



Markov Chain Modelling in Finance: Stock Valuation and Price Discovery

Riccardo De Blasis

This thesis is presented as part of the requirements for the conferral of the degree:

Doctor of Philosophy

Supervisor:
Prof. Alex Frino

Co-supervisor:
Prof. Terry Walter

The University of Wollongong
School of Accounting, Economics and Finance

July, 2019

This work © copyright by Riccardo De Blasis, 2019. All Rights Reserved.

No part of this work may be reproduced, stored in a retrieval system, transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of the author or the University of Wollongong.

Declaration

I, *Riccardo De Blasis*, declare that this thesis is submitted in partial fulfilment of the requirements for the conferral of the degree *Doctor of Philosophy*, from the University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. This document has been already submitted at the University of Chieti-Pescara as per the co-tutelle PhD agreement between the University of Wollongong and the University of Chieti-Pescara (Italy).

Riccardo De Blasis

July 31, 2019

Abstract

In this thesis we present three financial applications of Markov chain models based on three separate papers. The focus is about two important topics in finance, namely stock valuation and price discovery.

In the first paper, we propose further advancements in the Markov chain stock model. First, we provide a formula for the second order moment of the fundamental price process with transversality conditions that avoids the presence of speculative bubbles. Second, we assume that the process of the dividend growth is governed by a finite state discrete time Markov chain and, under this hypothesis, we are able to compute the moments of the price process. We impose assumptions on the dividend growth process that guarantee finiteness of price and risk and the fulfilment of the transversality conditions. Subsequently, we develop non parametric statistical techniques for the inferential analysis of the model. We propose estimators of price, risk and forecasted prices and for each estimator we demonstrate that they are strongly consistent and that properly centralised and normalised they converge in distribution to normal random variables, then we also give the interval estimators. An application that demonstrate the practical implementation of methods and results to real dividend data concludes the paper.

In the second paper, we propose a dividend stock valuation model where multiple dividend growth series and their dependencies are modelled using a multivariate Markov chain. Our model advances existing Markov chain stock models. First, we determine assumptions that guarantee the finiteness of the price and risk as well as the fulfilment of transversality conditions. Then, we compute the first and second

order price-dividend ratios by solving corresponding linear systems of equations and show that a different price-dividend ratio is attached to each combination of states of the dividend growth process of each stock. Subsequently, we provide a formula for the computation of the variances and covariances between stocks in a portfolio. Finally, we apply the theoretical model to the dividend series of three US stocks and perform comparisons with existing models.

In last paper, we propose a new measure to establish price leadership among multiple related price series using a Multivariate Markov Chain modelled through a Mixture Transition Model. This new measure, the Price Leadership Share (PLS), can easily process more than two price series simultaneously, offering an advantage over the existing price discovery measures. In addition, we propose a leadership concentration index for comparative analysis. An application to six gold contracts, including spot, futures, and ETF, over a 2-year period, shows that gold futures contracts, mainly CME contract, have a major role in price discovery confirming previous literature's findings. Besides, the PLS results are provided for different settings of the model parameters to test the validity of the model. Overall, our results show how the Price Leadership Share overcomes the limits of other price discovery measures.

A Python implementation of the three Markov chain applications is reported in a separate chapter with description of the procedures and routines useful to replicate the analysis.

Acknowledgements

I would like to thank my supervisor at the University of Wollongong, Professor Alex Frino for his mentorship and for giving me the opportunity to work in an international and diverse environment. I extend my gratitude to my co-supervisor, Professor Terry Walter, for his useful comments and suggestions, as well as his thorough revision of this manuscript. I also thank Professor Vito Mollica for his constant advice.

In addition, I would like to thank my supervisor at the University of Chieti-Pescara, Professor Guglielmo D'Amico, for introducing me to the world of the Markov chains and for a continuous and thorough guidance over the three years of my candidature.

I gratefully acknowledge the financial support provided by the International Postgraduate Tuition Awards (IPTA) of the University of Wollongong and my industry sponsors, the Australian Securities Exchange and the Market Quality Dashboard from the Capital Market Cooperative Research Centre (CMCRC). Finally, I would like to extend my gratitude to all academics and fellows at the CMCRC for interesting discussions and continual support.

List of Papers

The following papers are included as part of this thesis:

1. Barbu, V.S., D'Amico, G., and De Blasis, R., 2017, 'Novel advancements in the Markov chain stock model: Analysis and inference', *Annals of Finance*, vol. 13, no. 2, pp. 125–152. <https://doi.org/10.1007/s10436-017-0297-9>.
2. D'Amico, G. and De Blasis, R., 2018, 'A multivariate Markov chain stock model'. Available at SSRN: <http://ssrn.com/abstract=3296253>.
Under review in the *Scandinavian Actuarial Journal*.
3. De Blasis, R., 2018, 'A new measure of price discovery in financial markets'. Available at SSRN: <http://ssrn.com/abstract=3297869>.
Presented to *FMCG 2019* conference in Sydney and *INFINITI 2019* conference in Glasgow.
Accepted to *FMA 2019* conference in New Orleans.

Contents

Abstract	i
Acknowledgements	iii
List of Papers	iv
List of Figures	viii
List of Tables	ix
1 Introduction	1
1.1 Stock valuation	1
1.2 Price discovery	4
1.3 Summary	6
2 Literature Review	8
2.1 Markov Chain Model	8
2.1.1 Discrete-Time Markov chain model	9
2.1.2 Multivariate Discrete-Time Markov chain model	13
2.2 Dividend Discount Models	14
2.2.1 General Model	15
2.2.2 Gordon Growth Model and Extensions	17
2.2.3 Markov chain stock models	22
2.2.4 Research questions	27
2.3 Price Discovery	29

2.3.1	Informational efficiency	29
2.3.2	Measures of contribution to price discovery	32
2.3.3	Research questions	40
2.4	Summary	42
3	Novel advancements in the Markov chain stock model	44
3.1	The Markov chain dividend valuation model	44
3.1.1	The computation of the moments	46
3.2	The inferential analysis	50
3.2.1	Estimation of a Markov chain	50
3.2.2	Estimation of the financial quantities	52
3.3	Application to real dividend data	59
3.4	Conclusion	65
4	A multivariate Markov chain stock model	67
4.1	Model	67
4.1.1	The Basic Dividend Valuation Model	68
4.1.2	A Multivariate Markov Model of Value	71
4.2	Empirical application	78
4.3	Conclusion	87
5	A new measure of price discovery in financial markets	88
5.1	The Multivariate Markov Chain model	88
5.1.1	Discrete-time Markov chain	89
5.1.2	The Multivariate Markov process	90
5.2	Price Leadership Share	92
5.3	Empirical application	95
5.3.1	Data	95
5.3.2	Parameters estimation	98
5.3.3	Robustness tests	106
5.4	Conclusion	107

6	Algorithmic implementation	110
6.1	Introduction	110
6.2	Markov chain parameters estimation	111
6.3	Markov stock model implementation	116
6.4	Price Leadership Share implementation	123
7	Conclusion	125
A	Appendix: proofs	128
	Bibliography	153

List of Figures

2.1	Example of a Transition Graph	10
2.2	Expected growth in a 3-stage dividend discount model	20
3.1	Discretization result for five states	60
3.2	Dividend forecast	62
3.3	RMSE values for one step forecast	63
3.4	Stability test for every state.	64
3.5	RMSE values for one-step and all step prediction for different time ranges of analysis	66
4.1	Yearly dividend series	79
4.2	Simulation of portfolio variance	86
5.1	Discretisation of price log returns	99
5.2	Price Leadership Share	103
5.3	Weekly leadership concentration	104
5.4	Price Leadership Share Robustness	108

List of Tables

3.1	Descriptive statistics of the dividend series	59
3.2	Dividend growth discretization with five states	60
3.3	Transition probability matrix with five states	60
3.4	Price dividend ratio and second order moment values for the five states	63
3.5	Statistics related to the stability test.	64
4.1	Descriptive statistics of yearly dividends and yearly dividend returns from 1987 to 2018.	80
4.2	GPC, DOV, and PH betas and CAPM estimated rate of returns based on stock data from 1987 to 2018.	80
4.3	Dividend growth discretisation with 3 states and corresponding state values.	81
4.4	Transition matrices of the multivariate Markov stock model	82
4.5	Estimation of $\lambda_{\beta,\alpha}$	83
4.6	Price-dividend ratios ψ_1	84
4.7	Second order price-dividend ratio ψ_2	85
4.8	Variance-covariance matrix for GPC, DOV, and PH.	85
4.9	Price-dividend ratio calculated with the univariate Markov stock model in Barbu et al. (2017).	86
4.10	Stock valuation comparison	87
5.1	Market quality and efficiency metrics for six gold contracts	97

5.2	Example of transition probability matrices for the CME gold futures price series	101
5.3	Matrix of price influences for six gold contracts	102
5.4	Price Leadership Share	103
5.5	PLS comparison for three different time ranges	105
5.6	Robustness analysis of the PLS measure over different sampling intervals	106
6.1	Example of transition probabilities as output of the function in Listing 3	113
6.2	Example of frequencies dataframe as element of the output dictionary from function in Listing 4	114
6.3	Sample of the <i>combiP</i> output.	119

List of Listings

1	Preliminary imports.	110
2	Discretisation procedure.	111
3	Transition probabilities.	112
4	Computation of the frequencies for the $\lambda_{\beta,\alpha}$ estimation.	114
5	Estimation of $\lambda_{\beta,\alpha}$ through MLE.	115
6	Function <i>get_edges</i> adapted for the dividend growth series.	116
7	Computation of P, Pg, and Pg_sq.	118
8	Computation of prodP, prodPg, and prodPg_sq.	119
9	Solution of linear systems of equations to obtain ψ_1 and ψ_2	120
10	Solution of Theorem 4.14.	122
11	Computation of the PLS and concentration index.	123

1 Introduction

This thesis presents three studies on the application of Markov chain models in finance. The focus is on two major financial topics, namely stock valuation and price discovery. In the first two applications, we advance existing literature on dividend stock models, proposing general frameworks for Markov stock valuations, in both univariate and multivariate settings, and providing empirical applications on real dividend data. The last study proposes a new measure of price discovery, advancing the debate on existing measures and providing useful tools for analysing multiple price series and comparing results in a different context. An application to six gold contracts confirms the validity of the measure.

This dissertation is complemented with an introduction to Markov chain models and a comprehensive review of models in both topics where we identify gaps and pose questions. In addition, we describe an algorithmic implementation of proposed models using Python programming language. In the remainder of this chapter, we introduce the problems in both stock valuation and price discovery.

1.1 Stock valuation

One of the main approaches to calculating the value of a share of stock is to consider the discounted value of future cash flows of that asset. The obtained value is called the intrinsic value or fundamental price of the stock. In general, the future cash flows are represented by dividends, i.e., profits distributed to the shareholders. Some of the most influential research articles in corporate finance are based on this idea (see, e.g., Williams, 1938, Gordon and Shapiro, 1956, Gordon, 1959, 1962).

Considerable research has been dedicated to improving the Gordon model, which assumes the dividend growth rate is constant. This model has been frequently criticised for this assumption and, consequently, many variants have been suggested in the literature. These variants impose ever less constraining assumptions on the dividend process. For example, Brooks and Helms (1990), Barsky and De Long (1993) consider multistage models with dividend growth rates changing deterministically among the stages. The hypothesis of a deterministic nature of dividends was abandoned in favour of more realistic probabilistic assumptions. Models based on Markov chains were proposed in Harrison and Kreps (1978), Hurley and Johnson (1994, 1998), Yao (1997), Ghezzi and Piccardi (2003) and, in general, regime switching in the dividend process were advanced in Gutiérrez and Vázquez (2004). The results of these articles were integrated into a semi-Markovian framework as provided by D’Amico (2013) where the semi-Markov hypothesis was advanced and validated on real data. A further generalisation is given in D’Amico (2017) where a continuous state space semi-Markov model is considered for the computation of the fundamental price and risk of the stock.

In a recent paper (Agosto and Moretto, 2015) a formula for the variance of the fundamental price was obtained in a multinomial-based model of dividend dynamics. However some open problems remain and in the first paper of this thesis (Barbu, D’Amico, and De Blasis, 2017), discussed in Chapter 3, we provide effective answers to these. First, we extend the computation of the risk to the Markov chain model that incorporates the multinomial setting. Second, we determine a specific assumption that guarantees that the risk can be computed by means of a convergent series. The assumption also assures that a transversality condition for the risk process holds true and therefore that the risk depends only on the randomness in the dividend process (the fundamental variable) and is not related either to the future value of the price process or to the future risk of the stock.

Additionally, we develop non-parametric techniques for the estimation of the parameters of the model, and we obtain point and interval estimators of the price,

risk and price-dividend ratio. Statistical analysis of the dividend discount model is surprisingly missing in the finance literature and, therefore, we also fill this gap. More precisely, we obtain estimators of the first and second order moments of the price process and the forecasted fundamental price; these estimators are shown to be consistent, and we also derive their asymptotic distribution. It is worth noting that the development of these statistical procedures represents a crucial step for the application of this type of model to real data. Both theoretical and statistical results are applied in a numerical experiment based on data freely available at <http://www.econ.yale.edu/shiller/data.htm>, collected by Robert Shiller.

Chapter 4 discusses the second paper of this thesis D'Amico and De Blasis (2018) that extends the Markov stock model to a multivariate setting. The rationale is that two or more stocks may be correlated. Thus, it is important to be able to consider this effect in the valuation procedure of the stock or the portfolio of stocks. In a recent paper, Agosto et al. (2018) computed the covariance between two stocks that may be held in a portfolio by an investor. They considered a Markov chain with state space equal to the set of possible couples of growth-dividend values for both stocks. However, this strategy cannot be easily implemented in real applications, especially when we introduce dependencies between more than two stocks as the number of parameters to estimate increases drastically.

In Chapter 4, we propose an extension of the Markov stock valuation model to consider an effective multivariate model with multiple dividend series. To reduce the number of parameters to estimate, we model the dependencies between dividend growth series using the mixture transition distribution model, that was first introduced by Raftery (1985) for modelling high order Markov chains and extended by Ching and Ng (2006) to multivariate Markov models. To perform the multivariate valuation, first, we determine assumptions that guarantee that the price process, as well as the risk process, can be expressed by convergent series depending only on the dividend process. At the same time, the satisfaction of the transversality condition avoids the presence of speculative bubbles. Second, we compute the first and second

order price-dividend ratios by solving corresponding linear systems of equations. We show that a different price-dividend ratio is attached to each combination of states of the growth process of each stock. In addition, we propose formulas for the computation of the variances and covariances between stocks that can be used for portfolio selection and valuation purposes. Finally, we apply our theoretical results to three US stocks with a long history of dividend payments and compare results with univariate valuation models.

1.2 Price discovery

Information has a fundamental role in the formation of prices in secondary markets. Understanding how prices efficiently incorporate information about the fundamental value has been, and still is, one of the main interests of the market microstructure literature. Moreover, with the increase in market fragmentation, trading the same asset in different venues becomes possible. The speed at which related prices incorporate news is another fundamental variable that needs to be considered.

This process is called price discovery, and there are currently three measures that are widely used in the literature. They all base their analysis on structural models of cointegrated price series that share a common random-walk efficient price. The first metric aims at quantifying how much of the variance of the efficient price can be attributed to the different markets. Hasbrouck (1995) refers to this proportion as the Information Share (IS). The second measure is the Component Share (CS), applied by Booth et al. (2002), Chu et al. (1999) and Harris et al. (2002) on the basis of the econometric work of Gonzalo and Granger (1995). It focuses on the decomposition of the price series into a permanent component, that reflects the contribution of the efficient price, and a transitory component, that represents the deviation from the efficient price due to market microstructure frictions.

Following a recent debate about the two measures and their interpretation, Putniņš (2013) developed a third measure based on the insight of Yan and Zivot (2010). If price series have different levels of noise, information share and compo-

ment share result in different levels of leadership and relative avoidance of noise. To eliminate this misinterpretation, the Informational Leadership Share combines the two metrics, so the dependence on the level of noise cancels out.

All existing price discovery measures base their analysis on a permanent-transitory decomposition of innovations, modelling prices with structural equation models and with the assumption that cointegrated prices follow a common efficient price. The general idea is to understand how much of the long-run information is due to one market or another. However, these models present some limitations, for example, there is no clear consensus in what information share and component share really measure, and the informational leadership share can only measure price discovery between two price series, with only one permanent and one transitory shock and when the errors are uncorrelated. Therefore, we propose a new measure of price discovery modelling price changes of dependent price series with a discrete-time multivariate Markov chain.

While in all approaches prices track a common efficient price with different reaction times and levels of noise, we focus on how prices influence the formation of other prices. We argue that the fastest price to reflect new information releases a price signal to the other slower price series. Once there is new information, prices follow other observable and faster-adjusting prices. We are not able to directly observe the variation in the fundamental price, but we can estimate dependencies between price series based on observation of price changes. In this paper, we measure the portion of influence that a price series has on the others using a multivariate Markov chain to model the dynamics between price return series. Summarising the price influences, we define a measure that we call Price Leadership Share (PLS). Hasbrouck (1995) interprets the Information Share as “who moves first” in the process of price adjustment, with our Price Leadership Share we aim at finding “who is the price leader” among all the price series.

In Chapter 5, the model is tested on the price return series of six gold contracts around the globe, including one spot, four futures and one ETF contract. Our results

show a strong price leadership from CME gold futures contract on all the other contracts and a weaker leadership of the SGE spot contract. All the futures contracts analysed together with the ETF contract are quicker in their price adjustment than the spot, in accordance with the existent literature on price discovery (Bohl et al., 2011, Rosenberg and Traub, 2009, Hauptfleisch et al., 2016). We test our model under different combinations of parameters settings, e.g. increasing the number of states of the Markov chain or the sampling frequency of the observation. Our results show a trade-off between computational effort and measure's precision, suggesting a 3-state Markov chain to model the price changes is appropriate. Moreover, increasing the sampling frequency generates a contemporaneous correlation between price innovation reducing the reliability of the measure (see, e.g., Hasbrouck, 1995). Hence, the lower the sampling interval, the better the price leadership measure.

In general, we find that our model presents some advantage over existing measures of price discovery. First, we can apply our model to more than two price series simultaneously and generate a ranking of price leadership. Second, the price leadership share measure is not dependent on lagged observation. Therefore, it can benefit from the entire set of observations, this being useful in the particular case of illiquid stocks.

1.3 Summary

The three studies in this thesis examine applications of Markov chain models to two crucial issues in finance, stock valuation and price discovery. This chapter introduced problems from both financial perspectives and motivations for the analysis presented in the following chapters.

The remainder of this thesis is structured as follows. Chapter 2 presents a brief introduction to the theory of the Markov chain models, from a univariate and a multivariate perspective. Then, two sets of literature analysis follow. First, we present a review of the dividend discount model and its variations, with a focus on the Markov chain stock models. Finally, we include an analysis of the literature on price discov-

ery in financial markets and its applications. Both reviews end with a formulation of research questions that introduce the problem discussed in the remaining chapters. The core discussion is presented in three separate chapters. Chapter 3 and Chapter 4 illustrate the dividend discount model extended by a Markov chain model, from univariate and multivariate perspectives, respectively. Chapter 5 introduces the application of multivariate Markov chain models to market microstructure, specifically price discovery. Chapter 6 presents an algorithmic implementation of methods and techniques proposed in this thesis using Python programming language. Complete routines are accompanied by a full description of functions and variables. The thesis closes with some concluding remarks about the three applications. All mathematical proofs are reported in the appendix.

2 Literature Review

2.1 Markov Chain Model

The Markov chain is a stochastic model that is useful to describe a series of events in which the probability of an event depends only on the state of the previous event. The model was first introduced in the early 20th century by the Russian mathematician Andrey Markov (Markov, 1906) and it has been widely used for predictions in many fields, like meteorology, physics, biology, chemistry, and social sciences. Applications in finance include the modelling of financial markets (Chu et al., 1999), stock valuation (Hurley and Johnson, 1994, Yao, 1997, Hurley and Johnson, 1998, Ghezzi and Piccardi, 2003, Barbu et al., 2017), and credit rating modelling (Jarrow et al., 1997, D’Amico et al., 2006, Vasileiou and Vassiliou, 2006, D’Amico et al., 2016). Some authors focus on the multivariate analysis, like Nicolau (2014) for financial applications, Ching et al. (2002) for modelling multiple categorical data in demand predictions, Siu et al. (2005) for analysing credit risk, and D’Amico and De Blasis (2018) for stock valuation.

This section reviews the discrete-time Markov chain model on the basis of the extensive work of Brémaud (1999). The first part describes a simple univariate model, followed by an analysis of the multivariate setting, modelled via a Mixture Transition Distribution (MTD) model that was first introduced by Raftery (1985) for high-order Markov chains to reduce the number of parameters.

2.1.1 Discrete-Time Markov chain model

A categorical time series can be described as a sequence of independent and identically distributed random variables $\{S_t\}_{t \geq 0}$ taking values in the countable set $\mathcal{M} = \{1, 2, 3, \dots, m\}$, that is the set of the possible states of our sequence. In this context, t represents *time*, but in general, it could represent a space or something else. If time t is discrete then the process is a discrete-time stochastic process and we call it a *Markov Chain* if for each t , and $j, i_0, \dots, i_t \in \mathcal{M}$

$$\mathbb{P}(S_{t+1} = j | S_t = i_t, S_{t-1} = i_{t-1}, \dots, S_0 = i_0) = \mathbb{P}(S_{t+1} = j | S_t = i_t). \quad (2.1)$$

The *Markov Property* (2.1) indicates that the probability of being in the state j at time $t + 1$ depends only on the state i_t occupied by the series at time t , regardless of the previous history. So, the future behaviour of a Markov is independent of the past states given the current state.

When this condition is time independent, the process is called a *Homogeneous Markov Chain* (HMC), and the probability

$$\mathbb{P}(S_{t+1} = j | S_t = i) = p_{ij}, \quad (2.2)$$

represents the probability to move from state i to state j at any point in time.

Considering all the possible combinations of changing from one state to another, in the set of states \mathcal{M} , and in accordance to Formula (2.2), we can build the matrix $\mathbf{P} = \{p_{ij}\}_{i,j \in \mathcal{M}}$ with m^2 elements, that is the *transition probability matrix* of the

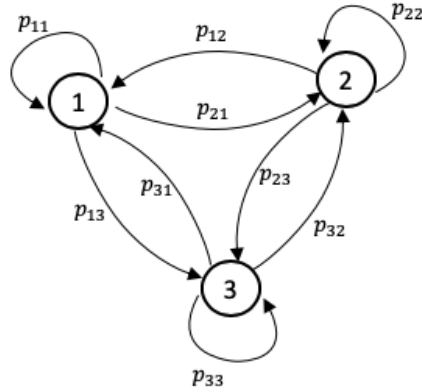


Figure 2.1: Example of a *Transition Graph* for a 3-state Markov chain. Nodes represent the states and edges have weights representing probabilities p_{ij} of the transition matrix.

HMC,

$$\mathbf{P} = \begin{array}{c} S_{t-1} \\ 1 \\ 2 \\ \vdots \\ m \end{array} \begin{array}{c} S_t \\ 1 \quad 2 \quad \dots \quad m \\ \left[\begin{array}{cccc} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{array} \right] \end{array}, \quad (2.3)$$

subject to

$$0 \leq p_{ij} \leq 1, \quad \forall i, j \in \mathcal{M}, \quad (2.4a)$$

$$\sum_{j=1}^m p_{ij} = 1, \quad \forall i \in \mathcal{M}. \quad (2.4b)$$

The matrix (2.3) is called a *stochastic matrix* because each element is a probability and every change from a state i must end in a state j .

A transition matrix can also be represented by a *Transition Graph*. The graph has each state represented by a node, and each probability by a link (or edge) that is oriented from state i to state j , according to the respective probability p_{ij} . Figure 2.1 is an example of a 3-state Markov chain with $\{p_{ij}\}_{i,j \in \mathcal{M}}$ transition probabilities.

A Markov chain is fully defined when we know the initial state with its probability distribution and the transition probability matrix.

Let

$$\mathbf{v}_t := [v_{t,1}, \dots, v_{t,m}], \quad (2.5)$$

be the probability vector where $v_{t,i} := \mathbb{P}(S_t = i)$ is the probability of being in state i at time t , with $i \in \mathcal{M}$, then

$$\mathbf{v}_{t+1} = \mathbf{v}_t \mathbf{P}, \quad (2.6a)$$

$$\mathbf{v}_t = \mathbf{v}_0 \mathbf{P}^t, \quad (2.6b)$$

where \mathbf{v}_0 is the initial distribution and \mathbf{P}^t is the t -step transition matrix, and in general, it is the t -th power of \mathbf{P} .

If the Markov property (2.1) is not independent of time, the Markov chain is called an *inhomogeneous Markov chain*, and there will be a transition matrix for each time step, $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_t$. Thus, relations (2.6) become

$$\mathbf{v}_{t+1} = \mathbf{v}_t \mathbf{P}_{t+1}, \quad (2.7a)$$

$$\mathbf{v}_t = \mathbf{v}_0 \mathbf{P}_1 \mathbf{P}_2 \dots \mathbf{P}_t. \quad (2.7b)$$

To study the asymptotic behaviour of the Markov chain, it is worth mentioning some useful concepts, that are irreducibility, aperiodicity, and stationary distribution.

A Markov chain, with its transition matrix, is said to be *irreducible* if every state communicates with all other states. If the Markov chain is homogeneous, a state i *communicates* with state j , if there exists $t \in \mathbb{N}$ such that

$$p_{ij}^{(t)} > 0, \quad (2.8)$$

where $p_{ij}^{(t)}$ is the element of the i -th row and j -th column of the matrix \mathbf{P}^t .

The communication property expressed in (2.8) can be denoted by $i \rightarrow j$ if the communication is bi-directional, then it is denoted by $i \leftrightarrow j$.

In general, if for all $i, j \in \mathcal{M}$ we have $i \leftrightarrow j$, then the Markov chain is *irreducible*,

otherwise is said to be *reducible*.

Let us define the *period* d_i of a state $i \in \mathcal{M}$ as

$$d_i = \gcd\{t \geq 1; p_{ii}^{(t)} > 0\}, \quad (2.9)$$

where $\gcd\{a_1, a_2, \dots\}$ is the greatest common divisor of a_1, a_2, \dots , and with $d_i = +\infty$ if there is no $t \geq 0$ with $p_{ii}^{(t)} > 0$.

Equation (2.9) indicates that the period is the greatest common divisor of the set of times that the model returns to state i , i.e., there is a probability of returning to state i . Then, if $d_i = 1$, then the state i is aperiodic, and if all states are aperiodic, then the Markov chain is *aperiodic*.

A row vector $\boldsymbol{\pi} = (\pi_1, \dots, \pi_m)$ is called a *stationary distribution* of the Markov chain, if

- (i) $\pi_i \geq 0$ for $i = 1, \dots, m$ and $\sum_{i=1}^m \pi_i = 1$, and
- (ii) $\boldsymbol{\pi} \mathbf{P} = \boldsymbol{\pi}$.

The last conditions mean that $\boldsymbol{\pi}$ is a left-eigenvector of the matrix \mathbf{P} with eigenvalue equal to 1.

In general, if (S_0, S_1, \dots) is a Markov chain, it is interesting to study its asymptotic behaviour, or more specifically understand what happens to the distribution of S_t when $t \rightarrow \infty$. Therefore, the analysis of the asymptotic behaviour is based on (i) the *existence* of the stationary distribution, (ii) the *uniqueness* of the stationary distribution, and (iii) the *convergence* to stationarity starting from an initial distribution.

If a Markov chain is irreducible and aperiodic, then it admits only one stationary distribution $\boldsymbol{\pi}$, and the distribution \mathbf{v}_t of the chain at time t approaches $\boldsymbol{\pi}$ as $t \rightarrow \infty$, regardless of the initial distribution \mathbf{v}_0 ,

$$\lim_{t \rightarrow \infty} p_{ij}^{(t)} = \pi_j \quad \forall i, j \in \mathcal{M}. \quad (2.10)$$

2.1.2 Multivariate Discrete-Time Markov chain model

The previous model can be extended to a multivariate setting, with more than one time series.

For every series in $\Gamma = \{1, 2, \dots, \gamma\}$, the probability of being in state j depends on the state i_1, \dots, i_γ occupied by all the available series one time step before. The Markov Property in (2.1) becomes:

$$\begin{aligned} \mathbb{P}[S_{t+1}^{(\alpha)} = j | (S_t^{(1)} = i_t^{(1)}, S_{t-1}^{(1)} = i_{t-1}^{(1)}, \dots, S_0^{(1)} = i_0^{(1)}), \dots, \\ (S_t^{(\gamma)} = i_t^{(\gamma)}, S_{t-1}^{(\gamma)} = i_{t-1}^{(\gamma)}, \dots, S_0^{(\gamma)} = i_0^{(\gamma)})] = \\ \mathbb{P}(S_{t+1}^{(\alpha)} = j | S_t^{(1)} = i_t^{(1)}, \dots, S_t^{(\gamma)} = i_t^{(\gamma)}), \end{aligned} \quad (2.11)$$

where $\alpha \in \Gamma$.

The new Property (2.11) shows that there are multiple dependencies between the series. Therefore, the transition probability matrix of the multivariate model must include each possible combination, m^γ , for the initial states, and every initial state must end in one of the possible final combinations. The result is $m^\gamma(m^\gamma - 1)$ total parameters to estimate for the multivariate Markov model, given that there are $m^\gamma - 1$ independent probabilities in each row. Such a configuration is not practical in a real-world application because the number of parameters will increase exponentially when the number of series and states increase.

Raftery (1985) proposed the Mixture Distribution Model (MTD) to reduce the number of parameters to estimate for high order Markov chains, and Ching et al. (2002) applied it to the multivariate Markov chains. A review of the MTD model and its application is available in Berchtold and Raftery (2002). Applying the MTD model the probability vector for series α at time $t + 1$ becomes

$$\mathbf{A}^{(\alpha)}(t + 1) = \sum_{\beta=1}^{\gamma} \mathbf{A}^{(\beta)}(t) \cdot \lambda_{\beta, \alpha} \cdot \mathbf{P}^{(\beta, \alpha)}, \quad (2.12)$$

where $\mathbf{A}^\alpha(t) := [A_1^{(\alpha)}, \dots, A_m^{(\alpha)}]$ and $A_i^{(\alpha)}(t) := Pr(S_t^{(\alpha)} = i)$.

According to this condition, we can build γ^2 transitions probability matrices $\mathbf{P}^{(\beta, \alpha)}$, each one containing the transition probabilities from state i in series β to state j in series α , with $\alpha, \beta \in \Gamma$,

$$\mathbf{P}^{(\beta, \alpha)} = \begin{array}{c} S_{t-1}^{(\beta)} \\ \begin{array}{c} 1 \\ 2 \\ \vdots \\ m \end{array} \end{array} \begin{array}{c} S_t^{(\alpha)} \\ \begin{array}{c} 1 \\ 2 \\ \dots \\ m \end{array} \end{array} \begin{bmatrix} p_{11}^{(\beta, \alpha)} & p_{12}^{(\beta, \alpha)} & \dots & p_{1m}^{(\beta, \alpha)} \\ p_{21}^{(\beta, \alpha)} & p_{22}^{(\beta, \alpha)} & \dots & p_{2m}^{(\beta, \alpha)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1}^{(\beta, \alpha)} & p_{m2}^{(\beta, \alpha)} & \dots & p_{mm}^{(\beta, \alpha)} \end{bmatrix}. \quad (2.13)$$

Parameters $\lambda_{\beta, \alpha}$ are the scalar weights that combine all the series, and are subject to:

$$\sum_{\beta=1}^{\gamma} \lambda_{\beta, \alpha} = 1, \quad (2.14)$$

$$\lambda_{\beta, \alpha} \geq 0. \quad (2.15)$$

The MTD model permits to reduce the total parameters to estimate from $m^\gamma(m^\gamma - 1)$ to $\gamma^2 m(m-1) + \gamma(\gamma-1)$, the first addend being the number of $p_{mm}^{(\beta, \alpha)}$ parameters and the second the number of weights $\lambda_{\beta, \alpha}$.

2.2 Dividend Discount Models

This section presents a review of the dividend discount models starting from basic models (Williams, 1938, Gordon and Shapiro, 1956) to more complex models (Brooks and Helms, 1990, Barsky and De Long, 1993). Extensive reviews of stock valuation methods can be found in Damodaran (2012) and Kamstra (2003). A particular focus is given to models that make use of the Markov chain (Hurley and Johnson, 1994, 1998, Yao, 1997, Ghezzi and Piccardi, 2003), as they represent the base of this

research. The section concludes with the identification of the open problems that, eventually, are addressed in Chapters 3 and 4. In the remainder of this section, the mathematical notation is slightly different from the previous analysis. The time is reported in parenthesis, for example, a price at time t is indicated as $\mathcal{P}(t)$ instead of \mathcal{P}_t .

