Master Thesis

Risk sensitivities in risk-neutral network valuations

Julian Stobbe

Frankfurt Institute For Advanced Studies
Johann Wolfgang Goethe-Universität
Frankfurt am Main, April 14, 2019

Supervisor: Prof. Dr. Nils Bertschinger

CONTENTS

1

INTRODUCTION

Cross-holding of financial assets among companies is a common practice in many markets providing a number of benefits such as protection from takeover or lower default risk for the participating firms [Suzuki 2002]. In the light of the financial crisis, which gained traction with the bankruptcy of Lehman Brothers in 2008, the question about contagious effects of such cross-ownership networks imposes itself.
Furthermore, effects like double counting of assets in market valuations can arise from cross-holdings of equity, which lead to over estimated market equity and undervalued financial leverage [Bøhren and Michalsen 1994; Fischer 2014]. Especially the first effect ties into the reasons behind the aforementioned financial crisis in 2008.
The full non-linear effects introduced by circular dependencies from cross-holdings of equity and bonds (debt), have been considered before by Suzuki [2002] and extended to more general liabilities by Fischer [2014]. Our goal is to quantify the risk of contagion, inherent in a cross-ownership network. To this end, the model derived by T. Suzuki, which is based on the famous Merton model [Black and Scholes 1973; Merton 1974], will be adapted.This Black-Scholes type model, provides a formula for a consistent valuation of firms in networks of cross equity and debt holdings. The risk for a crisis associated with a certain network is not obvious, especially for setups in which most firms are initially solvent. A well known concept to represent sensitivities of derivatives (in our case the firm equity and debt — i.e. financial stability) to changes in the underlyings, are the so called "Greeks". Since the derivatives are dependent on the network through the formulas obtained by T. Suzuki, the computation of the Greeks becomes more complicated. We will derive a, formally exact, scheme for their estimation and provide numerical examples for certain network types. Some of the mathematical approaches have been previously considered by Fischer. They are extended here, specifically by the calculation of the Greeks [2014].

2

MATHEMATICAL FOUNDATIONS

The evolution and prediction of stock prices requires somewhat advanced mathematical analysis. Looking at histories of stock prices, one quickly notices the rough trajectories. A reasonable model for these paths is the Brownian motion, originally used to describe the motion of a small particle in a suspension. We will look at some of the mathematical foundations required for the derivation of the network Greeks. First, we will look at the basics of probability theory and then introduce stochastic processes and special formulas needed for option pricing.

2.1 PROBABILITY THEORY AND STOCHASTIC PROCESSES

We begin by defining basic concepts from stochstatics that are essential for this thesis. Some of the definitions and derivations in this chapter are based on [Protter 2005] and on previous work [Stobbe 2018a].

Definition 1 (σ-algebra) Let Ω be a set, called universal set. Then $\Sigma \subseteq \mathcal{P}(\Omega)$ is called σ-algebra if:

(i) $\Omega \in \Sigma$

(ii) $A \in \Sigma \Longrightarrow X \ \ A \in \Sigma$

(iii) $\forall n \in \mathbb{N} \cup \{\infty\} : A_1, A_2, \cdots A_n \in \Sigma \Longrightarrow \bigcup_{i=1}^{n} A_i \in A$

We will use σ-algebras as sets of possible events which will have associated probabilities. These probabilities are given as a measure[1] induced by the σ-algebra:

Definition 2 (Filtered Complete Probability Space) A probability space $(\Omega, \mathcal{F}, P)$ is a measurable space $(\Omega, \mathcal{F})$ with probability measure P over an open set Ω (called sample space) with a σ-algebra $\mathcal{F}$ (called events) and probability measure $P\colon \mathcal{F} \to [0,1], P(\Omega) = 1$ [Dudley 2018].
Let $E \in \mathcal{F}$. A probability space is called complete if every set contained in a measurable null set is also measurable, i.e.: $\forall\, P(E) = 0 : \ \forall E' \subset E : E' \in \mathcal{F}$
A filtered complete probability space or stochastic basis $(\Omega, \mathcal{F}, \mathbb{F}, P)$ is a probability space equipped with a filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t \in \mathbb{T}}$. $\mathcal{F}_t$ is a sequence of σ-algebras with the condition:

$$\forall t_1, t_2 \in \mathbb{T} : t_1 \leq t_2 \Longrightarrow \mathcal{F}_{t_1} \subseteq \mathcal{F}_{t_2}$$

We will assume the *usual hypotheses* (i) and (ii). These allow us to handle some of the mathematical obstacles that arise from continuous time (the complete chain of arguments can be found in chapter 1 of [Protter 2005]):

(i) $\mathcal{F}_0$ contains all the P-null sets of $\mathcal{F}$

(ii) Right continuous: $\forall\, s, t \in I : \mathcal{F}_t = \bigcap_{t<s} \mathcal{F}_s$

Definition 3 (Random Variable) We call $X : \Omega \to S$ a S-valued random variable if it is a $\mathcal{F}$-measurable function[2] from a probability space $(\Omega, \mathcal{F}, P)$ to some measurable space $(S, \mathcal{S})$, which is called state space.

1 A measure is a function μ over a measurable space (A, Σ) with the properties: non-negativity $\forall E \in \Sigma : \mu(E) \geq 0$, contable additivity $\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i)$ and empty set measure $\mu(\emptyset) = 0$

2 A function $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ is called measurable if $\forall T \in \mathcal{S} : \{\omega \in \Omega \,|\, X(\omega) \in T\} \in \mathcal{F}$

Random variables assign certain events (which "live" in the state space) to configurations of the underlying model. For example an event may be a stock value and the underlying model could be made up of stock holdings of all investors.
For a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$, let $E \in \mathcal{F}$. We call $P(E)$ probability of the event E.

Definition 4 (Probability Distribution) We call μ_X distribution or law of X

$$\mu_X(B) = P(X \in B) = P(X^{-1}(B)),\ B \in \mathcal{S} \tag{1}$$

In case X is continuously distributed and real valued, we can also uniquely define the (cumulative) distribution function (CDF) as:

$$F_X(x) = P(X \leq x) = \mu_X((-\infty, x]) = \int_{-\infty}^{x} \rho_X(s)\,\mathrm{d}s \tag{2}$$

A probability density function (PDF) ρ_X for a random variable X and $B \in \mathcal{S}$ connects both measures μ_X and P and provides a transition to Lebesgue integration [Spreij 2010; Yeh and Vestrup 2003].

$$P(X \in B) = \int_B \rho_X(x)\,\mathrm{d}\mu_X(x) \quad \Longrightarrow \int_\Omega \rho_X(\omega)\,\mathrm{d}P(\omega) = \mathbb{1} \tag{3}$$

The notation in (4) will be used whenever we want to indicate that a random variable was drawn according to some PDF ρ_X.

$$X \sim \rho_X \tag{4}$$

For a monotonic transformation function $g(X) = Y$ and $\rho_{g(X)} = \rho_Y$ we can write the transformed probability density in terms of the old one and the inverse of g. Using $|\rho_Y(y)\,\mathrm{d}y| = |\rho_X(x)\,\mathrm{d}x|$, we can write the new density in terms of the original one:

$$\rho_Y(y) = \left|\frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}\right| \rho_X(g^{-1}(y)) \tag{5}$$

Before continuing with stochastic processes we need to introduce the conditional expected value with respect to a sub-σ-algebra.

Definition 5 (Conditional Expectation) For a probability space $(\Omega, \mathcal{F}, P)$, some associated random variable X and a sub-σ-algebra $\mathcal{F}' \subseteq \mathcal{F}$, we define the conditional expectation $E[X|\mathcal{F}']$ of X given $\mathcal{F}'$ as any $\mathcal{F}'$-measurable function satisfying for all $f \in \mathcal{F}'$:

$$\int_f E[X|\mathcal{F}']\,\mathrm{d}P = \int_f X\,\mathrm{d}P \tag{6}$$

Figure 1 shows an example of such a conditional expectation value (see [Wikipedia and Contributors 2019a]). $\mathcal{B}$ is the σ-algebra generated by the intervals with endpoints $\frac{n}{4}$ with $n = 0, 1, \ldots 4$ and $\mathcal{C}$ is the σ-algebra generated by the intervals with end points $0, \frac{1}{2}, 1$. The definitions until now only allow us to define random variables with a common density but not different densities depending on some parameter (e.g. time). In the next section we will amend this by introducing stochastic processes.

2.2 WIENER PROCESS AND BROWNIAN MOTION

We want to model the time development of some random quantity, namely the stock prices. This notion is formalized by stochastic processes over random variables. Adapted stochastic processes are additionally restricted by not being able to "see into the future", since the value at any time point t is restricted to the σ-algebra $\mathcal{F}_t$ fully contained within any future $\mathcal{F}_s$ (with $s > t$).

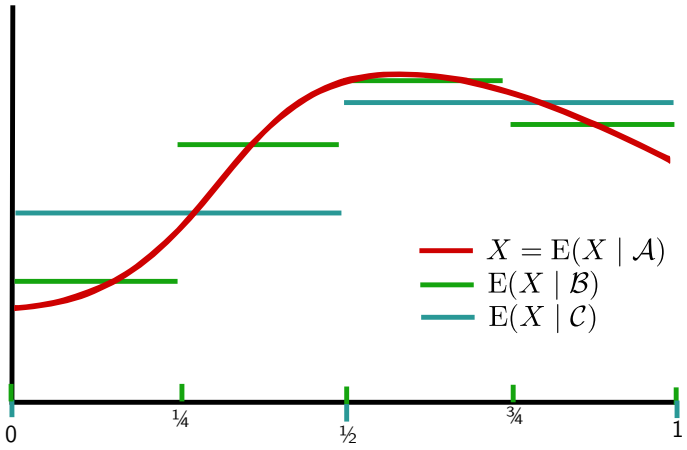


Figure 1: Example for conditional expectation, taken from [Wikipedia and Contributors 2019a]

Definition 6 (Adapted Stochastic Process) A stochastic process X_t on $(\Omega, \mathcal{F}, P)$ is a collection of random variables $(X_t)_{t\in\mathbb{T}}$ over some time interval $\mathbb{T}$[3] A process is called adapted to $\mathbb{F}$ if $\forall t \in \mathbb{T}$: $X_t \in \mathcal{F}_t$.

The stochastic process extends the random variables by the (continuous) time parameter. Going with our previous example, X_t could be the stock value giving a certain configuration at a specific time t. The introduction of filtrations allows to treat continuous time finite probability spaces within the context of countable sets. We can motivate some of the definitions by considering their concrete application in financial models. The σ-algebras $\mathcal{F}_t$ contain events available to investors at time t. While the information is the same for all investors, the probability assigned to events may differ [Elliott and Kopp 1999]. When modeling a market in such a system the natural question of optimal decisions, given all available information, can be formalized using stopping times.

Definition 7 (Stopping Times) Let $(X)_{t\in\mathbb{T}}$ be a random process with $\mathbb{T} = [0, \infty)$, then a random variable $\tau \in \mathbb{T} \cup \{\infty\}$ is called $\mathcal{F}_t$-stopping time, if:

$$\forall t \in \mathbb{T} : \{\tau \leq t\} \in \mathcal{F}_t$$

Examples for stopping times are fix times $\tau = t_0$ or profit thresholds $\tau = \inf\{t \geq 0 | W_t > v\}$. The second stopping time is the time at which a Brownian motion (see def 9) exceeds a threshold v. We continue with special classes of stochastic processes.

Definition 8 (Martingale) A martingale is an adapted stochastic process X_t over $(\Omega, \mathcal{F}, \mathbb{F}, P)$ with respect to a filtration $\mathbb{F}$ such that:

$$\mathbb{E}[X_t] < \infty \Rightarrow X_s = \mathbb{E}_P[X_t | \mathcal{F}_s] \tag{7}$$

Replacing the equal sign by $\leq$ or $\geq$, we obtain the definition for super- and submartignales. Note that the expected value $\mathbb{E}[X_t]$ is constant, monotonic increasing, monotonic decreasing for martingales, submartingales, supermartingales, accordingly.

The usual discussion about equivalence of two stochastic processes and derivation of stochastic integrals is left to the literature, e.g. chapter 6 [Elliott and Kopp 1999]. Martingales model random processes that are fair in the sense, that the expected value over time does not change. Super- and submartingales result in a loss or win over time. The standard method for the quantitative description of stock prices is the Wiener process.

Definition 9 (Wiener Process) A stochastic process W_t is called Wiener process or Brownian motion on $\mathbb{R}$ if it is a martingale and:

3 If not stated otherwise, we will use $\mathbb{T} = [0, \infty)$.

(i) $W_0 = 0$

(ii) $\forall s, t \in I, t \leq s : W_s - W_t \sim \mathcal{N}(0, s-t)$

(iii) W_{t+s} is independent of $\mathcal{F}_t^W = \sigma\{W_u : u \leq t\}$ with $\sigma(M) = \bigcap_{A \in \mathcal{F}} A$. This implies, using the usual rules for conditional expectation: $\mathbb{E}[W_{t+s}|\mathcal{F}_t^W] = \mathbb{E}[W_{t+s}]$

(iv) W_t is continuous in t almost surely

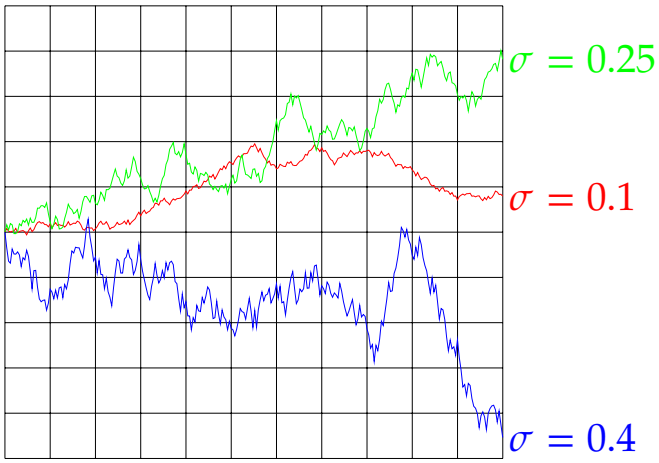


Figure 2: Wiener Process with different variances

The standard Brownian motion (W_t) with respect to $\mathcal{F}_t^W$ is a $\mathcal{F}_t^W$-martingale, because $\mathbb{E}[W_{t+s}|\mathcal{F}_t^W] = W_s$ almost surely, since:

$$\begin{aligned} \mathbb{E}[W_{t+s}|\mathcal{F}_t^W] &= \mathbb{E}[W_{t+s} - W_t + W_t|\mathcal{F}_t^W] \\ &= \mathbb{E}[W_{t+s} - W_t|\mathcal{F}_t^W] + \mathbb{E}[W_t|\mathcal{F}_t^W] \\ &\overset{\text{property 3}}{=} 0 + \mathbb{E}[W_t] \end{aligned} \tag{8}$$

It is also easy to show that for $\mathbb{T} = [0, \infty)$ and $\sigma \in \mathbb{R}^+$:

$$\left(W_t^2 - t\right)_{t \in \mathbb{T}} \tag{9}$$

$$\left(e^{\sigma W_t - \frac{\sigma^2}{2} t}\right)_{t \in \mathbb{T}} \tag{10}$$

are W_t-martingales [Elliott and Kopp 1999]. We wil later discover 10 to be a fitting description for changes in stock prices.

2.3 ITÔ CALCULUS

There are many processes in which the assumption of a fixed expected value over time does not hold. There we will need stochastic processes with a drift. The associated process is known as the Itô (or diffusion) process. The derivation of an appropriate integral type (called Itô or generalized Riemann–Stieltjes integral) is left to the literature[4], e.g. chapter 6 [Elliott and Kopp 1999] or chapter 25, 26 [Klenke 2013]. The most important definitions will be repeated here, as they are integral for the understanding of the financial model. Contrary to standard differential calculus, the integral is the fundamental part of the definition and the differential follows as inverse operation.

4 The subtleties arise from the conditions for the decomposition sequence $\mathcal{D}_T^n$ in $\int_0^T X_s \, \mathrm{d}M_s = \lim_{n \to \infty} \sum_{t \in \mathcal{D}_T^n} X_t (M_{t'} - M_t)$

Definition 10 (Itô Process) Let $(\Omega, \mathcal{F}, \mathbb{F}, P)$ be a filtered complete probability space and $(W_t)_{t\in\mathbb{T}}$ a $\mathbb{F}$-Brownian motion, $(K_s)_{s\in\mathbb{T}}$ and $(H_s)_{s\in\mathbb{T}}$ are $\mathbb{F}$-adapted stochastic processes with $\forall T \in \mathbb{T}$: $\int_0^T |A_s|^2 \,\mathrm{d}s < \infty$ almost surely ($A_s = K_s$ or H_s). A real valued Itô-process $(X_t)_{t\in\mathbb{T}}$ with X_0 $\mathcal{F}_0$-measurable has the form

$$X_t = X_0 + \int_0^t K_s(X_t)\,\mathrm{d}s + \int_0^t H_s(X_t)\,\mathrm{d}W_s$$

This process is the solution of the following stochastic differential equation:

$$\mathrm{d}X_t = K_t(X_t)\,\mathrm{d}t + H_s(X_t)\,\mathrm{d}W_s \tag{11}$$

Let $F : [0,\infty) \to \mathbb{R}$, $F \in C^2(\mathbb{R})$ and $(X_t)_{t\in\mathbb{T}}$ an Itô process, then (these conditions also guarantee existence of the Itô integral) we can state the Itô formula for the Brownian motion (proof chapter 25 [Klenke 2013]):

$$F(t, X_t) = F(0, X_0) + \int_0^t \frac{\partial F}{\partial s}(s, X_s)\,\mathrm{d}s + \int_0^t \frac{\partial F}{\partial x}(s, X_s)\,\mathrm{d}X_s + \frac{1}{2}\int_0^t \frac{\partial^2 F}{\partial^2 x}(s, X_s)\,\mathrm{d}\langle X\rangle_s \tag{12}$$

with $\langle X\rangle_t = \int_0^t H_s^2\,\mathrm{d}s$. In differential form:

$$\mathrm{d}F(t, X_t) = F(0, X_0) + \frac{\partial F}{\partial t}(t, X_s)\,\mathrm{d}t + \frac{\partial F}{\partial X_t}(t, X_s)\,\mathrm{d}X_t + \frac{\partial^2 F}{\partial^2 X_t}(t, X_s)\,\mathrm{d}X_t^2 \tag{13}$$

with $\mathrm{d}X_t^2 = \mathrm{d}t$ for $\mathrm{d}W_t^2$, 0 otherwise. Therefore:

$$\mathrm{d}X_t^2 = K^2\,\mathrm{d}t^2 + KH\,\mathrm{d}t\,\mathrm{d}W_t + H^2\,\mathrm{d}W_t^2 = H^2\,\mathrm{d}W_t^2$$

As an example we apply this formula to the following function $F(t, x)$:

$$F(t, x) = e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma x} \tag{14}$$

With $S_t = F(t, W_t)$

$$\begin{aligned} F(t, W_t) = W_0 &+ \int_0^t \left(\mu - \frac{\sigma^2}{2}\right) e^{\left(\mu - \frac{\sigma^2}{2}\right)s + \sigma W_s}\,\mathrm{d}s + \int_0^t \sigma e^{\left(\mu - \frac{\sigma^2}{2}\right)s + \sigma W_s}\,\mathrm{d}W_s \\ &+ \int_0^t \frac{\sigma^2}{2} e^{\left(\mu - \frac{\sigma^2}{2}\right)s + \sigma W_s}\,\mathrm{d}\langle X\rangle_s \end{aligned}$$

Using $\mathrm{d}\langle X\rangle_s = \mathrm{d}s$:

$$\begin{aligned} &= W_0 e + \int_0^t \mu e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma x}\,\mathrm{d}s + \int_0^t \sigma e^{\left(\mu - \frac{\sigma^2}{2}\right)s + \sigma W_s}\,\mathrm{d}W_s \\ &= W_0 + \int_0^t \mu S_t\,\mathrm{d}s + \int_0^t \sigma S_t\,\mathrm{d}W_s \end{aligned} \tag{15}$$

$$\Rightarrow \mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t \tag{16}$$

This is an Itô process, usually called goemetric Brownian motion, with $K_s = rS_s$ and $H_s = \sigma S_s$. Its solution is given in eq. 14. The fact, that eq. 14 is a solution to eq. 16 (which we just proofed) will be used when we identify eq. 16 as the stochastic process for asset prices in the Black-Scholes model.

The statement above can be extended to systems of stochastic differential equations with $X_t = (X_t^1, \cdots, X_t^n)$ and a m-dimensional Brownian motions $W_t^1, \cdots, W_t^m$:

$$\mathrm{d}X_t^i = \mu_t(X_t)\,\mathrm{d}t + \sum_{j=1}^{m} \sigma_{ij}(X_t)\,\mathrm{d}W_t^j \tag{17}$$

μ_i are the drift variables, σ_{ij} the diffusion coefficients. The diffusion matrix forms a covariance matrix (i.e. positive definite and symmetric). The very useful generalization of Itô's lemma (eq. 13) for any function $F : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}$ then becomes:

$$\begin{aligned} \mathrm{d}F(t, X_t) = &\left(\frac{\partial}{\partial t} + \sum_{i=1}^{n} \mu_i(X_t) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{n} \sigma_{ij}(X_t) \frac{\partial^2}{\partial x_i \partial x_j} \right) F(t, X_t)\, \mathrm{d}t \\ &+ \sum_{i=1}^{n} \sum_{j=1}^{m} \sigma_{ij}(X_t) \frac{\partial}{\partial x_i} F(t, X_t)\, \mathrm{d}W_t^j \end{aligned} \tag{18}$$

2.4 EQUIVALENT MARTINGALE MEASURE

We come to the important concept of the equivalent martingale measure (EMM), later also called risk neutral measure. The importance will become clear when we discuss the fundamental theorem of asset pricing in the next chapter. The mathematical foundation for this measure is given by the Girsanov theorem, for a proof see for example chapter 7 [Elliott and Kopp 1999].

Theorem 1 (Girsanov) *For an adapted $(X_t)_{t\in\mathbb{T}}$ stochastic process on $\left(\Omega, \mathcal{F}, \mathbb{F}^W, P\right)$ with the natural filtration for the Wiener process, $\mathbb{T} = [0, T]$, $\int_0^T X_s^2\, \mathrm{d}s < \infty$ almost surely and the Dolean-Dade exponent $\Lambda_t = e^{-\int_0^t X_s\, \mathrm{d}B_s - \frac{1}{2}\int_0^t X_s^2\, \mathrm{d}s}$ a martingale.*
One can define a new measure Q_X on $\mathcal{F}_T$:

$$\frac{\mathrm{d}Q_X}{\mathrm{d}P}\Big|_{\mathcal{F}_t} = \Lambda_t \qquad \forall t \leq T$$

It follows that

$$W_t = B_t + \int_0^t X_s\, \mathrm{d}s$$

is a standard Brownian motion on $\left(\Omega, \mathcal{F}, \mathbb{F}^W, Q_X\right)$.

Note that we require $T < \infty$. We can apply this change of measure to eq. 16. Let $r \geq 0$ (this will later become the risk free interest rate), then we can define:

$$\Lambda_t = e^{-\frac{t}{2}\left(\frac{r-\mu}{\sigma}\right)^2 + \frac{r-\mu}{\sigma} W_t} \tag{19}$$

Where $W_t^Q = \frac{\mu - r}{\sigma} t + W_t$ (i.e. $X_t = \frac{\mu - r}{\sigma}$ is a Brownian motion with measure Q and $S_t e^{-rt}$ is a Q-martingale

$$\mathrm{d}S_t = rS_t\, \mathrm{d}t + \sigma S_t\, \mathrm{d}W_t^Q \tag{20}$$

$$\implies S_t = S_0 e^{\left(r - \frac{\sigma^2}{2}\right)t + \sigma W_t^Q} \tag{21}$$

It also follows that we can define $\tilde{S}_t = e^{-rt} S_t$:

$$\begin{aligned} \mathrm{d}\tilde{S}_t &= -re^{-rt} S_t + e^{-rt}\, \mathrm{d}S_t \\ &= \tilde{S}_t \left(\mu\, \mathrm{d}t - r\, \mathrm{d}t + \sigma\, \mathrm{d}W_t\right) \\ \Rightarrow \mathrm{d}\tilde{S}_t &= \tilde{S}_t \sigma\, \mathrm{d}W_t^Q \\ \Rightarrow \tilde{S}_t &= S_0 e^{\sigma W_t^Q - \frac{\sigma^2}{2} t} \end{aligned}$$

P and Q are equivalent measures, i.e. they have the same null-space (an event with zero probability in P also has zero probability in Q and the other vice versa). This implies $P \leq Q$ and $Q \leq P$. This does however not mean, that every assigned probability greater zero is equivalent. We will later

use the equivalent martingale measure to simplify calculations. But since both measures do differ in probability weights it is important to keep the distingtion between the real world and artificial measure in mind. For example: the discounted stock price $e^{-rt}S_t$ and the discounted European option $e^{-rt}V_t$ (with $V_t = f(S_{t_0}, S_{t_1}, \cdots), t_i \in [0,t]$, see chapter 3) are martingales with respect to the filtration $\mathcal{F}_t^S$ with the risk neutral measure (EMM) with respect to Q, but not (necessarily) under the real world measure P.

2.5 FEYNMAN-KAC FORMULA

The Feynman-Kac formula provides an expression for solutions of certain stochastic differential equations. It is especially useful for European derivative pricing discussed in the next chapter. We consider the (multidimensional[5]) Itô process $X_t = (X_t^1, \cdots X_t^n)$, $W_t = (W_t^1, \cdots W_t^m)$ is a uncorrelated Wiener process as usual with $\mu(X_t, t) \in \mathbb{R}^n$ and $\sigma(X_t, t) \in \mathbb{R}^{n+m}$. Following [Gubner 2010, chapter 8] we start with the system of stochastic differential equations with correlations, i.e. $\mathrm{d}W_t^i \, \mathrm{d}W_t^j = \rho_{ij} \, \mathrm{d}t$, $t \in \mathbb{T} = [0,T]$ under measure Q:

$$\mathrm{d}X_t = \mu(t, X_t) \, \mathrm{d}t + \sigma(t, X_t) \, \mathrm{d}W_t^Q \tag{22}$$

The corresponding infinitesimal generator A with $f \in C_0^2(\mathbb{T} \times \mathbb{R}^n)$ is defined as [Gubner 2010, theorem 7.3.3.]:

$$Af(t,x) = \sum_{i=1}^{n} \mu_i(t,x) \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{ij} \left[\sigma\sigma^T\right]_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} \tag{23}$$

Theorem 2 (Feynman-Kac) *For the following partial differential equation with a bounded function $f(x,t)$, $\psi \in C_0^2(\mathbb{R})$ and μ, σ as usual, a solution*

$$\begin{aligned} &\frac{\partial u}{\partial t}(t,x) + \left(\frac{\sigma^2(t,x)}{2} \frac{\partial^2}{\partial x^2} + \mu(t,x) \frac{\partial}{\partial x} \right) u(t,x) + r(t,x)u(t,x) + f(t,x) = 0 \\ &\qquad \Leftrightarrow \frac{\partial u}{\partial t}(t,x) + Au(t,x) - r(t,x)u(t,x) + f(t,x) = 0 \end{aligned} \tag{24}$$

with the terminal condition

$$u(T,x,) = \psi(x)$$

The unknown function u can be expressed as follows:

$$u(t,x) = \mathbb{E}_Q \left[\int_t^T f(s, X_s) e^{-\int_t^s r(\tau, X_\tau) \, \mathrm{d}\tau} \, \mathrm{d}r + \int_t^T \psi(X_T) e^{-\int_t^s r(\tau, X_\tau) \, \mathrm{d}\tau} \, \mathrm{d}r \middle| X_t = x \right] \tag{25}$$

A proof for the multidimensional form can be found for example in [Pham 2009, chapter 9].
This formula provides us with a formal solution for the stochastic differential equation 22, which can in fact be evaluated analytically for european options in the Black Scholes model.

2.6 IMPLICIT FUNCTION THEOREM

We will later make use of the implicit function theorem (IFT) in order to rewrite the derivative of a fixed point function. Given a function $f(x,y)$, the IFT guarantees the existence of a function $g(y) = b$ around a point $f(a,b) = 0$ under some mild conditions (see below). Additionally the

5 We do not discuss the multidimensional Itô process as a system of stochastic differential equations, details can be found for example in [Gubner 2010, chapter 4]

theorem gives an explicit expression for the derivative of g. One can motivate this by considering the surface of a sphere in $\mathbb{R}^3$. In Cartesian coordinates this is explicitly written as $x^2 + y^2 + z^2 = C$. The implicit function defining this surface is $f(x, y, z) = x^2 + y^2 + z^2 - C = 0$. It is obviously not possible to define a global function $z = f(x, y)$ which uniquely defines the surface over the whole surface. However, the solution can be uniquely defined for $0 < x^2 + y^2 < C$ as $z = \sqrt{C - x^2 - y^2}$, see fig. 3. Recall the definition for the Jacobian $J_{f,a}$ of a function f at point a. $J_{f,a}$ is the matrix of all

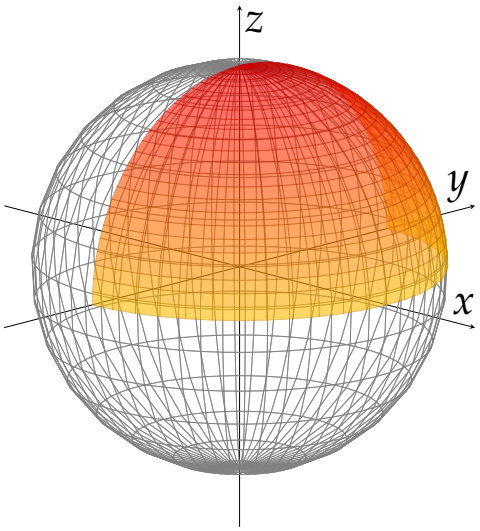


Figure 3: One of four possible regions in which a unique solution can be found. The region is limited by the lines with either $x = 0$ or $y = 0$

partial derivatives of all components for a function $f : \mathbb{R}^m \to \mathbb{R}^n$:

$$\left[J_{f,x}\right]_{ij} = \frac{\partial f_i(x)}{\partial x_j}$$

$$J_{f,x} = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n(x)}{\partial x_1} & \cdots & \frac{\partial f_n(x)}{\partial x_m} \end{pmatrix} \in \mathbb{R}^{n \times m}$$

Theorem 3 (Implicit Function Theorem) *Let $U \times V \subset \mathbb{R}^n \times \mathbb{R}^m$, $f : U \times V \to \mathbb{R}^n$ and $f \in C^1$. For a fixed point $a = (x_0, y_0)$, $x_0 \in U, y_0 \in V$ with $J_{f,y}(a)$ invertible. There exists a unique mapping $g : U \to V$ with $g(x_0) = y_0$ and $\forall x \in U : f(g(y), y)$.*
Additionally, for $x \in U$ and $y \in V$:

$$J_{g,y} = -\left[J_{f,x}(a)\right]^{-1} J_{f,y}(a) \tag{26}$$

$$= -\begin{pmatrix} \frac{\partial f_1(a)}{\partial x_1} & \cdots & \frac{\partial f_1(a)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n(a)}{\partial x_1} & \cdots & \frac{\partial f_n(a)}{\partial x_n} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial f_1(a)}{\partial y_1} & \cdots & \frac{\partial f_1(a)}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n(a)}{\partial y_1} & \cdots & \frac{\partial f_n(a)}{\partial y_m} \end{pmatrix} \tag{27}$$

A proof can be found in [theorem 2-12 Spivak 1965, chapter 2].

3 BLACK-SCHOLES MODEL

3.1 INTRODUCTION

We will now introduce the Black-Scholes model for financial markets, famously developed by Fischer Black, Myron Scholes and Robert Merton during the 1970s and rewarded with the Nobel price in 1997 [Black and Scholes 1973; Merton 1974]. The model allows the analytic treatment of European-style options (see sec. 3.2) and also provides a good basis for numerical simulations of more exotic option types. The central idea of Merton was the time continued hedging of the underlying asset. In other words, finding a ratio for the amount of bought and sold units of the underlying asset and the options in such a way, that the yield is exactly that of the risk free interest rate. More generally, we will see that this risk neutral pricing, justified by theorem 1, is provided by the mathematical formulation of the model [Elliott and Kopp 1999]. Mertons approach also provides a view of valuations for firm as options.
There are seven assumptions needed for the derivation of the central differential equation [Fallis and Hull 2006]:

(i) Geometric Brownian Motion — The stock prices follow a geometric Brownian motion. Since this stochastic process does not have any memory, this implies that consistent prediction of the next step is assumed to be impossible.

