.*

8.1 Results using the existence of a forward risk neutral probability measure

Part of our main results, namely asymptotic monotonicity in Theorem 8.17 and asymptotic minimality in Theorem 8.21 are based on the following two conditions:

Condition 8.10. *We say that this condition holds for times s and t with $0 \leq s < t$, if there exist a probability measure $\mathbb{Q}_{s,t}$ on $(\Omega, \mathcal{F}_t)$, which is equivalent to $\mathbb{P}|_{\mathcal{F}_t}$, and a $T_0 > t$ such that, for all $T \geq T_0$,*

$$P(s,T) \geq P(s,t)\, \mathbb{E}_{\mathbb{Q}_{s,t}}[P(t,T) \,|\, \mathcal{F}_s] \quad a.\,s. \tag{8.11}$$

This condition is sufficient for asymptotic monotonicity. For asymptotic minimality in Theorem 8.21 we need the stronger condition:

Condition 8.12 (Existence of forward (time s) risk neutral probability measure)**.** *We say that this condition holds for times s and t with $0 \leq s < t$, if there exists a probability measure $\mathbb{Q}_{s,t}$ on $(\Omega, \mathcal{F}_t)$, which is equivalent to $\mathbb{P}|_{\mathcal{F}_t}$ such that, for all $T > t$,*

$$P(s,T) \overset{a.s.}{=} P(s,t)\, \mathbb{E}_{\mathbb{Q}_{s,t}}[P(t,T) \,|\, \mathcal{F}_s]\,. \tag{8.13}$$

[1] In the discrete-time setting, the random maturities have to be integer-valued. This also applies to Remark 8.27, Definition 8.29, Theorem 8.34 and its corollaries. Since T_n attains only a finite number of values, $R(t, T_n)$ is $\mathcal{F}_t$-measurable.

We call $\mathbb{Q}_{s,t}$ the forward (time s) risk neutral probability measure for maturity t.

Condition 8.12 says that, simultaneously for all maturities $T > t$, the arbitrage-free forward price $P(s,T)/P(s,t)$, contracted at time s for the T-maturity zero-coupon bond at time t, can be expressed as the $\mathcal{F}_s$-conditional expectation of the price $P(t,T)$ at time t with respect to the measure $\mathbb{Q}_{s,t}$.

Remark 8.14. Suppose a money market account B_t with $t \in \mathbb{N}_0$ or $t \in [0,\infty)$ is given by a strictly positive and $\mathbb{F}$-adapted process. Then the following construction yields a model, where a forward risk neutral probability measure $\mathbb{Q}_{s,t}$ exists simultaneously for all times s and t with $0 \le s < t$. If $\mathbb{Q}$ is a probability measure such that B_0/B_T is $\mathbb{Q}$-integrable for every $T > 0$, then we can define zero-coupon bond prices by

$$P(t,T) = \mathbb{E}_{\mathbb{Q}}\Big[\frac{B_t}{B_T}\,\Big|\,\mathcal{F}_t\Big] = \frac{B_t}{B_0}\,\mathbb{E}_{\mathbb{Q}}\Big[\frac{B_0}{B_T}\,\Big|\,\mathcal{F}_t\Big], \qquad t \in [0,T], \tag{8.15}$$

and the forward (time s) risk neutral probability measure $\mathbb{Q}_{s,t}$ on $(\Omega, \mathcal{F}_t)$ by

$$\frac{d\mathbb{Q}_{s,t}}{d\mathbb{Q}} = \frac{B_s}{P(s,t)B_t}$$

for every $s \in [0,t)$. Since

$$\mathbb{E}_{\mathbb{Q}}\Big[\frac{B_s}{P(s,t)B_t}\,\Big|\,\mathcal{F}_s\Big] \overset{\text{a.s.}}{=} 1,$$

it follows by using Bayes' formula and the tower property, that

$$\mathbb{E}_{\mathbb{Q}_{s,t}}[P(t,T)\,|\,\mathcal{F}_s] \overset{\text{a.s.}}{=} \mathbb{E}_{\mathbb{Q}}\bigg[\frac{B_s}{P(s,t)B_t}\,\mathbb{E}_{\mathbb{Q}}\Big[\frac{B_t}{B_T}\,\Big|\,\mathcal{F}_t\Big]\,\bigg|\,\mathcal{F}_s\bigg] \overset{\text{a.s.}}{=} \frac{P(s,T)}{P(s,t)},$$

hence Condition 8.12 holds for all times s and t with $0 \le s < t$. We will use this construction for the examples in Chapter 9.

Example 8.16. In the discrete-time case, let $\{r_t\}_{t\in\mathbb{N}}$ be an interest rate process, which is $\mathbb{F}$-adapted and $(-1,\infty)$-valued. We define the money market account by

$$B_t = B_0 \prod_{i=1}^{t} (1 + r_i), \qquad t \in \mathbb{N}_0,$$

where B_0 is strictly positive and $\mathcal{F}_0$-measurable. For a probability measure $\mathbb{Q}$ such that B_0/B_T is $\mathbb{Q}$-integrable for every $T \in \mathbb{N}$, we define the corresponding zero-coupon bond prices by (8.15), which means

$$P(t,T) = \mathbb{E}_{\mathbb{Q}}\bigg[\prod_{i=t+1}^{T} \frac{1}{1+r_i}\,\bigg|\,\mathcal{F}_t\bigg], \qquad t \in \{0,1,\dots,T\}.$$

By Remark 8.14, a forward risk neutral probability measure exists in this model simultaneously for all times $s, t \in \mathbb{N}_0$ with $s < t$.

The following result, which we prove in Section 10.2, states that the long-term spot and forward rates, given by (8.5) and (8.6), respectively, almost surely never fall. This is also called *asymptotic monotonicity.* Under the assumption, that the long-term limits of the spot and forward rates exist, this is the so-called Dybvig–Ingersoll–Ross theorem.

Economically, from time s to a later time t, the available information increases, so a more informed decision concerning the best zero-coupon bonds for long-term investments can be made. However, to take advantage of this additional information, the gains during $[s,t]$ on zero-coupon bonds with a large maturity T should be negligible compared to the total gains until T, at least in the limit $T \to \infty$, see Example 8.18 for a counterexample. Therefore, in a reasonable economic environment as specified by Condition 8.10, the long-term spot rate at time t should be greater than the long-term spot rate at time s.

Theorem 8.17. *If Condition 8.10 holds for times s and t with $0 \le s < t$, then*

(i) $l(s) \le l(t)$ a. s. and

(ii) $l_F(s,s') \le l_F(t,t')$ a. s. for all $s' \ge s$ and $t' \ge t$.

Examples 9.16, 9.20 and 9.22 show that the inequalities can be strict everywhere on Ω.

Example 8.18. Given $0 \le s < t$, the deterministic, continuous-time example with $P(s,T) = e^{-(T-s)}$ for all $T \ge s$ and $P(t,T) = 1$ for all $T \ge t$ shows, that $l(s) = 1 > l(t) = 0$ can happen, if there is arbitrage by investing in the zero-coupon bonds with maturity $T > t$. To exploit the arbitrage in this example, sell at time s one t-maturity bond and buy e^{T-t} zero-coupon bonds of maturity T with $T > t$.

Asymptotic monotonicity raises the question, whether $l(s)$ is the largest $\mathcal{F}_s$-measurable random variable, which is almost surely dominated by $l(t)$. To discuss this question, we need the following definition.

Definition 8.19. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. For an $\overline{\mathbb{R}}$-valued random variable X, we define the upper $\mathcal{G}$-measurable envelope $X^{\mathcal{G}}$ as the essential infimum of all $\overline{\mathbb{R}}$-valued, $\mathcal{G}$-measurable random variables Z with $Z \ge X$ a. s. Similarly, we define the lower $\mathcal{G}$-measurable envelope $X_{\mathcal{G}}$ as the essential supremum of all $\overline{\mathbb{R}}$-valued, $\mathcal{G}$-measurable Z with $Z \le X$ a. s.*

Observe that $X_{\mathcal{G}} \le X \le X^{\mathcal{G}}$ a. s., and asymptotic monotonicity implies $l(s) \le l(t)_{\mathcal{F}_s}$ a. s. Note that even in case of convergence of the zero-coupon rates $R(t,T)$ to $l(t)$ as $T \to \infty$, the existence of a forward risk neutral probability measure does not imply asymptotic minimality in the sense that $l(s) \overset{\text{a.s.}}{=} l(t)_{\mathcal{F}_s}$, as Example 9.22 shows. In this example of a stochastic interest rate model, the long-term spot rate l jumps up from 0 to 1 at time $t = 1$ with probability 1. The following, purely analytical condition is a convenient additional assumption for proving asymptotic minimality in Theorem 8.21 below.

Definition 8.20. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. An $\overline{\mathbb{R}}$-valued random variable X is said to dominate the random variables $\{X_t\}_{t>0}$ in the $(\mathcal{G}, \mathbb{P})$-superexponential sense along a $\mathcal{G}$-measurable subsequence, if*[2]

$$\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{E}\big[(\max\{X_t - X, 0\})^t \,|\, \mathcal{G}\big] \overset{a.s.}{=} -\infty.$$

Theorem 8.21 (Asymptotic minimality)**.** *Let Condition 8.12 be satisfied for times s and t with $0 \le s < t$. Assume in addition that the upper $\mathcal{F}_s$-measurable envelope $V_t^{\mathcal{F}_s}$ of the*

[2] We use here the convention $\log 0 = -\infty$. Analogously to (8.5) and (8.6), the limit inferior is the limit as $n \to \infty$ of the essential infima over all $t > n$.

limiting annual discount factor at time t given by[3]

$$V_t = \begin{cases} 1/(l(t)+1) & \text{in the discrete-time case,} \\ \exp(-l(t)) & \text{in the continuous-time case,} \end{cases} \tag{8.22}$$

dominates $\{P(t,t+u)^{1/u}\}_{u>0}$ in the $(\mathcal{F}_s,\mathbb{Q})$-superexponential sense along an $\mathcal{F}_s$-measurable subsequence, which means that

$$\liminf_{T\to\infty} \bigl(\mathbb{E}_{\mathbb{Q}}\bigl[\max\{P(t,T)-(V_t^{\mathcal{F}_s})^{T-t},0\}\,\big|\,\mathcal{F}_s\bigr]\bigr)^{1/(T-t)} \overset{a.s.}{=} 0. \tag{8.23}$$

Then $l(s) \overset{a.s.}{=} l(t)_{\mathcal{F}_s}$ and $l_F(s,s') = l_F(t,t')_{\mathcal{F}_s}$ a. s. for all $s' \geq s$ and $t' \geq t$.

Remark 8.24. For asymptotic minimality we cannot weaken the requirements for the probability measure, because we use the probability measure in Condition 8.10 to show asymptotic monotonicity, but we need also the reversed inequality

$$P(s,T) \leq P(s,t)\,\mathbb{E}_{\mathbb{Q}_{s,t}}[P(t,T)\,|\,\mathcal{F}_s] \qquad \text{a. s.},$$

for the estimate in (10.15) in the proof of Theorem 8.21.

Remark 8.25. Note that $V_t^{\mathcal{F}_s} \overset{\text{a.s.}}{=} 1/(l(t)_{\mathcal{F}_s}+1)$ and $V_t^{\mathcal{F}_s} \overset{\text{a.s.}}{=} \exp(-l(t)_{\mathcal{F}_s})$, respectively, and since

$$V_t \overset{\text{a.s.}}{=} \begin{cases} \liminf_{T\to\infty} \frac{1}{R(t,T)+1} & \text{in the discrete-time case,} \\ \liminf_{T\to\infty} \exp(-R(t,T)) & \text{in the continuous-time case,} \end{cases}$$

we obtain

$$V_t \overset{\text{a.s.}}{=} \liminf_{T\to\infty} P(t,T)^{1/(T-t)}. \tag{8.26}$$

Remark 8.27. If there exists a sequence $\{T_n\}_{n\in\mathbb{N}}$ of $\mathcal{F}_s$-measurable random times, taking at most countable many values in (t,∞) and tending to infinity as $n\to\infty$, such that for every $\varepsilon>0$ there exists $n_\varepsilon \in \mathbb{N}$ satisfying

$$P(t,T_n)^{1/(T_n-t)} \leq V_t^{\mathcal{F}_s} + \varepsilon \qquad \text{a. s.} \tag{8.28}$$

for all $n \geq n_\varepsilon$, then (8.23) holds. Due to (8.26) and $V_t \leq V_t^{\mathcal{F}_s}$, this uniformity certainly holds for all $s\in[0,t]$ simultaneously when $\mathcal{F}_t$ is finite and the limit inferior in (8.26) is attained along a deterministic sequence $\{T_n\}_{n\in\mathbb{N}}$. The latter condition in turn is satisfied when $l(t) = \lim_{n\to\infty} R(t,T_n)$ a. s.

8.2 Results using different notions for absence of arbitrage

In the last section we assumed the existence of a forward risk neutral probability measure, resp. that Condition 8.10 holds. The second approach uses no-arbitrage arguments to show asymptotic monotonicity and minimality. The next definition gives two different notions of arbitrage and applies to discrete as well as continuous time. It is inspired by the definition of *arbitrage in the limit,* which is used by Schulze, and the definition of arbitrage used in Dybvig et al.

[3]We use here the conventions $1/0=\infty$, $1/\infty=0$, $\exp(\infty)=\infty$, and $\exp(-\infty)=0$.

