Revisiting the recovery theorem: Why does practice refute theory?

by

Okke P.H. Labout (SNR: 2053194)

A thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science in Quantitative Finance & Actuarial Science

Tilburg School of Economics and Management
Tilburg University

Supervised by: Prof. Dr. Bas J.M. Werker
Second Reader: Dr. Christoph Hambel

Date: July 23, 2024

Abstract

The recovery theorem by Ross (2015) introduced a way to estimate the real-world transition probabilities using state prices. However, for the recovery theorem to work, four assumptions need to hold, and empirical papers show that some of these four assumptions do not hold in reality.

In this thesis, I will refine the recovery theorem by adding one additional assumption: *"(strong) rational expectations"*. Furthermore, I find that if all five assumptions hold, the estimated pricing kernel becomes long-term risk neutral and that in a world where the risk-free (interest) rate is constant, the market is only influenced by idiosyncratic risk.

Contents

Acknowledgements

I would like to thank Prof. Dr. Bas J.M. Werker for not only supervising me, but also introducing me to the interesting research field that tries to elicit information from the risk neutral probabilities that could help us tell something about the real-world. Every meeting we had inspired and helped me understand this field and his feedback helped me elevate my thesis constantly.

Furthermore, I would like to thank the members of Tilburg University's department of finance for giving their feedback regarding my thesis, particularly Prof. Dr. Rik Frehen and Prof. Dr. Frank de Jong. During an online discussion, a few members of the department shared their opinions, suggestions and feedback and all of these helped me improve my thesis even further.

Finally, I would like to thank my friends and family for their support.

I am grateful for the various support provided by these people. They assisted me in completing this thesis, which allowed me to obtain a master's degree in Quantitative Finance and Actuarial Science.

Introduction

In asset pricing, a well-known problem is that asset prices are determined by supply and demand and consequently, (expected) asset returns consist out of the risk-free (interest) rate and a risk premium. That is, if μ is some return of an asset and r the risk-free rate, then $\mu = r + (\mu - r)$, where $\mu - r$ is the so called "*risk premium*"[1]. Since we can disentangle the risk-free rate and the risk premium from the return of an asset, it should be possible to fully eliminate this risk premium if we consider a market with only risk neutral agents. This lead economists to introduce the so called "*risk neutral probability measure*" ($\mathbb{Q}$), and by introducing this risk neutral probability measure, economist were able to construct the first fundamental theorem of asset pricing. The first fundamental theorem of asset pricing states that if the market is free of arbitrage, there exists a risk neutral probability measure ($\mathbb{Q}$), that makes sure that the price of an asset is equal to its discounted expected future value.
This fundamental theorem is on itself a nice result, but it became even more useful when Arrow and Debreu (1954) introduced their model about the existence of a competitive economy equilibrium. This paper introduces (Arrow-Debreu) state prices and these have since been applied in financial economics. An (Arrow-Debreu) state price is the price of a security that pays off 1 unit of currency if a particular state materialises and zero if one of the other states materialises.
A possible explanation why state prices are renowned in financial economics is that researchers have wondered if there is more that these state prices can explain, other than the (risk neutral) expectation of the discounted future value of an Arrow-Debreu security. The motivation for this reasoning is that state prices are, for example, related to the pricing kernel and the pricing kernel is important for pricing any asset. Hence, state prices might include more information about this one state that the security is linked to, other than the payoff.
In previous research, researchers have tried to combine the first fundamental theorem of asset pricing with observed asset prices in order to obtain these risk neutral probabilities from market data[2]. The problem is that these $\mathbb{Q}$-probabilities are the risk neutral probabilities and not the real-world probabilities. This is for example the reason why in credit risk management it is difficult to estimate (real-world) default probabilities, since extracting probabilities from insurance contracts yield the risk neutral probabilities and not the real-world probabilities. If an insurer wants to price insurance contract, retrieving (or calibrating) these $\mathbb{Q}$-probabilities are the sole requirement to then price insurance contracts using the first fundamental theorem of asset pricing. However, for management purposes, the real-world probabilities are needed. Hence, there is extensive previous research in the relation between the real-world probabilities and the risk neutral probabilities, in the hope of finding a method to go from these risk neutral probabilities to the real-world probabilities. For example, Hansen and Jagannathan (1991), Dybvig and Rogers (1997) and (Cox and Leland, 2000) try to use market data to extract agents utility function and then use these utility function to tell something about the subjective beliefs of the agents.

In 2015, Stephen Ross introduced the recovery theorem. In this theorem, Ross claims that under certain assumptions it is possible to estimate real-world probabilities, using Arrow-Debreu state prices. Despite the fact that the recovery theorem is a theoretically (and mathematically) sound theorem, there are however papers that empirically question its usefulness (for example van Ap-

[1] alternatively, one can consider $X_t = E[k_{t+1}X_{t+1}] \iff 1 = E[k_{t+1}R_{t+1}] \iff E_t[R_{t+1}] = \frac{1}{E_t[k_{t+1}]} + \frac{cov_t(k_{t+1}, R_{t+1})}{E_t[k_{t+1}]}$ where $\frac{cov_t(k_{t+1}, R_{t+1})}{E_t[k_{t+1}]}$ is the risk premium (Fletcher, 2007; Grossman and Shiller, 1981)

[2] for example Jackwerth (2004) shows how option data can be used to extract these risk neutral probabilities

pel and Maré (2018),Bakshi, Chabi-Yo, and Gao (2017) and Jackwerth and Menner (2020)), since they found that, after testing the underlying assumptions, some of these assumptions do not hold in reality. Furthermore, Borovicka, Hansen, and Scheinkman (2016) express their criticism about the recovery theorem with respect to its implication regarding long-term risk. In their paper, Borovicka et al. (2016) question whether some assumptions may overgeneralise or oversimplify the reality. Hence, theoretically, the recovery theorem might sound as a great theorem, however practical usefulness is still debatable. This brings me to the purpose of this thesis.

The primary objective of this thesis is to bridge the gap between the highly mathematical and empirical papers. As well as explaining (and proving) concepts used by Ross (2015), where the main focus lies on the intuition behind all the underlying assumptions.

I will start this thesis by deriving the recovery theorem. I will first introduce the framework and the underlying assumptions that are needed such that the recovery theorem, formulated by Ross (2015), can hold. Moreover, I will refine the theorem by adding one assumption: "*strong rational expectations*", such that the theorem, theoretically, better suits the objective of retrieving the real-world probability measure. Then, in the second chapter of this thesis, I will replicate the example done by Ross (2015), to show how the recovery theorem works in a data generating process. In the chapter that follows, I will explain what happens with the recovery theorem once I leave out any of the underlying assumptions. By elaborating on these assumptions, it becomes clear why this theoretically sound theorem does not work in reality, since the required assumptions are not present in reality. In the chapter that follows I will highlight two interesting implications that follow from the recovery theorem; it turns out that if all five assumptions hold, the constructed pricing kernel is *"long-term risk neutral"* and in special circumstances, the recovery theorem finds that the market only is influenced by idiosyncratic risk. Finally, I will end this thesis with a conclusion and I will give suggestions for possible future research.

The Recovery Theorem

In this chapter, I will derive the recovery theorem that was formulated by Ross (2015). The theorem states:

Theorem 1 (Recovery Theorem)**.** *If there is Absence of Arbitrage, if the pricing matrix is irreducible, and if it is generated by a transition independent kernel, then there exists a unique (positive) solution to the problem of finding the natural probability transition matrix, P, the discount rate, β, and the pricing kernel, K_t.*
In other words, for any given set of state prices there is a unique compatible natural measure and a unique pricing kernel.

To start the derivation of the theorem, consider a probability space $(\Omega, (\mathcal{F}_t)_{t\in\{0,...,\tau\}}, \mathbb{P})$. During this thesis, I will assume:

Assumption 1. *the (real-world) probability measure $\mathbb{P}$ can be characterized by a (finite state-space) Markov chain.*

Furthermore, I assume that there are m states within the Markov chain and that at time $t = 0, ..., \tau$ the corresponding (Markov) state is denoted by $j_t = 1, ..., m$. In other words, over time I move from a starting (Markov) state $j_0 = 1, ..., m$ to an ending (Markov) state $j_\tau = 1, ..., m$. This means that if I consider the entire time frame $\{0, ..., \tau\}$ I have in total $\tau + 1$ (Markov) states visited.
Using the above notation, I have $\Omega = \{1, ..., m\}^{\tau+1}$, meaning that state of the world $\omega \in \Omega$ represents a sequence of (Markov) states of length $\tau + 1$, where j_0 denotes the starting (Markov) state of the sequence and j_τ denotes the ending (Markov) state of the state of the world (so, we have $\omega = \{j_0, ..., j_\tau\} \in \Omega$). The filtration $(\mathcal{F}_t)_{t\in\{0,...,\tau\}}$ corresponds to the accumulation of information over time (so, $\mathcal{F}_0 \subset \mathcal{F}_1 \subset ... \subset \mathcal{F}_\tau$) and the (real-world) probability measure is denoted by $\mathbb{P}$, such that the probability that state $\omega = \{j_0, ..., j_\tau\}$ materializes is $\mathbb{P}[\{\omega\}]$.
I will assume that the above mentioned probability space is a representation of the real-world, in which (financial) assets are present. This means that I will assume that any (Markov) state j_t represents some return of an asset. Consequently, Once we arrive at some (Markov) state $j_t = 1, ..., m$, I can compute the price of the asset using the fact that

$$\frac{S_t}{S_{t-1}} = e^{R_t} \iff S_t = S_{t-1}e^{R_t}, \tag{1}$$

where R_t represents the return in (Markov) state $j_t = 1, ..., m$. Furthermore, I will assume that there is absence of arbitrage within the market:

Assumption 2. *The market obeys absence of arbitrage.*

Since Assumption 1 holds, I can denote the probability of moving from (Markov) state i to (Markov) state j as $p(i, j)$ and, consequently, let me denote the transition matrix under probability measure $\mathbb{P}$ as

$$P = \begin{bmatrix} p(1,1) & \dots & p(1,m) \\ \vdots & \ddots & \vdots \\ p(m,1) & \dots & p(m,m) \end{bmatrix}. \tag{2}$$

Furthermore, Assumption 2 implies that there exists a risk neutral probability measure $\mathbb{Q}$ that is equivalent to the real-world probability measure $\mathbb{P}$. Therefore, let me denote the transition probabilities under $\mathbb{Q}$ as $q(i, j)$ $(i, j = 1, ..., m)$. Next, consider the definition of equivalent probability measures:

Definition. *Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $(\Omega, (\mathcal{F}_t)_{t\in\{0,...,\mathcal{T}\}})$. $\mathbb{Q}$ is said to be equivalent to $\mathbb{P}$ ($\mathbb{P} \sim \mathbb{Q}$) iff. $\mathbb{P}$ and $\mathbb{Q}$ share the same null space. That is, $\forall A \in \mathcal{F}_t$, $\mathbb{P}[A] = 0 \iff \mathbb{Q}[A] = 0$*

The probabilities under $\mathbb{P}$ are given by $p(i, j)$ and since $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, it must hold that $p(i, j) = 0 \iff q(i, j) = 0$. Furthermore, since the structure of the Markov chain depends on whether certain (Markov) states communicate with each other ($p(i, j) > 0$), it must hold that if two (Markov) states under $\mathbb{P}$ communicate, these same (Markov) states must communicate under $\mathbb{Q}$, otherwise it would be possible that $p(i, j) > 0 \iff q(i, j) = 0$ for some $i, j = 1, ..., m$ and this violates the equivalence between probability measures $\mathbb{P}$ and $\mathbb{Q}$. Hence, under Assumption 1 and Assumption 2, the transition matrix under $\mathbb{Q}$ looks like:

$$Q = \begin{bmatrix} q(1,1) & \dots & q(1,m) \\ \vdots & \ddots & \vdots \\ q(m,1) & \dots & q(m,m) \end{bmatrix}. \tag{3}$$

I will continue by introducing the concept of "*irreducibility*" (see, for example (Resnick, 1992)).

Definition (Irredicibility)**.** *Consider a Markov chain with m (Markov) states. This Markov chain is said to be irreducible, if I can reach any state j from any state i within a finite number of movements.*

Consider the following two examples:

Example 1. *Consider the following Markov chain:*

$$\begin{bmatrix} 0.2 & 0.3 & 0.5 \\ 0.25 & 0.6 & 0.15 \\ 0.15 & 0.45 & 0.4 \end{bmatrix} \tag{4}$$

The Markov chain can be visualised using the following picture (or "graph"):

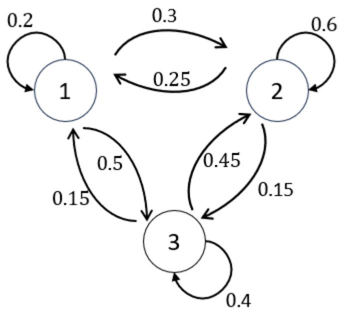

Figure 1: Visualisation of the Markov chain from (4)

In the Markov chain in (4) it is possible to reach any state $j = 1, 2, 3$ from any other state $i = 1, 2, 3$ (within one time step). This means that the Markov chain in (4) is irreducible.