(ii) No Arbitrage — No opportunity for riskless profit. This means for example, that there exists no opportunity to buy an asset in one market and immediately sell it in another at a higher value.

(iii) Frictionless Market — No fees, costs or taxes for transactions

(iv) No Dividends — Stocks do not pay any dividend until maturity of the derivatives

(v) Constant σ and r — The riskless interest rate r for lending and borrowing and the volatility σ are assumed to be constant. The risk-free interest rate is also assumed to be equal for all maturities. For many of the derivations here, this condition will be relaxed.

(vi) Short Sales — Short sale is the selling of a security not owned by the seller, i.e. the seller can borrow the security, possible fully using the earnings.

(vii) Perfect Liquidity — Buying and selling of stock can be done with any fractional amount, including short selling. Even very large transactions are assumed to have no implicit impact on the market, so there is no limit on the amount traded.

Some of these restrictions have been relaxed under extended models [Fallis and Hull 2006]. However, the original version of this model is a sufficient basis for our network extension.

3.2 CONTINGENT CLAIMS

A contingent claim is a derivative with a future payoff. Their value – which can be zero when no payoff occurs, hence the name – depends on the underlying asset. Derivatives which are not contingent claims are called forward commitment. The most prominent examples are options and futures. While futures are fixed contracts which both parties must abide by, the option is only exercised if it is to the advantage of the options owner.
Before discussing forward pricing, it is useful to define bonds. The price of any risky asset will

then be given relative to them, taking the interest rate into account. We will later see, that this choice is not fixed. A bond $A(t)$ is sold at a fixed price as a promise of repay at a certain time (maturity of the bond). Until maturity, the bond issuer pays interest $r(t)$ to the bond holder. We could also define it as a contingent claim with payoff 1 at maturity T. The pricing formula is:

$$S_0(t) = e^{\int_0^t r_s \, \mathrm{d}s} \tag{28}$$

A zero-coupon bond does not pay interest. We simplify calculations, by setting the face value (value at maturity) equal to 1. The bond holder makes a profit from the difference between the purchase price and the face value.

$$B(t,T) = S_0(t)\mathbb{E}\left[\frac{1}{S_0(T)}\middle|\mathcal{F}_t\right] \tag{29}$$

With the bond (for a motivation see sec 3.4) we can also write the payoff function as:

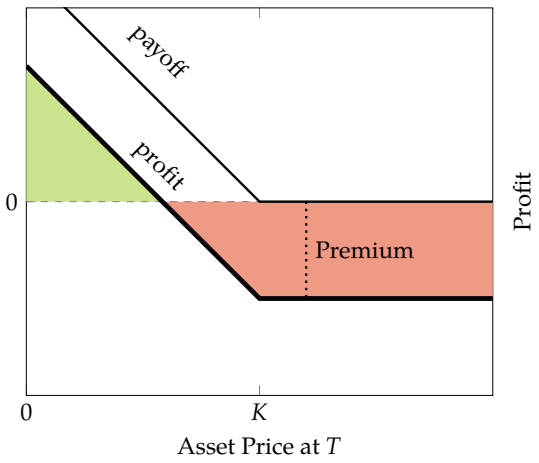
$$B(t,T) = \mathbb{E}\left[e^{-\int_t^T r_s \, \mathrm{d}s}\middle|\mathcal{F}_t\right] \tag{30}$$

These types of bonds differ in when they are worth the most. While a normal bond decreases in value over time, the value of the zero-coupon bond approaches that of its face value.

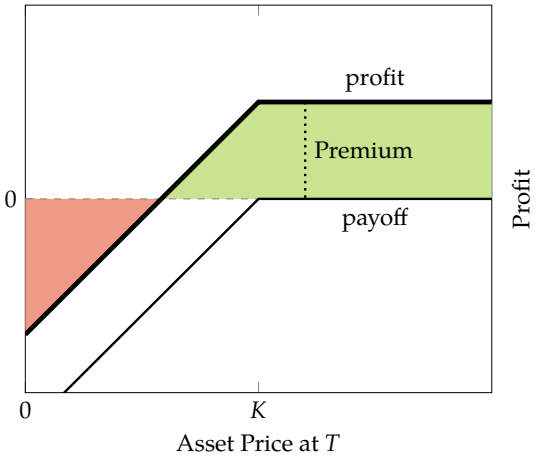
3.2.1 *Options*

We will now discuss options, specifically European style options, in detail. They can then be used to formulate a valuation model for firms.
An option is a contract allowing the buyer to either buy or sell the underlying asset for a specific amount of cash, called *strike price* K at or until a specified time. This time point is called *maturity* and denoted with T (in contrast to the time parameter t). We will consider European style options whenever not stated otherwise. These options can only be exercised at the date of maturity. American options can be exercised at any point before maturity. Besides these vanilla options, there are also so-called exotic options which depend on the stock price more indirectly. Examples are the Asian option (payoff is the average of the underlying price) or Barrier option (the underlying price must cross a threshold value before being exercised).
If the option allows selling of the asset, it is called *put*, otherwise it is called *call* option. The seller of an option is said to be in the *short* position, the buyer has the *long* position. The buyer of the option pays an additional premium, which is the profit for the trader writing the option. On one hand, this fee cannot be too small in order for the writer of the option to make a profit. On the other hand, a very large fee would only be bought by a trader expecting a very large increase of the stock price, which makes it unlikely that the option would actually be sold.
We will assume a discount rate r of zero for now. This results in four possible types of options for a trader:

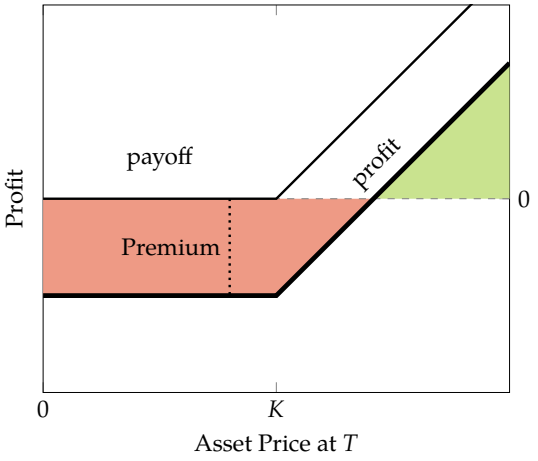
Long Put: To the right, the payoff and profit of a put option in the long position as function of the asset price at maturity T with strike price K are shown. The trader expecting the stock price to decrease, will buy a put option. If the spot price falls below the strike price minus the premium the trader can exercise his option, limiting his losses. Otherwise the option will expire without being used and the seller of the option makes a profit from the premium. The profitable region for the owner of the options are marked with green. The maximum loss is bounded by the premium and the potential profit achieved, when the asset price reaches zero.



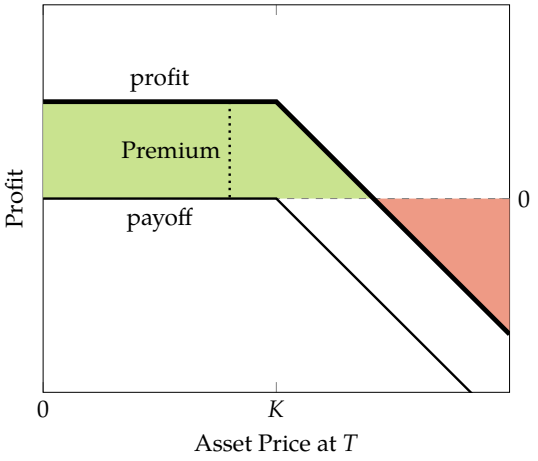
Short Put: To the right, the payoff and profit of a put option in the short position as function of the asset price at maturity T with strike price K. The trader selling (or writing) a put option, expects the stock price to increase, just as for the long call option. He will make a profit from the premium of the option, in case the spot price has risen by more than the premium when the option is exercised. The potential loss is bounded by an asset price of zero while the profit is at most that of the premium.



Long Call: To the right, the payoff and profit of a call option in the long position as function of the asset price at maturity T with strike price K is shown. The trader buying this option expects the stock price to increase. He makes a profit in case the spot price rises above the premium he paid for the option. An American option may be exercised at any point when the spot price has risen above the premium, a European option will be exercised at maturity under the same condition.
The maximum loss is again bounded by the premium, while the potential profit is not bounded.



Short Call: To the right, the payoff and profit of a call option in the short position as function of the asset price at maturity T with strike price K are shown. The trader selling (writing) a call option, expects the stock price to decrease, as for the long put option. He has the obligation to sell the shares at a fixed price. Again, the profit is bounded, but the potential loss is not. The buyer of a call option makes a profit of $S_T - K -$ Permium if $S_T > K$ and potentially loses the premium if $S_T < K$.



European style call options have a payoff of:

$$C_T = \max\{S_T - K, 0\} \tag{31}$$

European put options have a similar payoff function

$$P_T = \max\{K - S_T, 0\} \tag{32}$$

One can combine these options with put and calls in order to achieve certain risk boundaries. for example "selling a straddle" — selling both put and call option at the same time — exposes the trader to a possible loss at large decrease *or* increase of the asset price, but increases the profit margin for the expected asset price at maturity. A trader using such a strategy is called *speculator*,

valuing profit over security. Contrary to this, a *hedger* aims to maximize profit under the constraint of minimal risk. This is achieved by using hedging strategies which put minimal risk over larger profit margins. For example, one can limit the potential loss by using strategies such as butterfly spread or iron condor. Even more conservatively, there are *arbitrageurs*, only taking advantage of riskless profit. These opportunities are assumed to be temporary and short lived on the time scales we are interested in and are therefore neglected. In order to prevent arbitrage possibilities, the options must be priced correctly. Consider an example for a miss-priced option from [Bingham and Kiesel 2004, chapter 1]: An artificial financial market has one state at $t = 0$ and two possible states at maturity $t = T$ and interest rate $r = 0.25/T$. There are bonds $B(t)$, stock $S(t)$ and European call options $C(t)$ with values given in table 1. Let the probability for both events be $\frac{1}{2}$. Now we

$B(0) = \$1$	$S(0) = \$1$	$C(0) = \$0.2$
$B_1(T) = \$1.25$	$S_1(T) = \$1.75$	$C_1(T) = \$0.75$
$B_2(T) = \$1.25$	$S_2(T) = \$0.75$	$C_2(T) = \$0$

Table 1: Market at $t = 0$ and two different possibilities at $t = T$.

consider two different possible portfolios for an investor: $\#B = 10$, $\#S = 10$ and $\#C = 25$ and one with $\#B = 11.8$, $\#S = 7$ and $\#C = 29$. The first portfolio costs \$10 + \$10 + \$5 = \$25 and the second one \$24.6. The possible returns are \$48.75 in case 1 and \$20 in case 2. This means for equal probability of both events, there is a possibility for arbitrage from portfolio 1 to portfolio 2, by going short in (selling) 1.8 bonds and 4 call options, while at the same time going long in (buying) 3 stocks at $t = 0$. the difference between both portfolios at $t = 0$ is the arbitrage profit. For later reference we also write this in vector form, with the matrix $Z(t)$ containing all possible market states as columns:

$$
\begin{aligned}
S(0) &= \begin{pmatrix} \$1 \\ \$1 \\ \$0.2 \end{pmatrix} \\
Z(T) &= \begin{pmatrix} \$1.25 & \$1.25 \\ \$1.75 & \$0.75 \\ \$0.75 & \$0 \end{pmatrix} \\
H_1 &= \begin{pmatrix} 10 & 10 & 25 \end{pmatrix} \\
H_2 &= \begin{pmatrix} 11.8 & 7 & 29 \end{pmatrix}
\end{aligned}
$$

With initial cost for trading strategy 1 $H_1 S(0) = \$25$, trading strategy 2 $H_2 S(0) = \$24.6$ and value at maturity $H_1 Z(T) = \begin{pmatrix} \$48.75 & \$20 \end{pmatrix}$, $H_2 Z(T) = \begin{pmatrix} \$48.75 & \$20 \end{pmatrix}$ This illustrates the equivalence of correct option pricing and our assumption of no arbitrage possibilities.
A real market does not only have probabilities attached to every event, these probabilities are also subjective risk evaluations for every trader. However, we do not need to know them in order to classify whether or not arbitrage possibilities exist within the market. This notion will be made more precise when introducing the fundamental theorems of asset pricing in the next section.
Before continuing, we should formally define the riskless interest rate, allowing pricing of an asset at future times. The increase in value over time of an asset is given by the value times a riskless interest rate $r > 0$.

$$
\frac{\mathrm{d}S(t)}{\mathrm{d}t} = rS(t) \tag{33}
$$

We denote the price at maturity $t = T$ by S_T. The solution to this differential equation can be easily obtained by separation of variables(using $\tau = T - t$):

$$\begin{aligned} \int \frac{1}{S}\,\mathrm{d}S &= \int_t^T r\,\mathrm{d}t' \\ \Rightarrow \ln S(t) &= r\tau + C \\ \Rightarrow S(t) &= \tilde{C}e^{r\tau}, \qquad S(T) = \tilde{C}e^0 = S_T \\ \Rightarrow S(t) &= S_T e^{-r\tau} \end{aligned} \tag{34}$$

3.2.2 *Call-Put Parity*

We will now discuss how the no-arbitrage assumption leads to the fact, that we only need to consider European call *or* put options. More details are discussed in [Elliott and Kopp 1999; Bingham and Kiesel 2004]. Consider the difference between the payoff of a call and a put option at the same maturity:

$$\begin{aligned} C_T - P_T &= \max\{S_T - K, 0\} - \max\{K - S_T, 0\} \\ &= \begin{cases} S_T - K - 0 & S_T > K \\ 0 - (K - S_T) & S_T < K \\ 0 & S_T = K \end{cases} \\ &= S_T - K \end{aligned} \tag{35}$$

This implies for a portfolio consisting of call and put options as well as the underlying, that there is always have a payoff K at maturity. Using eq. 34, we can value this portfolio at any time for maturity T:

$$V(t) = S_0 + P(t) - C(t) = Ke^{-r(T-t)} \tag{36}$$

This equation is known as call-put parity, fixing the bounds for the payoff of European call options [Elliott and Kopp 1999]. Obviously an option should never cost less than zero since it provides only benefits to the buyer. Additionally, the option cannot be more expensive than the underlying, because directly owning the underlying is more valuable than only owning an option to buy it. Rearranging eq. 36 and using that $P_T \geq 0$ as discussed above, we get:

$$\max\{S_T - Ke^{-r(T-t)}, 0\} = C(t) - P(t) \leq C(t) \leq S(t) \tag{37}$$

This bound can be used for the estimation of options, for example in variance reduction methods 5.2.

3.3 FUNDAMENTAL THEOREMS OF ASSET PRICING

Above we have assumed no arbitrage for the Black Scholes model. This is a reasonable assumption, since any situation for which such opportunities arise would immediately be used by arbitrageurs. This opportunity has to be exhausted at some point, otherwise an infinite amount of profit would be generated by using it ad infinitum.
As mentioned before, we are interested in time scales at which such arbitrage opportunities are negligible. We have seen in table 1, that arbitrage possibilities are independent of the pricing probabilities. In fact, the risk neutral interest rate (in the context of mathematical derivations this was called equivalent martingale measure) will be constructed in such a way from the pricing of assets, that it prohibits possibility for arbitrage.
We will motivate this notion for discrete markets using an extension of our previous example[1].

1 The continuous time versions of these theorems will be used later, but not derived. Details and proofs were given by De Souza, Serrano Castañeda, and Ramos Morales [2014].

It is easy to verify that the correct pricing for the option at $t = 0$ should have been \$0.3. In that case the arbitrage opportunity would vanish, since $H_1 S(0) = H_2 S(0) = \$27.5$, while the profits at maturity remain unchanged. In our discrete, finite market with a single time step the existence and uniqueness of this no arbitrage pricing comes down to the existence and uniqueness of the solution for the system of equation:

$$\begin{aligned} A &= (S(0), Z(T)) \\ b &= \left(\$\text{investment}, \$\text{profit}_1, \$\text{profit}_2\right)^T \\ Ax &= b \end{aligned}$$

with $x, b \in \mathbb{R}^3$ in our example. If the set of equation admits multiple solutions for the investment with the same profit[2], there exists an opportunity for arbitrage.
There is however another way of revaluing the call option, by changing the probability measure P. Solving for the weights of the new measure Q under the constraint that S_1, S_2 is a martingale, one obtains the weights for the new measure:

$$\begin{aligned} \frac{S(0)}{B(T)} &= \mathbb{E}_Q\left[\frac{S(T)}{B(T)}\right] \\ \frac{S(0)}{B(0)} &= \frac{q}{B(T)} S_1(T) + \frac{1-q}{B(T)} S_2(T) \\ \Leftrightarrow 1 &= q1.4 + (1-q)0.6 \\ \Leftrightarrow q &= \frac{1}{2} \end{aligned}$$

This is purely coincidentally the same measure, as before. In general, the weights will of course not be the same. We have obtained a martingale S under the measure Q with $Q(Z_1) = \frac{1}{2}$ and $Q(Z_2) = \frac{1}{2}$. The correct pricing of the option can be found as

$$C_0 = B(0)\mathbb{E}_Q\left[\frac{C(T)}{B(T)}\right] = 1 \cdot 0.5 \cdot \frac{0.75}{1.25} + 0 = 0.3$$

This is the same result, that we have already verified as being the correct price at $t = 0$. Surprisingly, this results also holds for infinite and continuous markets. The pricing of an options at maturity can be obtained from the *trading strategy*, modelled by a predictable (see previous definition of self-financing portfolio and [Klenke 2013, chapter 25]) stochastic process $(H_t)_{0 \le t \le T}$.

$$C(0) = \mathbb{E}_Q[C(T)] \tag{38}$$

$$C(T) = C(0) + \int_0^T H_t \, \mathrm{d}S_t \tag{39}$$

The second line holds under restrictions, see [Freddy Delbaen and Walter Schachermayer 2006, chapter 1] for a full reference and proofs for the following theorems.

Theorem 4 (First Fundamental Theorem Of Asset Pricing) *A market which is modeled by a semi-martingale[3] $(S_t)_{0 \le t \le T}$ on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$ does not permit free lunch with vanishing risk (NFLVR), if and only if there exists (at least one) an equivalent martingale measure Q under which S is a σ-martingale.*

For finite markets, the working above can be changed in order to yield a simpler definition. "NFLVR" can be replaced with "no-arbitrage" and the generalized martingale with normal martingales. For our purposes the special version of this theorem will be sufficient (i.e. we model markets using non-generalized martingales). The risk neutral measure Q for market under our restriction, with

2 Note, that this argument is simplified by assuming equal probability for both final states.
3 Definitions for both generalized martingales can be found in [F. Delbaen and W. Schachermayer 1998].

assets $S_1, S_2, \cdots$ must satisfy the following equation for all S_i [F. Delbaen and W. Schachermayer 1998]:

$$S_i(t) = \mathbb{E}_Q \left[e^{-r(T-t)} S_i(T) | \mathcal{F}_t^S \right] \tag{40}$$

Definition 11 (Complete Market) A market is called complete, if it is frictionless (see (iii) above) and every possible future portfolio can be hedged.

The second condition means, that a trader can bet on any future state of the market, since all possible payoff vectors can be constructed. In our example, this correspeonds to the existence of a portfolio H for every possible payoff vector $(x, y) \in \mathbb{R}^2_{\geq 0}$. The completeness is easily verified by confirming that the span of the columns of $Z(T)$ is $\mathbb{R}^2$.
Using this definition it is possible to tighten the previous result:

Theorem 5 (Second Fundamental Theorem Of Asset Pricing) *A market on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$ is arbitrage free and unique, if and only if there exists exactly one equivalent martingale measure.*

Our example above is a complete market and therefore has a unique martingale measure. Proofs for both theorems for frictionless markets were originally given by Harrison and Pliska [1981] and many additional proofs for generalized conditions such as for continuous markets with small transaction costs [De Souza, Serrano Castañeda, and Ramos Morales 2014] have been found[4].
The second fundamental theorem of asset pricing allows us to compute expected values over derivatives regardless of individual risk averseness of traders. As a result one can compute the correct spot price of a derivative as the expected value at maturity, as we have done in our simple example in section 3.2.1.

3.3.1 *Change of Numéraire*

In the previous section we have used the bond price as a benchmark price process. However, any strictly positive process can be used for this purpose. Such a process is called *numéraire*. Note that a deterministic numéraire would not require any additional consideration, since it only appears as a constant facor in the random process. However, we want to price assets according to real market values with their associated risks. Using this technique we can (depending on the context) simplify calculations, for example by getting rid of the drift term. We will now give an abbreviated discussion of the impact, when changing the numéraire. This can then be used to write down the pricing of European options.
If for example the value of a dollar is given as $S_0(t) = e^{\int_0^t r_u \, \mathrm{d}u}$, then the discounted asset price is the following martingale with EMM P^*:

$$\mathrm{d}\left(\frac{S_1(t)}{S_0(t)} \right) = \frac{S_1(t)}{S_0(t)} \sigma(t) \, \mathrm{d}W_t \tag{41}$$

$$= e^{-\int_0^t r_u \, \mathrm{d}u} S_1(t) \sigma(t) \, \mathrm{d}W_t \tag{42}$$

We call $G(t) = e^{-\int_0^t r_u \, \mathrm{d}u} F(t)$ the discounted numéraire with a forward measure $\hat{P}$:

$$\frac{\mathrm{d}\hat{P}}{\mathrm{d}P^*} = \frac{G(T)}{G(0)} = e^{-\int_0^t r_u \, \mathrm{d}u} \frac{F(T)}{F(0)} \tag{43}$$

So we can indeed formulate our initial goal of evaluating $\mathcal{F}_t$ measurable random variables under different reference pricings (or measures) as:

$$\mathbb{E}_{\hat{P}} \left[X \middle| \mathcal{F}_t \right] = \mathbb{E}_{P^*} \left[e^{-\int_0^t r_u \, \mathrm{d}u} \frac{F(T)}{F(0)} X \middle| \mathcal{F}_t \right] \tag{44}$$

4 For example [W. Schachermayer 1992]. A complete review up to 2005 can be found in "The Mathematics of Arbitrage" by Freddy Delbaen and Walter Schachermayer [2006].

Contingent claims C can be priced in the following way, providing a link to path integrals [proposition 12.3 Privault 2016] and [Dash 2004]

$$\mathbb{E}_{P^*}\left[e^{-\int_0^t r_s \,\mathrm{d}u} C \,\middle|\, \mathcal{F}_t\right] = F(t)\mathbb{E}_{\hat{P}}\left[\frac{C}{F(T)}\,\middle|\,\mathcal{F}_t\right] \tag{45}$$

Choosing a suitable numériare can allow for easy evaluation of the right hand side. For example, using a bond with maturity T as before, we find (see also chapter 12 [Privault 2016] for a derivation)

$$\begin{aligned}\frac{\mathrm{d}\hat{P}}{\mathrm{d}P^*} &= \frac{e^{\int_0^T r_u \,\mathrm{d}u} S_0(T,T)}{S_0(0,T)} = \frac{e^{\int_0^T r_u \,\mathrm{d}u}}{S_0(0,T)} \\ \Rightarrow \mathbb{E}_{P^*}\left[e^{-\int_0^t r_s \,\mathrm{d}u} C \,\middle|\, \mathcal{F}_t\right] &= S_0(t)\mathbb{E}_{\hat{P}}\left[C\,\middle|\,\mathcal{F}_t\right]\end{aligned}$$

3.4 BLACK-SCHOLES DIFFERENTIAL EQUATION

In a self financing portfolio no additional cash is withdrawn or invested, but investment between different assets may be redistributed over time. They are used to model hedging against traiding risks (see discussions after options) and realistic investment strategies. In fact, any contingent claim can be generated by an admissable self-financing strategy in a complete market, see def. 11 [Elliott and Kopp 1999, chapter 4].

Definition 12 (Self-Financing portfolio) Let $\phi_t = (H_t^i)_{1\le i\le N}$ be a measurable adapted process and H_t^i represent the number of unit asset i held at time t [Elliott and Kopp 1999, chapter 7]. The value at time t is given by

$$V_t(\phi_t) = \sum_{i=0}^{N} H^i(t) S^i(t) \tag{46}$$

where S_0 is the bond price (or risk free bank account) with interest rate r_t and S_i are shares in risky assets. and ϕ_t is called self-financing if

$$\mathrm{d}V_t(\phi_t) = \sum_{i=0}^{N} H^i(t)\,\mathrm{d}S^i(t) \tag{47}$$

In this definition we assume that S_t^0 is a bond according to eq. 28 and S_t^1 is a risky asset. Furthermore only one additional asset is used in the following. S_t^1 is defined as in eq. 15 and eq. 16, such that:

$$\begin{aligned}S_t^1 &= S_0^1 + \int_0^t \mu(u) S_u^1 \,\mathrm{d}u + \int_0^t \sigma(u) S_u^1 \,\mathrm{d}B_u \\ \mathrm{d}X_t &= r(X_t - H_t^1 S_t^1)\,\mathrm{d}t + H_t^1 \,\mathrm{d}S_t^1 \\ \Rightarrow \mathrm{d}W_t^\theta &= \theta_t \,\mathrm{d}t + \mathrm{d}B_t\end{aligned}$$

with $\theta = \frac{\mu(t)-r_t}{\sigma(t)}$, this is a Wiener process and the discounted value process is given by:

$$\mathrm{d}\left(\frac{X_t}{S_t^0}\right) = H_t^1 \sigma(t) \frac{S_t^1}{S_t^0}\,\mathrm{d}W_t^\theta \tag{48}$$

As mentioned before, the usefulness of self-financing portfolios for our purposes comes from the fact, that we can easily formulate hedging strategies for contingent claims in terms of these portfolios. For example, a contingent claim with maturity T is called attainable, if it can be hedged from a self financing portfolio:

$$C(T) = \sum_{i=0}^{i} H_T^i S_T^i \tag{49}$$

Starting from assumption 1, i.e. the generalized Brownian motion, sec. 2.2 for the N underlying assets $S_i(t) \in \Omega, i \in [1, N]$ is assumed, we can summarize the above in terms of two differential equations:

$$\mathrm{d}S_0(t) = rS_0(t)\,\mathrm{d}t \tag{50}$$
$$\mathrm{d}S_1(t) = S_1(t)\mu\,\mathrm{d}t + S_1(t)\sigma\,\mathrm{d}W_t \tag{51}$$
$$\tag{52}$$

Let $(H_i^t)_{0\le i\le N}$ be a self financing strategy, defining the value according to eq. 46. Then for any function $f : T \times \Omega \to \mathbb{R}$ and $f \in C^{1,2}$ satifying the Black-Scholes equation, the following holds:

$$rf(t,x) = \frac{\partial f}{\partial t}(t,x) + rx\frac{\partial f}{\partial x}(t,x) + \frac{\sigma^2 x^2}{2}\frac{\partial^2 f}{\partial x^2}(t,x) \tag{53}$$

with

$$H_t^0 = \frac{\partial f}{\partial x}(t, S_1(t)) \tag{54}$$

This can be shown using Itô's lemma (eq. 13) and assumption (v). An example of such a function is the payoff function for an option. Consider for example a European style call option $C(x) = \max\{x - K, 0\}$, then the payoff function is given by the solution to the following boundary value problem:

$$rg_C(t,x) = \frac{\partial g_C}{\partial t}(t,x) + rx\frac{\partial g_C}{\partial x}(t,x) + \frac{\sigma^2 x^2}{2}\frac{\partial^2 g_C}{\partial x^2}(t,x) \tag{55}$$
$$g_C(T,x) = C(x) \tag{56}$$

This can indeed be solved analytically in this case [Privault 2016, section 5.5 and 5.6]

$$g_C(t,x) = x\Phi\left(d_+(T-t)\right) - Ke^{-(T-t)r}\Phi\left(d_-(T-t)\right) \tag{57}$$

with

$$d_+(s) = \frac{\ln(x/K) + (r + \frac{\sigma^2}{2}s)}{\sigma\sqrt{s}} \tag{58}$$
$$d_-(s) = \frac{\ln(x/K) + (r - \frac{\sigma^2}{2}s)}{\sigma\sqrt{s}} \tag{59}$$
$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{s^2}{2}}\,\mathrm{d}s \tag{60}$$

Inserting $x = S_t$, we can interpret the first term of eq. 57 as the risky and the second one as risk-free investment.
Even without the full proof, we can easily verify that the solution in eq. 57 recovers the correct payoff at maturity:

$$\lim_{s\to 0} d_+(s) = \lim_{s\to 0} d_-(s) = \begin{cases} +\infty & x > K \\ 0 & x = K \\ -\infty & x < K \end{cases} \tag{61}$$

$$\Rightarrow \lim_{t\to T} g_C(t, S_t) = \begin{cases} S_t \lim_{t\to\infty}\Phi(t) - K\lim_{t\to\infty}\Phi(t) = S_t - K & S_t > K \\ S_t - K = 0 & S_t = K \\ S_t \lim_{t\to-\infty}\Phi(t) - K\lim_{t\to-\infty}\Phi(t) = 0 & S_t < K \end{cases} \tag{62}$$
$$= \max\{S_t - K, 0\} = C_T \tag{63}$$

using $\lim_{t\to\infty}\Phi(t) = 1$ and $\lim_{t\to-\infty}\Phi(t) = 0$.

	Call	Put
V	$S_t\Phi(d_+) - Ke^{-r\tau}\Phi(d_-)$	$Ke^{-r\tau}\Phi(-d_-) - S_t\Phi(-d_-)-$
Δ	$\Phi(d_+)$	$-\Phi(-d_+)$
ν	$S_t\sqrt{\tau}\Phi'(d_+)$	$S_t\sqrt{\tau}\Phi'(d_+)$
Θ	$-\frac{S_t\sigma\Phi'(d_+)}{2\sqrt{\tau}} - rKe^{r\tau}\Phi(d_-)$	$-\frac{S_t\sigma\Phi'(d_+)}{2\sqrt{\tau}} + rKe^{r\tau}\Phi(-d_-)$
ρ	$K\tau e^{-r\tau}\Phi(d_-)$	$-K\tau e^{-r\tau}\Phi(-d_-)$

Table 2: Black-Scholes Greeks, [Wikipedia and Contributors 2019b]

3.5 THE GREEKS

Besides the price of derivatives (as before, we are interested in European options), one can also quantify the sensitivity to underlying parameters. These quantifiers are called Greeks and are useful in order to estimate the sensitivity of the asset price stemming from this parameter. The name is used because the resulting functions are usually denoted with Greek letters, or at least Greek sounding ones like "Vega".
They are regularly used as hedge parameter of portfolios in order to lower the total risk.
The Greeks are computed using the partial derivative of the asset price with respect to the underlying quantity, The most commonly used first order Greeks are:

$\Delta = \frac{\partial V}{\partial S_0}$ Change in option price with respect to the change in the price of the underlying asset S_0. This is the most commonly used Greek, since fixing this value to 0 for a portfolio provides some protection against changes in the underlying asset values (Delta hedging).

$\Theta = \frac{\partial V}{\partial T}$ Time decay of an option. This quantifies the decay in risk premium a seller might charge over time.

$\rho = \frac{\partial V}{\partial r}$ Dependency on the risk neutral interest rate.

(Vega) $\nu = \frac{\partial V}{\partial \sigma}$ Change in option price due to the change in volatility σ of the asset S_0. If, for example, Θ is much larger than 0, we expect the asset price to increase strongly if the volatility increases.

$\lambda = \Delta\frac{S}{V}$ While this is not a new variable in itself, its usefulness comes from the interpretation as leverage (the impact of debt capital can make on return of equity).

For European style options the Black-Scholes model permits analytic calculation of the Greeks. In table 2 an overview of the analytic formulas is given. In order to shorten the notation we also use $\tau := T - t$ and $d_{+/-} := d_{+/-}(T - t)$.

4 NETWORK MODEL

We will now introduce a model for networks of cross-ownerships between firms. Determining valuations of these firms at maturity has already been studied by Suzuki [2002] and Fischer [2014]. The novel contribution is an extension of the well known Greeks as risk measures for these types of netwok models.