Definition 8.29. *Given times $0 \le s < t$, the zero-coupon bonds with maturity $T \ge t$ provide an* arbitrage opportunity in the limit *for times s and t, if there exist a sequence $\{(\varphi_n, \psi_n)\}_{n\in\mathbb{N}}$ of $\mathcal{F}_s$-measurable, $\mathbb{R}^2$-valued portfolio compositions and a sequence $\{T_n\}_{n\in\mathbb{N}}$ of $\mathcal{F}_s$-measurable random maturities $T_n\colon \Omega \to (n \vee t, \infty)$, each one taking only a finite number of values, such that*

(i) $V_n(s) := \varphi_n P(s, T_n) + \psi_n P(s,t) \overset{a.s.}{=} 0$ *for all* $n \in \mathbb{N}$,

(ii) $\mathbb{P}(\liminf_{n\to\infty} V_n(t) > 0) > 0$, *where* $V_n(t) := \varphi_n P(t, T_n) + \psi_n$, *and*

(iii) $\liminf_{n\to\infty} V_n(t) \ge 0$ *a. s.*

We say that the zero-coupon bonds provide an arbitrage opportunity in the limit with vanishing risk *for times s and t, if (iii) is replaced by*

(iv) for every $\varepsilon > 0$ there exists $n_\varepsilon \in \mathbb{N}$ such that $V_n(t) \ge -\varepsilon$ a. s. for all $n \ge n_\varepsilon$.

Remark 8.30. Part (i) in Definition 8.29 always holds if $\psi_n := -\varphi_n P(s, T_n)/P(s,t)$ for all $n \in \mathbb{N}$.

Remark 8.31. Since (iv) implies (iii), the assumption of *no arbitrage opportunity in the limit* is stronger than *no arbitrage opportunity in the limit with vanishing risk.* If $\mathcal{F}_t$ is finite, then pointwise implies uniform convergence, hence (iii) implies (iv) and both notions of arbitrage are equivalent. Example 9.26 below shows that even the stronger assumption of *no arbitrage opportunity in the limit* does not imply the existence of a forward risk neutral probability measure in Condition 8.12. Even the weaker Condition 8.10 does not hold in this example.

Lemma 8.32 below shows that Condition 8.12 implies the weaker no-arbitrage condition, which by Theorem 8.33 is sufficient for asymptotic monotonicity. Actually, the no-arbitrage condition can be further weakened by excluding only arbitrage due to a positive investment in the long-term zero-coupon bonds. The lemma and the following theorems are proved in Section 10.3.

Lemma 8.32. *If there exists a forward time s risk neutral probability measure for maturity t as in Condition 8.12 with $0 \le s < t$, then there is no arbitrage opportunity in the limit with vanishing risk for times s and t.*

Theorem 8.33. *Consider times $0 \le s < t$. Assume that there is no arbitrage opportunity in the limit with vanishing risk for times s and t in the sense of Definition 8.29 by investing in the long-term zero-coupon bonds (with $\varphi_n \ge 0$ for all $n \in \mathbb{N}$). Then $l(s) \le l(t)$ a. s. and $l_F(s, s') \le l_F(t, t')$ a. s. for all $s' \ge s$ and $t' \ge t$.*

For the remaining results, we need the stronger assumption of *no arbitrage opportunity in the limit*, however, for Theorem 8.34 below we only have to exclude this limiting arbitrage by short-selling of the long-term zero-coupon bonds. The heuristic justification of the following theorem is as follows: If, with strictly positive probability, the worst long-term spot rate, which we will incur by placing our investment orders for time t already at an earlier time s based on the information available at s, is strictly larger than the best long-term spot rate we can earn by investing already at time s, then the prices of the long-term zero-coupon bonds must fall substantially during $[s,t]$, offering an arbitrage possibility in the limit by short-selling these bonds.

Theorem 8.34. *Consider times $0 \le s < t$. Assume that there is no arbitrage opportunity in the limit for times s and t in the sense of Definition 8.29 by short-selling the long-term zero-coupon bonds (with $\varphi_n \le 0$ for all $n \in \mathbb{N}$). Then for every sequence $\{T_n\}_{n\in\mathbb{N}}$ of $\mathcal{F}_s$-measurable random maturities $T_n\colon \Omega \to (n \vee t, \infty)$, each one taking only finitely many values,*

$$\Big(\liminf_{n\to\infty} R(t,T_n)\Big)_{\mathcal{F}_s} \le l(s) \quad \textit{a. s.} \tag{8.35}$$

Examples 9.3 and 9.20 show that, for certain sequences of random (or even deterministic) maturities, the inequality in (8.35) can be strict everywhere on Ω. In these models, the limit of $R(t,T_n)$ as $n \to \infty$ does not exist. Note that the deterministic model of Example 8.18, which admits an arbitrage possibility by investing in the zero-coupon bonds with maturity $T > t$, satisfies the assumptions of Theorem 8.34.

Using the definition of the long-term zero-coupon rate $l(t)$ from (8.5), the $\mathcal{F}_s$-measurability of $l(s)$ and the definition of the lower $\mathcal{F}_s$-measurable envelope in Definition 8.19, we obtain from Theorems 8.33 and 8.34:

Corollary 8.36 (Asymptotic minimality)**.** *Consider $0 \le s < t$. If there is no arbitrage opportunity in the limit for s and t in the sense of Definition 8.29, then*

$$\Big(\liminf_{T\to\infty} R(t,T)\Big)_{\mathcal{F}_s} \le l(s) \le \Big(\limsup_{T\to\infty} R(t,T)\Big)_{\mathcal{F}_s} \quad \textit{a. s.} \tag{8.37}$$

In particular, if $\lim_{T\to\infty} R(t,T)$ exists a. s., then $l(s) \overset{a.s.}{=} l(t)_{\mathcal{F}_s}$.

If the limit of $R(t,T)$ as $T \to \infty$ does not exist a. s., we might still get asymptotic minimality. Using asymptotic monotonicity and Theorem 8.34, each sequence $\{T_n\}_{n\in\mathbb{N}}$ of $\mathcal{F}_s$-measurable random maturities, each one taking only finitely many values, satisfies

$$\Big(\liminf_{n\to\infty} R(t,T_n)\Big)_{\mathcal{F}_s} \le l(s) \le l(t)_{\mathcal{F}_s}, \quad \text{a. s.},$$

if there is no arbitrage opportunity in the limit. If a special sequence of maturities satisfies additionally the reversed inequality, we have asymptotic minimality. Note that the sequence from Lemma 8.8 cannot be used in general, because these maturities are only $\mathcal{F}_t$-measurable.

Corollary 8.38 (Asymptotic minimality)**.** *Consider $0 \le s < t$. If there is no arbitrage opportunity in the limit for times s and t in the sense of Definition 8.29 and if there exists a sequence $\{T_n\}_{n\in\mathbb{N}}$ of $\mathcal{F}_s$-measurable random maturities $T_n\colon \Omega \to (n \vee t, \infty)$, each one taking only finitely many values, such that*

$$l(t)_{\mathcal{F}_s} \le \liminf_{n\to\infty} R(t,T_n) \quad \textit{a. s.,} \tag{8.39}$$

then $l(s) \overset{a.s.}{=} l(t)_{\mathcal{F}_s}$.

Remark 8.40. In Corollaries 8.36 and 8.38, it is actually sufficient to assume that there is no arbitrage opportunity in the limit with vanishing risk for times s and t by investing in the long-term zero-coupon bonds (with $\varphi_n \ge 0$ for all $n \in \mathbb{N}$) and that there is no arbitrage opportunity in the limit by short-selling the long-term zero-coupon bonds (with $\varphi_n \le 0$ for all $n \in \mathbb{N}$).

Remark 8.41. Using the almost sure equivalence of the long-term spot and forward rates in Lemma 8.9, we can also transfer Theorem 8.34 and its corollaries to the long-term forward rates. We refrain from spelling out the details.

Remark 8.42. We could relax Definition 8.29(iv) to

(v) there exists $n_0 \in \mathbb{N}$, such that the negative parts $V_n^-(t) := \max\{0, -V_n(t)\}$ for all $n \geq n_0$ are uniformly integrable and $\liminf_{n\to\infty} V_n(t) \geq 0$ a. s.

and get more limiting arbitrage opportunities in this way. This would strengthen the no-arbitrage assumption. However, using a more general version of Fatou's lemma for conditional expectations[4], the proof of Lemma 8.32 carries over, where the existence of a forward risk neutral probability measure for times s and t in Condition 8.12 implies no arbitrage in the limit with vanishing risk. Hence this Condition is still stronger. Therefore, this stronger no-arbitrage assumption would not be strong enough to imply asymptotic minimality, as Example 9.22 illustrates. In particular, the limiting arbitrage strategies given there cannot satisfy (v).

Remark 8.43. The proofs of the above theorems, corollaries and lemmas do not use path properties of the processes $\{P(t,T)\}_{0\leq t\leq T}$ (like being càdlàg or a semimartingale), and we also do not need a bank account process or additional assumptions on the filtration $\mathbb{F}$ (like containing all null sets of $\mathcal{F}$ or being right-continuous). Furthermore, we allow for $\mathbb{P}(P(t,T) > 1) > 0$, which can happen for models with negative interest rates like the Vašíček model or the Heath–Jarrow–Morton model.

[4]See *Fatou's lemma* at en.wikipedia.org/wiki/, version of October 11, 2008.

Chapter 9

Examples

In this chapter we show with illustrative examples first, that the long-term zero-coupon rates do not always exist and our generalization of the Dybvig–Ingersoll–Ross theorem is therefore useful. In these examples asymptotic monotonicity holds for the limit superior of the zero-coupon rates, resp. the forward rates. Four further examples illustrate the asymptotic minimality conditions as explained in Chapter 8. In Example 9.20 we describe a very simple stochastic interest rate model with $\Omega = \{0, 1\}$. Although this model provides no arbitrage opportunity in the limit and a forward risk neutral probability measure exists, asymptotic minimality does not hold. There *does not* exist a deterministic sequence $\{T_n\}_{n\in\mathbb{N}}$ with $T_n \to \infty$ such that (8.39) holds. Therefore, the absence of arbitrage opportunities in the limit for times s and t with $0 \le s < t$ or the existence of a forward risk neutral probability measure is not sufficient for asymptotic minimality in the sense of $l(s) \overset{\text{a.s.}}{=} l(t)_{\mathcal{F}_s}$. These conditions are not even necessary, see Example 9.23. Example 9.24 shows that asymptotic minimality is not an interval property, meaning that for times $0 < s < t < u$ the property $l(s) \overset{\text{a.s.}}{=} l(u)_{\mathcal{F}_s}$ *does not* imply $l(t) \overset{\text{a.s.}}{=} l(u)_{\mathcal{F}_t}$. Furthermore, the example shows that even if there is no arbitrage opportunity in the limit for times s and u, it is possible to have an arbitrage opportunity for times t and u.

All these examples are continuous-time short-rate models, and there exists a forward risk neutral probability measure for all times defined in Condition 8.12 by construction as pointed out in Remark 8.14. Hence by Lemma 8.32, these models do not provide an arbitrage opportunity in the limit with vanishing risk. For a model not satisfying Condition 8.12, see Example 9.26.

The general set-up of these models (with the exception of the last one) is given as follows. For a given $\mathbb{F}$-progressive interest rate intensity process $\{r_t\}_{t\ge 0}$ with locally integrable paths, we define the money market account by

$$B_t = \exp\Big(\int_0^t r_u\,du\Big), \qquad t \in [0, \infty).$$

Assume that $1/B_t$ is $\mathbb{Q}$-integrable for every $t > 0$. Using (8.15), the zero-coupon bond prices are given by

$$P(t,T) = \mathbb{E}_{\mathbb{Q}}\Big[\exp\Big(-\int_t^T r_u\,du\Big)\,\Big|\,\mathcal{F}_t\Big], \qquad 0 \le t \le T. \tag{9.1}$$

Therefore, the definition of $R(t,T)$ in (8.2) implies

$$R(t,T) = -\frac{1}{T-t}\log \mathbb{E}_{\mathbb{Q}}\Big[\exp\Big(-\int_t^T r_u\,du\Big)\,\Big|\,\mathcal{F}_t\Big], \qquad 0 \le t < T. \tag{9.2}$$

9.1 Models where the limits of the zero-coupon rates do not exist

In the following examples we discuss models, where the limits of the zero-coupon rates $R(t,T)$ for $T \to \infty$ do not exist. The idea is to vary the behaviour of the short rate on longer and longer time periods to get oscillating means. We illustrate this with a simple deterministic model and then with two short-rate models having an (exponentially) affine term structure. More specifically, we consider a variant of the familiar Vašíček model with time-dependent coefficients, which was proposed by Vašíček (1977) and Hull and White (1990). Secondly, we study the behaviour of the long-term spot rate in the model of Cox, Ingersoll and Ross (1985) with time-dependent coefficients (but constant dimension).

In both examples the mean level or the volatility of the short rate changes cyclically but decelerates over time. An economical justification for this behaviour can be the dependence on the business cycles, which become longer and longer. So, if the lengths of the business cycles increase exponentially, then the limits of the zero-coupon rates might not exist, as our examples show.

In our last example, we use the well-known Heath–Jarrow–Morton framework, proposed in [27], and choose an oscillating but decaying volatility function for the forward rates such that the limits of the zero-coupon rates do not exist, see Example 9.16 below. Since we specialize to a deterministic volatility function in product form, this example is related to the extended Vašíček model, cf. [46, Section 10.2].