Example 2. *Consider the following Markov chain:*

$$\begin{bmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 \\ 0 & 0.1 & 0.6 & 0.3 \\ 0 & 0 & 0.1 & 0.9 \end{bmatrix}. \tag{5}$$

In a graph, this Markov chain looks like:

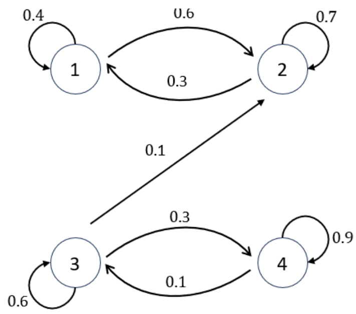

Figure 2: Visualisation of the Markov chain from (5)

Here, one can observe that it is impossible to reach (Markov) state 4 *if I start from either (Markov) state* 1 *or (Markov) state* 2*. This means that the Markov chain in (5) is reducible (not irreducible).*

In the above examples, it is relatively easy to draw the Markov chains, However, If the number of (Markov) states increases, it might not be easy to draw the Markov chain. Therefore, another way to check if a Markov chain is irreducible is to compute the following:

Proposition 1 (theorem 6.2.23 of Horn and Johnson (2013))**.** *Let* $P \in \mathbb{R}_+^{m \times m}$ *be a transition matrix of some Markov chain with m (Markov) states and let* I_m *be the* $m \times m$ *identity matrix. Then, the Markov chain is irreducible iff.* $(I_m + P)^{m-1}$ *only has positive (*> 0*) entries.*

For the full proof, see Horn and Johnson (2013), however, here, I want to give some intuition about why this proposition is correct. Consider the following scenario:
If the Markov chain is a loop (so $1 \to 2 \to ... \to m \to 1$) then this Markov chain is irreducible, since it takes m movements to get back to my starting (Markov) state and thus $P^m = I_m$ (for sure I have reached my starting state once I move m times). If I now compute P^{m+k} $(k > 0)$ I keep obtaining the identity matrix. So, just raising P to a large power and check if all elements are positive does not work in this specific Markov chain. If I compute $(I_m + P)^{m-1}$, then (sine P and I_m are square matrices) by the binomial of Newton, I get:

$$(I_m + P)^{m-1} = \sum_{k=0}^{m-1} \binom{m-1}{k} I_m^{(m-1)-k} P^k = \sum_{k=0}^{m-1} \binom{m-1}{k} P^k \tag{6}$$

notice, that $\binom{m-1}{k} > 0 \; \forall k = 0, ..., m-1$. So, the only way $(I_m + P)^{m-1} > 0$ can hold is if P^k has at some point positive entries for every combination of (i, j). Since then, the sum would give positive entries[3]. However, if P is irreducible, we can reach any state from any other state, so eventually, there should be some $k = 0, ..., m-1$ such that P^k results in positive entries at some specific combination of (i, j). And thus $(I_m + P)^{m+1} > 0 \forall (i, j)$ iff P (and thus the Markov chain) is irreducible.

This brings me to the third (important) assumption I need to make:

Assumption 3. *The Markov chain is irreducible*

[3]consider the matrices

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 5 \\ 2 & 3 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 0 & 2 & 3 \\ 5 & 0 & 0 \\ 0 & 1 & 4 \end{bmatrix}.$$

Both matrices have some zero elements (so both are non-negative), however, both matrices combined have for every combination of (i, j) at least 1 nonzero value. consequently, if I sum both matrices, I get:

$$A + B = \begin{bmatrix} 1 & 2 & 5 \\ 5 & 1 & 5 \\ 2 & 4 & 4 \end{bmatrix}$$

and this matrix has nonzero elements (is positive)

The recovery theorem (formulated by Ross (2015)) would like to investigate if it is possible to go from $\mathbb{Q}$ to $\mathbb{P}$. So, one might at first consider the Radon-Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$. The Radon-Nikodym derivative is used in asset pricing to transform from real-world probabilities to risk neutral probabilities, so, at first glance, this Radon-Nikodym derivative seems the solution to our problem. Notice, that since I use a finite sample space ($|\Omega| < \infty$), I can write the Radon-Nikodym derivative as the following:

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(\omega) = \frac{\mathbb{Q}[\{\omega\}]}{\mathbb{P}[\{\omega\}]} \tag{7}$$

So, let me compute both $\mathbb{Q}[\{\omega\}]$ and $\mathbb{P}[\{\omega\}]$.
Recall that the state of the world ω is a sequence of Markov states ($\omega = \{j_0, ..., j_\tau\}$). This means that we have:

$$\mathbb{P}[\{\omega\}] = \mathbb{P}[\{j_0, ..., j_\tau\}] = p_0(j_0) \cdot p(j_0, j_1) \cdot ... \cdot p(j_{\tau-1}, j_\tau) = p_0(j_0) \prod_{t=1}^{\tau} p(j_{t-1}, j_t) \tag{8}$$

where $p_0(j)$ is the starting distribution (the probability of starting in state j). Under the same reasoning, I find

$$\mathbb{Q}[\{\omega\}] = q_0(j_0) \prod_{t=1}^{\tau} q(j_{t-1}, j_t) \tag{9}$$

And so, the Radon-Nikodym derivative should equal:

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(\omega) = \frac{q_0(j_0)}{p_0(j_0)} \prod_{t=1}^{\tau} \frac{q(j_{t-1}, j_t)}{p(j_{t-1}, j_t)} \tag{10}$$

The problem with this Radon-Nikodm derivative is that I only know $q(i, j)$ (and $q_0(j)$) and thus I cannot use the Radon-Nikodym derivative to compute $\mathbb{P}$ (in general, I need to know both $d\mathbb{Q}$ and $\frac{d\mathbb{Q}}{d\mathbb{P}}$ in order to compute $d\mathbb{P}$).
A second guess might be to look at the *"stochastic discount factor"*, or *"pricing kernel"*. The first fundamental theorem of asset pricing (FFTAP) tells us that if absence of arbitrage holds, the price/value of an asset at time t (C_t) is equal to the expected discounted future value (C_{t+1}):

$$C_t = e^{-r} E_t^{\mathbb{Q}}[C_{t+1}] = E_t^{\mathbb{P}}\left[e^{-r}\frac{d\mathbb{Q}}{d\mathbb{P}} C_{t+1}\right] = E_t^{\mathbb{P}}[K_t C_{t+1}], \tag{11}$$

where $K_t := e^{-r}\frac{d\mathbb{Q}}{d\mathbb{P}}$ is the so called *"pricing kernel"* and r the risk-free rate. Note, that here, the risk-free rate (r) is constant, however, later on I will allow the risk-free rate to depend on the (starting) state (denoted as $r^{(i)}$) once I consider a multiperiod model. If one introduces the pricing kernel, then it is a small step towards Arrow-Debreu state prices, so let me first introduce this concept before I move on (since we need these state price later on).

Let me assume that we have one time period, so $t = \{0, 1\}$. Consider a Markov chain with m states. Let me assume that currently we are in state $i = 1, ..., m$ and let me define the price of an Arrow-Debreu security that pays off 1 (unit of currency) if state $j = 1, ..., m$ materializes as $c_j^{(i)}$. Then, assuming absence of arbitrage, the price of this security would be (using FFTAP):[4]

$$c_j^{(i)} = e^{-r} E^{\mathbb{Q}}\left[\mathbb{1}_{\{j_1=j\}} | j_0 = i\right] = e^{-r} \sum_{k=1}^{m} q(i, k) \mathbb{1}_{\{k=j\}} = e^{-r} q(i, j),\ \forall i, j = 1, ..., m \tag{12}$$

Notice, that using these Arrow-Debreu securities, we created a way to value the likelihood of a certain state materializing. Therefore, the price of an Arrow-Debreu securities is also called an

[4]note,

$$\mathbb{1}_{\{A\}} := \begin{cases} 1 \ \textit{in case of } A \\ 0 \ \textit{else} \end{cases}$$

(Arrow-Debreu) state price, as these securities give a way to price the likelihood of a state materializing ((Dybvig and Ross, 2003), (Adachi, 2021) and (Iwaki, Kijima, and Morimoto, 2001). Furthermore, since security $c_j^{(i)}$ pays off 1 if state $j = 1, ..., m$ materializes, we can construct some kind of insurance that always pays off 1, no matter which state materialises. If I construct a portfolio that holds one of each m Arrow-Debreu securities, then I can construct a portfolio that always pays off 1 unit of currency, no matter which state materializes. In other words, I can create a zero coupon bond (that pays off 1 at time $t = 1$) if I hold one of each of the securities, $c_1^{(i)}, ..., c_m^{(i)}$.
However, since absence of arbitrage holds, it must be true that the price/value of this portfolio of Arrow-Debreu securities is the same as the price of a zero coupon bond (law of one price). Let me define the price of the portfolio of m securities as $C^{(i)}$, then we get:

$$C^{(i)} := \sum_{j=1}^{m} c_j^{(i)} = \sum_{j=1}^{m} \left\{e^{-r} q(i,j)\right\} = e^{-r} \sum_{j=1}^{m} \{q(i,j)\} = e^{-r} \tag{13}$$

Indeed, we found that the price of this portfolio of Arrow-Debreu securities is equal to the price of a zero coupon bond (e^{-r}).
Notice, that if $r = 0$, then we have that $C^{(i)} = e^{-0} = 1$ and in this scenario, we have that $c_j^{(i)} \equiv q(i,j)$. Meaning that if I have the (state price) matrix

$$C := \begin{bmatrix} c_1^{(1)} & \dots & c_m^{(1)} \\ \vdots & \ddots & \vdots \\ c_1^{(m)} & \dots & c_m^{(m)} \end{bmatrix}, \tag{14}$$

this matrix becomes

$$C = \begin{bmatrix} q(1,1) & \dots & q(1,m) \\ \vdots & \ddots & \vdots \\ q(m,1) & \dots & q(m,m) \end{bmatrix} =: Q \tag{15}$$

And this matrix is a stochastic matrix (with row sum equal to 1).

Next, let me introduce some new notation. Let me define $c_j^{(i)}(t)$ as the state price of an Arrow-Debreu security that pays 1 if I reach, from the current state i, state j after t periods. Furthermore, to simplify the notation, let me define $c_j^{(i)}(1) \equiv c_j^{(i)}$ as the single period state price.
Let me assume that we have $t = \{0, ..., \tau\}$, $\tau > 1$ periods and, furthermore, let me assume that in every (starting) state we have a different interest rate, $r^{(i)}$. One can consider, for example, the scenario where in some (Markov) state returns turn out to be low, due to economic reasons and thus the central bank might decide to decrease interest rates. Then, if we move from this *"bad"* (Markov) state to an (economically speaking) *"better"* state, the central bank might respond by increasing their interest rates. So, to make the model a bit more realistic, I introduce this kind of dynamic interest rate.

Let me start with $\tau = 2$ (so $t = \{0, 1, 2\}$), then, one can view this problem as follows: At time $t = 0$, I am in state $j_0 = 1, ..., m$. The next period ($t = 1$), I move from this state $j_0 = 1, ..., m$ to some intermediate state j_1. Finally, at time $t = 2$, I move from this intermediary state j_1 to the desired (ending) state $j_2 = 1, ..., m$. How would the state price $c_j^{(i)}(2)$ look like in this specific scenario?
We can consider the following strategy: compute the value of this 1 unit of currency in the intermediary state and then use the intermediary values to compute the expected discounted intermediary values within state j_0 [5]. However, this intermediary value is nothing more than the state

[5] This strategy works similarly as how in asset pricing the price of an American options in a binomial tree is determined (see, for example,Cox, Ross, and Rubinstein (1979))

price that pays 1 in state j_2 starting in state j_1. So, I compute (using FFTAP)

$$\begin{aligned}
c_j^{(i)}(2) &= e^{-r^{(i)}} E_t^{\mathbb{Q}} \left[e^{r^{(j_{t+1})}} \mathbb{1}_{\{j_{t+2}=j\}} \right] \\
&= e^{-r^{(i)}} E_t^{\mathbb{Q}} \left[e^{-r^{(j_{t+1})}} E_{t+1}^{\mathbb{Q}} \left[\mathbb{1}_{\{j_{t+2}=j\}} \right] \right] \\
&= e^{-r^{(i)}} E_t^{\mathbb{Q}} \left[c_j^{(j_{t+1})} \right] \\
&= e^{-r^{(i)}} \sum_{k=1}^{m} q(i,k) c_j^{(k)} = \sum_{k=1}^{m} e^{-r^{(i)}} q(i,k) c_j^{(k)} \\
&= \sum_{k=1}^{m} c_k^{(i)} c_j^{(k)}
\end{aligned} \tag{16}$$

If I now use that

$$C := \begin{bmatrix} c_1^{(1)} & \dots & c_m^{(1)} \\ \vdots & \ddots & \vdots \\ c_1^{(m)} & \dots & c_m^{(m)} \end{bmatrix}, \tag{17}$$

(16) becomes:

$$c_j^{(i)}(2) = \begin{bmatrix} c_1^{(i)} & \dots & c_m^{(i)} \end{bmatrix} \cdot \begin{bmatrix} c_j^{(1)} \\ \vdots \\ c_j^{(m)} \end{bmatrix} =: C_{i,*} C_{*,j}. \tag{18}$$

Notice, that If I want to compute for every $j = 1, ..., m$ the two-period state price, I compute

$$\begin{bmatrix} c_1^{(i)}(2) & \dots & c_m^{(i)}(2) \end{bmatrix} = C_{i,*} C \tag{19}$$

And so, if I want to compute the (row) sum, I find:

$$\begin{aligned}
&= \begin{bmatrix} c_1^{(i)}(2) & \dots & c_m^{(i)}(2) \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = C_{i,*} C \iota = C_{i,*} \begin{bmatrix} e^{-r^{(1)}} \\ \vdots \\ e^{-r^{(m)}} \end{bmatrix} = e^{-r^{(i)}} \sum_{j=1}^{m} q(i,j) e^{-r^{(j)}} . \\
&= e^{-r^{(i)}} E^{\mathbb{Q}} \left[e^{-r^{(j_1)}} | j_0 = i \right] = e^{-r^{(i)}} E^{\mathbb{Q}} \left[e^{-r^{(j_1)}} E^{\mathbb{Q}} \left[1 | j_1 \right] \mid j_0 = i \right] = P(2)
\end{aligned} \tag{20}$$