2.2.1 General Model

Stock valuation is one the basic aspects of financial markets. Discussions about the fair price of a stock, or its overpricing and underpricing, have always been of paramount importance to investors. Williams (1938) was the first to recognise that market prices and fundamental values are “separate and distinct things not to be confused”. In his work, he states that an asset’s intrinsic long-term value is the present value of all future cash flows, i.e., dividends and future selling price.

Let $P(t)$ be the random variable giving the fundamental value of a stock at time $t \in \mathbb{N}$. Let $D(t)$ be the dividend at time $t \in \mathbb{N}$, also assumed to be a random variable, and denote by $k_e(t)$ the required rate of return on the stock at time t . If we buy a stock at time t and plan to sell it at time $t + 1$, the price $p(t) := \mathbb{E}_{(t)}[P(t)]$ that we pay is the expected value of the stock price at time $t + 1$ plus the cash flows distributed by the company, all discounted at the appropriate measure of risk $k_e(t)$,

$$p(t) = \mathbb{E}_{(t)} \left[\frac{P(t+1) + D(t+1)}{1 + k_e(t)} \right], \quad (2.16)$$

If we buy and hold the stock indefinitely, and assuming (see, e.g., Samuelson, 1973)

$$\lim_{i \rightarrow +\infty} \mathbb{E}_{(t)} \left[\frac{P(t+i)}{\prod_{j=0}^i [1 + k_e(t+j)]} \right] = 0, \quad (2.17)$$

then the price we pay is the expected value of all future cash flows in the form of dividends,

$$p(t) = \sum_{i=0}^{+\infty} \mathbb{E}_{(t)} \left[\frac{D(t+i+1)}{\prod_{j=0}^i [1+k_e(t+j)]} \right]. \quad (2.18)$$

If condition (2.17) is not assumed, then Blanchard and Watson (1982) proved that there could exist different solutions of the fundamental equation, i.e., there is the presence of bubbles in the stock market.

To solve equation (2.18), we have to identify the first input, namely future dividends. Because of the impossibility of making predictions of dividends through to infinity, many models make assumptions about the dividend growth, based on the expectation of growth rate of earnings and payout ratios, or apply specific stochastic processes to forecast dividends. The next two sections explain how the various models make assumptions about the dividend growth.

The other input of the equation is the discount factor $k_e(t)$, or cost of equity, that represents a measure of the asset's riskiness, that in most of the dividend discount models is assumed to be constant k_e . Traditionally, the estimation of k_e has been performed using the Capital Asset Pricing Model (CAPM). The model originates from the idea of mean-variance efficient portfolio of Markowitz (1952), and it is formalised by Sharpe (1964) and Lintner (1965) and extended by Black (1972). The rationale of the model is that risky investments R_i , for example, stocks in financial markets, are expected to be more remunerating than the risk-free assets

$$\mathbb{E}[R_i] = R_f + \beta_{im}(\mathbb{E}[R_m] - R_f), \quad (2.19)$$

$$\beta_{im} = \frac{Cov[R_i, R_m]}{Var[R_m]}, \quad (2.20)$$

where R_m is the return on the market portfolio, and R_f is the return on the risk-free asset. The Black (1972) version substitutes the risk-free rate with a zero-beta portfolio uncorrelated with the market. The coefficient β_{im} represents the correlation of the stock with the market, and can be estimated as slope coefficient of the OLS regression

$$Z_{it} = \alpha_{im} + \beta_{im}Z_{mt} + \epsilon_{it}, \quad (2.21)$$

where Z_{it} is the excess return of the stock on the risk-free asset, or equity premium, and Z_{mt} is the market risk premium, $\mathbb{E}[R_i] - R_f$. In practice, the market return and the risk-free rate are proxied by a market index, e.g., S&P 500 Index, and government treasury bonds, respectively, over a period of time that generally extends to about five years of historical data (Campbell et al., 1997). Many authors provides empirical evidence on the CAPM application (see, e.g., Jensen et al., 1972, Fama and MacBeth, 1973, Blume and Friend, 1973, Basu, 1977, Fama and French, 1992, 1993), while Roll (1977) criticise it because the market portfolio is not observable and therefore the model is not testable. For a comprehensive description of the CAPM models and its variations with econometrics analysis see, e.g., Campbell et al. (1997), Cochrane (2009).

In general, the dividend discount model is very attractive because it is intuitive and easy to implement. Nevertheless, it encounters much criticism because of the limits it poses. The main argument is the applicability of the model only to certain firms with stable, high-paying dividend policy. Moreover, firms recent practice is to perform share buybacks instead of paying dividends, for obvious tax reasons, reducing the dividend cash flow and resulting in an underestimation of the value of the firm. The same principle applies to other assets that are ignored in the model, e.g., the value of brand names. However, share buybacks and values of other assets can be included in the dividends flow and treated as such with adequate adjustments (see, e.g., Damodaran, 2012).

2.2.2 Gordon Growth Model and Extensions

Equation (2.18) can be rewritten in terms of dividend growth, defining

$$g(t) = \frac{D(t+1) - D(t)}{D(t)}, \quad (2.22)$$

as the growth rate of dividends from time t to time $t + 1$, so that $D(t + 1) = D(t)(1 + g(t))$ and $D(t + 2) = D(t)(1 + g(t))(1 + g(t + 1))$. Then, the price becomes,

$$p(t) = D(t) \sum_{i=0}^{+\infty} \mathbb{E}_{(t)} \left[\prod_{j=0}^i \frac{1 + g(t+j)}{1 + k_e(t+j)} \right]. \quad (2.23)$$

Assuming a constant dividend growth rate $g(t+j) = g$ and a constant discounting factor $k_e(t+j) = k_e$, equation (2.23) becomes

$$p(t) = D(t) \sum_{i=0}^{+\infty} \frac{(1+g)^i}{(1+k_e)^i}, \quad (2.24)$$

and summing the geometric progression, we obtain the *Gordon fundamental price estimate* (Gordon, 1962)

$$p^G(t) = D(t) \frac{1+g}{k_e-g}, \quad \text{or} \quad p^G(t) = \frac{D(t+1)}{k_e-g}, \quad (2.25)$$

with the constraint $g < k_e$ to obtain a finite price.

The Gordon model is straightforward because it requires only estimates of the dividend growth rate and discount rate, that are both easily obtained from a company's historical data. Nevertheless, it has some limitations. The model can result in incorrect estimations of the price when the growth rate approaches the discount rate, as the price tends to grow up to infinity. Therefore, this model is more suitable for companies with a stable dividend policy with a growth that is less than the growth of the economy. Moreover, empirical applications of the Gordon model show that dividends tend to grow exponentially, meaning that a linear growth model is not suitable for the stock valuation (see, e.g., Campbell and Shiller, 1987, West, 1988).

The assumption of constant growth of the dividends forever is not realistic. To relax this assumption, Malkiel (1963) introduces a 2-stage model, with the first period of n years of extraordinary growth followed by a stable growth forever. The value of a stock can be obtained as the sum of first years values, calculated from the general model plus a discounted value of the Gordon growth model at year n :

$$p^{2st}(t) = \mathbb{E}_{(t)} \left[\sum_{i=0}^n \frac{D(t+i+1)}{\prod_{j=0}^i [1+k_e(t+j)]} + \frac{P^G(n)}{\prod_{j=0}^n [1+k_e(t+j)]} \right], \quad (2.26)$$

where $P^G(n)$ is the Gordon growth fundamental price estimate (2.25) at year n .

A further assumption of constant growth in the first phase, g_h , and constant discount rate $k_{e,h}$, simplifies equation (2.26) to

$$p^{2st}(t) = \frac{D(t)(1 + g_h) \left[1 - \frac{(1+g_h)^n}{(1+k_{e,h})^n} \right]}{k_{e,h} - g} + \frac{P^G(n)}{(1 + k_{e,h})^n}, \quad (2.27)$$

This model is suitable for valuing companies that expect to have an initial growth period higher than normal, because of a specific investment or a patent right, that will result in higher profits. At the same time, it presents some limits. First, the growth rate is expected to drop drastically from high to normal level, and second, it is hard to define the length of the high growth period in practical terms.

To avoid the sharp drop from high to stable growth rate, Fuller and Hsia (1984) propose a linear decline of the growth in their ‘‘H’’ model. The high growth phase with decline is assumed to last $2H$ periods up to the stable growth phase g_n , with an initial growth rate g_a . The model assumes that the discount rate k_e is constant over time, as well as the dividend payout ratio.

$$p^H(t) = \frac{D(t)(1 + g_a)}{k_e - g_n} + \frac{D(t)H(g_a - g_n)}{k_e - g_n}, \quad (2.28)$$

A constant payout ratio assumption poses some limits to this model. Generally, a company is expected to have lower payout ratios in high growth phases and higher payout ratios in the stable growth phase, as shown in Figure 2.2.

A 3-stage model, initially formulated by Molodovsky et al. (1965) and derived from a combination of the H model and the 2-stage model, with the inclusion of a variable payout policy and different discount factors for the various phases, overcomes the limits of previous models, but it requires a larger number of inputs. Let $k_{e,h}$, $k_{e,d}$, and $k_{e,st}$ be the discount factors for high, declining, and stable phases, respectively. Let g_a and g_n be the growth rate at the beginning and the end of the period. Let EPS be the earnings per share, and Π_a and Π_n the payout ratios at the beginning and end of the period, respectively. The stock valuation for the 3-stage

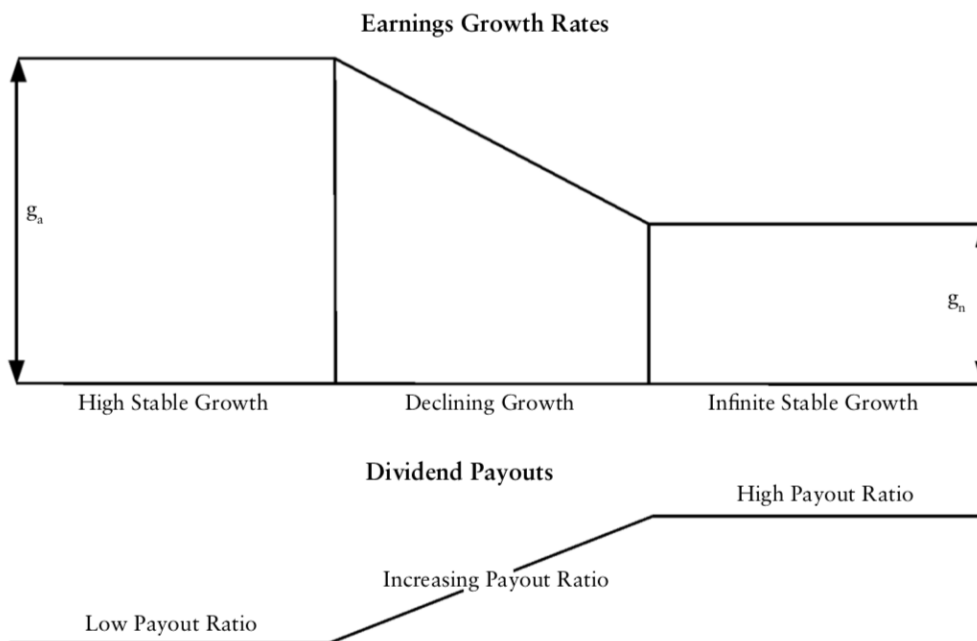


Figure 2.2: Expected growth in a 3-stage dividend discount model (Figure from Damodaran, 2012)

model is

$$\begin{aligned}
 p^{3st}(t) = & \sum_{i=0}^{n1} \frac{EPS(t)\Pi_a(1+g_a)^i}{(1+k_{e,h})^i} + \sum_{i=n1+1}^{n2} \frac{D(t+i)}{(1+k_{e,d})^i} \\
 & + \frac{EPS(t+n2)\Pi_n(1+g_n)}{(k_{e,st}-g_n)(1+k_{e,h})^{n1}(1+k_{e,d})^{n2-n1}},
 \end{aligned} \tag{2.29}$$

An empirical comparison of the Gordon model and its variations is in Sorensen and Williamson (1985). The authors analyse the intrinsic value of a random sample of 150 firms from the S&P 400 using data available in 1981, from four different valuation models, price/earning model, constant growth model, two-period, and three-period model. They base the analysis on normalised earnings and a dividend payout ratio of approximately 45 per cent. The discount factor is calculated using the CAPM model for the growth period, according to the beta of the stock and the high growth period is assumed to last five years for all the stock. Then, based on the assumption that all mature firms look alike, an equal risk measure of 8% among all the stocks is adopted for the stable phase.

For every model, the authors generate five portfolios of 30 stocks each, ordered

from undervalued to overvalued securities, estimating returns for two years. Results show that the increased complexity of the model improves the annualised returns. As well as looking at the risk characteristics of the portfolios, the 3-stage model outperforms the other model.

Brooks and Helms (1990) generalise the 2-stage model from Malkiel (1963). They propose an N-stage model, with quarterly dividends and fractional periods. Within each stage, dividends growth is assumed constant, and the discount rate is based on quarterly compounding $r_e = (1 - k_e)^{\frac{1}{4}} - 1$. They test the model on the case of Commonwealth Edison Company (CWE), an electricity supplier, estimating the required rate of return for three cases: (a) annual dividends, no fractional period; (b) quarterly dividends, no fractional periods; and (c) quarterly dividends, fractional periods. They show that ignoring quarterly compounding and fractional periods the results present a downward bias.

Another extension of the Gordon growth models is given in Barsky and De Long (1993). The authors propose to model the permanent dividend growth as a geometric average of past dividend changes:

$$g(t) = (1 - \theta) \sum_{i=0}^t \theta^i \Delta D(t - i) + \theta^t g(0) \quad (2.30)$$

with $g(t)$ following a random walk process and, thus, change in dividends following an IMA(1,1).

Donaldson and Kamstra (1996) generalise the Gordon growth model allowing for arbitrary dividend growth and discount rates. Their methodology involves a Monte Carlo simulation and numerical integration of the random joint process of dividend growth and discount rates

$$y(t + 1) = \frac{1 + g(t + j)}{1 + k_e(t + j)}. \quad (2.31)$$

They forecast a range of possible evolution of the process $y(t + 1)$ up to a certain point in the future, $t + I$, and calculate the average of several estimations of the

present stock value

$$p(t) = D(t) \sum_{i=0}^I \prod_{j=0}^i y(t+1). \quad (2.32)$$

2.2.3 Markov chain stock models

According to equation (2.23), the stock valuation is obtained through two inputs, namely the dividend growth and the discount factor. The idea of the Markov chain stock models is to describe the dividend growth rate as a sequence of independent, identically distributed, discrete random variables, and model it as a Markov process. In all these models, the discount factor k_e is kept constant.

Hurley and Johnson (1994) model the dividend growth as a *Markov dividend stream*. They assume that in each period the dividend can increase with probability q , be the same with probability $1 - q$, to resemble a step pattern in the long term. Moreover, they include the possibility for the firm to go bankrupt, with probability q_B . They propose two variations of the model, an additive model and a geometric model, both giving an estimation of the value, along with a lower bound estimation for each of these values.

In the *additive model*, the dividend at time $t + 1$ increase by the amount Δ with probability q , and assuming a constant discount rate k_e , the value of the firm is

$$p(t) = \begin{cases} D(t) + \Delta + p(t+1) \frac{D(t)+\Delta}{1+k_e} & \text{with prob } q \\ D(t) + p(t+1) \frac{D(t)}{1+k_e} & \text{with prob } 1 - q - q_B \\ 0 & \text{with prob } q_B \end{cases}, \quad (2.33)$$

and the closed form solutions for the value and the lower bound are

$$p^A(t) = \frac{D(t)}{k_e} + \left[\frac{1}{k_e} + \frac{1}{k_e^2} \right] q \Delta, \quad (2.34)$$

and

$$p_{low}^A(t) = \frac{D(t)(1 - q_B)}{k + q_B} + \left[\frac{1}{k + q_B} + \frac{1}{(k + q_B)^2} \right] q \Delta. \quad (2.35)$$

Note that, when $q_b = 0$, $p_{low}^A = p^A$.

The *geometric model* assumes that the dividend increases with a growth rate g and with a probability q

$$D(t+1) = \begin{cases} D(t)(1+g) & \text{with prob } q \\ D(t) & \text{with prob } 1 - q - q_B \end{cases}. \quad (2.36)$$

The closed form solutions for the value and the lower bound become

$$p^G(t) = \frac{D(t)(1 + qg)}{k_e - qg}, \quad (2.37)$$

and

$$p_{low}^G(t) = D(t) \left[\frac{1 + qg - q_B}{k_e - (qg - q_B)} \right]. \quad (2.38)$$

It is worth noting that the geometric model reduces to the Gordon model, setting the expected growth rate to $qg - q_B$, or, if we exclude the possibility of bankruptcy, setting the expected growth rate to qg .

An empirical application to three stocks, provided in Hurley and Johnson (1994), shows that the geometric method performs well when the dividend series is erratic and does not always show increases. The model gives an estimation that is very close to the actual stock prices.

Hurley and Johnson (1998) formulate a generalised version of their model to include the possibility of a decrease in the dividends, so the dividend at time t is $D(t) = D(t-1) + \Delta_i$ for the additive model, and $D(t) = D(t-1)(1 + g_i)$ with probability q_i for the geometric model. Both Δ_i and g_i include the possibility of dividends reduction, or suspensions. Under the condition $q_0 + \sum_{i=1}^n q_i = 1$, the closed form solution for both models are

$$p^A(t) = \frac{D(t)}{k_e} + \left[\frac{1}{k_e} + \frac{1}{k_e^2} \right] \sum_{i=1}^n q_i \Delta_i, \quad (2.39)$$

and

$$p^G(t) = D(t) \frac{1 + \sum_{i=1}^n q_i g_i}{k_e - \sum_{i=1}^n q_i g_i}. \quad (2.40)$$

When $n = 1$, the models reduces to Hurley and Johnson (1994) models.

The same proposal for dividend reduction to extend Hurley and Johnson (1994) models is advanced by Yao (1997). The author introduces a trinomial dividend valuation model and extends the additive model, where the dividend at time $t + 1$ is

$$D(t+1) = \begin{cases} D(t) + \Delta & \text{with prob } q^u \\ D(t) - \Delta & \text{with prob } q^d \\ D(t) & \text{with prob } q^c = 1 - q^u - q^d \end{cases}, \quad (2.41)$$

with closed solution for the stock value

$$p^A(t) = \frac{D(t)}{k_e} + \left[\frac{1}{k_e} + \frac{1}{k_e^2} \right] (q^u - q^d) \Delta. \quad (2.42)$$

Then, the geometric model, with

$$D(t+1) = \begin{cases} D(t)(1+g) & \text{with prob } q^u \\ D(t)(1-g) & \text{with prob } q^d \\ D(t) & \text{with prob } q^c = 1 - q^u - q^d \end{cases}, \quad (2.43)$$

and closed solution

$$p^G(t) = D(t) \frac{1 + (q^u - q^d)g}{k_e - (q^u - q^d)g}. \quad (2.44)$$

Lower bounds for both models are also given by the author. Moreover, a practical

application on five firms, provided in Yao (1997), shows that the model produces better estimates than Hurley and Johnson (1994).

Ghezzi and Piccardi (2003) start from the previous Markov models to formulate a Markov chain stock model. The authors begin with a description of the simple model for the dividend growth rate using a 2-state discrete Markov chain, and a constant discount rate $r = 1 + k_e$. Finally, they extend the model to an n -state Markov chain and define a vector of price-dividend ratios as the solution of a system of linear equations.

In previous models, Hurley and Johnson (1994, 1998) and Yao (1997) assume that the dividend growth rates are independent, identically distributed, discrete random variables, thus obtaining one closed form solution irrespective of the state of the dividend. On the contrary, Ghezzi and Piccardi (2003) relax the i.i.d. assumption and obtain a different price-dividend solution for each state of the dividends. This variety allows the Markov chain stock model to be closer to reality.

The dividend series obeys the difference equation

$$D(k+1) = G(k+1)D(k), \quad k = t, t+1, \dots, \quad (2.45)$$

where $G(k+1)$ is the dividend growth factor described by a Markov chain.

Relation (2.45) asserts that given an initial and known value of the dividend $D(0) = d \in \mathbb{R}$, we can obtain next random dividend $D(1)$ by multiplication with the random growth factor from time zero to time one, that is $D(1) = G(1)D(0) = G(1)d$. A repetition of this operation gives $D(2) = G(2)D(1) = G(2)G(1)d$ and more generally $D(n) = \prod_{i=1}^n G(i)d$.

The combination of the dividend discount model equation (2.23) and (2.45), with a constant discount factor r , i.e., one plus the required rate of return, yields

$$p(k) = d(k) \sum_{i=1}^{+\infty} \frac{\mathbb{E}_{(k)}[\prod_{j=1}^i G(k+j)]}{r^i} =: d(k)\psi_1(g(k)), \quad (2.46)$$

where $d(k)$ and $g(k)$ are the values at time k of the dividend process and of the

growth dividend process, respectively. It should be noticed that they are known quantities given the information available up to time k . The quantity $\psi_1(g(k))$ is the so called price-dividend ratio.

The simple case is modelled with a 2-state Markov chain taking values in the state space $E = \{g_1, g_2\}$. Let $\mathbf{P} = (p_{ij})_{i,j \in E}$ be the one-step transition probability matrix of this Markov chain, and let

$$\mathbf{A1} : \bar{g} := \max(p_{11}g_1 + p_{12}g_2, p_{21}g_1 + p_{22}g_2) < r, \quad (2.47)$$

be the largest one step conditional expectation on the dividend growth rate.

If A1 holds true, then the series $p(k) = \sum_{i=1}^{+\infty} \frac{\mathbb{E}_{(k)}[D(k+i)]}{r^i}$ converges and satisfies the asymptotic condition in (2.17), and the pair $(\psi_1(g_1), \psi_1(g_2))$ is the unique and non-negative solution of the linear system

$$\begin{aligned} \psi_1(g_1) &= p_{11} \frac{\psi_1(g_1)g_1 + g_1}{r} + p_{12} \frac{\psi_1(g_2)g_2 + g_2}{r} \\ \psi_1(g_2) &= p_{21} \frac{\psi_1(g_1)g_1 + g_1}{r} + p_{22} \frac{\psi_1(g_2)g_2 + g_2}{r}. \end{aligned} \quad (2.48)$$

Assuming that for any given $D(k)$ we obtain the same $\mathbb{E}[D(t+1)]$, irrespective of the initial states g_1, g_2 , then $p_{11} = q$ and $p_{22} = 1 - q$ so the solution to (2.48) becomes

$$\psi_1(g_1) = \psi_1(g_2) = \frac{qg_1 + (1-q)g_2}{r - 1g_1 - (1-q)g_2}, \quad (2.49)$$

thus implying that the same price-dividend ratio is attached to each state, sharing the same results as Hurley and Johnson (1994, 1998) and Yao (1997).

The results have a straightforward extension to the case of an s -state Markov chain with state space $E = \{g_1, g_2, \dots, g_s\}$. Assumption A1 becomes,

$$\bar{g} := \max_{i \in E} \left(\sum_{j=1}^s p_{ij}g_j \right) < r. \quad (2.50)$$

If $\bar{g} < r$ the series (2.46) converges and the unique and non-negative solution to

the linear system is

$$\psi(g_i) = \sum_{j=1}^s p_{ij} \frac{\psi(g_j)g_j + g_j}{r}, \quad i = 1, 2, \dots, s. \quad (2.51)$$

This model has the advantage of assigning a different price-dividend ratio to each value of the states. Forecasts on the dividend growth rate are updated based on the previous value of the state, according to the Markov property. On the contrary, all previous models make assumptions on forecasts once and for all, thus obtaining a unique valuation.

Agosto and Moretto (2015) complement the model calculating a closed-form expression for the variance of random stock prices in a multinomial setting. The authors argue that for proper investment decisions a measure of risk should be taken into consideration. Thus applying the standard mean-variance analysis, an investor can deal with financial decisions under uncertainty. In their model, they relate the variance of stock prices with the variance of the dividend rate of growth, obtaining a measure of the stock riskiness. Nevertheless, in the case of portfolio selection, the variance itself is not enough, but the covariance is needed. Agosto et al. (2018) provide an explicit formula for the covariance between random stock prices with correlated random growth rates.

A further generalisation of Ghezzi and Piccardi (2003) is available in D'Amico (2013). The author models the dividend growth rate as a semi-Markov chain. In this setting, prices become duration dependent. Therefore, they are influenced by the current state of the dividend growth process and by the elapsed time in the state. The same author proposes another extension of the model describing the dividend growth series via a continuous state space semi-Markov model (D'Amico, 2017).

2.2.4 Research questions

The review of the Markov chain stock model shows many advancements in the literature on stock valuation. However, some problems remain open. While Ghezzi and Piccardi (2003) introduce the Markov chain stock model along with the computa-

tion of the first moment, and Agosto and Moretto (2015) calculate the variance of random stock prices in a multinomial setting, there is a lack of a unified framework that effectively describes the Markov chain stock model. Surprisingly, a statistical analysis of the dividend discount model is missing as well, as too is an application to real data. Moreover, Agosto et al. (2018) provide a formula for the covariance, but the existent literature is missing a proper formalisation of the dividend discount model in a multivariate setting.

This thesis aims to contribute to the literature on Markov chain stock model, proposing general frameworks for both univariate and multivariate stock valuation, providing estimators of the first and second moments. Chapters 3 and 5 describe the two approaches and provide empirical applications.

Chapter 3 is based on the work of Barbu, D'Amico, and De Blasis (2017) that extends Ghezzi and Piccardi (2003) and includes the computation of the risk of the Markov chain model integrating the multinomial setting from Agosto and Moretto (2015). The paper answers the following questions:

- (i). How to determine a specific assumption that guarantees that the risk can be computed by means of a convergent series?
- (ii). Under which hypothesis are the transversality conditions satisfied?
- (iii). How is it possible to compute the first and second moments of the price process?
- (iv). Are the moment estimators consistent? What is their asymptotic distribution?

Chapter 4 is based on the work of D'Amico and De Blasis (2018) that extends the model in Barbu, D'Amico, and De Blasis (2017) proposing a multivariate Markov chain stock model. The paper answers the following questions:

- (i). How to define an effective multivariate model?
- (ii). Under which hypothesis are the transversality conditions of the multivariate model satisfied?

- (iii). How is it possible to compute the first and second moments (variances and covariances) of the price processes?

2.3 Price Discovery

According to Lehmann (2002, p. 259), price discovery is “the efficient and timely incorporation of the information implicit in investor trading into market prices” and is one of the main functions of secondary markets. Therefore, it is essential to study how new information is impounded into prices of securities, and with the increasing market fragmentation, determine “where the price information and price discovery are being produced” (Hasbrouck, 1995, p. 1175).

In this section, we first review the Efficient Markets Hypothesis and how some high-frequency metrics can measure informational efficiency. Then, we describe three common measures of contribution to price discovery, namely Information Share (IS) by Hasbrouck (1995), Component Share (CS) by Gonzalo and Granger (1995), and Information Leadership Share introduced by Putniņš (2013) based on the work of Yan and Zivot (2010). The section ends with a formalisation of the research questions with the proposal of a new measure for price discovery based on the Multivariate Markov Chain Model, that is developed in Chapter 5, based on De Blasis (2018).

2.3.1 Informational efficiency

The general idea of the price discovery process stems from the Efficient Markets Hypothesis, originated by the seminal work of Fama (1963, 1965a,b). Under the Efficient Markets Hypothesis, prices fully reflect all available information and should follow a random walk trajectory. It means that in a perfectly efficient market, price changes are completely random, and prices do not deviate from fundamental value to a sufficient extent to allow excess returns to be obtained from trading on available information. Deviations from these three conditions, measured by low- and high-

frequency market microstructure metrics, demonstrate a certain level of inefficiency. In recent analysis, high-frequency measures are preferred over the low-frequency ones, because they allow for a better understanding of the dynamics of market efficiency and, at the same time, they correlate with the low-frequency metrics (Rösch et al., 2016).

The Random Walk Hypothesis (RWH), developed by Cowles and Jones (1937) in their analysis of sequences and reversals in historical stock returns, finds some early empirical confirmation in Cootner (1962, 1964), Fama (1963, 1965a), Fama and Blume (1966), and Osborne (1959). The test of the RWH is performed observing deviations from the expected results (under the hypothesis) of return variances or serial correlations. Under the RWH, the variance should increase linearly with increases of the time horizon analysed, i.e., $\sigma_{kT}^2 = k\sigma_T^2$, and there should be no serial correlation. Positive or negative autocorrelations demonstrate deviations from RWH and allow for short-term predictability, both inconsistent with the Efficient Market Hypothesis.

The variance ratio test is defined by Lo and MacKinlay (1988)

$$VarianceRatio_{kI} = \left| \frac{\sigma_{kI}^2}{k\sigma_I^2} - 1 \right|, \quad (2.52)$$

σ_I^2 is the variance of midquote returns sampled at interval I . The authors analyse US stock returns indexes from 1962 to 1985 and find that variances grow more than linearly, thus rejecting the hypothesis of random walk and also implying positive serial correlation. At high-frequency level, the variance ratio can be analysed in combinations of one second to ten seconds, ten seconds to 60 seconds, and one minute to five minutes sampling intervals (see, e.g., O'Hara and Ye, 2011). Large values of variance ratio indicate a greater inefficiency.

Analysing US indexes returns from 1926 to 1986, Fama and French (1988) and Poterba and Summers (1988) show a negative serial correlation in returns. Serial

correlation is measured by first order autocorrelation

$$Autocorrelation_{k,t} = Corr(r_{k,t}, r_{k,t-1}), \quad (2.53)$$

where $r_{k,t}$ is the t -th midquote return of length k for one stock. At high-frequency level, it is measured at different intraday frequencies, e.g., $k \in \{15sec, 30sec, 60sec\}$ (see, e.g., Hendershott and Jones, 2005). Because, both positive and negative value signal a deviation from the RWH, the test uses absolute values of autocorrelations, with large values showing informational inefficiency (see, e.g., Boehmer et al., 2013).

Excessive fluctuations from the fundamental value due to trading frictions are another signal of informational inefficiency. They are proxied using the short-term midquote volatility, that is the first principal component of the combination intraday midquote returns standard deviations taken at different intervals (O'Hara and Ye, 2011).

Finally, the predictability is assessed through delay in impounding market-wide information and lagged order imbalance. The delay measure from Hou and Moskowitz (2005) can be adapted to intraday data, taking the R^2 of the regression of 1-minute midquote returns for stock i on the index returns with ten lags of index returns

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \sum_{k=1}^{10} \delta_{i,k} r_{m,t-k} + \epsilon_{i,t}, \quad (2.54)$$

and the R^2 of the constrained regression without lags. The *Delay* metric is obtained as

$$Delay = 1 - \frac{R_{Constrained}^2}{R_{Unconstrained}^2}. \quad (2.55)$$

When variations in stock returns are explained by the lagged market returns, the delay measure is close to 1, and this means that there is lower informational efficiency.

Also, predictability of returns can be tested using lagged order imbalance (see, e.g., Chordia et al., 2005, 2008). The idea is based on over-reactions and under-

reactions to information released by order flows. The informational inefficiency is tested analysing the R^2 of the regression of midquote returns of stock i on lagged order imbalances

$$r_{i,t} = \alpha_i + \sum_{k=1}^{10} \beta_{i,k} OIB_{i,t-k} + \epsilon_{i,t}, \quad (2.56)$$

where OIB_t is buyer-initiated less seller-initiated dollar volumes or trades. Higher values of R^2 measure higher inefficiency.

Price discovery measures reviewed in this section are a good proxy for testing the Efficient Markets Hypothesis, especially at a high-frequency level because they provide more granularity and a better description of the markets dynamics. However, they present a limitation. Together with informational efficiency, they also measure liquidity in the market. Therefore, it is difficult to separate the measure of informational efficiency from the measure of liquidity.

2.3.2 Measures of contribution to price discovery

Market fragmentation is gaining great relevance in the studies of financial markets. Initial evidence can be found in Garbade and Silber (1979) in which the authors study the NYSE and regional exchange trading patterns. Many other authors analyse this aspect as well (see, e.g., Garbade et al., 1979, Grünbichler et al., 1994, Werner and Kleidon, 1996). The rationale of these studies is that different market prices follow a common efficient price. Therefore, there is the possibility of short-term arbitrage if small deviations from the efficient price occur.

Initially, the focus of the researchers has been on the relationship between futures and spot markets, with the conclusion that futures prices incorporate most of new information (see, e.g., Garbade and Silber, 1983). The main tool used for the analysis was the *lead-lag* methodology (see, e.g., Chan, 1992, Quan, 1992, Tse, 1995, Fleming et al., 1996, Frino et al., 2000). This technique is based on the regression of one

stock's returns on leads and lags of another stock's returns

$$r_{A,t} = \alpha + \sum_{k=-n}^{+n} \beta_k r_{B,t+k} + \epsilon_t. \quad (2.57)$$

Although this methodology gives reasonable outcomes, it can also result in misspecifications and produce spurious relations (Hasbrouck, 1995). On the contrary, most recent literature focuses on three measures of price discovery, (i) Information Share, (ii) Component Share, and (iii) Information Leadership Share, based on price dynamics described by structural equation models. Hasbrouck (1995) claims that this approach is able to replicate the lead-lag results, without incurring the problems of this earlier methodology.