4.1 INTRODUCTION

We need to introduce a new index to identify each firm, changing the definition of the diffusion slightly. Note especially that we now use $a_i(t)$ instead of $S(t)$ for the asset prices, to avoid confusion with the equity part. All important variables are summarized in table 3.

$$\begin{aligned} \frac{\mathrm{d}a_i(t)}{a_i(t)} &= r\,\mathrm{d}t + \sigma_i\,\mathrm{d}W_i^Q \\ \Rightarrow a_i(t) &= a_{0,i}\,e^{\left(r-\frac{1}{2}\sigma_i^2\right)t+\sigma_i W_i^Q(t)} \end{aligned} \tag{64}$$

Let $Z_i = \sigma_i W_i^Q$ and $\mathbb{E}[W_i^Q W_j^Q] = \rho_{ij}$

$$a_i(t) = a_{0,i}e^{\left(r-\frac{\sigma_i^2}{2}\right)t+\sqrt{t}Z_i} \tag{65}$$

$$\Leftrightarrow Z_i = \frac{1}{\sqrt{t}}\left(\ln(a_i(t)) - \ln(a_{0,i}) - \left(r-\frac{\sigma_i^2}{2}\right)t\right) \tag{66}$$

Therefore the random variable Z is distributed according to a multivariate Gaussian:

$$\mathbf{Z} \sim \mathcal{N}(0, \mathbf{\Sigma}), \qquad \Sigma_{ij} = \rho_{ij}$$

We will use the Cholesky decomposed covariance matrix (eq. 67) since it allows us to sample from multivariate Gaussian random variables during simulations:

$$\Sigma_{ij} = \left(L^T L\right)_{ij} \tag{67}$$

$$\Rightarrow Z_i = \mu + L^T Z_{0,i} \tag{68}$$

where $Z_{0,i}$ is the i-th component of a $\mathcal{N}(0, \mathbb{1})$ and Z_i is the i-th component of a $\mathcal{N}(\mu, \Sigma)$ distributed random variable.

4.2 VALUATION

As discussed above, we can write equity and debt of firms at time T as options on the firms value [Merton 1974].

$$V_i = e^{-r\tau}\mathbb{E}_Q\left[\max\{0, a_i(T) - K\}\right] \tag{69}$$

$$= \mathbb{E}_Z\left[e^{-r\tau}\max\{0, a_{0,i}e^{(r-\frac{\sigma_i^2}{2})\tau+\sqrt{\tau}L^T Z_i} - K\}\right] \tag{70}$$

r	interest rate
σ_i	volatility
$\Sigma_{ij} = \left(L^T L\right)_{ij} \sigma_j$	correlation between Brownian motion of firms
s_i	equity
r_i	debt payout
$v_i = r_i + s_i$	total value
x_i	combined equity and debt
a_i	assets
d_i	value of debt
$g(\mathbf{a}, \mathbf{x})$	fixed point function

Table 3: Variables for the network model

Let N be the number of firms with cross holdings. Then equity s_i and debt r_i part of the value are defined as call and put options:

$$\begin{aligned} s_i &= \max\{0, a_i + M^s_{ij} s_j + M^d_{ij} r_j - d_i\} \\ r_i &= \min\{d_i, a_i + M^s_{ij} s_j + M^d_{ij} r_j\} \\ v_i &= s_i + r_i \end{aligned}$$

The value under the risk-neural measure $\mathbb{P}_Q$ is

$$v^Q_i = \mathbb{E}_Q \left[v_i(a_1, \cdots a_N) \right]$$

This recurrent definition can be solved by finding the fixed point of the value function. Accounting for the networkstructe of the cross holdings, we need to introduce a matrix connecting equity and debt part of different firms, following the original work from Suzuki [2002]:

$$\begin{aligned} M_{ij} &= \begin{cases} M^s_{ij} & \text{if } j \leq N \\ M^d_{i(j-N)} & \text{if } j > N \end{cases} \\ x_i &= \begin{cases} s_i = \max\{0, a_i + M_{ij} x_j - d_i\} & \text{if } i \leq N \\ r_{i-N} = \min\{d_{i-N}, a_{i-N} + M_{i-N,j} x_j\} & \text{if } i > N \end{cases} \\ v_i &= \max\{0, a_i + M_{ij} x_j - d_i\} + \min\{d_i, a_i + M_{ij} x_j\} \\ &= a_i + M_{ij} x_j \end{aligned}$$

A more concise notation can be obtained by stacking $\mathbf{s}$ and $\mathbf{r}$ into a single vector, as well as defining a $N \times 2N$ matrix M: We then use the common valuation function $g(\mathbf{a}, \mathbf{d}, \mathbf{x})$:

$$g_i(\mathbf{a}, \mathbf{x}) = \begin{cases} \max\{0, a_i + M_{ij} x_j - d_i\} & \text{if } i \leq N \\ \min\{d_i, a_i + M_{ij} x_j\} & \text{if } i > N \end{cases} \tag{71}$$

Since the valuation function eq. 71 only depends on the difference of $\mathbf{a}$ and $\mathbf{d}$, we will leave d as a fixed parameter, simplifying the fixed point function:

$$f\left(\mathbf{a}, , \mathbf{x}\left(\mathbf{a}, \right)\right) \equiv g\left(\mathbf{a}, \mathbf{x}\left(\mathbf{a}\right)\right) - \mathbf{x} = 0 \tag{72}$$

4.3 FIRST ORDER GREEKS

The definition above allows for the application of the implicit function theorem, sec. 2.6:

$$\begin{aligned}\left[J_{x,a}\right]_{ij} = \frac{\partial}{\partial a_j} x_i(\mathbf{a}) &= -\left[J_{f,x}^{-1}\left(\mathbf{a}, \mathbf{x}\left(\mathbf{a}\right)\right)\right]_{ik}\left[J_{f,a}\left(\mathbf{a}, \mathbf{x}\left(\mathbf{a}\right)\right)\right]_{kj} \\ &= \left[\left[\mathbb{1}_{2N\times 2N} - J_{g,x}\left(\mathbf{a}, \mathbf{x}\left(\mathbf{a}\right)\right)\right]^{-1}\right]_{ik} \frac{\partial g_k\left(\mathbf{a}, \mathbf{x}\left(\mathbf{a}\right)\right)}{\partial a_j} \qquad (73)\end{aligned}$$

$$B(\mathbf{a}, \mathbf{x}(\mathbf{a})) = \left[\mathbb{1}_{2N\times 2N} - J_{g,x}\left(\mathbf{a}, \mathbf{x}\left(\mathbf{a}\right)\right)\right]^{-1} \qquad (74)$$

With the usual Jacobian matrix:

$$\left[J_{f,a}(x,y)\right]_{ij} = \frac{\partial f_i}{\partial a_j}\Big|_{f(x,y)}$$

Let $\zeta_i \in \{0,1\}$ be an indicator variable which is equal to one if firm i is solvent ($v_i \geq d_i$) and zero otherwise. Using this definition the Jacobian matrix has the following form:

$$\frac{\partial}{\partial x_j} g_i = \begin{cases} \zeta_i M_{ij} & \text{if } i \leq N \\ (1-\zeta_i) M_{ij} & \text{if } i > N \end{cases} \qquad (75)$$

$$\frac{\partial}{\partial a_j} g_i = \begin{cases} \delta_{ij}\zeta_i & \text{if } i \leq N \\ \delta_{ij}(1-\zeta_i) & \text{if } i > N \end{cases} \qquad (76)$$

We assume that the asset values at maturity are written as a function of a random variable Z. There are two different approaches which allow the estimation of the first order Greeks (i.e. sensitivity to exogenous assets, interest rate, maturity and volatility) for networks of arbitrary size.

4.3.1 *Log Derivative Trick*

The first approach makes use of the fact that the derivative of the logarithm of a function is sometimes easier to compute than the direct derivative.

$$\nabla_x \log p(x) = \frac{\nabla_x p(x)}{p(x)} \qquad (77)$$

$$\Leftrightarrow \nabla_x p(x) = p(x)\nabla_x \log p(x) \qquad (78)$$

This is especially useful when computing derivatives of parameters of expected values. Let $x \sim p(x|\theta)$, and $x \in \Omega$ discrete (otherwise replace the sum by an integral):

$$\begin{aligned}\nabla_\theta \mathbb{E}\left[f(x)\right] &= \nabla_\theta \sum_{x\in\Omega} p(x|\theta) f(x) \\ &= \sum_{x\in\Omega} \nabla_\theta p(x|\theta) f(x) \\ &= \sum_{x\in\Omega} p(x|\theta) f(x) \nabla_\theta \log p(x|\theta) \\ &= \mathbb{E}\left[f(x)\nabla_\theta \log p(x|\theta)\right]\end{aligned}$$

Let $\Theta \in \{\mathbf{S}_0, \sigma, r, \tau\}$. Using eq. 80 we can write:

$$\frac{\partial}{\partial \theta_j}\mathbb{E}_Q\left[e^{-rT} x_i(\mathbf{a}_T)\right] = \mathbb{E}_Q\left[e^{-rT}\frac{\partial}{\partial\theta}\ln\left(p(\mathbf{a}_T|\theta)\right) x_i(\mathbf{a}_T)\right] \qquad (79)$$

$$
\begin{aligned}
p(\mathbf{a}(t)) &= p(\mathbf{Z}) \det\left(J_{Z,a(t)}\right) = p(\mathbf{Z}) \det\left(\delta_{ij}(\sqrt{t}a_{t,i})^{-1}\right) \\
&= -\frac{\partial}{\partial a_k}\left(\frac{1}{2} Z_i \Sigma_{ij}^{-1} Z_j\right) \\
&= -\frac{1}{2T}\left(\frac{\delta_{ik}}{a_i}\Sigma_{ij}^{-1} Z_j + \frac{\delta_{jk}}{a_j} Z_i \Sigma_{ij}^{-1}\right) \\
&= -\frac{1}{t a_k}\Sigma_{kj}^{-1} Z_j \qquad (80)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial a_{0,k}} \ln\left(p(\mathbf{a}(t))\right) &= -\frac{1}{2T}\left(\frac{\partial Z_i}{\partial a_{0,k}}\Sigma_{ij}^{-1} Z_j + \frac{\partial Z_j}{\partial a_{0,k}} Z_i \Sigma_{ij}^{-1}\right) \\
&= \frac{1}{T a_{0,k}}\Sigma_{kj}^{-1} Z_j \qquad (81)
\end{aligned}
$$

We used the symmetry of the (inverse) covariance matrix.
For example, Δ can be computed as follows:

$$
\Delta_{ij} = \mathbb{E}_Z\left[e^{-rT}\frac{1}{T a_{0,j}}\Sigma_{jk}^{-1}\left(\ln(a_k(T)) - \ln(a_{0,k}) - \left(r - \frac{\sigma_k^2}{2}\right)T\right) x_i(\mathbf{a}_T)\right]
$$

However, this simple approach has undesirably slow convergence and has been disregarded in favor of the approach described in the next section. The variance as function of the number of

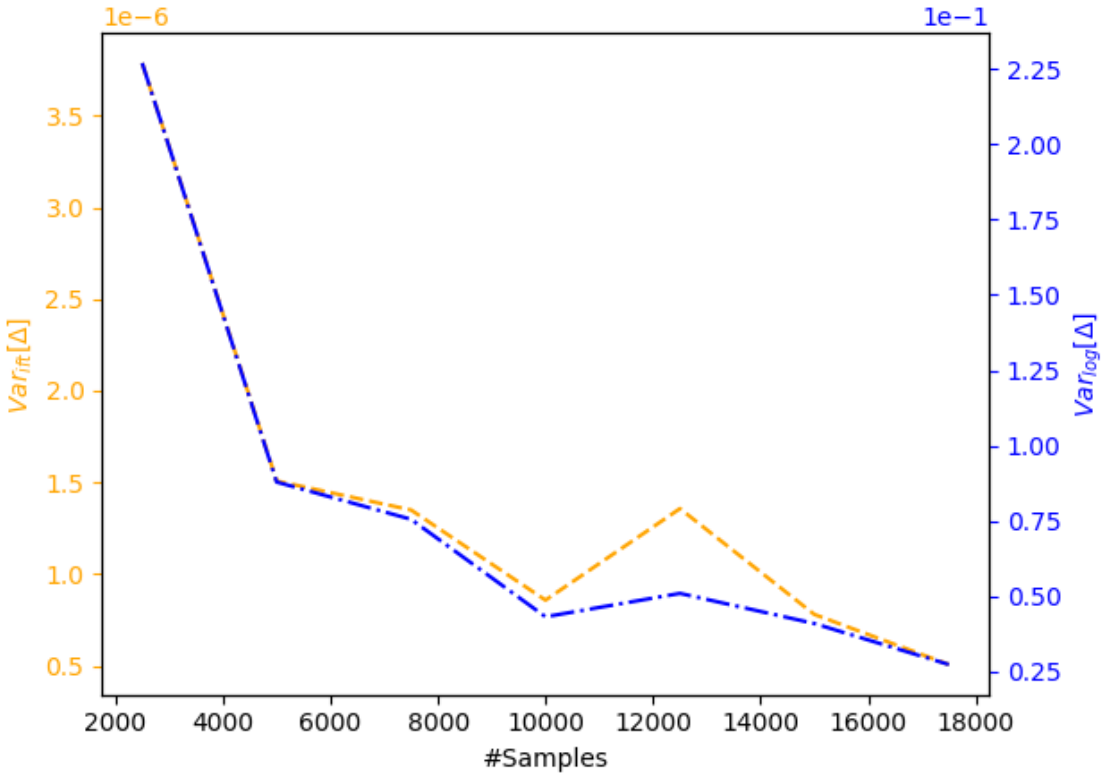


Figure 4: Comparison of variance between log-derivative trick and inverse function theorem. The same sequence of PRNs have been used for both simulations.

samples is shown in fig. 4. One can observe several orders of magnitude difference in the variance. A strong correlation between both variances for the same PRNs can be observed as well, leaving the inverse function theorem method as the better option.

4.3.2 *Implicit Function Theorem*

It is also possible to compute the Greeks from our fixed point iteration using the implicit function theorem eq. 73.

$$\frac{\partial}{\partial \theta_j}\mathbb{E}_Q\left[e^{-rT}x_i\left(\mathbf{a}\left(Z\right)\right)\right] = \mathbb{E}_Z\left[e^{-rT}\left[B\left(\mathbf{a},\mathbf{x}\left(\mathbf{a}\right)\right)\right]_{ik}\frac{\partial g_k\left(\mathbf{a},\mathbf{x}\left(\mathbf{a}\right)\right)}{\partial a_l}\frac{\partial a_l\left(Z\right)}{\partial \theta_j}\right] \tag{82}$$

The result is a $2N \times M$ matrix for $1 \leq l \leq M$.
We now compute the first order Greeks using eq. 82.

$$\begin{aligned}
\Delta_{ij} &= \frac{\partial}{\partial a_{0,j}}\mathbb{E}\left[e^{-rT}x_i(\mathbf{a}_T)\right] \\
&= \mathbb{E}_Z\left[e^{-rT}\left[B\left(\mathbf{a},\mathbf{x}\left(\mathbf{a}\right)\right)\right]_{ik}\frac{\partial g_k\left(\mathbf{a},\mathbf{x}\left(\mathbf{a}\right)\right)}{\partial a_l}\frac{\partial a_l\left(Z\right)}{\partial a_{0,j}}\right] \\
&= \mathbb{E}_Z\left[e^{-rT}\left[B\left(\mathbf{a},\mathbf{x}\left(\mathbf{a}\right)\right)\right]_{ik}\frac{\partial g_k}{\partial a_l}\delta_{jl}e^{(r-\sigma_l^2/2)T+\sqrt{T}L_{lm}Z_m}\right] \\
\Rightarrow \Delta &= \mathbb{E}_Z\left[e^{-rT}\left[B\left(\mathbf{a},\mathbf{x}\left(\mathbf{a}\right)\right)\right]J_{g,\mathbf{a}}\mathrm{diag}\left(e^{(r-\boldsymbol{\sigma}^2/2)T+\sqrt{T}L\mathbf{Z}}\right)\right] \\
&= \mathbb{E}_Z\left[e^{-rT}\left(\mathbb{1}-J_{g,\mathbf{a}}M\right)^{-1}J_{g,\mathbf{a}}\mathrm{diag}\left(e^{(r-\boldsymbol{\sigma}^2/2)T+\sqrt{T}L\mathbf{Z}}\right)\right]
\end{aligned}$$

For Vega we get:

$$\begin{aligned}
\nu_{ij} &= \frac{\partial}{\partial \sigma_j}\mathbb{E}\left[e^{-rT}x_i(\mathbf{a}_T)\right] \\
&= \mathbb{E}\left[e^{-rT}\left[B\left(\mathbf{a},\mathbf{x}\left(\mathbf{a}\right)\right)\right]_{ik}\frac{\partial g_k}{\partial a_l}\frac{\partial a_l(Z)}{\partial \sigma_j}\right] \\
&= \mathbb{E}\left[e^{-rT}\left[B\left(\mathbf{a},\mathbf{x}\left(\mathbf{a}\right)\right)\right]_{ik}\frac{\partial g_k}{\partial a_l}\left(-T\delta_{lj}\sigma_l+\sqrt{T}\delta_{lj}L_{lm}Z_m\right)a_{0,l}e^{(r-\sigma_l^2/2)T+\sqrt{T}L_{lm}Z_m}\right] \\
\Rightarrow \nu &= \mathbb{E}_Z\left[e^{-rT}B\left(\mathbf{a},\mathbf{x}\left(\mathbf{a}\right)\right)J_{g,\mathbf{a}}\mathrm{diag}\left(a_{0,l}\left(-T\boldsymbol{\sigma}+\sqrt{T}LZ\right)e^{(r-\boldsymbol{\sigma}^2/2)T+\sqrt{T}L\mathbf{Z}}\right)\right]
\end{aligned}$$

For θ:

$$\begin{aligned}
\theta_i &= \frac{\partial}{\partial T}\left[e^{-rT}x_i(\mathbf{a}_T)\right] \\
&= -r\mathbb{E}\left[e^{-rT}x_i(\mathbf{a}_T)\right]+\mathbb{E}\left[e^{-rT}\left[B\left(\mathbf{a},\mathbf{x}\left(\mathbf{a}\right)\right)\right]_{ik}\frac{\partial g_k}{\partial a_l}\frac{\partial a_l(Z)}{\partial T}\right] \\
&= -r\mathbb{E}\left[e^{-rT}x_i(\mathbf{a}_T)\right]+\mathbb{E}\left[e^{-rT}\left[B\left(\mathbf{a},\mathbf{x}\left(\mathbf{a}\right)\right)\right]_{ik}\frac{\partial g_k}{\partial a_l}\left[\left(r-\frac{\sigma_l^2}{2}\right)+\frac{L_{lm}Z_m}{2\sqrt{T}}\right]a_l(Z)\right] \\
\Rightarrow \theta &= \mathbb{E}\left[e^{-rT}\left[B\left(\mathbf{a},\mathbf{x}\left(\mathbf{a}\right)\right)J_{g,\mathbf{a}}\left(r-\frac{\boldsymbol{\sigma}^2}{2}+\frac{L\mathbf{Z}}{2\sqrt{T}}\right)\mathbf{a}(\mathbf{Z})-rx(\mathbf{a}_T)\right]\right]
\end{aligned}$$

And finally for ρ :

$$\begin{aligned}
\rho_i &= \frac{\partial}{\partial r}\left[e^{-rT}x_i(\mathbf{a}_T)\right] \\
&= -T\mathbb{E}\left[e^{-rT}x_i(\mathbf{a}_T)\right]+\mathbb{E}\left[e^{-rT}\left[B\left(\mathbf{a},\mathbf{x}\left(\mathbf{a}\right)\right)\right]_{ik}\frac{\partial g_k}{\partial a_l}Ta_l(Z)\right] \\
\Rightarrow \rho &= \mathbb{E}\left[Te^{-rT}\left[B\left(\mathbf{a},\mathbf{x}\left(\mathbf{a}\right)\right)J_{g,\mathbf{a}}\mathbf{a}(\mathbf{Z})-x(\mathbf{a}_T)\right]\right]
\end{aligned}$$

These estimators can be used for any network structure with either analytically calculatable Jacobians or numerical ones. We will later also compare the results with the well known Black-Scholes Greeks for a suitable edge case.

5 MONTE CARLO SAMPLING OF RANDOM NETWORKS

5.1 INTRODUCTION

While analytic methods are detrimental to the understanding of a subject, they can lead to intractable expressions such as complicated integrals or may just not be available at all. Often numerous techniques for a numerical or analytical approximation of the original problem are developed. However, one of the most robust and widely usable options are Monte Carlo methods, only rivaled in popularity by finite difference schemes. Both approaches are popular not because of their efficiency or error robustness, but because they are applicable to almost any problem. This also lends them the title of "brute force methods". For this work, we will only need to briefly discuss the main aspects of Monte Carlo sampling, because we only need a basic expected value estimate and well know convergence improvements (variance reduction). Let $f : \Omega \to \mathbb{R}$ and $\Omega \subset \mathbb{R}$ and $X \sim \rho_X$, then the expected value is written as, so the general expected value

$$\mathbb{E}_{\rho_X}[f] = \int f(x)\, \mathrm{d}\mu_X(x) \tag{83}$$

becomes the well known Lebesgue integral:

$$\mathbb{E}_{\rho}[f] = \int f(x)\rho(x)\, \mathrm{d}x \tag{84}$$

We can estimate this integral by drawing samples $x \in S$ with $S \subset \Omega$ from the distribution $\rho_X(x)$, so that

$$\langle f \rangle = \frac{1}{|S|} \sum_{x \in S} f(x) \tag{85}$$

From the central limit theorem it is clear that this is convergent with a rate of $O\left(\frac{1}{\sqrt{N}}\right)$:

$$\begin{aligned} \left| \left(\frac{1}{N} \sum_{x \in S} f(x) \right) - \mathbb{E}[f] \right| &= \left| \sqrt{\frac{\mathrm{Var}[f]}{N}} \frac{\sum_{x \in S} \left(f(x) - N\mathbb{E}[f] \right)}{\sqrt{N \cdot \mathrm{Var}[f]}} \right| \\ &= \sqrt{\frac{\mathrm{Var}[f]}{N}} |\zeta| \end{aligned}$$

with $\zeta \sim \mathcal{N}(0, 1)$.
Often the distribution of X is unknown and difficult to approximate. The simplest and therefore often also fastest method for an approximation is rejection sampling[1]. Let $p(x)$ be the distribution we want to approximate. Then we can find a suitable envelop distribution $q(x)$ over the same support Ω and a scaling factor c, such that

$$p(x) \leq cq(x), \qquad \forall x \in \Omega$$

We now draw samples $x \sim q$ and subsequently from the uniform distribution over $[0, cq(x)]$:

$$\zeta \sim \mathcal{U}(0, c \cdot q(x))$$

The resulting 2-dimensional random variable (x, ζ) has a uniform distribution under the graph of $c \cdot q(x)$. We now only accept (i.e. don't throw away) samples, that also lie below the graph of $f(x)$, thereby keeping the uniform distribution. The resulting algorithm is given in alg. 1. Of course the convergence drastically depends on the similarity between p and q, especially in high dimensions,

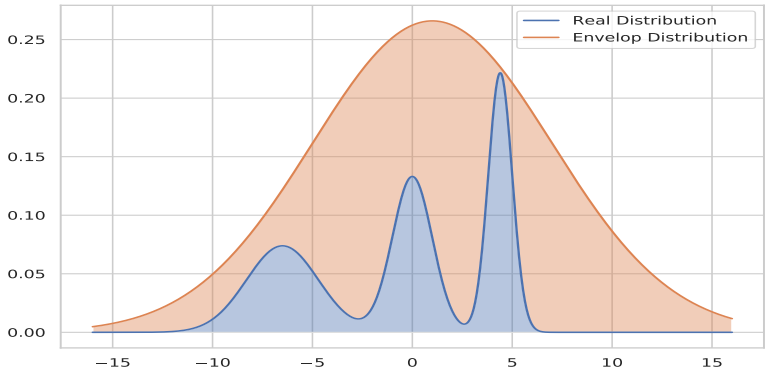


Figure 5: Envelop and real distribution for rejection sampling. The red samples falling into the red area are rejected.

Pseudocode 1 Rejection Sampling

function RejectionSampling(distribution p, envelope q, factor c, sampler for $\mathcal{U}[0, c \cdot q(x)]$, number of samples N)
  repeat
    Draw $x_i \sim q$
    Draw $\zeta \sim U(0, c \cdot q(x_i))$
    if $\zeta < p(x)$ **then**
      Add x_i to p_{sample}
    end if
  until size of $p_{\text{sample}} > N$
return p_{sample}
end function

where much of the weight is concentrated near the surface of the sampled manifold. It is possible to improve this method by making the envelope adaptive, i.e. decreasing $q(x_i)$ whenever x_i is rejected. Rejection sampling can be surprisingly efficient for computationally inexpensive function evaluation, due to the simplicity of the algorithm (e.g. only a single call of f).

5.2 VARIANCE REDUCTION

For sufficiently computationally inexpensive functions f rejection sampling is surprisingly fast (see also alg. 3 for an example). However, for cases with undesirably slow convergence, there is a class of algorithms available, called variance reduction methods. They are not applicable in all situations and can actually increase the variance when applied badly. We will now discuss the methods employed during this thesis briefly.

1 Another popular method is the use of Markov chains to approximate an arbitrary distribution.

5.2.1 *Control Variates*

Arguably the simplest method for variance reduction are control variates. Here, one introduces a function g which is similar[2] to f with known integral $I(g)$.

$$\begin{aligned} \mathbb{E}[f(X)] &= \mathbb{E}[f(X)] + \beta\mathbb{E}[g(X) - g(X)] \\ \Rightarrow \langle f(X)\rangle_{\mathrm{CV}_g} &= \langle f(X)\rangle + \beta\left(\mathbb{E}[g(X)] - \langle g(X)\rangle\right) \end{aligned}$$

With $\beta \in \mathbb{R}$ some arbitrary parameter. This reduces the variance of the sampled quantity if the variance of the CV estimator is lower than that of the original one:

$$\sigma^2_{\mathrm{CV}f} = \left(\sigma_f^2 + \beta^2\sigma_g^2 - 2\beta\rho_{fg}\sigma_f\sigma_g\right) < \sigma_f^2 \tag{86}$$

$$\Leftrightarrow \sigma_X^2 < 2\rho_{fg}\sigma_f\sigma_g \tag{87}$$

Where we used the Pearson correlation coefficient as an estimator for the true covariance:

$$\rho_{XY} \approx \frac{\mathrm{Cov}[X,Y]}{\sigma_X\sigma_Y} \tag{88}$$

Minimization yields:

$$\beta^* = \frac{\rho_{fg}\sigma_f}{\sigma_g} \tag{89}$$

$$\Rightarrow \sigma^2_{\mathrm{CV}f} = \left(1 - \rho_{fg}^2\right)\sigma_f^2 \tag{90}$$

Eq. 90 shows, that we need a highly correlated control variate in order to minimize the variance. A well known example, closely related to our application, is the estimation of pricing for basket options [Du, Liu, and Gu 2014]. Let W_i with $1 \leq i \leq N$ be independent Wiener processes:

$$S_i(T) = x_i \exp\left\{\left(r - \sum_{j=1}^{N}\frac{\sigma_{ij}^2}{2}\right)T + \sum_{j=1}^{N}\sigma_{ij}^2 W_T^j\right\} \tag{91}$$

Suppose we want to compute the basket option:

$$\mathbb{E}[\max\{K - X, 0\}] \tag{92}$$

with $\sum_i^M a_i = 1$, $m = \sum_i^M a_i x_i$ and

$$X = \sum_i^N a_i S_i(T) \tag{93}$$

$$\Rightarrow \frac{X}{m} = \sum_i^M \frac{a_i x_i}{m}\exp\left\{rT + \sum_{j=1}^{N}\sigma_{ij}W_T^j\right\} \tag{94}$$

The expected value can now be estimated using Monte Carlo sampling. However, there is an analytically solvable edge case, by replacing the stochastic volatilities σ_{ij} with square-integrable functions:

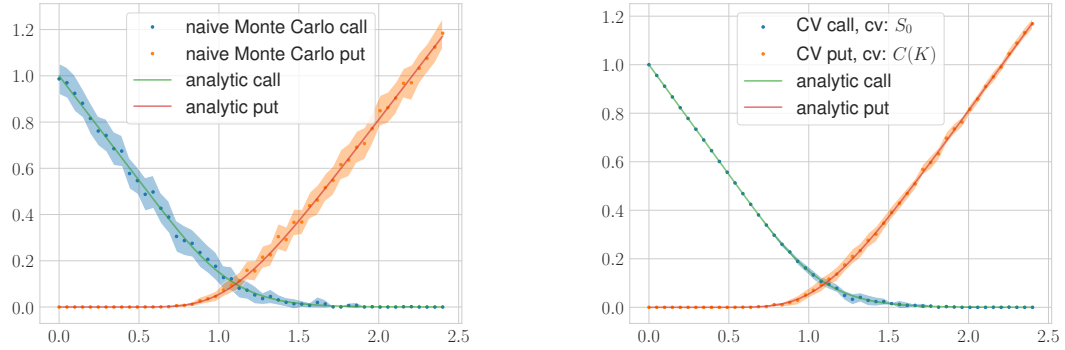
$$\mathbb{E}[\max\{K - Y, 0\}] = \mathbb{E}\left[\max\left\{K - m\prod_{i=1}^{M}\exp\left\{a_i x_i\left(rT + \sum_{j=1}^{N}\sigma_{ij}W_T^j\right)\right\}^{\frac{1}{m}}, 0\right\}\right] \tag{95}$$

This can be evaluated analytically using the log-normal valuation formula and allows for application of the control variate method by sampling from:

$$\max\{K - X, 0\} - \max\{K - Y, 0\} \tag{96}$$

As an even more basic example, we show the sampling of a European call option with the spot price S_0 and the payoff of the put option as control variate in fig. 6.

2 Here similar only means that the new estimator has a smaller variance under Monte Carlo sampling than the old one, for example by following the fluctuations of f.



(a) naive Monte Carlo sampling

(b) control variates: S_0 for call, $C(K)$ for put

Figure 6: European call and put payoffs, variance comparison with and without control variates. Notice, that different control variates were used for call and put.

5.2.2 *Antithetic variates*

Whenever we can transform the sampled random variable without changing its law, these additional samples can be used to lower the variance. For example, we generate i.i.d. PRNs for our simulation, and can modify the estimator as follows:

$$X_i \sim \mathcal{N}(0,1), \qquad X_i' = -X_i, \quad 1 \leq i \leq N \tag{97}$$

$$\Rightarrow \langle f \rangle = \frac{1}{2N}\left[\sum_{X_i} f(X_i) + \sum_{X_i'} f(X_i')\right] \tag{98}$$

While not always applicable or efficient (both in the number of calls to f and the resulting variance reduction), this method is very easy to implement. Figure 7 shows a comparison of the antithetic sampling method with the previously discussed control variate method.