Example 9.3 (Deterministic model)**.** Define the set

$$A = \left[\tfrac{1}{3}, 1\right) \cup \bigcup_{k=0}^{\infty} \left[2^{2k+1}, 2^{2k+2}\right), \tag{9.4}$$

the càdlàg interest rate intensity $r_t = 1_A(t)$ for $t \geq 0$, and the continuous function

$$R_A(t,T) = \frac{1}{T-t} \int_t^T 1_A(u)\, du = \frac{\lambda(A \cap [t,T])}{T-t}, \qquad 0 \leq t < T, \tag{9.5}$$

where λ denotes the Lebesgue measure. Since $\{r_t\}_{t \geq 0}$ is deterministic, (9.2) implies $R(t,T) = R_A(t,T)$ for all $0 \leq t < T$. Note that $T \geq 1$ is a local minimum of $R_A(0,\cdot)$ if and only if there exists $n \in \mathbb{N}_0$ with $T = 2^{2n+1}$. Since

$$\lambda(A \cap [0, 2^{2n+1}]) = \frac{2}{3} + \sum_{k=0}^{n-1} 2^{2k+1} = \frac{2}{3} + 2\frac{4^n - 1}{3} = \frac{2^{2n+1}}{3}, \qquad n \in \mathbb{N}_0,$$

we have $R_A(0, 2^{2n+1}) = 1/3$. Furthermore, $T \geq 2$ is a local maximum of $R_A(0,\cdot)$ if and only if there exists $n \in \mathbb{N}_0$ with $T = 2^{2n+2}$. Since $\lambda(A \cap [0, 2^{2n+2}]) = \lambda(A \cap [0, 2^{2n+1}]) + 2^{2n+1} = 2^{2n+3}/3$, we get $R_A(0, 2^{2n+2}) = 2/3$. Hence, we have $R_A(0,T) \in [1/3, 2/3]$ for all $T \geq 1$, and the interval $[1/3, 2/3]$ is also the set of all accumulation points of $\{R_A(0,T)\}_{T>0}$. Since $|R_A(t,T) - R_A(0,T)| \leq 2t/T$ for $0 \leq t < T$, the latter is also true for $\{R_A(t,T)\}_{T>t}$, in particular the limit of $R(t,T)$ as $T \to \infty$ does not exist. Since $l(t) = \limsup_{T\to\infty} R_A(t,T) = 2/3$ for all $t \in [0,\infty)$, asymptotic minimality holds. This can also be shown by verifying the assumptions of Corollary 8.38. For $t \in [0,\infty)$ and $T_n := 2^{2n+2}$ with $n \in \mathbb{N}_0$ such that $T_n > t$,

$$|R_A(t,T_n) - l(t)| = |R_A(t,T_n) - R_A(0,T_n)| \leq \frac{2t}{2^{2n+2}} \overset{n\to\infty}{\longrightarrow} 0,$$

hence (8.39) is satisfied. Since the model is deterministic, the σ-algebra $\mathcal{F}_t$ is finite. Remark 8.31 implies that no arbitrage opportunity in the limit is equivalent to no arbitrage opportunity in the limit with vanishing risk. Furthermore, the example illustrates that the inequality (8.35) in Theorem 8.34 can be strict, because $l(0) = 2/3$ but $\liminf_{n\to\infty} R(0, T_n) = 1/3$ for $T_n := 2^{2n+1}$ with $n \in \mathbb{N}$. Note that this example can be generalized to an interest intensity process $r_t = a + b1_A(t)$ for $t \geq 0$, where $a, b \in \mathbb{R}$ and $b \neq 0$.

A broad class of interest rate models have an (exponentially) affine term structure, i. e., the price process of a zero-coupon bond with maturity $T > 0$ admits the representation

$$P(t, T) = \exp\bigl(A(t, T) + B(t, T)r_t\bigr), \qquad t \in [0, T),$$

with deterministic real-valued functions A and B, cf. [8, Chapter 22.3]. Hence, the zero-coupon rate process for $T > 0$ is given by

$$R(t, T) = -\frac{A(t, T) + B(t, T)r_t}{T - t}, \qquad t \in [0, T). \tag{9.6}$$

Therefore, if for $t \geq 0$ the short rate r_t is not deterministic, then the limit of $\{R(t, T)\}_{T>t}$ exists a. s. if and only if the limits of $A(t, T)/T$ and $B(t, T)/T$ for $T \to \infty$ exist. In the following we consider generalizations of the familiar Vašíček and Cox–Ingersoll–Ross models, which both belong to the (exponentially) affine term structure models. In these generalized models we show that, with appropriate choices of time-dependent coefficients, the limit of $A(t, T)/T$ as $T \to \infty$ does not exist.

Example 9.7 (Vašíček model with time-dependent coefficients)**.** Let $\alpha > 0$ be a parameter for the mean reverting strength. Suppose the mean level μ: $[0, \infty) \to \mathbb{R}$ is a locally integrable function and the volatility σ: $[0, \infty) \to \mathbb{R}$ is a locally square-integrable function. Let $\{W_t\}_{t\geq 0}$ be a standard Brownian motion under $\mathbb{Q}$, and let the initial value r_0 be normally distributed (possibly with zero variance) and independent of the Brownian motion. Define the interest rate intensity process by

$$r_t = e^{-\alpha t}\Bigl(r_0 + \alpha \int_0^t e^{\alpha s}\mu_s \, ds + \int_0^t e^{\alpha s}\sigma_s \, dW_s\Bigr), \qquad t \geq 0. \tag{9.8}$$

Using Itô's formula, it follows that $\{r_t\}_{t\geq 0}$ is a strong solution of the stochastic differential equation

$$dr_t = \alpha(\mu_t - r_t)\, dt + \sigma_t \, dW_t, \qquad t \geq 0,$$

with initial value r_0. Note that $\{r_t\}_{t\geq 0}$ is a Gaussian process with continuous paths, see e. g. [2, Chapter 8]. It follows from (9.8) that

$$r_u = e^{-\alpha(u-t)} r_t + \alpha \int_t^u e^{-\alpha(u-s)}\mu_s \, ds + \int_t^u e^{-\alpha(u-s)}\sigma_s \, dW_s, \qquad 0 \leq t \leq u,$$

hence the conditional distribution of the integral $I_{t,T} = \int_t^T r_u \, du$ given r_t is a normal one. In particular, the process $\{r_t\}_{t\geq 0}$ is Markovian and (9.2) simplifies to

$$R(t, T) = -\frac{1}{T - t} \log \mathbb{E}_{\mathbb{Q}}\bigl[\exp\bigl(-I_{t,T}\bigr) \,\big|\, r_t\bigr], \qquad 0 \leq t < T. \tag{9.9}$$

Using the stochastic Fubini theorem, see e. g. Protter (2004), we obtain

$$\begin{aligned}I_{t,T} - r_t \int_t^T e^{-\alpha(u-t)}\,du - \alpha \int_t^T \mu_s \int_s^T e^{-\alpha(u-s)}\,du\,ds \\ = \int_t^T \sigma_s \underbrace{\int_s^T e^{-\alpha(u-s)}\,du}_{=(1-e^{-\alpha(T-s)})/\alpha}\,dW_s, \qquad 0 \le t \le T.\end{aligned}$$

Since the stochastic integral on the right-hand side is independent of r_t with zero expectation, it follows that

$$\mathbb{E}_{\mathbb{Q}}\big[I_{t,T}\,\big|\,r_t\big] = r_t\,\frac{1-e^{-\alpha(T-t)}}{\alpha} + \int_t^T \big(1-e^{-\alpha(T-s)}\big)\mu_s\,ds, \qquad 0 \le t \le T,$$

and, using the Itô isometry,

$$\mathrm{Var}_{\mathbb{Q}}(I_{t,T}\,|\,r_t) = \frac{1}{\alpha^2}\int_t^T \big(1-e^{-\alpha(T-s)}\big)^2\sigma_s^2\,ds, \qquad 0 \le t \le T.$$

If X has a normal distribution, then $\log\mathbb{E}\big[e^{-X}\big] = -\,\mathbb{E}[X] + \frac{1}{2}\,\mathrm{Var}(X)$. Applying these results to (9.9) leads to

$$\begin{aligned}R(t,T) = r_t\,\frac{1-e^{-\alpha(T-t)}}{\alpha(T-t)} + \frac{1}{T-t}\int_t^T \big(1-e^{-\alpha(T-s)}\big)\mu_s\,ds \\ - \frac{1}{2\alpha^2(T-t)}\int_t^T \big(1-e^{-\alpha(T-s)}\big)^2\sigma_s^2\,ds, \qquad 0 \le t < T.\end{aligned} \tag{9.10}$$

Given $t \ge 0$, the limit of the zero-coupon rates $\{R(t,T)\}_{T>t}$ exists in $\mathbb{R}$ if and only if the limit of the difference of the last two terms in (9.10) exists in $\mathbb{R}$ as $T \to \infty$. It remains to choose suitable time-dependent functions for the mean level μ or the volatility σ such that this is not the case. Let us discuss three specific choices.

If μ is bounded and $\lim_{s\to\infty}\sigma_s^2 = \infty$, then, for every $n \in \mathbb{N}$, there exists $T_n \ge t$ such that $\sigma_s^2 \ge n$ for all $s \ge T_n$. Since $1 - e^{-\alpha(T-s)} \ge 1/2$ for $s \le T - 1/\alpha$, we get for all $T \ge T_n + 1/\alpha$,

$$\frac{4}{T-t}\int_t^T \big(1-e^{-\alpha(T-s)}\big)^2\sigma_s^2\,ds \ge \frac{1}{T-t}\int_{T_n}^{T-1/\alpha}\sigma_s^2\,ds \ge n\frac{T-T_n-1/\alpha}{T-t} \overset{T\to\infty}{\longrightarrow} n.$$

Hence $l(t) = \limsup_{T\to\infty} R(t,T) = -\infty$ by (9.10) and, in particular, asymptotic minimality holds. Note, that (8.39) is satisfied. A similar argumentation shows that $l(t) = \pm\infty$ if σ is bounded and $\lim_{s\to\infty}\mu_s = \pm\infty$, respectively.

We now discuss cases where the mean level μ and the volatility σ remain bounded. Note that, for every $a > 0$ and bounded measurable function g: $[0,\infty) \to \mathbb{R}$,

$$\left|\int_t^T e^{-a(T-s)}g(s)\,ds\right| \le \frac{1-e^{-a(T-t)}}{a}\|g\|_\infty \le \frac{\|g\|_\infty}{a}, \qquad 0 \le t \le T.$$

Therefore, if follows from (9.10) that, for every $t \ge 0$,

$$R(t,T) = \frac{1}{T-t}\int_t^T \mu_s\,ds - \frac{1}{2\alpha^2(T-t)}\int_t^T \sigma_s^2\,ds + O\Big(\frac{1}{T}\Big) \tag{9.11}$$

as $T > t$ tends to infinity.

We first consider a constant volatility $\sigma \in \mathbb{R}$ and a time-dependent mean level $\mu_s := a + b1_A(s)$ for $s \geq 0$ with A given by (9.4), $a \in \mathbb{R}$ and $b > 0$. Using (9.11) we obtain, for every $t \geq 0$,

$$R(t,T) = a + b R_A(t,T) - \frac{\sigma^2}{2\alpha^2} + O\Big(\frac{1}{T}\Big) \qquad \text{as } T \to \infty,$$

with $R_A(t,T)$ given by (9.5). It is shown in Example 9.3 that the limit of $R_A(t,T)$ as $T \to \infty$ does not exist, hence the limit of $\{R(t,T)\}_{T>t}$ does not exist either. Since $\limsup_{T\to\infty} R_A(t,T) = 2/3$ by the results from Example 9.3, we see that

$$l(t) = \limsup_{T\to\infty} R(t,T) = a + \frac{2b}{3} - \frac{\sigma^2}{2\alpha^2}, \qquad t \geq 0,$$

hence asymptotic minimality holds for all $0 \leq s < t$. A similar result can be obtained, if we choose a constant mean level $\mu \in \mathbb{R}$ and a time-dependent volatility $\sigma_s := a + b1_A(s)$ with $a, b \in \mathbb{R}$ satisfying $2ab + b^2 \neq 0$.

To illustrate explicitly that a bounded, continuously varying volatility function σ can also lead to oscillating zero-coupon rates, we consider a constant mean level $\mu \in \mathbb{R}$ and a volatility function of the form

$$\sigma_t = \sqrt{a + b\sin(\log(t+1)) + b\cos(\log(t+1))}, \qquad t \geq 0,$$

with $a, b \in (0,\infty)$ satisfying $a \geq \sqrt{2}\,b$. Then $0 \leq \sigma_t \leq \sqrt{2a}$ for all $t \geq 0$. Furthermore, for all $T > 0$ and $t \in [0,T)$,

$$\frac{1}{T-t}\int_t^T \sigma_s^2\,ds = a + b\,\frac{(T+1)\sin(\log(T+1)) - (t+1)\sin(\log(t+1))}{T-t}. \tag{9.12}$$

Together with (9.11) we obtain for the long-term spot rate process

$$l(t) = \limsup_{T\to\infty} R(t,T) = \mu - \frac{a-b}{2\alpha^2}, \qquad t \geq 0, \tag{9.13}$$

but for the limes inferior of the zero coupon rates

$$\liminf_{T\to\infty} R(t,T) = \mu - \frac{a+b}{2\alpha^2}, \qquad t \geq 0.$$

Hence, the limit of $\{R(t,T)\}_{T>t}$ as $T \to \infty$ does not exist. Since the long-term spot-rate process given by (9.13) is a deterministic constant, asymptotic minimality holds for all times $0 \leq s < t$.