For simplicity, we now consider a structural equation model with two stocks that trade in different markets and follow a common efficient price like in Harris et al. (2002) however, this analysis can be extended to n stocks.

The common efficient price m_t can be expressed by a random walk process

$$m_t = m_{t-1} + \eta_t^P, \quad \eta_t^P \stackrel{iid}{\sim} N(0, \sigma_{\eta^P}^2), \quad (2.58)$$

where η_t^P is the random new information representing the permanent shock to the efficient price.

Stock prices of different markets follow the efficient price but will be different by a random disturbance with mean equal to zero and stationary covariance

$$p_{1,t} = m_t + e_{1,t}, \quad \text{and} \quad p_{2,t} = m_t + e_{2,t}, \quad (2.59)$$

where $e_{1,t}$ and $e_{2,t}$ are transitory disturbances due to microstructure frictions that do not affect the dynamic of η_t^P .

If we rewrite equations (2.59) in terms of price changes

$$\begin{aligned}\Delta p_{1,t} &= \Delta m_t + \Delta e_{1,t} = \eta_t^P + \Delta e_{1,t}, \\ \Delta p_{2,t} &= \Delta m_t + \Delta e_{2,t} = \eta_t^P + \Delta e_{2,t},\end{aligned}$$

then we can express the two prices as

$$p_{1,t} = p_{1,t-1} + \eta_t^P + \Delta e_{1,t}, \quad \text{and} \quad p_{2,t} = p_{2,t-1} + \eta_t^P + \Delta e_{2,t}. \quad (2.60)$$

Considering a realisation of the two prices at time $t = T$, both prices impound a common factor, that is the common efficient price,

$$p_{1,T} = p_{1,0} + \sum_{t=1}^T \eta_t^P + e_{1,T}, \quad \text{and} \quad p_{2,T} = p_{2,0} + \sum_{t=1}^T \eta_t^P + e_{2,T}. \quad (2.61)$$

Therefore, each price is the sum of:

- (i). an initial non-stochastic value such that $p_{1,0} = p_{2,0}$,
- (ii). the cumulated new information deriving from the efficient price,
- (iii). a market specific transitory disturbance.

Considering that the difference between the two prices is the difference between the transitory disturbances $p_{1,t} - p_{2,t} = e_{1,t} - e_{2,t}$, and this difference being a stationary time series, then the two prices are cointegrated. Therefore, $\Delta p_{1,t}$ and $\Delta p_{2,t}$ can be estimated as a vector error correction model (VECM) explained in a general form in the remainder of the discussion where we follow the notation from Yan and Zivot (2010).

Let $\mathbf{p}_t = (p_{1,t}, p_{2,t}, \dots)'$ be the vector of two log prices of two stocks from different markets¹. Let us assume that \mathbf{p}_t is an integrated process of order 1, $I(1)$, in other words, contains a random walk component, and $\Delta \mathbf{p}_t$ is stationary process $I(0)$.

¹Prices can be futures and spot prices, or they could be related to the same stock traded in different venues, linked by arbitrage.

Thus, the price changes, $\Delta \mathbf{p}_t$, can be expressed with a Wold representation as

$$\Delta \mathbf{p}_t = \Psi(L)\mathbf{e}_t = \mathbf{e}_t + \Psi_1\mathbf{e}_{t-1} + \Psi_2\mathbf{e}_{t-2} + \dots, \quad (2.62)$$

where $\Psi(L) = \sum_{k=0}^{\infty} \Psi_k L^k$ is the lag polynomial operator with $\Psi_0 = \mathbf{I}_2$, and $\mathbf{e}_t = (e_{1,t}, e_{2,t}, \dots)'$ is the vector of serially uncorrelated errors with $\mathbb{E}[\mathbf{e}_t] = 0$ and

$$\mathbb{E}[\mathbf{e}_t \mathbf{e}_s'] = \begin{cases} \mathbf{0} & \text{if } t \neq s \\ \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} & \text{otherwise,} \end{cases}$$

where Σ is a time-invariant covariance matrix with ρ being the correlation between the error terms.

If prices follow the same efficient price, then \mathbf{p}_t is cointegrated and we can assume that the cointegrating vector is $\beta = (1, -1)'$ so that $\beta'\mathbf{p}_{t-1} = p_{1,t} - p_{2,t}$ is $I(0)$. With cointegration, \mathbf{p}_t can be described with a VECM(K-1) model of the form

$$\Delta \mathbf{p}_t = \alpha(\beta'\mathbf{p}_{t-1} - \mathbb{E}[\beta'\mathbf{p}_{t-1}]) + \sum_{k=1}^{K-1} \Gamma_k \Delta \mathbf{p}_{t-k} + \mathbf{e}_t, \quad (2.63)$$

where α is the vector of the error correction coefficients measuring how fast each price eliminates differences from the efficient price, and Γ_k is a matrix with price changes coefficients. The term $\mathbb{E}[\beta'\mathbf{p}_{t-1}]$ represents the systematic differences in prices, e.g., differences between a bid and offer quote.

Applying the permanent-transitory decomposition of Beveridge and Nelson (1981) to the VMA in (2.62) we obtain

$$\mathbf{p}_t = \mathbf{p}_0 + \Psi(1) \sum_{j=1}^t \mathbf{e}_j + \mathbf{s}_t, \quad (2.64)$$

where $\Psi(1) = \sum_{k=0}^{\infty} \Psi_k$, $\mathbf{s}_t = \Psi^*(L)\mathbf{e}_t$, and $\Psi_k^* = -\sum_{j=k+1}^{\infty} \Psi_j$, $k = 0, \dots, \infty$. Hasbrouck (1995) shows that the first term of the r.h.s of equation (2.64) is constant

and reflects non-stochastic differences between price variables. In the second term, the matrix $\Psi(1)$ is the sum of the moving average coefficients and it is a measure of the impact of innovations \mathbf{e}_t , i.e., the long-run impact of innovations on prices. The full term, $\Psi(1) \sum_{j=1}^t \mathbf{e}_j$, captures the random walk component that belongs to the efficient price. The third term is a zero-mean covariance stationary process.

Because of the cointegrating vector $\boldsymbol{\beta} = (1, -1)'$, the rows of the impact matrix, $\Psi(1)$, are all the same, therefore we can define one impact vector $\boldsymbol{\psi}$ and describe the permanent innovation as

$$\eta_t^P = \boldsymbol{\psi} \mathbf{e}_t. \quad (2.65)$$

Considering the efficient price in (2.58), and assuming $\mathbf{p}_0 = \mathbf{0}$, then we can rewrite equation (2.64) as

$$\mathbf{p}_t = \mathbb{1} m_t + \mathbf{s}_t, \quad (2.66)$$

where $\mathbb{1} = (1, 1)'$. Equation (2.66) expresses that prices $p_{i,t}$ share the common underlying asset that incorporate the full permanent information, but have different transitory effects, \mathbf{s}_t , capturing deviations from the efficient price. This last term can be a consequence of many market microstructure frictions, e.g., bid-ask bounce, price discreteness, or illiquidity effects.

In this setting, where multiple cointegrated prices follow a common efficient price and each market produces its variations, Hasbrouck (1995) develops the Information Share (IS), that is a measure of the contribution to price discovery from each market. IS measures how much of the variation in the efficient price, $Var(\boldsymbol{\psi} \mathbf{e}_t) = \boldsymbol{\psi}' \boldsymbol{\Sigma} \boldsymbol{\psi}$, can be attributed to one market.

If $\boldsymbol{\Sigma}$ is diagonal, the Information Share is

$$IS_i = \frac{\psi_i^2 \sigma_i^2}{\boldsymbol{\psi}' \boldsymbol{\Sigma} \boldsymbol{\psi}} = \frac{\psi_i^2 \sigma_i^2}{\psi_1^2 \sigma_1^2 + \psi_2^2 \sigma_2^2}, \quad i = 1, 2, \quad (2.67)$$

with $IS_1 + IS_2 = 1$. On the contrary, if the price innovations are correlated across markets, i.e., $\boldsymbol{\Sigma}$ is not diagonal, the IS does not produce the proper results. Using

the Cholesky factorization,

$$\Sigma = \mathbf{M}\mathbf{M}', \quad (2.68)$$

where M is a lower triangular matrix, we can rewrite equation (2.67) as

$$IS_i = \frac{([\boldsymbol{\psi}\mathbf{M}]_i)^2}{\boldsymbol{\psi}'\boldsymbol{\Sigma}\boldsymbol{\psi}}. \quad (2.69)$$

Equation (2.69) produces different results according to the ordering of prices. In the case of two prices Hasbrouck (1995) proposes to define a low (high) bound with i -th price ordered first (last), and to sample at high frequencies to reduce contemporaneous correlation. He reports that in his application to price discovery between NYSE market and other exchanges with prices sampled at one second, the range of the bounds is very narrow and prices have a low contemporaneous residual correlations. Similar results are shown in Hendershott and Jones (2005) and Tse (1995). However, the sampling frequency is data dependent and IS produces different results at different sampling frequencies, as well as extensive ranges for the bounds, thus creating some difficulties in the interpretation of results (see, e.e., Grammig et al., 2005, Theissen, 2002, Sapp, 2002, Huang, 2002).

Booth et al. (2002), Chu et al. (1999) and Harris et al. (2002) propose to measure the contribution to price discovery using the Permanent-Transitory decomposition of Gonzalo and Granger (1995) of the type

$$\mathbf{p}_t = \mathbf{A}_1 f_t + \mathbf{A}_2 \mathbf{z}_t. \quad (2.70)$$

Equation (2.70) expresses cointegrated prices as a sum of common factors $f_t = \boldsymbol{\gamma}'\mathbf{p}_t \sim I(1)$, indicating a permanent component, plus stationary error correction terms $\mathbf{z}_t = \boldsymbol{\alpha}'\mathbf{p}_{t-1} \sim I(0)$, that is the transitory component that does not Granger-cause f_t in the long run. Gonzalo and Granger show that $\boldsymbol{\gamma} = (\boldsymbol{\alpha}'_{\perp}\boldsymbol{\beta}_{\perp})^{-1}\boldsymbol{\alpha}'_{\perp}$ is orthogonal to the error correction coefficient vector in the VECM model (2.63). Since the cointegrating vector is $\boldsymbol{\beta} = (1, -1)'$, we can choose $\boldsymbol{\beta}_{\perp} = \mathbb{1} = (1, 1)'$, so that $\boldsymbol{\beta}'_{\perp}\boldsymbol{\beta} = 0$, and thus $\boldsymbol{\gamma} = (\boldsymbol{\alpha}'_{\perp}\mathbb{1})^{-1}\boldsymbol{\alpha}'_{\perp}$, that means that the permanent

component can be interpreted as weighted average of observed prices with weights γ_i with $i = 1, 2$.

Booth et al. (2002), Chu et al. (1999) and Harris et al. (2002) interpret the weights γ_i as a measure of the contribution of one market to the permanent component of prices. This contribution to price discovery is called Component Share (CS)

$$CS_i = \gamma_i = \frac{\alpha_{\perp,i}}{\alpha_{\perp,1} + \alpha_{\perp,2}}, \quad i = 1, 2. \quad (2.71)$$

Both the IS and CS rely on the decomposition of innovations. The former is based on the attribution of the shares of the common factor variance $Var(\boldsymbol{\psi}\mathbf{e}_t) = \boldsymbol{\psi}'\boldsymbol{\Sigma}\boldsymbol{\psi}$ to each market, while the latter is based on the distribution of shares of the common factor price to each market. CS metric presents the advantage of measuring the market's price discovery contribution directly. However, there is considerable debate in the literature on what these metrics really measure.

Baillie et al. (2002) reconcile the two measures. Starting from the work of Johansen (1991), they state that both models derive from $\boldsymbol{\alpha}_{\perp}$, and prove that

$$\boldsymbol{\Psi}(1) = \begin{pmatrix} \psi \\ \psi \end{pmatrix} = \boldsymbol{\Pi} \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_1 & \gamma_2 \end{pmatrix}, \quad (2.72)$$

where $\boldsymbol{\Pi} = (\boldsymbol{\alpha}'_{\perp}(\mathbf{I} - \sum_{k=1}^{K-1} \boldsymbol{\Gamma}_k)\boldsymbol{\beta}_{\perp})^{-1}$, with \mathbf{I} being the identity matrix.

Therefore, there exist the relationship

$$\frac{\psi_1}{\psi_2} = \frac{\gamma_1}{\gamma_2}, \quad (2.73)$$

so the measures CS and IS can be expressed in relative values.

Substituting (2.73) into (2.67), when $\boldsymbol{\Sigma}$ is diagonal, we obtain

$$IS_i = \frac{\gamma_i^2 \sigma_i^2}{\gamma_1^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2}, \quad i = 1, 2. \quad (2.74)$$

In the same way, it is possible to express CS in terms of ψ_i

$$CS_i = \frac{\psi_i}{\psi_1 + \psi_2}, \quad i = 1, 2, \quad (2.75)$$

or directly in terms of elements of the error correction coefficient vector α

$$CS_1 = \frac{\alpha_2}{\alpha_2 - \alpha_1}, \quad CS_2 = \frac{-\alpha_1}{\alpha_2 - \alpha_1}. \quad (2.76)$$

In the case of correlated errors across markets, using the Cholesky factorisation (2.68) with the lower triangular matrix

$$\mathbf{M} = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2(1 - \rho^2)^{\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} m_{11} & 0 \\ m_{12} & m_{22} \end{pmatrix},$$

and substituting (2.73) into (2.69), we obtain

$$IS_1 = \frac{(\psi_1 m_{11} + \psi_2 m_{12})^2}{(\psi_1 m_{11} + \psi_2 m_{12})^2 + (\psi_2 m_{22})^2}, \quad (2.77a)$$

$$IS_2 = \frac{(\psi_2 m_{22})^2}{(\psi_1 m_{11} + \psi_2 m_{12})^2 + (\psi_2 m_{22})^2}. \quad (2.77b)$$

Yan and Zivot (2010) advance the discussion about the interpretation of the two measures. They argue that Information Share, when the errors are uncorrelated, is more appropriate to measure the contribution to price discovery, as it measures the response of one market to innovations. However, if the errors are serially correlated, i.e., there is the presence of microstructure frictions, IS becomes ambiguous, requiring one to calculate the upper-lower bounds. In fact, a high IS can be attributed to a strong response to either new information or frictions. On the contrary, Component Share can measure the response to transitory frictions. Therefore, the authors propose to combine the two metrics to compensate the ambiguity of the IS, in a way that a high IS with a low CS shows a very good response to innovations, while this is not true for a high IS with a high CS.

Putniņš (2013) formalises the intuition of Yan and Zivot (2010) into a new measure, the Informational Leadership Share (ILS)

$$ILS_1 = \frac{IL_1}{IL_1 + IL_2}, \quad ILS_2 = \frac{IL_2}{IL_1 + IL_2}, \quad (2.78)$$

where

$$IL_1 = \left| \frac{IS_1 CS_2}{IS_2 CS_1} \right|, \quad \text{and} \quad IL_2 = \left| \frac{IS_2 CS_1}{IS_1 CS_2} \right|.$$

The author argues that IS and CS are proper measures when the prices share the same level of noise. However, when price errors are different, “IS and CS both measure a combination of leadership (relative speed) in impounding new information and relative avoidance of noise, to different extents” (Putniņš, 2013, p. 81). Conversely, ILS helps to cancel the noise component to identify the price leadership if the following conditions are respected:

- (i). the analysis involves only two price series,
- (ii). there is only one permanent and one transitory shock,
- (iii). the errors are uncorrelated.

2.3.3 Research questions

The review of the price discovery literature shows that researchers have dedicated much effort in understanding the price discovery dynamics. Many measures have been created, from low frequency to high frequency for a general understanding, and more complex ones to define one market’s contribution to price discovery in the scenario of cointegrated prices.

The three common measures from the literature are the Information Share (IS), the Component Share (CS), and the Informational Leadership Share (ILS). They are all based on structural equation systems to model the prices dynamics, with the idea that prices in different markets follow a common efficient price. They rely on

the permanent-transitory decomposition of new information and try to understand how much of the long-run information is due to one market or another.

However, all proposed metrics present several limitations due to assumptions in the models' definitions, and there is no precise specification on what they really measure when the assumptions are relaxed. In trying to balance the IS and CS, the ILS finds a compromise between the two metrics, but its results are limited to two price series with uncorrelated errors. In conclusion, many questions are still open.

In this thesis, we address these problems establishing a new measure for price discovery that does not rely on the definition of the structural models, typical of the other measures. On the contrary, we use a multivariate Markov chain model to establish the contribution to price discovery. We still assume that prices are cointegrated, following a common efficient price, but since we cannot observe the common price, we have to rely entirely on what we directly observe.

As noted in previous sections, the timing of incorporation of new information is of primary importance in defining the contribution to price discovery. It means that some prices update faster than others when there is new information, and these new prices are observable in the market. Thus, a price change is a signal to other markets that, in turn, can adapt their prices based on the common efficient price plus the signalled price change.

Because we can directly observe the dynamics of the price change series from multiple markets, we develop a metric that models the price changes in terms of dependencies between markets. For this purpose, the multivariate Markov chain represents an optimal tool to understand these dynamics and, hence, the contribution to price discovery.

In this thesis, we answer the following question:

- (i). How to integrate the Markov chain model into the framework of price discovery?
- (ii). How to define a new measure of price discovery based on the Markov chain model?

- (iii). Is it possible to have a summary measure that is comparable with the others?
- (iv). Is it possible to overcome the limitation of the number of markets in the analysis?
- (v). Can we overcome the limits due to the serial correlation of the errors?

Chapter 5, based on the work of De Blasis (2018) answers these questions and presents the formulation of a new measure of price discovery based on a multivariate Markov chain.

2.4 Summary

This chapter presented an overview of the Markov chain model, in both its univariate and multivariate aspects, followed by reviews of two important topics in finance, stock valuation and price discovery. The first topic is addressed in terms of the dividend discount model and presents many valuations procedures that are present in literature, starting from the basic Gordon model to more complex methodologies, based on the Markov chain model. The second topic is more related to financial markets and market microstructure. The review discusses why price discovery is important and how it is measured.

Both reviews highlight that some problems are still open and not entirely addressed by the literature. Therefore, some research questions are posed and the following chapters propose solutions based on the application of Markov chain models.

Chapter 3 and 4 address questions on the dividend discount model and formalise two frameworks, univariate and multivariate, for stock valuation when the dividend growth is modelled through a Markov chain. Empirical applications with algorithmic implementations complete the research and demonstrate the validity of the model.

Chapter 5 proposes a new measure of price discovery based on a multivariate Markov chain model, to overcome other measures' limitations. The new measure is tested and compared with existent measures.

The algorithmic implementation of the methods presented in this thesis is reported in Chapter 6.

3 Novel advancements in the Markov chain stock model

This chapter presents further advancements in the Markov chain stock model proposing a general framework for valuation and provides estimators for the first and second moments of the price process, as well as an application to real dividend data to demonstrate the practical implementation of these methods. The chapter is organised as follows: first, in Section 3.1, we present the Markov chain dividend valuation model and we derive the results relating to the risk process. Next, Section 3.2 presents results on the statistical estimation of the price-dividend ratio of the first and second order. In Section 3.3 we present an application to real dividend data and discuss some practical problems to be dealt with when executing an application to real data. All proofs are deferred to the Appendix.

3.1 The Markov chain dividend valuation model

Let $P(k)$ be the random variable giving the fundamental value of a stock at time $k \in \mathbb{N}$. Let $D(k)$ be the dividend at time $k \in \mathbb{N}$, also assumed to be a random variable, and denote by r one plus the required rate of return on the stock, assumed to be constant. The fundamental valuation analysis states that $p(k) := \mathbb{E}_k [P(k)]$ obeys the equation

$$p(k) = \frac{\mathbb{E}_k [D(k+1) + P(k+1)]}{r}, \quad (3.1)$$

where \mathbb{E}_k is the conditional expectation given the information available up to time k . As it is well known, see for example Samuelson (1973), if we assume that

$$\lim_{i \rightarrow +\infty} \frac{\mathbb{E}_k[P(k+i)]}{r^i} = 0, \quad (3.2)$$

then the solution of (3.1) is expressed by the series

$$p(k) = \sum_{i=1}^{+\infty} \frac{\mathbb{E}_k[D(k+i)]}{r^i}. \quad (3.3)$$

If condition (3.2) is not assumed, then Blanchard and Watson (1982) proved that there can exist different solutions of the fundamental equation, i.e., in the stock market there is the presence of bubbles.

Note that the uncertainty in the dividend process propagates into the price process. In order to have a quantification of this effect, we analyse the second order moment of the price process. To this end, according to D'Amico (2017), if we set

$$P^2(k) := \left(\frac{D(k+1) + P(k+1)}{r} \right)^2, \quad (3.4)$$

then, by means of successive substitutions we get

$$\begin{aligned} P^2(k) &= \sum_{i=1}^N \frac{D^2(k+i)}{r^{2i}} + 2 \sum_{i=1}^N \sum_{j>i}^N \frac{D(k+i)D(k+j)}{r^{i+j}} \\ &\quad + 2 \sum_{i=1}^N \frac{D(k+i)P(k+N)}{r^{i+N}} + \frac{P^2(k+N)}{r^{2N}}. \end{aligned}$$

Therefore, applying the conditional expectation \mathbb{E}_k we obtain

$$\begin{aligned} p^{(2)}(k) &:= \mathbb{E}_k[P^2(k)] = \sum_{i=1}^N \frac{\mathbb{E}_k[D^2(k+i)]}{r^{2i}} + 2 \sum_{i=1}^N \sum_{j>i}^N \frac{\mathbb{E}_k[D(k+i)D(k+j)]}{r^{i+j}} \\ &\quad + 2 \sum_{i=1}^N \frac{\mathbb{E}_k[D(k+i)P(k+N)]}{r^{i+N}} + \frac{\mathbb{E}_k[P^2(k+N)]}{r^{2N}}. \end{aligned} \quad (3.5)$$

Equation (3.5) explains that to guarantee the dependence of the risk measure

($p^{(2)}(k)$) only on the dividend process, it is necessary that both $\lim_{N \rightarrow +\infty} \frac{\mathbb{E}_k[P^2(k+N)]}{r^{2N}} = 0$ and $\lim_{N \rightarrow +\infty} \sum_{i=1}^N \frac{\mathbb{E}_k[D(k+i)P(k+N)]}{r^{i+N}} = 0$. In this case, the solution of (3.5) would be

$$p^{(2)}(k) = \sum_{i=1}^{+\infty} \frac{\mathbb{E}_k[D^2(k+i)]}{r^{2i}} + 2 \sum_{i=1}^{+\infty} \sum_{j>i} \frac{\mathbb{E}_k[D(k+i)D(k+j)]}{r^{i+j}}. \quad (3.6)$$

Formula (3.6) is the fundamental formula for the risk of the price process.

In order to be able to evaluate (3.3) and (3.6), we need to specify a model for the dividend process. For example, in the paper of Gordon and Shapiro (1956) is assumed a constant growth rate of dividends. Many variants of the Gordon and Shapiro (1956) model have been suggested in the finance literature. These variants come from a common need of imposing less restrictive assumptions on the dividend process. For example, Brooks and Helms (1990) and Barsky and De Long (1993) consider a multistage model with dividend growth rates changing deterministically among stages. Models based on Markov chains were proposed by Hurley and Johnson (1994, 1998) and by Yao (1997). In a more recent paper of Ghezzi and Piccardi (2003) it is assumed that dividends satisfy the difference equation

$$D(k+1) = G(k+1)D(k), \quad (3.7)$$

where $\{G(k)\}$ is the dividend growth factor described by a Markov chain.

Relation (3.7) asserts that given an initial and known value of the dividend $D(0) = d \in \mathbb{R}$, we can obtain next random dividend $D(1)$ by multiplication with the random growth factor from time zero to time one, that is $D(1) = G(1)D(0) = G(1)d$. A repetition of this operation gives $D(2) = G(2)D(1) = G(2)G(1)d$ and more generally $D(n) = \prod_{i=1}^n G(i)d$.

3.1.1 The computation of the moments

For clarity of exposition, we limit ourselves for the moment to the simplest case of a two state Markov chain with state space $E = \{g_1, g_2\}$. The generalisation to a general finite state space Markov chain is straightforward and will be discussed at

the end of this section. Let $\mathbf{P} = (p_{ij})_{i,j \in E}$ be the one step transition probability matrix of this Markov chain. The combination of Equations (3.3) and (3.7) yields

$$p(k) = d(k) \sum_{i=1}^{+\infty} \frac{\mathbb{E}_k[\prod_{j=1}^i G(k+j)]}{r^i} =: d(k)\psi_1(g(k)), \quad (3.8)$$

where $d(k)$ and $g(k)$ are the values at time k of the dividend process and of the growth dividend process, respectively. It should be noticed that they are known quantities given the information available up to time k . The quantity $\psi_1(g(k))$ is the so called price-dividend ratio.

It should be noted that, where needed in the following, we will use the notation $p(d(k), g(k))$ to denote the price at time k in order to stress the dependence on the value of the dividend process and of the growth dividend process at that time.

The following assumption will be needed in the sequel

$$\mathbf{A1} : \bar{g} := \max(p_{11}g_1 + p_{12}g_2, p_{21}g_1 + p_{22}g_2) < r. \quad (3.9)$$

Note that \bar{g} is the largest one step expectation of the dividend growth rate.

Proposition 3.1. *(see Ghezzi and Piccardi, 2003) If A1 holds true, then the series $p(k) = \sum_{i=1}^{+\infty} \frac{\mathbb{E}_k[D(k+i)]}{r^i}$ converges and satisfies the asymptotic condition*

$$\lim_{i \rightarrow +\infty} \frac{\mathbb{E}_k[P(k+i)]}{r^i} = 0.$$

Proposition 3.2. *(see Ghezzi and Piccardi, 2003) If A1 holds true, the pair $(\psi_1(g_1), \psi_1(g_2))$ is the unique and nonnegative solution of the linear system*

$$\begin{aligned} \psi_1(g_1) &= p_{11} \frac{\psi_1(g_1)g_1 + g_1}{r} + p_{12} \frac{\psi_1(g_2)g_2 + g_2}{r} \\ \psi_1(g_2) &= p_{21} \frac{\psi_1(g_1)g_1 + g_1}{r} + p_{22} \frac{\psi_1(g_2)g_2 + g_2}{r}. \end{aligned} \quad (3.10)$$

In order to compute $p^{(2)}(k)$, we need an additional assumption

$$\mathbf{A2} : \bar{g}^{(2)} := \max(p_{11}g_1^2 + p_{12}g_2^2, p_{21}g_1^2 + p_{22}g_2^2) < r^2. \quad (3.11)$$

$\bar{g}^{(2)}$ is the largest one step second order moment of the dividend growth rate.

Proposition 3.3. *Assume that hypotheses A1 and A2 hold true. Then, the series*

$$p^{(2)}(k) = \sum_{i=1}^{+\infty} \frac{\mathbb{E}_k[D^2(k+i)]}{r^{2i}} + 2 \sum_{i=1}^{+\infty} \sum_{j>i} \frac{\mathbb{E}_k[D(k+i)D(k+j)]}{r^{i+j}} \quad (3.12)$$

converges and the following asymptotic conditions are satisfied:

$$\lim_{N \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[P^2(k+N)]}{r^{2N}} = 0, \quad \lim_{N \rightarrow +\infty} \sum_{i=1}^N \frac{\mathbb{E}_k[D(k+i)P(k+N)]}{r^{i+N}} = 0.$$

Proof. See appendix.

Proposition 3.4. *Assume that hypotheses A1 and A2 hold true. Then, the pair $(\psi_2(g_1), \psi_2(g_2))$ is the unique and nonnegative solution of the linear system*

$$\begin{aligned} \psi_2(g_1)(r^2 - p_{11}g_1^2) - \psi_2(g_2)p_{12}g_2^2 &= p_{11}g_1^2(1 + 2\psi_1(g_1)) + p_{12}g_2^2(1 + 2\psi_1(g_2)) \\ \psi_2(g_2)(r^2 - p_{22}g_2^2) - \psi_2(g_1)p_{21}g_1^2 &= p_{21}g_1^2(1 + 2\psi_1(g_1)) + p_{22}g_2^2(1 + 2\psi_1(g_2)). \end{aligned} \quad (3.13)$$

Proof. See appendix.

The results have a straightforward extension to the case of an s -state Markov chain with state space $E = \{g_1, g_2, \dots, g_s\}$. Note that the assumptions A1 and A2 should be formulated as follows:

$$\bar{g} := \max_{i \in E} \left(\sum_{j=1}^s p_{ij}g_j \right) < r \quad (3.14)$$

$$\bar{g}^{(2)} := \max_{i \in E} \left(\sum_{j=1}^s p_{ij}g_j^2 \right) < r^2. \quad (3.15)$$

In this more general case, Propositions 3.2, 3.3 and 3.4 remain valid and the systems (3.10) and (3.13) can be conveniently represented in matrix form. To this end let us introduce some matrix notation. Let \mathbf{I} be the identity matrix of dimension $s \times s$. For any $r \in \mathbb{R}^* := \mathbb{R} - \{0\}$, we define $\mathbf{I}_r := r\mathbf{I}$ and, more generally, $\mathbf{I}_r^n = \mathbf{I}_{r^n}$

and $\mathbf{I}_r^{-1} = \mathbf{I}_{r-1}$. Moreover, for any $\mathbf{g} = (g_1, \dots, g_s)^\top$, $\mathbf{g}^n = (g_1^n, \dots, g_s^n)^\top \in (\mathbb{R}^*)^s$ with $()^\top$ denoting the transpose of a vector, we denote by

$$\mathbf{I}_{\mathbf{g}} = (I_g(i, j))_{i, j \in E}, \quad I_g(i, j) = \begin{cases} g_i, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \quad (3.16)$$

More generally, it results that $\mathbf{I}_{\mathbf{g}}^n = \mathbf{I}_{\mathbf{g}^n}$ and $\mathbf{I}_{\mathbf{g}}^{-1} = \mathbf{I}_{\mathbf{g}^{-1}}$.

Finally let $\Psi_1 = (\psi_1(g_1), \dots, \psi_1(g_n))^\top$ and $\Psi_2 = (\psi_2(g_1), \dots, \psi_2(g_n))^\top$ be the vectors of the price-dividend ratio of first and second order. Then, the systems (3.10) and (3.13) have the following matrix representation:

$$(\mathbf{I}_r - \mathbf{P} \cdot \mathbf{I}_{\mathbf{g}}) \cdot \Psi_1 = \mathbf{P} \cdot \mathbf{g}, \quad (3.17)$$

$$\left(\mathbf{I}_r^2 - \mathbf{P} \cdot \mathbf{I}_{\mathbf{g}}^2 \right) \cdot \Psi_2 = \mathbf{P} \cdot \left((\mathbf{g} \diamond \mathbf{g}) + 2\Psi_1 \diamond (\mathbf{g} \diamond \mathbf{g}) \right), \quad (3.18)$$

where \cdot denotes the usual row by column matrix product and \diamond denotes the Hadamard element by element product. When no confusion is possible, we will omit writing \cdot for the usual row by column matrix product.

Note that, according to Proposition 3.2, the system (3.10) (or equivalently (3.17)) has a unique solution. Consequently, the matrix $(\mathbf{I}_r - \mathbf{P} \cdot \mathbf{I}_{\mathbf{g}})$ is invertible and the solution is given by

$$\Psi_1 = (\mathbf{I}_r - \mathbf{P} \cdot \mathbf{I}_{\mathbf{g}})^{-1} \cdot \mathbf{P} \cdot \mathbf{g}. \quad (3.19)$$

Similarly, according to Proposition 3.4, the system (3.13) (or equivalently (3.18)) has a unique solution. Consequently, the matrix $(\mathbf{I}_r^2 - \mathbf{P} \cdot \mathbf{I}_{\mathbf{g}}^2)$ is invertible and the solution is given by

$$\begin{aligned} \Psi_2 &= (\mathbf{I}_r^2 - \mathbf{P} \cdot \mathbf{I}_{\mathbf{g}}^2)^{-1} \cdot \mathbf{P} \cdot \left((\mathbf{g} \diamond \mathbf{g}) + 2\Psi_1 \diamond (\mathbf{g} \diamond \mathbf{g}) \right) \\ &= (\mathbf{I}_{r,2} - \mathbf{P} \cdot \mathbf{I}_{\mathbf{g}^2})^{-1} \cdot \mathbf{P} \cdot \left((\mathbf{g} \diamond \mathbf{g}) + 2\Psi_1 \diamond (\mathbf{g} \diamond \mathbf{g}) \right), \end{aligned} \quad (3.20)$$

where we used the fact that $\mathbf{I}_{\mathbf{g}}^2 = \mathbf{I}_{\mathbf{g}^2}$ and $\mathbf{I}_r^2 = \mathbf{I}_{r^2}$. It should be remarked that relation (3.20) gives an explicit formula for the second-order price-dividend ratio that in turn, after multiplication with $d^2(t)$ gives a formula for the second moment of the price process that is expressed in function of the model parameters \mathbf{P} and \mathbf{g} .