5.2.3 *Importance Sampling*

Instead of changing the sampled quantity, it is also possible to change the underlying law from which samples are generated. This is done by introducing a new density $\tilde{\rho}(x)$ with $\tilde{\rho}(x) > 0$ for all x in the support of ρ:

$$\mathbb{E}[f(X)]_\rho = \int f(x)\rho(x)\,\mathrm{d}x = \int f(x)\frac{\rho(x)}{\tilde{\rho}(x)}\tilde{\rho}(x)\,\mathrm{d}x \tag{99}$$

$$= \int f(x)w(x)\tilde{\rho}(x)\,\mathrm{d}x \tag{100}$$

$$= \mathbb{E}[w(X)f(X)]_{\tilde{\rho}} \tag{101}$$

One extremely valuable property of this method (making this the almost exclusively used variance reduction method in statistical physics) is the ability to estimate densities with unknown normalization constants Z directly.

$$1 = \frac{1}{Z}\int \rho(x)\,\mathrm{d}x = \frac{1}{Z}\int w(x)\tilde{\rho}(x)\,\mathrm{d}x$$
$$= \int p(x)\tilde{\rho}(x)\,\mathrm{d}x$$

The overall estimator is then given as:

$$\mathbb{E}[f(X)] = \frac{\int w(x)f(x)\,\mathrm{d}x}{\int w(x)\,\mathrm{d}x} = \int p(x)f(x)\,\mathrm{d}x \tag{102}$$

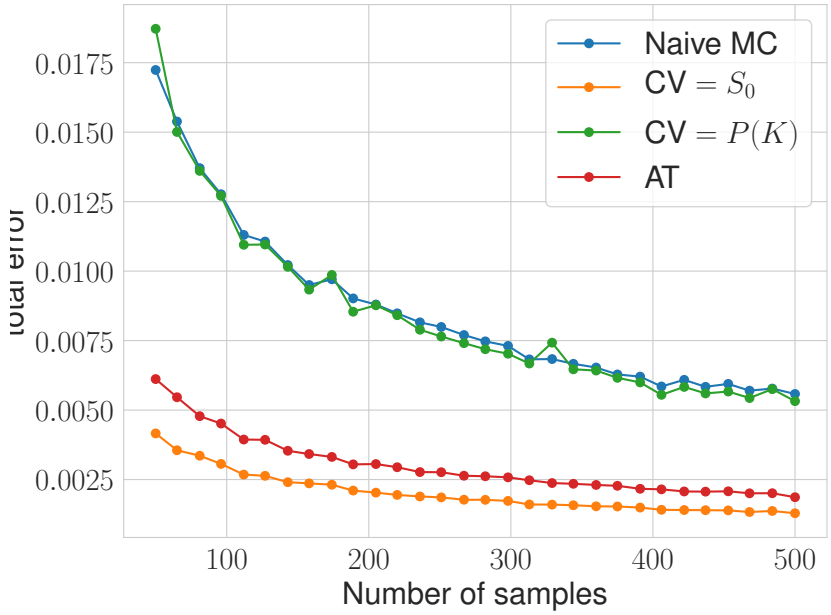


Figure 7: Comparison of variance reduction for control variate and antithetic sampling

The variance is reduced if $\text{Var}_{\tilde{\rho}}[w(X)f(X)] < \text{Var}_{\rho}[g(X)]$

$$\text{Var}_{\tilde{\rho}}[w(X)f(X)] = \int w(x)f^2(x)\rho(x)\,\mathrm{d}x - \mathbb{E}^2[f(X)] \tag{103}$$

The advantage of importance sampling comes from the fact, that the weights can amplify the contributions for which $f(x)$ is large, but $\rho(x)$ is small and vice versa. A usual example is the sampling of rare events with large impact, for example financial crises. The full algorithm is given in alg. 2. For the computation of the weight the log-sum-exp trick is used. This reduces the numerical inaccuracy which would be introduced by dividing very small probabilities. In fact, the optimal choice

Pseudocode 2 Importance Sampling

```
function ImportanceSample(p, q, f, sampler for X ~ q, N)
    repeat
        Draw X_i ~ q
        Set w_i = exp{log(p(X_i)) − log(q(X_i))}
        Set s_i = f(X_i)
        Set res += exp{ log(p(X_i)) − log(q(X_i)) + log(f(X_i)) }
    until size of i > N
return res
end function
```

for $\tilde{\rho}^\star$ (i.e. a density minimizing $\text{Var}_{\tilde{\rho}}[f]$) is proportional to $|f(x)|\rho(x)$ [Kahn and Marshall 1953]. For example, we can chose

$$\tilde{\rho}^\star(x) = \frac{|f(x)|\,\rho(x)}{\mathbb{E}_{\rho}[|f(X)|]} \tag{104}$$

Then, we can use the usual variance formula in terms of raw moments to show that the optimality of the density:

$$\begin{aligned}
\mathrm{Var}_{\tilde{\rho}^\star}[f] + \mathbb{E}^2_{\tilde{\rho}^\star}[f] = \mathbb{E}_{\tilde{\rho}^\star}\left[f^2\right] &= \int \frac{f^2(x)\rho^2(x)}{\tilde{\rho}^\star(x)}\,\mathrm{d}x && (105)\\
&= \int \frac{\mathbb{E}_\rho[|f(X)|]f^2(x)\rho^2(x)}{|f(x)|\rho(x)}\,\mathrm{d}x \\
&= \mathbb{E}^2_\rho[|f(X)|] \\
&= \mathbb{E}^2_{\tilde{\rho}}[|w(X)f(X)|] \\
&\leq \mathbb{E}_{\tilde{\rho}}\left[w^2(X)f^2(X)\right] && (106)\\
&= \mathrm{Var}_{\tilde{\rho}^\star}[f] + \mathbb{E}^2_{\tilde{\rho}}[f] && (107)
\end{aligned}$$

In eq. 106 we have used Jensen's inequality. This proofs that $\rho^\star$ is indeed the minimal choice of $\tilde{\rho}$. The insight we gain from this proof however, lies not in in the fact that we can give such a formula for the minimal choice (since the normalizing factor is precisely the quantity we wanted to compute in the first place), but rather in the influence of $\tilde{\rho}$ on the variance. Since $\tilde{\rho}(x)$ appears in the denominator of the expected value, small values of the importance density can be problematic. In general this means that we want heavy tailed importance distributions[3], in order to prevent large variances. We will make use of this observation when choosing the Student's t-distribution as an importance density for the normal distribution.
There are a couple of diagnostic measures, quantifying the quality of the importance sampled estimator. Consider the situation, where very few weights are many orders of magnitude larger, than the rest. In this case it is safe too assume that importance sampling did more harm than good, since we are effectively only using these few samples in the computation of the expected value. One can extend this notion to more general cases by starting with assumption of N i.i.d. samples X_i:

$$R_w = \frac{\sum_i w_i X_i}{\sum_i w_i} \tag{108}$$

One can now set the variance to the assumed variance of X_i,i.e. $\frac{\sigma^2}{n_e}$, and solve for the effective sample size n_{eff}:

$$n_{\text{eff}} = \frac{\left(\sum_i w_i\right)^2}{\sum_i w_i^2} \tag{109}$$

Another often used measure [Gubernatis, Kawashima, and Werner 2016]:

$$n_{\text{eff}} = \frac{N}{1 + cv^2(w)} \tag{110}$$

$$cv(w) = \sqrt{\frac{1}{N-1}\sum_i \left(w_i - \langle w\rangle\right)^2} \tag{111}$$

Where cv is the coefficient of variation. The easier to compute variance of the estimated mean can lull into false security, because heavily skewed weights may produce a small but false variance.

5.3 PSEUDO RANDOM NUMBERS

Computers are deterministic Turing machines and can therefore not produce true randomness without the help of external devices. These devices are available for specialized applications, but

3 A distribution function $F_X(x)$ is said to have heavier tails than a distribution function $F_Y(x)$ if the survival function of X, $S_X(x) = 1 - F_X(x)$, is larger than the survival function of Y for large x.

require a lot of additional consideration to ensure uncorrelated samples. Monte Carlo applications require a lot of random numbers, distributed over multiple instances of the program. This is due to the fact, that Monte Carlo programs are naively parallelizable, using the linearity of the expected value. These random numbers must not only be uncorrelated among themselves, but also between different processes, without introducing a large overhead due to inter-process communications[4]. Uncorrelated samples are needed in order to prevent a systematic bias of the Monte Carlo results. There are some options to detect correlations between successive samples (see sec. 5.5) but in general there is no guarantee for that. However, there exists a family of algorithms which can produce pseudo random numbers (PRN) from deterministic operations. They are called pseudo random number generators (PRNGs).
We will now discuss how to generate uniformly distributed PRNs and the transforms to other distributions, needed for this thesis. In order to demonstrate problematic behavior and possible improvements, it is sufficient to consider one of the simplest PRNGs. The linear congruential generator is based on the well known primitive roots. b is a primitive root modulo n iff:

$$b^i \equiv j \pmod{n}, \qquad 1 \leq i \leq n \tag{112}$$

i is called discrete logarithm and an example of a candidate for a one-way function[5]. j will take all values between 1 and n in a pseudo-random order. We can now define the Lehmer random number generator:

$$X_{k+1} \equiv aX_k \pmod{n} \tag{113}$$

with n prime, g a primitive root of n and the seeds X_0 coprime to n. Because of eq.112, the pseudo random numbers X_k have a period l (i.e. $X_{k+l} = X_k$), of $n - 1$. This algorithm is a special case of the well known linear congruential generator (LCG) with offset 0.

$$X_{k+1} \equiv aX_k + b \pmod{n} \tag{114}$$

Period and conditions for the parameters can be obtained by virtue of the Hull-Dobell Theorem. The LCG class of generators already provides some useful properties. One of the most important ones in the context of large scale Monte Carlo simulations is the ability to jump ahead without generating all numbers in between. For the Lehmer generator we have:

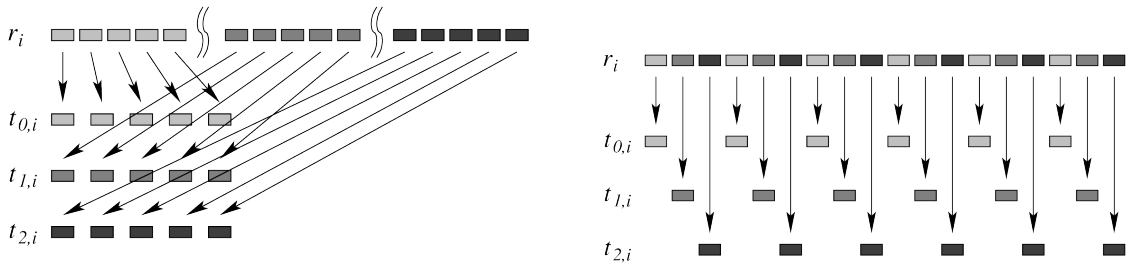
$$X_k = a^n X_0 \tag{115}$$

$$\Rightarrow X_k \equiv a^n X_0 \pmod{n} \tag{116}$$

With an additional offset b it is still possible to fast forward and directly generate a PRN multiple steps ahead, but the formula becomes slightly more complicated. We will use this property in order to "jump ahead" and generate multiple streams of random numbers for different instances of the program. This is important, since other methods (e.g. choosing different seeds) could in general lead to correlated behavior. For example, the compiler could optimize similar code and force all seeds to be the same or the instantiation time of all processes (which is sometimes used as seed) could be the same, see also fig. 9. In order to distribute random numbers over different processes, the main two approached are block splitting and leapfrogging, see fig. 8b. If the number of needed random numbers is not known in advance the leapfrog scheme is of course more advantageous, since we can exclude the possibility of collisions from an overlap of streams. Still, the ability to generate independent streams is valuable and the periodicity is long enough, to make block splitting a good choice whenever the random number generator allows inexpensive jump ahead. For this reason I have chosen to use the block splitting for the internal MPI parallelization. Behavior like shown in fig. 9 is extremely hard to spot and can lead to systematic bias. For example,

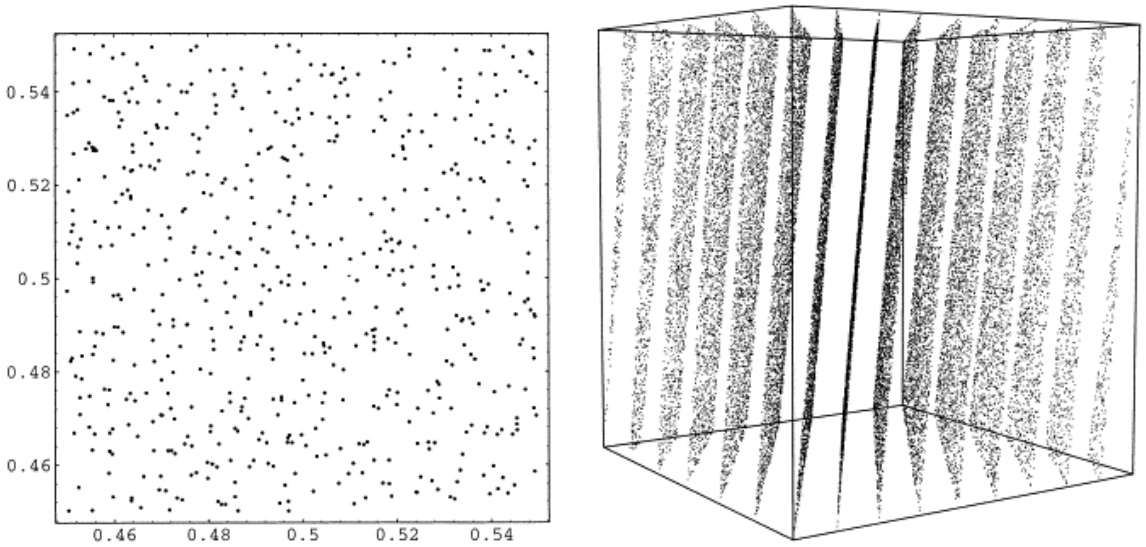
4 For the purpose of Monte Carlo applications, we do not require quite as many properties as for encryption. For example, the seed of linear congruential generators can be reconstructed using the LLL-algorithm [A. M. Frieze et al. 1988], but these types of generators are still suitable for Monte Carlo applications.

5 As mentioned before, the discrete logarithm may actually turn out to be not a one-way function and is subject to ongoing research. This gives a hint, why the LCG class of generators is not considered difficult to predict and therefore unsuitable for cryptography applications.



(a) Blocking scheme (b) Leapfrog scheme

Figure 8: Parallelization approaches for streams of PRNGs, from [Hellekalek 1998]



(a) LCG in two dimensions (b) LCG in three dimensions

Figure 9: Problematic behavior with LCG type random numbers in high dimensions [Hellekalek 1998]

if a lot of weight is concentrated in gaps that can never be sampled, due to gaps in the space of possible PRNs. There are established test suites available that are specifically designed to test new algorithms for known shortcomings and old ones for newly discovered flaws. The best known ones are Diehard and TestU01 [Robert G. Brown 2017; L'Ecuyer and Simard 2007]. I have decided to use TRNG, due to the statistics framework I have developed in a previous project around this library [Bauke and Mertens 2007]. This library has proven work for large scale physics simulations and explicitly tested against some of them. Since we only need a comparatively small amount of random numbers for this work, the library can be considered stable enough.

5.4 TRANSFORMATION OF RANDOM VARIABLES

Until now we have only considered generation of uniformly distributed pseudo random numbers. In our application we will however need variables, that are distributed according to the log-normal and student t-distribution. Usually, transformation algorithms come packaged with PRNG libraries such as TRNG. We do however require more control over the transformation, in order to apply importance sampling. Additionally, the multivariate versions of these distributions are rarely available in libraries.

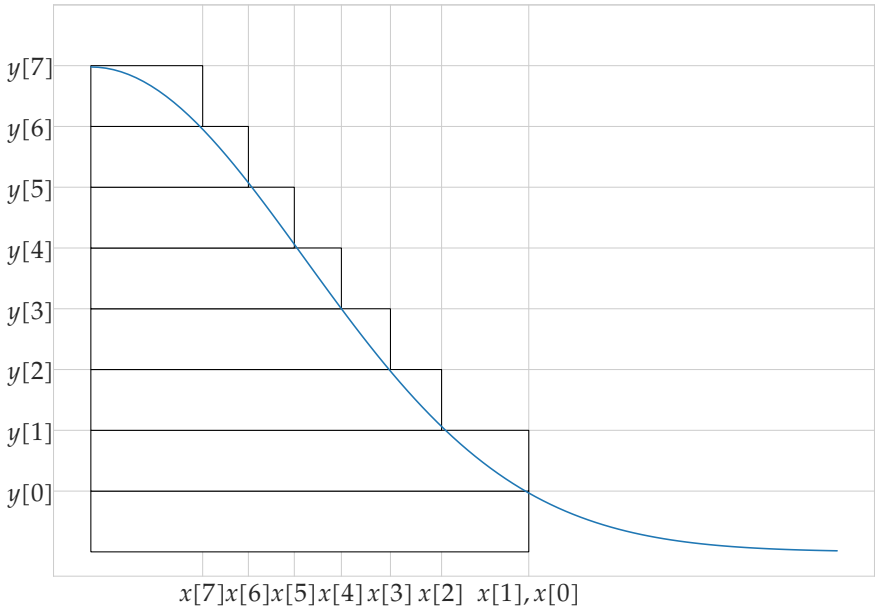


Figure 10: Ziggurat construction

5.4.1 *Transformation to the normal distribution*

Of course any transformation can naively be done using rejection sampling. But especially for high dimensions, this approach can become arbitrarily inefficient. The direct transformation from uniform to normally distributed random variables can be done more efficiently by using for example the Marsaglia polar, Box-Muller or Ziggurat algorithm. The first two algorithms use special formulas obtained by analytic solution of inversion sampling. For example two independent normally distributed samples $Z_0, Z_1 \sim \mathcal{N}(0,1)$ can be obtained through the Box-Muller transform from $U_0, U_1 \sim U(0,1)$ as follows:

$$Z_0 = \sqrt{-2 \ln U_0} \cos(2\pi U_1) \tag{117}$$

$$Z_1 = \sqrt{-2 \ln U_0} \sin(2\pi U_1) \tag{118}$$

Interestingly, both direct algorithms are considered slow alternatives to a slightly improved version of rejection sampling, the Ziggurat algorithm [Malik and Hemani 2016]. We will briefly discuss the Ziggurat algorithm as a successful example of the rejection sampling method. Here the right half of the probability density function[6] is segmented into horizontal, rectangular slices.

A slice i is randomly selected, then a uniform random number U_0 is generated and a rejection candidate $x = U_0 x_i$ is computed. If $x < x_{i+1}$ then the candidate x is rejected. This can be seen in fig. 10. Any slice $x[x+1]$ must be fully under the curve. The rejection rate is controlled by the number of slices and is therefore substantially lower than for the naive rejection algorithm. The comparison is also faster, because the comparison with a computationally expensive function is replaced by the comparison with a floating point value. In case the sample is larger than the right edge of $x[i+1]$, there is still the possibility of x lying under the curve. In this case, the normal rejection comparison has to be used. TRNG comes with an implementation of such a transformation algorithm, the value in discussing the Ziggurat algorithm comes from the fact, that this is an example for a fast and reliable rejection sampling algorithm.

Having obtained $\mathcal{N}(0,1)$ distributed PRNGs, the transformation to multivariate normal distribu-

6 Due to the symmetry of the normal distribution, one can randomly flip the sign of the generated sample.

Pseudocode 3 Ziggurat Algorithm

```
procedure ZIGGURAT(uniform PRNG, normal pdf f, lists of rectangle borders x[] and y[])
    Set r = None
    repeat
        Randomly choose slice i
        Generate U_0 ~ U(0,1)
        Set r = U_0 x[i]
        if r < x[i+1] then
            return r
        end if
        Generate U_1 ~ U(0,1)
        Set y = y[i] + y[i+1]U_1 - y[i]U_1
        if y < f(r) then
            return r
        end if
    until Sample is generated
return r
end procedure
```

tions with different means and variances is straight forward by virtue of the change of variables formula:

$$f_Y(y) = \left|\det J_{g^{-1},y}\right| \cdot f_X(g^{-1}(y)) \tag{119}$$

The multivariate density $\mathcal{N}(0_N, \mathbb{1}_N)$ is easy to calculate. Let f_{Z_i} be the density for a univariate normal distribution, then the multivariate version $f_{\mathbf{Z}}(\mathbf{x})$ can be constructed as a product:

$$\begin{aligned} f_{\mathbf{Z}_1}(\mathbf{x}) &= \prod_{i=1}^{N} f_{Z_i}(x_i) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} \\ &= \frac{1}{(2\pi)^{N/2}} e^{-\frac{1}{2}\mathbf{x}\cdot\mathbf{x}} \end{aligned}$$

This can now be transformed to arbitrary mean $\boldsymbol{\mu} \in \mathbb{R}^N$ and and covariance matrix $\Sigma \in \mathbb{R}^{N\times N}$. First we decompose Σ using the Cholesky decomposition $\Sigma = LL^T$. This is always possible due to the properties of the covariance matrix. Now we let $g(\mathbf{Z}) = L\mathbf{Z} + \boldsymbol{\mu}$. With the transformation formula (eq. 119) and the previously obtained transformation for $\mathcal{N}(0_N, \mathbb{1}_N)$, we obtain the density for multivariate normal distributions $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$:

$$\begin{aligned} f_{\mathbf{Z}} &= f_{\mathbf{Z}_1}\left(g^{-1}(\mathbf{x})\right) \cdot \left|\det J_{g^{-1},y}\right| \\ &= f_{\mathbf{Z}_1}\left(L^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right)\right) \frac{1}{\sqrt{\det \Sigma}} \\ &= \frac{1}{\sqrt{(2\pi)^N \det \Sigma}} \exp\{\frac{1}{2}\left(L^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \cdot \left(L^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)\} \\ &= \frac{1}{\sqrt{(2\pi)^N \det \Sigma}} \exp\{\frac{1}{2}\left((\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)\} \end{aligned}$$

5.4.2 *Transformation to log-normal distribution*

For our simulations we also require the log-normal and student t-distribution. The first one is easily obtained from normally distributed samples following its definition: We call a random variable X log-normally distributed iff

$$\log(X) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \tag{120}$$

This is of course easily inverted to yield an expression that transforms a $Z \sim \mathcal{N}(\mathbf{0}, 1)$ into $X \sim \text{Lognormal}(\boldsymbol{\mu}, \Sigma)$:

$$X = e^{\mu + LZ} \tag{121}$$

With $\Sigma = LL^T$ as before.
The log-normal distribution has the following properties:

PDF $p(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$

expected value $\mathbb{E}[X] = e^{\mu + \frac{\sigma^2}{2}} \Rightarrow \mathbb{E}[S_t] = S_0 e^{\mu\tau}$

variance $\text{Var}[X] = \left(e^{\sigma^2} - 1\right) e^{2\mu+\sigma^2} \Rightarrow \text{Var}[S_t] = \left(e^{\sigma^2\tau} - 1\right) S_0^2 e^{2\mu\tau}$

In order to fix the mean asset price at $\langle S_t \rangle \approx 1$, we also fix $S_0 = \frac{1}{\mu} = e^{-rt + t\frac{\sigma^2}{2}}$ according to the definition of the expected value above.

5.4.3 *Transformation to student-t distribution*

The transformation from normal to t-distribution is a bit more involved but still feasible with only a few operations. In addition to the normally distributed random variable Z we also need $W \sim \chi^2_\nu$. The χ^2 distribution with ν degrees of freedom has the following probability density function:

$$f_W(w) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} w^{\nu/2-1} e^{-w/2} \tag{122}$$

For a proof of the transformation, we also need the χ_ν distribution with ν degrees of freedom. Its PDF is:

$$f_V(v) = \frac{1}{2^{\nu/2-1}\Gamma(\nu/2)} v^{\nu-1} e^{-v^2/2} \tag{123}$$

Theorem 6 (Transformation to student-t distribution) *A random variable T, distributed according to the student t-distribution with ν degrees of freedom and PDF*

$$f_T(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\mu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

can be obtained from $Z \sim \mathcal{N}(0, 1)$ and $W \sim \chi^2_\nu$ using:

$$T = \left(\frac{W}{\nu}\right)^{-\frac{1}{2}} Z \tag{124}$$

Proof Of course $\sqrt{W}$ is χ_ν distributed and we can immediately apply the transformation formula 119 with $\frac{1}{\sqrt{\nu}}$ into eq. 123. Let $V \sim \chi_\nu$, then:

$$\begin{aligned}
\left(\frac{W}{\nu}\right)^{-\frac{1}{2}} = R = \frac{V}{\sqrt{\nu}} \Rightarrow \det \frac{\partial V}{\partial W} &= \sqrt{\nu} \\
\Rightarrow f_R(r) &= f_V(\sqrt{\nu}r) \left| \det \frac{\partial V}{\partial W} \right| \\
&= \frac{1}{2^{\nu/2-1}\Gamma(\nu/2)} (\sqrt{\nu}r)^{\nu-1} e^{-(\sqrt{\nu}r)^2/2} \sqrt{\nu}\Theta(r) \\
&= \frac{1}{2^{\nu/2-1}\Gamma(\nu/2)} \sqrt{\nu}^{\frac{\nu}{2}} r^{\nu-1} e^{-\nu r^2/2} \Theta(r)
\end{aligned}$$

We can now use the transformation of densities for a ratio of two random variables[Curtiss 1941]:

$$f_T(t) = \int_\infty^\infty |r| f_Z(rt) f_R(r)\,\mathrm{d}r \tag{125}$$

Inserting the previously obtained density and using the integral representation of the gamma function: $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\,\mathrm{d}x$.

$$\begin{aligned} f_T(t) &= \int_\infty^\infty |r| f_Z(rt) f_R(r)\,\mathrm{d}r \\ &= \int_\infty^\infty |r| \frac{1}{\sqrt{2\pi}} e^{-\frac{(rt)^2}{2}} \frac{1}{2^{\nu/2-1}\Gamma(\nu/2)} \sqrt{\nu}^{\frac{\nu}{2}} r^{\nu-1} e^{-\nu r^2/2} \Theta(r) \\ &= \frac{\nu^{\frac{\nu}{2}}}{\sqrt{2^{\nu-1}\pi}\Gamma(\frac{\nu}{2})} \int_0^\infty r^\nu e^{-\frac{r^2(\nu+t^2)}{2}}\,\mathrm{d}x \end{aligned}$$

At this point we substitute $x = \frac{1}{2}(n+t^2)r^2$ so that $\mathrm{d}r = 2\left(\frac{2}{n+t^2}\right)^{\frac{1}{2}}\left(\frac{2}{n+t^2}\right)^{-1}$

$$\begin{aligned} &= \frac{\nu^{\frac{\nu}{2}}}{\sqrt{2^{\nu-1}\pi}\Gamma(\frac{\nu}{2})} \int_0^\infty \left(\frac{2}{\nu+t^2}\right)^{\frac{n-1}{2}} y^{\frac{\nu}{2}} e^{-x}\,\mathrm{d}x \\ &= \frac{\nu^{\frac{\nu}{2}}}{\sqrt{2^{\nu-1}\pi}\Gamma(\frac{\nu}{2})} 2^{\frac{n-1}{2}} \nu^{-\frac{n-1}{2}} \left(1+\frac{t^2}{\nu}\right)^{-\frac{n-1}{2}} \Gamma\left(\frac{\nu+1}{2}\right) \\ &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1+\frac{t^2}{\nu}\right)^{-\frac{n-1}{2}} \end{aligned}$$

which is indeed the probability density function of the student-t distribution.

5.5 DATA ANALYSIS AND RESAMPLING METHODS

While Monte Carlo methods are exact in the limit of infinitely many samples, we do only have a finite amount of runtime available. This begs the question, how many samples are sufficient. Sadly, there is no easy answer such as a residual term for analytically controlled methods. In fact, we may be stuck in a local pocket of the sampled probability density (this applies even more directly to the, here mostly omitted, Markov chain Monte Carlo methods) and falsely assume our simulation to be converged according to a usual estimator such as the sample variance. There is a family of algorithms available for the estimation of the precision of statistics such as mean and variance, called resampling methods. For this thesis I used a previously implemented version of such a method, called blocking or binning analysis [Stobbe 2018a]. These methods do not only allow for the estimation of confidence intervals of parameters, but also autocorrelation times in Markov Chain Monte Carlo simulations with low impact on computation times. One could also resort to Baysian statistics. However, bootstrap methods, such as the binning analysis discussed in the following, can be implemented with almost no overhead and are in fact a numerical implementation of non parametric or parametric maximum likelihood methods, without the use of analytic formulas [Sapatinas 2004, p. 8.2.3].

5.6 BLOCKING ANALYSIS

We start by binning our series of samples into blocks, giving us a histogram of the data, i.e. estimated distribution. Our intention is to sample from this approximated distribution and com-

pare results between different configurations of blocks. Let X_k with $1 \leq k \leq N$ be a sample and $pi(i) \in S_n$ some permutation. We define the i-th bin of size l as $S_i^{(l)}$ and initial block size (number of child blocks) b as [Ambegaokar and Troyer 2009]:

$$S_i^{(l)} = \sum_{k=l\cdot i}^{(l+1)i} X_{\pi(k)}, \qquad 0 \leq l \leq \log_b(N) - \log_b(N_{\text{blocks}}) \tag{126}$$

Expected value and variance can be estimated from this block construction as follows (note that we do not gain any benefit from this construction for now):

$$\mathbb{E}\left[\bar{X}_i\right] = \frac{1}{N} \sum_{\pi(j)\in\text{blocks}_i} X_{\pi(j)} \tag{127}$$

$$\text{Var}\left[\bar{X}\right] = \text{Var}\left[\frac{1}{N}\sum_i^N X_i\right] = \text{Var}\left[\frac{1}{N_{\text{blocks}}} \sum_i^{N_{\text{blocks}}} \frac{N_{\text{blocks}}}{N} \sum_{\pi(j)\in\text{blocks}_i}^{N} X_{\pi(j)}\right] \tag{128}$$

$$\approx \text{Var}\left[\frac{1}{N_{\text{blocks}}} \sum_{i=1}^{N_{\text{blocks}}} \bar{X}_i\right] \tag{129}$$

$$\approx \frac{1}{N_{\text{blocks}}^2} \sum_{i=1}^{N_{\text{blocks}}} \left(\bar{X}_i - \bar{X}\right)^2 \tag{130}$$

We can now treat each bin as a new random variable and calculate the desired quantity over this new sets of samples.

$$Z_i = \frac{S_i^{(l)}}{2^l} \tag{131}$$

Note, that we can accumulate samples into a tree hierarchy of bins, as shown in fig. 11. For $\pi(i) \equiv \text{id}(i)$ the space requirements are only logarithmic in the number of samples. Especially

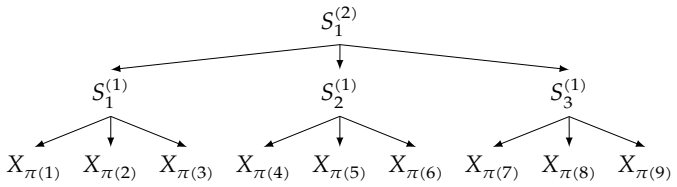


Figure 11: Tree binning structure for blocking analysis

autocorrelation times (number of samples until Markov chain samples become uncorrelated) are easily calculated from the ratio between the variance over all samples and the blocking estimate. However, since we will not be using Markov chain methods, we omit the discussion here and directly discuss the precision estimation of parameters.
We write sampled estimates with a hat, so that an estimated parameter of the distribution P

$$\hat{\theta} = s(P) \tag{132}$$

becomes in the bootstrap approximation

$$\hat{\theta}^\star = s(X) \tag{133}$$

The Bootstrap estimate of the probability P is denoted by $\hat{P}$.
This means, that the naive way to obtain a feature, whereby one samples from an unknown distribution and applies a function s to obtain this feature from the samples, is modified by first generating an empirical distribution from which samples are drawn. One then then generates the feature by repeatedly drawing new samples, obtaining a posterior distribution for θ, due to the knowledge of an estimated prior distribution $\hat{P}$.

Our special case uses blocks of a given size l with i bootstrap replications. The blocks $S_i^{(l)}$ are defined as above and $S^{(l)} = \bigcup_i S_i^{(l)}$ to estimate a distribution $\hat{P}$ in the following form:

$$\hat{\theta}_i^{(l)\star} = s\left(S_i^{(l)}\right) \tag{134}$$

$$\hat{\theta}_{[\cdot]}^{(l)\star} = \sum_{i=1}^{N_{\text{blocks}}} \hat{\theta}_i^{(l)\star} \tag{135}$$

$$\hat{P}(\hat{\theta} \in S^{(l)}) = \frac{1}{N_{\text{blocks}}} \sum_{i=1}^{N_{\text{blocks}}} \mathbb{1}_{S^{(l)}}\left(\hat{\theta}_i^{(l)\star}\right) \tag{136}$$

We can now estimate the standard bootstrap error $\hat{\sigma}$ and from that the standard confidence intervals of the parameter, [Demirözer, Karaca, and Karsavuran 2011, chapter 6, 7].

$$\hat{\sigma}_{\text{boot}}\left(\hat{\theta}\right) = \left[\frac{1}{N_{\text{blocks}} - 1} \sum_{i=1}^{N_{\text{blocks}}} \left(\hat{\theta}_i^{\star} - \hat{\theta}_{[\cdot]}^{\star}\right)^2\right]^{\frac{1}{2}} \tag{137}$$

$$\text{90\% confidence interval: } \left[\hat{\theta} - z_{\alpha/2}\hat{\sigma}_{\text{boot}}\left(\hat{\theta}\right), \hat{\theta} + z_{\alpha/2}\hat{\sigma}_{\text{boot}}\left(\hat{\theta}\right)\right] \tag{138}$$

Bootstrap estimators for the t intervals are derived for example in [Demirözer, Karaca, and Karsavuran 2011, chapter 8].
We now have a setup to quantify data sufficiency and posterior distributions of parameters based on our Monte Carlo samples. These are both critical in order to validate results quantify errors in the simulation. For Markov Chain Monte Carlo, the blocking analysis above also provides convenient estimation of the correlation between samples.
For random variables with finite expected value and variance (i.e. the central limit theorem is applicable), the data sufficiency can be checked by comparing skewness and kurtosis (the third and fourth central moments) to that of the normal distribution. Similarity to the normal distribution is not a rigorous condition for data sufficiency, but in light of the fact that there are no good general estimations available, a good heuristic.

$$\text{skewness:} \quad \mathbb{E}\Big[(X - \mathbb{E}[X])^3\Big] \tag{139}$$

$$\approx \frac{\sum_i^N (X_i - \mathbb{E}[X])^3}{(N-1)\text{Var}^{\frac{3}{2}}[X]} \tag{140}$$

$$\text{kurtosis:} \quad \mathbb{E}\Big[(X - \mathbb{E}[X])^4\Big] \tag{141}$$

$$\approx \frac{\sum_i^N (X_i - \mathbb{E}[X])^4}{(N-1)\text{Var}^2[X]} \tag{142}$$

The same heuristic can recursively be applied to the block construction above (i.e. $X \to \bar{X}$ and $N \to N_{\text{blocks}}$), in order to check whether coarser accumulation still yields desirable statistics.