Example 9.14 (Cox–Ingersoll–Ross model with time-dependent coefficients)**.**
Let α: $[0,\infty) \to (0,\infty)$ and β: $[0,\infty) \to (-\infty,0)$ be two continuously differentiable functions and let $\{W_t\}_{t\geq 0}$ be a standard Brownian motion under $\mathbb{Q}$. Analogously to [46, Sections 10.3.2 and 10.3.3], by considering a squared Bessel process of dimension $\delta \in (0,\infty)$ with respect to some probability measure $\mathbb{P}$, applying a suitable measure change to $\mathbb{Q}$ using Girsanov's theorem, and rescaling the state space by the function α, we can construct an interest rate intensity process $\{r_t\}_{t\geq 0}$ which solves the stochastic differential equation

$$dr_t = \Big(\delta\alpha(t) + \Big(2\beta(t) + \frac{\alpha'(t)}{\alpha(t)}\Big) r_t\Big)\,dt + 2\sqrt{\alpha(t) r_t}\,dW_t, \qquad t \geq 0,$$

with deterministic initial value $r_0 \geq 0$. If for given $0 \leq t < T$ there is a solution $F_T : [t,T] \to \mathbb{R}$ to the Riccati equation

$$F_T^2(u) + F_T'(u) = 2\alpha(u) + \beta^2(u) + \beta'(u), \qquad u \in [t,T],$$

with the terminal condition $F_T(T) = \beta(T)$, then it follows as in [46, Sections 10.3.3 and 10.3.4] that the corresponding zero-coupon rate is given by

$$R(t,T) = -\frac{1}{2(T-t)}\biggl(\frac{F_T(t)-\beta(t)}{\alpha(t)}r_t + \delta\int_t^T \bigl(F_T(u)-\beta(u)\bigr)\,du\biggr),$$

which corresponds to (9.6) resulting from an (exponentially) affine term structure.

We now make specific choices for α and β. For $b > 0$ and $a > \sqrt{2}b$ define the function

$$\beta(t) = -a + b\sin(\log(t+1)) + b\cos(\log(t+1)), \qquad t \geq 0.$$

Note that β is continuously differentiable and that $-a - \sqrt{2}b \leq \beta(t) < 0$ for all $t \geq 0$. Furthermore, for $c > 0$ with $c^2 > (a + \sqrt{2}b)^2 + \sqrt{2}b$, define the function

$$\alpha(t) = \frac{1}{2}\bigl(c^2 - \beta^2(t) - \beta'(t)\bigr), \qquad t \geq 0.$$

Since $\beta^2(t) \leq (a+\sqrt{2}b)^2$ and $\beta'(t) \leq \sqrt{2}b$, it follows that $\alpha(t) > 0$ for all $t \geq 0$. For these functions α and β, the Riccati equation simplifies, for each $T > 0$, to

$$F_T^2(u) + F_T'(u) = c^2, \qquad u \in [0,T].$$

The solution for the terminal condition $F_T(T) = \beta(T)$ is

$$F_T(u) = c\tanh(cu + g_T), \qquad u \in [0,T],$$

where $g_T := \operatorname{artanh}(\beta(T)/c) - cT$. Since $|\beta(T)| \leq a + \sqrt{2}b < |c|$ for all $T > 0$, the area tangents hyperbolicus of $\beta(T)/c$ is well-defined and bounded with respect to T. Note that $\frac{d}{dx}\log(\cosh(cx+g_T)) = c\tanh(cx+g_T)$ for all $x \in \mathbb{R}$. Therefore,

$$\int_t^T F_T(u)\,du = \log\frac{\cosh\bigl(\operatorname{artanh}\frac{\beta(T)}{c}\bigr)}{\cosh\bigl(\operatorname{artanh}\bigl(\frac{\beta(T)}{c}\bigr) - c(T-t)\bigr)}, \qquad 0 \leq t \leq T.$$

Using $\cosh x = (e^x + e^{-x})/2$ for $x \in \mathbb{R}$ and the boundedness of $\operatorname{artanh}(\beta(T)/c)$, it follows that

$$\lim_{T\to\infty}\frac{1}{T-t}\int_t^T F_T(u)\,du = -c, \qquad t \geq 0.$$

Finally, integration of β, cf. (9.12), yields

$$l(t) = \limsup_{T\to\infty} R(t,T) = \frac{\delta}{2}(c-a+b), \qquad t \geq 0 \tag{9.15}$$

but

$$\liminf_{T\to\infty} R(t,T) = \frac{\delta}{2}(c-a-b), \qquad t \geq 0.$$

Since l in (9.15) is a deterministic constant, asymptotic minimality holds.

Our next example is the well-known Gaussian Heath–Jarrow–Morton model, cf. [46, Chapter 11], with deterministic but time-dependent volatility of the forward rates. For this volatility we choose a non-negative function, which fluctuates over time but converges to zero when the maturities tend to infinity. Again, we assume, that the volatility varies with the business cycles of exponentially increasing lengths.

Example 9.16 (Gaussian Heath–Jarrow–Morton model)**.** Let $\sigma_1, \sigma_2\colon [0,\infty) \to \mathbb{R}$ denote bounded measurable functions. Define the volatility $\sigma\colon [0,\infty)^2 \to \mathbb{R}$ of the forward rates by $\sigma(u,v) = \sigma_1(u)\sigma_2(v)$ for all $u, v \geq 0$. Suppose that the deterministic forward rate curve $f(0,\cdot)\colon [0,\infty) \to \mathbb{R}$ at time zero is locally integrable. We set up the model directly using the spot martingale measure $\mathbb{Q}$, under which zero-coupon bond prices, discounted by the bank account process, are martingales. Therefore, let $\{W_t\}_{t\geq 0}$ be a standard Brownian motion under $\mathbb{Q}$ and let the forward rates satisfy

$$f(t,T) = f(0,T) + \int_0^t \sigma(u,T)\sigma^*(u,T)\,du + \int_0^t \sigma(u,T)\,dW_u, \qquad 0 \leq t \leq T,$$

with integrated volatility $\sigma^*(u,T) := \int_u^T \sigma(u,v)\,dv$ so that they obey the Heath–Jarrow–Morton drift condition. The short-term interest rate intensity process is given by $r_t = f(t,t)$ for $t \geq 0$. Then, for each maturity $T > 0$ and time $t \in [0,T)$, the zero-coupon rate is given by

$$R(t,T) = \frac{1}{T-t}\biggl(\int_0^T f(0,u)\,du - \int_0^t \bigl(r_u - \frac{1}{2}(\sigma^*(u,T))^2\bigr)\,du + \int_0^t \sigma^*(u,T)\,dW_u\biggr),$$

see e. g. [46, Chapter 11, pp. 388–389].

Given $t \geq 0$, the limit of the zero-coupon rates $\{R(t,T)\}_{T>t}$ as $T \to \infty$ might not exist, if the averages of the initial forward rates $\{f(0,u)\}_{u\in[t,T]}$ do not converge, see Examples 9.3 and 9.7 for such functions. In the following, we therefore assume the existence of the limit of these averages so that we can define

$$f^* = \lim_{T\to\infty} \frac{1}{T}\int_0^T f(0,u)\,du.$$

To further discuss the limiting behaviour of the zero-coupon rates, we first consider their stochastic component. Substituting the stochastic integral from the short-rate r_v into the formula for $R(t,T)$ and using the stochastic Fubini theorem, we obtain

$$\begin{aligned}\frac{1}{T-t}\biggl(\int_0^t \sigma^*(u,T)\,dW_u - \int_0^t\int_0^v \sigma(u,v)\,dW_u\,dv\biggr)& \\ = \frac{1}{T-t}\int_t^T \sigma_2(v)\,dv \int_0^t \sigma_1(u)\,dW_u,& \qquad 0 \leq t < T,\end{aligned} \tag{9.17}$$

which converges to zero as $T \to \infty$ whenever the averages of $\{\sigma_2(v)\}_{v\in[t,T]}$ do. To obtain the oscillating behaviour of the zero-coupon rates, define

$$\sigma_2(v) = \frac{1}{2\sqrt{v+1}}\bigl(a + \sin(b\log(v+1)) + 2b\cos(b\log(v+1))\bigr), \qquad v \geq 0,$$

with parameters[1] $a, b \in \mathbb{R}$. Then $\lim_{v\to\infty}\sigma_2(v) = 0$, hence the stochastic part given in (9.17) tends to zero as $T \to \infty$. For all $T > 0$ and $u \in [0,T]$,

$$\int_u^T \sigma_2(v)\,dv = \sqrt{T+1}\bigl(a + \sin(b\log(T+1))\bigr) - \sqrt{u+1}\bigl(a + \sin(b\log(u+1))\bigr).$$

[1] If we choose $a \geq \sqrt{1_{b\neq 0} + 4b^2}$, then $\sigma_2(v) \geq 0$ for all $v \geq 0$.

Hence, using the above expression for $R(t,T)$, the long-term spot rate is given by

$$\begin{aligned} l(t) &= \limsup_{T\to\infty} R(t,T) \\ &= f^* + \limsup_{T\to\infty} \frac{1}{2(T-t)} \int_0^t \sigma_1^2(u)\Big(\int_u^T \sigma_2(v)\,dv\Big)^2\,du \\ &= f^* + \frac{1}{2}\limsup_{T\to\infty}\big(a+\sin(b\log(T+1))\big)^2 \int_0^t \sigma_1^2(u)\,du \\ &= f^* + \frac{1}{2}(|a|+1_{b\neq 0})^2 \int_0^t \sigma_1^2(u)\,du, \qquad t\geq 0, \end{aligned} \tag{9.18}$$

where $1_{b\neq 0}$ equals 1 if $b\neq 0$ and 0 otherwise. In particular, if a and b are not both zero, then, for all times $0\leq s<t$ with $\int_s^t \sigma_1^2(u)\,du>0$, asymptotic monotonicity in the sense $l(s)=l(t)_{\mathcal{F}_s}$ does not hold. The same reasoning as above yields

$$\liminf_{T\to\infty} R(t,T) = f^* + \frac{1}{2}\big(\max\{|a|-1_{b\neq 0},0\}\big)^2 \int_0^t \sigma_1^2(u)\,du, \qquad t\geq 0, \tag{9.19}$$

hence, for $t\geq 0$, the limit of the zero-coupon rates does not exist if $\int_0^t \sigma_1^2(u)\,du>0$ and $b\neq 0$. Furthermore, if $|a|>1_{b\neq 0}$ and σ_1 is not the zero function, then this model provides arbitrage opportunities in the limit for all times $0\leq s<t$ satisfying

$$(|a|+1_{b\neq 0})^2 \int_0^s \sigma_1^2(u)\,du < (|a|-1_{b\neq 0})^2 \int_0^t \sigma_1^2(u)\,du$$

by short-selling the long-term zero-coupon bonds: For every deterministic sequence $\{T_n\}_{n\in\mathbb{N}}$ tending to infinity, we have $l(s)<\liminf_{n\to\infty} R(t,T_n)$ by (9.18) and (9.19), hence the assumptions of Theorem 8.34 have to be violated.

9.2 Models violating the asymptotic minimality

We now present four short-rate models in continuous time, which illustrate the link between asymptotic minimality and the existence of a forward risk neutral probability measure in Condition 8.12, no arbitrage in the limit and convergence of the spot rate.

Example 9.20. On $\Omega=\{0,1\}$ consider $X(\omega)=\omega$ for $\omega\in\Omega$, let $\mathbb{Q}$ denote the uniform distribution, $\mathcal{F}_t=\{\varnothing,\Omega\}$ for $t\in[0,1/3)$ and $\mathcal{F}_t$ equal to the power set of Ω for $t\geq 1/3$. With A given by (9.4), define the interest rate intensity process by

$$r_t = X1_A(t) + (1-X)1_{A^c\cap[1/3,\infty)}(t), \qquad t\in[0,\infty).$$

Note that $\{r_t\}_{t\geq 0}$ is adapted and càdlàg. Using (9.2) and Jensen's inequality, we get for all $t\in[0,1/3)$ and $T>1/3$,

$$\begin{aligned} R(t,T) &= -\frac{1}{T-t}\log\frac{\exp(-\lambda(A\cap[t,T]))+\exp(-\lambda(A^c\cap[1/3,T]))}{2} \\ &\leq \frac{\lambda(A\cap[t,T])+\lambda(A^c\cap[1/3,T])}{2(T-t)} = \frac{T-1/3}{2(T-t)} \leq \frac{1}{2}, \end{aligned}$$

hence $l(t)\leq 1/2$. For $t\geq 1/3$, X is $\mathcal{F}_t$-measurable and we get from (9.2)

$$\begin{aligned} R(t,T) &= \frac{X\lambda(A\cap[t,T])+(1-X)\lambda(A^c\cap[t,T])}{T-t} \\ &= XR_A(t,T)+(1-X)(1-R_A(t,T)), \qquad T>t, \end{aligned}$$

with $R_A(t,T)$ given by (9.5). Therefore, $l(t) = \limsup_{T\to\infty} R(t,T) = 2/3$ for all $t \geq 1/3$, because the points in $[1/3, 2/3]$ are the accumulation points of $R_A(t,T)$ as $T \to \infty$, see Example 9.3.

In this example asymptotic minimality fails for all times $s \in [0, 1/3)$ and $t \geq 1/3$. By construction there exists a forward risk neutral probability measure, which implies with Lemma 8.32, that there is no arbitrage opportunity in the limit with vanishing risk. Since $\mathcal{F}_t$ is finite for each $t \geq 0$, the model provides also no arbitrage opportunity in the limit by Remark 8.31. Therefore Condition 8.12, resp. the weaker Condition 8.10, and the two different notions of no-arbitrage are not sufficient for asymptotic minimality. Furthermore, for $s \in [0, 1/3)$ and $t \geq 1/3$ the inequality $l(s) \leq l(t)$ for asymptotic minimality is strict on Ω.

Indeed, the inequality (8.39) fails, which is sufficient for asymptotic minimality in combination with no arbitrage opportunity in the limit. Consider an arbitrary deterministic sequence $\{T_n\}_{n\in\mathbb{N}}$ tending to infinity. Then

$$\liminf_{n\to\infty} R(t,T_n) = X \liminf_{n\to\infty} R_A(t,T_n) + (1-X)\Big(1 - \limsup_{n\to\infty} R_A(t,T_n)\Big). \tag{9.21}$$

Assume (8.39) holds for $\omega = 1$, then $\liminf_{n\to\infty} R_A(t,T_n) \geq l(t)_{\mathcal{F}_s} = 2/3$. Therefore, $\limsup_{n\to\infty} R_A(t,T_n) \geq 2/3$. By (9.21) follows that the inequality (8.39) fails for $\omega = 0$.