The last point we have to deal with is about the forecasting of future fundamental prices. To this end, let us denote by $E^{(n)}p(d(k), g_a) := \mathbb{E}_{(d(k), g_a)}[P(D(k+n), G(k+n))]$ the forecasted fundamental price within n units of time given that at current time k the dividend is $D(k) = d(k)$ and the growth dividend process is in state $G(k) = g_a$. The following proposition gives an explicit formula for $E^{(n)}p(d(k), g_a)$.

Proposition 3.5. *The expected forecast of fundamental price, given that $D(k) = d_k$ and $g(k) = g_a$, is given by*

$$E^{(n)}p(d_k, g_a) = \sum_{j_1, \dots, j_n \in E} \prod_{i=1}^n \left(p_{j_{i-1}j_i} g_{j_i} \right) p(d_k, g_{j_n}), \quad (3.21)$$

where $g_{i_0} = g_a$. Note that we can write formula (3.21) in the following matrix form:

$$\begin{pmatrix} E^{(n)}p(d_k, g_1) \\ \vdots \\ E^{(n)}p(d_k, g_s) \end{pmatrix} = d_k \mathbf{P}^n \mathbf{I}_{\mathbf{g}^n} \begin{pmatrix} \psi_1(g_1) \\ \vdots \\ \psi_1(g_s) \end{pmatrix}. \quad (3.22)$$

Proof. See appendix.

3.2 The inferential analysis

3.2.1 Estimation of a Markov chain

Let $X = (X_n, n \in \mathbb{N})$ be an homogeneous ergodic Markov chain (see e.g., Brémaud, 1999) defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a finite state space $E = \{g_1, g_2, \dots, g_s\}$ and transition probability matrix $\mathbf{P} = (p_{ij})_{i,j \in E}$. By $\alpha = (\alpha_i)_{i \in E}$ we denote the initial distribution, $\alpha_i := \mathbb{P}(X_0 = g_i), i \in E$, and by $\boldsymbol{\pi} = (\pi_i)_{i \in E}$, the unique stationary distribution (assumed to exist, due to the ergodicity of the chain). The

stationary distribution is determined by the equations $\pi_j = \sum_{i \in E} \pi_i p_{ij}$, for all $j \in E$.

Note that, for notational convenience, for any $i, j \in \{1, \dots, s\}$ we set p_{ij} , α_i and π_i instead of $p_{g_i g_j}$, α_{g_i} and π_{g_i} , respectively.

In this paper we consider the estimation problem when only one trajectory of the Markov chain is observed. This is in agreement with the data used in the application where a time series of dividend constitutes the sample. The case when the analyst observes several sample paths of the dividend process can be dealt with using similar techniques. These two different sampling schemes have been deeply studied by Anderson and Goodman (1957), Billingsley (1961a,b) and Sadek and Linnios (2002).

Assume that $\mathbf{x} = (x_0, x_1, \dots, x_m)$ is a sample path of the Markov chain observed up to time m , i.e. a realisation of $\mathbf{X} = (X_0, X_1, \dots, X_m)$. Let us define the counting processes

$$N_{ij}(m) := \sum_{u=1}^m \mathbb{1}_{\{X_{u-1}=i, X_u=j\}}, \quad N_i(m) := \sum_{u=0}^{m-1} \mathbb{1}_{\{X_u=i\}}.$$

They represent the number of transitions from i to j observed up to time m and the number of visits to state i observed up to time m , respectively.

The maximum likelihood estimator of p_{ij} (see, e.g., Billingsley, 1961a,b) is

$$\hat{p}_{ij}(m) = \frac{N_{ij}(m)}{N_i(m)}. \quad (3.23)$$

We will recall now some asymptotic results related to the estimators of Markov chains.

Proposition 3.6. *(see, e.g., Billingsley, 1961a) Let $(X_n)_{n \in \mathbb{N}}$ be an ergodic Markov chain defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with finite state space $E = \{g_1, g_2, \dots, g_s\}$, transition probability matrix $\mathbf{P} = (p_{ij})_{i,j \in E}$ and stationary distribution $\boldsymbol{\pi} = (\pi_i)_{i \in E}$. Then:*

1. $\hat{p}_{ij}(m) \xrightarrow[m \rightarrow \infty]{a.s.} p_{ij}$.

2. The s^2 dimensional random vector $(\sqrt{m}(\widehat{p}_{ij}(m) - p_{ij}))_{i,j \in E}$ converges in distribution, as m tends to infinity, to a normally distributed random variable $Z \sim \mathcal{N}_{s^2}(\mathbf{0}, \widetilde{\Gamma})$, where $\widetilde{\Gamma} \in \mathcal{M}_{s^2 \times s^2}$ is the covariance matrix of dimension $s^2 \times s^2$ defined under diagonal form by blocks as follows:

$$\widetilde{\Gamma} := \text{diag} \left(\frac{\mathbf{\Lambda}_i}{\pi_i} \mid i \in E \right) = \begin{pmatrix} \frac{1}{\pi_1} \mathbf{\Lambda}_1 & 0 & \cdots & 0 \\ 0 & \frac{1}{\pi_2} \mathbf{\Lambda}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\pi_s} \mathbf{\Lambda}_s \end{pmatrix}, \quad (3.24)$$

where, for any $i \in E$, the matrix $\mathbf{\Lambda}_i$ is given by $\mathbf{\Lambda}_i = (\delta_{jl}p_{ij} - p_{ij}p_{il})_{j,l \in E}$, with δ_{ij} the Kronecker symbol, i.e., $\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ if $i \neq j$.

3.2.2 Estimation of the financial quantities

Here, using the results listed in Proposition 3.6 we propose estimators of the financial quantities computed in Section 3.1 and we also derive their asymptotic properties when the length of the sample path goes to infinity. It is to be remarked again that the results of this section have a direct extension to the case when several trajectories of the growth dividend process are observed.

Estimation of the price-dividend ratio

As the price-dividend ratio is a function of the transition probability matrix \mathbf{P} cf. formula (3.19), the general idea is to use this expression in order to: (i) build a plug-in estimator of the price-dividend ratio using the estimator of the transition matrix $\widehat{p}_{ij}(m)$ given in (3.23); (ii) use the asymptotic results given in Proposition 3.6 in order to derive the consistency of the estimator of the price-dividend ratio (by means of the continuous mapping theorem) and the asymptotic normality (applying the delta method, see e.g. Van der Vaart (1998)).

The application of the continuous mapping (respectively of the delta method) requires that the objective function be a continuous (respectively differentiable)

function of the parameters. In our model the parameter space is given by $\{p_{ij} : 0 \leq p_{ij} \leq 1, \sum_{j \in E} p_{ij} = 1, \forall i \in E\}_{i,j \in E}$. As the variables p_{ij} are not independent because each row of the matrix \mathbf{P} sums to one, it will be convenient to express Ψ_1 as a function of independent variables. Consequently, let us consider the function

$$\Phi_1 = (\Phi_1^1, \dots, \Phi_1^s) : \mathbb{R}^{s(s-1)} \rightarrow \mathbb{R}^s$$

defined by

$$\Phi_1(p_{ij}, i = 1, \dots, s, j = 1, \dots, s-1) = (\mathbf{I}_r - \mathbf{P}\mathbf{I}_g)^{-1} \mathbf{P}\mathbf{g} = \Psi_1, \quad (3.25)$$

where, for any $i \in E$, we express p_{is} as a function of the arguments of Φ_1 in the obvious way, $p_{is} = 1 - \sum_{j=1}^{s-1} p_{ij}$. In this way we reduce the parameter space deleting the variables $p_{is}, \forall i \in E$ that are constrained by the values of the others parameters; thus we can proceed to apply our methodology on the reduced parameter space, i.e., $\{p_{ij}, i = 1, \dots, s, j = 1, \dots, s-1\}$.

First, we will give some preliminary results, useful for obtaining the asymptotic normality of the estimators; in all these results, $\mathbf{P}^0 := \mathbf{I}$.

Lemma 3.7. (see, e.g., Sadek and Limnios, 2002) For any $i, j \in E$ we have

$$\frac{\partial \mathbf{P}^n}{\partial p_{ij}} = \sum_{k=1}^n \mathbf{P}^{k-1} \frac{\partial \mathbf{P}}{\partial p_{ij}} \mathbf{P}^{n-k}, \quad (3.26)$$

where

$$\frac{\partial \mathbf{P}}{\partial p_{ij}} = (a_{lk})_{l,k \in E}, \text{ with } a_{lk} := \begin{cases} 1, & \text{if } l = i, k = j, \\ -1, & \text{if } l = i, k = s, \\ 0, & \text{otherwise.} \end{cases} \quad (3.27)$$

Lemma 3.8. Let $n \in \mathbb{N}$ and assume that the inverse $(\mathbf{I}_r^n - \mathbf{P}\mathbf{I}_g^n)^{-1}$ exists. Then, for any $i, j \in E$, we have

$$\frac{\partial (\mathbf{I}_r^n - \mathbf{P}\mathbf{I}_g^n)^{-1}}{\partial p_{ij}} = \mathbf{I}_g^n (\mathbf{I}_r^n - \mathbf{P}\mathbf{I}_g^n)^{-1} \cdot \frac{\partial \mathbf{P}}{\partial p_{ij}} \cdot (\mathbf{I}_r^n - \mathbf{P}\mathbf{I}_g^n)^{-1}. \quad (3.28)$$

Proof. See appendix.

We are able now to propose an estimator of the price-dividend ratio. The estimator of the price-dividend ratio is obtained by plug-in of the transition probability matrix estimator in (3.19). Thus we have

$$\begin{aligned}\widehat{\Psi}_1(m) &= (\widehat{\psi}_1(g_1; m), \dots, \widehat{\psi}_1(g_n; m))^\top := (\mathbf{I}_r - \widehat{\mathbf{P}}(m)\mathbf{I}_g)^{-1} \widehat{\mathbf{P}}(m)\mathbf{g} \\ &= \left(\mathbf{I}_r^{-1} \sum_{n=0}^{\infty} \widehat{\mathbf{P}}^n(m) \mathbf{I}_g \mathbf{I}_r^{-n} \right) \widehat{\mathbf{P}}(m)\mathbf{g},\end{aligned}\quad (3.29)$$

where $\widehat{\mathbf{P}}(m) = (\widehat{p}_{ij}(m))_{i,j \in E}$, with $\widehat{p}_{ij}(m)$ the classical MLE given in (3.23).

Remark 3.9. Note that it is possible that the inverse $(\mathbf{I}_r - \widehat{\mathbf{P}}(m)\mathbf{I}_g)^{-1}$ does not exist for some m ; nonetheless, taking into account that the inverse $(\mathbf{I}_r - \mathbf{P}\mathbf{I}_g)^{-1}$ exists (this is equivalent to the fact that the system (3.10) has a unique solution, cf. Proposition 3.2) and that $\widehat{\mathbf{P}}(m)$ is *a.s.* convergent to $\mathbf{P}(m)$, as m goes to infinity (cf. Proposition 3.6), we get that $(\mathbf{I}_r - \widehat{\mathbf{P}}(m)\mathbf{I}_g)^{-1}$ exists starting from a certain m^* .

Let $\mathcal{M}_{s \times s}$ be the set of real matrices of order $s \times s$. The following asymptotic results hold true.

Theorem 3.10. *The estimator of the price-dividend ratio proposed in (3.29) is:*

1. *strongly consistent, as m goes to infinity, i.e.,*

$$\widehat{\Psi}_1(m) \xrightarrow[m \rightarrow \infty]{a.s.} \Psi_1; \quad (3.30)$$

2. *asymptotically normal, as m goes to infinity, i.e.,*

$$\sqrt{m} \left(\widehat{\Psi}_1(m) - \Psi_1 \right) \xrightarrow[m \rightarrow \infty]{\mathcal{D}} \mathcal{N}_s(\mathbf{0}, \Sigma_1), \quad (3.31)$$

where the covariance matrix Σ_1 has the form

$$\Sigma_1 = \Phi'_1 \Gamma(\Phi'_1)^\top \in \mathcal{M}_{s \times s}, \quad (3.32)$$

where:

- $\mathbf{\Gamma} \in \mathcal{M}_{s(s-1) \times s(s-1)}$ is the restriction of $\tilde{\mathbf{\Gamma}}$ given in (3.24) to $s(s-1) \times s(s-1)$, in the sense that $\mathbf{\Gamma}$ is the asymptotic covariance matrix of the vector $(\sqrt{m}(\hat{p}_{ij}(m) - p_{ij}))_{i=1, \dots, s, j=1, \dots, s-1}$;
- $\mathbf{\Phi}'_1 = \left(\frac{\partial \Phi_1^i}{\partial p_{lk}} \right)_{i,l=1, \dots, s, k=1, \dots, s-1} \in \mathcal{M}_{s \times s(s-1)}$ is the partial derivative matrix of Φ_1 with respect to $(p_{ij}, i = 1, \dots, s, j = 1, \dots, s-1)$; detailed expression of this matrix will be given later in (A.18) and (A.19).

Proof. See appendix.

Estimation of the second-order price-dividend ratio

The estimation of the second order price-dividend ratio is obtained by plug-in of the transition probability matrix estimator and of the estimator of the first-order price-dividend ratio in (3.20). Thus we have

$$\begin{aligned}
\widehat{\Psi}_2(m) &= (\widehat{\psi}_2(g_1; m), \dots, \widehat{\psi}_2(g_n; m))^\top \\
&:= \left(\mathbf{I}_r^2 - \widehat{\mathbf{P}}(m) \cdot \mathbf{I}_g^2 \right)^{-1} \cdot \widehat{\mathbf{P}}(m) \cdot \left(\mathbf{g} \diamond \mathbf{g} + 2\widehat{\Psi}_1(m) \diamond \mathbf{g} \diamond \mathbf{g} \right) \\
&= \left(\mathbf{I}_{r^2}^{-1} \sum_{n=0}^{\infty} \widehat{\mathbf{P}}^n(m) \mathbf{I}_{g^2} \mathbf{I}_{r^2}^{-n} \right) \widehat{\mathbf{P}}(m) \cdot \left(\mathbf{g} \diamond \mathbf{g} + 2\widehat{\Psi}_1(m) \diamond \mathbf{g} \diamond \mathbf{g} \right), \quad (3.33)
\end{aligned}$$

where $\widehat{\mathbf{P}}(m) = (\widehat{p}_{ij}(m))_{i,j \in E}$, with $\widehat{p}_{ij}(m)$ the classical MLE given in (3.23) and $\widehat{\Psi}_1(m)$ given in (3.29).

We will state here a remark similar to Remark 3.9.

Remark 3.11. Note that it is possible that the inverse $\left(\mathbf{I}_r^2 - \widehat{\mathbf{P}}(m) \cdot \mathbf{I}_g^2 \right)^{-1}$ does not exist for some m ; nonetheless, taking into account that the inverse $\left(\mathbf{I}_r^2 - \mathbf{P} \cdot \mathbf{I}_g^2 \right)^{-1}$ exists (this is equivalent to the fact that the system (3.13) has unique solution, cf. Proposition 3.4) and that $\widehat{\mathbf{P}}(m)$ is *a.s.* convergent to $\mathbf{P}(m)$, as m goes to infinity (cf. Proposition 3.6), we get that $\left(\mathbf{I}_r^2 - \widehat{\mathbf{P}}(m) \cdot \mathbf{I}_g^2 \right)^{-1}$ exists starting from a certain m^* .

In order to investigate the asymptotic properties of $\widehat{\Psi}_2(m)$ as m goes to infinity, note that, according to (3.20), Ψ_2 is expressed as a continuous and differentiable function of \mathbf{P} . Nonetheless, it will be convenient to express Ψ_2 as a function of independent variables. Consequently, let us consider

$$\Phi_2 = (\Phi_2^1, \dots, \Phi_2^s) : \mathbb{R}^{s(s-1)} \rightarrow \mathbb{R}^s$$

defined by

$$\begin{aligned} & \Phi_2(p_{ij}, i = 1, \dots, s, j = 1, \dots, s-1) \\ &= (\mathbf{I}_r^2 - \mathbf{P} \cdot \mathbf{I}_g^2)^{-1} \cdot \mathbf{P} \cdot (\mathbf{g} \diamond \mathbf{g} + 2\Psi_1 \diamond \mathbf{g} \diamond \mathbf{g}) = \Psi_2, \end{aligned} \quad (3.34)$$

where, for any $i \in E$, we express p_{is} as a function of the arguments of Φ_2 , $p_{is} = \sum_{j=1}^{s-1} p_{ij}$. Let us now give the analogous of Theorem 3.10 for the second order price-dividend ratio.

Theorem 3.12. *The estimator of the second order price-dividend ratio proposed in (3.33) is:*

1. *strongly consistent, as m goes to infinity, i.e.,*

$$\widehat{\Psi}_2(m) \xrightarrow[m \rightarrow \infty]{a.s.} \Psi_2; \quad (3.35)$$

2. *asymptotically normal, as m goes to infinity, i.e.,*

$$\sqrt{m} \left(\widehat{\Psi}_2(m) - \Psi_2 \right) \xrightarrow[m \rightarrow \infty]{\mathcal{D}} \mathcal{N}_s(\mathbf{0}, \Sigma_2), \quad (3.36)$$

where the covariance matrix Σ_2 has the form

$$\Sigma_2 = \Phi_2' \Gamma(\Phi_2')^\top \in \mathcal{M}_{s \times s}, \quad (3.37)$$

where:

- $\mathbf{\Gamma} \in \mathcal{M}_{s(s-1) \times s(s-1)}$ is the restriction of $\tilde{\mathbf{\Gamma}}$ given in (3.24) to $s(s-1) \times s(s-1)$, in the sense that $\mathbf{\Gamma}$ is the asymptotic covariance matrix of the vector $(\sqrt{m}(\hat{p}_{ij}(m) - p_{ij}))_{i=1, \dots, s, j=1, \dots, s-1}$;
- $\mathbf{\Phi}'_2 = \left(\frac{\partial \Phi_2^i}{\partial p_{ik}} \right)_{i,l=1, \dots, s, k=1, \dots, s-1} \in \mathcal{M}_{s \times s(s-1)}$ is the partial derivative matrix of Φ_2 with respect to $(p_{ij}, i = 1, \dots, s, j = 1, \dots, s-1)$; detailed expression of this matrix will be given later in (A.21) and (A.22).

Proof. See appendix.

Estimation of the forecasted fundamental prices

Using similar techniques it is possible to give an estimator of the forecasted fundamental prices obtained in (3.22). This estimator is given by

$$\begin{pmatrix} \widehat{E}^{(n)}(m)p(d_k, g_1) \\ \vdots \\ \widehat{E}^{(n)}(m)p(d_k, g_s) \end{pmatrix} = d_k \widehat{\mathbf{P}}^n(m) \mathbf{I}_{\mathbf{g}^n} \begin{pmatrix} \widehat{\psi}_1(g_1; m) \\ \vdots \\ \widehat{\psi}_1(g_s; m) \end{pmatrix}, \quad (3.38)$$

where $\widehat{\mathbf{P}}(m) = (\widehat{p}_{ij}(m))_{i,j \in E}$, with $\widehat{p}_{ij}(m)$ the classical MLE given in (3.23) and $\widehat{\Psi}_1(m)$ given in (3.29).

For any $n \in \mathbb{N}^*$, $(E^{(n)}p(d_k, g_1), \dots, E^{(n)}p(d_k, g_s))^\top$ can be expressed as a differentiable function of \mathbf{P} . Nonetheless, it is convenient to express it as a function of independent variables. Consequently, let us consider

$$\Theta = (\Theta^1, \dots, \Theta^s) : \mathbb{R}^{s(s-1)} \rightarrow \mathbb{R}^s$$

defined by

$$\begin{aligned} \Theta(p_{ij}, i = 1, \dots, s, j = 1, \dots, s-1) &= d_k \mathbf{P}^n(m) \mathbf{I}_{\mathbf{g}^n} \Psi_1 \\ &= (E^{(n)}p(d_k, g_1), \dots, E^{(n)}p(d_k, g_s))^\top, \end{aligned} \quad (3.39)$$

where, for any $i \in E$, we express p_{is} as a function of the arguments of Θ , $p_{is} =$

$$\sum_{j=1}^{s-1} p_{ij}.$$

Theorem 3.13. *The estimator of the forecasted fundamental prices given in (3.38)*

is:

1. *strongly consistent, as m goes to infinity, i.e.,*

$$\begin{pmatrix} \widehat{E}^{(n)}(m)p(d_k, g_1) \\ \vdots \\ \widehat{E}^{(n)}(m)p(d_k, g_s) \end{pmatrix} \xrightarrow[m \rightarrow \infty]{a.s.} \begin{pmatrix} E^{(n)}p(d_k, g_1) \\ \vdots \\ E^{(n)}p(d_k, g_s) \end{pmatrix}; \quad (3.40)$$

2. *asymptotically normal, as m goes to infinity, i.e.,*

$$\sqrt{m} \left[\begin{pmatrix} \widehat{E}^{(n)}(m)p(d_k, g_1) \\ \vdots \\ \widehat{E}^{(n)}(m)p(d_k, g_s) \end{pmatrix} - \begin{pmatrix} E^{(n)}p(d_k, g_1) \\ \vdots \\ E^{(n)}p(d_k, g_s) \end{pmatrix} \right] \xrightarrow[m \rightarrow \infty]{\mathcal{D}} \mathcal{N}_s(\mathbf{0}, \Sigma_{\Theta}), \quad (3.41)$$

where the covariance matrix Σ_{Θ} has the form

$$\Sigma_{\Theta} = \Theta' \Gamma (\Theta')^{\top} \in \mathcal{M}_{s \times s}, \quad (3.42)$$

where:

- $\Gamma \in \mathcal{M}_{s(s-1) \times s(s-1)}$ is the restriction of $\tilde{\Gamma}$ given in (3.24) to $s(s-1) \times s(s-1)$, in the sense that Γ is the asymptotic covariance matrix of the vector $(\sqrt{m}(\widehat{p}_{ij}(m) - p_{ij}))_{i=1, \dots, s, j=1, \dots, s-1}$;
- $\Theta' = \left(\frac{\partial \Theta^i}{\partial p_{lk}} \right)_{i,l=1, \dots, s, k=1, \dots, s-1} \in \mathcal{M}_{s \times s(s-1)}$ is the partial derivative matrix of Θ with respect to $(p_{ij}, i = 1, \dots, s, j = 1, \dots, s-1)$; detailed expression of this matrix will be given later in (A.24) and (A.25).

Proof. See appendix.

Obs	Mean	Median	Min	Max	Std
1739	13.9200	11.9105	4.7315	43.5155	7.1223

Table 3.1: Descriptive statistics of the dividend series

3.3 Application to real dividend data

The Markov stock model developed in this paper is illustrated by analysing a sample of Stock Market Data given in Shiller (2005). The data can be downloaded from <http://www.econ.yale.edu/~shiller/data.htm>. It has been updated by Robert J. Shiller in order to cover S&P 500 index data on composite stock prices, dividends and earnings on a monthly basis since January 1871¹. The values of the nominal dividends included in the data represent 12 months moving sums adjusted to index for the last quarter of the year. For our purpose, we use the real dividends computed from the nominal ones adjusted for inflation. The Table 3.1 shows some descriptive statistics of the dataset.

The dividend growth sequence has been calculated according to the equation (3.7).

To apply the Markov model it is necessary to discretize the dividend growth distribution in as many bins as the desired number of states. We fix the width of the discretization intervals to a value that we identify with the standard deviation of the dividend growth σ , centering the discretization around the value 1, that is the absence of growth. Indeed, using this approach of discretization makes evident that the maximum allowable states with the available data is $\frac{\max(g)-\min(g)}{\sigma}$ where g is the dividend growth sequence; in our dataset this ratio is $\cong 11$. Moreover, in order to include all the data, the width of the external bins will be different from σ . The discretization result for a five states model is shown in Figure 3.1.

Once the number of states has been selected, we can calculate their values through the median value of the dividend growth in every interval and also the

¹“Monthly dividend and earnings data are computed from the S&P four-quarter totals for the quarter since 1926, with linear interpolation to monthly figures. Dividend and earnings data before 1926 are from Cowles and associates (Common Stock Indexes, 2nd ed. [Bloomington, Ind.: Principia Press, 1939]), interpolated from annual data”

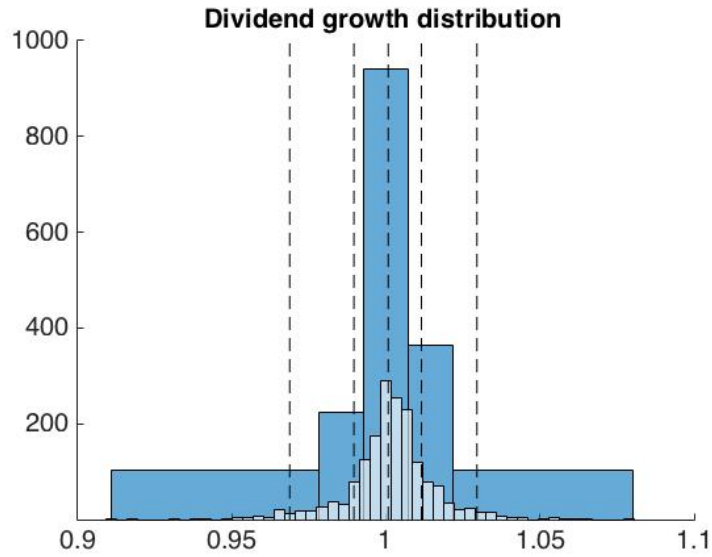


Figure 3.1: Discretization result for five states

State	Dividend growth range	State value
1	0.911 - 0.978	0.969
2	0.979 - 0.993	0.990
3	0.994 - 1.007	1.001
4	1.008 - 1.022	1.011
5	1.023 - 1.080	1.030

Table 3.2: Dividend growth discretization with five states

one step transition probabilities matrix using the MLE given in (3.23).

Setting, for example, the number of states equal to 5, the edges of the intervals identifying the dividend growth range and the value of the states are indicated in Table 3.2 and the estimated transition probability matrix, with the log-likelihood value of -1776.733 , is shown in Table 3.3.

From the transition probabilities we calculate the one step ahead forecast starting from January 1871 to the end of the series. Considering that the first rate of growth $g = 0.9704$ falls into the first state, we can use the first row of the matrix to predict

P_{ij}	$j = 1$	2	3	4	5
$i = 1$	0.495	0.267	0.152	0.057	0.029
2	0.134	0.424	0.335	0.067	0.040
3	0.015	0.076	0.740	0.144	0.024
4	0.016	0.066	0.368	0.484	0.066
5	0.019	0.049	0.194	0.311	0.427

Table 3.3: Transition probability matrix with five states

the following growth rate, thus the next dividend value. In general, if at current time t the growth value of dividend is in state $i \in E$, then next dividend growth is calculated as:

$$g(t+1) = p_{i1}g_1 + p_{i2}g_2 + p_{i3}g_3 + p_{i4}g_4 + p_{i5}g_5 \quad (3.43)$$

and the dividend value by

$$D_f^{\text{one}}(t+1) = D_r(t)g(t+1) \quad (3.44)$$

where $D_r(t)$ represents the real dividend observed from the data at time t and $D_f^{\text{one}}(t+1)$ is the one step forecasted dividend at time $t+1$ using the Markov chain model.

We also computed the all steps forecast $D_f^{\text{all}}(t+1)$ at time $t+1$ using the forecasted dividend $D_f^{\text{all}}(t)$ at time t through the relation

$$D_f^{\text{all}}(t+1) = D_f^{\text{all}}(t)g(t+1), \quad (3.45)$$

where $D_f^{\text{all}}(0)$ is $D_r(0)$.

The results are shown in Figure 3.2. The right panel represents the same data as the left panel, zoomed from time 925 to 975, in order to show the extremely good predictability of the model using the one step ahead forecasting.

Then, in order to obtain an indicator of the quality of the prediction we calculated the Root Mean Square Error (RMSE) which is a measure of the distance between the forecasted data and the real data:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (D_r(t) - D_f(t))^2}, \quad (3.46)$$

where N is the length of the series, $D_r(t)$ represents the real dividend, while $D_f(t)$ represents either the one step forecasted dividend at time t , $D_f^{\text{one}}(t)$, or the all step forecasted dividend at time t , $D_f^{\text{all}}(t)$. In the model with five states the value of the RMSE is 0.095 for the one step analysis and 4.325 for the all step analysis.

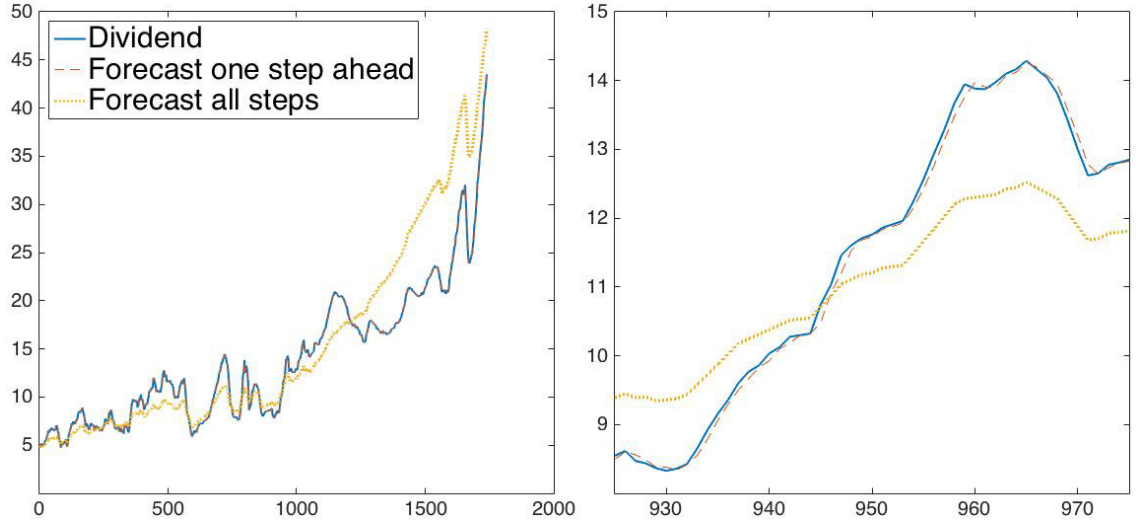


Figure 3.2: Dividend forecast

A repetition of the computation for different values of the number of states allows us to set the optimal number of states for the model as the value that minimize the RMSE. The values are represented in Figure 3.3.

The figure shows a decreasing trend for the one step forecast that suggests that the prediction becomes more accurate when increasing the numbers of states of the model. In this way, we decided to work with a five state model because models with a higher number of states improve the results very modestly and, at the same time, could make the estimation results worse due to increasing the parameter space of the model.

Once the optimal number of states for these available data has been chosen, the next step is the computation of the price dividend ratio $\psi_1(g_k)_{k \in E}$ and of the second order moment of the price $\psi_2(g_k)_{k \in E}$. To this end, it is necessary to consider the proper value of the rate of return of the analyzed data. Therefore, we calculate the proper discount factor that coincides with the rate of return of the S&P 500. For our analysis, we calculated the historical rate of return of the S&P 500 since January 1871, as Compound Annual Growth Rate, corresponding to the geometric mean of the annual market returns. Its value, adjusted for the inflation, is 6.3264%. This value verifies both assumptions A1 and A2.

The computation of the estimators of the price-dividend ratio and of the second-

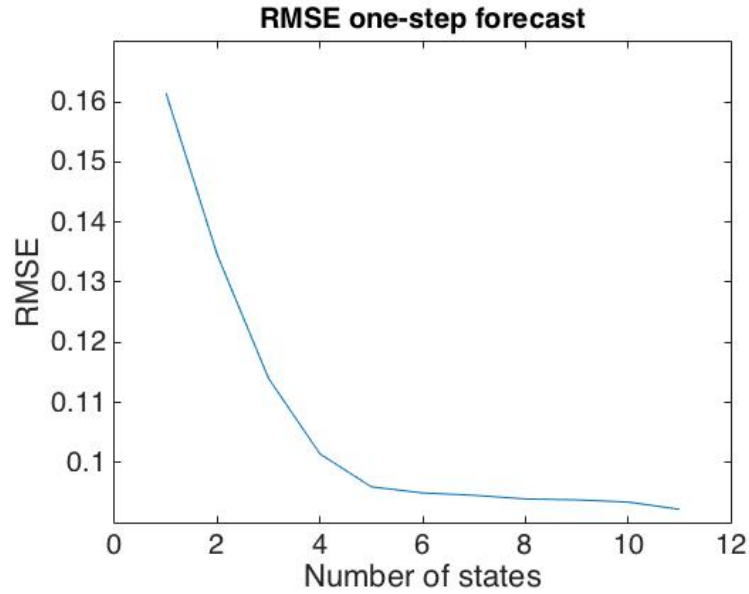


Figure 3.3: RMSE values for one step forecast

State	Price dividend ratio $\psi_1(g)$	2^{nd} order Price dividend ratio $\psi_2(g)$
1	15.516 ± 0.190	242.028 ± 5.742
2	15.856 ± 0.163	252.559 ± 5.010
3	16.155 ± 0.140	261.893 ± 4.348
4	16.300 ± 0.149	266.662 ± 4.659
5	16.560 ± 0.177	275.403 ± 5.564

Table 3.4: Price dividend ratio and second order moment values for the five states (Intervals at confidence level $\alpha = 5\%$)

order moment is done by means of formulas (3.29) and (3.33) and the values are shown in Table 3.4, with the lower values associated to state 1 and the greater values to state 5. In the same table there are also reported the confidence intervals for the first and second-order price-dividend ratios that are evaluated using the asymptotic covariance matrices calculated in Theorems 3.10 and 3.12. As a matter of example, the confidence interval for $\psi_1(g_i)$ at a confidence level $(1 - \alpha)$ is:

$$\hat{\psi}_1(g_i; m) - z_{1-\frac{\alpha}{2}} \frac{\sqrt{\Sigma_1(1,1)}}{\sqrt{m}} < \psi_1(g_i) < \hat{\psi}_1(g_i; m) + z_{1-\frac{\alpha}{2}} \frac{\sqrt{\Sigma_1(1,1)}}{\sqrt{m}},$$

where z_α is the quantile of order α of the standard normal distribution.