5.7 RANDOM NETWORKS

In order to generalize specific network configurations, we will use random networks with certain parameters. A fitting abstraction for the network of firms is the Erdős–Rényi model, where the nodes represent firms and the edges cross holdings Erd\H os and Rényi 1959.

Definition 13 (Graph) We call $G = (V, E)$ a graph with edges E and vertices V. V is a finite set of vertices, $E \subset V \times V$ is a set of tuples called edges for a directed graph. $E \subset \{u, v \,|\, u, v \in V, u \neq v\}$ for undirected graphs.

We call $u, v \in V$ adjacent if $u, v \in E$. The degree of a vertex in an undirected graph is the number adjacent vertices, i.e.

$$\deg(v) = |\{u \in V | u, v \in E\}| \tag{143}$$

For a directed graph we distinguish in and out degree, with

$$\deg_{\text{in}}(v) = |\{u \in V | (u, v) \in E\}| \tag{144}$$

$$\deg_{\text{out}}(v) = |\{u \in V | (v, u) \in E\}| \tag{145}$$

Two vertices u, v are called connected if there exists a number n of vertices w_i, such that $\{u, w_1\}$, $\{u, w_2\}$, $\cdots \{w_n, v\} \in E$. A (sub-) graph $G(V, E)$ is called connected if $\forall u, v \in V$ u and v are connected. A directed graph is called connected if such a path exists for either $u \to v$ or $v \to u$ and strongly connected if both hold. One can define functions on graphs. The most common one is the addition of an edge weight. Let $G(V, E)$ be an undirected graph, then die edge weights are defined as a function of the edges $w : E \to \mathbb{R}^+$. The restriction to positive weights is not necessary, but sufficiently general for our proposes. Before arriving at the full model, it is useful to state some basic properties of easier special cases. We start by considering properties of an undirected random graph.

The Erdős–Rényi model $G(n, p)$ is constructed from n vertices by inserting every possible edge with probability p. A fully connected graph has $\binom{n}{2}$ edges. Inserting each edge independently with probability p leads to a binomial distribution of the degree for each vertex in the graph:

$$P(\deg(v) = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k} \tag{146}$$

$$\longrightarrow \frac{(\langle k \rangle)^k e^{-\langle k \rangle}}{k!} \tag{147}$$

For large networks (in the limit $n \to \infty, p \to 0, np = \langle k \rangle$) with constant connectivity $\langle k \rangle$, the degree approaches the usual Poisson distribution. There are two boundary values of the connectivity, at which a qualitative transition in the graph occurs [Joho et al. 1990]:

$\langle k \rangle \sim 1$: For $\langle k \rangle < 1$ to $\langle k \rangle > 1$ a transition from a mostly disconnected to an almost fully connected graph happens.

$\langle k \rangle \sim \ln n$: There will almost surely[7] be at least one disconnected vertex until $\langle k \rangle > (1 + \epsilon) \ln n$.

This behavior as well as three typical networks in the different regions are illustrated in fig. 12
Surprisingly enough both transition thresholds are the same for directed graphs, see theorem 12.2 and 12.9 [A. Frieze and Karonski 2016].

The directed graph model with the additional restriction that $\nexists v \in V : (v, v) \in E$ can now be used to represent our network of firms. We split the graph $G(n, p)$ into cross holdings of debt and assets, i.e. $n = 2N$, with cross holding probability of p. The graph is then encoded as an adjacency matrix $M_{\text{s}} \in \{0, 1\}^{N \times N}$, $M_{\text{d}} \in \{0, 1\}^{N \times N}$ with $M_{ii} = 0$. In and out degree can then be computed as the sum over columns and rows. We model the percentage values of cross holding between firms as edge weights. This requires us to transform the adjacency matrix to a bistochastic one (i.e. the sum over all rows and columns is equal to 1). The last additional requirement is an adjustable percentage for the out-degree percentage to some value, because at most 100 percent of shares of a firm can be distributed across the network. Matrices of this type are also known as generalized left (sub-) stochastic matrices. In order to be able, to apply Sinkhorns algorithm 4, we generally fix rows and columns simultaniously (i.e. force a double (sub-) stochastic matrix).

Our simple model is assumed to be a good representation of networks of cross holdings. However, one may want to impose a more realistic model of actual dependency structures. For these cases more sophisticated graph models such as Price's model could easily be used instead.

For the initial test with the Erdős–Rényi model a grid with refinement at the interesting transition points will be used. A uniform spacing over $\langle k \rangle \in [a, b]$ with N points is extended by additional N points in a region of length $\frac{(b-a)}{10}$ around $\langle k \rangle \approx 1$ and $\langle k \rangle \approx \frac{\ln(n)}{n}$.

7 In the limit of $n \to \infty$, the probability of a randomly sampled graph having this property approaches 1

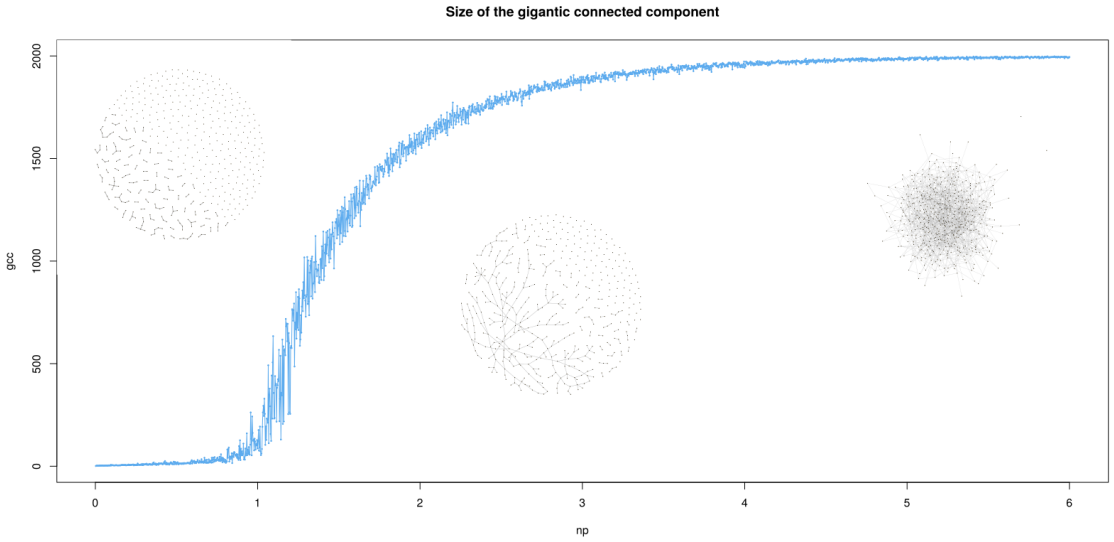


Figure 12: Transition to giant connected component at connectivity ~ 1. The figure shows the size of the largest conencted sub-graph as a function of the connectivity $\langle k \rangle = np$

5.8 CONFIGURATION MODEL

As an additional model we discuss the configuration model, which receives an input containing the degree of every node and produces multi graphs (i.e. a graph containing multi-edges and self loops). The input is usually encoded using a list of the node labels v_i, each repeated k_i times, with k_i being the degree of v_i. For our directed graph with n nodes of degree k, we generate two lists l, r containing each of the n nodes k times. We can then generate a random permutation of one of the lists (for example l) and view each pair (l_i, r_i) as a non zero entry in the adjacency matrix at A_{ij}. In order to obtain simple graphs from the configuration model, one can either reject a proposed

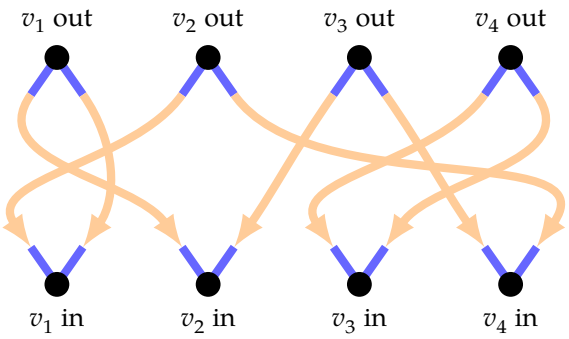


Figure 13: Configuration Model Stub Construction

network or delete all loops and multi edges. Since both methods did not result in an observable change of the degree distribution, the latter was used for improved performance. In the limit of $N \to \infty$, loops and multi edges become increasingly irrelevant, confirming the validity of this step Angel, Hofstad, and Holmgren 2016. In fig. 13 the initial generation of the model is shown. The resulting graph (including one loop, multi edges are omitted) after scaling is shown in fig. 14. Details about the scaling will be discussed in the next section.

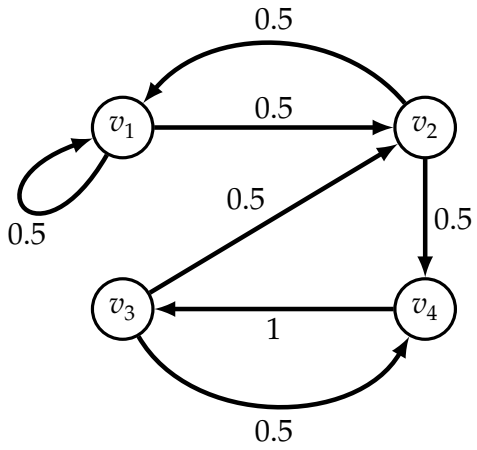


Figure 14: Resulting Graph after applying the Sinkhorn algorithm to the input from fig. 13

5.9 RANDOM SCALED MATRICES

The network of N firms will be represented by a $N \times 2N$ matrix as discussed in chapter 4 and the previous section. While the ER-model is well understood and the adjacency matrix can be generated quite easily, the transformation to a generalized doubly stochastic (g.d.s.) matrix leads to unsuspected difficulties.
The Birkhoff-von Neumann theorem states, that any $n \times n$ bistochastic matrix A_n can be decomposed into a linear combination of $n \times n$ permutation matrices [Dufossé et al. 2018]. The set of all A_n, called Birkhoff polytope B_n, is therefore the convex hull of all permutation matrices.

$$A_n = \sum_{i=1}^{m} c_i P_i \tag{148}$$

However, the proof is not constructive and the research regarding the structure of the Birkhoff polytope is still ongoing [Dufossé et al. 2018]. This also means, that there is no algorithm known which allows sampling of matrices from a specified distribution. Permutation matrices also form a basis of g.d.s. matrices [Lai 1986]. The lack of knowledge about the precise structure of the Birkhoff polytope forces us to carefully observe the distribution of the sampled matrices. Initial tests using rejection sampling[8] revealed a skewed distribution and high rejection rates at some connectivities.
We start with a random ER type graph with edge weights all equal to one. There exist so called matrix scaling algorithms, that are able to transform the adjacency matrix of such a graph (containing only zero and ones) to an adjacency matrix containing the edge weights. We also require this matrix to have fixed row sums and zero diagonal elements. These algorithms are primarily used in image processing, specifically the problem of discrete optimal transport [Ferradans et al. 2013]. They were originally motivated by Sinkhorn's theorem [Toro et al. 2016].

Theorem 1 *Sinkhorn Given a strictly positive square matrix A, there exist two strictly positive diagonal matrices D_1, D_2 with $B = D_1 A D_2$ bistochastic. D_1 and D_2 a unique up to a scalar factor.*

This theorem gives rise to algorithm 4 for the generation of bistochastic matrices. Algorithm 4 is not fully applicable to our case for two reasons: We require arbitrary row and column sums and our matrix contains zeros. The first constraint can be easily met by multiplying the resulting matrix by our scalar sum value. The second condition does not hold, since at least the diagonal is zero (no self holdings). However, it is known, that the matrix can still be rescaled to be (sub) stochastic, if and only if A is fully irreducible. A matrix A is said to be (fully) indecomposable

8 ER networks are generated, rescaled to the row sum α and rejected if any column sum is not a probability

Pseudocode 4 Sinkhorn's Aglorithm

procedure GENERATE B(strictly positive matrix A)
  $B \leftarrow A$
  repeat
    $c_i \leftarrow \sum_i B_{ij}$
    $B'_{ij} \leftarrow B_{ij}/c_i$
    $r_i \leftarrow \sum_j B'_{ij}$
    $B'_{ij} \leftarrow B'_{ij}/r_i$
  until $B \approx B'$
return B
end procedure

if there exist no matrices P and Q (for not fully indecomposable, also called irreducible, we set $Q = P^T$) such that

$$PAQ = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} \tag{149}$$

Irreducibility is equivalent to the condition, that the digraph constructed from A as an adjacency matrix, is fully connected [Fenner and Loizou 1971]. Fully indecomposable matrices can also be associated with connected Graphs. The set of edges $K^\sigma(A)$ for a Graph with adjacency matrix $A_{[n\times n]}$ and with permutation $\sigma \in S_n$, is constructed from nodes v_i, $1 \leq i \leq n$, and edges $E = \{(x_i, x_{\sigma(i)})|a_{ij} > 0\}$. A is fully indecomposable if and only if all $n!$ possible $K^\sigma(A)$ are strongly connected. This can also be expressed more informally as follows: The total support property of a matrix is necessary and sufficient for the existence of the Sinkhorn fixed point (i.e. convergence of the algorithm) with the same zero pattern as the starting point [Sinkhorn and Knopp 1967]. A matrix is said to have a total support if all non-zero elements can be brought onto a strictly positive (no zero elements) main diagonal by permutation of columns. For example A_1 below, cannot have a strictly positive main diagonal. A_2 does also not have a total support, because the a_{11} entry (marked in blue) is either on a non strictly positive main diagonal or not on the main diagonal at all, under both possible column permutations. One can also distinguish between support and total support. A matrix with support has a strictly positive diagonal under column permutations. This is the case for A_2 but not for A_1.
The second condition above, is of course significantly harder to check than the first one, for $\sigma(i) \equiv i$. It is therefore more useful to test algorithm 4 for one of two possible failure states: the solution can start to oscillate, or the support can change, i.e. one or more elements converge to a weight of zero.
Toro et al. provided two examples for both cases in his original paper [2016]:

$$A_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \qquad A_2 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

Note that we only need to consider matrices with all zero rows and columns removed, since they do not contribute or interfere with the scaling algorithm. In the example above, A_1 results in oscillations with no convergence to a bistochastic matrix. A_2 does converge, but the diagonal matrices do not exist. In other words: the in/out degree of some nodes change by applying the algorithm. For the example A_2, the a_{11} entry approaches zero.
While the second case can be tolerated as a model limitation (the cross asset or debt holding is of negligible amount), the first one poses a substantial problem and the initial guess has to be rejected. However, a problematic initial guess is easily detectable. One only needs to check whether a non zero element has a value smaller than row sum divided by $N - 1$.
Another necessary condition for (r, c) scalability is that $\|r\|_1 = \|c\|_1$ [Rothblum and Schneider 1989; Allen-Zhu et al. 2017]. We therefore only allow for this kind of constraint. During tests the limit of 1000 rejection steps proofed sufficient for all parameters with floating point equivalency precision

set to 10^{-8}.
In fact, after testing the results, there appeared no visible differences in the resulting Greeks by omitting the re-normalizing of row sums. This results in the algorithm terminating after the first step. However, the results have been periodically compared with the Sinkhorn algorithm to confirm consistency with results from controlled row and column sums.

5.9.1 *Sampling Procedure*

Each valuation sample for a given adjacency matrix is generated by drawing $Z \sim \mathcal{N}(0, \mathbb{1}_{N\times N})$ and subsequently generating the assets according to 119. If the importance sampler is active, the weight is also computed at this point according to sec. 5.2.3. The asset value, as well as Z (required for the computation of the Greeks) is handed off to the fixpoint iterator.
Additionally a random network is generated as well and used for a given number of Monte Carlo samples, until a new network is generated[9]. Since neither the ER-model nor the scaling algorithms guarantee fixed connectivity, but our results directly depend on it, the results are binned according to their actual average degree. The coarseness of the binning is another implicit parameter, that is adjusted according to the number of samples and size of the connectivity interval. This histogram is then returned as result for the run and has to be treated accordingly when simulating multiple parameter combinations.

9 This is done for performance reasons, since generating a new network requires considerable resources compared to a single valuation sample. Linearity of the expected value guarantees convergence towards the same value as long as enough networks are generated.

6 SETUP AND ANALYTIC RESULTS

Some of the results in this chapter are taken from [Bertschinger and Stobbe 2018].
All figures have been computed using scripts provided in the `scripts` sub directory, available here [Stobbe 2018b], partly also using a C++ back-end library (included in the `src` directory). A cleaner re-implementation written in Julia is also available [Bertschinger and Stobbe 2019].
Our goal is to first obtain the valuations of all firms by fixed point iteration of the valuation function. The simplest method is direct iteration i.e. starting from a random starting vector $x^{(0)} = (s, r)$, inserting x_0 into $x^{(1)} = g(a, x^{(0)})$ (see eq. 71) and so on. Since g is known to have a unique fixed point, this method will converge [Fischer 2014; Suzuki 2002]. Even though there are more sophisticated methods available, such as Anderson acceleration, the simpler naive iteration is sufficient, as the bulk of the computational power is spent on other parts (e.g. matrix inversion for computation if the Greeks) of the simulation.
Having obtained the valuation, we can then find the solvent firms and in turn the derivatives of g according to eq. 75 and eq. 76, using $v = s + r$ and the condition for solvency of firm i: $v_i > d_i$. This gives access to the Greeks, specifically Δ, as a measure of risk.
It is straight forward to see that the network amplifies the value of firms beyond the sum of their exogenous assets. Note that $x, s, r, a, d \in \mathbb{R}^N$ for N firms, whenever these variables appear without index.

$$v_i = \max\{0, a_i + (Mx)_i - d_i\} \min\{d_i, a_i + (Mx)_i\}$$
$$a_i + (Mx)_i \geq a_i$$

This hold, since only strictly positive matrices M are meaningful in our context (i.e. no "negative" cross holdings) and x is strictly positive by definition. The complement of the internally held assets for firm i is given by

$$m_i^{\text{out}} = 1 - \sum_j M_{ji} \tag{150}$$
$$x_i^{\text{out}} = m_i^{\text{out}} x_i \tag{151}$$

The total value held by outside investors must be equal to the sum of the exogenous assets a_i, since the cross holdings only redistribute value across the network, but can not change the total amount. In other words: $m^{\text{out}} \cdot x = \sum_i a_i$. This can also be confirmed by direct evaluation of the dot product [Bertschinger and Stobbe 2018]. However, the total internal value of the firms increases, as shown in eq.150. This means, that the network itself gives rise to an increase in value over the sum of the external investments (since $\sum_i v_i \geq \sum_i v_i^{\text{out}}$).

6.1 EFFECTIVE ONE DIMENSIONAL CASE

Before considering the general case with arbitrary network types, some insights can be gained by considering the completely homogeneous case:

$$M_{ij} = \alpha \left(1 - \delta_{ij}\right) \tag{152}$$
$$r_i = \min d_i, a_i + M^d r \tag{153}$$

Due to the symmetry of M, we can ignore the index i and solve the effective one firm case.

$$s = \max\{0, a + \alpha x - d\} \tag{154}$$
$$r = \min\{d, a + \alpha x\} \tag{155}$$

This effective single firm model can in fact be solved analytically and yields the fixed points:

$$\zeta = \begin{cases} 1 & \text{if } (1 - w^d)d < a \\ 0 & \text{otherwise} \end{cases} \tag{156}$$

$$s^\star = \begin{cases} \frac{a-(1-w^d)d}{1-w^s} & \text{if } \zeta = 1 \\ 0 & \text{otherwise} \end{cases} \tag{157}$$

$$r^\star = \begin{cases} d & \text{if } \zeta = 1 \\ \frac{a}{1-w^d} & \text{otherwise} \end{cases} \tag{158}$$

$$\Rightarrow v^\star = \begin{cases} \frac{a+(w^s-w^d)d}{1-w^s} & \text{if } \zeta = 1 \\ \frac{a}{1-w^d} & \text{otherwise} \end{cases} \tag{159}$$

This result can be obtained by separately considering all four cases of $s > 0$, $s = 0$ and $\zeta = 1$, $\zeta = 0$ for both cases of s [Bertschinger and Stobbe 2018]. A visualization for fixed asset and debt as a function of percentage equity and debt cross-holdings is show in fig. 15. It can be seen from eq. 156, that increased debt cross-holding shifts the default boundary to lower values of the external assets. We can also observe this shift in eq. 157 to eq. 159, where w^d results in higher equity and higher recovery value of debt. However, it reduces the value of the firms in the solvent case. In this

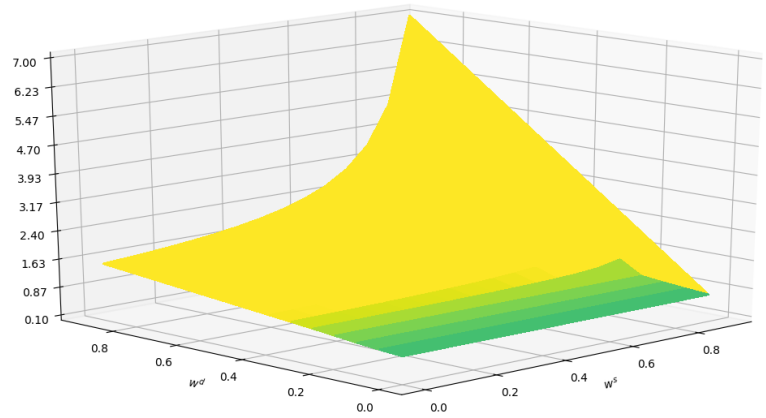


Figure 15: Value of firm in the homogeneous case as function of w^s and w^d. assets are set to 0.7, debt to 0.9.

simple model, the equity cross-holdings only have an effect on the value of the firm when it is solvent, without any effect on the default boundary.
Note that we can assume $d \equiv 1$ w.l.o.g., which will be done in the following. We can also relax the constraint of a fixed asset value. This will turn out to introduce an impact of the equity cross-holding term on the default threshold. a is assumed to follow a geometric Brownian motion, such

that $a \to a_t$ and $s \to s_t$, etc. One can still obtain closed form solutions in that case (see appendix C.1 [Bertschinger and Stobbe 2018]):

$$s_t = \frac{a_t \Phi(d_+) - (1 - w^d) d e^{-r-(T-t)} \Phi(d_-)}{1 - w^s} \tag{160}$$

$$= \frac{1}{1 - w^s} C_T \tag{161}$$

$$r_t = \frac{a_t \Phi(-d_+) - (1 - w^d) d e^{-r-(T-t)} \Phi(d_-)}{1 - w^d} \tag{162}$$

$$= \frac{1}{1 - w^d} (e^{-r(T-t)} (1 - w^d) d - P_T) \tag{163}$$

with the usual

$$d_\pm = \frac{\ln \frac{a_t}{d(1-w^d)} + (r \pm \frac{\sigma^2}{2})(T - t)}{\sqrt{T - t}\sigma} \tag{164}$$

And C_T, P_T being the usual Black-Scholes call and put payoffs at maturity T, with strike price equal to $(1 - w^d)d$ and spot price a_t. Here we can observe the amplification of r and s by cross-holdings

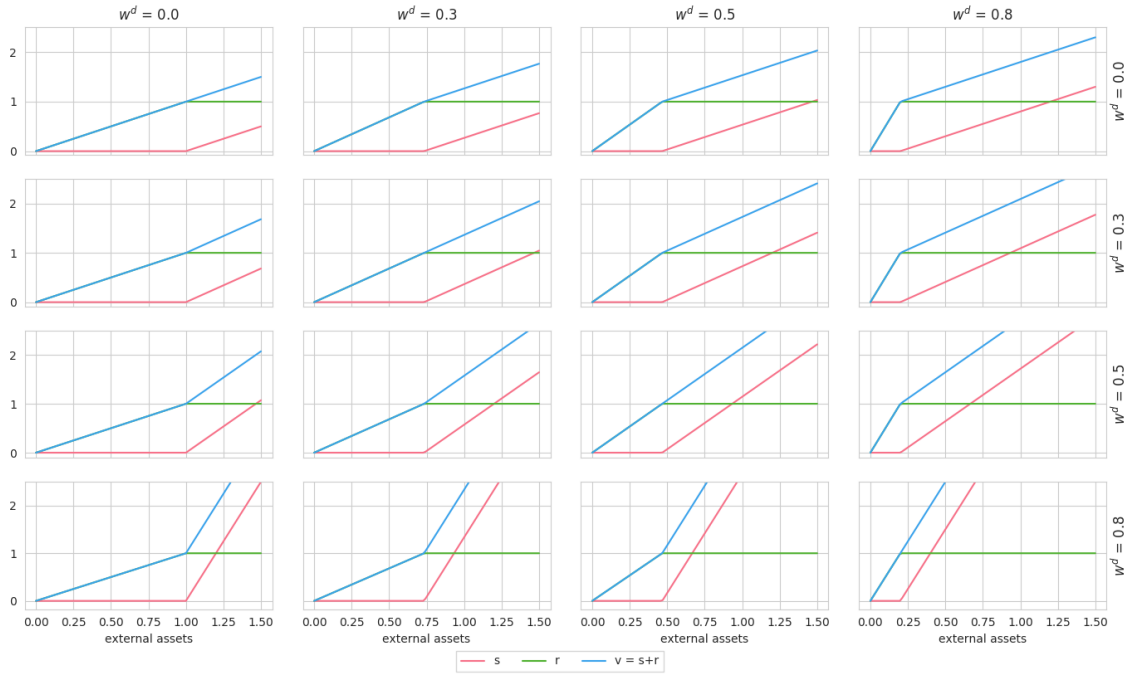
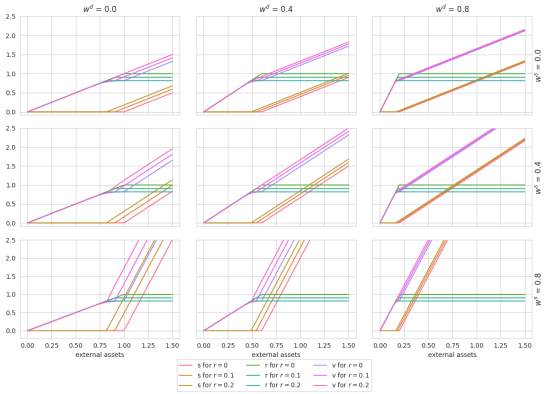
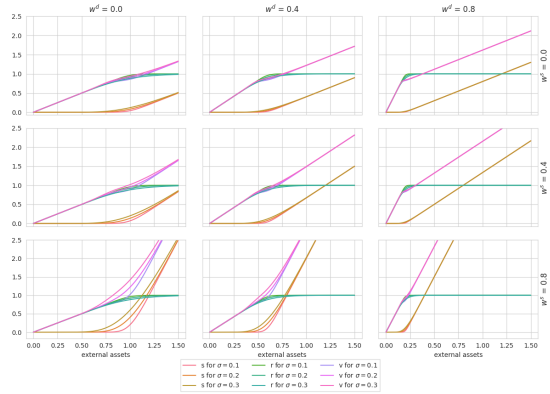


Figure 16: Equity and debt value over external asset price for different values of debt and equity cross holdings. $\sigma = 0.1$, $r = 0$, $\tau = 1$

of equity and debt. Furthermore, the default boundary is also lowered by debt cross-holdings (i.e. $d \to (1 - w^d)d$). The corresponding values for r , s and v are shown in fig. 16 for $r = 0$ and $\sigma = 0.1$. in fig. 17a and fig. 17b we can also observe, that a change in the volatility σ smoothes out the curves, while increase of the interest rate introduces a shift. Note, that an increase of the debt cross-holdings reduces the difference between equity and debt of varying r, while an increase in equity cross-holdings increases the shift. This has already been observed in the derivation of the simple effective model, but also holds true here.
Since we have reduced the model to the Black-Scholes model with modified asset and strike price, it is now also possible to use the standard derivations for the Greeks and write down their closed



(a) Equity and debt value over external asset price for different values of debt and equity cross holdings for varying r. $\sigma = 0.01$, $\tau = 1$



(b) Equity and debt value over external asset price for different values of debt and equity cross holdings for varying σ. $r = 0$, $\tau = 1$

form representations using eq. 163 and eq. 161. For the equity part we get, using the chain rule and the derivative of the cumulative distribution function (i.e. the density function) $\Phi'(x) = \varphi(x)$:

$$\Delta_s = \frac{\partial s_t}{\partial a_t} = \frac{1}{1 - w^s}\Phi(d_+) \tag{165}$$

$$\nu_s = \frac{\partial s_t}{\partial \sigma} = \frac{a_t\sqrt{\tau}}{1 - w^s}\varphi(d_+) \tag{166}$$

$$-\Theta_s = \frac{\partial s_t}{\partial \tau} = -\frac{a_t\sigma}{2(1 - w^s)\sqrt{\tau}}\varphi(d_+) - \frac{rd(1 - w^d)}{1 - w^s}\Phi(d_-)e^{-r\tau} \tag{167}$$

$$\rho_s = \frac{\partial s_t}{\partial r} = \frac{\tau d(1 - w^d)}{1 - w^s}e^{-r\tau} \tag{168}$$

For the debt part, we obtain similar results:

$$\Delta_r = \frac{\partial s_t}{\partial a_t} = \frac{1}{1 - w^d}\Phi(-d_-) \tag{169}$$

$$\nu_r = \frac{\partial s_t}{\partial \sigma} = -\frac{a_t\sqrt{\tau}}{1 - w^d}\varphi(d_+) \tag{170}$$

$$-\Theta_r = \frac{\partial s_t}{\partial \tau} = rde^{-r\tau}\Phi(d_-) + \frac{a_t\sigma}{2\sqrt{\tau}(1 - w^d)}\varphi(d_+) \tag{171}$$

$$\rho_r = \frac{\partial s_t}{\partial r} = d\tau\Phi(d_-)e^{-t\tau} \tag{172}$$

$$\tag{173}$$

In figure 18 and 19 we can see the behavior of the Greeks for differing values of cross-holdings in the effective one firm model. We observe that an increase in the interest rate results in a shift of the transition to higher spot prices. An increase in the volatility results in broader features and smoother transitions. Having seen that both parameters do not change the qualitative behavior of the Greeks, we will keep them fixed to $r = 0$ and $\sigma = 0.3$ throughout most of the remaining thesis.