From (20) it becomes clear that the row sum $\sum_{j=1}^{m} c_j^{(i)}(2) = P(2)$ is the two period discount rate (price of a zero coupon bond that pays off 1 after two periods).
Let me now turn towards the scenario in which $\tau > 2$. In this scenario, the value $c_j^{(i)}(\tau)$ indicates the state price of an Arrow-Debreu security that pays 1 unit of currency in case I reach state $j = 1, ..., m$ from state $i = 1, ..., m$, after $\tau > 2$ periods. We can consider this scenario as follows: At time $t = 0$, I am in state $i = 1, ..., m$. In the next period ($t = 1$), I move from this starting state to a new state, say $j_1 = 1, ..., m$. Then, in the following period ($t = 2$) I move from state j_1 to a new state, say $j_2 = 1, ..., m$ and this continues until I reach the penultimate period ($t = \tau - 1$), because at time $t = \tau$, I must move from this penultimate state ($j_{\tau-1}$) to the ending state $j = 1, ..., m$.
Notice, that in the above mentioned scenario, the only states that are important are the starting and ending state (i and j). I do not care which states I visit during $t = \{1, ..., \tau - 1\}$, only that at time $t = 0$, I am in state i and at time $t = \tau$ I am in state j. So, we get the following notation for the state price $c_j^{(i)}(\tau)$:

Proposition 2. *Let*

$$C := \begin{bmatrix} c_1^{(1)} & \dots & c_m^{(1)} \\ \vdots & \ddots & \vdots \\ c_1^{(m)} & \dots & c_m^{(m)} \end{bmatrix} \tag{21}$$

with (see (12))

$$c_j^{(i)} := e^{-r^{(i)}} q(i,j),\ \forall i,j = 1,...,m \tag{22}$$

be the state price matrix of the Markov chain, such that $c_j^{(i)}$ *represents the state price of an Arrow-Debreu security that pays 1 unit of currency if, in the next period* ($t = 1$)*, we reach state* $j = 1,...,m$*, starting from (current) state* $i = 1,...,m$ *(so the single period state price). Then,*

$$C^\tau, \tag{23}$$

represents the state price matrix of Arrow-Debreu securities that pay 1 unit of currency if after τ *periods, I reach state* j *(column* j*) from state* i *(row* i*).*

Proof. any row of C looks like:

$$C_{i,*} = \begin{bmatrix} c_1^{(i)} & \dots & c_m^{(i)} \end{bmatrix}$$

and any column of C looks like:

$$C_{*,j} = \begin{bmatrix} c_j^{(1)} \\ \vdots \\ c_j^{(m)} \end{bmatrix}$$

This means, that if I multiply any row of C with any column of C (which I do in case of C^2), I get:

$$C_{i,*}C_{*,j} := \begin{bmatrix} c_1^{(i)} & \dots & c_m^{(i)} \end{bmatrix} \cdot \begin{bmatrix} c_j^{(1)} \\ \vdots \\ c_j^{(m)} \end{bmatrix} = \sum_{k=1}^{m} c_k^{(i)} c_j^{(k)} =: c_j^{(i)}(2) = \text{"state price at } \tau = 2\text{"} \tag{24}$$

Now, consider $C^3 = C \cdot C^2$. Any column of C^2 looks like

$$\begin{bmatrix} c_j^{(1)}(2) \\ \vdots \\ c_j^{(m)}(2) \end{bmatrix}$$

So, we have that

$$\begin{aligned} \begin{bmatrix} c_1^{(i)} & \dots & c_m^{(i)} \end{bmatrix} \cdot \begin{bmatrix} c_j^{(1)}(2) \\ \vdots \\ c_j^{(m)}(2) \end{bmatrix} &= \sum_{k=1}^{m} c_k^{(i)} c_j^{(k)}(2) := \sum_{k=1}^{m} e^{-r^{(i)}} q(i,k) c_j^{(k)}(2) \\ &= e^{-r^{(i)}} E_t^{\mathbb{Q}} \left[c_j^{(j_{t+1}=k)}(2) \right] = c_j^{(i)}(3) \end{aligned}$$

since, i and j are arbitrary rows and columns, this must hold for every element of C^3. Let me now assume that for $t = \tau - 1$ the proposition holds (proof by induction). Does it then hold for $t = \tau$? we can write $C^\tau = C \cdot C^{\tau-1}$. Any column of $C^{\tau-1}$ looks like:

$$C_{*,j}^{\tau-1} = \begin{bmatrix} c_j^{(1)}(\tau-1) \\ \vdots \\ c_j^{(m)}(\tau-1) \end{bmatrix}.$$

This means, that for any row i and column j of C^τ, it must hold that:

$$\begin{aligned} C_{i,j}^\tau = C_{i,*} \cdot C_{*,j}^{\tau-1} = \begin{bmatrix} c_1^{(i)} & \dots & c_m^{(i)} \end{bmatrix} \cdot \begin{bmatrix} c_j^{(1)}(\tau-1) \\ \vdots \\ c_j^{(m)}(\tau-1) \end{bmatrix} &= \sum_{k=1}^{m} c_k^{(i)} c_j^{(k)}(\tau-1) \\ := \sum_{k=1}^{m} e^{-r^{(i)}} q(i,k) c_j^{(k)}(\tau-1) = e^{-r^{(i)}} E_t^{\mathbb{Q}} \left[c_j^{(j_{t+1}=k)}(\tau-1) \right] &= c_j^{(i)}(\tau) \end{aligned} \tag{25}$$

Since i and j are chosen arbitrarily, it must hold that C^τ is equal to the state price of an Arrow-Debreu security that pays off 1 unit of currency if I reach state j, starting from state i, after τ periods. Hence, Proposition 2 holds for any $\tau \geq 1$ □

This brings me to the final notation of the state prices:

Definition (Multiperiod state prices)**.** *Let $t = \{0, ..., \tau\}$ be my time horizon (with $\tau \geq 1$) and let me denote the probability of moving from (Markov) state $i = 1, ..., m$ to (Markov) state $j = 1, ..., m$ under the risk neutral probability measure as $q(i, j)$. Then, the (state) price of an Arrow-Debreu security that pays off 1 unit of currency if I reach state $j = 1, ..., m$ from state $i = 1, ..., m$ after τ periods, is given by:*

$$c_j^{(i)}(\tau) = \begin{cases} e^{-r^{(i)}} q(i,j) \text{ if } \tau = 1 \\ C_{i,*} C^{\tau-2} C_{*,j} \text{ if } \tau \geq 2 \end{cases} \tag{26}$$

where

$$C := \begin{bmatrix} c_1^{(1)}(1) & \dots & c_m^{(1)}(1) \\ \vdots & \ddots & \vdots \\ c_1^{(m)}(1) & \dots & c_m^{(m)}(1) \end{bmatrix},$$

$C_{i,*}$ corresponds to the i-th row of matrix C, $C_{*,j}$ corresponds to the j-th column of matrix C and $r^{(i)}$ is the risk-free rate in (Markov) state i.

Recall, that state price matrix is given by

$$C = \begin{bmatrix} c_1^{(1)} & \dots & c_m^{(1)} \\ \vdots & \ddots & \vdots \\ c_1^{(m)} & \dots & c_m^{(m)} \end{bmatrix} = \begin{bmatrix} e^{-r^{(1)}} q(1,1) & \dots & e^{-r^{(1)}} q(1,m) \\ \vdots & \ddots & \vdots \\ e^{-r^{(m)}} q(m,1) & \dots & e^{-r^{(m)}} q(m,m) \end{bmatrix}. \tag{27}$$

This means that if I want to compute the pricing kernel, I get:

$$e^{-r^{(i)}} \frac{q(i,j)}{p(i,j)} = \frac{e^{-r^{(i)}} q(i,j)}{p(i,j)} = \frac{C_{i,j}}{p(i,j)}. \tag{28}$$

So, the pricing kernel can be written in terms of the state price divided by the real-world probability.
Working directly with this pricing kernel, does not work (for the same reason why working with the Radon-Nikodym derivative did not work). However, there is another formulation of the pricing kernel that might be useful. This formulation uses utility of consumption to derive the pricing kernel. Consider the following example:

Example 3. *Consider a market in which M assets are traded. I now introduce an agent who needs to choose an investment strategy, $\xi := [\xi_1, ..., \xi_M]^T$, where ξ_k is the fraction of wealth invested in asset $k = 1, ..., M$. Furthermore, let me assume that currently we live in (Markov) state i and the next (Markov) state j is random with some transition probability $p(i, j)$. The agent can decide to either invest (parts of) its wealth in the assets $X := [X_1, ..., X_M]^T$ with price $X(i) := [X_1(i), ..., X_M(i)]^T$ or consume $c(i)$. The agent knows that in the next (Markov) state the assets have some return $X(j)$, but at the decision time, he does not know this value, but only the distribution of the future price. Consequently, this agent aims to solve*

$$\begin{aligned} \max_{\xi} \; & \{u(c(i)) + E_t[\beta \cdot u(c(j))]\} \\ s.t. \; & c(i) = w(i) - \xi^T X(i) = w(i) - \sum_k X_k(i)\xi_k \\ & c(j) = w(j) + \xi^T X(j) = w(j) + \sum_k X_k(j)\xi_k \end{aligned} \tag{29}$$

one can now compute the first order condition with respect to ξ_k to find

$$X_k(i)u'(c(i)) = E_t[\beta X_k(j)u'(c(j))] \iff X_k(i) = E_t\left[\beta \frac{u'(c(j))}{u'(c(i))} X_k(j)\right]$$

Using the First Fundamental Theorem of Asset Pricing, one can easily see that a pricing kernel should equal $K_t^{(k)} = \beta \frac{u'(c(j))}{u'(c(i))}$. Interestingly, this formulation of the pricing kernel shows that the pricing kernel only depends on the marginal rate of substitution between the consumption in the current (Markov) state $(c(i))$ and the consumption in the future (Markov) state $(c(j))$.

This example introduces a new way of writing the pricing kernel, namely the notion of a "transition independent kernel".

Definition (Transition Independent Kernel)**.** *A (pricing) kernel is said to be transition independent if there is a positive function of the (Markov) states, $h : \{1, ..., m\} \to \mathbb{R}^+$, and a positive constant β such that, for any transition from i to j, the kernel has the form*

$$e^{-r^{(i)}} \frac{q(i,j)}{p(i,j)} := \frac{C_{i,j}}{p(i,j)} = \beta \frac{h(j)}{h(i)} \tag{30}$$

Two remarks about the transition independent kernel:

i) Note, that when I am in the next state (j) I will not invest further, but I fully consume my remaining wealth. This follows from the second budget constraint $(c(j) = w(j) + \sum_k X_k(j)\xi_k)$. This means that when I make the investment decision in state i my underlying objective is that in the next/final (Markov) state I fully consume everything that I have. In other words, the agent does not have a bequest motive (he does not want to maximize his terminal wealth). In Appendix A one can find an example that shows that if an agent has some kind of bequest motive, the kernel will no longer be transition independent.

ii) Furthermore, notice that consumption in each state is determined by the (Markov) state itself $(c(j) \leftrightarrow h(j)$ and $h : \{1, ..., m\} \to \mathbb{R}^+)$ and every (Markov) state contains all the necessary information. However, every (Markov) state represents some return and consequently, the agent's consumption is (fully) determined by the return of the asset and not the value (or wealth) of the agent.

Notice, however, that by writing the kernel based on an agent's optimization, we need to make the assumption that every agent agrees upon the distribution of possible state outcomes. We will estimate the real-world probability measure $\mathbb{P}$ using the risk neutral measure $\mathbb{Q}$, that is calibrated by prices and these prices are found by having a market equilibrium. In other words, the prices that we use to calibrate the risk neutral measure $\mathbb{Q}$ are based on the subjective beliefs of agents. Consequently, If I do not impose any assumption regarding this subjective belief, I am no longer estimating the real-world probability measure $(\mathbb{P})$, but the subjective belief of the agents $(\mathcal{P}_n)$. Therefore, I need to make the assumption:

Assumption 4. *All agents within the market obey strong rational expectations. That is, assume that there are N agents acting on the market and that $\mathcal{P}_n$ $(n = 1, ..., N)$ is the subjective belief of agent $n = 1, ..., N$ and $\mathbb{P}$ the real-world probability measure, then it must hold that $\mathcal{P}_n[A] = \mathbb{P}[A]\, \forall n = 1, ..., N,\ \forall A \in \mathcal{F}_t$.*

By making Assumption 4, I guarantee that once I obtain estimates for the probabilities that the agent uses to find his optimum, I obtain at the same time estimates for the real-world probability measure $\mathbb{P}$. This brings me to the following assumption:

Assumption 5. *The pricing kernel is transition independent. That is, I can write the pricing kernel as in (30).*

With Assumption 5, I get:

$$\frac{C_{i,j}}{p(i,j)} = \beta \frac{h(j)}{h(i)} \iff h(i)C_{i,j} = \beta h(j)p(i,j) \tag{31}$$

using matrix notation this becomes:

$$HC = \beta PH \tag{32}$$

with

$$H = \begin{bmatrix} h(1) & & 0 \\ & \ddots & \\ 0 & & h(m) \end{bmatrix},$$

P the transition matrix under $\mathbb{P}$ and C the state price matrix.