Finally, suppose $T_n := 2^n$ for $n \in \mathbb{N}$. We have seen in Example 9.3, that

$$\liminf_{n\to\infty} R_A(t,T_n) = 1/3, \quad \text{and} \quad \limsup_{n\to\infty} R_A(t,T_n) = 2/3 \quad \text{for all } t \geq 0.$$

Hence, for $1/3 \leq s \leq t$ the inequality (8.35) is strict on Ω, i.e.

$$\liminf_{n\to\infty} R(t,T_n) = 1/3 < 2/3 = l(s).$$

Even, if the limit of the zero-coupon bonds and a forward risk neutral probability measure exist, this is not sufficient for asymptotic minimality, which is shown by the following example.

Example 9.22. Consider $\Omega = (0,1]$ with Lebesgue measure $\mathbb{Q}$, define $\mathcal{F}_t = \{\varnothing, \Omega\}$ for $t \in [0,1)$ and let $\mathcal{F}_t$ denote the Borel σ-algebra of $(0,1]$ for $t \geq 1$. Let $\tau(\omega) = 1/\omega$ for $\omega \in \Omega$ denote the random time, when the interest rate intensity jumps to 1, i.e., we define the interest rate intensity process by $r_t = 1_{[\tau,\infty)}(t)$ for $t \geq 0$. Then τ is $\mathcal{F}_1$-measurable and (9.2) implies for $T > 1$

$$R(1,T) = \frac{1}{T-1}\int_1^T r_u\,du = \frac{T-(T\wedge\tau)}{T-1} \overset{T\to\infty}{\longrightarrow} 1$$

everywhere on Ω, hence $l(1) = 1$. For $t \in [0,1)$ and $T \geq 1$, (9.2) implies that

$$R(t,T) = -\frac{1}{T-t}\log \mathbb{E}_{\mathbb{Q}}\Bigg[\underbrace{\exp\Big(-\int_1^T r_u\,du\Big)}_{\geq 1_{\{\tau\geq T\}}}\Bigg] \leq -\frac{1}{T-t}\log\frac{1}{T} \overset{T\to\infty}{\longrightarrow} 0,$$

hence $l(t) = 0$ due to non-negative interest rates. Therefore, asymptotic minimality does not hold. This does not contradict Corollary 8.36, because this model provides an arbitrage opportunity in the limit for the times $s \in [0,1)$ and $t = 1$ by short-selling long-term zero-coupon bonds. Choose a $(1,\infty)$-valued deterministic sequence $\{T_n\}_{n\in\mathbb{N}}$ tending to infinity, define $\varphi_n = -\exp((T_n-1)/2)$ for each $n \in \mathbb{N}$ and fix $\{\psi_n\}_{n\in\mathbb{N}}$ according to Remark 8.30 so that Definition 8.29(i) holds. Then

$$V_n(1) = -\exp\Big((T_n-1)\Big(\frac{1}{2} - R(1,T_n)\Big)\Big) + \frac{\exp\big(\frac{1}{2}(T_n-1) - R(s,T_n)(T_n-s)\big)}{P(s,1)},$$

hence $\liminf_{n\to\infty} V_n(1) \overset{\text{a.s.}}{=} \infty$ and parts (ii) and (iii) of Definition 8.29 hold.

The existence of a forward risk neutral probability measure or the absence of arbitrage opportunities in the limit, is not even necessary for asymptotic minimality.

Example 9.23. Consider $\Omega = (0,1]$ with Lebesgue measure $\mathbb{Q}$, define $\mathcal{F}_t = \{\varnothing, \Omega\}$ for $t \in [0,1)$ and let $\mathcal{F}_t$ denote the Borel σ-algebra of $(0,1]$ for $t \geq 1$. Let $\tau(\omega) = 1/\omega$ for $\omega \in \Omega$ be a random time. Define the interest rate intensity process[2] by $r_t = 1 - \frac{1}{t}1_{[1,\tau)}(t)$ for $t \geq 0$. With (9.1) the zero-coupon bond price for maturity $T \geq 1$ is given by

$$P(t,T) = e^{-(T-t)}\,\mathbb{E}_{\mathbb{Q}}[\tau \wedge T] = e^{-(T-t)}(1+\log T), \qquad t \in [0,1).$$

Using the definition of the zero-coupon rates in (8.2), we obtain for every $t \in [0,1)$ and $T \geq 1$

$$R(t,T) = 1 - \frac{\log(1+\log T)}{T-t} \overset{T\to\infty}{\longrightarrow} 1.$$

For $t = 1$ the zero-coupon prices for $T \geq 1$ are given by $P(1,T) = e^{-(T-1)}(\tau \wedge T)$ and therefore the zero-coupon rates equal

$$R(1,T) = 1 - \frac{\log(\tau \wedge T)}{T-1} \overset{T\to\infty}{\longrightarrow} 1.$$

Hence asymptotic minimality holds.

On the other hand, we can construct an arbitrage opportunity in the limit for the times $s = 0$ and $t = 1$ according to Definition 8.29 by short-selling long-term zero-coupon bonds. For this define $T_n = n+1$,

$$\varphi_n = -\frac{1}{e\,P(0,T_n)} = -\frac{e^{T_n-1}}{1+\log T_n},$$

and $\psi_n = 1$ for all $n \in \mathbb{N}$. Then $V_n(0) = 0$ for all $n \in \mathbb{N}$ and

$$V_n(1) = -\frac{(\tau \wedge T_n)}{1+\log T_n} + 1 \overset{n\to\infty}{\longrightarrow} 1 \qquad \text{on } \Omega.$$

Example 9.24. Consider $\Omega = \mathbb{N}$, define the filtration

$$\mathcal{F}_t = \begin{cases} \{\varnothing, \Omega\} & \text{for } t \in [0,1), \\ \{\varnothing, \{1\}, \Omega \setminus \{1\}, \Omega\} & \text{for } t \in [1,2), \\ \mathcal{P}(\Omega) & \text{for } t \in [2,\infty), \end{cases}$$

where $\mathcal{P}(\Omega)$ denotes the power set, and the probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{P}(\Omega))$ by $\mathbb{Q}(\{\omega\}) = 1/\omega - 1/(\omega+1)$ for all $\omega \in \Omega$. Let $\tau(\omega) = \omega$ for $\omega \in \Omega$ denote the random time, when the

[2]This example can be slightly simplified if we omit the 1 and allow negative interest rates.

interest rate intensity jumps to $1 - 1/\omega$, i. e., we define the interest rate intensity process by $r_t = (1 - 1/\tau)1_{[\tau,\infty)}(t)$ for $t \geq 0$. Then τ is $\mathcal{F}_2$-measurable and (9.2) implies for $T > 2$

$$R(2,T) = \frac{1}{T-2}\int_2^T r_u\,du = \Big(1 - \frac{1}{\tau}\Big)\frac{T - (T \wedge (\tau \vee 2))}{T-2} \overset{T\to\infty}{\longrightarrow} 1 - \frac{1}{\tau} \tag{9.25}$$

everywhere on Ω, hence $l(2) = 1 - 1/\tau$. Therefore $l(2)_{\mathcal{F}_0} = 0$ and $l(2)_{\mathcal{F}_1} = \frac{1}{2}1_{\Omega\setminus\{1\}}$.

For $T > 0$, we always have that $R(0,T) \geq 0$ and (9.2) implies that

$$R(0,T) = -\frac{1}{T}\log \mathbb{E}_{\mathbb{Q}}\Big[\underbrace{\exp\Big(-\int_0^T r_u\,du\Big)}_{\geq 1_{\{\tau \geq \lceil T\rceil\}}}\Big] \leq -\frac{1}{T}\log\frac{1}{\lceil T\rceil} \overset{T\to\infty}{\longrightarrow} 0,$$

hence $l(0) = 0$ and asymptotic minimality holds for times 0 and 2.

For $T > 1$, (9.2) implies as in (9.25) that $l(1) = 0$ on $\{1\}$ and that on the complement $\Omega\setminus\{1\}$

$$R(1,T) = -\frac{1}{T-1}\log \mathbb{E}_{\mathbb{Q}}\Big[\underbrace{\exp\Big(-\int_1^T r_u\,du\Big)}_{\geq 1_{\{\tau \geq \lceil T\rceil\}}}\Big|\,\tau \geq 2\Big] \leq -\frac{1}{T-1}\log\frac{2}{\lceil T\rceil} \overset{T\to\infty}{\longrightarrow} 0,$$

hence $l(1) = 0$ on Ω and asymptotic minimality *does not* hold for times 1 and 2. To construct an arbitrage opportunity in the limit for times 1 and 2 according to Definition 8.29 by short-selling the long-term zero-coupon bonds, define for each $n \in \mathbb{N}$ the deterministic maturity $T_n = n + 2$ and the strategy by

$$\varphi_n = -1_{\Omega\setminus\{1\}}\exp\big((T_n - 1)R(1,T_n) - R(1,2)\big)$$

and $\psi_n = 1_{\Omega\setminus\{1\}}$. Then (φ_n, ψ_n) is $\mathcal{F}_1$-measurable and $V_n(1) = 0$ for all $n \in \mathbb{N}$. Furthermore, we obtain for all $n \in \mathbb{N}$

$$\varphi_n P(2,T_n) = -1_{\Omega\setminus\{1\}}\exp\big((T_n - 2)(R(1,T_n) - R(2,T_n)) + R(1,T_n) - R(1,2)\big).$$

Since $l(1) = \lim_{n\to\infty} R(1,T_n) = 0$ on Ω as well as $l(2) = \lim_{n\to\infty} R(2,T_n) = 1 - 1/\tau \geq 1/2$ on $\Omega\setminus\{1\}$ by (9.25), we get $\liminf_{n\to\infty}\varphi_n P(2,T_n) = 0$. Therefore, $\liminf_{n\to\infty} V_n(2) = \psi_n \geq 0$ and with probability $\mathbb{Q}(\Omega\setminus\{1\}) = 1/2$ the limes inferior is strictly greater than zero.

9.3 A model without forward risk neutral probability measure and without limiting arbitrage opportunities

The following example is inspired by the infinite-horizon model considered in Example 7.2 in Pliska (1997). It shows that in general for a model, which does not provide an arbitrage opportunity in the limit, there must not exist a forward risk neutral probability measure. The other implication is also not true, see Example 9.22.

Example 9.26. Define $\Omega = \mathbb{N}$, $\mathcal{F}_0 = \{\varnothing, \mathbb{N}\}$ and $\mathcal{F}_1 = \mathcal{P}(\mathbb{N})$, the set of all subsets of $\mathbb{N}$. Let $\mathbb{P}$ be any probability measure on $\mathcal{F}_1$ with $\mathbb{P}(\{\omega\}) > 0$ for all $\omega \in \Omega$. Define zero-coupon bond prices by $P(0,n) = 1$ for all $n \in \mathbb{N}$ and

$$P(1,n)(\omega) = \begin{cases} 1 & \text{if } \omega \leq n-2, \\ (n^2+1)/2 & \text{if } \omega = n-1, \\ 1/2 & \text{if } \omega \geq n, \end{cases}$$

for all $\omega \in \Omega$ and integers $n \geq 2$. By (8.1) and (8.5), it follows that $l(0) = l(1) = 0$, hence asymptotic monotonicity and minimality hold for times $s = 0$ and $t = 1$.

To verify that there is no arbitrage opportunity in the limit for times $s = 0$ and $t = 1$, consider an $\mathcal{F}_0$-measurable, hence deterministic sequence $\{T_n\}_{n\in\mathbb{N}}$ of maturities with $T_n > n$ for all $n \in \mathbb{N}$ and deterministic portfolios (φ_n, ψ_n) with $\varphi_n = -\psi_n$ for all $n \in \mathbb{N}$ so that Definition 8.29(i) is satisfied. Then $V_n(1)(\omega) = 0$ for all $\omega \in \{1, 2, \dots, n-2\}$, hence $\liminf_{n\to\infty} V_n(1) = 0$ on Ω. Therefore, part (iii) of Definition 8.29 is satisfied, but part (ii) does not hold.

To show that Condition 8.10 for times $s = 0$ and $t = 1$ is not satisfied (and there does not exist a forward risk neutral measure for times $s = 0$ and $t = 1$ like in Condition 8.12), we argue by contradiction. Assume that there exists an equivalent probability measure $\mathbb{Q} = \mathbb{Q}_{0,1}$ such that $P(0,n) \geq \mathbb{E}_{\mathbb{Q}}[P(1,n)]$ for all integers $n \geq n_0 \geq 2$. This implies

$$\mathbb{Q}(\{n-1, n, \dots\}) \geq \frac{n^2+1}{2}\mathbb{Q}(\{n-1\}) + \frac{1}{2}\mathbb{Q}(\{n, n+1, \dots\}),$$

hence $n^2\mathbb{Q}(\{n-1\}) \leq \mathbb{Q}(\{n-1, n, \dots\})$ for all integers $n \geq n_0$. Define the constant $c = (n_0 - 1)\,\mathbb{Q}(\{n_0 - 1, n_0, \dots\})/n_0 > 0$. Then we get by induction

$$\mathbb{Q}(\{n-1, n, \dots\}) \geq \frac{cn}{n-1} \tag{9.27}$$

for all integers $n \geq n_0$, because

$$\begin{aligned}\mathbb{Q}(\{n, n+1, \dots\}) &= \mathbb{Q}(\{n-1, n, \dots\}) - \mathbb{Q}(\{n-1\}) \\ &\geq \Big(1 - \frac{1}{n^2}\Big)\mathbb{Q}(\{n-1, n, \dots\}) \geq \Big(1 - \frac{1}{n^2}\Big)\frac{cn}{n-1} = c\frac{n+1}{n}\end{aligned}$$

for all integers $n \geq n_0 + 1$. However, (9.27) for $n \to \infty$ implies $\mathbb{Q}(\varnothing) = c > 0$, which is impossible for a probability measure.