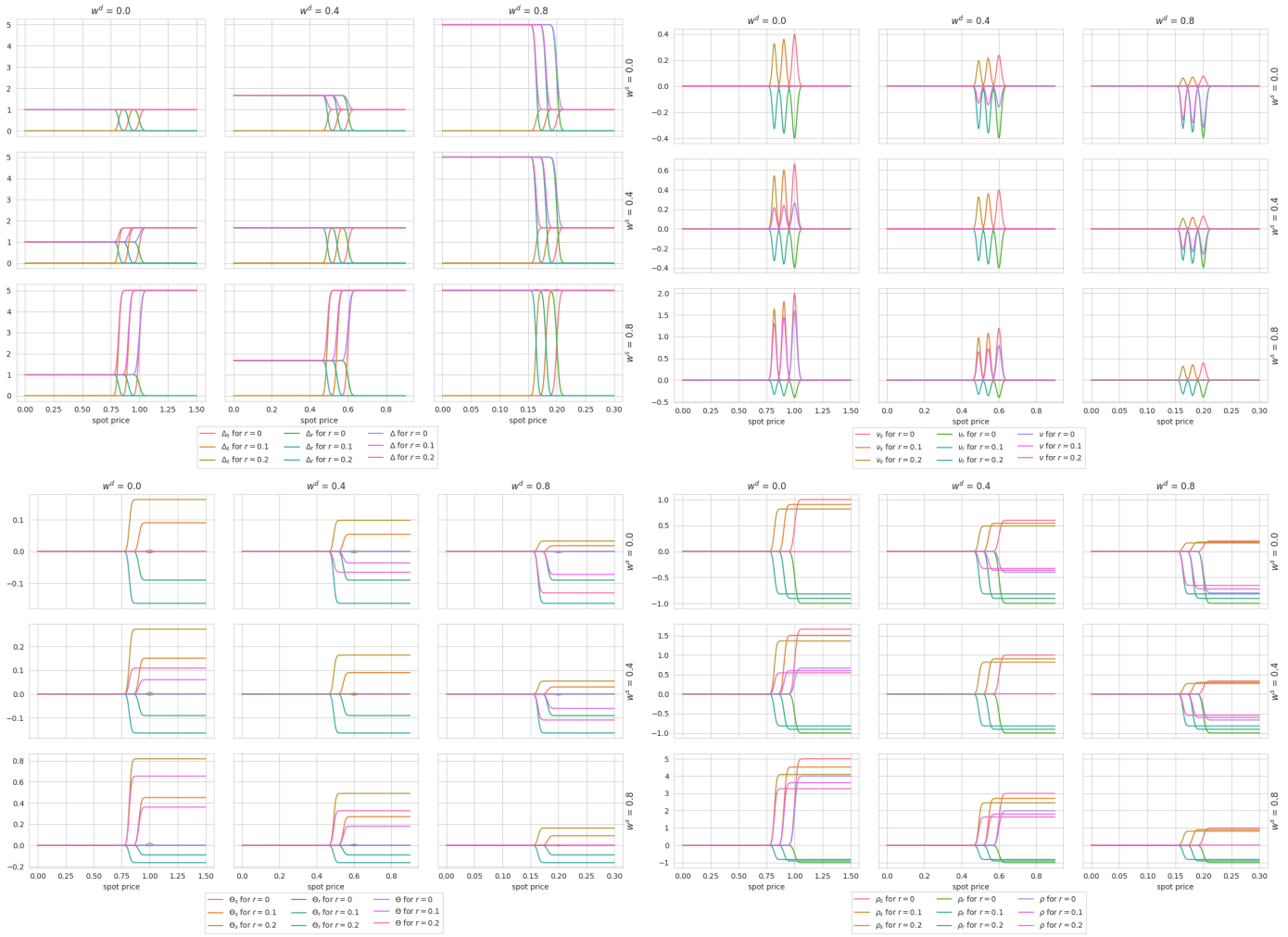


Figure 18: Greeks for fixed $\sigma = 0.02$, $\tau = 1$, $d = 1$ and varying interest rate.

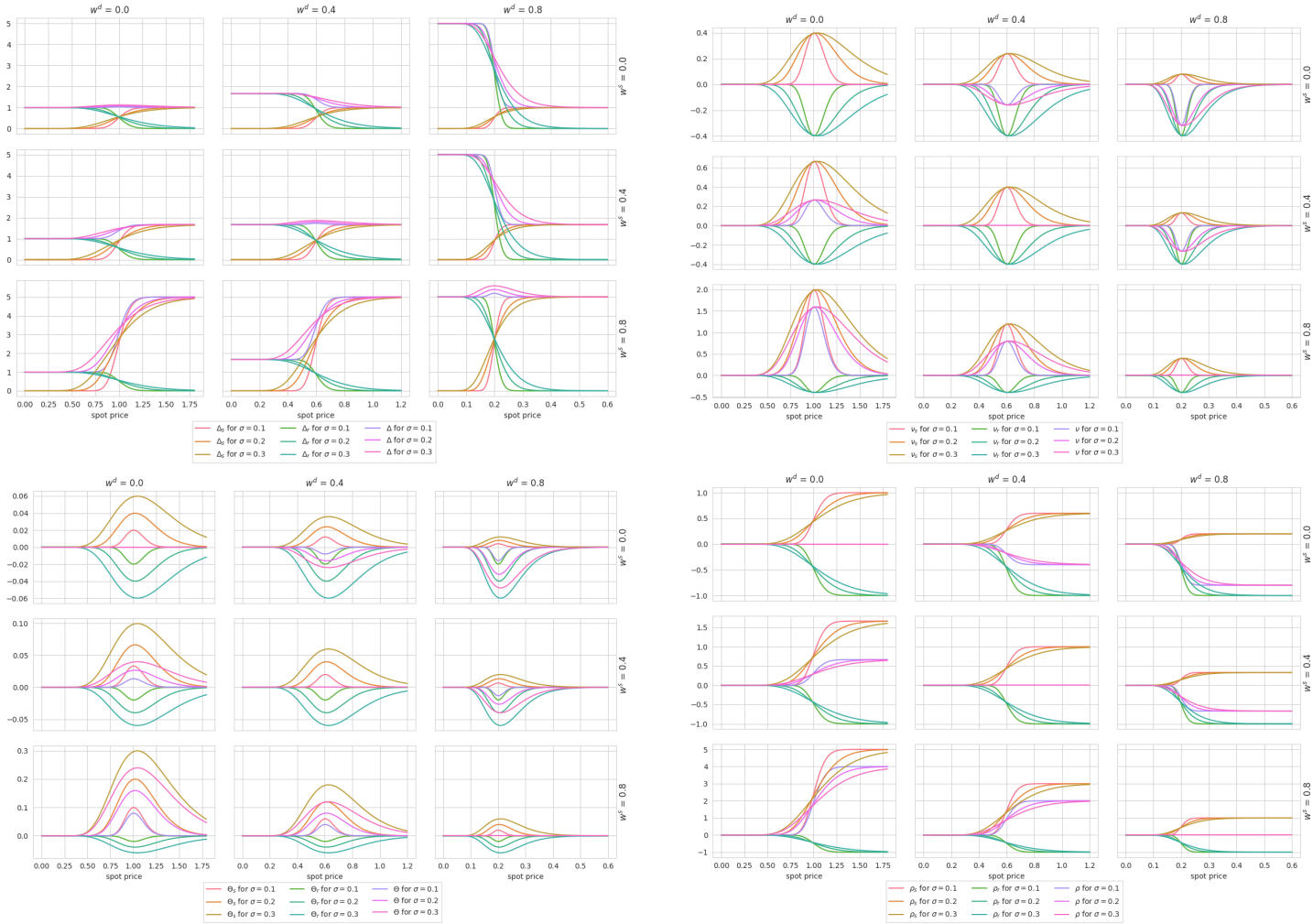


Figure 19: Greeks for fixed $r = 0$, $\tau = 1$, $d = 1$ and varying volatility.

6.2 TWO DIMENSIONAL CASE AND FIXED POINT BEHAVIOR

Before introducing random networks, we consider the case of fixed two-dimensional networks. This will not only serve as a verifiable test case (since the solution for the solvency regions is known [Suzuki 2002]) but also give some insight into skewed networks. We will assume log-normally distributed spot prices S_0. These are mapped onto different values S_T at maturity T. The transformation behavior of the fixed point is directly dependent on the form of the adjacency matrix for equity and debt. Figures 20 shows the distortion of the spot price and solvency before and after the transformation. One can observe a heavily skewed distribution that cannot be described by the — otherwise usual for asset prices — log-normal distribution. The default bound-

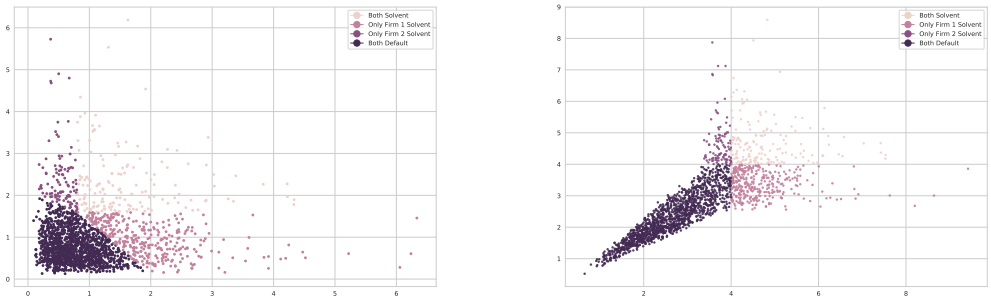


Figure 20: Distortion of the Log-Normally distributed exogenous assets by application of the fixed point iteration. Parameters: $\sigma^2 = 0.4$, $M^s \equiv 0$, $M^d_{01} = 0.8$, $M^d_{10} = 0.6$, $d = 4$, $r = 0$, $T = 1$. Reproduced from [Fischer 2014]

aries (borders between different regions of solvency) and distortions under cross holdings have been obtained analytically for two firms by Suzuki [Suzuki 2002]. For comparison, the mapping from a uniform grid of spot prices is also shown in fig. 21, using only cross-debt holdings (this leaves the solvent region unchanged). Since we will later use equal spot prices for all firms, the diagonal line is drawn in as well. We observe that the correlation between defaults is heavily in-

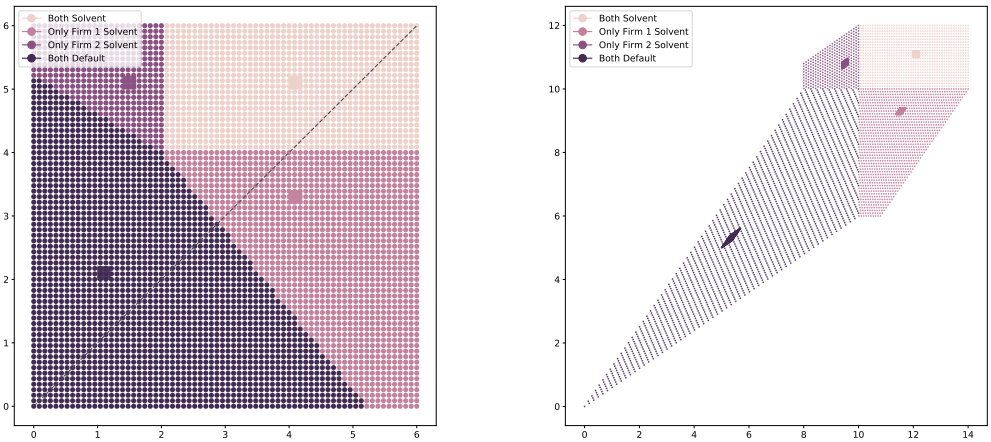
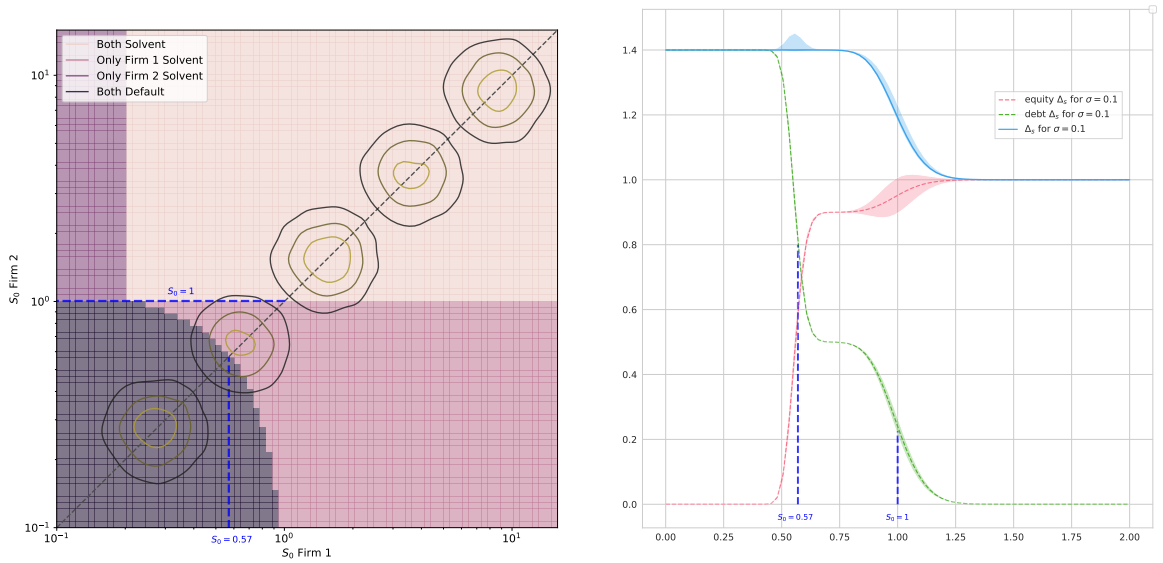


Figure 21: Distortion of squares according to the Suzuki regions. Parameters: $\sigma^2 = 0.8$, $M^s \equiv 0$, $M^d_{01} = 0.8$, $M^d_{10} = 0.6$, $d = 10$, $r = 0$, $T = 1$

creased by cross-debt holdings. The same can be seen for correlations of solvency and cross-equity

holdings. Since the model exhibits all relevant features for $M^s \equiv 0$, we will mostly use this case for performance reasons. In the next chapter, the Erdős Rényi model will be used for the generation of networks of size larger than two. This will result in homogeneous observables, as the Monte Carlo average over all firms is taken. It is therefore valuable to consider how asymmetries affect the observables (i.e. Greeks), before considering the N-dimensional case. To this end we again generate log-normally distributed assets and run the previously described Monte Carlo algorithm for differing adjacency matrices of debt and equity, focusing on cases with large differences between the values of cross holdings.
In fig. 27, all Greeks are shown for $M^s \equiv 0$ and varying off-diagonals and volatilities. The obvious difference compared to the effective one dimensional model before is the bimodal structure in equity and debt part of the Greeks. Note, that this is not necessarily reflected in the total value, especially for larnge volatility. The reason for this behavior becomes obvious when looking at the Suzuki areas again. For $M^d_{01} \gg M^d_{10}$, the point at which all four areas meet, see fig. 22a, is shifted



(a) Solvency for both firms as function of the spot price. Model only generates samples from die diagonal, 5 of the S_t distributions are shown.

(b) Δ as function of the spot price. Transitions between solvency regions are marked.

Figure 22: Transition of Δ from double solvent over only one firm solvent to both default, visible as step in equity and debt (but not value) part of Δ.

away from the diagonal. This means, that the intermediate region of only one firm being solvent becomes increasingly visible, see fig. 22b We define:

$$\Delta_s = \frac{1}{n} \sum_{j}^{N} \sum_{i=1}^{N} G_{i,j} \tag{174}$$

$$\Delta_r = \frac{1}{n} \sum_{j}^{N} \sum_{i=N+1}^{2N} G_{i,j} \tag{175}$$

$$\Delta = \Delta_r + \Delta_s \tag{176}$$

The distribution of S_T is shown as well in fig. 22a for the same parameters as used for the computation of (with the sum either over the equity, debt or both parts) in fig. 22b and fig. 27 in order to show the transition regions and the effect of volatility. Note, that we distinguish between the parts of Δ belonging to equity and debt part of the fixed point. The sum of both will be called value Δ.
Having established the overall behavior of the Greeks as function of the spot price for different connectivities, we can now focus on more general networks. In this work I will solely focus on

simple network models and computation of first order Greeks. More general network structures such as scale-free networks, different risk measures, etc are left for future research.

7 ERDŐS RÉNYI RESULTS

7.1 GENERAL CONSIDERATIONS

We will use the previously discussed Erdős–Rényi (ER) model to create networks of cross-holdings. The ER model results will approach the simple effective one firm model in the infinite firm limit, since firms are indistinguishable in this model. However, the two transition connectivities lead to qualitative changes of the network structure. Additionally, the change in the number of cross-holdings is expected to result in a different redistribution of exogenous assets. It will therefore be more interesting to consider the Greeks as functions of the connectivity rather than of the spot price. Using the convergence property, the results for the inhomogeneous case can be used to indicate insufficient sample size: For example, we do not expect vega to be bimodal in the ER network for a large number of firms. The simulations will require a couple of parameters which are summarized here again: As before, we will assume Log-normal distribution for S_0.

$M^{d/s}$	full network Structure as adjacency matrix with weights
N	size of Network i.e. number of firms
p	probability of cross holding between any two firms
$w^s = \sum_i M^s_{ij}$	equity cross holdings
$w^d = \sum_i M^d_{ij}$	debt cross holdings
S_0 or a_0	spot price
$N_{\text{net samples}}$	number of different network configurations
N_{samples}	Monte Carlo samples per network configuration
$\langle k \rangle = N \cdot p$	connectivity of the ER model graph. Fixed in/out degree for configuration model.
σ_i	Volatility of firm i
β	default probability scale
T	maturity
r	risk free interest rate
L	$L^T L = \Sigma$ Cholesky decomposed volatility covariance of the underlying Wiener process

From the previous chapter we know that the columns of the cross holding matrix correspond to the proportion of (equity or debt) value that is distributed over the network. It is therefore nescessary to fix the column sums to a probability (i.e. $0 \leq w^d \leq 1$). While not strictly needed, it can also be desirable to fix the value of the row sums. According to Sinkhorn und subsequent results (theorem 1) the row and column sums cannot be rescaled independently in general. If a rescaling of rows is intended the value will therefore also be set to w^d.
Drawing uniformly from the set of random scalable matrices (without the relaxation of $w^d \leq 1$ and the constraint $M^d_{ii} = 0$) is debate of ongoing research [Dyer, A. Frieze, and Kannan 1991; Diaconis, Lebeau, and Michel 2012]. Different algorithms that sample from the set of possible M with different constraints were implemented for this thesis. Each of them generates an adjacency matrix without weights which is then rescaled as needed. For example, the ER model sets each entry M^d_{ij} is set to 1 with probability p_d. For the rescaling of the adjacency matrix, we have the following options:

REJECTION SAMPLING After the columns are normalized to w^d, the matrix is rejected if any row sum is greater than the given constraint. There are some behaviours that one may want to reject. One is the suppression of one element (see sec 5.9) of one connection, because one may want to fix the in and out degree. This suppression also introduces the possibility of asymmetric edge weights. One can also

fix the sum over the second index of M.
Specific to the configuration model, one can also disallow self loops.
Not controlling for these situations allows for a higher sampling rate but possibly leads to biased results. The observables have therefore been carefully checked with small batch sizes before disregarding some of the constraints for larger simulations.

SIMPLIFIED NORMALIZATION The columns are normalized to w^d without regard for the distribution of the row sums.

SINKHORN-KNOPP (SK) ALGORITHM Based on the Sinkhorn-Knopp theorem, this provides an iterative algorithm with some known results about convergence behavior and row sum distribution.

Since few analytic results are available for our sampling algorithms, one has to check carefully for the actual distribution of random matrices generated by the algorithms. The most important requirement is, that no systematic bias is introduced by sampling certain types of networks more often than others. For example, non-trivial correlations between matrix elements, i.e. non-uniform sampling over the set of all possible random matrices under our constraints, could be one such biased sample set.
As seen in fig. 22b it is often useful to fix the default probability at about $\frac{1}{2}$ in order to see the general behavior of the Greeks for different spot prices. We therefore fix the face value of debt to:

$$\langle d \rangle = \frac{\beta}{1-\alpha}\left(\langle S_0 \rangle - \frac{\sigma^2}{2}T\right) \tag{177}$$

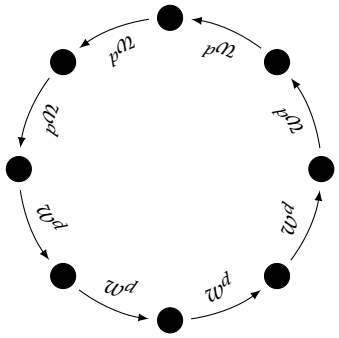
7.2 PROGRAM VERIFICATION RESULTS

While the properties of the ER model are valid for graphs of all sizes, it is not clear from the outset how different networks sizes effect the observables. Furthermore, the convergence behavior of the Monte Carlo algorithm depends on the parameters of the model, specifically $\langle k \rangle$ and w^d (or w^s) may cause problematic behaviour in different regions. This can be seen in fig. 29 and 29, where the smallest and largest networks are the most unstable ones at equal sample parameters. While the $N = 5$ network exhibits obvious finit size effects, the larger networks tend to have a higher variance (especially for increasing connectivity) due to the larger sample space. It is therefore desirable to find a network size with sufficiently many firms to observe the transition effect, but few enough to make sampling feasible. Both effects are more visible for ν and Θ, due to the smaller absolute value. However, since we are mostly interested in Δ as a risk measure, network sizes of $N = 40$ to $N = 80$ were chosen in the following, demanding sample sizes of 10^5 to 10^7.
Note, that the results for connectivities are only approximated. Due to the rescaling algorithm, it is disadvantageous to only take exact connectivities into account. Instead, networks are generated and then binned according to their connectivity. This results in some contribution of $\langle k \rangle > 0$ for $N \gg 1$, as can be observed in fig. 29. The overall effect of this approximation is assumed to be negligible at larger connectivities, but the transition at $\langle k \rangle = 1$ (see discussion of ER model) will be slightly blurred by this method.

7.3 SPECIAL GEOMETRIES

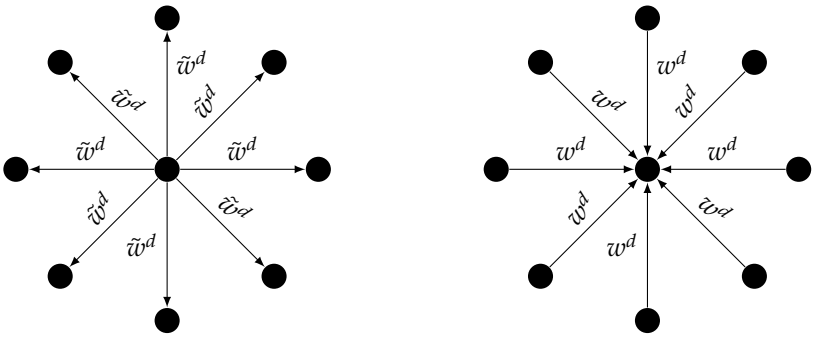
There are a couple of special network structures that have been investigated as an example for possible future research, to shorten the notation, we use $\tilde{w}^d = \frac{w^d}{N}$:

RING There is a circular dependency between all N firms. In our example (fig. 30 and fig. 30), each firm i has a debt cross-holding w^d of exactly one other firm $i \to i+1$, with $N \to 1$.



UNIFORM A fully connected network with equal crossholdings is called uniform in the following. This case has already been studied as the limiting case of the ER model at $p \to 1$ above. It is included here as comparison to the other network types. A

STAR One firm (called center) has a debt cross-holding with all other firms which have no direct edge between each other. The risks of both type of firms obviously differs and the Greeks have therefore been plotted for both separately. Below, the network structure of cross-holdings for debt which will be used in the following, are shown. The network is split into the view of the central and the view of the remaining firms. This is done to emphasize the partitioning of the network into two different types of firms.



The Greeks for all three cases have been plotted in fig. 30 and fig. 30. Since no difference has been observed between networks of different sizes only one is shown here.
Interestingly, the ring network remains unchanged under increase of w^d, while the risk (we use Δ as a measure) is amplified at lower spot prices for all other networks. The transition to lower risk is also shifted to lower spot prices for increasing w^d, an example of a diversification benefit.
In fig. 23 the value of the firms as function of the external assets is show, as well as the total impact of change to the external assets π:

$$\pi_i = \mathbb{E}\left[\sum_j \frac{\partial v_j}{\partial a_{i,T}}\right] - 1 \tag{178}$$

$$= \mathbb{E}_Z\left[e^{-rT}\left[B\left(\mathbf{a}, \mathbf{x}\left(\mathbf{a}\right)\right)\right]_{ik} \frac{\partial g_k\left(\mathbf{a}, \mathbf{x}\left(\mathbf{a}\right)\right)}{\partial a_l} \frac{\partial a_l\left(Z\right)}{\partial a_{0,j}}\right] - 1 \tag{179}$$

$$= \mathbb{E}_Z\left[e^{-rT}\left[B\left(\mathbf{a}, \mathbf{x}\left(\mathbf{a}\right)\right)\right]_{ik} \frac{\partial g_k}{\partial a_l} \cdot 1\right] - 1 \tag{180}$$

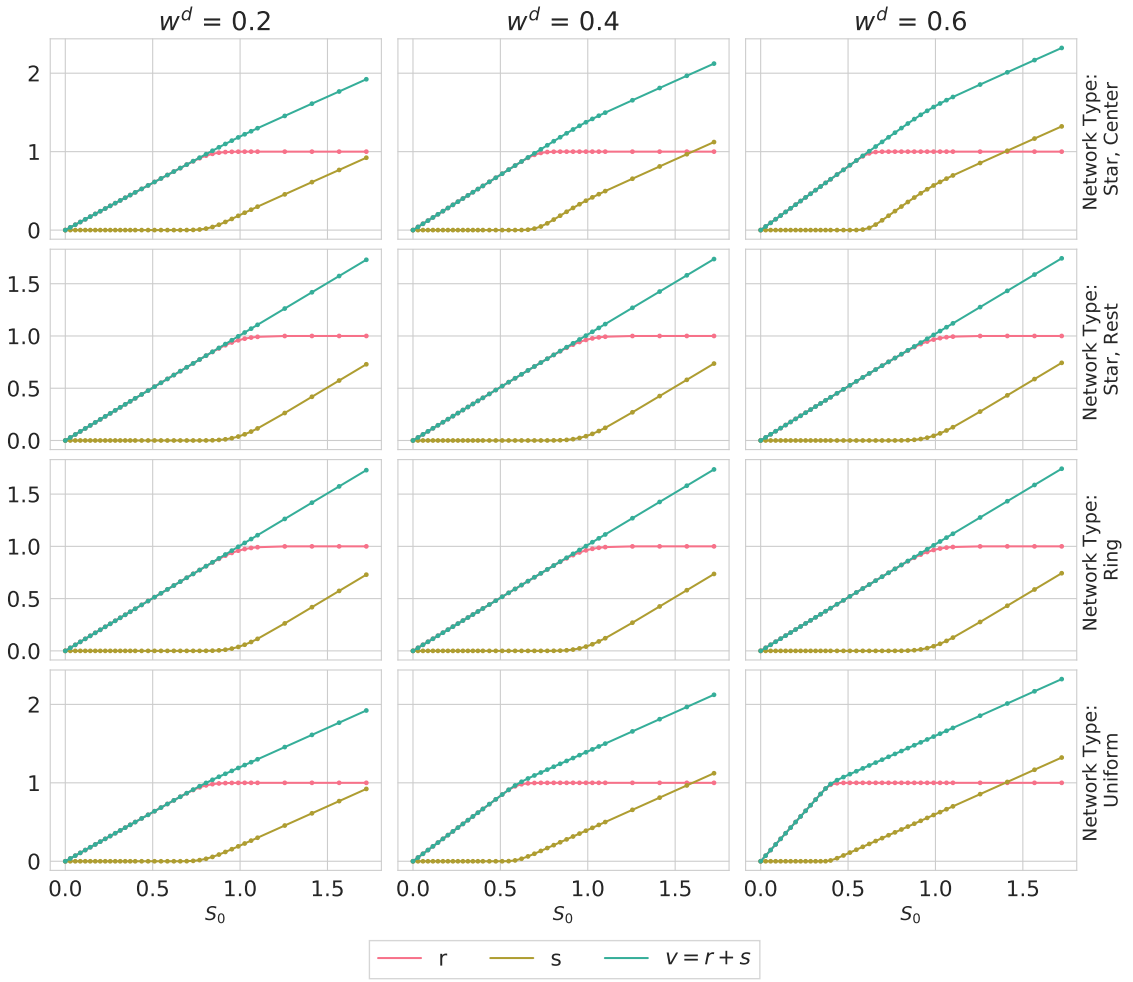


Figure 23: equity, recovery value, value and π

While π can be understood as a risk measure as well (since it measures the impact of change in assets at maturity), the Greeks — specifically Δ — is proposed as a better measure, since it does depend on the change of asset price depending on the spot price. In fig. 30 and fig. 30 we can see, that $\pi(S_0)$ has the same shape as Δ but does capture the transition less accurately for star and ring networks. Especially the center of the star network suffers from the approximation $\frac{\partial a_0(Z)}{\partial a_{0,0}} = 1$.

7.4 ER NETWORKS

The goal of this section is to quantify the effect of the phase transitions in the ER model on the Greeks (specifily Δ as the primary risk measure). Each firm i is has $\langle k_i^{\text{in}} \rangle = \langle k_i^{\text{out}} \rangle = np$ counterparties for debt cross-holdings, each with a weight of $\frac{w^d}{k_i}$. If not mentioned otherwise, the average debt value will be fixed in this section to a value at which about half of the firms are expected to default. For the following simulation we use $N_{\text{net samples}} = 5000$ and $N_{\text{samples}} = 500$. The external assets are drawn independently from a log-normal distribution as before.
Fig. 24 shows the advantage of debt cross-holdings: the solvency region is shifted to lower spot prices. Furthermore, the network reacts more uniformly (i.e. for increasing connectivity, the firms default more similarly) for large connectivities, as can be seen from the steeper slope. As expected,

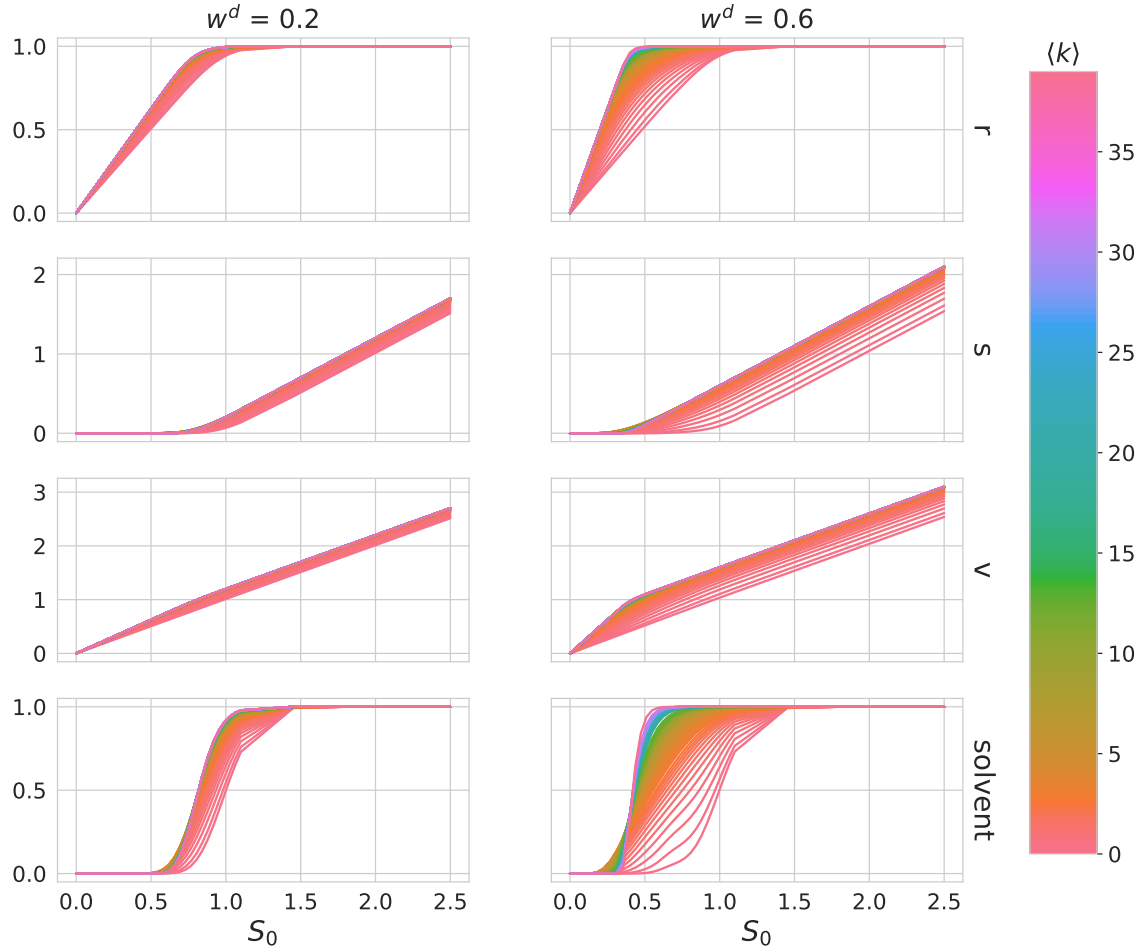


Figure 24: Greeks, value and solvency as functions of the spot price for different connectivities. Parameters: $N = 40$, $\sigma = 0.15$, $r = 0$, $T = 1$.

this is a result of an increased value, see also fig. 24. In fig. 25, the Greeks are shown as functions of the connectivity. One can easily observe a transition at $S_0 \approx 0.5$ for $w^d = 0.6$, when the firms enter the solvency region (compare fig. 24) in all Greeks. Note, that for barely solvent firms under strong debt cross-holdings (in fig. 25, the green lines at $0.5 < S_0 < 1.0$), Δ exhibits non monotonic behavior. This maximum in risks stems from the tradeoff between increasing contagion probability due to more available conenctions and increased value leading to lower default probability (see above). Before the increase in connections can lead to enough diversification benefit for the firms, the contagious effect of only a few connections increases the risk, leading to a maximum (in case of the parameters above at $\langle k \rangle \approx 2$). Only strong debt cross-holdings will drive such an effect, since the contagion of defaulting firms is not strong enough otherwise. This can be seen on the left side in fig. 25.