To assess the stability of this model we undertook a test on the width of the discretization bins, that is on the sensitivity of the model with respect to the edges that denote the state space. This test consists in modifying the widths of the bin

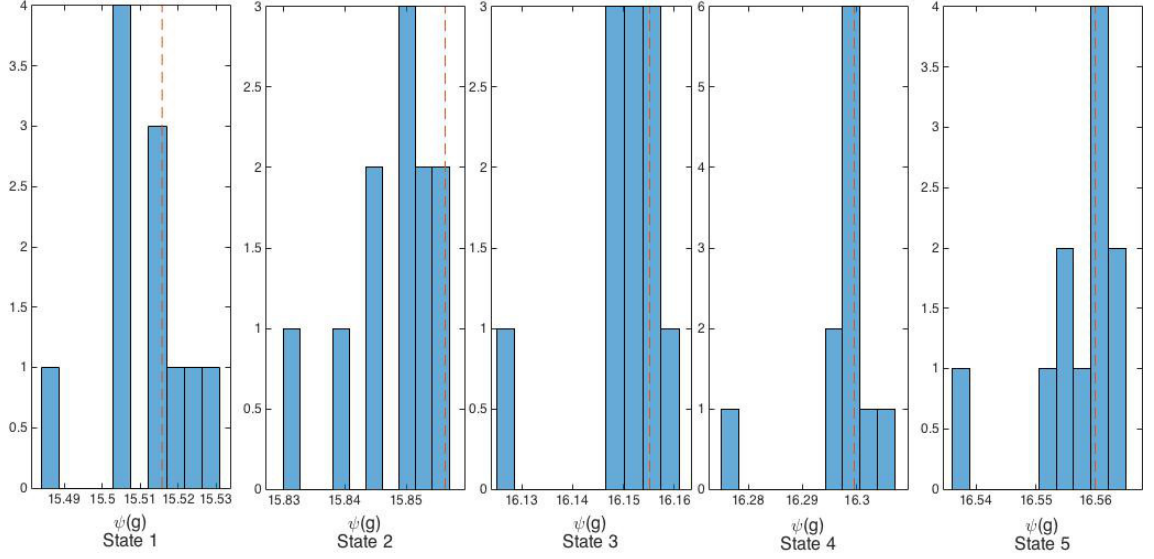


Figure 3.4: Stability test for every state.

State	$\min(\psi_1(g))$	$\max(\psi_1(g))$	$\text{mean}(\psi_1(g))$	$\text{std}(\psi_1(g))$	$\psi_1(g), \epsilon = 0$
1	15.4845	15.5306	15.5114	0.0125	15.5157
2	15.8304	15.8563	15.8478	0.0076	15.8563
3	16.1273	16.1600	16.1506	0.0084	16.1551
4	16.2761	16.3061	16.2970	0.0075	16.2996
5	16.5379	16.5644	16.5573	0.0075	16.5600

Table 3.5: Statistics related to the stability test.

with values $\sigma \pm \epsilon$. For the computation, we took 100 values of ϵ ranging from -10^{-4} and 10^{-4} at fixed intervals to have the corresponding price dividend ratio $\psi_1(g(k))$ for every state. The results of the test show that the model is quite stable to little variation of σ . Figure 3.4 shows the distribution of the values of the price-dividend ratio for all the five states. The red striped vertical line represents the price-dividend ratio when $\epsilon = 0$. The statistics for the stability test, shown in Table 3.5, clearly states the effectiveness of the discretization method based on the σ value.

The tested model shows a good predictability of future dividend, hence of future prices when analyzing all the dividend series from 1871. To make the test more realistic we now apply the model considering a limited amount of historical data of different time ranges in order to predict future values. First we fix the time interval of prediction, as a typical application would suggest, around 40 periods. Hence we chose the threshold date. From this threshold, going backward, we can calculate the

transition probabilities matrices based on the last 100 observations, the last 200 up to the last 500 past observations. Then we calculate the RMSE variations for the different time ranges in order to assess the effect of the inclusion of more historical data in the model. The RMSE values are then compared and plotted in Figure 3.5. Depending on the growth values included in the restricted dataset, the constraint given from $\frac{\max(g)-\min(g)}{\sigma}$ puts a limit to the maximum states allowed, like the 100, 200 and 300 periods cases. For every chosen period, the RMSE values decrease when increasing the number of states, especially for the one-step analysis, following the behavior that we have already analyzed. Moreover, the results reveal that small periods of analysis have approximately the same level of predictability as bigger periods.

3.4 Conclusion

In this chapter, we extended the Markov chain based dividend discount model introduced in Ghezzi and Piccardi (2003) by computing the second-order price-dividend ratio, that is a measure of risk to attach to the price-dividend ratio for measuring the profitability of an investment in a stock. In the second section, we developed non-parametric statistical techniques to estimate the financial quantities and the corresponding confidence interval starting from a time series of dividend data. Also, along with the estimators, we proposed their asymptotic properties. The chapter concludes with a real problem application where we analysed many practical problems that an analyst can face when applying the model, e.g. the determination of the number of states for the Markov chain, the determination of the states and of their values, the stability of the results with respect to the choice of the state space and the forecasting of dividend and fundamental value and risk.

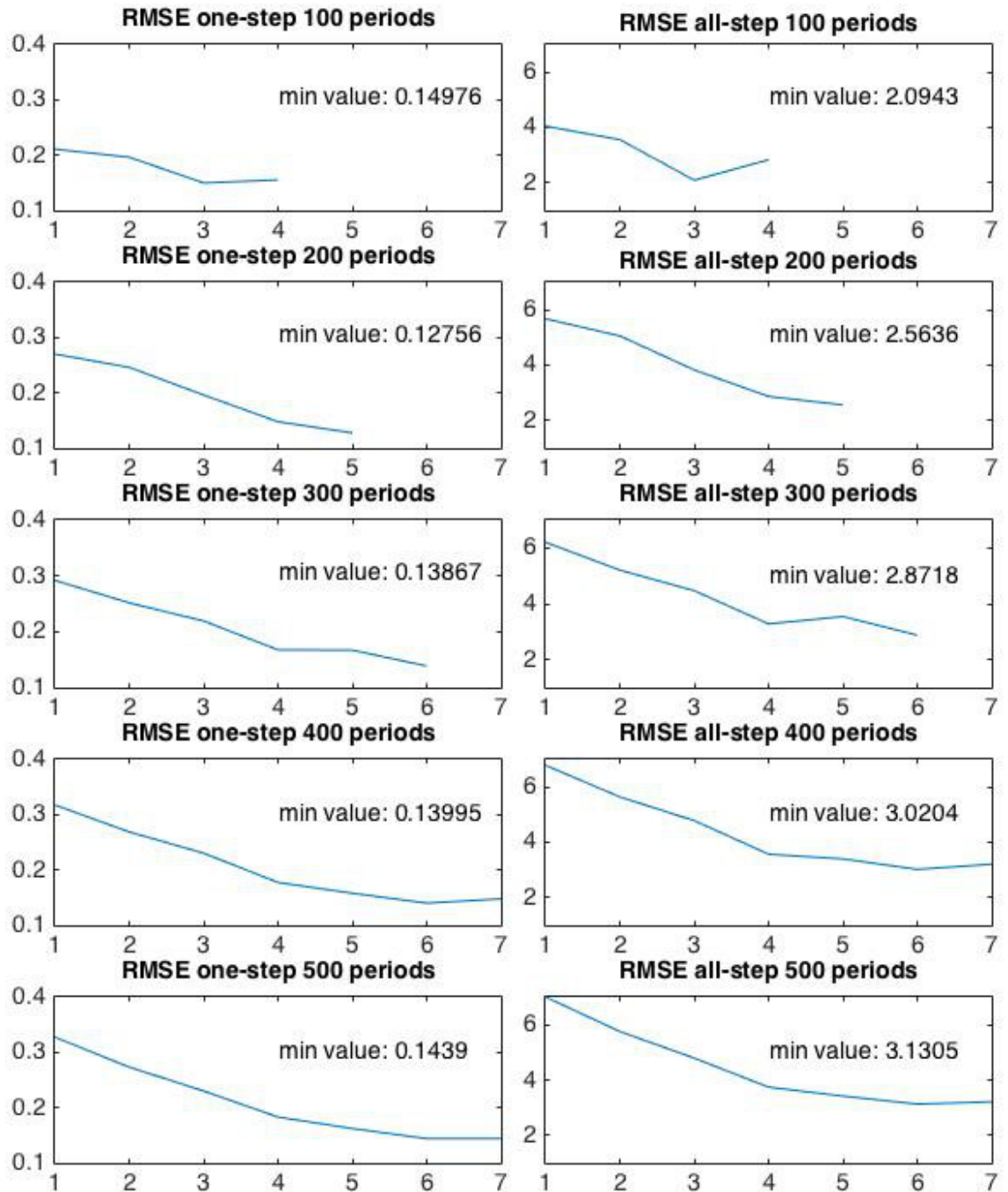


Figure 3.5: RMSE values for one-step and all step prediction for different time ranges of analysis

4 A multivariate Markov chain stock model

In this chapter we present a dividend stock valuation model where multiple dividend growth series and their dependencies are modelled using a multivariate Markov chain, and extend the model proposed in previous chapter to a multivariate setting. Like in previous model, we compute the first and second order price-dividend ratios by solving corresponding linear systems of equations and show that a different price-dividend ratio is attached to each combination of states of the dividend growth process of each stock. Subsequently, we provide a formula for the computation of the variances and covariances between stocks in a portfolio. Finally, we apply the theoretical model to the dividend series of three US stocks and perform comparisons with existing models.

This chapter is organised as follows: in Section 4.1 we review the basic dividend valuation model and then we define the multivariate Markov model. Then, Section 4.2 presents an application of the theoretical model to three US stocks and discusses the validity of the results in a comparison with other dividend valuation models. All proofs are left in the Appendix.

4.1 Model

In this section we first present general information on fundamental analysis, i.e. how to price firms on the basis of fundamentals and successively we advance a multivariate Markov chain model as a suitable environment for the valuation of the

returns and risk of a firm. The multivariate model allows us to consider a dependence structure between a pool of stocks in the pricing mechanism in order to measure the influence of each stock on the other stocks.

4.1.1 The Basic Dividend Valuation Model

Suppose a stock is paying dividends in time. The value of the dividends is not known in advance and then a common choice is to consider it as generated by a discrete time random process $\{D(k)\}_{k \in \mathbb{N}}$. Since dividends are a measure of profitability of the firm, the value of the corresponding stock can be expressed in terms of discounted future dividends and therefore is a realisation of a discrete time stochastic process $\{\mathcal{P}(k)\}_{k \in \mathbb{N}}$. In an efficient market, price and dividend obey the relation

$$\mathcal{P}(k) = \frac{D(k+1) + \mathcal{P}(k+1)}{r}, \quad (4.1)$$

where r is one plus the required rate of return on the stock. Relation (4.1) suggests that at time k the stock is valued after the dividend $D(k) = d(k)$ has been paid.

This formula equates the value at time k to the dividend at $k+1$, which is due to the ownership of the stock during the interval $[k, k+1)$, plus the value of the stock at time $k+1$. These two quantities are divided by r in order to discount them at time k .

Let $p(k) := \mathbb{E}_{(k)}[\mathcal{P}(k)]$ be the fundamental price at time k . From relation (4.1) we get

$$p(k) = \frac{\mathbb{E}_{(k)}[D(k+1) + \mathcal{P}(k+1)]}{r}, \quad (4.2)$$

which expresses the value of the stock in a dividend based present value model. As it is well known, see for example Samuelson (1973), if we assume as transversality condition that

$$\lim_{i \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[\mathcal{P}(k+i)]}{r^i} = 0, \quad (4.3)$$

the unique solution of (4.2) is expressed by the series

$$p(k) = \sum_{i=1}^{+\infty} \frac{\mathbb{E}_{(k)}[D(k+i)]}{r^i}. \quad (4.4)$$

Formula (4.4) evaluates the stock as a function of expected future dividend stream and discount rates.

If condition (4.3) is not assumed, Blanchard and Watson (1982) proved that different solutions of the fundamental equation can exist, thus revealing the presence of bubbles.

Since dividends are random, they are uncertain and accordingly prices will show deviations from their expected value. For this reason it is important to introduce a risk measure able to express the effects of dividends uncertainty into fundamental prices. This problem has been recently investigated in a multinomial model and in Markov and semi-Markov chain model by Agosto and Moretto (2015), Barbu et al. (2017), D'Amico (2017). To quantify this effect, according to Barbu et al. (2017) we introduce a risk measure:

$$\mathcal{P}_2(k) := (\mathcal{P}(k))^2 = \left(\frac{D(k+1) + \mathcal{P}(k+1)}{r} \right)^2. \quad (4.5)$$

Successive substitutions in future prices $\mathcal{P}(k+N)$ of relation (4.1) yield

$$\begin{aligned} \mathcal{P}_2(k) &= \sum_{i=1}^N \frac{D^2(k+i)}{r^{2i}} + 2 \sum_{i=1}^N \sum_{j>i}^N \frac{D(k+i)D(k+j)}{r^{i+j}} \\ &+ 2 \sum_{i=1}^N \frac{D(k+i)\mathcal{P}(k+N)}{r^{i+N}} + \frac{\mathcal{P}_2(k+N)}{r^{2N}}. \end{aligned} \quad (4.6)$$

Now, set $p_2(k) := \mathbb{E}_{(k)}[\mathcal{P}_2(k)]$ and assume that

$$\lim_{N \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[\mathcal{P}_2(k+N)]}{r^{2N}} = 0, \quad (4.7)$$

and

$$\lim_{N \rightarrow +\infty} \sum_{i=1}^N \frac{\mathbb{E}_{(k)}[D(k+i)\mathcal{P}(k+N)]}{r^{i+N}} = 0, \quad (4.8)$$

hold true, then we can deduce that:

$$p_2(k) = \sum_{i=1}^{+\infty} \frac{\mathbb{E}_{(k)}[D^2(k+i)]}{r^{2i}} + 2 \sum_{i=1}^{+\infty} \sum_{j>i} \frac{\mathbb{E}_{(k)}[D(k+i)D(k+j)]}{r^{i+j}}. \quad (4.9)$$

Expression (4.9) is called the fundamental formula of the risk. It expresses the risk in term of the second order moment of the value process, see Barbu et al. (2017), D'Amico (2017). From the formula it is possible to recognise risk as a function of the second order and product moments of the dividend process.

In a recent paper, Agosto et al. (2018) considered the problem of computing the covariance between two stocks that may be held in a portfolio by an investor. In general the two stocks may be correlated and it is important to be able to consider this effect in the financial evaluation of a stock or of a portfolio. In this respect, let us denote by $\mathcal{M}^{(\alpha,\beta)}(k)$ the price product between stocks α and β , i.e.

$$\mathcal{M}^{(\alpha,\beta)}(k) := \mathcal{P}^{(\alpha)}(k) \cdot \mathcal{P}^{(\beta)}(k) = \prod_{a \in \{\alpha,\beta\}} \left(\frac{D^{(a)}(k+1) + \mathcal{P}^{(a)}(k+1)}{r_a} \right),$$

where r_a denotes one plus the required rate of return for stock a . Successive substitutions of relation (4.1) for each stock into future prices $\mathcal{P}^a(k+N)$, $a \in \{\alpha,\beta\}$ together with the fulfilment of condition

$$\lim_{N \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[\mathcal{P}^{(\alpha)}(k+N)\mathcal{P}^{(\beta)}(k+N)]}{r_\alpha^N \cdot r_\beta^N} = 0,$$

and condition

$$\lim_{N \rightarrow +\infty} \sum_{i=1}^N \mathbb{E}_{(k)} \left[\frac{D^{(\alpha)}(k+i)\mathcal{P}^{(\beta)}(k+N)}{r_\alpha^i \cdot r_\beta^N} \right] = \lim_{N \rightarrow +\infty} \sum_{i=1}^N \mathbb{E}_{(k)} \left[\frac{D^{(\beta)}(k+i)\mathcal{P}^{(\alpha)}(k+N)}{r_\alpha^N \cdot r_\beta^i} \right] = 0,$$

yield

$$\begin{aligned}
p_2^{(\alpha,\beta)}(k) &:= \mathbb{E}_{(k)}[\mathcal{M}^{(\alpha,\beta)}(k)] \\
&= \sum_{i=1}^{+\infty} \mathbb{E}_{(k)} \left[\frac{D^{(\alpha)}(k+i)D^{(\beta)}(k+i)}{r_\alpha^i \cdot r_\beta^i} \right] \\
&+ \sum_{i=1}^{+\infty} \sum_{j>i} \mathbb{E}_{(k)} \left[\frac{D^{(\alpha)}(k+i)D^{(\beta)}(k+j)}{r_\alpha^i \cdot r_\beta^j} \right] \\
&+ \sum_{i=1}^{+\infty} \sum_{j>i} \mathbb{E}_{(k)} \left[\frac{D^{(\alpha)}(k+j)D^{(\beta)}(k+i)}{r_\alpha^j \cdot r_\beta^i} \right].
\end{aligned} \tag{4.10}$$

We call this expression *the fundamental formula of the price-product*. The computation of formula (4.10) depends on the joint dynamics of the two stocks. A first attempt in this direction was done by Agosto et al. (2018) where a Markov chain with state space equal to the set of possible couples of growth-dividend values for both stocks is considered. We know, however, that this simple strategy cannot be implemented in real applications especially when multiple stocks are considered as dependent because there is a dramatic increase in the number of parameters to be estimated. As a matter of example, if we consider a portfolio of m stocks and each one may assume d different values of the growth-dividend process then a total of $d^m \cdot (d^m - 1)$ transition probabilities have to be estimated.

4.1.2 A Multivariate Markov Model of Value

In this section, we propose a multivariate Markov chain model for stock valuation. The model belongs to the class of mixture transition distribution models which go back to the pioneering work of Raftery (1985). The idea of creating a mixture of Markov transition probabilities to represent a multivariate Markov model has found a detailed theoretical and applied description in the book by Ching and Ng (2006) where an almost complete list of then available references to this class of stochastic processes can be recovered.

Let us consider the problem of how to price a stock when we have at our disposal information about several stocks that may constitute our financial portfolio. As it is well known in finance, it is important to apply a model that considers the

dependence structure that eventually characterise the pool of stocks. Let us assume that our portfolio is constituted of γ stocks. For each $\alpha = 1, 2, \dots, \gamma$ we denote by $\{D^{(\alpha)}(k)\}_{k \in \mathbb{N}}$ the dividend process. We assume that

$$D^{(\alpha)}(k+1) = G^{(\alpha)}(k+1) \cdot D^{(\alpha)}(k), \quad (4.11)$$

where $\{G^{(\alpha)}\}_{k \in \mathbb{N}}$ is the growth-dividend random process for stock α . Relation (4.11) affirms that we are considering a geometric model of increases/decreases in the dividends. Some recent contributions were based on the same idea of geometric model of Markov and semi-Markov type, see e.g. Ghezzi and Piccardi (2003), Barbu et al. (2017), D'Amico (2017). Although widely used, the Markov and semi-Markov model are subject to an important limitation, i.e. independence of the dividend processes of the considered stocks. For this reason, we consider a generalisation where the growth-dividend processes $\{G^{(1)}(k), G^{(2)}(k), \dots, G^{(\gamma)}(k)\}_{k \in \mathbb{N}}$ form a multivariate Markov chain.

Let $A_i^{(\alpha)}(n) := \mathbb{P}[G^{(\alpha)}(n) = i]$ be probability of growth-dividend of stock α to be at time n in state i . Define the corresponding vector of probability distribution $\mathbf{A}^\alpha(n) := [A_1^{(\alpha)}, \dots, A_m^{(\alpha)}]$.

In the multivariate Markov chain model the following relationship is formulated

$$\mathbf{A}^{(\alpha)}(n+1) = \sum_{\beta=1}^{\gamma} \mathbf{A}^{(\beta)}(n) \cdot \lambda_{\beta,\alpha} \cdot \mathbf{P}^{(\beta,\alpha)}, \quad (4.12)$$

where $\lambda_{\beta,\alpha} \in [0, 1]$, $\sum_{\beta=1}^{\gamma} \lambda_{\beta,\alpha} = 1$ and $\mathbf{P}^{(\beta,\alpha)}$ is the transition probability matrix of stock α given the state occupied one time step before by stock β , i.e.

$$\mathbf{P}_{i,j}^{(\beta,\alpha)} = \mathbb{P}[G^{(\alpha)}(n+1) = j \mid G^{(\beta)}(n) = i]. \quad (4.13)$$

According to equation (4.12) the probability distribution function of the growth-dividend process at time $n+1$ for the α stock depends not only on the state of the growth-dividend process of the same stock at time n , but on the set of states visited

by every stock in the portfolio at time n .

A consequence of the adopted multivariate model is that the price of the α stock at any time k , given all information available in the market at that time, depends on the vector of states of the growth-dividend processes of the pool of stocks, i.e. $\mathbf{G}(k) = \mathbf{g}(k)$. Indeed, as a direct generalisation of (4.4) we have

$$\begin{aligned} p^{(\alpha)}(\mathbf{g}(k)) &= \sum_{i=1}^{+\infty} \frac{\mathbb{E}_{(k)}[D^{(\alpha)}(k+i)]}{r_{\alpha}^i} \\ &= \sum_{i=1}^{+\infty} \left(\frac{\mathbb{E}_{(k)}[\prod_{j=1}^i G^{(\alpha)}(k+j)]}{r_{\alpha}^i} \right) d^{(\alpha)}(k). \end{aligned} \quad (4.14)$$

Let $\mathbf{e}^{(\alpha)} \in E$ be the vector denoting the state of the α -stock, i.e. $\mathbf{e}^{(\alpha)} = (0, \dots, 0, 1, 0, \dots, 0)$ with the non zero element in the j -th position, means that the stock α occupies a state of the growth-dividend process equal to g_j . The state of the growth-dividend process for the pool of stocks can be represented by means of $E = \{\mathbf{e}_1^{(\alpha)}, \mathbf{e}_2^{(\alpha)}, \dots, \mathbf{e}_m^{(\alpha)}\}$, which is the set of unit vectors in \mathbb{R}^m .

Relation (4.14) provide a representation of prices as a series. Therefore, it is important to determine sufficient conditions on the model parameters such that prices are finite. To this end, let us assume that

Assumption 1. *For every stock $\alpha \in \{1, \dots, \gamma\}$ we assume that the following condition holds true:*

$$\bar{g}^{(\alpha;1)} := \max_{\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(\gamma)}} \left(\sum_{j=1}^m \sum_{\beta=1}^{\gamma} \sum_{h=1}^m e_h^{(\beta)} \lambda_{\beta, \alpha} \mathbf{P}_{h,j}^{(\beta, \alpha)} g_j \right) < r_{\alpha}$$

Theorem 4.1. *Let $\mathbf{g}(k)$ denotes the vector of states of the multivariate growth-dividend process of the pool of stocks at current time $k \in \mathbb{N}$. Let us assume that Assumption 1 holds true. Then:*

$$\begin{aligned} i) \quad p^{(\alpha)}(\mathbf{g}(k)) &= \sum_{i=1}^{+\infty} \frac{\mathbb{E}_{(k)}[\prod_{j=1}^i G^{(\alpha)}(k+j)]}{r_{\alpha}^i} d^{(\alpha)}(k) < +\infty \\ ii) \quad \lim_{i \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[\mathcal{P}^{(\alpha)}(k+i)]}{r_{\alpha}^i} &= 0. \end{aligned}$$

Proof. See appendix.

Remark 4.2. *The theorem gives a sufficient condition under which, first, prices are expressed by convergent series, and second, the transversality condition avoiding the presence of speculative bubbles is satisfied.*

Definition 4.3. *Define the price-dividend ratio for the α -th stock as follows:*

$$\psi_1^{(\alpha)}(\mathbf{g}(k)) := \frac{p^{(\alpha)}(\mathbf{g}(k))}{d^{(\alpha)}(k)} = \sum_{i=1}^{+\infty} \frac{\mathbb{E}_{(k)}[\prod_{j=1}^i G^{(\alpha)}(k+i)]}{r_\alpha^i}. \quad (4.15)$$

Theorem 4.4. *Let $\mathbf{G}(k) = \mathbf{g}_a(k) = (g_{a_1}^{(1)}, \dots, g_{a_\gamma}^{(\gamma)})$ denotes the vector of states of the multivariate growth-dividend process of the pool of stocks at current time $k \in \mathbb{N}$. Let us assume that Assumption 1 holds true. Then, for every stock $\alpha \in \{1, \dots, \gamma\}$ we have that:*

$$\begin{aligned} \psi_1^{(\alpha)}(g_{a_1}^{(1)}, \dots, g_{a_\gamma}^{(\gamma)}) = & \frac{1}{r_\alpha} \left\{ \sum_{j_\alpha=1}^m \sum_{\beta=1}^{\gamma} \sum_{h=1}^m e_h^{(\beta)}(k) \lambda_{\beta, \alpha} \mathbf{P}_{h, j_\alpha}^{(\beta, \alpha)} g_{j_\alpha}^{(\alpha)} + \right. \\ & \left. \sum_{j_1, \dots, j_\gamma=1}^m \psi_1^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) \cdot g_{j_\alpha}^{(\alpha)} \cdot \prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w, f} \mathbf{P}_{c, j_f}^{(w, f)} \right\} \end{aligned} \quad (4.16)$$

This linear system of m^γ equations in m^γ unknown admits a unique solution.

Proof. See appendix.

Remark 4.5. *Theorem 4.4 gives an effective way to get prices by solving a linear system of equations and by multiplying the price-dividend ratio with the initial value of the dividend process. It shows that a different price-dividend ratio is attached to each vector $(g_{a_1}^{(1)}, \dots, g_{a_\gamma}^{(\gamma)})$ of the multivariate growth-dividend process $\mathbf{G}(k)$.*

Remark 4.6. *If $\lambda_{w, f} = \delta_{w, f}$ then the stocks are independent of each other and we recover the linear system of equations established in Ghezzi and Piccardi (2003) for the uni-dimensional Markov model, i.e.*

$$\psi(g_i)r = \sum_{h=1}^m p_{i, h} g_h + \sum_{h=1}^m \psi(g_h) p_{g_i, g_h}$$

The evaluation of the second order moment of the price process for any stock $\alpha \in \{1, \dots, \gamma\}$ based on the MVMC model shares a similar strategy to that used for the computation of the first order moment. Anyway, the valuation of $p_2^{(\alpha)}(\mathbf{g}(k))$ requires control for the second order moment of the growth-dividend process as well as the first order moment as done with Assumption 1. For this reason we introduce the following additional assumption:

Assumption 2. For every stock $\alpha \in \{1, \dots, \gamma\}$ we assume that:

$$\bar{g}^{(\alpha;2)} := \max_{\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(\gamma)}} \left(\sum_{j=1}^m \sum_{\beta=1}^{\gamma} \sum_{h=1}^m e_h^{(\beta)} \lambda_{\beta, \alpha} \mathbf{P}_{h,j}^{(\beta, \alpha)} (g_j)^2 \right) < r_{\alpha}^2$$

Theorem 4.7. Let $\mathbf{g}(k)$ denotes the vector of states of the multivariate growth-dividend process of the pool of stocks at current time $k \in \mathbb{N}$. Let us assume that Assumptions 1 and 2 hold true. Then:

$$p_2^{(\alpha)}(\mathbf{g}(k)) = \sum_{i=1}^{+\infty} \frac{\mathbb{E}_{(k)}[(D^{(\alpha)}(k+i))^2]}{r_{\alpha}^{2i}} + 2 \sum_{i=1}^{+\infty} \sum_{j>i} \frac{\mathbb{E}_{(k)}[D^{(\alpha)}(k+i)D^{(\alpha)}(k+j)]}{r_{\alpha}^{i+j}} < +\infty,$$

$$\lim_{N \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[(\mathcal{P}^{(\alpha)}(k+N))^2]}{r_{\alpha}^{2N}} = 0, \quad (4.17a)$$

$$\lim_{N \rightarrow +\infty} \sum_{i=1}^N \frac{\mathbb{E}_{(k)}[D^{(\alpha)}(k+i)\mathcal{P}^{(\alpha)}(k+N)]}{r_{\alpha}^{i+N}} = 0. \quad (4.17b)$$

Proof. See appendix.

Remark 4.8. Theorem 4.7 presents conditions under which the transversality conditions are satisfied so that the presence of speculative bubbles is avoided and the representation of the risk as a convergent series that depends only on the the dividend process is permitted. This results extend the corresponding result established for univariate Markov chain model in Barbu et al. (2017).

Definition 4.9. Define the second-order price-dividend ratio for the α -th stock as follows:

$$\psi_2^{(\alpha)}(\mathbf{g}(k)) := \frac{p_2^{(\alpha)}(\mathbf{g}(k))}{(d^{(\alpha)}(k))^2}.$$

Theorem 4.10. Let $\mathbf{G}(k) = \mathbf{g}_a(k) = (g_{a_1}^{(1)}, \dots, g_{a_\gamma}^{(\gamma)})$ denotes the vector of states of the multivariate growth-dividend process of the pool of stocks at current time $k \in \mathbb{N}$. Let us assume that Assumptions 1 and 2 hold true. Then, for every stock $\alpha \in \{1, \dots, \gamma\}$ we have that:

$$\begin{aligned}
& r_\alpha^2 \psi_2^{(\alpha)}(g_{a_1}^{(1)}, \dots, g_{a_\gamma}^{(\gamma)}) - \sum_{j_1, \dots, j_\gamma=1}^m \psi_2^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) (g_{j_\alpha}^{(\alpha)})^2 \left(\prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w,f} \mathbf{P}_{c,j_f}^{(w,f)} \right) \\
&= 2 \sum_{j_1, \dots, j_\gamma=1}^m \psi_1^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) \cdot (g_{j_\alpha}^{(\alpha)})^2 \cdot \left(\prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w,f} \mathbf{P}_{c,j_f}^{(w,f)} \right) \\
&+ \sum_{j=1}^m \sum_{\beta=1}^{\gamma} \sum_{h=1}^m e_h^{(\beta)}(k) \lambda_{\beta,\alpha} \mathbf{P}_{h,j}^{(\beta,\alpha)} (g_j^{(\alpha)})^2
\end{aligned} \tag{4.18}$$

This linear system of m^γ equations in m^γ unknown admits a unique solution.

Proof. See appendix.

Since the stocks are possibly correlated it makes sense to compute the covariance function between two stocks, which in turns allows us to compute the variance of the portfolio. In this respect we can claim the following result.

Theorem 4.11. Let $\mathbf{g}(k)$ denotes the vector of states of the multivariate growth-dividend process of the pool of stocks at current time $k \in \mathbb{N}$. Let us assume that Assumptions 1 and 2 hold true. Then:

$$\begin{aligned}
p_2^{(\alpha,\beta)}(k) &= \sum_{i=1}^{+\infty} \mathbb{E}_{(k)} \left[\frac{D^{(\alpha)}(k+i) D^{(\beta)}(k+i)}{r_\alpha^i r_\beta^i} \right] + \sum_{i=1}^{+\infty} \sum_{j>i} \mathbb{E}_{(k)} \left[\frac{D^{(\alpha)}(k+i) D^{(\beta)}(k+j)}{r_\alpha^i r_\beta^j} \right] \\
&+ \sum_{i=1}^{+\infty} \sum_{j>i} \mathbb{E}_{(k)} \left[\frac{D^{(\alpha)}(k+j) D^{(\beta)}(k+i)}{r_\alpha^j r_\beta^i} \right] < +\infty
\end{aligned}$$

$$\lim_{N \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[\mathcal{P}^{(\alpha)}(k+N) \mathcal{P}^{(\beta)}(k+N)]}{r_\alpha^N r_\beta^N} = 0 \tag{4.19a}$$

$$\lim_{N \rightarrow +\infty} \sum_{i=1}^N \mathbb{E}_{(k)} \left[\frac{D^{(\alpha)}(k+i) \mathcal{P}^{(\beta)}(k+N)}{r_\alpha^i r_\beta^N} \right] = 0 \tag{4.19b}$$

$$\lim_{N \rightarrow +\infty} \sum_{i=1}^N \mathbb{E}_{(k)} \left[\frac{D^{(\beta)}(k+i) \mathcal{P}^{(\alpha)}(k+N)}{r_\alpha^i r_\beta^N} \right] = 0 \tag{4.19c}$$

Proof. See appendix.