The problem with (32) is that I still need to know $\frac{h(j)}{h(i)}$ in order to compute $p(i,j)$. So, let me use the fact that P is a stochastic (transition) matrix. That is, I know that

$$P\iota := P \cdot \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} =: \iota \tag{33}$$

Since I have:

$$HC = \beta PH \iff P = \frac{1}{\beta} HCH^{-1} \tag{34}$$

I can write:

$$P\iota = \frac{1}{\beta} HCH^{-1}\iota = \iota \tag{35}$$

$$\iff CH^{-1}\iota = \beta H^{-1}\iota \tag{36}$$

$$\iff C\nu = \beta\nu \tag{37}$$

In (37), one can recognise that ν is the eigenvector of C, for eigenvalue β. The question that needs to be answered is whether there exist such an eigenvalue β for which the corresponding eigenvector only has positive entries. The reason why this is important is because ν depends on $h(j)$ and (according to the transition independent kernel) $h(j) > 0\, \forall j = 1, ..., m$.
I can use the Perron-Frobenius theorem to find an eigenvalue with positive eigenvector entries.

Theorem 2 (Perron-Frobenius)**.** *If a matrix $A \in \mathbb{R}^{n\times n}$ is non-negative and irreducible, there is a (unique) non-negative eigenvector w such that $Aw = \rho(A)w$, where $\rho(A) := max\{\ |\lambda|\ :\ \lambda \text{ is eigenvalue of } A\} \in \mathbb{R}$ is the spectral radius or "Perron root" of A.*

recall Assumption 1 and Assumption 3. Since

$$C = \begin{bmatrix} e^{-r^{(1)}}q(1,1) & \dots & e^{-r^{(1)}}q(1,m) \\ \vdots & \ddots & \vdots \\ e^{-r^{(m)}}q(m,1) & \dots & e^{-r^{(m)}}q(m,m) \end{bmatrix}, \tag{38}$$

and $e^{-r^{(i)}} > 0$, it must be that by definition C is non-negative, since for any probability it holds that $q(i,j) \in [0,1]$. Furthermore, by assuming that $\mathbb{P}$ is characterised by a Markov chain (Assumption 1), $\mathbb{Q}$ will depend on the same Markov chain and this Markov chain is irreducible (Assumption 3). Consequently, it is sufficient to conclude that C is an irreducible matrix, because Q is irreducible and multiplying Q with a (positive) scalar does not affect this property. Hence, by Assumption 1 and Assumption 3 I can conclude that C is a non-negative irreducible matrix and thus I can apply the Perron-Frobenius theorem to matrix C. Finally, if I choose the eigenvalue $\beta = \rho(C)$ and the corresponding eigenvector $\nu = w$, then by using the Perron-Frobenius theorem, I know that (37) has a solution.

Once I have found $\nu = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}$, I can compute P by using that

$$(H)^{-1} = \begin{bmatrix} \frac{1}{h(1)} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{h(m)} \end{bmatrix} \tag{39}$$

And thus,

$$(H)^{-1}\iota = \begin{bmatrix} \frac{1}{h(1)} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{h(m)} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{h(1)} \\ \vdots \\ \frac{1}{h(m)} \end{bmatrix} \tag{40}$$

Consequently, we know that

$$\nu = (H)^{-1}\iota \iff \frac{1}{h(i)} = \nu_i \iff h(i) = \frac{1}{\nu_i} \tag{41}$$

However, recall that we had (see (31)):

$$h(i)C_{i,j} = \beta p(i,j)h(j) \iff p(i,j) = C_{i,j}\frac{1}{\beta}\frac{h(i)}{h(j)}$$

which can be rewritten as:

$$p(i,j) = C_{i,j}\frac{1}{\beta}\frac{h(i)}{h(j)} \tag{42}$$

$$\iff p(i,j) = C_{i,j}\frac{1}{\beta}\frac{\frac{1}{\nu_i}}{\frac{1}{\nu_j}} \tag{43}$$

$$\iff p(i,j) = C_{i,j}\frac{1}{\beta}\frac{\nu_j}{\nu_i} \tag{44}$$

Hence, To compute $\hat{p}(i,j)$, one can apply the following algorithm:

1) construct C
2) find eigenvalues of C and their corresponding eigenvectors (ν)
3) set $\beta = \rho(C) = max\{|\lambda| : \lambda \textit{ is eigenvalue of } C\}$
4) compute $\hat{p}(i,j)$ using $\hat{p}(i,j) = C_{i,j}\frac{1}{\beta}\frac{\nu_j}{\nu_i}$

This specific method to obtain real-world probabilities was introduced by Ross (2015) and led to the recovery theorem (Theorem 1)

Example of the Recovery Theorem

In this chapter, I will demonstrate the example that was done by Ross (2015). This example uses a Data Generating Process (DGP) to generate the real-world probabilities. Then, it applies the recovery theorem and checks, using the generated real-world probabilities, how accurate the estimated probabilities estimate the (generated) real-world probabilities.

Example 4 (Ross (2015))**.** *Assume that we are in a Black-Scholes market. In this market we have (one) risky asset with dynamics:*

$$dS_t = S_t(\mu dt + \sigma dW_t^{\mathbb{P}}) = S_t(rdt + \sigma dW_t^{\mathbb{Q}}) \tag{45}$$

where $W_t^{\mathbb{P}}$ is a Brownian motion with no drift under $\mathbb{P}$ and $W_t^{\mathbb{Q}}$ is a Brownian motion with no drift under $\mathbb{Q}$. Furthermore, μ is the return of the risky asset, σ is the volatility of the asset and r is the risk-free rate. We limit the time frame to one period such that we have $t \in [0, 1]$.
Next, I introduce some agent that has CRRA utility function with risk aversion $\gamma = 3$ and subjective discount factor $\beta = e^{-\delta} = e^{-0.02} \approx 0.9802$. Since the CRRA utility function has the form

$$u(x) = \begin{cases} ln(x) \text{ if } \gamma = 1 \\ \dfrac{x^{1-\gamma} - 1}{1-\gamma} \text{ if } \gamma \neq 1 \end{cases} \tag{46}$$

the transition independent kernel assumption implies that

$$\frac{C_{i,j}}{p(i,j)} = \beta \frac{u'(c(j))}{u'(c(i))} \iff C_{i,j} = \beta \frac{u'(c(j))}{u'(c(i))} \cdot p(i,j) \tag{47}$$

The only thing that remains, in order to compute the state prices ($C_{i,j}$), is determining what the agent's consumption will be. Since we only have one period, Ross (2015) assumes that in the final period ($t = 1$) the agent fully consumes the value of the asset and in the first period ($t = 0$) he consumes some hypothetical $t = 0$ return of the asset that was realised by making a hypothetical investment at time $t = -1$ (Let me call this return S_H)[6]*. If I fill in everything that we know, we get as state prices:*

$$C_{S_H,S_1} = e^{-0.02} \left(\frac{S_1}{S_H}\right)^{-\gamma} \cdot p(S_H, S_1) \tag{48}$$

and using the dynamics of S_t I know that (normalize $S_0 = 1$)

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma\sqrt{t}W_t^{\mathbb{P}}} = e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma\sqrt{t}W_t^{\mathbb{P}}} \tag{49}$$

$$\iff ln(S_t) = (\mu - \frac{1}{2}\sigma^2)t + \sigma\sqrt{t}W_t^{\mathbb{P}} \tag{50}$$

[6]The reason why I formulate the $t = 0$ consumption in this way is because we are considering an irreducible Markov chain. If I reach (in the future) the same state as the current state, I still want to consume the same way as I did at the start. Hence, to guarantee consistency I need to introduce this hypothetical $t = 0$ return.

since $W_t^{\mathbb{P}} \sim N(0,1)$ *(due to the one period), I have that* $ln(S_t) \sim N((\mu - \frac{1}{2}\sigma^2)t, \sigma^2 t)$*. This means that I have that (using the pdf of a log-normal distribution)*

$$p(S_H, S_1) = \frac{1}{S_1 \sigma \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{ln(S_1)-(\mu-\frac{1}{2}\sigma^2)}{\sigma}\right)^2} = \frac{1}{S_1}\phi\left(\frac{ln(S_1)-(\mu-\frac{1}{2}\sigma^2)}{\sigma}\right) \tag{51}$$

Where $\phi(.)$ *is the pdf of a standard normal distribution.*
Using everything until now, gives:

$$C_{S_H,S_1} = e^{-0.02}\left(\frac{S_1}{S_H}\right)^{-3} \cdot \frac{1}{S_1}\phi\left(\frac{ln(S_1)-(\mu-\frac{1}{2}\sigma^2)}{\sigma}\right) \tag{52}$$

Even though I am able to compute the state prices using (52), I still need to do one more step. In the (52), there is still a continuous density function $(\phi(.))$*, which I need to discretise. Recall that I have a Markov chain and a Markov chain is in discrete time (not continuous). To discretise my continuous density function, I need to introduce the concept of "sigma distances (from the mean)":*

Intermezzo (Sigma distances)**.** *Consider a normally distributed random variable* X *(so* $X \sim N(\mu, \sigma^2)$*). A property of the normal distribution is that it is symmetric around the mean* (μ) *and that*

$$\begin{cases} \mathbb{P}[\mu - \sigma \leq X \leq \mu + \sigma] \approx 68\% \\ \mathbb{P}[\mu - 2\sigma \leq X \leq \mu + 2\sigma] \approx 95\% \\ \mathbb{P}[\mu - 3\sigma \leq X \leq \mu + 3\sigma] \approx 99.7\% \end{cases} \tag{53}$$

Consider a new random variable, say Y*. Then I will call the values* $Z_1^{(1)}$ *and* $Z_2^{(1)}$ *a sigma distance of 1 away from the mean if it holds that* $\mathbb{P}[Z_1^{(1)} \leq Y \leq Z_2^{(1)}] \approx 68\%$*. In the same way, the values* $Z_1^{(3)}$ *and* $Z_2^{(3)}$ *are a sigma distance of 3 away from the mean if* $\mathbb{P}[Z_1^{(3)} \leq Y \leq Z_2^{(3)}] \approx 99.7\%$*. Hence,*

Definition. *Let* X *and* Y *be two random variables and assume that* $X \sim N(\mu, \sigma^2)$*. then,* $Z_1^{(k)}$ *and* $Z_2^{(k)}$ *are a sigma distance of* k *away from the mean* $\iff \mathbb{P}[Z_1^{(k)} \leq Y \leq Z_2^{(k)}] = \mathbb{P}[\mu - k\sigma \leq X \leq \mu + k\sigma]$.
If I define the CDF of a standard normal distribution as $\Phi(.)$*, then I can rephrase this definition as:*
$Z_1^{(k)}$ *and* $Z_2^{(k)}$ *are a sigma distance of* k *away from the mean* $\iff \mathbb{P}[Z_1^{(k)} \leq Y \leq Z_2^{(k)}] = \Phi(k) - \Phi(-k)$.

I will assume that there are 17 possible outcomes for the risky asset, that are based on sigma distances from the expected return of the asset. Stated differently, I assume that the entire support of the risky asset can be approximated by sigma distances from the mean. I will use values of $k = \{0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4\}$ *such that I get:*

$$S = \left\{Z_1^{(k)}, Z_2^{(k)} : \mathbb{P}\left[Z_1^{(k)} \leq S_1 \leq Z_2^{(k)}\right] = \Phi(k) - \Phi(k), k = \{0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4\}\right\} \tag{54}$$

If I use (just as Ross (2015)) $\mu = 8\%$ *and* $\sigma = 20\%$*. Then I get the following sigma distances:* [7]

Table 1: sigma distances for different k

k	0	0.5	1	1.5	2	2.5	3	3.5	4
$Z_1^{(k)}$	1.062	0.961	0.869	0.787	0.712	0.644	0.583	0.527	0.477
$Z_2^{(k)}$	1.062	1.174	1.297	1.433	1.584	1.751	1.935	2.138	2.363

[7]Note, that the reason why I have slightly different sigma distances is because of the following: *"While sigma can be chosen as the standard deviation of the derived martingale measure from P, we chose the current at-the-money implied volatility from option prices on the S&P 500 index as of March 15, 2011"* (Ross, 2015). In other words, I use the "standard deviation of the derived martingale measure", whereas Ross uses the "implied volatility from option prices". This discrepancy, however, does not affect the clarifying objective of this example.

To see how accurate this approximation is, I will plot the non-approximated and the approximated pdf's of the log-normal distribution :

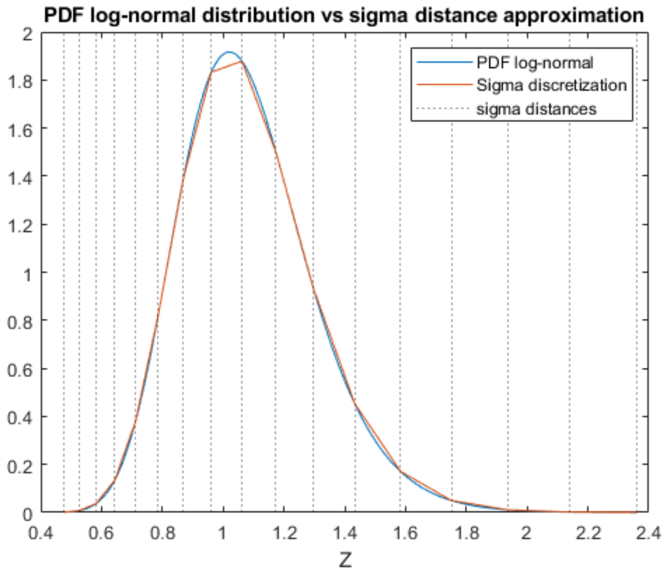

Figure 3: pdf's of both the non-approximated and the approximated log-normal distribution

From Figure 3 it becomes clear that the sigma distance approximation estimates reasonably accurate the continuous log-normal pdf.
Once, I know the sigma distances, I can compute the discretised real-world transition matrix (P) *and the state price matrix* (C)*. Doing so, yields the following two* 17×17 *tables:*[8]

Table 2: real-world transition matrix P

		S_1																
	$p(S_H, S_1)$	0.477	0.527	0.583	0.644	0.712	0.787	0.869	0.961	1.062	1.174	1.297	1.433	1.584	1.751	1.935	2.138	2.363
S_H	0.477	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	0.527	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	0.583	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	0.644	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	0.712	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	0.787	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	0.869	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	0.961	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	1.062	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	1.174	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	1.297	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	1.433	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	1.584	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	1.751	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	1.935	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	2.138	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	2.363	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.156	0.097	0.047	0.018	0.005	0.001	0	0

[8] note, that I rounded all probabilities to 3 decimals, which causes certain elements of the matrix to be 0.