Fig. 26 shows the Greeks as functions of the connectivity, again confirming the risk lowering benefit of increased connectivities. The contagious effect of the network is far less visibile here, since the larger connectivities have a bigger impact on the overall change in shape of the functions. However, the exact effect of contagion even at large connectivities is not obvious, since it clearly does not vanish but is merely obfuscated by the increased firm values. An introduction of default costs into the model could give insight into this phenomenon, since one could then tune the propagation of defaults through the network. The extension of the model is left for future research.

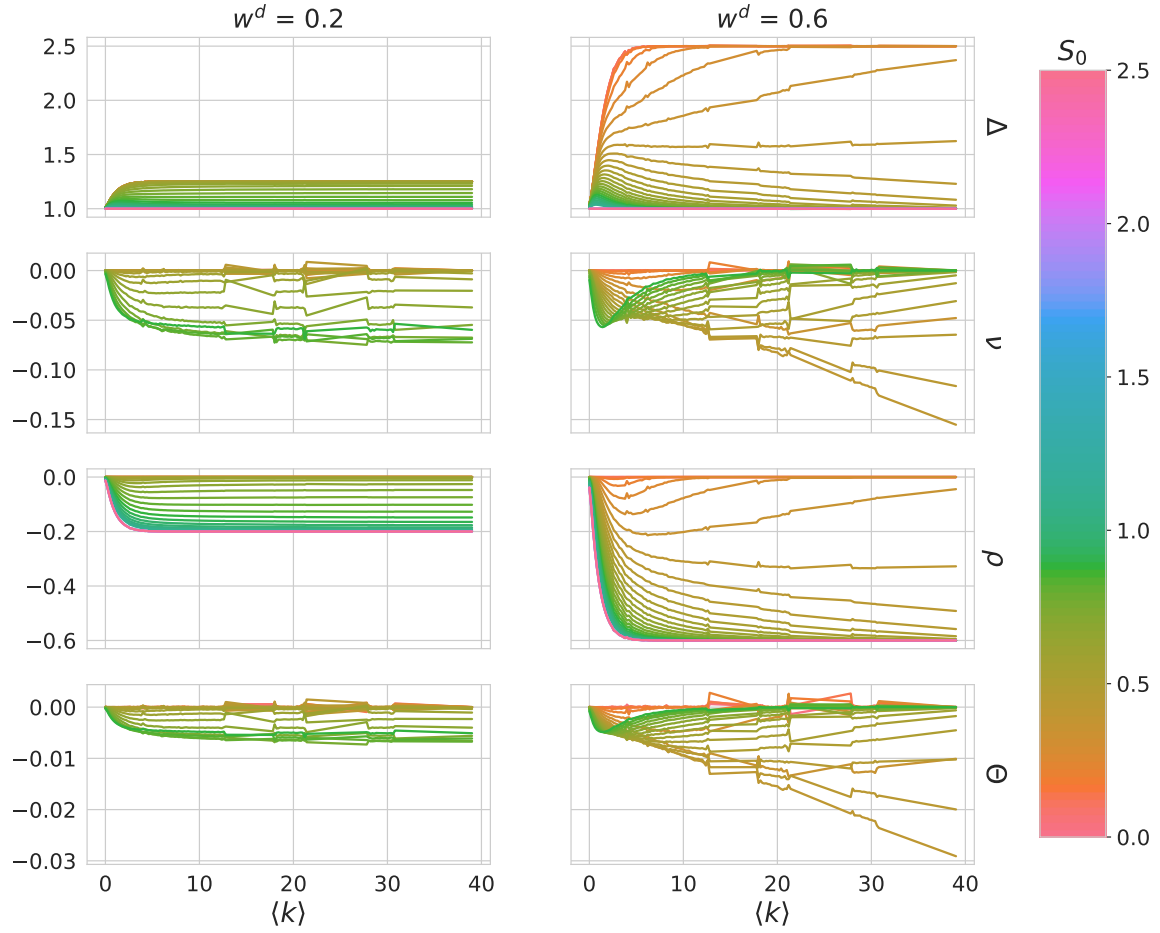


Figure 25: Greeks as functions of the connectivity for different spot prices. Parameters: $N = 40$, $\sigma = 0.15$, $r = 0$, $T = 1$

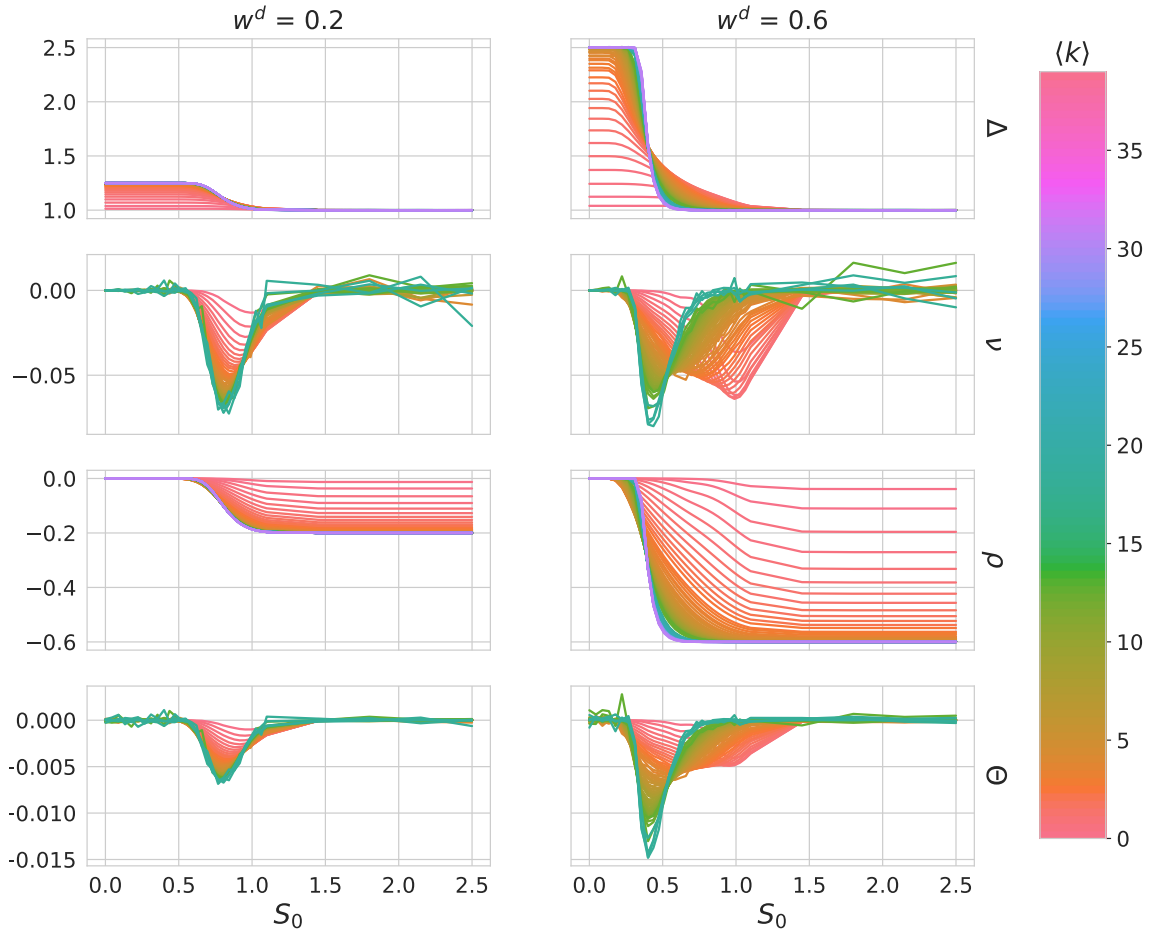


Figure 26: Greeks, value and solvency as functions of the spot price for different connectivities. Parameters: $N = 40$, $\sigma = 0.15$, $r = 0$, $T = 1$.

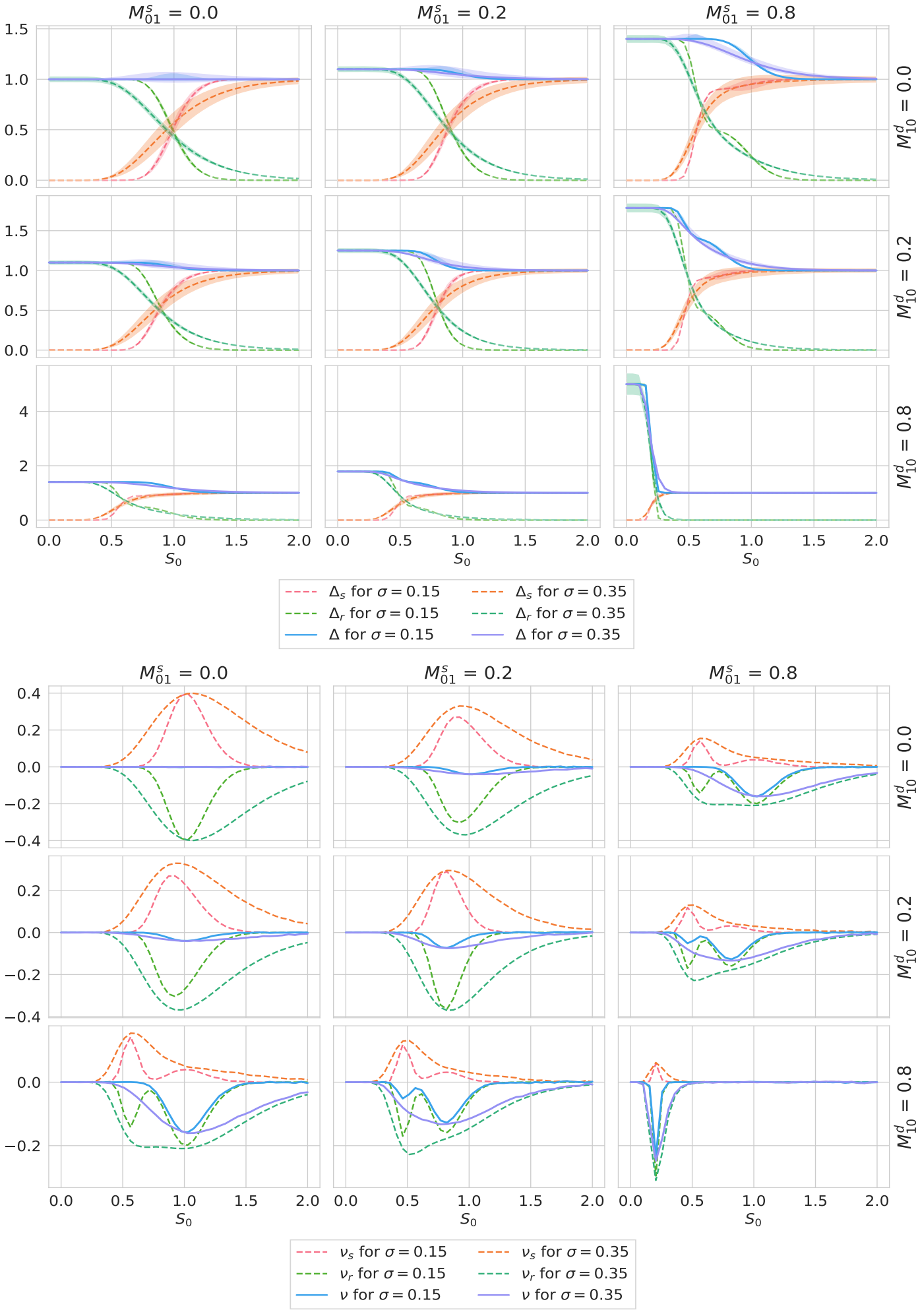


Figure 27: Δ and ν for fixed $r = 0$, $\tau = 1$, $d = 1$ and varying volatility and values of debt cross-holdings.

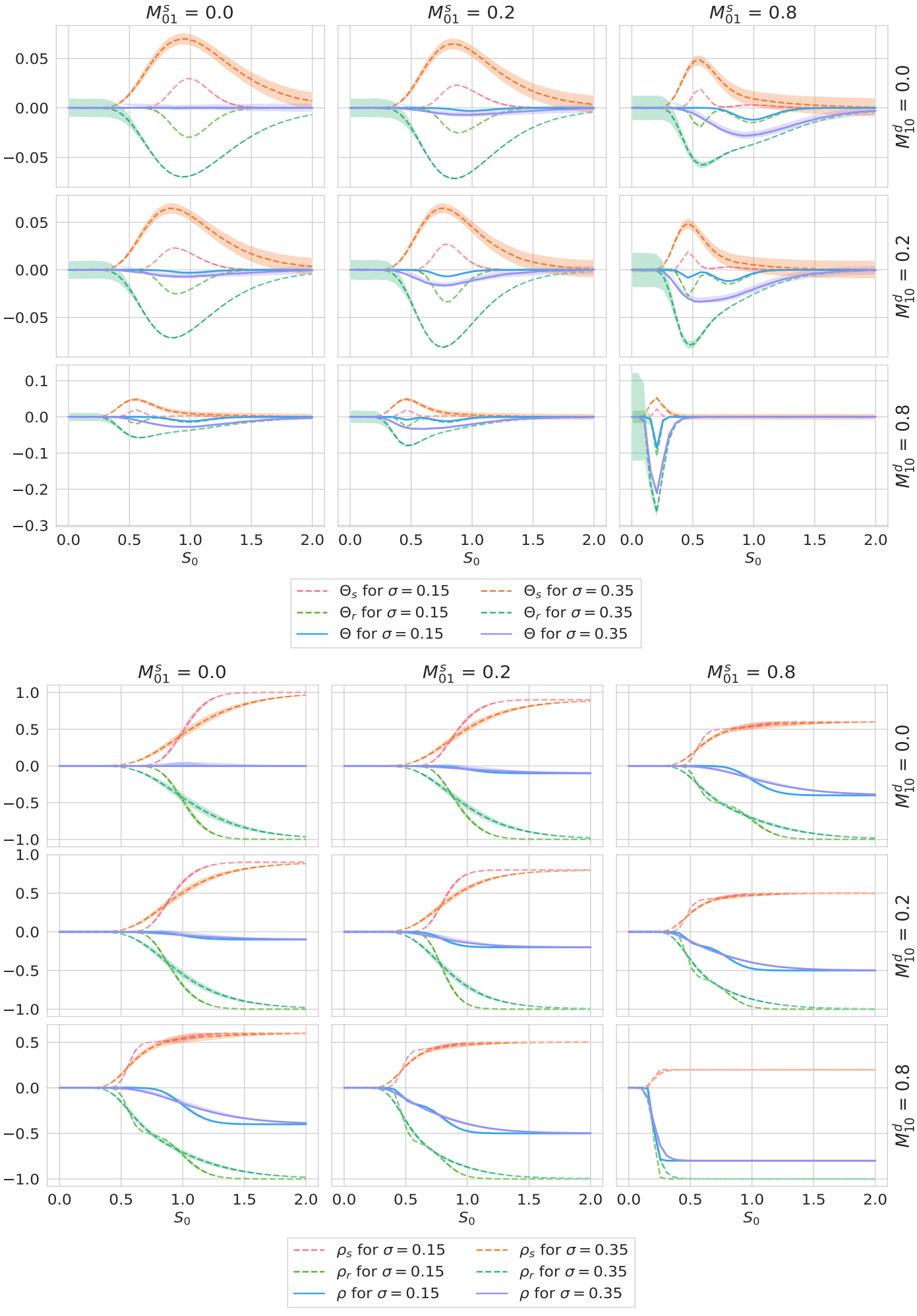


Figure 27: Θ and ρ for fixed $r = 0$, $\tau = 1$, $d = 1$ and varying volatility and values of debt cross-holdings.

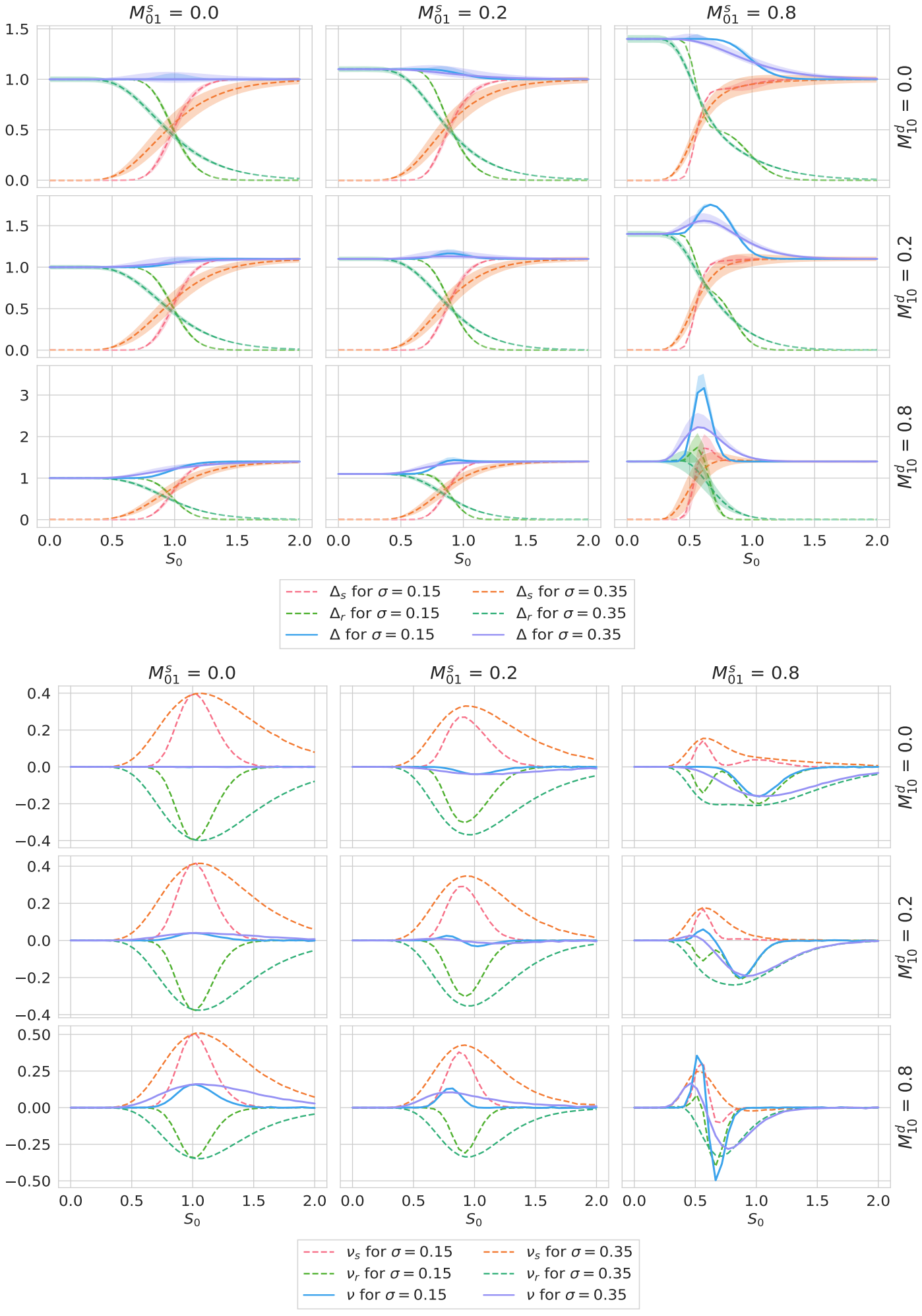


Figure 28: Δ and ν for fixed $r = 0$, $\tau = 1$, $d = 1$ and varying volatility and values of equity debt cross-holdings.

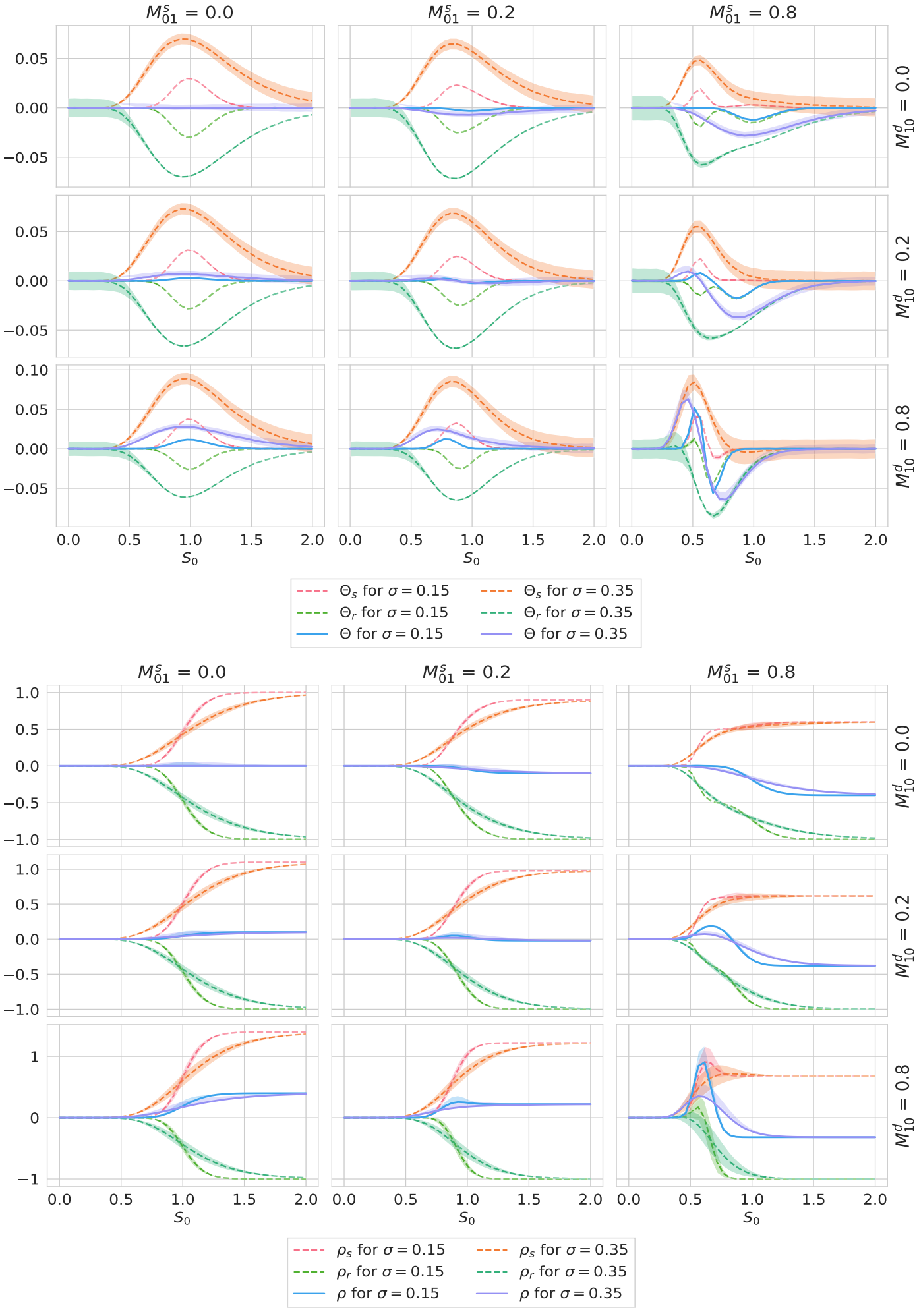


Figure 28: Θ and ρ for fixed $r = 0$, $\tau = 1$, $d = 1$ and varying volatility and values of equity and debt cross-holdings.

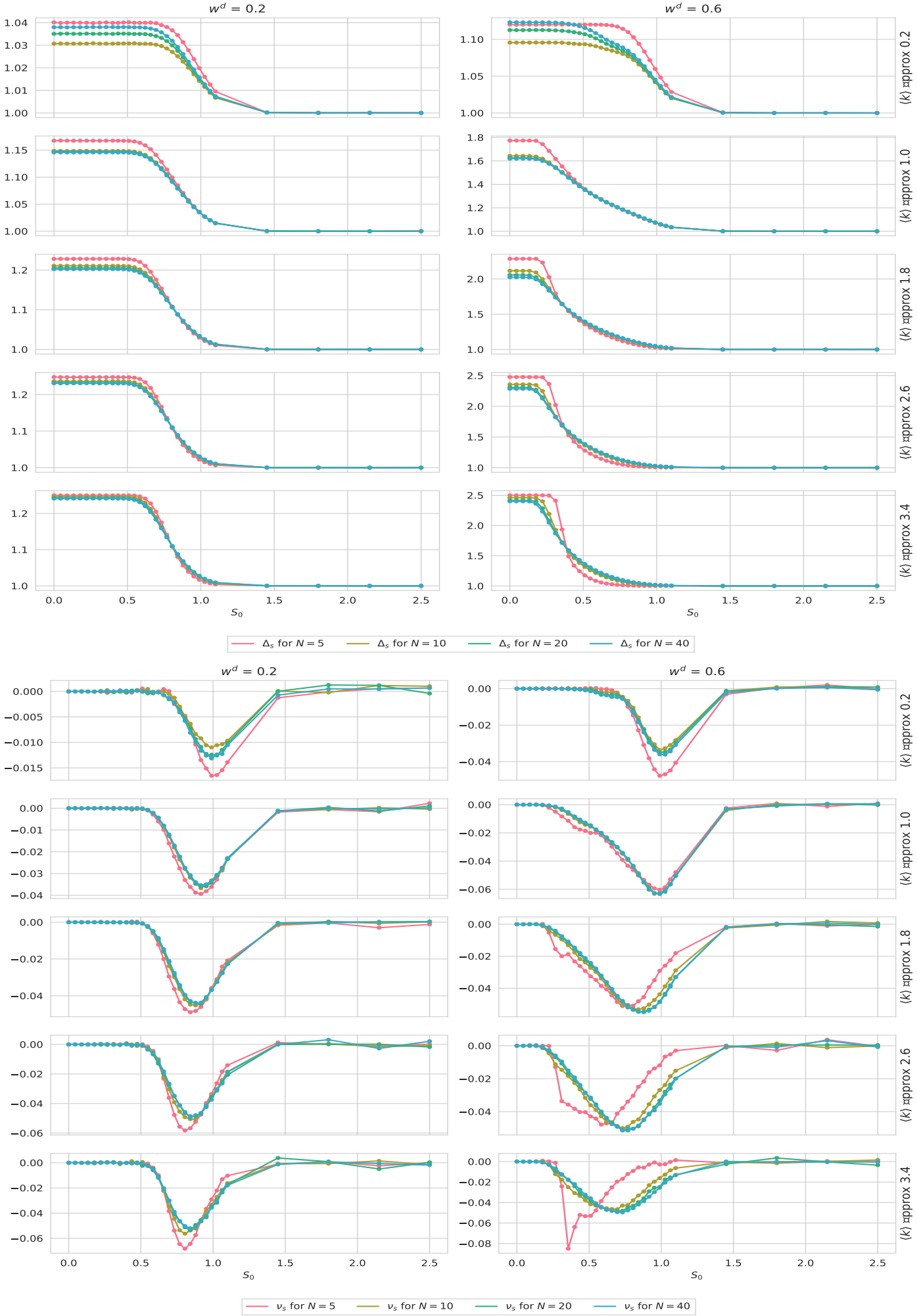


Figure 29: $\Delta(S_0)$ and $\nu(S_0)$ for differing networks sizes with $N_{\text{net samples}} = 2000$ and $N_{\text{MC}} = 100$

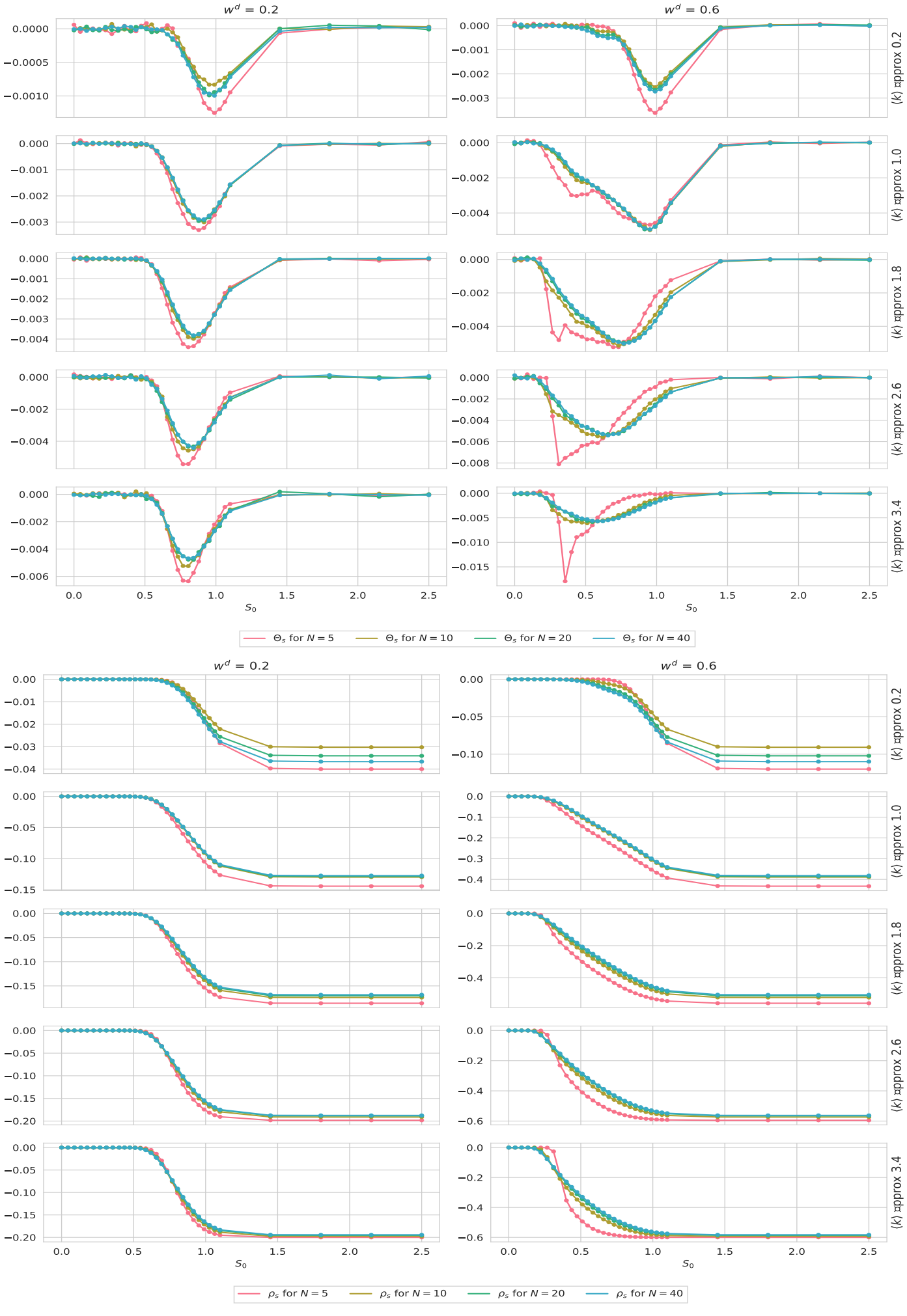


Figure 29: $\Theta(S_0)$ and $\rho(S_0)$ for differing networks sizes with $N_{\text{net samples}} = 2000$ and $N_{\text{MC}} = 100$