Chapter 10

Proofs for asymptotic monotonicity and minimality

10.1 Proofs of auxiliary results

Proof of Lemma 8.8. Consider a finite non-empty set $I \subset (n \vee t, \infty)$ of zero-coupon bond maturities, which is required to be also a subset of $\mathbb{N}$ in the discrete-time case. Let $M_I := \max_{u \in I} R(t,u)$ denote the maximal available zero-coupon rate. Define the random maturity $T_I \colon \Omega \to I$ as the first one realizing this maximal rate, i. e.

$$T_I = \sum_{u \in I} u\, 1_{\{R(t,u)=M_I,\, R(t,v)<M_I \text{ for all } v \in I, v<u\}}.$$

Note that T_I is $\mathcal{F}_t$-measurable and that $R(t,T_I) = M_I$. By [22, Theorem A.32(b)], there exists, for every $n \in \mathbb{N}$, an increasing sequence $\{I_{k,n}\}_{k\in\mathbb{N}}$ of finite subsets of $(n \vee t, \infty)$, which are also subsets of $\mathbb{N}$ in the discrete-time case, such that

$$S_n := \operatorname*{ess\,sup}_{T>n\vee t} R(t,T) = \lim_{k\to\infty} R(t, T_{I_{k,n}}) \qquad \text{a. s.}$$

Hence, for every $n \in \mathbb{N}$, there exists $k_n \in \mathbb{N}$ such that the essential supremum is nearly reached with high probability, e. g. with the abbreviation $T_n := T_{I_{k_n,n}}$,

$$\mathbb{P}\big(\min\{S_n, n\} - 2^{-n} \le R(t,T_n) \le S_n\big) \ge 1 - 2^{-n}.$$

The a. s. limit of $\{S_n\}_{n\in\mathbb{N}}$ exists due to the monotonicity of the essential suprema. Hence, using the first Borel–Cantelli lemma, the a. s. limit of $\{R(t,T_n)\}_{n\in\mathbb{N}}$ exists and agrees with the one of $\{S_n\}_{n\in\mathbb{N}}$. $\square$

Proof of Lemma 8.9. Fix $0 \le s \le t$. In the continuous-time case, using the definition of the arbitrage-free forward rate in (8.4) and the zero-coupon rate in (8.2),

$$F(s,t,T) = \frac{\log P(s,t)}{T-t} + \frac{T-s}{T-t} R(s,T), \qquad T \in (t,\infty).$$

Since the first summand tends to zero almost surely as $T \to \infty$, it follows that

$$l_F(s,t) = \limsup_{T\to\infty} F(s,t,T) = \limsup_{T\to\infty} \frac{T-s}{T-t}\, R(s,T) = l(s) \qquad \text{a. s.,}$$

by the definition of the long-term forward rate in (8.6) and the long-term zero-coupon rate in (8.5). In the discrete-time case, using the definition (8.3) of the arbitrage-free forward rate and the definition (8.1) of the zero-coupon rate, we see that it is enough to prove

$$\limsup_{T\to\infty} \log\left(\frac{P(s,t)}{P(s,T)}\right)^{1/(T-t)} \overset{\text{a.s.}}{=} \limsup_{T\to\infty} \log P(s,T)^{-1/(T-s)}.$$

However, that is what we just verified for the continuous-time case. □

10.2 Proofs assuming existence of a forward risk neutral probability measure

The key observation for our generalization is the following lemma, which uses notation introduced in Definition 8.19.

Lemma 10.1. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* $\mathcal{G}$ *a sub-*σ*-algebra of* $\mathcal{F}$*.*

(i) For every non-negative random variable X *on* $(\Omega, \mathcal{F}, \mathbb{P})$*, the function* $(0,\infty) \ni t \mapsto \mathbb{E}\big[X^t\big|\mathcal{G}\big]^{1/t}$ *is non-decreasing a. s. and*

$$X \le \lim_{t\to\infty} \mathbb{E}\big[X^t\big|\mathcal{G}\big]^{1/t} = X^{\mathcal{G}} \qquad a.\,s. \tag{10.2}$$

(ii) Let $\{X_t\}_{t>0}$ *be a collection of non-negative random variables on* $(\Omega, \mathcal{F}, \mathbb{P})$*. For each* $n \in \mathbb{N}$ *let* Y_n *denote the essential infimum of* $\{X_t\}_{t>n}$*. Then*

$$X := \liminf_{t\to\infty} X_t \le \lim_{n\to\infty} \mathbb{E}[Y_n^n|\mathcal{G}]^{1/n} = X^{\mathcal{G}} \le \liminf_{t\to\infty} \mathbb{E}[X_t^t|\mathcal{G}]^{1/t} \quad a.\,s. \tag{10.3}$$

(iii) If in (ii) the random variable $X^{\mathcal{G}}$ *dominates* $\{X_t\}_{t>0}$ *in the* $(\mathcal{G}, \mathbb{P})$*-superexponential sense along a subsequence according to Definition 8.20, then the last inequality in* (10.3) *is an a. s. equality.*

Remark 10.4. For the trivial case $\mathcal{G} = \{\varnothing, \Omega\}$, Lemma 10.1(i) implies the well-known result $\lim_{p\to\infty} \|X\|_{L^p} = \|X\|_{L^\infty}$.

Remark 10.5. For a non-negative random variable Z with $\mathbb{E}[Z] = \infty$, we define $\mathbb{E}[Z|\mathcal{G}] = \sup_{n\in\mathbb{N}} \mathbb{E}[\min\{Z, n\}|\mathcal{G}]$. For a σ-integrable random variable with respect to $\mathcal{G}$, the generalization of the conditional expectation is given in [26, Chapter 4].

Remark 10.6. Part (ii) of Lemma 10.1 was introduced by Hubalek et al. (2002) with the additional assumption that the sequence $\{X_t\}_{t>0}$ converges. A further application of the lemma is to prove that the long volatilities, implied by the Black–Scholes formula, cannot fall, which was done by Rogers and Tehranchi (2006).

Example 10.7. Note that $X < X^{\mathcal{G}}$ is possible in (10.2), even for a bounded X. As an example, consider $\Omega = (0,1)$ with Lebesgue measure and Borel σ-algebra, $\mathcal{G} = \{\varnothing, \Omega\}$ and $X(\omega) = \omega$ for $\omega \in \Omega$. Then $\mathbb{E}[X^n|\mathcal{G}]^{1/n} = (n+1)^{-1/n}$ and $X^{\mathcal{G}} = 1$.

Example 10.8. Note that the last inequality in (10.3) can be strict for a bounded sequence $\{X_n\}_{n\in\mathbb{N}}$ which converges everywhere in a monotone way. In the setting of Example 10.7, consider $X_n = 1_{(0,1/n)}$ for $n \in \mathbb{N}$ with pointwise limit $X = 0$, hence $X^{\mathcal{G}} = 0$. However, $\mathbb{E}[X_n^n|\mathcal{G}]^{1/n} = n^{-1/n} \to 1$ as $n \to \infty$.

Proof of Lemma 10.1. (i) Consider $0 < s < t < \infty$. Jensen's inequality for conditional expectations, applied to the convex function $\varphi(x) = x^{t/s}$ implies

$$\mathbb{E}[X^s|\mathcal{G}]^{1/s} = \bigl(\varphi(\mathbb{E}[X^s|\mathcal{G}])\bigr)^{1/t} \le \mathbb{E}[\varphi(X^s)|\mathcal{G}]^{1/t} = \mathbb{E}\bigl[X^t|\mathcal{G}\bigr]^{1/t} \quad \text{a. s.},$$

hence the almost sure limit $C := \lim_{n\to\infty} \mathbb{E}\bigl[X^{t_n}|\mathcal{G}\bigr]^{1/t_n}$ exists along every sequence $t_n \nearrow \infty$ and every other sequence gives a. s. the same limit. Note that C is $\mathcal{G}$-measurable. If Z is a $\mathcal{G}$-measurable random variable with $\mathbb{P}(X \le Z) = 1$, then $\mathbb{E}\bigl[X^t|\mathcal{G}\bigr]^{1/t} \le \mathbb{E}\bigl[Z^t|\mathcal{G}\bigr]^{1/t} = Z$ a. s. for all $t > 0$.

It remains to show that $X \le C$ a. s., which we do by contradiction. We assume for the set $A := \{X > C\}$ that $\mathbb{P}(A) > 0$. Since $A \subset \{C < \infty\}$ there exist $k \in \mathbb{N}$ with $\mathbb{P}(A \cap \{C \le k\}) > 0$. Furthermore, there exists $l \in \mathbb{N}$ such that $\mathbb{P}(B) > 0$ for $B := \{X \ge C + 1/l,\, C \le k\}$. We obtain

$$\mathbb{E}[X 1_B] \ge \mathbb{E}[C 1_B] + \mathbb{P}(B)/l > \mathbb{E}[C 1_B]\,, \tag{10.9}$$

because $\mathbb{P}(B) > 0$ and $\mathbb{E}[C1_B] \le k\,\mathbb{P}(B) < \infty$. In the remaining part of the proof, we use the convention $\infty \cdot 0 = 0$ for products. Using the conditional Hölder inequality[1] and the fact that $\mathbb{N} \ni n \mapsto x^{1-1/n}$ is non-increasing for every $x \in [0,1]$, it follows for all $m, n \in \mathbb{N}$ with $m \le n$ that

$$\begin{aligned}\mathbb{E}[X1_B|\mathcal{G}] &\le \mathbb{E}[X^n|\mathcal{G}]^{1/n}\,\mathbb{E}[1_B|\mathcal{G}]^{1-1/n}\\ &\le \mathbb{E}[X^n|\mathcal{G}]^{1/n}\,\mathbb{E}[1_B|\mathcal{G}]^{1-1/m} \le C\,\mathbb{E}[1_B|\mathcal{G}]^{1-1/m} \qquad \text{a. s.}\end{aligned}$$

Passing to the limit $m \to \infty$ and using the $\mathcal{G}$-measurability of C,

$$\mathbb{E}[X1_B|\mathcal{G}] \le C\,\mathbb{E}[1_B|\mathcal{G}] = \mathbb{E}[C1_B|\mathcal{G}] \qquad \text{a. s.}$$

Taking expectations gives $\mathbb{E}[X1_B] \le \mathbb{E}[C1_B]$, which is a contradiction to (10.9).

(ii) Since $Y_m \le Y_n \le \sup_{k\in\mathbb{N}} Y_k = X$ for all $m, n \in \mathbb{N}$ with $m \le n$, we obtain

$$\mathbb{E}[Y_m^n|\mathcal{G}]^{1/n} \le \mathbb{E}[Y_n^n|\mathcal{G}]^{1/n} \le \mathbb{E}[X^n|\mathcal{G}]^{1/n} \le X^{\mathcal{G}} \qquad \text{a. s.},$$

using part (i) for the last inequality. Hence, by part (i), for every $m \in \mathbb{N}$,

$$\begin{aligned}Y_m \le Y_m^{\mathcal{G}} = \lim_{n\to\infty} \mathbb{E}[Y_m^n|\mathcal{G}]^{1/n} &\le \sup_{n\in\mathbb{N}} \mathbb{E}[Y_n^n|\mathcal{G}]^{1/n}\\ &\le \lim_{n\to\infty} \mathbb{E}[X^n|\mathcal{G}]^{1/n} = X^{\mathcal{G}} \qquad \text{a. s.}\end{aligned}$$

Therefore,

$$X = \sup_{m\in\mathbb{N}} Y_m \le \sup_{m\in\mathbb{N}} Y_m^{\mathcal{G}} \le \lim_{n\to\infty} \mathbb{E}[Y_n^n|\mathcal{G}]^{1/n} \le X^{\mathcal{G}} \qquad \text{a. s.}$$

Since $\sup_{m\in\mathbb{N}} Y_m^{\mathcal{G}}$ is $\mathcal{G}$-measurable and dominates X a. s., it also dominates $X^{\mathcal{G}}$ a. s., hence the last two inequalities are a. s. equalities.

For all $t > n$ we have $Y_n \le X_t$ a. s. Using Jensen's inequality for conditional expectations

$$\mathbb{E}[Y_n^n|\mathcal{G}]^{1/n} \le \mathbb{E}\bigl[Y_n^t|\mathcal{G}\bigr]^{1/t} \le \mathbb{E}\bigl[X_t^t|\mathcal{G}\bigr]^{1/t} \qquad \text{a. s.},$$

hence

$$\mathbb{E}[Y_n^n|\mathcal{G}]^{1/n} \le \liminf_{t\to\infty} \mathbb{E}\bigl[X_t^t|\mathcal{G}\bigr]^{1/t} \qquad \text{a. s.}$$

[1]For a proof, cf. *Hölder's inequality* at en.wikipedia.org/wiki/, version of April 6, 2008.

Passing to the limit $n \to \infty$ gives the last inequality in (10.3).

(iii) Since $X_t \le X^{\mathcal{G}} + \max\{X_t - X^{\mathcal{G}}, 0\}$ for all $t > 0$, the conditional Minkowski inequality and the $\mathcal{G}$-measurability of $X^{\mathcal{G}}$ imply for all $t \ge 1$ that

$$\mathbb{E}\big[X_t^t \,|\, \mathcal{G}\big]^{1/t} \le X^{\mathcal{G}} + \mathbb{E}\big[(\max\{X_t - X^{\mathcal{G}}, 0\})^t \,|\, \mathcal{G}\big]^{1/t} \qquad \text{a. s.}$$

By the assumption and Definition 8.20, the limit inferior of the last term is zero. □

With this lemma we show that the long-term spot rates never fall without the assumption that the spot rates converge.