Table 3: State price matrix C

	C_{S_H,S_1}	S_1 0.477	0.527	0.583	0.644	0.712	0.787	0.869	0.961	1.062	1.174	1.297	1.433	1.584	1.751	1.935	2.138	2.363
S_H	0.477	0	0.001	0.002	0.006	0.012	0.019	0.024	0.023	0.017	0.010	0.005	0.002	0	0	0	0	0
	0.527	0	0.001	0.003	0.008	0.016	0.025	0.032	0.031	0.023	0.014	0.006	0.002	0.001	0	0	0	0
	0.583	0	0.001	0.004	0.010	0.022	0.034	0.043	0.042	0.032	0.019	0.009	0.003	0.001	0	0	0	0
	0.644	0	0.002	0.005	0.014	0.029	0.046	0.058	0.056	0.043	0.025	0.012	0.004	0.001	0	0	0	0
	0.712	0	0.002	0.007	0.019	0.039	0.062	0.078	0.076	0.058	0.034	0.016	0.006	0.002	0	0	0	0
	0.787	0	0.003	0.010	0.025	0.053	0.084	0.106	0.103	0.078	0.046	0.021	0.008	0.002	0	0	0	0
	0.869	0	0.004	0.013	0.034	0.071	0.113	0.142	0.138	0.105	0.062	0.029	0.010	0.003	0.001	0	0	0
	0.961	0	0.006	0.018	0.046	0.096	0.153	0.192	0.187	0.142	0.084	0.039	0.014	0.004	0.001	0	0	0
	1.062	0	0.008	0.024	0.062	0.130	0.207	0.259	0.253	0.192	0.113	0.052	0.019	0.005	0.001	0	0	0
	1.174	0	0.011	0.032	0.083	0.176	0.280	0.350	0.341	0.260	0.153	0.071	0.025	0.007	0.001	0	0	0
	1.297	0	0.015	0.043	0.112	0.237	0.377	0.473	0.460	0.350	0.206	0.095	0.034	0.010	0.002	0	0	0
	1.433	0	0.020	0.058	0.151	0.320	0.509	0.637	0.621	0.472	0.278	0.128	0.046	0.013	0.003	0	0	0
	1.584	0	0.027	0.079	0.204	0.432	0.687	0.861	0.838	0.637	0.376	0.173	0.062	0.018	0.004	0.001	0	0
	1.751	0	0.036	0.106	0.276	0.583	0.928	1.163	1.132	0.861	0.507	0.234	0.084	0.024	0.005	0.001	0	0
	1.935	0	0.049	0.143	0.372	0.787	1.253	1.569	1.528	1.162	0.685	0.316	0.113	0.032	0.007	0.001	0	0
	2.138	0	0.065	0.193	0.502	1.062	1.690	2.117	2.062	1.568	0.924	0.426	0.153	0.043	0.009	0.001	0	0
	2.363	0	0.088	0.261	0.678	1.433	2.282	2.858	2.783	2.116	1.247	0.575	0.207	0.059	0.012	0.002	0	0

At first glance, Table 2 might seem strange. However, recall that every (Markov) state represents some return. And in the Black-Scholes worlds, returns are (iid) log-normally distributed. In other words, the likelihood of going from a high value to a low value is not large, however, the likelihood of a certain return in a low state or a high state is just the same.

Now that I have obtained the state price matrix (C), I will apply the second step and the third step of the algorithm: find the eigenvalues of C and compute the spectral radius of C ($\rho(C)$). It turns out that I get $\rho(C) = 0.980 \approx e^{-0.02}$. Finally, I will compute the estimated real-world transition matrix using (44):

$$\hat{p}(i,j) = C_{i,j}\frac{1}{\beta}\frac{\nu_j}{\nu_i}$$

This gives the matrix:

Table 4: Estimated transition matrix $\hat{P}$

	$\hat{p}(S_H,S_1)$	S_1 0.477	0.527	0.583	0.644	0.712	0.787	0.869	0.961	1.062	1.174	1.297	1.433	1.584	1.751	1.935	2.138	2.363
S_H	0.477	0	0.001	0.004	0.014	0.039	0.087	0.146	0.193	0.195	0.160	0.097	0.044	0.020	0	0	0	0
	0.527	0	0.001	0.004	0.015	0.040	0.088	0.147	0.192	0.198	0.161	0.096	0.043	0.015	0	0	0	0
	0.583	0	0.001	0.004	0.014	0.040	0.086	0.144	0.190	0.196	0.157	0.100	0.047	0.021	0	0	0	0
	0.644	0	0.001	0.004	0.014	0.040	0.087	0.147	0.193	0.197	0.157	0.097	0.047	0.016	0	0	0	0
	0.712	0	0.001	0.004	0.014	0.040	0.086	0.145	0.190	0.195	0.157	0.096	0.048	0.017	0.008	0	0	0
	0.787	0	0.001	0.004	0.014	0.040	0.086	0.145	0.190	0.195	0.156	0.098	0.048	0.017	0.006	0	0	0
	0.869	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.195	0.156	0.097	0.048	0.019	0.004	0	0	0
	0.961	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.195	0.157	0.097	0.048	0.017	0.006	0	0	0
	1.062	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.196	0.157	0.097	0.047	0.018	0.005	0	0	0
	1.174	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.194	0.156	0.097	0.047	0.018	0.005	0.002	0	0
	1.297	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.195	0.156	0.097	0.047	0.017	0.005	0.002	0	0
	1.433	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.195	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	1.584	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.195	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	1.751	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.195	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	1.935	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.195	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	2.138	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.195	0.156	0.097	0.047	0.018	0.005	0.001	0	0
	2.363	0	0.001	0.004	0.014	0.040	0.086	0.145	0.191	0.195	0.156	0.097	0.047	0.018	0.005	0.001	0	0

To check how accurately I estimated the real-world transition matrix, I computed the squared error between the estimated real-world transition matrix ($\hat{P}$) and the (true) real-world probabilities that follow from the dynamics of S_t (P). It turns out that when I compute:

$$\mathcal{E} := \begin{bmatrix} (p(1,1)-\hat{p}(1,1))^2 & \dots & (p(1,m)-\hat{p}(1,m))^2 \\ \vdots & \ddots & \vdots \\ (p(m,1)-\hat{p}(m,1))^2 & \dots & (p(m,m)-\hat{p}(m,m))^2 \end{bmatrix} \tag{55}$$

I get (rounded to 4 decimals)

$$\mathcal{E} = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} \tag{56}$$

This shows that The recovery theorem accurately estimates the real-world probabilities (once all necessary assumptions are satisfied).

Critical Appraisal of Underlying Assumptions

The recovery theorem introduced by Ross (2015), builds upon 5 assumptions, and in this section I will explain the assumptions and show what happens with the recovery theorem once I leave out any of them. The underlying assumptions are of course important (otherwise they would not be made), but the goal of this chapter is to give intuition about the economic purpose/meaning of these assumption. This might help explain why empirical studies that test the recovery theorem on real-world data conclude that it does not accurately reflect reality.

$\mathbb{P}$ characterized by Markov chain

In this section I will elaborate on the first assumption that I made: the event space is a Markov chain.
The definition of a Markov chain[9] is the following:

Definition (Markov Process[10])**.** *Let $(X_t)_{t\geq 0}$ be a stochastic process with state space S. Then, this stochastic process is said to be a Markov process if:*

$$\mathbb{P}\left[X_{t+s} = j \mid X_s = i,\ X_u = x_u,\ 0 \leq u < s\right] = \mathbb{P}\left[X_{t+s} = j \mid X_s = i\right] \tag{57}$$

for all $s, t \geq 0$, and all $i, j, x_u \in S$.

In other words (57) implies that *the future, given the present, does not depend on the past.* This means that if there is no Markov chain, It must hold that the future may depend on the past. The problem now is that by doing this, we increased the degrees of freedom with respect to how the probabilities relate to one another.
By increasing the degrees of freedom, it is more difficult to find the initial state price matrix C and thus it becomes more difficult to estimate the real-world probabilities, not only because we do not know whether the matrix C is correct, but also because we do not know whether the dependency within the estimated real-world probabilities hold in the future. Consider the following; assume that currently we are at time $t = 0$. We estimate the real-world probabilities and then try to explain what happens at time $t = 4$. I now observe a problem since these $t = 4$ probabilities might depend on the realisations at time $t = 2$ and $t = 3$ (future depends on the past if there is no Markov chain). However, at time $t = 0$ I cannot know what would happen at time $t = 2$ and $t = 3$ (these events are not in the information set $\mathcal{F}_0$). Hence, the estimation that I found cannot be used for the future and thus using the recovery theorem in this setting does not work.
Furthermore, recall that the Radon-Nikodym derivative was defined as (see (7)):

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(\omega) = \frac{\mathbb{Q}[\{\omega\}]}{\mathbb{P}[\{\omega\}]}$$

[9] note: a Markov chain is a discrete time Markov process
[10] see, for example Resnick (1992)

If the state space is characterised by a Markov chain, then we are able to compute the real-world probability in terms of a product ($\mathbb{P}[\{\omega\}] = \mathbb{P}[\{j_0, ..., j_\tau\}] = p_0(j_0)\prod_{t=1}^{\tau} p(j_{t-1}, j_t)$). However, if we do not have a Markov chain, this expression becomes:

$$\mathbb{P}[\{\omega\}] = \mathbb{P}[\{j_0, ..., j_\tau\}] = \mathbb{P}[j_0] \cdot \mathbb{P}[j_1|j_0] \cdot \mathbb{P}[j_2|j_0,\, j_1] \cdot ... \cdot \mathbb{P}[j_\tau|j_0,\, ...\, ,\, j_{\tau-1}] \tag{58}$$

This, we cannot write as a product of transition probabilities since there might be dependence between two future (Markov) states. If I assume that the event space is a Markov chain, then all these conditional probabilities can be reduced to transition probabilities. Consequently, if I do not assume that I have a Markov chain, I cannot reduce the pricing kernel to the ratio

$$e^{-r^{(i)}}\frac{d\mathbb{Q}}{d\mathbb{P}}(\omega) = e^{-r^{(i)}}\frac{p(i,j)}{q(i,j)}$$

and so, the recovery theorem does not work. Now that I have shown why the Markov chain assumption is required, I can discuss how realistic this expression is.

It is not uncommon in financial economics to assume that stochastic processes are Markov processes. If we examine market dynamics, one could reason that due to for example noise traders, future returns might not depend on previous values. These noise traders base their decisions on current news or current beliefs, so they might not really consider historic behavior of the market. That said, It might be that noise traders base themselves on past (i.e 2 years) behavior, but not that much in the past. This means that assuming that *the future, given the present, does not depend on the past* might not be that far-fetched.

Once someone agrees upon this Markov process, then how does someone go towards a Markov chain? Consider the following strategy: recall that each (Markov) state represents some return. Of course, theoretically, the returns can range from negative infinity to positive infinity[11]. I can then make a (finite) partition of this interval, such that I am left with $\Xi < \infty$ *"(Markov) states"* (for example $R \in (-\infty, \infty)$ becomes $R \in \{\Lambda_1, ..., \Lambda_\Xi\}$ where $\Lambda_1 = \{R : R < l_1\}$, $\Lambda_2 = \{R : l_1 \leq R < l_2\}$,...,$\Lambda_\Xi = \{R : R \geq l_{\Xi-1}\}$). Then, I no longer have the probability that the return is equal to some value, but I get the probability that a return lies in the partition. Consequently, the transition probabilities correspond to the probability that a return lies in some partition rather than the probability that a return takes some value. This shows that the Markov chain assumption is important, and furthermore reasonable realistic.

Absence of Arbitrage

In this second section, I will elaborate on the next assumption that I made: the market obeys absence of arbitrage.

Absence of arbitrage is a well-known concept within financial economics. It means that it is not possible to make a riskless profit. Consider the formal definition of absence of arbitrage:

Definition (Arbitrage Opportunity). *Let ϕ_t be a self-financing trading strategy and let $(X_t)_{t\geq 0}$ be a ($\mathcal{F}_t$-adapted) stochastic process. The trading strategy ϕ_t is an arbitrage opportunity if for the value $V_t := \sum_{a\in\mathcal{A}} \phi_t^a X_t^a = \phi_t^T X_t$ where $\mathcal{A}$ is the set of all (tradeable) assets:*

i) $V_0 \leq 0$ *(no initial cashflow/investment)*

ii) $\mathbb{P}[V_\tau \geq 0] = 1$ *(riskless investment)*

iii) $\mathbb{P}[V_\tau > 0] > 0$ *(there is a probability of making a profit)*

Definition (Absence of Arbitrage). *A market satisfies absence of arbitrage if no arbitrage opportunity exist.*

To explain the relevance of Assumption 2, consider the scenario in which absence of arbitrage does not hold. In this scenario, there must exist an arbitrage opportunity and hence the three

[11]where the likelihood of a really high or really low return is near zero

conditions in the definition must hold. Since absence of arbitrage does not hold, The first fundamental theorem of asset pricing implies that there does not exist a risk neutral probability measure $\mathbb{Q}$.
Since, this probability measure does not exist, the whole recovery theorem cannot be applied. After all, the objective was to recover the real-world probabilities from the risk neutral probabilities (or observed prices), however if these risk neutral probabilities do not exist, the recovery theorem cannot be applied.
Just as with the previous assumption, this assumption is not that far-fetched since assuming absence of arbitrage is quite common within financial economics. A quite often used argument is that if there exist some arbitrage opportunity, these will be rapidly traded away by arbitrageurs. Hence, assuming absence of arbitrage is also reasonable realistic. Therefore, it is not surprising that the empirical papers do not contradict this assumption.