Remark 4.12. *Theorem 4.11 presents conditions under which the transversality conditions are satisfied so that the presence of speculative bubbles is avoided and the representation of the price-product between two stocks as a convergent series that depends only on the the dividend processes is permitted.*

Definition 4.13. *Define the product price-dividend ratio for the α -th and β -th stocks as follows:*

$$\psi_2^{(\alpha;\beta)}(\mathbf{g}(k)) := \frac{p_2^{(\alpha;\beta)}(\mathbf{g}(k))}{d^{(\alpha)}(k) \cdot d^{(\beta)}(k)}. \quad (4.20)$$

Theorem 4.14. *Let $\mathbf{G}(k) = \mathbf{g}_a(k) = (g_{a_1}^{(1)}, \dots, g_{a_\gamma}^{(\gamma)})$ denotes the vector of states of the multivariate growth-dividend process of the pool of stocks at current time $k \in \mathbb{N}$. Let us assume that Assumptions 1 and 2 hold true. Then, for every couple of stocks $\alpha, \beta \in \{1, \dots, \gamma\}$ we have that:*

$$\begin{aligned} r_\alpha r_\beta \psi_2^{(\alpha;\beta)}(g_{a_1}^{(1)}, \dots, g_{a_\gamma}^{(\gamma)}) &= \sum_{j_\alpha, j_\beta=1}^m g_{j_\alpha}^{(\alpha)} g_{j_\beta}^{(\beta)} \left(\prod_{f \in \{\alpha, \beta\}} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w,f} \mathbf{P}_{c,j_f}^{(w,f)} \right) \\ &+ \sum_{j_1, \dots, j_\gamma=1}^m \psi_1^{(\beta)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) (g_{j_\alpha}^{(\alpha)}) (g_{j_\beta}^{(\beta)}) \left(\prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w,f} \mathbf{P}_{c,j_f}^{(w,f)} \right) \\ &+ \sum_{j_1, \dots, j_\gamma=1}^m \psi_1^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) (g_{j_\alpha}^{(\alpha)}) (g_{j_\beta}^{(\beta)}) \left(\prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w,f} \mathbf{P}_{c,j_f}^{(w,f)} \right) \\ &+ \sum_{j_1, \dots, j_\gamma=1}^m (g_{j_\alpha}^{(\alpha)}) (g_{j_\beta}^{(\beta)}) \psi_1^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) \psi_1^{(\beta)}(g_{j_1}, \dots, g_{j_\gamma}) \left(\prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w,f} \mathbf{P}_{c,j_f}^{(w,f)} \right) \end{aligned} \quad (4.21)$$

Proof. See appendix.

Once we obtain the product price-dividend ratio for any couple (α, β) of stocks, it is simple to compute the covariance function between the prices of two stocks:

$$\begin{aligned} &Cov(\mathcal{P}^{(\alpha)}(\mathbf{g}(k)), \mathcal{P}^{(\beta)}(\mathbf{g}(k))) \\ &= \mathbb{E}_{(k)}[\mathcal{P}^{(\alpha)}(\mathbf{g}(k)) \cdot \mathcal{P}^{(\beta)}(\mathbf{g}(k))] - \mathbb{E}_{(k)}[\mathcal{P}^{(\alpha)}(\mathbf{g}(k))] \cdot \mathbb{E}_{(k)}[\mathcal{P}^{(\beta)}(\mathbf{g}(k))] \quad (4.22) \\ &= d^{(\alpha)}(k) d^{(\beta)}(k) \left(\psi_2^{(\alpha;\beta)}(\mathbf{g}(k)) - \psi_1^{(\alpha)}(\mathbf{g}(k)) \psi_1^{(\beta)}(\mathbf{g}(k)) \right). \end{aligned}$$

The knowledge of the covariance function give us the possibility to use the model for portfolio selection purposes when prices are assumed to behave according to the Dividend Valuation Model.

4.2 Empirical application

In this section, we propose an application of our multivariate Markov chain model to three stocks. To perform our valuation, we have to select stocks with a stable dividend policy, i.e. regular dividend payments, and a sufficiently long history of dividends payments. For this purpose, our attention goes to the so called *Dividend Kings*. They are firms that mostly operate in consumer and industrial goods or utilities sectors and aim at securing stable dividend payments to investors.

We source dividends data and end-of-day stock prices for the years 1987 to 2018 from Thomson Reuters Tick History (TRTH). Data are processed within the Market Quality Dashboard¹ developed and managed by Capital Markets CRC. To perform a consistent valuation with yearly discount rates, i.e. company's cost of capital, we aggregate quarterly dividends to obtain yearly dividend series.

Among the group of Dividend Kings, we select three companies with the highest correlations between dividend series to test how results change according to correlations. However, the model can also be used when the correlation between stocks is low or zero. In our application, Genuine Parts Company (GPC), Dover Corporation (DOV), and Parker-Hannifin Corporation (PH), all traded in NYSE market, show the highest correlations, e.g. 0.92 between DOV and GPC, 0.94 between DOV and PH, and 0.97 between GPC and PH. These high correlations are also observable in Figure 4.1 that shows the yearly dividend history for the three stocks. The chart highlights a consistent upward trend after an initial period of stable dividends for all series. This trend reflects positive dividend growth rates for all three stocks, i.e. about 3% for GPC and DOV, and 5% for PH, as reported in Table 4.1. Panel A of the table shows general descriptive statistics for the dividend series, while Panel

¹MQD website: <http://www.mqdashboard.com>

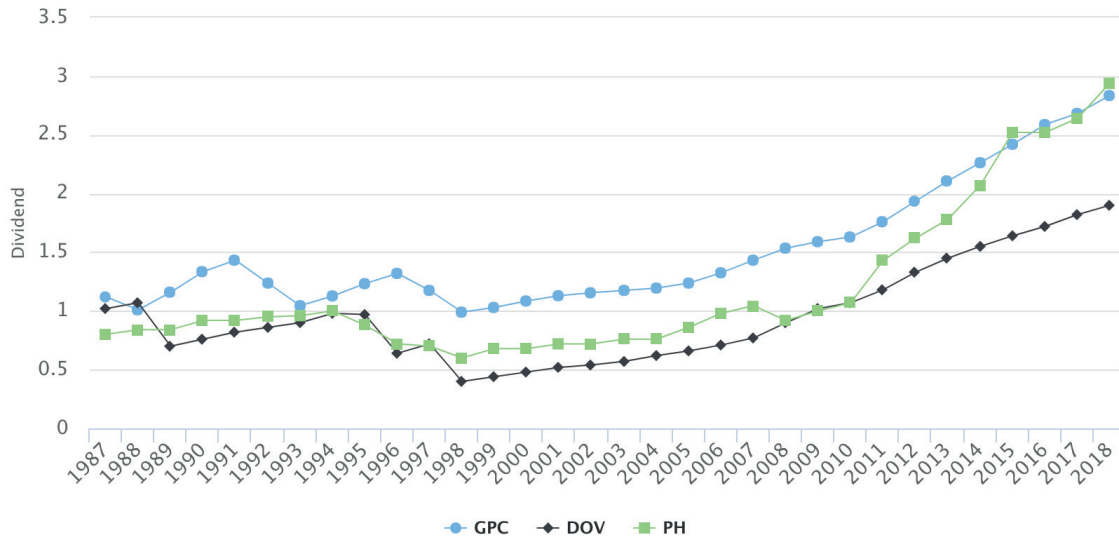


Figure 4.1: Yearly dividend series of Genuine Parts Company (GPC), Dover Corporation (DOV), and Parker-Hannifin Corporation (PH) from 1987 to 2018.

B reports the same analysis for the dividend growth rate series that are obtained according to equation (4.11).

The required rate of returns for solving the valuation equations is calculated using the Capital Asset Pricing Model (CAPM). The model originates from the idea of mean-variance efficient portfolio of Markowitz (1952), formalised by Sharpe (1964) and Lintner (1965). The rationale is that risky stocks are expected to be more remunerating than the risk free assets. Therefore, an investor expects a premium to be paid for purchasing a risky stock,

$$\pi_e = \mathbb{E}[R] - \mathbb{E}[R_f]. \tag{4.23}$$

The premium π_e is called the ex-ante equity risk premium and it is generally estimated using historical data. In our analysis, market and risk free returns are proxied by S&P 500 and US treasury bills returns, respectively. Using the full period from 1987 to 2018, we calculate the average of yearly returns for both series, as well as a unique value for stock's beta, based on the entire period consistently with the Markov chain estimation approach. Thus, our discounting factor r , that is one plus

<i>Panel A: Yearly dividends</i>								
	count	mean	std	min	25%	50%	75%	max
GPC	32	1.51	0.53	0.99	1.15	1.28	1.66	2.84
DOV	32	0.96	0.42	0.40	0.66	0.88	1.10	1.90
PH	32	1.18	0.65	0.60	0.76	0.92	1.16	2.94

<i>Panel B: Yearly dividend returns</i>								
	count	mean	std	min	25%	50%	75%	max
GPC	31	1.03	0.08	0.84	1.02	1.06	1.08	1.15
DOV	31	1.03	0.14	0.56	1.05	1.07	1.09	1.17
PH	31	1.05	0.11	0.82	1.00	1.05	1.11	1.34

Table 4.1: Descriptive statistics of yearly dividends and yearly dividend returns from 1987 to 2018.

	GPC	DOV	PH
β	0.749	0.970	1.060
r	1.069	1.075	1.077

Table 4.2: GPC, DOV, and PH betas and CAPM estimated rate of returns based on stock data from 1987 to 2018.

the required rate of return, can be estimated as

$$r_i = 1 + \bar{R}_f + \beta_{im}(\bar{R}_m - \bar{R}_f), \quad (4.24)$$

$$\beta_{im} = \frac{Cov[R_i, R_m]}{Var[R_m]}, \quad (4.25)$$

where \bar{R}_m is the yearly average return of the market portfolio, and \bar{R}_f is the yearly average return of the risk free asset.

Table 4.2 shows the estimated beta and discount factors for the three stocks.

The multivariate Markov stock model produces a price-dividend ratio for each combination of series and states. Therefore, to enhance readability in this application, we decided to model the dividend growth series with 3-state Markov chains. Hence, results will contain 27 different price-dividend ratios for each series.

The first step in our analysis is to discretise the dividend returns into three categories, i.e. the states of the Markov chain. We adopt the same approach of Barbu et al. (2017), and in this 3-state scenario, we assume that the central state

State	Dividend growth range			State value		
	GPC	DOV	PH	GPC	DOV	PH
1	0.843-0.959	0.556-0.918	0.818-0.949	0.877	0.657	0.880
2	0.960-1.041	0.919-1.088	0.950-1.051	1.036	1.058	1.000
3	1.042-1.151	1.089-1.169	1.052-1.336	1.074	1.103	1.114

Table 4.3: Dividend growth discretisation with 3 states and corresponding state values.

is the zero dividend growth and the external states are the negative and positive returns. Also, we allow for the zero dividend growth to include returns that are close to the null return, both negative and positive values. Thus, the central state includes all the observations that are within a length of $\sigma/2$ from the zero return. The edges of the discretisation procedures for each stock are reported in Table 4.3. Moreover, the table shows the assumed dividend growth value for each state that corresponds to the median of the observations falling inside each interval.

From the categorised series, we calculate the transition matrices of the multivariate model. According to equation (4.13), we have a matrix for each combination of two dividend series, $\mathbf{P}_{i,j}^{(\beta,\alpha)}$, including the combination of a series with itself. Therefore, with three dividend growth series modelled through a 3-state Markov chain, there are nine transition matrices with nine elements each. Considering that each matrix is a stochastic matrix, there are $\gamma^2 m(m-1) = 54$, parameters to estimate. The estimation of the transition probabilities can be obtained using the maximum likelihood estimator from Billingsley (1961a)

$$\hat{\mathbf{P}}_{ij}^{(\beta,\alpha)} = \frac{n_{ij}^{(\beta,\alpha)}}{\sum_{j=1}^m n_{ij}^{(\beta,\alpha)}}, \quad (4.26)$$

where $n_{ij}^{(\beta,\alpha)}$ is the occurrences of transitions from state i in series β to state j in series α .

Table 4.4 reports nine 3×3 transition matrices. Each matrix contains nine elements, each one representing the probability of moving from state i of series β to state j of series α . Starting states i are read by rows and arriving state j are read by columns. For the stochastic characteristic of the matrix, each row sums to one.

$\begin{matrix} GPC \rightarrow GPC \\ \left(\begin{array}{ccc} 0.4 & 0.2 & 0.4 \\ 0. & 0.556 & 0.444 \\ 0.125 & 0.188 & 0.688 \end{array} \right) \end{matrix}$	$\begin{matrix} GPC \rightarrow DOV \\ \left(\begin{array}{ccc} 0.4 & 0.2 & 0.4 \\ 0. & 0.778 & 0.222 \\ 0.062 & 0.625 & 0.312 \end{array} \right) \end{matrix}$	$\begin{matrix} GPC \rightarrow PH \\ \left(\begin{array}{ccc} 0.2 & 0.6 & 0.2 \\ 0. & 0.333 & 0.667 \\ 0.188 & 0.312 & 0.5 \end{array} \right) \end{matrix}$
$\begin{matrix} DOV \rightarrow GPC \\ \left(\begin{array}{ccc} 0.333 & 0.333 & 0.333 \\ 0.111 & 0.278 & 0.611 \\ 0.111 & 0.333 & 0.556 \end{array} \right) \end{matrix}$	$\begin{matrix} DOV \rightarrow DOV \\ \left(\begin{array}{ccc} 0. & 0.333 & 0.667 \\ 0.111 & 0.722 & 0.167 \\ 0.111 & 0.444 & 0.444 \end{array} \right) \end{matrix}$	$\begin{matrix} DOV \rightarrow PH \\ \left(\begin{array}{ccc} 0. & 0.333 & 0.667 \\ 0.111 & 0.5 & 0.389 \\ 0.222 & 0.111 & 0.667 \end{array} \right) \end{matrix}$
$\begin{matrix} PH \rightarrow GPC \\ \left(\begin{array}{ccc} 0.25 & 0.5 & 0.25 \\ 0.25 & 0.333 & 0.417 \\ 0. & 0.214 & 0.786 \end{array} \right) \end{matrix}$	$\begin{matrix} PH \rightarrow DOV \\ \left(\begin{array}{ccc} 0.25 & 0. & 0.75 \\ 0.167 & 0.75 & 0.083 \\ 0. & 0.643 & 0.357 \end{array} \right) \end{matrix}$	$\begin{matrix} PH \rightarrow PH \\ \left(\begin{array}{ccc} 0.25 & 0.25 & 0.5 \\ 0.167 & 0.417 & 0.417 \\ 0.071 & 0.357 & 0.571 \end{array} \right) \end{matrix}$

Table 4.4: Transition matrices of the multivariate Markov stock model. Each matrix contains nine transition probability indicating the probability of moving from state i for series β to state j for series α . Starting states i are on rows and arriving state j are on columns. Each row of each matrix sums to one.

As an example, the matrix $PH \rightarrow GPC$, at bottom-left corner of Table 4.4, shows the transitions probabilities when $\beta = PH$ and $\alpha = GPC$. It shows the dependence of GPC dividends from PH dividends. Specifically, the first top-left element of the matrix, i.e. 0.25, is the probability of arriving in state 1 of series GPC leaving state 1 of series PH.

The next step is the estimation of parameters $\lambda_{\beta,\alpha}$ that represent the weights of the dependencies between the series. We perform a maximum likelihood estimation of the function

$$\log L = \sum_{i_1, i_2, \dots, i_\gamma, j} n_{i_1, i_2, \dots, i_\gamma, j} \log P_\alpha^{MTD}, \quad (4.27)$$

where $P_\alpha^{MTD} = \lambda_{1\alpha} \mathbf{P}_{i_1, j}^{(1, \alpha)} + \dots + \lambda_{\gamma\alpha} \mathbf{P}_{i_\gamma, j}^{(\gamma, \alpha)}$ is the linear combination of the transition probabilities from states i_γ of series γ to state j of series α . Values $n_{i_1, i_2, \dots, i_\gamma, j}$ represents the number of observed transition combinations of the type $(i_1, i_2, \dots, i_\gamma)$ at time $t - 1$ to state j of series α at time t . The maximisation is constrained by the conditions

$$\sum_{\beta=1}^{\gamma} \lambda_{\beta,\alpha} = 1, \quad \text{and} \quad \lambda_{\beta,\alpha} \geq 0.$$

Computationally, the maximum likelihood estimation is obtained by minimising

	GPC	DOV	PH
GPC	0.45	0	0.55
DOV	0.15	0	0.85
PH	0.49	0.51	0

Table 4.5: Estimation of $\lambda_{\beta,\alpha}$. Values indicate the influence of stock β (read by columns) to stock α (read by rows). Each row sums to one.

the negative likelihood function, using a minimisation algorithm under constraints, namely *Sequential Least Squares Programming* (SLSQP) optimisation subroutine originally implemented by Kraft (1988). Table 4.5 reports the estimation of $\lambda_{\beta,\alpha}$ for the dividend growth series of GPC, DOC, and PH.

After calculating the transition matrices and the values of $\lambda_{\beta,\alpha}$ and the required discount factor, we can solve the system of linear equations in formula (4.16). In our example, after verifying Assumption 1, we obtain a price-dividend ratio for each combination of starting states and for each destination series, for a total of 81 values, as reported in Table 4.6.

Finally, after verifying Assumption 2, we can solve the system of equations in (4.18) to obtain the second order moment of the price dividend ratio for all three series. Then, we solve equation (4.21) to calculate the second order moment of the price-dividend ratio for the mixed combinations of the stock. Results are in Table 4.7.

After observing the state occupied from last dividend growth rates, i.e. state 3 for GPC, state 2 for DOV, and state 3 for PH, we calculate the variance-covariance matrix for the three stocks according to equation (4.22). Results are reported in Table 4.8 and can be used for portfolio selection purposes. Figure 4.2 shows a simulation of the variance of portfolio that includes all three stock with different weights. The portfolio variance is reported on z axis and weights of GPC and DOV on x and y axis respectively. The weight of PH is not reported as the sum of the three weights is one. From the figures, at different angles, it is evident that there exists a specific combination of weights that minimises the portfolio variance. While, including only one of the stocks in the portfolio will result in a higher variance.

GPC	<i>States</i>		GPC	ψ_1	
	DOV	PH		DOV	PH
1	1	1	32.8693	24.3190	29.1173
		2	32.8649	24.2142	28.8670
		3	33.9509	25.9960	28.8668
	2	1	32.7255	24.0914	28.4871
		2	32.7211	23.9876	28.2423
		3	33.8027	25.7532	28.2421
	3	1	32.8699	24.3300	28.7594
		2	32.8656	24.2251	28.5123
		3	33.9516	26.0077	28.5120
2	1	1	34.2770	25.4251	30.3092
		2	34.2697	25.3181	30.0499
		3	35.3628	27.1403	30.0496
	2	1	34.1290	25.1915	29.6758
		2	34.1216	25.0854	29.4219
		3	35.2103	26.8914	29.4217
	3	1	34.2778	25.4364	29.9495
		2	34.2705	25.3293	29.6933
		3	35.3636	27.1523	29.6930
3	1	1	33.9322	25.1327	29.6280
		2	33.9255	25.0264	29.3735
		3	35.0161	26.8333	29.3733
	2	1	33.7849	24.9003	28.9973
		2	33.7781	24.7949	28.7483
		3	34.8644	26.5857	28.7481
	3	1	33.9329	25.1439	29.2698
		2	33.9262	25.0375	29.0185
		3	35.0169	26.8453	29.0183

Table 4.6: Price-dividend ratios ψ_1 . Each combination of starting states for the three series and each destination series produces a different price-dividend ratio.

<i>States</i>			ψ_2					
GPC	DOV	PH	GPC	DOV	PH	GPC:DOV	GPC:PH	DOV:PH
1	1	1	1179.55	749.42	931.79	801.05	959.14	709.00
		2	1180.36	734.69	915.26	797.30	950.69	699.76
		3	1255.18	830.86	915.82	883.91	981.61	751.88
	2	1	1169.89	737.68	892.15	790.03	934.20	687.02
		2	1170.69	723.17	876.33	786.33	925.96	678.06
		3	1244.94	817.88	876.86	871.77	956.09	728.58
	3	1	1179.50	750.92	911.14	801.42	947.55	700.84
		2	1180.31	736.16	894.98	797.67	939.19	691.70
		3	1255.13	832.52	895.53	884.32	969.74	743.22
2	1	1	1274.34	806.45	1006.88	872.59	1040.28	771.26
		2	1274.79	791.19	989.17	868.59	1031.11	761.43
		3	1350.53	891.00	989.77	960.17	1063.26	816.96
	2	1	1264.13	794.19	966.84	860.89	1014.35	748.52
		2	1264.57	779.16	949.84	856.94	1005.41	738.98
		3	1339.74	877.50	950.41	947.32	1036.77	792.88
	3	1	1274.28	808.01	986.02	872.98	1028.21	762.80
		2	1274.74	792.72	968.68	868.99	1019.15	753.08
		3	1350.48	892.72	969.27	960.61	1050.93	808.00
3	1	1	1252.34	791.96	964.88	854.14	1007.07	745.62
		2	1252.86	776.87	947.80	850.22	998.18	736.06
		3	1328.28	875.37	948.37	940.39	1029.70	789.91
	2	1	1242.22	779.80	925.18	842.58	981.43	723.12
		2	1242.72	764.94	908.80	838.71	972.77	713.84
		3	1317.57	861.97	909.35	927.69	1003.49	766.08
	3	1	1252.29	793.51	944.20	854.53	995.15	737.26
		2	1252.81	778.39	927.48	850.61	986.36	727.81
		3	1328.22	877.08	928.05	940.82	1017.50	781.05

Table 4.7: Second order price-dividend ratio ψ_2 . Values represents a measure of the risk. Each combination of starting states for the three series and each destination series produces a different value for the second order price-dividend ratio.

	GPC	DOV	PH
GPC	820.16		
DOV	4.28	560.18	
PH	10.06	10.02	716.54

Table 4.8: Variance-covariance matrix for GPC, DOV, and PH.

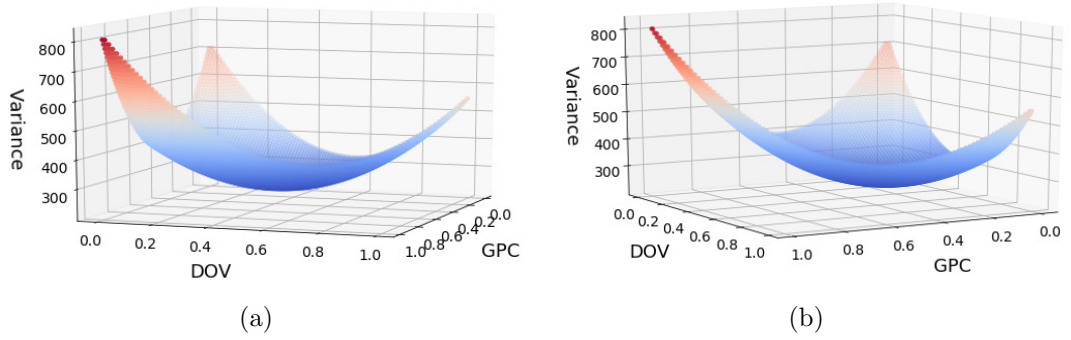


Figure 4.2: Simulation of the variance of a hypothetical portfolio with different weights of the three stocks. GPC and DOV weights are reported in x and y axis respectively. PH weight is not shown as the three weights sum to one. The portfolio variance is reported on z axis. Both subfigures shows the portfolio variance from different angles. From the figures, it is clear that there is a specific combination of weights that minimises the variance of the portfolio.

<i>State</i>	ψ_1		
	GPC	DOV	PH
1	33.08	23.78	30.19
2	36.41	22.24	30.11
3	35.47	22.62	31.10

Table 4.9: Price-dividend ratio calculated with the univariate Markov stock model in Barbu et al. (2017).

Finally, we perform a comparison of our multivariate Markov stock model with the univariate model proposed in Barbu et al. (2017). The univariate model yields price-dividend ratios shown in Table 4.9. Performing the calculation for single dividend growth series, we obtain only three price-dividend ratios for each series, corresponding to each state of the Markov chain. Then, we calculate the estimated price for the three companies, using both the multivariate and univariate models, and we compare these values with the actual stock price. Table 4.10 reports the comparison results.

It is interesting to note how dependencies between series tend to reduce the variability in the stock valuation compared to the univariate analysis. For example, GPC value reduces from the univariate model to the multivariate as well as PH, while DOV increases its value. Comparing the stock price at 31 October 2018, GPC stock price appears to be almost in line with the fundamental valuation performed with

	GPC	DOV	PH
Multivariate	98.84	50.51	84.52
Univariate	100.55	42.26	91.44
Stock price	97.92	82.84	151.63

Table 4.10: Estimated firms prices using the proposed multivariate Markov chain model compared with univariate model estimations, and with the actual stock prices at 31 October 2018.

the multivariate model, while DOV and PH are considerably overpriced, even though DOV reduces its overpricing in the multivariate model. In general, the multivariate model offers a wider range of valuations, depending on the combination of states occupied by the dividend growth rates of all series, thus it allows for more dynamics, including dependencies between series, and results in a better valuation.

4.3 Conclusion

This chapter presented a multivariate extension of the Markov chain stock model introduced in Chapter 3. We focused on how dividend forecasts are updated taking into account the possible dependencies between the state of the dividend processes for each stock. In this model, each update of a state depends on a vector of states of all the growth series. The model introduces a linear system of equations for the first and second order price-dividend ratios that are attached to the vector of states. Further, we proposed a formula for the computation of the variances and covariances between stocks for portfolio selection and valuation purposes. The chapter closes with an application to dividend growth series from three US stocks with a long history of dividend payments and correlation between the series. The application shows how to practically implement the model and how our proposed multivariate model performs better than other dividend valuation models.

5 A new measure of price discovery in financial markets

In this chapter, we propose a new measure to establish price leadership among multiple related price series using a Multivariate Markov Chain. This new measure, the Price Leadership Share (PLS), can easily be calculated with more than two price series simultaneously, offering an advantage over the existing price discovery measures. We test our model to six gold contracts, including spot, futures, and ETF, over a 2-year period, showing that gold futures contracts, mainly CME contract, have a major role in price discovery confirming previous literature's findings.

This chapter is organised as follows: Section 5.1 briefly reviews the multivariate Markov chain model. Section 5.2 defines the Price Leadership Share measure. Finally, the empirical application on gold contracts is presented in section 5.3 along with procedures on how to estimate the parameters of the model.

5.1 The Multivariate Markov Chain model

This section presents a brief review of the multivariate Markov chain model. The first part describes a simple model with a single time series, followed by an analysis of the multivariate setting, modelled via a Mixture Transition Distribution (MTD) model that was first introduced by Raftery (1985) for high-order Markov chains to reduce the number of parameters. For a comprehensive review on Markov chain models see, e.g. Brémaud (1999).

subject to $0 \leq p_{ij} \leq 1$, $\forall i, j \in \mathcal{M}$ and $\sum_{j=1}^m p_{ij} = 1$, $\forall j \in \mathcal{M}$. Each element of the matrix is a probability and every change from a state i must end in a state j , therefore the matrix is a *stochastic matrix*.

A Markov chain is fully defined when we know the initial probability distribution and the transition probability matrix.

Let

$$\mathbf{A}(t) := [A_1, \dots, A_m], \quad (5.4)$$

be the probability vector where $A_i(t) := \Pr(S_t = i)$ is the probability of being in state i at time t , with $i \in \mathcal{M}$, then

$$\mathbf{A}(t+1) = \mathbf{A}(t)\mathbf{P}, \quad (5.5)$$

$$\mathbf{A}(t+1) = \mathbf{A}(0)\mathbf{P}^t. \quad (5.6)$$

5.1.2 The Multivariate Markov process

The previous model can be extended to a multivariate setting, with more than one time series. For every series α in $\Gamma = \{1, 2, \dots, \gamma\}$, the probability of being in state j depends on the state i_1, \dots, i_γ occupied by all the available series one time step before. The Markov Property in (5.1) becomes:

$$\begin{aligned} \Pr[S_{t+1}^{(\alpha)} = j | (S_t^{(1)} = i_t^{(1)}, S_{t-1}^{(1)} = i_{t-1}^{(1)}, \dots, S_0^{(1)} = i_0^{(1)}), \dots, \\ (S_t^{(\gamma)} = i_t^{(\gamma)}, S_{t-1}^{(\gamma)} = i_{t-1}^{(\gamma)}, \dots, S_0^{(\gamma)} = i_0^{(\gamma)})] = \\ \Pr(S_{t+1}^{(\alpha)} = j | S_t^{(1)} = i_t^{(1)}, \dots, S_t^{(\gamma)} = i_t^{(\gamma)}), \end{aligned} \quad (5.7)$$

where $\alpha \in \Gamma$. The new Property (5.7) shows that there are multiple dependencies between the series. Therefore, the transition probability matrix of the multivariate model must include each possible combination, m^γ , for the initial states, and every

initial state must end in one of the possible final combinations. The result is $m^\gamma(m^\gamma - 1)$ total parameters to estimate for the multivariate Markov model, given that there are $m^\gamma - 1$ independent probabilities in each row. Such a configuration is not practical in a real-world application because the number of parameters will increase exponentially when the number of series and states increase.

Raftery (1985) proposed the Mixture Distribution Model (MTD) to reduce the number of parameters to estimate for high order Markov chains, and Ching et al. (2002) applied it to the multivariate Markov chains. A review of the MTD model and its application is available in Berchtold and Raftery (2002). Applying the MTD model the probability vector for series α at time $t + 1$ becomes

$$\mathbf{A}^{(\alpha)}(t + 1) = \sum_{\beta=1}^{\gamma} \mathbf{A}^{(\beta)}(t) \cdot \lambda_{\beta,\alpha} \cdot \mathbf{P}^{(\beta,\alpha)}, \quad (5.8)$$

where $\mathbf{A}^\alpha(t) := [A_1^{(\alpha)}, \dots, A_m^{(\alpha)}]$ and $A_i^{(\alpha)}(t) := Pr(S_t^{(\alpha)} = i)$.

According to this condition, we can build γ^2 transitions probability matrices $\mathbf{P}^{(\beta,\alpha)}$, each one containing the transition probabilities from state i in series β to state j in series α , with $\alpha, \beta \in \Gamma$,

$$\mathbf{P}^{(\beta,\alpha)} = \begin{array}{c} S_{t-1}^{(\beta)} \\ \begin{array}{c} 1 \\ 2 \\ \vdots \\ m \end{array} \end{array} \begin{array}{c} S_t^{(\alpha)} \\ \begin{array}{c} 1 \\ 2 \\ \dots \\ m \end{array} \end{array} \begin{bmatrix} p_{11}^{(\beta,\alpha)} & p_{12}^{(\beta,\alpha)} & \dots & p_{1m}^{(\beta,\alpha)} \\ p_{21}^{(\beta,\alpha)} & p_{22}^{(\beta,\alpha)} & \dots & p_{2m}^{(\beta,\alpha)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1}^{(\beta,\alpha)} & p_{m2}^{(\beta,\alpha)} & \dots & p_{mm}^{(\beta,\alpha)} \end{bmatrix}. \quad (5.9)$$

Parameters $\lambda_{\beta,\alpha}$ are the scalar weights that combine all the series, and are subject to:

$$\sum_{\beta=1}^{\gamma} \lambda_{\beta,\alpha} = 1, \quad (5.10)$$

$$\lambda_{\beta,\alpha} \geq 0. \tag{5.11}$$

The MTD model permits to reduce the total parameters to estimate from $m^\gamma(m^\gamma - 1)$ to $\gamma^2 m(m - 1) + \gamma(\gamma - 1)$, the first addend being the number of $p_{mm}^{(\beta,\alpha)}$ parameters and the second the number of weights $\lambda_{\beta,\alpha}$.

5.2 Price Leadership Share

To overcome the limits of existing price discovery measures, we rely entirely on our direct observations of stock prices, specifically on dependencies between price changes. We notice that some prices update faster than others when there is new information, and these changes are directly observable in the market. Thus, a price change can be interpreted as a signal to other markets that, in turn, update their prices based on that signal. In general, we argue that price changes dependencies exist between price series in related markets.

If we assume that price changes from dependent price series are expressed as sequences of random variables that satisfy the multivariate Markov property in (5.7), then the realisation of a price change series α at time $t + 1$ depends on price changes at time t from all the related series, $1, \dots, \gamma$. More specifically, equation (5.8) tells us that the probability for a price change in series α of being in a specific state (e.g., negative, positive or null) is a linear combination of all γ series of weighted transition probabilities from each series initial states to the arrival state in series α . In other words, the $\lambda_{\beta,\alpha}$ weights indicate how much a series influences other series in changing the price.