Proof of Theorem 8.17. (i) In the discrete- and continuous-time case, it is sufficient to show that

$$\liminf_{T\to\infty} P(t,T)^{\frac{1}{T-t}} \le \liminf_{T\to\infty} P(s,T)^{\frac{1}{T-s}} \qquad \text{a. s.} \tag{10.10}$$

by definition of the zero-coupon rate in (8.1) and (8.2), respectively. Note that

$$\lim_{T\to\infty} P(s,t)^{1/(T-t)} = 1. \tag{10.11}$$

Using (10.3) from Lemma 10.1 (with $X_u = P(t, t+u)^{1/u}$ for $u > 0$ and $\mathcal{G} = \mathcal{F}_s$) and afterwards (10.11), it follows that

$$\begin{aligned} \liminf_{T\to\infty} P(t,T)^{\frac{1}{T-t}} &\le \liminf_{T\to\infty} \big(\mathbb{E}_{\mathbb{Q}_{s,t}}\big[P(t,T)\,\big|\,\mathcal{F}_s\big]\big)^{\frac{1}{T-t}} \\ &= \liminf_{T\to\infty} \big(P(s,t)\,\mathbb{E}_{\mathbb{Q}_{s,t}}\big[P(t,T)\,\big|\,\mathcal{F}_s\big]\big)^{\frac{1}{T-t}} \qquad \text{a. s.} \end{aligned} \tag{10.12}$$

For $\varepsilon > 0$ define $f_\varepsilon \colon [0,\infty] \to [0,\infty]$ by

$$f_\varepsilon(x) = \begin{cases} x & \text{for } x \in [0,1], \\ x^{1+\varepsilon} & \text{for } x > 1. \end{cases}$$

Then $x^{1+\delta} \le f_\varepsilon(x)$ for all $x \in [0,\infty)$, uniformly in $\delta \in [0,\varepsilon]$. Using the property in Condition 8.10 and this estimate, we obtain for all $T \ge t + (t-s)/\varepsilon$

$$\begin{aligned} \big(P(s,t)\,\mathbb{E}_{\mathbb{Q}_{s,t}}\big[P(t,T)\,\big|\,\mathcal{F}_s\big]\big)^{\frac{1}{T-t}} &\le P(s,T)^{\frac{1}{T-s}(1+\frac{t-s}{T-t})} \\ &\le f_\varepsilon\big(P(s,T)^{\frac{1}{T-s}}\big) \qquad \text{a. s.} \end{aligned} \tag{10.13}$$

Since f_ε is continuous and monotone increasing, we obtain with (10.12) that

$$\liminf_{T\to\infty} P(t,T)^{\frac{1}{T-t}} \le f_\varepsilon\Big(\liminf_{T\to\infty} P(s,T)^{\frac{1}{T-s}}\Big) \qquad \text{a. s.}$$

for all $\varepsilon > 0$, which implies (10.10).

(ii) This follows from part (i) and the equivalence of the long-term forward and spot rates in Lemma 8.9. □

The following proof combines the ideas from the proofs of Lemma 10.1(iii) and Theorem 8.17.

Proof of Theorem 8.21. Since $l(s)$ is $\mathcal{F}_s$-measurable by definition (8.5), Theorem 8.17 implies that $l(s) \leq l(t)_{\mathcal{F}_s}$ a. s., hence it remains to show that $l(s) \geq l(t)_{\mathcal{F}_s}$ a. s. According to the definition (8.22) of the limiting annual discount factor V_t and Remark 8.25, it suffices to show that

$$\liminf_{T\to\infty} P(s,T)^{1/(T-s)} \leq V_t^{\mathcal{F}_s} \qquad \text{a. s.} \tag{10.14}$$

For $\varepsilon \in (0,1)$ define g_ε: $[0,\infty] \to [0,\infty]$ by

$$g_\varepsilon(x) = \begin{cases} x^{1-\varepsilon} & \text{for } x \in [0,1],\\ x, & \text{for } x > 1. \end{cases}$$

The $x^{1-\delta} \leq g_\varepsilon(x)$ for all $x \in [0,\infty)$, uniformly in $\delta \in [0,\varepsilon]$. Using the property in (8.11) and this estimate for $T \geq s + (t-s)/\varepsilon$, we obtain

$$\begin{aligned} P(s,T)^{\frac{1}{T-s}} &= P(s,t)^{\frac{1}{T-s}} \big(\mathbb{E}_{\mathbb{Q}_{s,t}}\big[P(t,T)\,\big|\,\mathcal{F}_s\big]\big)^{\frac{1}{T-s}} \\ &\leq P(s,t)^{\frac{1}{T-s}} g_\varepsilon\Big(\big(\mathbb{E}_{\mathbb{Q}_{s,t}}\big[P(t,T)\,\big|\,\mathcal{F}_s\big]\big)^{\frac{1}{T-t}}\Big) \qquad \text{a. s.} \end{aligned} \tag{10.15}$$

Using (10.11), Lemma 10.1(iii) (with $X_u = P(t,t+u)^{1/u}$ for $u > 0$ and $\mathcal{G} = \mathcal{F}_s$) and (8.26), it follows that

$$\liminf_{T\to\infty} P(s,T)^{1/(T-s)} \leq g_\varepsilon\big(V_t^{\mathcal{F}_s}\big) \qquad \text{a. s.}$$

for every $\varepsilon \in (0,1)$, which implies (10.14). Using Lemma 8.9, the result for the long-term forward rates follows. □

10.3 Proofs assuming absence of arbitrage in the limit

Proof of Lemma 8.32. Let $\{\varphi_n\}_{n\in\mathbb{N}}$ be a real-valued $\mathcal{F}_s$-measurable sequence, and $\{T_n\}_{n\in\mathbb{N}}$ be a sequence of $\mathcal{F}_s$-measurable random maturities $T_n : \Omega \to (n \vee t, \infty)$, each one taking only a finite number of values. Define the sequence $\{\psi_n\}_{n\in\mathbb{N}}$ as in Remark 8.30 to ensure part (i) of Definition 8.29. Then Condition 8.12 implies

$$\mathbb{E}_{\mathbb{Q}_{s,t}}[V_n(t)\,|\,\mathcal{F}_s] \overset{\text{a.s.}}{=} V_n(s)/P(s,t) \overset{\text{a.s.}}{=} 0, \qquad n \in \mathbb{N}. \tag{10.16}$$

Assume part (iv) of Definition 8.29, in particular $\liminf_{n\to\infty} V_n(t) \geq 0$ a. s. In addition, there exists $n_1 \in \mathbb{N}$ such that $V_n(t) \geq -1$ a. s. for all $n \geq n_1$. Using Fatou's lemma for conditional expectations and (10.16), we obtain

$$\mathbb{E}_{\mathbb{Q}_{s,t}}\Big[\liminf_{n\to\infty} V_n(t)\,\Big|\,\mathcal{F}_s\Big] \leq \liminf_{n\to\infty}\, \mathbb{E}_{\mathbb{Q}_{s,t}}[V_n(t)\,|\,\mathcal{F}_s] = 0 \qquad \text{a. s.}$$

So we must have $\liminf_{n\to\infty} V_n(t) \overset{\text{a.s.}}{=} 0$, hence part (ii) of Definition 8.29 fails. Hence, there is no arbitrage opportunity in the limit with vanishing risk. □

Proof of Theorem 8.33. We prove asymptotic monotonicity, i. e. $l(s) \leq l(t)$ a. s. It suffices to prove $l(s) \leq l(t)_{\mathcal{F}_s}$ a. s., which we do by contradiction. Assume for the event $A := \{l(s) > l(t)_{\mathcal{F}_s}\}$ that $\mathbb{P}(A) > 0$. Since $A \subset \{l(s) > -\infty,\, l(t)_{\mathcal{F}_s} < \infty\}$, there exists $k \in \mathbb{N}$ such that $B := \{X > l(t)_{\mathcal{F}_s}\}$ with $X := \min\{k, l(s) - 2/k\}$ satisfies $\mathbb{P}(B) > 0$. Note that X is a real-valued, $\mathcal{F}_s$-measurable random variable and that $B \in \mathcal{F}_s$. Let $\{\tilde{T}_n\}_{n\in\mathbb{N}}$ denote an $\mathcal{F}_s$-measurable sequence satisfying Lemma 8.8, in particular

$$l(s) \overset{\text{a.s.}}{=} \lim_{n\to\infty} R(s,\tilde{T}_n).$$

Without loss of generality we assume that $\tilde{T}_n > t$ for all $n \in \mathbb{N}$. Since $l(s) > X + 1/k$ by the definition of X, there exists $m \in \mathbb{N}$ such that

$$C := B \cap \bigcap_{n=m}^{\infty} \Big\{ R(s, \tilde{T}_n) \geq X + \frac{1}{k} \Big\} \tag{10.17}$$

satisfies $\mathbb{P}(C) > 0$. Note that $C \in \mathcal{F}_s$. Define $D = \{X > l(t)\} \cap C$. If $\mathbb{P}(D) = 0$, then $X \leq l(t)$ a. s. on C, hence $X \leq l(t)_{\mathcal{F}_s}$ a. s. on C by the $\mathcal{F}_s$-measurability of C and X. This contradicts the strict inequality in the definition of B, which contains C, hence $\mathbb{P}(D) > 0$.

In the continuous-time case define

$$\varphi_n = 1_C \exp((\tilde{T}_n - s)X) P(s,t) \qquad \text{and} \qquad \psi_n = -1_C \exp((\tilde{T}_n - s)X) P(s, \tilde{T}_n),$$

in the discrete-time case, noting that $X > l(t)_{\mathcal{F}_s} \geq -1$ on C, define

$$\varphi_n = 1_C (X+1)^{\tilde{T}_n - s} P(s,t) \qquad \text{and} \qquad \psi_n = -1_C (X+1)^{\tilde{T}_n - s} P(s, \tilde{T}_n)$$

for all $n \in \mathbb{N}$. Then every (φ_n, ψ_n) is $\mathcal{F}_s$-measurable, the corresponding portfolio value $V_n(s) := \varphi_n P(s, \tilde{T}_n) + \psi_n P(s,t)$ is zero, and $V_n(t) = \varphi_n P(t, \tilde{T}_n) + \psi_n$. In the continuous-time case, using the definition of the zero-coupon rates in (8.2) these summands equal

$$\begin{aligned} \varphi_n P(t, \tilde{T}_n) &= 1_C \exp\big((\tilde{T}_n - t)(X - R(t, \tilde{T}_n)) + (t-s)(X - R(s,t))\big), \\ \psi_n &= -1_C \exp\big((\tilde{T}_n - s)(X - R(s, \tilde{T}_n))\big) \end{aligned}$$

for all $n \in \mathbb{N}$. Using the definition of C in (10.17) and $\tilde{T}_n > n$ from Lemma 8.8,

$$1_C \exp\big((\tilde{T}_n - s)(X - R(s, \tilde{T}_n))\big) \leq \exp\Big(-\frac{n-s}{k}\Big) \qquad \text{for all } n \geq m.$$

Since $\varphi_n \geq 0$ for all $n \in \mathbb{N}$, part (iv) of Definition 8.29 holds. Since $X > l(t)$ on D and $l(t) \geq \limsup_{n\to\infty} R(t, \tilde{T}_n)$ a. s. by the definition of the long-term zero-coupon rate in (8.5), we obtain that

$$\liminf_{n\to\infty} 1_C \exp\big((\tilde{T}_n - t)(X - R(t, \tilde{T}_n))\big) \overset{\text{a.s.}}{=} \infty \qquad \text{on } D.$$

Therefore, we have an arbitrage opportunity in the limit with vanishing risk for times s and t, which is the desired contradiction. In the discrete-time case, we proceed in a similar way. The result for the long-term forward rates follows by using Lemma 8.9. □

The following proof has some similarities with the preceding one, however, the stronger no-arbitrage assumption from Definition 8.29 is needed, because the downside risk of the constructed portfolios might be unbounded.

Proof of Theorem 8.34. We want to show that the set, where (8.35) is violated, is a $\mathbb{P}$-null set. For this purpose, define the event

$$C = \{Y > l(s)\}, \qquad \text{where} \qquad Y := \Big(\liminf_{n\to\infty} R(t, T_n)\Big)_{\mathcal{F}_s}, \tag{10.18}$$

and the portfolio compositions

$$\varphi_n = -1_C \frac{P(s,t)}{P(s,T_n)} \qquad \text{and} \qquad \psi_n = 1_C,$$

for all $n \in \mathbb{N}$. Then every (φ_n, ψ_n) is $\mathcal{F}_s$-measurable, the corresponding portfolio value $V_n(s) := \varphi_n P(s, T_n) + \psi_n P(s, t)$ is zero, and $V_n(t) = \varphi_n P(t, T_n) + \psi_n$. In the continuous-time case, the first summand can be rewritten using (8.2) as

$$\varphi_n P(t, T_n) = -1_C \exp\bigl((T_n - t)(R(s, T_n) - R(t, T_n)) + (t - s)(R(s, T_n) - R(s, t))\bigr),$$

for all $n \in \mathbb{N}$. Since $\limsup_{n\to\infty} R(s, T_n) \le l(s) < \infty$ a. s. on C and

$$\limsup_{n\to\infty}\bigl(R(s, T_n) - R(t, T_n)\bigr) \le l(s) - \liminf_{n\to\infty} R(t, T_n) \le l(s) - Y < 0 \quad \text{a. s. on } C,$$

we get $\liminf_{n\to\infty} \varphi_n P(t, T_n) \overset{\text{a.s.}}{=} 0$. Since $\psi_n = 1_C \ge 0$ for all $n \in \mathbb{N}$, this implies part (iii) of Definition 8.29. Since an arbitrage opportunity in the limit for times s and t is excluded by assumption, Definition 8.29(ii) implies $\mathbb{P}(C) = 0$. In the discrete-time case, using the representation

$$\varphi_n P(t, T_n) = -1_C \left(\frac{1 + R(s, T_n)}{1 + R(t, T_n)}\right)^{T_n - t} \left(\frac{1 + R(s, T_n)}{1 + R(s, t)}\right)^{t-s},$$

we conclude in a similar way that $\mathbb{P}(C) = 0$. □

Bibliography

[1] A. Albert, *Estimating the infinitesimal generator of a continuous time, finite state Markov process*, The Annals of Mathematical Statistics **33** (1962), 727–753.