Irreducible Markov chain

In Assumption 3, I assumed that the Markov chain is irreducible. What happens if Assumption 3 no longer applies?
Mathematically, I observe a problem if I do not have an irreducible matrix. If the matrix is reducible, I cannot apply the Perron-Frobenius theorem to find a solution for (37). And so the question arises whether I am still able to find a (feasible) solution to (37)

$$C\nu = \beta\nu \iff CH^{-1}\iota = \beta H^{-1}\iota,$$

once I omit Assumption 3?
Recall that any solution to the above equation must satisfy that $\nu_i > 0$ since $\nu_i = \frac{1}{h(i)}$ and due to the transition independent kernel, $h(i) > 0$.
Consider the following example:

Example 5. *Let me assume* $\mathcal{R} = \begin{bmatrix} r^{(1)} & \dots & r^{(4)} \end{bmatrix}^T = \begin{bmatrix} 0.08 & 0.02 & 0 & -0.02 \end{bmatrix}^T$ *and recall the reducible matrix from Example 2. In that example, We had the matrix*

$$Q = \begin{bmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 \\ 0 & 0.1 & 0.6 & 0.3 \\ 0 & 0 & 0.1 & 0.9 \end{bmatrix}.$$

In Example 2 I concluded that the above matrix is reducible by looking at the graph. However, I can also validate this by computing

$$(I_4 + Q)^{4-1} = \begin{bmatrix} 3.5540 & 4.4460 & 0 & 0 \\ 2.2230 & 5.7770 & 0 & 0 \\ 0.1410 & 0.8380 & 4.2490 & 2.7720 \\ 0.0030 & 0.0520 & 0.9240 & 7.0210 \end{bmatrix}. \tag{59}$$

Now, I can conclude that there are elements equal to zero and so the matrix is reducible. Then, I compute the state price matrix C, by computing $C_{i,j} = \mathcal{R}_i Q_{i,j}$
If I now compute the eigenvalues of the matrix C, I find that the largest eigenvalue (in absolute value) is 1.0672 and if I compute all four eigenvectors, I find $\begin{bmatrix} 0.890 \\ -0.445 \\ 0.093 \\ -0.011 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -0.971 \\ 0.241 \end{bmatrix}, \begin{bmatrix} -0.691 \\ -0.702 \\ -0.098 \\ 0.143 \end{bmatrix},$ $\begin{bmatrix} 0 \\ 0 \\ 0.576 \\ 0.817 \end{bmatrix}$. *Clearly, these eigenvectors do not yield feasible estimates, since firstly, not all eigenvectors are positive and secondly, using $\nu_i = 0$ implies that $h(i) \equiv \frac{1}{\nu_i} = \frac{1}{0}$ and this number is not defined. So I cannot use (37) to estimate my real-world probabilities.*

In general, the recovery theorem uses the Perron-Frobenius theorem, and this theorem requires a matrix to be irreducible (and non-negative) in order to be applied. Therefore, I need to make this assumption.
Moreover, recall that a Markov chain is irreducible if it is possible to reach any (Markov) state from any other (Markov) state within a finite number of movements. If the Markov chain is no longer irreducible, it must hold that there exists a (Markov) state such that once I leave this (Markov) state, I will no longer return to it (within a finite number of movements). Since a (Markov) state corresponds to some return, this would translate to a certain return that cannot be reached once I leave it.
The most intuitive example would be an extremely low return. Once a return is extremely low, the company that sells the asset probably will default and thus we will no longer reach any other return. This shows that we cannot find real-world default probabilities using the recovery theorem, since introducing a (Markov) state that represents the company to default, causes the recovery theorem to no longer work, since this kind of Markov chain will be reducible.

Strong rational expectation

The fourth assumption I made was that all agents obey strong rational expectations. In this section I will consider what happens if this assumption is not made. Notice, that this assumption was not made in the original paper of Ross (2015).

Recall, the definition of strong rational expectations:

Definition. *Assume that there are N agents acting on the market and that $\mathcal{P}_n$ $(n = 1, ..., N)$ is the subjective belief of agent $n = 1, ..., N$ and $\mathbb{P}$ the real-world probability measure, then it must hold that $\mathcal{P}_n[A] = \mathbb{P}[A]\ \forall n = 1, ..., N,\ \forall A \in \mathcal{F}_t$.*

Without strong rational expectations, we just have N agents all with their own subjective beliefs, $\mathcal{P}_n$, with respect to what the current prices will do in the future. How will this influence the recovery theorem?
The agents observe the current (asset) prices and by assuming absence of arbitrage, they thus agree upon the risk neutral probability measure ($\mathbb{Q}$). If there is no strong rational expectations, then all these subjective beliefs could be different than the real-world probability measure $\mathbb{P}$ (so, $\mathbb{P}_n[A] \neq \mathbb{P}[A]$). This automatically, implies that there cannot be one (unique) pricing kernel, since now there are multiple different pricing kernels in the form of $\frac{C_{i,j}}{\mathcal{P}_n(i,j)}$.
Furthermore, to recover the real-world probabilities based on Arrow-Debreu state prices, I use the equation:

$$k_{i,j} := \frac{C_{i,j}}{p(i,j)} \iff C_{i,j} = k_{i,j}p(i,j) \tag{60}$$

The only restriction that I have, is that $\sum_{j=1}^{m} p(i,j) = 1$. At the same time, I can construct many other kernels that obey (60), for example:

$$\tilde{k}_{i,j} := \frac{C_{i,j}}{\mathcal{P}_n(i,j)} \iff C_{i,j} = \tilde{k}_{i,j}\mathcal{P}_n(i,j) \tag{61}$$

where $\mathcal{P}_n(i,j) := \eta_{i,j}p(i,j)$ such that $\sum_{j=1}^{m} \eta_{i,j}p(i,j) = 1$. If I now continue with this logic, I can still apply the transition independent kernel assumption to obtain an estimation of the form:

$$\mathcal{P}_n(i,j) = C_{i,j}\frac{1}{\beta}\frac{v_j}{v_i} \iff p(i,j) = C_{i,j}\frac{1}{\beta}\frac{v_j}{v_i}\frac{1}{\eta_{i,j}}. \tag{62}$$

Hence, the recovery theorem does no longer predict the real-world probability measure ($\mathbb{P}$), but the subjective belief of the agent ($\mathcal{P}_n$), once I discard the strong rational expectation assumption.
Note that in the original paper of Ross (2015), this strong rational expectations assumption was not imposed, and this led other researchers (specifically Borovicka et al. (2016)) to state the following:

"Interestingly, this recovery does not impose rational expectations, and thus the resulting Markov evolution could reflect investors' subjective beliefs and not necessarily the actual time-series evolution" (Borovicka et al., 2016). Hence, the assumption of strong rational expectation is important to include, even though, in the original paper of Ross (2015), this was not the case.

Transition independent kernel

The final assumption that I made was that the pricing kernel is transition independent. The transition independent kernel assumption, states that the pricing kernel must satisfy:

$$\frac{C_{i,j}}{p(i,j)} = \beta \frac{h(j)}{h(i)}$$

This, I could rewrite to (see (32))

$$HC = \beta PH \iff P = \frac{1}{\beta} HCH^{-1}$$

and using the fact the P is a stochastic matrix I eventually arrived at (see (37))

$$C\nu = \beta\nu$$

By making an assumption about the structure of a pricing kernel, you impose restrictions on the structure of the market. So, this assumption requires special attention. Consequently, most empirical papers test whether this assumption holds in reality. It turns out that in reality, pricing kernels do not obey this kind of structure (Bakshi et al., 2017; Jackwerth and Menner, 2020), and in this section I will explain why empirical papers find their results. I will start by showing that there is no difference in calculating the long-term real-world probabilities based on long-term state prices or short-term state prices if the kernel is transition independent.

Consider the following theorem:

Theorem 3. *Let λ be an eigenvalue of the square matrix $A \in \mathbb{R}^{n\times n}$ and let w be the corresponding eigenvector, such that $Aw = \lambda w$. Then, λ^k is an eigenvalue of A^k and w is the corresponding eigenvector, such that $A^k w = \lambda^k w$, $\forall k \in \mathbb{N} \setminus \{0\}$.*

Proof. Note that, if $k = 1$, we have $A^1 w = Aw = \lambda w = \lambda^1 w$. So, the statement holds if $k = 1$. Let me now assume that $k = \kappa + 1$ and that it holds that $A^\kappa w = \lambda^\kappa w$, for some $\kappa > 1$ (proof by induction). Then, I have:

$$A^{\kappa+1} w = AA^\kappa w = A\lambda^\kappa w = \lambda^\kappa Aw = \lambda^\kappa \lambda w = \lambda^{\kappa+1} w \tag{63}$$

So, the statement holds as well if $\kappa > 1$. Hence, by induction, the statement must hold for every $k \in \mathbb{N} \setminus \{0\}$. □

Recall, that C is the (one-period) state price matrix and that $C_{i,j} \equiv c_j^{(i)}$ represents the state price of an Arrow-Debreu security that pays off 1 in case I reach (Markov) state j, starting from (Markov) state i, the next period.
Furthermore, C^τ is the (τ-period) state price matrix and $C_{i,j}^\tau \equiv c_j^{(i)}(\tau)$ represents the state price of an Arrow-Debreu security that pays of 1 in case I reach (Markov) state j, starting from (Markov) state i, after τ periods.
In the recovery theorem, I compute $\hat{P}$ based on the eigenvalues and eigenvectors of C, but with the same reasoning, I can estimate the probabilities $\hat{P^\tau}$, by computing the eigenvalues and eigenvectors of C^τ and using Theorem 3, I know that if ν is the eigenvector of C that corresponds to eigenvalue β, then it must hold that ν is also the eigenvector of C^τ that corresponds to eigenvalue β^τ. Consequently, I find that

$$\hat{P^\tau} = \frac{1}{\beta^\tau} HC^\tau H^{-1} \tag{64}$$

Notice, however, that I can also compute $\hat{P}^\tau$ from knowing $\hat{P}$ and then just compute $\hat{P}^\tau = \hat{P} \cdot ... \cdot \hat{P}$. Doing this yields,

$$\hat{P} \cdot ... \cdot \hat{P} = \frac{1}{\beta} HCH^{-1} \cdot ... \cdot \frac{1}{\beta} HCH^{-1} = \left(\frac{1}{\beta}\right)^\tau HC^\tau H^{-1}. \tag{65}$$

Interestingly, I get the same result as in (64). This implies that, for long-term probabilities, it does not matter If I compute my estimates for the real-world probabilities based on one period state prices or based on long-term state prices.

Another interesting result from assuming that the kernel is transition independent is the fact that not only long-term and short-term state prices yield the same estimated real-world probabilities, but also that these estimated real-world probabilities turn out to be long-term constant. That is, in the long-run the real-world transition probabilities no longer depend on the (Markov) state that you leave, but only depend on the (Markov) state that you go to.
To see this, let me introduce an alternative formulation of the Perron-Frobenius theorem (see Meyer (2000) (8.3.10)):

Theorem 4 (Alternative Perron-Frobenius Theorem). *let w^T and v be the left and right eigenvalues (respectively) of some square matrix A, that corresponds to eigenvalue $\rho(A) := \max\{|\lambda| : \lambda \text{ is eigenvalue of } A\}$, such that $w^T v = 1$. Then,*

$$\lim_{k\to\infty} \left(\frac{1}{\rho(A)}\right)^k A^k = vw^T \tag{66}$$

Recall that C is the (one-period) state price matrix and that C^τ represents the τ-period state price matrix. If I now apply this alternative Perron-Frobenius theorem to matrix C, I get:

$$\lim_{\tau\to\infty} \left(\frac{1}{\rho(C)}\right)^\tau C^\tau = \nu l, \tag{67}$$

where l is the left eigenvector and ν the right eigenvector, both corresponding to eigenvalue $\rho(C)$. Recall, however, that this $\rho(C)$ represents the β in the transition independent kernel and ν_i represents $\frac{1}{h(i)}$.
Furthermore, recall that $\hat{P} = \frac{1}{\beta} HCH^{-1}$ and $\hat{P}^\tau = \left(\frac{1}{\beta}\right)^\tau HC^\tau H^{-1}$. Consequently,

$$\begin{aligned} \lim_{\tau\to\infty} \hat{P}^\tau &= \lim_{\tau\to\infty} \left(\frac{1}{\beta}\right)^\tau HC^\tau H^{-1} = H\nu l H^{-1} \\ &= \begin{bmatrix} \nu_1 h(1) \\ \vdots \\ \nu_m h(m) \end{bmatrix} \begin{bmatrix} \frac{l_1}{h(1)} & \cdots & \frac{l_m}{h(m)} \end{bmatrix} \\ &= \begin{bmatrix} \nu_1 h(1) \frac{l_1}{h(1)} & \dots & \nu_1 h(1) \frac{l_m}{h(m)} \\ \vdots & \ddots & \vdots \\ \nu_m h(m) \frac{l_1}{h(1)} & \dots & \nu_m h(m) \frac{l_m}{h(m)} \end{bmatrix}. \end{aligned} \tag{68}$$