In general, there are γ^2 values of $\lambda_{\beta,\alpha}$ subject to condition (5.10) that can be organised in a matrix form,

$$\begin{array}{rcc}
& & \text{Series } \beta \\
\text{Series } \alpha & 1 & 2 & \dots & \gamma \\
1 & \left[\begin{array}{cccc} \lambda_{1,1} & \lambda_{2,1} & \dots & \lambda_{\gamma,1} \\ \lambda_{1,2} & \lambda_{2,2} & \dots & \lambda_{\gamma,2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{1,\gamma} & \lambda_{2,\gamma} & \dots & \lambda_{\gamma,\gamma} \end{array} \right. & & \\
2 & & & & \\
\vdots & & & & \\
\gamma & & & &
\end{array} \quad (5.12)$$

Each element of the matrix (5.12) measures the price change influence that series β has on series α . For example, element $\lambda_{1,1}$ is the portion of influence of series 1 on series 1, element $\lambda_{2,1}$ is the portion of influence of series 2 on series 1, and so on. Each row of the matrix contains the influence shares from all series to a specific series α , including the self-influence, and the sum of all row's elements is equal to one.

To understand which price series is leading the others and to what extent, we have to summarise values in the matrix (5.12). First, we exclude the self-influence and focus only on the external influence, setting the diagonal elements of the matrix to zero and, if $\gamma > 2$, normalising the row elements to sum to one,

$$\begin{array}{rcc}
& & \text{Series } \beta \\
\text{Series } \alpha & 1 & 2 & \dots & \gamma \\
1 & \left[\begin{array}{cccc} 0 & \lambda_{2,1}^* & \dots & \lambda_{\gamma,1}^* \\ \lambda_{1,2}^* & 0 & \dots & \lambda_{\gamma,2}^* \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{1,\gamma}^* & \lambda_{2,\gamma}^* & \dots & 0 \end{array} \right. & & \\
2 & & & & \\
\vdots & & & & \\
\gamma & & & &
\end{array} \quad (5.13)$$

where

$$\lambda_{\beta,\alpha}^* = \begin{cases} \frac{\lambda_{\beta,\alpha}}{\sum_{\beta=1, \beta \neq \alpha}^{\gamma} \lambda_{\beta,\alpha}} & \text{if } \gamma > 2 \\ \lambda_{\beta,\alpha} & \text{if } \gamma = 2 \end{cases} \quad (5.14)$$

Then, summing the elements of the matrix (5.13) by columns and normalising the results to one, we obtain a probability distribution of external price change influences, that we define Price Leadership Share (PLS),

$$PLS = \left[\frac{\sum_{\alpha=1}^{\gamma} \lambda_{1,\alpha}^*}{\sum_{\alpha=1}^{\gamma} \sum_{\beta=1}^{\gamma} \lambda_{\beta,\alpha}^*}, \dots, \frac{\sum_{\alpha=1}^{\gamma} \lambda_{\gamma,\alpha}^*}{\sum_{\alpha=1}^{\gamma} \sum_{\beta=1}^{\gamma} \lambda_{\beta,\alpha}^*} \right], \quad (5.15)$$

where $\sum_{\alpha=1}^{\gamma} \sum_{\beta=1}^{\gamma} \lambda_{\beta,\alpha}^* = \gamma$ if all series have at least one external influence.

From the PLS distribution, we can identify the price series that is the leader in impounding new information by simply calculating the mode of the distribution. Moreover, we can measure the quantity of information that is carried by the probability distribution, and understand how the leadership is concentrated. We adopt the entropy measure proposed by Theil (1967) and derived from the mathematical theory of communication by Shannon (1948). The entropy of the price leadership share can be expressed as

$$T = \sum_{i=1}^{\gamma} PLS_i \log(\gamma PLS_i), \quad (5.16)$$

where PLS_i are the elements of the price leadership share distribution (5.15), and $0 \leq T \leq \log(\gamma)$. If one element of the PLS vector has probability equal to one and all others zero, then the full information is conveyed by that element, i.e. it is the sole leader, and the entropy value is equal to its highest value, $\log(\gamma)$. On the contrary, when all elements of PLS have the same probability $1/\gamma$, it means that the information is equally distributed and there is no price leader. In this last case, the entropy value is equal to zero.

We can generalise the entropy indicator to compare PLS distributions with different lengths. For example, if we apply our methodology to several contexts with different numbers of price series and want to understand whether the price leadership is stronger in one setting or another, we can compute a price leadership

concentration index, that is the ratio of the entropy and its maximum value

$$L = \frac{T}{\log(\gamma)}. \quad (5.17)$$

Higher values of the index L mean higher price leadership concentration.

5.3 Empirical application

5.3.1 Data

We test our Price Leadership Share on six gold contracts across the world. Although the major venue for trading physical gold is the London OTC market, we focus on international exchange-traded contracts for the availability of high-frequency price quotes. The data include one spot contract, from the Shanghai Gold Exchange (SGE) matching market, four futures contracts from the Chicago Mercantile Exchange (CME), Shanghai Futures Exchange (SHFE), Tokyo Commodity Exchange (TOCOM), and Dubai Gold & Commodities Exchange (DGCX), and one Exchange Traded Fund, the SPDR gold shares (GLD) traded in NYSE Arca. The choice of comparing the futures against the spot market is a common practice in price discovery studies, with a major role attributed to futures contracts (Bohl et al., 2011, Rosenberg and Traub, 2009). Specifically, Hauptfleisch et al. (2016) find that the gold futures contract has a more important role in incorporating new information about the value of gold than the London OTC market, despite the huge difference in market share, 8% against 78% respectively. Moreover, we include the ETF because some studies suggest that ETFs might have a role in the price discovery process (see, e.g., Marshall et al., 2013).

Because of the discrete-time setting and consistent with other price discovery measures, we use intraday prices series at different sampling intervals² and perform the analysis on a daily basis. For example, a price series of a 9-hour trading day,

²A common procedure includes the use of a 1-second interval midquote returns

with sampling at a 1-second interval, would result in 32,400 observations, allowing for extensive possibilities in setting the model parameters, number of series and number of states, without incurring any estimation limitations. The estimation of the measure on a daily basis permits us to understand the dynamics of the price influence over time and avoid the effect of potential seasonal patterns in the price series. Moreover, like other measures, we use midquote prices to capture the price adjustment dynamics (Goettler et al., 2009).

In this application, we use intraday midquote log returns sampled at a 1-second interval from January 2016 to December 2017. The data are sourced from Thomson Reuters Tick History (TRTH) and processed by the Market Quality Dashboard³ developed and managed by Capital Markets CRC. The Price Leadership Share measure is obtained on a daily basis from intraday returns, for the complete overlapping trading hours of all contracts, that is between 1.30 pm and 6.30 pm UTC. Days of holidays in any of the markets are excluded from the computation. The final sample consists of 424 trading days.

For a better comprehension of markets analysed in this application, we compute some market quality and efficiency metrics. Table 5.1 gives a summary of basic metrics for all contracts. Total volume is the aggregated volume over the entire period, expressed in millions of ounces. CME gold futures shows the highest activity followed by SHFE. All other contracts have a very low trading activity. It is important to keep in mind that the spot contract includes only quotes in the matching market and not those from the OTC market. Other metrics are averages of the daily weighted average across all trades and quotes. The effective spread is calculated as the difference between midquote and actual transaction price, and it measures the liquidity of the market. The lower the spread, the more liquid is the market. On average, the ETF appears to be the most liquid contract, followed by CME. On the contrary, DGCX and TOCOM futures contracts are the least liquid.

In addition, we use the variance ratio as a proxy for market efficiency (Lo and

³MQD website: <http://www.mqdashboard.com>

	Total Volume (Mil oz)	Effective Spread (bps)	Variance Ratio	Order Imbalance	Tick Size
CME	21.18	1.29	0.12	0.12	0.1 USD/oz
DGCX	0.04	2.71	0.15	-1.22	0.1 USD/oz
GLD	0.21	0.74	0.18	-0.34	0.1 USD/oz
TOCOM	0.87	2.55	0.23	-0.40	1 JPY/g
SGE	0.60	1.82	0.29	-2.59	0.01 CNY/g
SHFE	3.57	2.38	0.15	-0.03	0.05 CNY/g

Table 5.1: Market quality and efficiency metrics for six gold contracts for the period January 2016 - December 2017. Figure are averages of daily metrics. Total volume is aggregated over the entire period. All other metrics are the averages of the daily weighted averages across all trades and quotes for the day.

MacKinlay, 1988, O’Hara and Ye, 2011). If a stock follows a random walk, returns variance is a linear function of the measurement frequency. The variance ratio measures deviations from efficiency, meaning that lower numbers correspond to more efficient markets. Overall, all contracts show low values of variance ratio. However, CME contract has a slightly higher efficiency. Finally, the order imbalance, calculated as the volume difference between buyer and seller initiated trades, is a measure of information asymmetry and it has effects on returns and liquidity. An order imbalance close to zero is an optimal value. Again, in this case, SHFE shows a value closer to zero, followed by CME.

From the analysis of Table 5.1, CME contract appears to be the ideal candidate as the price leader, as the market quality and efficiency metrics are the best among all contracts. It is followed by the gold ETF, while the Asian contracts show some level of inefficiency and we do not expect them to have a significant share of price leadership. However, we have to take into account the minimum tick size allowed in each market for a fair comparison of price changes. For example, the US and Dubai contracts share the same tick size, while the Asian contracts have different sizes, that are bigger, in absolute terms, compared to the others. This difference will be reflected in the price change dynamics. In fact, a slight change in the US market cannot always be followed by a similarly small change in markets with higher tick size because of the constrained minimum step for the price change. Nevertheless, we can address this issue in the discretisation of price changes in case it turns to be

a limit to the application of the model.

5.3.2 Parameters estimation

In the discrete-time multivariate Markov chain model, the series take values in the set $\mathcal{M} = \{1, 2, 3, \dots, m\}$, therefore we have to accurately categorise every price change into a possible state of the Markov Chain. This discretisation process can be performed by identifying some critical thresholds, according to the number of states, in a way that it is possible to describe the dynamics of the returns entirely. For example, in a 3-state scenario with log-returns, we can assume that the central state is the zero return and the external states are the negative and positive returns. Moreover, following Barbu et al. (2017), to make the discretisation more realistic, we can also assume that the price variations around the zero can be all considered as null returns. In our setting, we include in the zero return all the observations that are within a length of $\sigma/2$ from the null return and all the remaining returns are associated to the external states. If the data are normally distributed, there will be a 38% probability that a random observation will fall in the zero return area. Successive thresholds can be set at σ distance from the others, in both directions, up to the limit of the distribution to increase the number of states. We use an even number of states always including the null return to better represent the stationary series of price returns.

This discretisation technique is well suited for approximately normally distributed time series of price returns and performs well when tested on market data. However, it can be easily adapted with the modification of the bin width according to the distribution of the observed data. For example, when comparing multiple instruments with very different tick sizes, we can adapt the central bin size to include small changes that happen in one market but cannot be reflected in the market with wider tick size. A simulation of this simple discretisation method applied to the CME gold futures returns is provided in Figure 5.1. A 3-state categorisation with a central bin width equal to σ has been applied to the time series of price log returns.

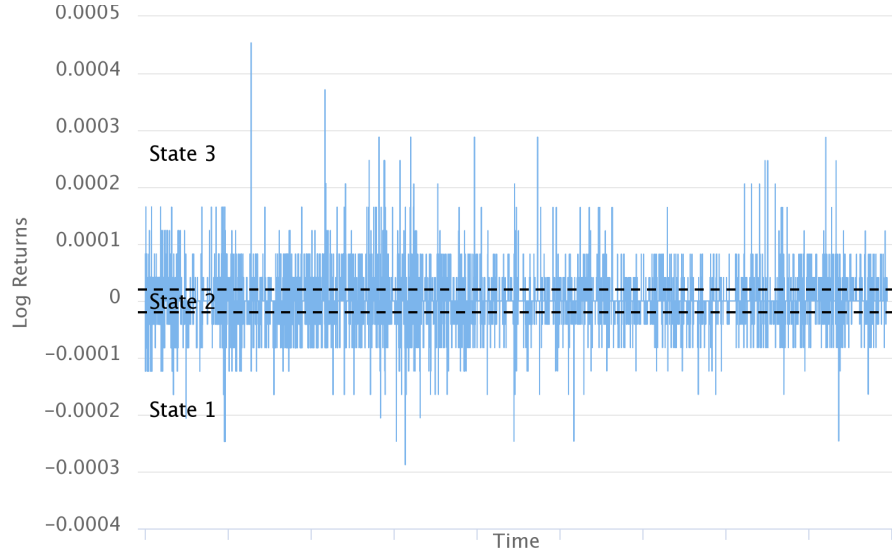


Figure 5.1: This figure shows how to model a price series with a Markov chain model. The price log returns of the CME gold futures contract are categorised into three states: the central state coincides with the null return, while the external states represent the positive and negative returns. The central state includes all the price returns that are within a $\sigma \setminus 2$ distance from the zero return, to allow for small variations around the null return.

Once a categorical time series from the price returns is obtained, we estimate the probabilities matrices. The probabilities p_{ij} for a simple Markov chain can be estimated from the data using the Maximum Likelihood Estimation (see, e.g., Billingsley (1961a,b)). Given the stochastic nature of the matrix, we have to estimate $m(m - 1)$ probabilities.

If we consider an observed time series $s_1^n = s_1, s_2, \dots, s_n$ with n observations as an outcome of the random process S^n , the probability of its realisation is:

$$\Pr(S^n = s^n) = \Pr(S_1 = s_1) \prod_{t=2}^n \Pr(S_t = s_t | S_{t-1} = s_{t-1}). \quad (5.18)$$

We can rewrite this probability in terms of transition probabilities to obtain the likelihood function of a sample given the Markov chain model:

$$L(p) = \Pr(S_1 = s_1) \prod_{i=1}^m \prod_{j=1}^m p_{ij}^{n_{ij}}, \quad (5.19)$$

where n_{ij} is the occurrences of transitions from state i to state j . Maximising the logarithm of the likelihood function,

$$\mathcal{L}(p) = \log L(p) = \log \Pr(S_1 = s_1) + \sum_{i,j} n_{ij} \log p_{ij}, \quad (5.20)$$

we obtain the maximum likelihood estimator for the transition probabilities

$$\hat{p}_{ij} = \frac{n_{ij}}{\sum_{j=1}^m n_{ij}}. \quad (5.21)$$

The extension to the multivariate Markov chain is straightforward. The estimator in (5.21) becomes

$$\hat{p}_{ij}^{(\beta,\alpha)} = \frac{n_{ij}^{(\beta,\alpha)}}{\sum_{j=1}^m n_{ij}^{(\beta,\alpha)}}, \quad (5.22)$$

where i is the initial state of series β , and j is the arrival state of series α .

In this application with six series, modelled with a 3-state Markov chain, we have 36 matrices of nine elements each for each day of the analysed period. Because every matrix is a stochastic matrix of the type in (5.3), we have to estimate six elements in each matrix with the estimator (5.22). Every price change series will have six transition matrices, where each element is the probability of transitioning from state i of series α at time t to state j of series β at time $t+1$. For every matrix, states i are read by rows and states j are read by columns, therefore each row sums to one. An example of transition probability matrices for the price influence of all series on CME contract, for the data relative to 23 February 2016, is presented in Table 5.2. In each matrix, we note a higher probability of arriving in a central state from external states, showing a mean reversion pattern of the return series.

According to Equation (5.8), the probability of being in one of the states of a specific price series is given by a linear combination of the vector of initial states of all series, price influence weights $\lambda_{\beta,\alpha}$ and transition probabilities. Therefore, the next step is to estimate the portions of price influence. The estimation of the parameters $\lambda_{\beta,\alpha}$ is obtained maximising the likelihood function of the MTD model (see Berchtold and Raftery (2002)),

$CME \rightarrow CME$	$DGEX \rightarrow CME$	$GLD \rightarrow CME$
$\begin{pmatrix} 0.221 & 0.501 & 0.279 \\ 0.183 & 0.629 & 0.189 \\ 0.281 & 0.488 & 0.231 \end{pmatrix}$	$\begin{pmatrix} 0.23 & 0.506 & 0.264 \\ 0.204 & 0.592 & 0.204 \\ 0.25 & 0.492 & 0.257 \end{pmatrix}$	$\begin{pmatrix} 0.261 & 0.486 & 0.252 \\ 0.195 & 0.604 & 0.201 \\ 0.256 & 0.48 & 0.265 \end{pmatrix}$
$TOCOM \rightarrow CME$	$SGE \rightarrow CME$	$SHFE \rightarrow CME$
$\begin{pmatrix} 0.255 & 0.506 & 0.239 \\ 0.207 & 0.581 & 0.212 \\ 0.23 & 0.514 & 0.256 \end{pmatrix}$	$\begin{pmatrix} 0.352 & 0.5 & 0.148 \\ 0.211 & 0.572 & 0.217 \\ 0.217 & 0.478 & 0.304 \end{pmatrix}$	$\begin{pmatrix} 0.245 & 0.507 & 0.247 \\ 0.209 & 0.58 & 0.212 \\ 0.226 & 0.508 & 0.266 \end{pmatrix}$

Table 5.2: Example of transition probability matrices for the CME gold futures price series for 23 February 2016, picked at random. All the contracts are modelled with a 3-state Markov chain, with the central state being the zero log return, and the external states the positive and negative log return. Every matrix represents the transition probabilities of being in a state i in a specific contract at time $t - 1$ and ending in a state j of the CME contract at time t . For every matrix, states i are read by rows and states j are read by columns. Each row sums to one.

$$\log L(MTD) = \sum_{i_1, \dots, i_\gamma, j=1}^m n_{i_1, \dots, i_\gamma, j} \log \left(\sum_{\beta=1}^{\gamma} \lambda_{\beta, \alpha} \mathbf{P}_{i_\beta, j}^{(\beta, \alpha)} \right), \quad (5.23)$$

where $n_{i_1, \dots, i_\gamma, j}$ is the observed number of sequences of the type $S_{t-1}^{(1)} = i_1, \dots, S_{t-1}^{(\gamma)} = i_\gamma, S_t^{(\alpha)} = j$, respecting constraints (5.10) and (5.11).

The software implementation performs a minimisation of the negative likelihood function, using a minimisation algorithm under constraints, namely *Sequential Least Squares Programming* (SLSQP) optimisation subroutine originally implemented by Kraft (1988).

The maximisation of the log-likelihood of the MTD function for the gold contracts results in a matrix of price influence shares for each day of the analysis, like the matrix in (5.12). Averaging the lambda matrices across the whole period, we obtain a summary measure of price influence, as shown in Table 5.3. The first row indicates that the CME price series is influenced on average by CME itself for 81%, DGX for 12%, GLD for 6%, and SHFE for 1%. The second row represents the influence on DGX price change series from all the series, and so on. Elements on the diagonal represent the self-influence, and they are generally the highest values, showing that most of the price update comes from the previous time price change

	CME	DGCX	GLD	TOCOM	SGE	SHFE
CME	0.81	0.12	0.06	0.00	0.00	0.01
DGCX	0.45	0.38	0.15	0.00	0.00	0.02
GLD	0.31	0.13	0.55	0.00	0.00	0.01
TOCOM	0.22	0.00	0.13	0.65	0.00	0.00
SGE	0.11	0.12	0.13	0.08	0.43	0.13
SHFE	0.23	0.02	0.26	0.00	0.00	0.49

Table 5.3: Matrix of price influences for six gold contracts traded in different markets and averaged across the period January 2016 to December 2017. The selected time range is from 1.30PM to 6.30PM UTC time to allow for full overlap of all price series. Days of holidays in any of the markets are excluded from the computation for a total of 424 trading days. Each row represents the share of price influence for each contract (in top row) on a specific price change series (in first column). The sum of each row is equal to 1. Results show a general strong price change influence from CME gold futures contract on all the other contracts, and a weak influence of the SGE spot contracts.

of the same series.

Finally, from the daily price influence matrices, we can derive the Price Leadership Share measure according to equation (5.15). Recall that the PLS is the total external price change influence. Thus we have to exclude the self-influence, and it represents the probability that a series is the price leader. Figure 5.2 shows the dynamic of the PLS measure, averaged at monthly sampling, over the two years. Results are reasonably stable across the period, with CME maintaining its leadership position and DGCX gaining almost the same share as GLD after October 2017. The other series are quite stable and close to zero price leadership. SGE share is zero most of the period except for July 2017.

Averaging the daily PLS vectors across all period, we obtain a further summary measure for the PLS as reported in Table 5.4. Similarly to Figure 5.2, results show a stronger price leadership role from the CME gold futures contract, CME series being the mode of the probability distribution of PLS. The ETF contract, i.e. GLD, appears to have a quite important position in the leadership together with DGCX. On the contrary, SHFE, TCE, and SGE occupy the lowest position, demonstrating an almost total absence of leadership. Overall, all futures contracts lead the SGE spot market, confirming previous literature findings on price discovery.

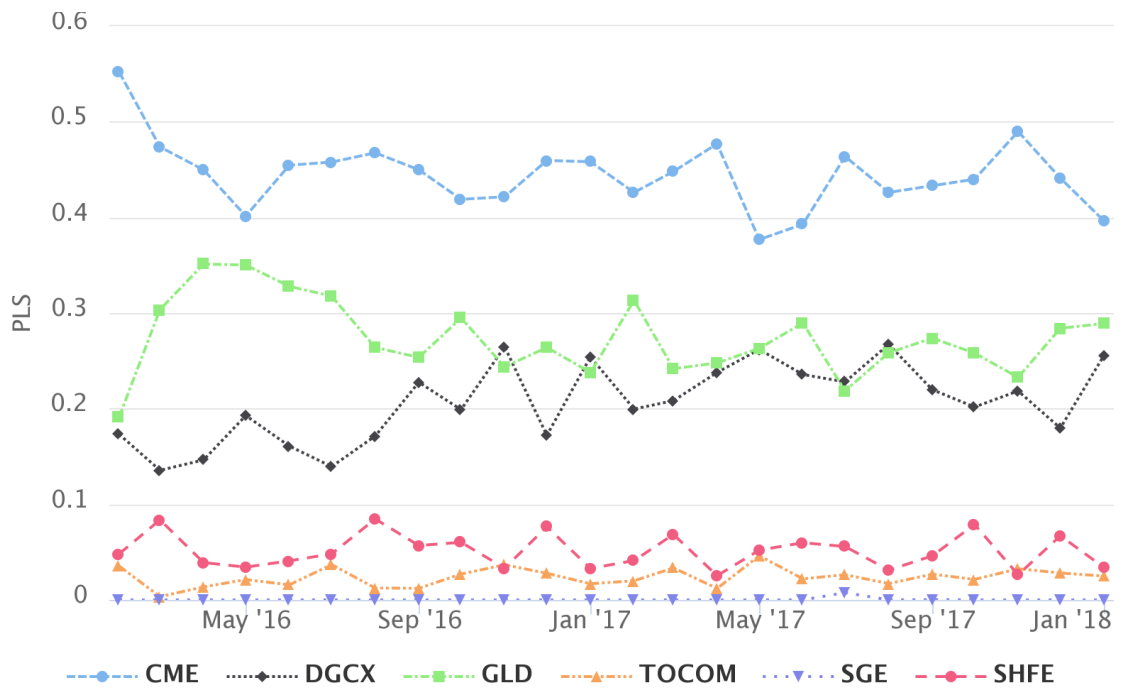


Figure 5.2: Price Leadership Share over the period January 2016 to December 2017, for the time range from 1.30PM to 6.30PM UTC to allow for full overlap of all price series. PLS are sampled monthly. CME series appears to lead all other series for the entire period. After an initial difference, DGXCX and GLD are almost aligned in term of price leadership, while the other series do not show relevant PLS values.

Contract	PLS
CME	0.447
GLD	0.274
DGCX	0.205
SHFE	0.051
TOCOM	0.023
SGE	0.000

Table 5.4: Price Leadership Share for the seven gold contracts traded in different markets across the period January 2016 - December 2017. The selected time range is from 1.30PM to 6.30PM UTC time to allow for full overlap of all price series. On average, there is a strong price leadership from the CME gold futures contract, followed by the ETF contract. Overall, all the futures contract lead the spot markets, SGE and LME



Figure 5.3: Weekly leadership concentration over the period January 2016 to December 2017, for the time range from 1.30PM to 6.30PM UTC to allow for full overlap of all price series. The value of the concentration L can range between zero, i.e. absence of leadership, and 1, i.e. leadership fully concentrated in one series.

Furthermore, we can compute a single indicator that indicates how concentrated is the leadership, using formula (5.17), with $0 \leq L \leq 1$. Zero means total absence of leadership when all series have 1/6 probability of being the leader, and one means that the leadership is entirely concentrated in one series. The evolution of the concentration index for six gold contracts is shown in Figure 5.3. The price leadership concentration is calculated as a monthly average of the concentration series and shows a slight downward trend, signalling that the leadership becomes less concentrated towards the end of the period, due to an increase of leadership share of DGCX, this being observable from Figure 5.2.

The analysis on six gold contracts shows a strong price leadership from CME with a weaker, almost close to zero, share of leadership for the Asian contracts. One of the reasons for this high difference is the selected time range for the analysis, i.e. 13:30 to 18:30 UTC. During these hours, in the all US markets are open for the day session, while in the Asian markets there is the night session open and in Dubai the evening one. Therefore, it quite natural to expect a higher activity and a higher efficiency in the US contracts.

In the remainder of this section, we analyse how price leadership shares change if we perform the estimation selecting a time range in which the Asian markets are

Contract	PLS		
	13:30-18:30	5:30-6:15	1:00-3:30
CME	0.447	0.507	0.348
GLD	0.274	n/a	n/a
DGCX	0.205	0.237	n/a
SHFE	0.051	0.143	0.257
TOCOM	0.023	0.076	0.198
SGE	0.000	0.036	0.197
\bar{L}	0.286	0.204	0.021
\bar{L}	0.366	0.360	0.071

Table 5.5: PLS comparison for three different time ranges. All hours in UTC time. Some of the contracts are not available because they are closed during selected time range

open for the day session, and the US is in the night session. We first select a time range between 5:30 and 6:15 UTC, that corresponds to the second morning session of SGE and SHFE. For this selection, we have to drop the ETF contract because it is closed. Then, we select the first session of the morning for SGE and SHFE, between 1:00 and 3:30 UTC, and in this case, we have to drop also the Dubai contract.

Table 5.5 reports a comparison of PLS calculated for the three selected time ranges, along with the concentration index in two versions, (i) as a measure of the summarised PLS vector over the period, and (ii) as the average of the daily concentration measures.

Results from Table 5.5 evidence a confirmed price leadership from CME contracts when considering the time range 5:30-6:15, instead we observe a slight reduction in its share in the third interval analysed. It is also noticeable the increase in price leadership share for the Asian contracts, especially when we consider the first session of the morning (third column), where SHFE, TOCOM, and SGE gain a consistent share. Also, looking at the concentration index, we notice a considerable redistribution of the leadership for the same session.

Analysing the PLS for the morning session of SHFE and SGE, we cannot attribute a clear leadership to CME, especially if we take into account that the tick size of the Asian contract is slightly bigger than the CME contract.

Contract	PLS			
	1-second 3-state	1-second 5-state	10-second 3-state	1-minute 3-state
CME	0.447	0.519	0.354	0.119
GLD	0.274	0.211	0.119	0.131
DGCX	0.205	0.227	0.120	0.131
SHFE	0.051	0.033	0.177	0.181
TOCOM	0.023	0.001	0.103	0.187
SGE	0.000	0.000	0.058	0.251
L	0.286	0.350	0.084	0.020
\bar{L}	0.366	0.418	0.286	0.236

Table 5.6: Robustness analysis of the PLS measure over different sampling intervals, i.e. 10 seconds and one minute, and modelling the price change series with a 5-state Markov chain.

5.3.3 Robustness tests

In addition to the previous estimation, we perform a further analysis estimating the PLS and respective concentration index values at different sampling intervals, i.e. 10 seconds and one minute, and with a 5-state Markov chain. Results reported in Table 5.6 highlight that increasing the number of states from three to five produces better results in term of price leadership share, even though we note a swap of positions between GLD and DGX, possibly due to the averaging of the measure over the period. In fact, looking at Figure 5.2, the PLS evolution between these two series appears to converge to similar values. This trend is better defined when modelling price changes with a 5-state chain. Also, the concentration index registers a slight increase denoting a more concentrated leadership. However, because the calculation with five states requires more transitions probabilities to estimate and more computational time, there is a trade-off between computational effort and precision of the measure. Therefore, we suggest performing the analysis with a 3-state Markov chain.

On the contrary, looking at different sampling frequencies, we note a complete reversal of the leadership when we sample price changes with 1-minute interval and a small change when sampling at a 10-second interval. At first sight, these figures seem to show a general instability of the model. However, if we observe the concentration

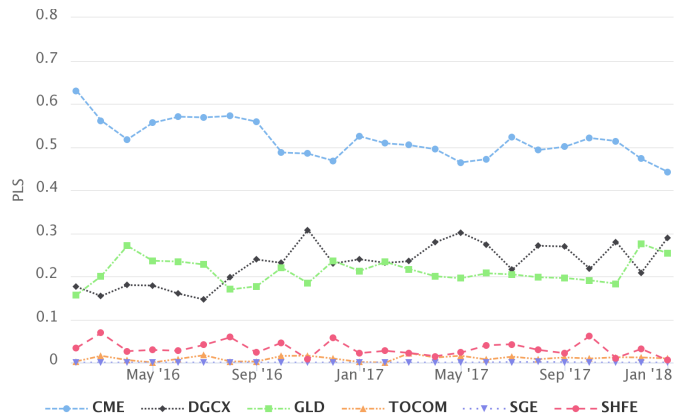
values for both frequencies, we cannot consider the PLS results as being valid as the 1-second interval ones. The leadership concentration of the PLS vector, i.e. 0.02, is a value close to zero and expresses a complete lack of concentration and a signal of an absence of clear price leadership. Instead, the average of the concentration indices seems to propose a leadership concentration similar to 1-second sampling interval. Nevertheless, this is a consequence of the increased variability in the PLS measure when sampling at lower frequencies as reported in Figure 5.4(c).

A reason for this high discrepancy of the PLS results sampled at a lower frequency is due to loss of information about price changes that occur within the interval. Time aggregation generally produces a contemporaneous correlation between price innovations. The rationale of our model is that once a market updates its price, the other market responds to that signal with a price change. If the interval is too long, all changes in the same interval appear to be contemporaneous, and the model loses the ability to capture the price change dynamics (see, e.g. Hasbrouck, 1995). Therefore, our model performs better at a shorter sampling interval, ideally up to event time. However, because we model the price changes with a discrete-time homogeneous Markov chain we have to limit our analysis to fixed time intervals.

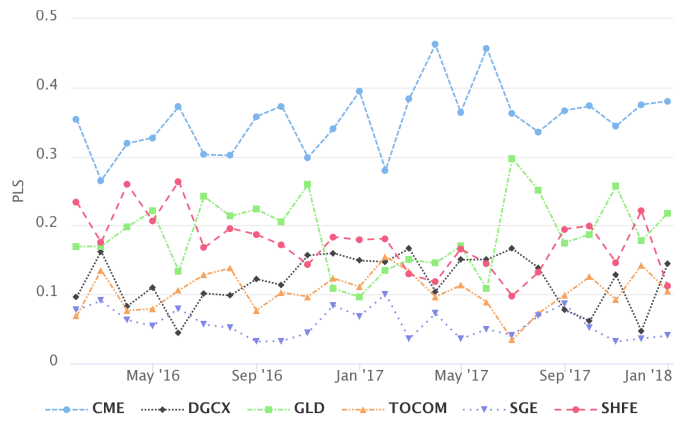
5.4 Conclusion

In this chapter we proposed a new measure of price discovery using a multivariate Markov chain model. This new measure, that we called the price leadership share, is based on the observation of the dependencies between price changes of cointegrated price series. It captures the dynamics of price change and, therefore, the timing of incorporating new information. Further, along with the price leadership share, we proposed a concentration index that measures the concentration of price leadership and it is useful to make a comparison between different PLS outcomes.