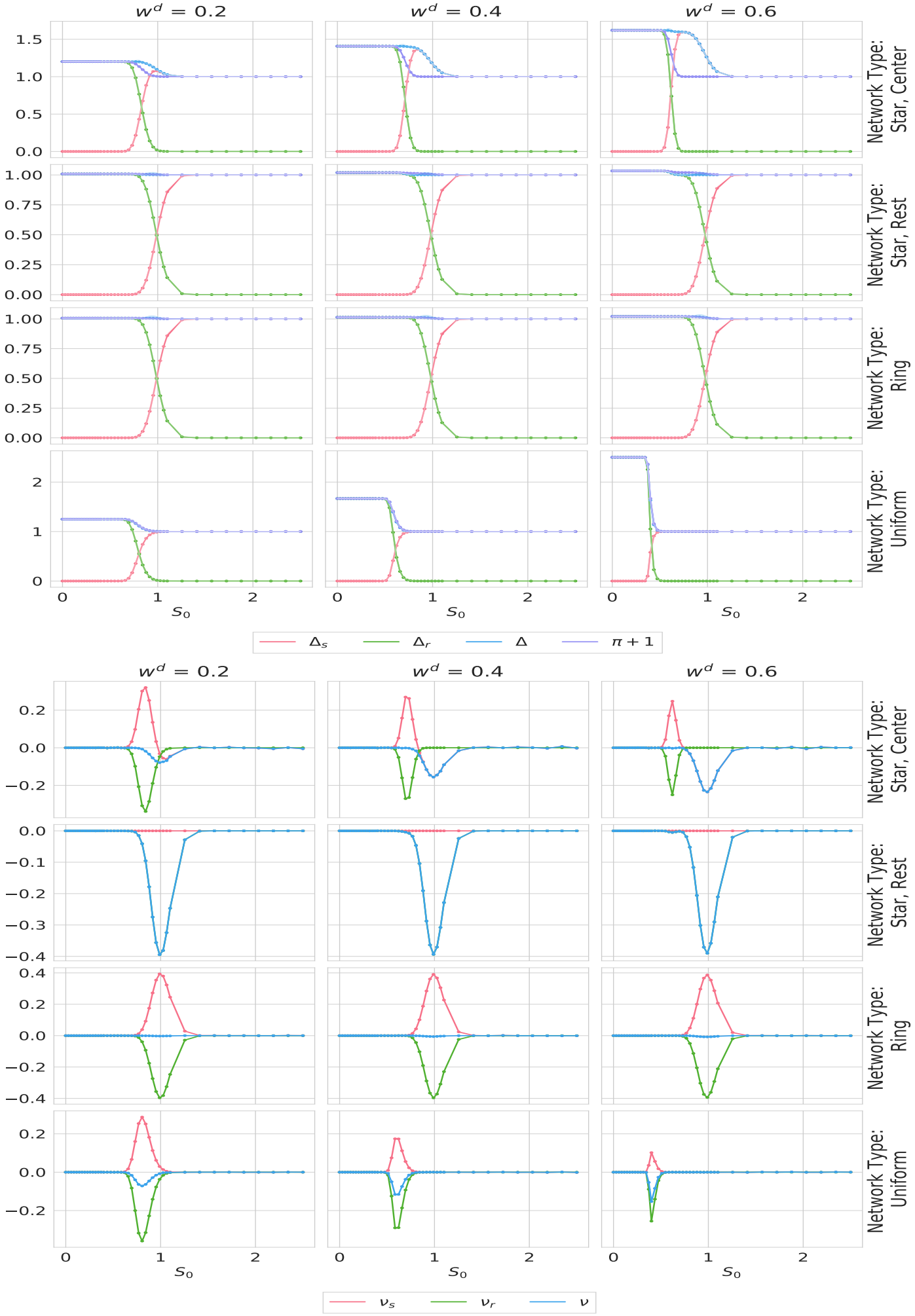


Figure 30: $\Delta(S_0)$ and $\nu(S_0)$ for different network topologies with $N_{\text{MC}} = 5000$, $r = 0$, $T = 1$, $\sigma = 0.1$, $N = 30$

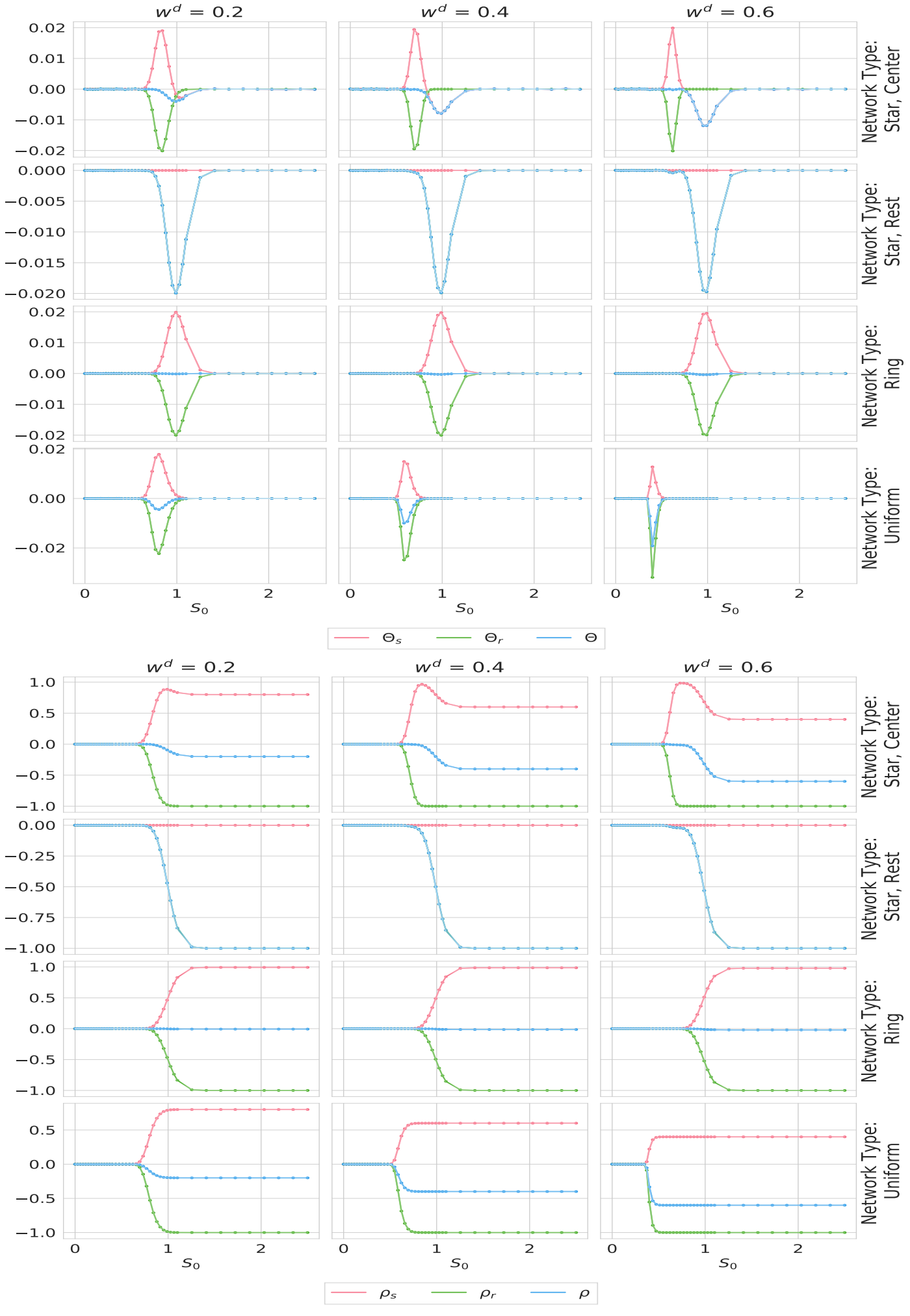


Figure 30: $\Theta(S_0)$ and $\rho(S_0)$ for different network topologies with $N_{\text{MC}} = 5000$

8

PROGRAM

During the course of the thesis, a simulation was implemented (the code is available here [Stobbe 2018b]). In this chapter I will give a short overview of the general structure and dependencies. While the reimplementation in `Julia` is likely the better choice in order to reproduce results or possible extensions, only the original code will be discussed here [Bertschinger and Stobbe 2019]. The simulation code consists of a back-end written in `C++`, that is accessed through a number of `Python` scripts. The `C++` code consists of five distinct parts: graph model generators, network valuation and network Greeks computation, Monte Carlo accumulators and statistical analysis tools, local front-end (in order to run the program without the `Python` front-end) and testing toolkits. The `Python` scripts consist of example simulation runs. All figures in this thesis can be directly reproduced (or are mostly already contained in examples), using the files in the "scripts" subdirectory.
In fig. 31, the code structure is shown as a dependency graph on the data accumulation code (which accumulates different kinds of statistics from samples). The `BlackScholesNetwork.hpp`

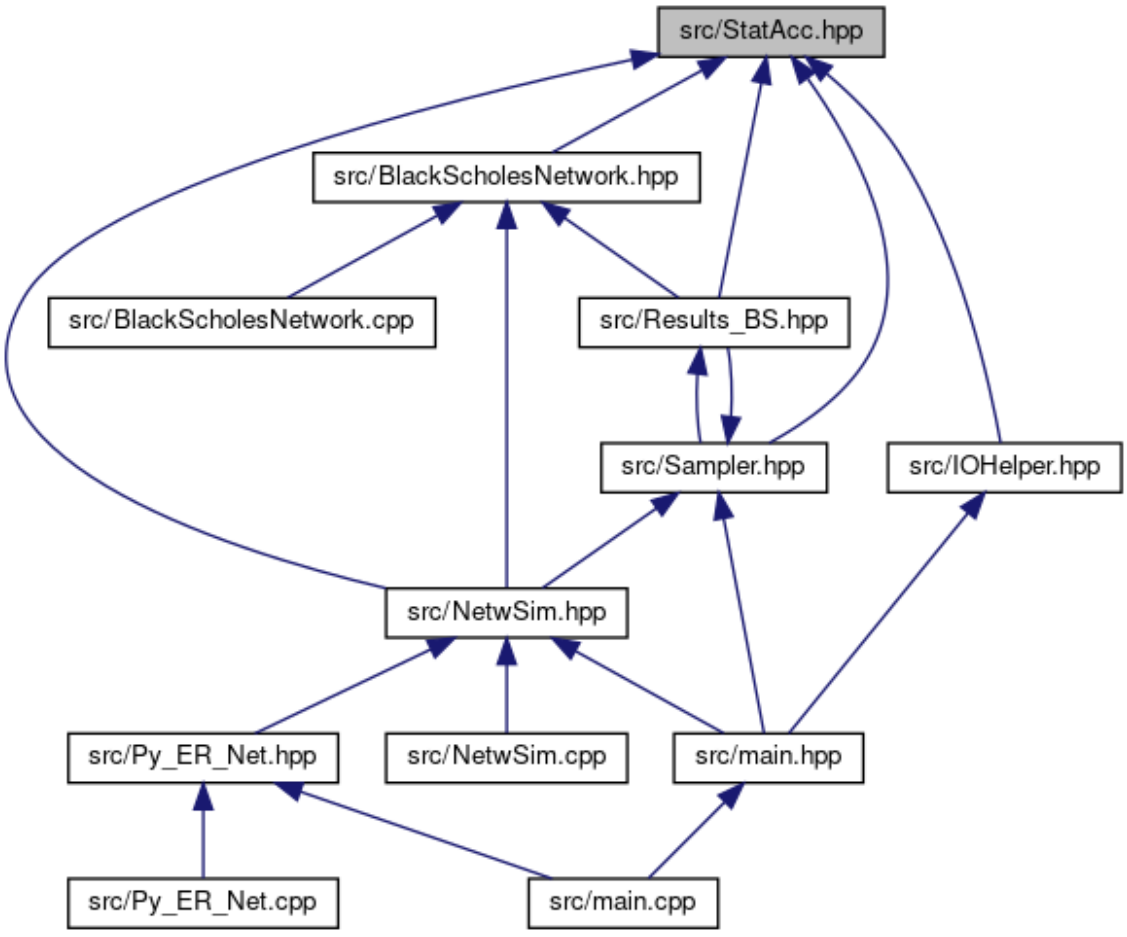


Figure 31: Code structure as dependency graph of the data accumulation class

contains the valuation algorithms for the network (i.e. fixed point iteration of $x = Mx$). The prob-

ability distributions[1] from which exogenous assets and adjacency matrices should be drawn are specified in the `NetwSim.hpp`. This functionality is encapsulated in a class, that is exposed through a `Python` API. Results and corresponding statistics are accumulated in `Sampler.hpp` and stored in a data structure, defined in `Results_BS.hpp`.
The source code is compiled using cmake, producing an executable for testing purposes and a shared library that can be used with a standard `Python` import statement. Compilation flags for cmake control, whether internal MPI parallelization during sampling is used.
Several external (but commonly used) packages are needed in order to run theses scripts. For example `IPyParallel` is used in order to provide parallelization over different parameter sets. The dependencies of the code include

`pybind11` Header library, exposing the top level functions that handle the setup of the simulation and extraction of the results from `C++` to `Python`

`eigen3` Internal data storage of vector valued quantities, exposed to `Python` through `PyBind11` without additional copy operations via `Numpy` arrays.

`trng` Threadsafe pseudo random number generator, supports MPI and can sample from all distributions required for this thesis except the multivariate student-T distribution

`easylogging++` Automated logging and error reporting support, including performance timer and automated log file generation with minimal performance overhead. Logging is also supported for the functions exported to the `Python` module.

`numpy` This is the only required library on the `Python` end, as simulation results are returned in form of `Numpy` arrays.

Since valuation and Greeks for any given set of parameters are often computed within seconds, the parallelization defaults to the scripting stage and used to computed several models with different parameters at the same time to avoid synchronization overhead.

1 More specifically, the algorithms to generate samples, since the exact distribution may not be known (see previous chapter).

9 SUMMARY AND OUTLOOK

In this thesis I have implemented and tested a numerical model for the quantification of systemic risks in financial networks. The limits of one firm and infinitely many equivalent firms have been discussed and used as verification of the more general network model. The code, computing valuations for more general networks has been tested against theses special cases. For the analytically inaccessible network structures, numerical results on the basis of the Erdős Rényi model have been obtained, providing insight into more realistic network structures.

There are several possible extensions of this project: Instead of generating random networks with symmetric firms, one could consider community structures or special topologies, possibly resembling real world networks more closely. Furthermore, one can relate other known methods for risk in financial networks to ours. The analytic considerations have already been discussed [Bertschinger and Stobbe 2018], but a comparison of simulated quantities could also be insteresting from a numerical performance standpoint.

As of now, our model only uses a single, fixed maturity, uncorrelated volatilities and European options. A more general model could be relatively easy to define, in cases where only the term in the expected value changes and the implicit function theorem trick is still applicable, the calculation of the Greeks still remains valid. One example could be inhomogeneous contracts between firms (i.e. multiple maturities). The overall structure of the program remains the same under this extension, but the fixed point calculation has to be carried out at multiple maturities, leaving the existence and uniqueness unclear.

Another unrealistic feature of the current model is the neglect of default costs. This additional strain on the network under stress could make certain network topologies more viable than others.

comment

BIBLIOGRAPHY

Allen-Zhu, Zeyuan et al. (Oct. 2017). "Much faster algorithms for matrix scaling". In: *Annual Symposium on Foundations of Computer Science - Proceedings*. Vol. 2017-Octob. IEEE, pp. 890–901. ISBN: 9781538634646. DOI: 10.1109/FOCS.2017.87. URL: http://ieeexplore.ieee.org/document/8104119/.

Ambegaokar, Vinay and Matthias Troyer (Feb. 2009). "Estimating errors reliably in Monte Carlo simulations of the Ehrenfest model". In: *American Journal of Physics* 78.2, pp. 150–157. ISSN: 0002-9505. DOI: 10.1119/1.3247985. URL: http://arxiv.org/abs/0906.0943%7B%5C%%7D0Ahttp://dx.doi.org/10.1119/1.3247985.

Angel, Omer, Remco van der Hofstad, and Cecilia Holmgren (Mar. 2016). "Limit laws for self-loops and multiple edges in the configuration model". In: URL: https://arxiv.org/pdf/1603.07172.pdf%20http://arxiv.org/abs/1603.07172.

Bauke, Heiko and Stephan Mertens (June 2007). "Random numbers for large-scale distributed Monte Carlo simulations". In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 75.6, p. 066701. ISSN: 15393755. DOI: 10.1103/PhysRevE.75.066701. URL: https://link.aps.org/doi/10.1103/PhysRevE.75.066701.

Bertschinger, Nils and Julian Stobbe (Oct. 2018). "Systemic Greeks: Measuring risk in financial networks". In: URL: http://arxiv.org/abs/1810.11849.

– (2019). *FinNetValu*. URL: https://github.com/bertschi/FinNetValu.

Bingham, Nicholas H. and Rüdiger Kiesel (2004). *Risk-Neutral Valuation*. Springer Finance. London: Springer London, p. 456. ISBN: 978-1-84996-873-7. DOI: 10.1007/978-1-4471-3856-3. URL: http://link.springer.com/10.1007/978-1-4471-3856-3.

Black, Fischer and Myron Scholes (May 1973). "The Pricing of Options and Corporate Liabilities". In: *Journal of Political Economy* 81.3, pp. 637–654. ISSN: 0022-3808. DOI: 10.1086/260062. URL: http://www.journals.uchicago.edu/doi/10.1086/260062.

Bøhren, Øyvind and Dag Michalsen (Sept. 1994). "Corporate cross-ownership and market aggregates: Oslo stock exchange 1980-1990". In: *Journal of Banking and Finance* 18.4, pp. 687–704. ISSN: 03784266. DOI: 10.1016/0378-4266(94)00015-8. URL: https://ac.els-cdn.com/0378426694000158/1-s2.0-0378426694000158-main.pdf?%7B%5C_%7Dtid=9db13c23-711f-4c34-a765-b74e8a48c2a2%7B%5C&%7Dacdnat=1546426719%7B%5C_%7D70164a1cb4c3296f389e1217348908ec%20http://linkinghub.elsevier.com/retrieve/pii/0378426694000158.

Curtiss, J (Dec. 1941). "On the Distribution of the Quotient of Two Chance Variables". In: *The Annals of Mathematical Statistics* 12.4, pp. 409–421. ISSN: 00034851. DOI: 10.1214/aoms/1177731679. URL: http://projecteuclid.org/euclid.aoms/1177731679%7B%5C%%7D5Cnpapers2://publication/uuid/1307DCE4-140F-43D5-825F-5A587B6CE055.

Dash, Jan W. (June 2004). *Quantitative Finance and Risk Management*. Ed. by Dash Wu. Vol. 91. Berlin, Heidelberg: WORLD SCIENTIFIC, pp. 1–802. ISBN: 9812387129. DOI: 10.1142/9003. URL: http://link.springer.com/10.1007/978-3-642-19339-2%20https://www.worldscientific.com/worldscibooks/10.1142/9003.

De Souza, Elizeu Clementino, José Antonio Serrano Castañeda, and Juan Mario Ramos Morales (2014). "Autobiografía y educación: Tradiciones, diálogos y metodologías". In: *Revista Mexicana de Investigacion Educativa* 19.62, pp. 683–694. ISSN: 14056666. DOI: 10.2139/ssrn.1784440. URL: http://www.ssrn.com/abstract=1784440.

Delbaen, F. and W. Schachermayer (Oct. 1998). "The fundamental theorem of asset pricing for unbounded stochastic processes". In: *Mathematische Annalen* 312.2, pp. 215–250. ISSN: 00255831. DOI: 10.1007/s002080050220. URL: http://link.springer.com/10.1007/s002080050220.

Delbaen, Freddy and Walter Schachermayer (2006). *The Mathematics of Arbitrage*. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-540-21992-7. DOI: 10.1007/978-3-540-31299-4. URL: http://link.springer.com/10.1007/978-3-540-31299-4.

Demirözer, Ozan, Ismail Karaca, and Yusuf Karsavuran (2011). *Population fluctuations of some important pests and natural enemies found in Oil-bearing rose (Rosa damascena Miller) production areas in Isparta province (Turkey)*. Vol. 35. 4. Boston, MA: Springer US, pp. 539–558. ISBN: ISSN 1010-6960. DOI: 10.1007/978-1-4899-4541-9. URL: http://link.springer.com/10.1007/978-1-4899-4541-9.

Diaconis, Persi, Gilles Lebeau, and Laurent Michel (Apr. 2012). "Gibbs/Metropolis algorithms on a convex polytope". In: *Mathematische Zeitschrift* 272.1-2, pp. 109–129. ISSN: 00255874. DOI: 10.1007/s00209-011-0924-5. URL: https://arxiv.org/pdf/1104.0749.pdf%20http://arxiv.org/abs/1104.0749.

Du, Kun, Guo Liu, and Guiding Gu (2014). "Accelerating Monte Carlo Method for Pricing Multi-asset Options under Stochastic Volatility Models". In: *International Journal of Applied Mathematics* 44.2, pp. 62–70. ISSN: 19929978. URL: http://www.iaeng.org/IJAM/issues%7B%5C_%7Dv44/issue%7B%5C_%7D2/IJAM%7B%5C_%7D44%7B%5C_%7D2%7B%5C_%7D01.pdf.

Dudley, Richard M. (2018). *Real analysis and probability*. Cambridge University Press, pp. 1–436. ISBN: 9781351084642. DOI: 10.1201/9781351076197.

Dufossé, Fanny et al. (May 2018). "Further notes on Birkhoff–von Neumann decomposition of doubly stochastic matrices". In: *Linear Algebra and Its Applications* 554, pp. 68–78. ISSN: 00243795. DOI: 10.1016/j.laa.2018.05.017. URL: http://dx.doi.org/10.1016/j.laa.2016.02.023%20https://linkinghub.elsevier.com/retrieve/pii/S0024379518302568%20http://linkinghub.elsevier.com/retrieve/pii/S0024379516001257.

Dyer, Martin, Alan Frieze, and Ravi Kannan (Jan. 1991). "A random polynomial-time algorithm for approximating the volume of convex bodies". In: *Journal of the ACM* 38.1, pp. 1–17. ISSN: 00045411. DOI: 10.1145/102782.102783. URL: http://portal.acm.org/citation.cfm?doid=102782.102783.

Elliott, Robert J. and P. Ekkehard Kopp (1999). *Mathematics of Financial Markets*. 2nd Editio. Springer Finance. New York: Springer-Verlag. ISBN: 978-1-4757-7148-0. DOI: 10.1007/978-1-4757-7146-6. URL: http://link.springer.com/10.1007/978-1-4757-7146-6.

Erd\H os, Paul and Alfréd Rényi (1959). "On Random Graphs I." In: *Publicationes Mathematicae (Debrecen)* 6, pp. 290–297.

Fallis, A.G and John C Hull (2006). *Options, futures, and other derivatives*. 8th. Vol. 53. 9. Prentice Hall, pp. 1689–1699. ISBN: 9788578110796. DOI: 10.1017/CBO9781107415324.004.

Fenner, Trevor I. and Georghios Loizou (Dec. 1971). "On fully indecomposable matrices". In: *Journal of Computer and System Sciences* 5.6, pp. 607–622. ISSN: 10902724. DOI: 10.1016/S0022-0000(71)80018-1. URL: http://linkinghub.elsevier.com/retrieve/pii/S0022000071800181.

Ferradans, Sira et al. (July 2013). "Regularized discrete optimal transport". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7893 LNCS, pp. 428–439. ISSN: 03029743. DOI: 10.1007/978-3-642-38267-3_36. URL: http://arxiv.org/abs/1307.5551.

Fischer, Tom (2014). "No-arbitrage pricing under systemic risk: Accounting for cross-ownership". In: *Mathematical Finance* 24.1, pp. 97–124. ISSN: 09601627. DOI: 10.1111/j.1467-9965.2012.00526.x.

Frieze, Alan M. et al. (Apr. 1988). "Reconstructing Truncated Integer Variables Satisfying Linear Congruences". In: *SIAM Journal on Computing* 17.2, pp. 262–280. ISSN: 0097-5397. DOI: 10.1137/0217016. URL: http://epubs.siam.org/doi/10.1137/0217016.

Frieze, Alan and Michal Karonski (2016). *Introduction to Random Graphs*. Cambridge: Cambridge University Press. ISBN: 9781316339831. DOI: 10.1017/CBO9781316339831. URL: http://ebooks.cambridge.org/ref/id/CBO9781316339831.

Gubernatis, J. E., N. (Naoki) Kawashima, and P. Werner (2016). *Quantum Monte Carlo methods : algorithms for lattice models*. 1st ed. Cambridge University Press, p. 488. ISBN: 9781107006423. URL: https://www.amazon.com/Quantum-Monte-Carlo-Methods-Algorithms/dp/1107006422.

Gubner, John A. (2010). *Stochastic differential equations*. Universitext. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1333–1352. ISBN: 9781420073652. DOI: 10.1201/b10384. URL: http://th.

if.uj.edu.pl/%7B~%7Dgudowska/dydaktyka/Oksendal.pdf%20http://link.springer.com/10.1007/978-3-642-14394-6.

Harrison, J. Michael and Stanley R. Pliska (Aug. 1981). "Martingales and stochastic integrals in the theory of continuous trading". In: *Stochastic Processes and their Applications* 11.3, pp. 215–260. ISSN: 03044149. DOI: 10.1016/0304-4149(81)90026-0. URL: http://linkinghub.elsevier.com/retrieve/pii/0304414981900260.

Hellekalek, P. (June 1998). "Good random number generators are (not so) easy to find". In: *Mathematics and Computers in Simulation* 46.5-6, pp. 485–505. ISSN: 03784754. DOI: 10.1016/S0378-4754(98)00078-0. URL: http://linkinghub.elsevier.com/retrieve/pii/S0378475498000780.

Joho, Rolf H. et al. (Feb. 1990). "Toxin and kinetic profile of rat brain type III sodium channels expressed in Xenopus oocytes". In: *Molecular Brain Research* 7.2, pp. 105–113. ISSN: 0169328X. DOI: 10.1016/0169-328X(90)90087-T. URL: http://www.ncbi.nlm.nih.gov/pubmed/2160038.

Kahn, H. and A. W. Marshall (1953). "Methods of Reducing Sample Size in Monte Carlo Computations". In: *Journal of the Operations Research Society of America* 1.5, pp. 263–278. ISSN: 0096-3984. DOI: 10.1287/opre.1.5.263. URL: http://pubsonline.informs.org/doi/abs/10.1287/opre.1.5.263.

Klenke, Achim (2013). *Wahrscheinlichkeitstheorie*. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-642-36017-6. DOI: 10.1007/978-3-642-36018-3. URL: http://link.springer.com/10.1007/978-3-642-36018-3.

L'Ecuyer, Pierre and Richard Simard (Aug. 2007). "TestU01". In: *ACM Transactions on Mathematical Software* 33.4, 22–es. ISSN: 00983500. DOI: 10.1145/1268776.1268777. URL: http://portal.acm.org/citation.cfm?doid=1268776.1268777.

Lai, Hang Chin (Dec. 1986). "On the linear algebra of generalized doubly stochastic matrices and their equivalence relations and permutation basis". In: *Japan Journal of Applied Mathematics* 3.2, pp. 357–379. ISSN: 09102043. DOI: 10.1007/BF03167108. URL: http://link.springer.com/10.1007/BF03167108.

Malik, Jamshaid Sarwar and Ahmed Hemani (Nov. 2016). "Gaussian Random Number Generation". In: *ACM Computing Surveys* 49.3, pp. 1–37. ISSN: 03600300. DOI: 10.1145/2980052. URL: http://dl.acm.org/citation.cfm?doid=2988524.2980052.

Merton, Robert C. (May 1974). "on the Pricing of Corporate Debt: the Risk Structure of Interest Rates*". In: *The Journal of Finance* 29.2, pp. 449–470. ISSN: 00221082. DOI: 10.2307/2978814. URL: http://doi.wiley.com/10.1111/j.1540-6261.1974.tb03058.x.

Pham, Huyên (2009). *Continuous-time Stochastic Control and Optimization with Financial Applications*. Vol. 61, p. 249. ISBN: 978-3-540-89499-5. DOI: 10.1007/978-3-540-89500-8. URL: http://link.springer.com/10.1007/978-3-540-89500-8.

Privault, N (2016). *Notes on Stochastic Finance: Stochastic Calculus for JumpStochastic*, p. 932. URL: http://www.ntu.edu.sg/home/nprivault/indext.html.

Protter, Philip E. (2005). *Stochastic Integration and Differential Equations*. Vol. 21. Stochastic Modelling and Applied Probability. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-642-05560-7. DOI: 10.1007/978-3-662-10061-5. URL: http://link.springer.com/10.1007/978-3-662-10061-5.

Robert G. Brown (2017). *Diehard Test*. URL: http://webhome.phy.duke.edu/%7B~%7Drgb/General/dieharder.php (visited on 08/09/2017).

Rothblum, Uriel G. and Hans Schneider (Mar. 1989). "Scalings of matrices which have prespecified row sums and column sums via optimization". In: *Linear Algebra and Its Applications* 114-115.C, pp. 737–764. ISSN: 00243795. DOI: 10.1016/0024-3795(89)90491-6. URL: http://linkinghub.elsevier.com/retrieve/pii/0024379589904916.

Sapatinas, Theofanis (2004). *The Elements of Statistical Learning*. Vol. 167. Springer Series in Statistics 1. New York, NY: Springer New York, pp. 192–192. ISBN: 978-0-387-84857-0. DOI: 10.1111/j.1467-985X.2004.298_11.x. URL: http://doi.wiley.com/10.1111/j.1467-985X.2004.298%7B%5C_%7D11.x.

Schachermayer, W. (Dec. 1992). "A Hilbert space proof of the fundamental theorem of asset pricing in finite discrete time". In: *Insurance Mathematics and Economics* 11.4, pp. 249–257. ISSN: 01676687.

DOI: 10.1016/0167-6687(92)90013-2. URL: http://linkinghub.elsevier.com/retrieve/pii/0167668792900132.

Sinkhorn, Richard and Paul Knopp (May 1967). "Concerning nonnegative matrices and doubly stochastic matrices". In: *Pacific Journal of Mathematics* 21.2, pp. 343–348. ISSN: 0030-8730. DOI: 10.2140/pjm.1967.21.343. URL: http://msp.org/pjm/1967/21-2/p14.xhtml.

Spivak, Michael (Sept. 1965). "Calculus on Manifolds". In: *American Mathematical Monthly*. Vol. 75. 5. WORLD SCIENTIFIC, p. 567. ISBN: 0805390219. DOI: 10.2307/2314769. URL: http://www.jstor.org/stable/2314769?origin=crossref.

Spreij, PJC (2010). *Measure Theoretic Probability*. Tech. rep., pp. 2010–2010.

Stobbe, Julian (2018a). "Master Thesis Implementation of continuous time quantum Monte Carlo solvers for the infinite dimensional Hubbard model at half filling Julian Stobbe Johann Wolfgang Goethe-Universität". Master Thesis. Goethe University Frankfurt am Main, p. 165.

– (2018b). *MSc Thesis*. URL: https://github.com/Atomtomate/sys%7B%5C_%7Drisk.

Suzuki, Teruyoshi (2002). "Valuing corporate debt: The effect of cross-holdings of stock and debt". In: *Journal of the Operations Research Society of Japan* 45.2, pp. 123–144. ISSN: 04534514. DOI: 10.15807/jorsj.45.123. URL: http://www.orsj.or.jp/%7B~%7Darchive/pdf/e%7B%5C_%7Dmag/Vol.45%7B%5C_%7D02%7B%5C_%7D123.pdf%20https://www.jstage.jst.go.jp/article/jorsj/45/2/45%7B%5C_%7DKJ00001203152/%7B%5C_%7Darticle.

Toro, Mauricio et al. (June 2016). "Gelisp: A framework to represent musical constraint satisfaction problems and search strategies". In: *Journal of Theoretical and Applied Information Technology* 86.2, pp. 327–331. ISSN: 18173195. DOI: 10.1214/aoms/1177705148. URL: http://www.jstor.org/stable/2238545.

Wikipedia and Contributors (2019a). *Conditional expectation — Wikipedia, The Free Encyclopedia*. URL: https://en.wikipedia.org/wiki/Conditional%7B%5C_%7Dexpectation.

– (2019b). *Greeks (finance) — Wikipedia, The Free Encyclopedia*. URL: https://en.wikipedia.org/wiki/Greeks%7B%5C_%7D(finance).

Yeh, J. and E.M. Vestrup (2003). "Real Analysis: Theory of Measure and Integration". In: *World Scientific*. 3rd ed. Vol. 101. 473. World Scientific, p. 840. ISBN: 9812566538. DOI: 10.1198/jasa.2006.s64. URL: http://books.google.com/books?id=4k2uKwwelVUC%7B%5C%%7D5Cnhttps://books.google.ca/books?id=%7B%5C_%7DjKjngEACAAJ%7B%5C&%7Ddq=real+analysis+yeh%7B%5C&%7Dhl=en%7B%5C&%7Dsa=X%7B%5C&%7Dei=m5%7B%5C_%7DSVLnwAtHToATM6IHADA%7B%5C&%7Dved=0CB4Q6AEwAA.