Since $h(i) = \frac{1}{\nu_i}$, I have that

$$\lim_{\tau\to\infty} \hat{P}^\tau = \begin{bmatrix} \nu_1 h(1) \frac{l_1}{h(1)} & \dots & \nu_1 h(1) \frac{l_m}{h(m)} \\ \vdots & \ddots & \vdots \\ \nu_m h(m) \frac{l_1}{h(1)} & \dots & \nu_m h(m) \frac{l_m}{h(m)} \end{bmatrix} = \begin{bmatrix} l_1 \nu_1 & \dots & l_m \nu_m \\ \vdots & \ddots & \vdots \\ l_1 \nu_1 & \dots & l_m \nu_m \end{bmatrix}. \tag{69}$$

Hence, $\lim_{\tau\to\infty} \hat{P}^\tau(i,j) = l_j \nu_j$, where $\nu \in \mathbb{R}^{m\times 1}$ and $l \in \mathbb{R}^{1\times m}$ are such that $lC = l\rho(C)$ and $C\nu = \rho(C)\nu$. In other words, if $\tau \to \infty$, the transition probabilities only depend on the (Markov) state I go to and not the current state. Recall that the (Markov) states represent returns and that the result in (69) resembles some stationary/limiting distribution of a Markov chain. Consequently, the transition independent kernel is called *"trend stationary"* by Borovicka et al. (2016). Furthermore, notice that in Theorem 4, one of the requirements is that $w^T v = 1$. If I translate this to the matrix C, we have that $l\nu = \sum_{i=1}^{m} l_i \nu_i = 1$ and this means that $\lim_{\tau \to \infty} \hat{P}^{\tau}$ has row sum equal to 1 (or $\lim_{\tau \to \infty} \hat{P}^{\tau} \iota = \iota$). Notice, that this result is also found in Example 4. In that chapter I estimated the real-world probabilities and we also found that $\hat{P}$ only changes in the columns, not the rows even though the matrix C did change in both the column and row direction.

These two reasons are the most probable explanations why empirical papers, such as, for example, Jackwerth and Menner (2020) and Bakshi et al. (2017) find that the transition independent kernel assumption does not work in reality. They find that this specific structure of a pricing kernel does not find accurate transition probabilities, and consequently, that assuming this kind of structure most likely explains why the recovery theorem does not work in reality.

Implications of the Recovery Theorem

In this chapter I will highlight some implications that follow if all underlying assumptions of the recovery theorem hold. The results in this chapter are not a direct consequence of one specific assumption, however, the combination all the assumptions causes the following results to appear.

Constant risk-free rate

In this section, I would like to emphasise one particular scenario, in which the risk-free rate in each state is constant. That is, in this section I consider what happens in case $r^{(1)} = ... = r^{(m)} = r$. In this scenario, the state price matrix, C, is given by

$$C = \begin{bmatrix} e^{-r^{(1)}}q(1,1) & \dots & e^{-r^{(1)}}q(1,m) \\ \vdots & \ddots & \vdots \\ e^{-r^{(m)}}q(m,1) & \dots & e^{-r^{(m)}}q(m,m) \end{bmatrix} = \begin{bmatrix} e^{-r}q(1,1) & \dots & e^{-r}q(1,m) \\ \vdots & \ddots & \vdots \\ e^{-r}q(m,1) & \dots & e^{-r}q(m,m) \end{bmatrix} = e^{-r} \cdot Q. \quad (70)$$

Recall that the recovery theorem builds upon trying to find a solution to (37):

$$C\nu = \beta\nu \iff e^{-r}Q\nu = \beta\nu$$

Notice, however, that Q is a stochastic matrix and thus $Q\iota = \iota$ must hold. However, the Perron-Frobenius theorem tells us that any nonnegative irreducible matrix A has a (unique) positive eigenvector that corresponds to eigenvalue $\rho(A)$. So, this implies that $\iota := \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ is this unique positive eigenvector of Q.

Consider the following property of eigenvalues:

Property 1. *Let $A \in \mathbb{R}^{m\times m}$ be a square matrix and let w be the eigenvector that corresponds to eigenvalue λ, such that $Aw = \lambda w$. Then, for every scalar $\alpha \in \mathbb{R} \setminus \{0\}$ it must hold that $\tilde{A} := \alpha A$ has as eigenvector w and corresponding eigenvalue $\tilde{\lambda} = \alpha\lambda$, such that $\tilde{A}w = \alpha Aw = \alpha\lambda w = \tilde{\lambda}w$*

And so, using this property, if 1 is an eigenvalue of Q, then it must hold that e^{-r} must be the eigenvalue of $e^{-r}Q = C$. This means, that $\nu = \iota$ and $\beta = e^{-r}$ and this implies that the estimated probabilities become:

$$\hat{p}(i,j) = C_{i,j}\frac{1}{\beta}\frac{\nu_j}{\nu_i} = e^{-r}q(i,j)\frac{1}{e^{-r}}\frac{1}{1} = q(i,j) \quad (71)$$

Hence, if I have constant risk-free rate in each (Markov) state, the recovery theorem finds that $\hat{p}(i,j) = q(i,j)$. Therefore, the real-world probabilities (in this scenario) are solely determined by the risk neutral measure, implying that the market only consists out of idiosyncratic risk; no external event can influence the real-world probabilities and there cannot be a risk premium.

long-term risk neutral probability measure

In this second section, I will explain another implication of the recovery theorem. It turns out that the recovery theorem yields an interesting result with respect to long-term risk. In the previous chapter, I already showed two interesting long-term results, however in this section I will make a general conclusion regarding the special behavior of the recovery theorem with respect to long-term risk in accordance with what Borovicka et al. (2016) conclude in their paper.

To explain this special behaviour of the transition independent kernel, recall that the transition independent kernel assumption yields as a kernel:

$$k_{i,j} := \frac{C_{i,j}}{p(i,j)} = \beta\frac{h(j)}{h(i)} = \rho(C)\frac{\nu_i}{\nu_j}$$

where ν is the eigenvector corresponding to eigenvalue $\rho(C) := max\{|\lambda| : \lambda \textit{ is eigenvalue of } C\}$. Notice, however, that I can change this formulation of the kernel as follows:

$$C_{i,j} = k_{i,j}p(i,j) \cdot \frac{\tilde{p}(i,j)}{\tilde{p}(i,j)} = k_{i,j}\tilde{p}(i,j)\theta(i,j) = \rho(C)\frac{\nu_i}{\nu_j}\tilde{p}(i,j)\theta(i,j) \tag{72}$$

where $\theta(i,j) := \frac{p(i,j)}{\tilde{p}(i,j)}$. And so, I arrive at a new kernel, that obeys all other restrictions (other than a transition independent kernel) that follows:

$$C_{i,j} = \tilde{k}_{i,j}\tilde{p}(i,j) = \left(\rho(C)\frac{\nu_i}{\nu_j}\theta(i,j)\right)\tilde{p}(i,j) \tag{73}$$

Since the transition independent kernel has the structure:

$$k_{i,j} = \frac{C_{i,j}}{p(i,j)} = \beta\frac{h(j)}{h(i)},$$

this new kernel must obey:

$$\tilde{k}_{i,j} = \frac{C_{i,j}}{\tilde{p}(i,j)} = \beta\frac{h(j)}{h(i)}\theta(i,j), \tag{74}$$

where $\theta(i,j) := \frac{p(i,j)}{\tilde{p}(i,j)}$. This way of writing the pricing kernel was introduced by Hansen and Scheinkman (2009). If I now compare the transition independent kernel of Ross (2015) with this new kernel by Hansen and Scheinkman (2009), it is easily seen that the main difference lies in this new factor $\theta(i,j)$. In fact, the transition independent kernel is a special case of the kernel by Hansen and Scheinkman (2009), where Ross (2015) sets $\theta(i,j) \equiv 1\ \forall i,j = 1,...,m$. So, for a better understanding of the implication of a transition independent kernel, I will elaborate on what this $\theta(i,j)$ means and what the result of setting this $\theta(i,j) \equiv 1$ entails.
In (74), I mentioned the general pricing kernel:

$$\tilde{k}_{i,j} = \frac{C_{i,j}}{\tilde{p}(i,j)} = \beta\frac{h(j)}{h(i)}\frac{p(i,j)}{\tilde{p}(i,j)}$$

Here, there is a new probability $\tilde{p}(i,j)$ that is different than the real-world probabilities $p(i,j)$. In fact, I now have a kernel that is influenced by two probability measures $\mathbb{P}$ and $\tilde{P}$ and the fraction $\theta(i,j) = \frac{p(i,j)}{\tilde{p}(i,j)}$ is a change of measure from $\mathbb{P}$ to $\tilde{P}$. Clearly, the transition independent kernel, $k_{i,j} := \frac{C_{i,j}}{p(i,j)} = \beta\frac{\nu_i}{\nu_j}$, is a pricing kernel under probability measure $\mathbb{P}$ and so it must hold that:

$$\pi_0 = E^{\mathbb{P}}\left[k_{i,j}^{(\tau)}\Pi_\tau | j_0 = i\right], \tag{75}$$

where π_0 is the price of some asset and Π_τ the (random) payoff of this security at time $t = \tau$. However, what do we get if I do not use the real-world probability measure $\mathbb{P}$ (that we find using the recovery theorem), but the new probability measure $\tilde{P}$? In this case, I have a new *"price"* $\tilde{\pi}_0$ that corresponds to the (random) payoff Π_τ[12]:

$$\tilde{\pi}_0 = E^{\tilde{P}}\left[k_{i,j}^{(\tau)}\Pi_\tau|j_0 = i\right] = E^{\tilde{P}}\left[\beta^\tau\frac{\nu_i}{\nu_j}\Pi_\tau|j_0 = i\right] = \beta^\tau\nu_i E^{\tilde{P}}\left[\frac{\Pi_\tau}{\nu_j}|j_0 = i\right] \tag{76}$$

Next, I will compute the τ-period yield of this asset:

$$\begin{aligned} e^{\tilde{Y}_\tau\tau} = \frac{E^{\tilde{P}}\left[\Pi_\tau\right]}{\tilde{\pi}_0} &\iff \tilde{Y}_\tau = \frac{1}{\tau}\left(ln\left(E^{\tilde{P}}\left[\Pi_\tau\right]\right) - ln(\beta^\tau) - ln\left(E^{\tilde{P}}\left[\nu_i\frac{\Pi_\tau}{\nu_j}|j_0 = i\right]\right)\right) \\ \iff \tilde{Y}_\tau = -ln(\beta) + \frac{1}{\tau}&\left(ln\left(E^{\tilde{P}}\left[\Pi_\tau\right]\right) - ln\left(E^{\tilde{P}}\left[\nu_i\frac{\Pi_\tau}{\nu_j}|j_0 = i\right]\right)\right) \end{aligned} \tag{77}$$

Now, Observe that as $\tau \to \infty$,

$$\lim_{\tau\to\infty} \tilde{Y}_\tau = -ln(\beta) \tag{78}$$

If I do the same for the yield under the real-world probability, I get:

$$e^{Y_\tau\tau} = \frac{E^{\mathbb{P}}\left[\Pi_\tau\right]}{\pi_0} \iff Y_\tau = \frac{1}{\tau}\left(ln(E^{\mathbb{P}}\left[\Pi_\tau\right]) - ln(\pi_0)\right) \tag{79}$$

Now, I find that

$$\lim_{\tau\to\infty} Y_\tau = 0 \tag{80}$$

This shows that if I use the probability measure that is estimated using the recovery theorem, I construct a probability measure that in the long-term results in yields that are equal to zero. In other words, the probability measure $\mathbb{P}$ from the recovery theorem implies that there is no long-term yield. On the other hand, If I use another probability measure $\tilde{P}$, that does not follow from the recovery theorem, then I do observe long-term yields.
Within finance it is well-known that there is a long-term yield, due to the risk premium that people demand (Jawadi and Prat, 2017). So the fact that the recovery theorem refutes this phenomena, shows that in practice the theorem might not work. This is the reason why Borovicka et al. (2016) refer to the estimated real-world probability measure as *"long-term risk neutral"*.

[12]notice, that $k_{i,j}$ is a pricing kernel under $\mathbb{P}$, since it holds that $k_{i,j} = \frac{C_{i,j}}{p(i,j)}$. Under this new probability measure $\tilde{P}$, this $k_{i,j}$ is no longer a pricing kernel. Under $\tilde{P}$ I do have a pricing kernel, but this pricing kernel is $\tilde{k}_{i,j} = \frac{C_{i,j}}{\tilde{p}(i,j)}$.

Conclusion

In this thesis I revisited the recovery theorem. I found that there are five assumptions that are required such that the recovery theorem works. Of these five assumptions there are two that could be true in reality (Markov chain and no arbitrage), however there are also two assumptions that are *"bold"* in the sense that it is quite unlikely that these hold in reality (transition independent kernel and (strong) rational expectations). Furthermore, I introduced some explanations why empirical papers conclude that the recovery theorem does not work in reality.

That said, even though the recovery theorem does not work in practice, it does not (automatically) mean that it is useless, since there are plenty of commonly used models that do not fully replicate reality either. For example, the Black-Scholes model oversimplifies reality, by assuming that there is a constant risk-free rate, and that the risky asset has constant expected return and constant volatility. Or the Vasicek model is not capable of replicating the current term structure of interest rate, however it is still used sometimes to estimate future term structures of interest rates. So, while the recovery theorem cannot be applied in reality, there might still be certain aspects that can help researchers understand what is required to make a model that, in the future, could be useful. After all, as George Box quoted: *"All models are wrong, but some might be useful"* (Box, 1976).