[2] L. Arnold, *Stochastic Differential Equations: Theory and Applications*, Wiley, New York, 1974.

[3] M. Avellaneda and L. Wu, *Credit contagion: pricing cross-country risk in brady debt markets*, International Journal of Theoretical and Applied Finance **4** (2001), 921–938.

[4] T. R. Bielecki, S. Crèpey, M. Jeanblanc, and M. Rutkowski, *Valuation of basket credit derivatives in the credit migration environment*, in [7] (2008), 471–507.

[5] T. R. Bielecki and M. Rutkowski, *Credit Risk: Modeling, Valuation and Hedging*, Springer, Berlin, Heidelberg, 2002.

[6] T. R. Bielecki, A. Vidozzi, and L. Vidozzi, *An efficient approach to valuation of credit basket products and ratings triggered step-up bonds*, Working paper, Illinois Institute of Technology, 2006.

[7] J. R. Birge and V. Linetsky (eds.), *Handbooks in Operations Research and Management Science: Financial Engineering*, vol. 15, Elsevier, 2008.

[8] T. Björk, *Arbitrage Theory in Continuous Time*, 2. ed., Oxford University Press, New York, 2004.

[9] A. J. G. Cairns, *Interest Rate Models: An Introduction*, Princeton University Press, New Jersey, 2004.

[10] J. F. Carriere, *Long-term yield rates for actuarial valuations*, North American Actuarial Journal **3** (1999), no. 3, 13–24.

[11] P. Collin-Dufresne, R. S. Goldstein, and J. Helwege, *Is credit event risk priced? Modeling contagion via the updating of beliefs*, Working paper, University California Berkley, 2003.

[12] F. Couderc and O. Renault, *Times-to-default: life cycle, global and industry cycle impacts*, Working paper, University of Geneva, 2005.

[13] J. C. Cox, J. E. Ingersoll, and S. A. Ross, *A theory of the term-structure of interest rates*, Econometrica **53** (1985), 385–407.

[14] P. Dai Pra, W. Runggaldier, E. Sartori, and M. Tolotti, *Large portfolio losses: a dynamic contagion model*, Annals of Applied Probability **19** (2009), 347–394.

[15] M. Davis and V. Lo, *Infectious defaults*, Quantitative Finance **1** (2001), 382–387.

[16] D. Duffie, A. Eckner, G. Horel, and L. Saita, *Frailty correlated default*, Working paper, Stanford University, 2006.

[17] D. Duffie and K. J. Singleton, *Modeling term structures of defaultable bonds*, Review of Financial Studies **12** (1999), 687–720.

[18] ______, *Credit Risk. Pricing, Measurement, and Management*, Princeton University Press, New Jersey, 2003.

[19] P. Dybvig, J. Ingersoll, and S. Ross, *Long forward and zero coupon rates can never fall*, Journal of Business **69** (1996), 1–25.

[20] D. Egloff, M. Leippold, and P. Vanini, *A simple model of credit contagion*, Journal of Banking and Finance **31** (2007), 2475–2492.

[21] S. N. Ethier and T. G. Kurtz, *Markov Processes*, Wiley, New York, 1986.

[22] H. Föllmer and A. Schied, *Stochastic Finance. An Introduction in Discrete Time*, 2. ed., Walter de Gruyter, Berlin, New York, 2004.

[23] R. Frey and J. Backhaus, *Pricing and hedging of portfolio credit derivatives with interacting default intensities*, Working paper, University of Leipzig, 2007.

[24] K. Giesecke and S. Weber, *Cyclical correlation, credit contagion, and portfolio losses*, Journal of Banking and Finance **28** (2003), 3009–3036.

[25] ______, *Credit contagion and aggregated losses*, Journal of Economic Dynamics and Control **30** (2006), 741–767.

[26] S. He, J. Wang, and J. Yan, *Semimartingale Theory and Stochastic Calculus*, CRC Press, Beijing, 1992.

[27] D. Heath, R. Jarrow, and A. Morton, *Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation*, Econometrica **60** (1992), 77–105.

[28] U. Horst, *Stochastic cascades, credit contagion, and large portfolio losses*, Journal of Economic Behavior and Organization **63** (2007), 25–54.

[29] F. Hubalek, I. Klein, and J. Teichmann, *A general proof of the Dybvig–Ingersoll–Ross theorem: long forward rates can never fall*, Mathematical Finance **12** (2002), 447–451.

[30] J. Hull and A. White, *Pricing interest–rate–derivative securities*, The Review of Financial Studies **3** (1990), 573–592.

[31] R. Israel, J. Rosenthal, and J. Wei, *Finding generators for Markov chains via empirical transition matrices, with applications to credit rating*, Mathematical Finance **11** (2001), 245–265.

[32] R. A. Jarrow, D. Lando, and S. M. Turnbull, *A markov model of the term structure of credit risk spreads*, Review of Financial Studies **10** (1997), 481–523.

[33] R. A. Jarrow and F. Yu, *Counterparty risk and the pricing of defaultable securities*, The Journal of Finance **56** (2001), 1765–1799.

[34] C. Kardaras and E. Platen, *On the Dybvig–Ingersoll–Ross theorem*, Working paper, arXiv:0901.2080v2, 2010.

[35] N. El Karoui, A. Frachot, and H. Geman, *On the behavior of long zero coupon rates in a no arbitrage framework*, Review of Derivatives Research **1** (1998), no. 4, 351–369.

[36] D. Kavvathas, *Estimating credit rating transition probabilities for corporate bonds*, Working paper, University of Chicago, 2000.

[37] A. Klenke, *Wahrscheinlichkeitstheorie*, 1. ed., Springer, Berlin, Heidelberg, 2006.

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

[39] ______, *Credit Risk Modeling: Theory and Applications*, Princton University Press, New Jersey, 2004.

[40] D. Lando and T. Skødeberg, *Analyzing rating transitions and rating drift with continuous observations*, Journal of Banking and Finance **26** (2002), 423–444.

[41] G. Last and A. Brandt, *Marked Point Processes on the Real Line: The Dynamic Approach*, Springer, New York, 1995.

[42] J. P. Laurant and J. Gregory, *Basket default swaps, CDOs and factor copulas*, The Journal of Risk **7** (2005), 103–122.

[43] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, Springer, New York, 1998.

[44] D. X. Li, *On default correlation: a copula function approach*, Working paper, RiskMetrics Group, 2000.

[45] J. H. McCulloch, *Long forward and zero-coupon rates indeed can never fall, but are indeterminate: a comment on Dybvig, Ingersoll and Ross*, Working paper, Ohio State University, 2000.

[46] M. Musiela and M. Rutkowski, *Martingale Methods in Financial Modelling*, 2. ed., Springer-Verlag, Berlin, Heidelberg, 2005.

[47] S. R. Pliska, *Introduction to Mathematical Finance: Discrete Time Models*, Blackwell, Malden, Oxford, 1997.

[48] P. E. Protter, *Stochastic Integration and Differential Equations*, 2. ed., Springer-Verlag, New York, 2004.

[49] L. C. G. Rogers and M. R. Tehranchi, *The implied volatility surface does not move by parallel shifts*, Working Paper, University of Cambridge, 2006.

[50] P. Schönbucher, *Information-driven default contagion*, Working Paper, ETH Zurich, 2003.

[51] P. Schönbucher and D. Schubert, *Copula-dependent default risk in intensity models*, Working paper, University of Bonn, 2001.

[52] K. Schulze, *Asymptotic maturity behavior of bond markets*, Working paper, University of Bonn, 2007.

[53] F. Spitzer, *Infinite systems with locally interacting components*, The Annals of Probability **9** (1981), 349–364.

[54] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, Cambridge, 2000.

[55] O. Vašíček, *An equilibrium characterization of the term structure*, Journal of Financial Economics **5** (1977), 177–188.

[56] D. Vazza, D. Aurora, and J. Torres, *Annual European corporate default study and rating transitions*, Standard & Poors, New York, 2007.

[57] Y. Yao, *Term structure models: a perspective from the long rate*, North American Journal **3** (1999), 122–138.

[58] G. Yin and Q. Zhang, *Continuous-Time Markov Chains and Applications: A Singular Perturbation Approach*, Springer, New York, 1998.

[59] F. Yu, *Correlated defaults in intensity-based models*, Mathematical Finance **17** (2007), 155–173.

Verena Goldammer

born in Alzenau on August 5th, 1980
Nationality: German

Geologengasse 8/14
A-1030 Wien
☎ +43 1 58801 10536 (office)
✉ vgoldamm@fam.tuwien.ac.at
www.fam.tuwien.ac.at/~vgoldamm/

Education

since Feb. 2007	**Ph.D. in Financial Mathematics**, *Vienna University of Technology*, Austria. Supervisor: Prof. Dr. Uwe Schmock Current research: Dependent credit rating transitions, risk measures on L^p random variables, behaviour of long-term yields.
Oct. 2000–Dec. 2006	**Diploma in Mathematics**, *Johannes Gutenberg University Mainz*, Germany. Studies of *mathematics with economics* as subsidiary subject. Diploma exam passed on December 21st, 2006, grade: very good. Diploma thesis *Spread of an Infection in a System of Free Random Walks* in stochastics under supervision of Prof. Dr. Achim Klenke.
Aug. – Dec. 2004	**Erasmus Studies**, *Royal Institute of Technology*, Stockholm, Sweden.
Sept. 1991–June 2000	**Secondary School**, *Spessart-Gymnasium Alzenau*, Germany. Final exam passed on June 30th, 2000, grade: very good.

Professional Experiences

since May 2010	**Research Assistant**, *Vienna University of Technology*, Institute of Mathematical Methods in Economics, Austria. Research assistant in the WWTF project *Mathematics of Financial Risk Measurement and Stochastic Dependence* of Dr. Acciaio, Prof. Dr. Schachermayer, Dr. Schaller and Prof. Dr. Schmock.
Jan. 2009 – Apr. 2010	**Research Assistant**, *Vienna University of Technology*, Institute of Mathematical Methods in Economics, Austria. Research assistant in the WWTF project *Mathematics and Credit Risk* of Prof. Dr. Schachermayer, Ao.-Prof. Dr. Fulmek, Prof. Dr. Pichler and Prof. Dr. Schmock.
Feb. 2007 – Dec. 2008	**Research Assistant**, *Vienna University of Technology*, Institute of Mathematical Methods in Economics, Austria. Research assistant in the Christian–Doppler Laboratory for Portfolio Risk Management (PRisMa Lab), working in cooperation with *UniCredit Bank Austria*.
Jan. – Apr. 2005	**Internship/Working Student**, *Siemens AG*, Munich, Germany. Analysis of a huge sensor data set using MATLAB scripts.

Teaching Experiences

March – June 2009	**Exercises *Stochastic Analysis II***, *Vienna University of Technology*, Austria.
Oct. 2007 – Jan. 2008	**Exercises *Life Insurance Mathematics***, *Vienna University of Technology*, Austria.
Nov. 2005 – Mar. 2006	**Exercises *Stochastics II***, *Johannes Gutenberg University Mainz*, Germany.

Publications and Preprints

V. Goldammer, *Modeling and Estimation of Dependent Credit Rating Transitions*, in preparation.

V. Goldammer, *Asymptotic Properties of the Maximum Likelihood Estimator for the Strongly Coupled Random Walk with Application to Credit Risk*, in preparation.

V. Goldammer and U. Schmock, *Generalization of the Dybvig–Ingersoll–Ross Theorem*, to appear in Mathematical Finance, 2010.

V. Goldammer, *Ausbreitung einer Infektion in einem System freier Irrfahrten*, Diploma thesis, Johannes Gutenberg University Mainz, 2006.

Talks

December 2009 **QMF Conference**, *Modeling and Estimation of Dependent Credit Rating Transitions*, Sydney, Australia.

November 2009 **LMUexcellent Symposium**, *Modeling and Estimation of Dependent Credit Rating Transitions*, invited, LMU Munich, Germany.

September 2009 **PRisMa Day 2009**, *Generalization of the Dybvig–Ingersoll–Ross Theorem and Asymptotic Minimality*, invited, Vienna University of Technology, Austria.

December 2008 **EBIM Doctoral Workshop**, *Modeling and Estimation of Dependent Credit Rating Transitions*, invited, University of Bielefeld, Germany.

December 2008 **Special Semester on Stochastics with Emphasis on Finance – Concluding Workshop**, *Modeling and Estimation of Dependent Credit Rating Transitions*, invited, RICAM Linz, Austria.

November 2008 **Internal Workshop with PRisMa Lab's Industry Partners**, *Implied Rating Models*, Vienna University of Technology, Austria.

September 2008 **PRisMa Day 2008**, *Modeling and Estimation of Dependent Credit Rating Transitions*, invited, Vienna University of Technology, Austria.

September 2008 **European Summer School in Financial Mathematics**, *Modeling and Estimation of Dependent Credit Rating Transitions*, Paris, France.