To conclude my thesis, I would like to suggest possible future research based on this thesis:

1. In the final chapter, I introduced the *"general"* transition independent kernel

$$\tilde{k}_{i,j} = \frac{C_{i,j}}{\tilde{p}(i,j)} = \beta \frac{h(j)}{h(i)} \theta(i,j).$$

   There has been ample of empirical research that concludes that the transition independent kernel (with $\theta(i,j) \equiv 1$) does not work in reality. However, no research has yet been done to check if this *"general"* kernel is more realistic, let alone if the recovery theorem works using this kernel. Hence, investigating if this "general" transition independent kernel works might be interesting future research.

2. Empirically test if the irreducible Markov chain assumption holds. Theoretically, it seems improbable that this assumption is reasonable, since an example of a Markov chain that is reducible could be a Markov chain in which there is a (Markov) state that implies the default of a company and we know that any company has a probability of default, which implies that assuming an irreducible Markov chain might not be reasonable. However, as long as there is no empirical proof, the assumption that the Markov chain is irreducible is still debatable. It might even be the case that for some markets this assumption is reasonable and for other markets, it should not be assumed.

3. How would the recovery theorem be formulated in continuous time? Recall the first assumption. This assumption assumes that the event space is characterised by a Markov chain. A Markov chain is a discrete time finite state space stochastic process. This meant that in order to use the recovery theorem, I first needed to discretize my event space. However, we also know that a Markov chain is a discrete time Markov process and so, we might be able to construct a continuous time recovery theorem.

Bibliography

T. Adachi. On an aggregate state-price defator in a multi-period market model. *Japan Journal of Industrial and Applied Mathematics*, 38(3):1063–1074, 2021. 10.1007/s13160-021-00476-w.

K. Arrow and G. Debreu. Existence of an equilibrium for a competitive economy. *Econometrica*, 22(3):265–290, 1954. 10.2307/1907353.

G. Bakshi, F. Chabi-Yo, and X. Gao. A recovery that we can trust? deducing and testing the restrictions of the recovery theorem. *Review of Financial Studies*, 9(1):532–555, 2017. 10.1093/rfs/hhx108.

B. Borovicka, L. Hansen, and J. Scheinkman. Misspecified recovery. *Journal of Finance*, 71(6): 2493–2544, 2016. 10.1111/jofi.12404.

G. Box. Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799, 1976. 10.2307/2286841.

J. Cox and H. Leland. On dynamic investment strategies. *Journal of Economic dynamics & control*, 24(11):1859–1880, 2000. 10.1016/S0165-1889(99)00095-0.

J. Cox, S. Ross, and M. Rubinstein. option pricing: A simplified approach. *Journal of Financial Economics*, 7(3):229–263, 1979. 10.1016/0304-405X(79)90015-1.

P. Dybvig and C. Rogers. Recovery of preferences from observed wealth in a single realization. *Review of Financial Studies*, 10(1):151–174, 1997.

P. Dybvig and S. Ross. Arbitrage, state prices and portfolio theory. *Handbook of the Economics of Finance*, 1(B):605–637, 2003.

J. Fletcher. Can asset pricing models price idiosyncratic risk in u.k. stock returns? *The Financial Review*, 42(4):507 – 535, 2007.

S. Grossman and R. Shiller. The determinants of the variability of stock market prices. *American Economic Review*, 71(2):222–227, 1981.

L. Hansen and R. Jagannathan. Implications of security market data for models of dynamic economies. *Journal of Political Economy*, 99(2):225–262, 1991.

L. Hansen and J. Scheinkman. Long-term risk: An operator approach. *Econometrica*, 77(1): 177–234, 2009. 10.3982/ECTA6761.

R. Horn and C. Johnson. *Matrix Analysis*. Cambridge university press, New York, 2nd edition, 2013.

H. Iwaki, M. Kijima, and Y. Morimoto. An economic premium principle in a multiperiod economy. *Insurance: Mathematics and Economics*, 28(3):325–339, 2001. 10.1016/S0167-6687(00)00081-0.

J. Jackwerth. *Option-Implied Risk-Neutral Distributions and Risk Aversion*. CFA Institute Research Foundation, Charlottesville, 2004.

J. Jackwerth and M. Menner. Does the ross recovery theorem work empirically? *Journal of Financial Economics*, 137(3):723–739, 2020. 10.1016/j.jfineco.2020.03.006.

F. Jawadi and G. Prat. Equity prices and fundamentals: a ddm–apt mixed approach. *Review of Quantitative Finance and Accounting*, 49(3):661–695, 2017. 10.1007/s11156-016-0604-y.

R. Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 51(3):247–257, 1969.

C. Meyer. *Matrix analysis and applied linear algebra*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2000.

S. Resnick. *Adventures in Stochastic Processes*. Birkhäuser Verlag, Boston, 1992.

S. Ross. The recovery theorem. *Journal of Finance*, 70(2):615–648, 2015. 10.1111/jofi.12092.

V. van Appel and E. Maré. The ross recovery theorem with a regularised multivariate markov chain. *ORiON*, 34(2):133–155, 2018. 10.5784/34-2-594.

Appendix A

Example of a kernel that is not transition independent

The following example shows a kernel that is not transition independent. This example is based on Merton's portfolio problem (Merton, 1969), which is a renowned problem in asset-liability management. Since, I work with a finite state-space Markov chain, I will change this Merton problem to a discrete time version:

Example 6. *Let me assume that there is a riskless asset in the form of a bond (B_t) and a risky asset in the form of a stock (S_t). My time frame is a partition of the interval $[0, \tau]$ of the form $t = \{t_0, ..., t_n, ..., t_N\}$, such that $t_n = n\Delta t$ and $t_N := N\Delta t = \tau$. Furthermore, the agent has at every time point some income denoted by y_t and the consumption at time t is denoted by c_t. The number of units invested (at time t) in some asset is M_t^a with $a = \{B, S\}$. Lastly, the agent has an utility function regarding consumption $u(c(.))$ and an utility function regarding terminal wealth $\bar{u}(.)$ (or bequest motive).*
At time $t = t_n$ the agent gets his income (y_{t_n}) and he uses this income (fully) to consume or to invest. This means that we have a budget constraint:

$$(y_{t_n} - c_{t_n})\,\Delta t = \left(B_{t_n}\left(M^B_{t_{n-\Delta t}} - M^B_{t_n}\right) + S_{t_n}\left(M^S_{t_{n-\Delta t}} - M^S_{t_n}\right)\right) \tag{A.1}$$

Since, my wealth is equal the the value of my assets, I get

$$W_{t_n} = B_{t_n} M^B_{t_{n-\Delta t}} + S_{t_n} M^S_{t_{n-\Delta t}} \tag{A.2}$$

And thus the wealth dynamics become:

$$\begin{aligned} W_{t_{n+\Delta t}} - W_{t_n} &= B_{t_{n+\Delta t}} M^B_{t_n} + S_{t_{n+\Delta t}} M^S_{t_n} - \left(B_{t_n} M^B_{t_{n-\Delta t}} + S_{t_n} M^S_{t_{n-\Delta t}}\right) \\ &= \left(B_{t_n}\left(M^B_{t_{n-\Delta t}} - M^B_{t_n}\right) + S_{t_n}\left(M^S_{t_{n-\Delta t}} - M^S_{t_n}\right)\right) \\ &\quad + M^B_{t_n}(B_{t_{n+\Delta t}} - B_{t_n}) + M^S_{t_n}(S_{t_{n+\Delta t}} - S_{t_n}) \\ &= (y_{t_n} - c_{t_n})\Delta t + M^B_{t_n}(B_{t_{n+\Delta t}} - B_{t_n}) + M^S_{t_n}(S_{t_{n+\Delta t}} - S_{t_n}) \end{aligned} \tag{A.3}$$

The agent aims to solve an indirect utility function of the form:

$$\begin{aligned} J_{t_i} &= \sup_{(c_{t_n}, M_{t_n})_{n=i}^{N-1}} E_{t_i}\left[\sum_{n=i}^{N-1}\left(e^{-\delta(t_n - t_i)} u(c_{t_n})\right) + e^{-\delta(\tau - t_i)}\bar{u}(W_\tau)\right] \\ s.t.\ c_{t_n} &= y_{t_n} + \frac{1}{\Delta t}\left(B_{t_n} M^B_{t_{n-\Delta t}} - B_{t_n} M^B_{t_n} + S_{t_n} M^S_{t_{n-\Delta t}} - S_{t_n} M^S_{t_n}\right) \\ W_{t_{n+\Delta t}} &= W_{t_n} + (y_{t_n} - c_{t_n})\Delta t + M^B_{t_n}(B_{t_{n+\Delta t}} - B_{t_n}) + M^S_{t_n}(S_{t_{n+\Delta t}} - S_{t_n}) \end{aligned} \tag{A.4}$$

If I now take the derivative with respect to $M^S_{t_n}$, and solve the first order condition, I get:

$$
\begin{aligned}
\frac{\partial}{\partial M^S_{t_n}} J_{t_i} &= E_{t_i}\left[-e^{-\delta(t_n-t_i)}u'(c_{t_n})\frac{S_{t_n}}{\Delta t} + e^{-\delta(t_{n+\Delta t}-t_i)}u'(c_{t_{n+\Delta t}})\frac{S_{t_{n+\Delta t}}}{\Delta t} + e^{-\delta(\tau-t_i)}\bar{u}'(W_\tau)(S_{t_{n+\Delta t}} - S_{t_n})\right] = 0 \\
\iff S_{t_n} &\left(e^{-\delta(t_n-t_i)}u'(c_{t_n})\frac{1}{\Delta t} + e^{-\delta(\tau-t_i)}\bar{u}'(W_\tau)\right) \\
&= E_{t_n}\left[\left(e^{-\delta(t_{n+\Delta t}-t_i)}u'(c_{t_{n+\Delta t}})\frac{1}{\Delta t} + e^{-\delta(\tau-t_i)}\bar{u}'(W_\tau)\right) S_{t_{n+\Delta t}}\right] \\
\iff S_{t_n} &= E_{t_n}\left[\frac{e^{-\delta(t_{n+\Delta t}-t_i)}u'(c_{t_{n+\Delta t}})\frac{1}{\Delta t} + e^{-\delta(\tau-t_i)}\bar{u}'(W_\tau)}{e^{-\delta(t_n-t_i)}u'(c_{t_n})\frac{1}{\Delta t} + e^{-\delta(\tau-t_i)}\bar{u}'(W_\tau)} S_{t_{n+\Delta t}}\right] \\
\iff S_{t_n} &= E_{t_n}\left[K_{t_n} S_{t_{n+\Delta t}}\right]
\end{aligned}
\tag{A.5}
$$

Using FFTAP, we know that (in this setting), the pricing kernel must equal

$$
\begin{aligned}
K_{t_n} &= \frac{e^{-\delta(t_{n+\Delta t}-t_i)}u'(c_{t_{n+\Delta t}})\frac{1}{\Delta t} + e^{-\delta(\tau-t_i)}\bar{u}'(W_\tau)}{e^{-\delta(t_n-t_i)}u'(c_{t_n})\frac{1}{\Delta t} + e^{-\delta(\tau-t_i)}\bar{u}'(W_\tau)} \\
&= \frac{e^{-\delta t_{n+\Delta t}}\left[u'(c_{t_{n+\Delta t}})\frac{1}{\Delta t} + e^{-\delta(\tau-t_{n+\Delta t})}\bar{u}'(W_\tau)\right]}{e^{-\delta t_n}\left[u'(c_{t_n})\frac{1}{\Delta t} + e^{-\delta(\tau-t_n)}\bar{u}'(W_\tau)\right]} \\
&= e^{-\delta(t_{n+\Delta t}-t_n)}\frac{u'(c_{t_{n+\Delta t}})\frac{1}{\Delta t} + e^{-\delta(\tau-t_{n+\Delta t})}\bar{u}'(W_\tau)}{u'(c_{t_n})\frac{1}{\Delta t} + e^{-\delta(\tau-t_n)}\bar{u}'(W_\tau)} \\
&= \tilde{\beta}\frac{g(j_{n+\Delta t}, j_\tau)}{g(j_n, j_\tau)}
\end{aligned}
\tag{A.6}
$$

And this is not a transition independent kernel, since the kernel cannot be written as $K_{t_n} = \beta\frac{h(j)}{h(i)}$, because in both the numerator and denominator, there is a (positive) function $g(j_m, j_\tau)$ that depends on two (Markov) states, not one. This function depends on the terminal wealth, which we only know in the final (Markov) state, and some other Markov state that we can move to.
In fact, if we are interested in the terminal wealth at any (nonconsecutive) future (Markov) state, we can solve the same optimization as in (A.4), where we change τ to the time that we are interested in and the summation we change such that is goes until the time point just before our desired ending point. Then, the resulting FOC is approximately the same and so the pricing kernel will be of the form $K_{t_n} = \bar{\beta}\frac{\tilde{g}(j_{n+\Delta t}, j_{\tilde{\tau}})}{\tilde{g}(j_n, j_{\tilde{\tau}})}$, where $\tilde{\tau} = \{j_{n+2\Delta t}, ..., j_\tau\}$ is the ending point of our interest. And using this generalization, one can easily see that the kernel is only transition independent if we are interested in maximizing consumption, and disregard the value of my terminal wealth.

Hence, as soon as I introduce some kind of concern regarding terminal wealth, the corresponding pricing kernel will no longer be transition independent.