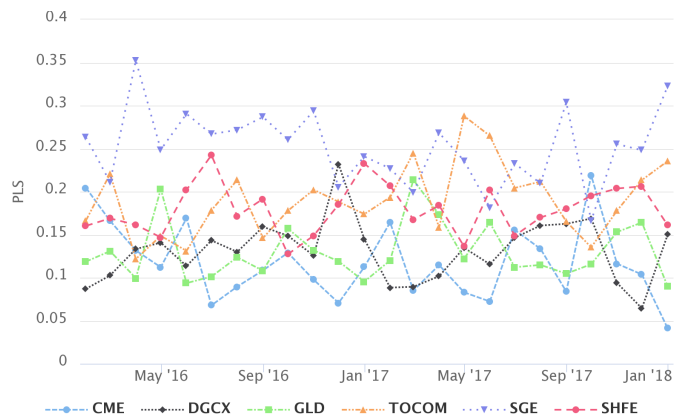
The model is tested with an empirical application using six gold contracts across the world, such as spot, futures and ETF contracts. The test confirms results that were expected from the simple microstructure analysis, as well as previous literature



(a) 5 states, 1 second interval



(b) 3 states, 10 second interval



(c) 3 states, 1 minute interval

Figure 5.4: Price Leadership Share sampled monthly over the period January 2016 - December 2017, time range 1.30PM to 6.30PM UTC time. (a) Price changes modelled with a 5-state Markov chain and midquote sampled at 1-second interval. (b) Price changes modelled with a 3-state Markov chain and midquote sampled at 10-second interval. (c) Price changes modelled with a 3-state Markov chain and midquote sampled at 1-minute interval.

findings on price discovery. Moreover, we show that the PLS offers advantages over the existent measures, i.e. information share, component share, and informational leadership share, because it can easily process more than two price series simultaneously and produces a ranking of the price leadership and a model of the price change dependencies between series. Also, the price leadership share measure does not require us to include lagged observations in the model, in contrast to what happens with other measures. Hence, our model can benefit from a bigger sample for the analysis that proves to be a strength in the case of illiquid stocks with a small set of observations.

6 Algorithmic implementation

6.1 Introduction

In this chapter, we present a computational implementation of the models proposed in this thesis. First, we analyse how to estimate the parameters of a multivariate Markov chain. We skip the univariate case as it is straightforward and a special case of the multivariate one. Then, we describe the stock valuation problem, from the multivariate perspective. Again, the univariate case can be solved inputting only one series to the multivariate model. Finally, we discuss the computation of the price leadership share measure.

All procedures are introduced in Python¹ programming language because it is an open source language and permits an easy readability and reusability of the code. Besides, we use dedicated packages for an efficient implementation of our algorithms, such as Numpy² and Scipy³ for scientific computations, and Pandas⁴ for data analysis and manipulation. The import routine of mentioned packages is describe in Listing 1.

```
1 import pandas as pd
2 import numpy as np
3 from scipy.optimize import Bounds
4 from scipy.optimize import minimize
```

Listing 1: Preliminary imports.

¹<https://www.python.org/>

²<http://www.numpy.org/>

³<https://www.scipy.org/>

⁴<https://pandas.pydata.org/>

6.2 Markov chain parameters estimation

When modelling financial series with a Markov chain, the first step to perform the computation is to create a correspondence between returns of the series and states of the chain. This discretisation procedure is explained in Listing 2. After importing the return series⁵ into a Pandas dataframe, identified in listings as *df*, we have to identify the categories' edges. The function *get_edges* accepts two inputs, namely the dataframe and the number of states. First, it initialises a dictionary to store the edges by series, then it loops over the series to find specific edges for each series α . The function calculates $m + 1$ edges, with m being the number of states of the Markov chain. Each edges has a width of size σ , and it is centred around the zero. Finally, the external edges are changed according to the minimum and maximum of the distribution. Once the edges are ready, we can use the *cut* function from Pandas to assign each return to the corresponding state for each series.

```
1 def get_edges(df, m):
2     edges = {}
3     series = df.columns
4     for alpha in series:
5         std = df[alpha].std()
6         edges[alpha] = [std*(-m/2+x) for x in range(m+1)][1:-1]
7         edges[alpha].append(df[alpha].max())
8         edges[alpha].insert(0,df[alpha].min())
9     return edges
10
11 edges = get_edges(df, m)
12 g = pd.DataFrame()
13 for key, value in edges.items():
14     g[key] = pd.cut(df[key],value,labels=False, include_lowest=True)
```

Listing 2: Discretisation procedure.

From the categorised returns, we can estimate the parameters of the multivariate Markov chain. First, we estimate the transition probabilities of equation (2.13). We

⁵The series has to be sorted from the oldest observation to the most recent one to run the code correctly.

recall that the estimator for the transition probabilities is

$$\hat{p}_{ij}^{(\beta,\alpha)} = \frac{n_{ij}^{(\beta,\alpha)}}{\sum_{j=1}^m n_{ij}^{(\beta,\alpha)}}, \quad (6.1)$$

where $n_{ij}^{(\beta,\alpha)}$ is the occurrences of transitions from state i of series β to state j of series α .

Because these probabilities are generally presented in form of a matrix, an easy way to build an estimation routine is to create nested loops with a counter of the observations' frequencies. However, considering that Pandas has extremely efficient management of the indices, we found useful to build the transition probabilities in a columnar form instead than a matrix one. To this extent, we create a dataframe with a double index containing all combinations of starting-ending states of the transitions, i.e. $i \rightarrow j$. The probability dataframe contains as many columns as all combinations of series β and α , i.e. γ^2 according to condition (2.12). In the case of the univariate model, the dataframe will have only one column.

```

1 def get_p(g, m):
2     list_ = []
3     for i in range(m):
4         for j in range(m):
5             list_.append([i,j])
6     index_col = ['i','j']
7     p_index = pd.DataFrame(list_, columns=index_col)
8
9     series = g.columns
10    f = lambda x: x/x.sum()
11    list_ = []
12    for beta in series:
13        for alpha in series:
14            p = pd.concat([g[beta],g[alpha].shift(-1)],
15                          axis=1)[: -1].astype(dtype='int')
16            p.columns = index_col
17            p = pd.DataFrame(p.groupby(index_col).size(),
18                          columns=['freq']).reset_index()
19            p = p.merge(p_index, on=index_col, how='right').fillna(0)
20            p[f'p{beta}_{alpha}'] = p['freq'].groupby(p['i']).transform(f)
21            p = p.drop('freq', axis=1).set_index(index_col)
22            list_.append(p)
23    return pd.concat(list_, axis=1)

```

Listing 3: Transition probabilities.

For example, applying the function `get_p` to the same gold data of Chapter 5, we

		pCME_CME	pCME_DGX	pCME_GLD	pCME_TCE	pCME_SGE	...
<i>i</i>	<i>j</i>						
0	0	0.221	0.222	0.190	0.150	0.008	...
	1	0.501	0.701	0.677	0.796	0.985	...
	2	0.279	0.078	0.133	0.055	0.007	...
1	0	0.183	0.080	0.114	0.057	0.005	...
	1	0.629	0.838	0.765	0.886	0.990	...
	2	0.189	0.082	0.121	0.057	0.005	...
2	0	0.281	0.081	0.138	0.046	0.007	...
	1	0.488	0.694	0.667	0.818	0.985	...
	2	0.231	0.225	0.196	0.135	0.008	...

Table 6.1: Example of transition probabilities as output of the function in Listing 3

obtain an output similar to Table 6.1. It is worth mentioning that Python counters always start from zero, thus, in the remainder of this chapter, the states of the Markov chain will be considered starting from zero as well, e.g. a 3-state Markov chain will have states 0, 1, and 2. This does not affect the final results as it is an internal computation that can be adapted at the end adding one to all states values.

Next step is the estimation of the weights $\lambda_{\beta,\alpha}$ of the mixture transition distribution model. This estimation is performed maximising the log likelihood function of the MTD model

$$\log L(MTD) = \sum_{i_1, \dots, i_\gamma, j=1}^m n_{i_1, \dots, i_\gamma, j} \log \left(\sum_{\beta=1}^{\gamma} \lambda_{\beta, \alpha} \mathbf{P}_{i_\beta, j}^{(\beta, \alpha)} \right), \quad (6.2)$$

where $n_{i_1, \dots, i_\gamma, j}$ is the observed number of sequences of the type $S_{t-1}^{(1)} = i_1, \dots, S_{t-1}^{(\gamma)} = i_\gamma, S_t^{(\alpha)} = j$, respecting constraints (2.14) and (2.14).

Listing 4 presents a function to obtain frequencies $n_{i_1, \dots, i_\gamma, j}$, along with transition probabilities $p_{i_\gamma, j}^{(\beta, \alpha)}$ that will be used in the MLE function.

The function *get_frequencies* accepts the returns series g and the transition probabilities as inputs, and creates a dictionary of dataframes, one for each return series (first loop of the function). Each dataframe contains a left section with γ columns of starting states i_γ , one for each series, plus a column for the arrival state j in the specific series α , named for example *CME+1* if we consider CME series. Then, the dataframe contains a column with frequencies of the combinations i_1, \dots, i_γ, j . Frequencies are computed grouping the observations by combinations, and calculating

```

1 def get_frequencies(g, p):
2     f = {}
3     series = g.columns.tolist()
4     for alpha in series:
5         temp = g[series]
6         temp[alpha+'+1'] = temp[alpha].shift(-1)
7         temp = temp[:-1].astype(dtype='int')
8         series_plus = temp.columns.tolist()
9         temp = pd.DataFrame(temp.groupby(series_plus).size(),
10                             columns=['freq']).reset_index()
11     for beta in series:
12         p_ba = p[f'p{beta}_{alpha}'].reset_index()
13         p_ba = p_ba.rename(columns={'i':beta,
14                                     'j':alpha+'+1'})
15         temp = temp.merge(p_ba,
16                             on=[beta,alpha+'+1'],
17                             how='left')
18     f[alpha] = temp
19     return f

```

Listing 4: Computation of the frequencies for the $\lambda_{\beta,\alpha}$ estimation.

CME	DGX	GLD	TCE	SGE	SFE	CME+1	freq	pCME_CME	pDGX_CME	pGLD_CME	pTCE_CME	pSGE_CME	pSFE_CME
0	0	0	0	0	1	0	1	0.220592	0.230071	0.261424	0.254682	0.351852	0.208820
0	0	0	0	0	1	1	1	0.500655	0.505550	0.486208	0.506367	0.500000	0.579642
0	0	0	0	1	0	0	22	0.220592	0.230071	0.261424	0.254682	0.211228	0.245455
0	0	0	0	1	0	1	30	0.500655	0.505550	0.486208	0.506367	0.571975	0.507273
0	0	0	0	1	0	2	19	0.278753	0.264379	0.252367	0.238951	0.216797	0.247273
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 6.2: Example of frequencies dataframe as element of the output dictionary from function in Listing 4

the size of each the group. Eventually, the function attaches to each dataframe γ values of transition probabilities that are specific to each combination i_γ, j (second loop of the function).

Considering the data from Chapter 5 and specifically CME series, the outcome of the function is reported in Table 6.2.

The frequency dataframe serves as input of the routine that maximises the log likelihood function in (6.2). In practice, the procedure performs a minimisation of the negative log likelihood function using the *minimize* routine of the package Scipy with the method SLSQP, i.e. *Sequential Least Squares Programming* algorithm to minimize a function of multiple variables with different combinations of bounds or constraints (Kraft, 1988). Listing 5 describes the *MLE* function that reproduces the equation in (6.2) using as inputs an initial lambda vector, a frequency dataframe, and the targeted series. Then, the *get.lambda* function, looping over the series,

uses the *MLE* routine as objective function of the minimisation procedure for each specific destination series α , after setting the initial bounds (2.14), initial lambda vector, e.g. $[0,0,0,0,0,1]$, and constraint (2.15). Finally, results are aggregated into a matrix form. In the case of the univariate Markov chain model, the routing will output a single value of lambda equal to one.

```

1 def MLE(lambda0, f, alpha):
2     L = f[f.columns]
3     gamma = len(lambda0)
4     series = L.columns[:gamma].tolist()
5     L['p'] = 0
6     for i, beta in enumerate(series):
7         L['p'] += lambda0[i]*L[f'p{beta}_{alpha}']
8     L['MLE'] = L['freq']*np.log(L['p']+.0000001)
9     return -L['MLE'].sum()
10
11 def get_lambda(series, f):
12     gamma = len(series)
13     # setup of bounds (2.14) and initial vector lambda
14     lambda0 = []
15     bounds = ((0,None),)
16     for beta in range(gamma-1):
17         lambda0.append(0)
18         bounds += ((0,None),)
19     lambda0.append(1)
20     lambda0 = np.array(lambda0)
21
22     # setup of condition (2.15)
23     cond = "lambda x: np.array(["
24     cond += ' + '.join(['x[' + str(beta) + ']'
25                         for beta in range(gamma)]) + ' - 1])'
26     cond = eval(cond)
27
28     # minimisation function
29     lambda_mat = []
30     for alpha in series:
31         eq_cons = {'type': 'eq', 'fun' : cond}
32         logL = minimize(MLE, lambda0,
33                        args=(f[alpha], alpha),
34                        jac="2-point",
35                        method='SLSQP',
36                        constraints=eq_cons, bounds=bounds)
37         lambda_row = []
38         for beta in range(gamma):
39             lambda_row.append(round(logL.x[beta],3))
40         lambda_mat.append(lambda_row)
41     return pd.DataFrame(lambda_mat, index=series, columns=series)

```

Listing 5: Estimation of $\lambda_{\beta,\alpha}$ through MLE.

6.3 Markov stock model implementation

In this section we propose an implementation of the Markov stock model from Chapters 3 and 4. As previous section, we discuss only the multivariate implementation as the univariate one is a special case of the multivariate with one dividend series.

First, we have to adapt the discretisation procedure to calculate the edges according to the dividend growth process in (3.7). The *get_edges* function in Listing 2 is useful for log returns. Therefore, for this application of the Markov stock model, we convert our dividend returns in log returns, get the edges and take the exponential of the results. In the same function, we include the calculation of the values of the growth for each state of the Markov chain and for every series α , using the median of observed growths for each category.

```
1 def get_edges(df, m):
2     # get the logarithm of the process
3     df = df.apply(lambda x: np.log(x))
4     edges = {}
5     state_value = {}
6     series = df.columns
7     for alpha in series:
8         std = df[alpha].std()
9         edges[alpha] = [std*(-m/2+x) for x in range(m+1)][1:-1]
10        edges[alpha].insert(0, df[alpha].min())
11        edges[alpha].append(df[alpha].max())
12        # compute growth value of the state
13        state_value[alpha] = []
14        for i in range(m):
15            median = df[(df[alpha]>edges[alpha][i]) &
16                       (df[alpha]<=edges[alpha][i+1])][alpha].median()
17            state_value[alpha].append(median)
18        # get the exponential value
19        edges[alpha] = [np.exp(x) for x in edges[alpha]]
20        state_value[alpha] = [np.exp(x) for x in state_value[alpha]]
21    return edges, state_value
```

Listing 6: Function *get_edges* adapted for the dividend growth series.

After obtained the transition probabilities and the lambda values using routines in previous section, we define algorithms for implementing Theorem 4.4, 4.10, and 4.14 to obtain first- and second-order price-dividend ratios.

For computing the price-dividend ratios, we solve linear systems of equations

of the type $B = xA$, where B is a vector of known terms and A is a matrix of coefficients of the unknown terms.

We start rewriting equation from Theorem 4.4 and identify the blocks that we need to compute in the algorithm

$$\begin{aligned}
& \underbrace{\sum_{j_\alpha=1}^m \sum_{\beta=1}^{\gamma} \sum_{h=1}^m e_h^{(\beta)}(k) \lambda_{\beta,\alpha} \mathbf{P}_{h,j_\alpha}^{(\beta,\alpha)} g_{j_\alpha}^{(\alpha)}}_{Pg} = r_\alpha \cdot \underbrace{\psi_1^{(\alpha)}(g_{a_1}^{(1)}, \dots, g_{a_\gamma}^{(\gamma)})}_x \\
& - \sum_{j_1, \dots, j_\gamma=1}^m \underbrace{\psi_1^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)})}_x \cdot g_{j_\alpha}^{(\alpha)} \cdot \underbrace{\prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w,f} \mathbf{P}_{c,j_f}^{(w,f)}}_{prodP} .
\end{aligned} \tag{6.3}$$

The same applies to equation of Theorem 4.10 for the second-order price-dividend ratios.

$$\begin{aligned}
& \underbrace{\sum_{j=1}^m \sum_{\beta=1}^{\gamma} \sum_{h=1}^m e_h^{(\beta)}(k) \lambda_{\beta,\alpha} \mathbf{P}_{h,j}^{(\beta,\alpha)} (g_j^{(\alpha)})^2}_{Pg_sq} \\
& + 2 \sum_{j_1, \dots, j_\gamma=1}^m \underbrace{\psi_1^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)})}_{\text{Solution from (6.3)}} (g_{j_\alpha}^{(\alpha)})^2 \underbrace{\left(\prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w,f} \mathbf{P}_{c,j_f}^{(w,f)} \right)}_{prodP} \\
& \underbrace{\hspace{15em}}_{prodPg_sq} \\
& = r_\alpha^2 \cdot \underbrace{\psi_2^{(\alpha)}(g_{a_1}^{(1)}, \dots, g_{a_\gamma}^{(\gamma)})}_{x^2} \\
& - \sum_{j_1, \dots, j_\gamma=1}^m \underbrace{\psi_2^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)})}_{x^2} (g_{j_\alpha}^{(\alpha)})^2 \underbrace{\left(\prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w,f} \mathbf{P}_{c,j_f}^{(w,f)} \right)}_{prodP} .
\end{aligned} \tag{6.4}$$

It is worth noting that the matrix $prodP$ is constant for every series α and it reduces to matrix P if we consider the univariate case, and both systems have m^γ equations in m^γ unknown. To solve these systems, we start computing the known term Pg at the left side of equation (6.3) and the term Pg_sq from left side of

equation (6.4).

```

1 def get_combiP(m, p, state_value, lambda_values, series):
2     gamma = len(series)
3     # create a list of ranges
4     # for example: s=3 and m=5 ->
5     # [range(0, 5), range(0, 5), range(0, 5), range(0, 5)]
6     ranges = [range(m) for x in range(gamma+1)]
7     # dataframe with all combinations of states and series
8     combinations = list(itertools.product(*ranges))
9
10    combiP = {}
11    Pg = {}
12    Pg_sq = {}
13    for alpha in series:
14        df = pd.DataFrame(combinations, columns=series+[alpha+'+1'])
15        # attach transition probabilities and lambda values
16        for beta in series:
17            p_tmp = p[f'p{beta}_{alpha}'].reset_index()
18            p_tmp = p_tmp.rename(columns={'start':beta, 'end':alpha+'+1'})
19            df = df.merge(p_tmp, on=[beta,alpha+'+1'], how='left')
20            df['lambda '+beta+'_'+alpha] = lambda_values.loc[alpha,beta]
21        # compute P looping over beta series
22        df['P'] = 0
23        for beta in series:
24            df['P'] += df['p'+beta+'_'+alpha]*df['lambda '+beta+'_'+alpha]
25        # attach states values g
26        g = pd.DataFrame(state_value[alpha], columns=['g'])
27        df = df.merge(g, how='left', left_on=alpha+'+1', right_index=True)
28        # compute Pg and Pg_sq
29        df['Pg'] = df['P']*df['g']
30        df['Pg_sq'] = df['P']*df['g']**2
31        Pg[alpha] = df.groupby(series)['Pg'].sum().values
32        Pg_sq[alpha] = df.groupby(series)['Pg_sq'].sum().values
33        combiP[alpha] = df
34    return combiP, Pg, Pg_sq

```

Listing 7: Computation of P, Pg, and Pg_sq.

Function `get_pg` in Listing 7 generates a dataframe (`combiP`) that includes all combinations of starting states i and ending states j . The function attaches to each series α and to each combination specific transition probabilities and lambda values corresponding to the combination. Finally, it computes the probability P looping over the β series and multiplies results by the states values and their squared values. From the `combiP` dataframe, grouping by starting states and summing the group results, the function computes the blocks Pg and Pg_sq .

Table 6.3 presents a sample of the output of `combiP` when $\alpha = GPC$ from application in Chapter 4. Columns Pg and Pg_sq from the dataframe corresponds

GPC	DOV	PH	GPC+1	pGPC_GPC	lambda GPC_GPC	pDOV_GPC	lambda DOV_GPC	pPH_GPC	lambda PH_GPC	P	g	Pg	Pg_sq
0	0	0	0	0.4	0.45	0.333333	0.0	0.250000	0.55	0.317500	0.876915	0.278420	0.244151
			1	0.2	0.45	0.333333	0.0	0.500000	0.55	0.365000	1.035565	0.377981	0.391424
			2	0.4	0.45	0.333333	0.0	0.250000	0.55	0.317500	1.073547	0.340851	0.365920
		1	0	0.4	0.45	0.333333	0.0	0.250000	0.55	0.317500	0.876915	0.278420	0.244151
			1	0.2	0.45	0.333333	0.0	0.333333	0.55	0.273333	1.035565	0.283054	0.293121
			2	0.4	0.45	0.333333	0.0	0.416667	0.55	0.409167	1.073547	0.439260	0.471566
		2	0	0.4	0.45	0.333333	0.0	0.000000	0.55	0.180000	0.876915	0.157845	0.138416
			1	0.2	0.45	0.333333	0.0	0.214286	0.55	0.207857	1.035565	0.215250	0.222905
			2	0.4	0.45	0.333333	0.0	0.785714	0.55	0.612143	1.073547	0.657164	0.705496
	1	0	0	0.4	0.45	0.111111	0.0	0.250000	0.55	0.317500	0.876915	0.278420	0.244151
:	:	:	:	:	:	:	:	:	:	:	:	:	:

Table 6.3: Sample of the *combiP* output.

to single values before the grouping function.

Next step is to compute blocks $prodP$, $prodPg$, and $prodPg_sq$ from equations (6.3) and (6.4). The function get_prodP is based on the calculation of the product between transition probabilities weighted with lambda values across all series. Initially, we create a matrix of products of transition probabilities calculating all possible combinations of states for all series using the Python routine *itertools.product*. Then, we loop over all combinations and multiply the elements from combinations of probabilities P . Therefore, we produce a matrix with $m^\gamma \times m^\gamma$ elements.

```

1 def get_prodP(m, series, combiP):
2     ranges = [range(m) for x in range(len(series))]
3     list_prod = []
4     for i in itertools.product(*ranges):
5         list_p = []
6         for alpha in series:
7             list_p.append(combiP[alpha].set_index(series).loc[i]['P'].tolist())
8         product = list(itertools.product(*list_p))
9         list_ = []
10        for item in product:
11            list_.append(np.prod(item))
12        list_prod.append(list_)
13    prodP = np.asarray(list_prod)
14
15    prodPg = {}
16    prodPg_sq = {}
17    for j, alpha in enumerate(series):
18        g = []
19        g = []
20        for i in itertools.product(*ranges):
21            g1.append(state_value[alpha][i[j]])
22            g2.append(state_value[alpha][i[j]**2])
23        prodPg[alpha] = prodP*g1
24        prodPg_sq[alpha] = prodP*g2
25    return prodP, prodPg, prodPg_sq

```

Listing 8: Computation of $prodP$, $prodPg$, and $prodPg_sq$.

For example, analysing the output in Table 6.3, for every combination i in the first loop of function *get_prodP*, the routine select three probabilities P for each series (i.e. from each dataframe *combiP* corresponding to a specific series). In this particular case, we have 3×3 probabilities, that can be combined and multiplied to generate 27 elements that constitute the first row of the matrix.

Finally, each row of the matrix *prodP* is multiplied by a vector of state values, as well as their squared values, ordered according to the sequence of states in series α . The sequence of states corresponds to the sequence observable in the dataframe index for any specific series α .

```

1 def get_psi(series, Pg, Pg_sq, prodP, prodPg, prodPg_sq, r):
2     ranges = [range(m) for x in range(len(series))]
3     combinations = list(itertools.product(*ranges))
4     index = pd.DataFrame(combinations, columns=columns)+1
5     # check conditions A1 and A2
6     for alpha in series:
7         if r[alpha]<max(Pg[alpha]):
8             print('Condition not verified for: '+alpha)
9         if r[alpha]<np.sqrt(max(Pg_sq[alpha])):
10            print('Condition not verified for: '+alpha)
11
12    psi = []
13    for alpha in series:
14        x = np.linalg.solve((np.eye(len(prodPg[alpha]))*(r[alpha])) -
15                            prodPg[alpha],Pg[alpha])
16        psi.append(pd.DataFrame(x, columns=['$\psi$ '+alpha]))
17    psi = pd.concat(psi, axis=1)
18    psi = pd.concat([index, psi], axis=1).set_index(columns)
19
20    B2, A2 = {}, {}
21    for alpha in series:
22        B2[alpha] = 2*(np.sum(psi['$\psi$ '+alpha].values*prodPg_sq[alpha],
23                              axis=1))+Pg_sq[alpha]
24        A2[alpha] = np.eye(len(prodPg_sq[alpha]))*(r[alpha]**2) - prodPg_sq[alpha]
25    psi2 = []
26    for alpha in series:
27        x2 = np.linalg.solve(A2[alpha],B2[alpha])
28        psi2.append(pd.DataFrame(x2, columns=['$\psi^2$ '+alpha+' '+alpha]))
29    psi2 = pd.concat(psi2, axis=1)
30    psi2 = pd.concat([index, psi2], axis=1).set_index(columns)
31
32    return psi, psi2

```

Listing 9: Solution of linear systems of equations to obtain ψ_1 and ψ_2 .

Once all blocks are obtained, we solve both linear systems of equations, creating the vector B and the matrix A and using the Numpy routine *linalg.solve*. It is

important to mention that for the second-order price-dividend ratios we include solutions of the first equation in the computation of vector B , and the required rate of return is stored in a Python dictionary with keys being the series names.

Function *get_psi* from Listing 9 starts with the creation of the dataframe index containing combinations of states for all series, and a check for conditions A1 and A2. Then, we compute psi_1 for every series α and combine the results from all series by columns in a Pandas dataframe. The function follows with the preparation of the vector $B2$ and matrix $A2$ and concludes with computation of psi_2 with all series combined in a single dataframe.

To compute the covariance between stock, we implement an algorithm to solve equation (6.5) from theorem 4.14.

$$\begin{aligned}
r_\alpha r_\beta \psi_2^{(\alpha, \beta)}(g_{a_1}^{(1)}, \dots, g_{a_\gamma}^{(\gamma)}) &= \underbrace{\sum_{j_\alpha, j_\beta=1}^m g_{j_\alpha}^{(\alpha)} g_{j_\beta}^{(\beta)} \left(\underbrace{\prod_{f \in \{\alpha, \beta\}} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w, f} \mathbf{P}_{c, j_f}^{(w, f)}}_{\text{prodP}} \right)}_{\text{prodPg_cov}} \\
&\quad \underbrace{\hspace{10em}}_{\text{add1}} \\
&+ \underbrace{\sum_{j_1, \dots, j_\gamma=1}^m \underbrace{\psi_1^{(\beta)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)})}_{\text{Solution from (6.3) } \beta} (g_{j_\alpha}^{(\alpha)})(g_{j_\beta}^{(\beta)}) \left(\underbrace{\prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w, f} \mathbf{P}_{c, j_f}^{(w, f)}}_{\text{prodP}} \right)}_{\text{prodPg_cov}}}_{\text{add2}} \\
&+ \underbrace{\sum_{j_1, \dots, j_\gamma=1}^m \underbrace{\psi_1^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)})}_{\text{Solution from (6.3) } \alpha} (g_{j_\alpha}^{(\alpha)})(g_{j_\beta}^{(\beta)}) \left(\underbrace{\prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w, f} \mathbf{P}_{c, j_f}^{(w, f)}}_{\text{prodP}} \right)}_{\text{prodPg_cov}}}_{\text{add3}} \\
&+ \underbrace{\sum_{j_1, \dots, j_\gamma=1}^m \underbrace{psi_1^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)})}_{\text{Solution from (6.3) } \alpha} \underbrace{\psi_1^{(\beta)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)})}_{\text{Solution from (6.3) } \beta} (g_{j_\alpha}^{(\alpha)})(g_{j_\beta}^{(\beta)}) \left(\underbrace{\prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w, f} \mathbf{P}_{c, j_f}^{(w, f)}}_{\text{prodP}} \right)}_{\text{prodPg_cov}}}_{\text{add4}}
\end{aligned} \tag{6.5}$$

Listing 10 shows the function *get_psi_cov* in which we first compute the block

$prodPg_cov$ multiplying the matrix $prodP$ by the product of the values of the states for both series, α and β . This computation is done in the same way as for block $prodPg$ and $prodPg_sq$ from Listing 8. Then, we calculate all four addends from equation (6.5) that are finally summed and divided by the required rate of returns for both stocks. This operation is repeated for any combination of the covariance matrix between stocks, without considering repetitions. Results are joined in a single dataframe psi_cov .

```

1 def get_psi_cov(series, prodP, psi, r):
2     ranges = [range(m) for x in range(len(series))]
3     combinations = list(itertools.product(*ranges))
4     index = pd.DataFrame(combinations, columns=columns)+1
5     psi_cov = []
6     # compute the combination of series for covariance
7     # ('GPC', 'DOV')
8     # ('GPC', 'PH')
9     # ('DOV', 'PH')
10    for alpha_beta in itertools.combinations(series,2):
11        alpha = alpha_beta[0]
12        beta = alpha_beta[1]
13        g = []
14        for i in combinations:
15            # get the state for series alpha and beta based on combination i
16            i_alpha = i[series.index(alpha)]
17            i_beta = i[series.index(beta)]
18            # get the value of the state corresponding to combination i
19            g_alpha = state_value[alpha][i_alpha]
20            g_beta = state_value[beta][i_beta]
21            g.append(g_alpha*g_beta)
22        prodPg_cov = prodP*g
23
24        add1 = np.sum(prodPg_cov, axis=1)
25        add2 = np.sum(psi['$\psi$ '+alpha].values*
26                    prodPg_cov, axis=1)
27        add3 = np.sum(psi['$\psi$ '+beta].values*
28                    prodPg_cov, axis=1)
29        add4 = np.sum(psi['$\psi$ '+alpha].values*
30                    psi['$\psi$ '+beta].values*
31                    prodPg_cov, axis=1)
32        cov = (add1+add2+add3+add4)/(r[alpha]*r[beta])
33        psi_cov.append(pd.DataFrame(cov, columns=['$\psi^2$ '+alpha+' '+beta]))
34    psi_cov = pd.concat(psi_cov, axis=1)
35    psi_cov = pd.concat([index, psi_cov], axis=1).set_index(series)
36    return psi_cov

```

Listing 10: Solution of Theorem 4.14.

The output of the dataframe psi is reported in Table 4.6, and joint output from dataframes $psi2$, and psi_cov is in Table 4.7.

Eventually, we use results from *psi_cov* to compute the covariance function in equation (4.22).

6.4 Price Leadership Share implementation

For the PLS measure, first we have to estimate the parameters of the multivariate Markov chain. Once the lambda values are calculated, we can compute the Price Leadership Share implementing equation (5.15) and the concentration index in (5.17).

Listing 11 includes two functions, one to obtain the PLS vector and the other for the *L* index. The *PLS* function has the lambda matrix as input. First, we set the diagonal values to zero, then we normalise the rows to one (see line 6). Finally, we sum the values by columns and normalise to one. A vector of PLS is returned by the function.

Function *L_index* takes the PLS vector as input and compute the entropy of the probability distribution according to equation (5.16) Then, the entropy result is normalised to its maximum value $\log(\gamma)$.

```

1 def PLS(lambda_mat):
2     mat = lambda_mat.values.copy()
3     np.fill_diagonal(mat,0)
4     if len(mat)>2:
5         col_sum = np.sum(mat, axis=1)[: ,None]
6         mat = np.nan_to_num(mat/col_sum)
7     PLS = np.sum(mat, axis=0)/np.sum(mat)
8     return pd.Series(PLS, index=lambda_mat.columns)
9
10 def L_index(PLS):
11     gamma = len(PLS)
12     entropy = 0
13     for i in PLS:
14         entropy += i*np.log(gamma*i+0.000000001)
15     L = entropy/np.log(gamma)
16     return L

```

Listing 11: Computation of the PLS and concentration index.

The entire implementation is optimised to produce the PLS results in a very short time without dedicated hardware but merely using a personal computer. As demon-

stration, the data analysed in this implementation includes one day of midquote returns for six series sampled at a 1-second interval between 13:30 and 18:30 UTC, for a total of 108 thousand observations and 246 parameters to estimate for the multivariate Markov chain, and the computation time is 3.39 seconds. The speed is due to the indexing facility from the Pandas package.

7 Conclusion

This thesis examined three applications of Markov chain models to financial issues. Two chapters focused on the problem of stock valuation, extending existing dividend discount models and proposing a new framework to valuation using Markov chains from both univariate and multivariate perspectives. The last application focused on the proposal of a new methodology to measure price discovery to overcome the limits of existing measures.

The review proposed in Chapter 2 illustrates the extensive research that has been undertaken in both topics and highlights the strengths and weaknesses of existing models with a brief formulation of questions left unanswered by academics. With this dissertation we filled these gaps proposing solutions that further advance research in stock valuation and market microstructure.

In Chapter 3 we extended previous results on the Markov chain stock model by computing for the first time ever the so called second-order price-dividend ratio. This provides the analyst with a measure of risk to juxtapose to the price-dividend ratio for measuring the profitability of an investment in a stock. Furthermore we developed non-parametric statistical techniques useful for the estimation of the financial quantities starting from a time series of dividend data. We proposed estimators of the first and second-order price-dividend ratios and we established their asymptotic properties that are fundamental for the computation of interval confidence of the considered financial quantities. Finally, in the application we considered many practical problems the analyst would encounter when applying our model. Namely, the determination of a suitable number of states for the Markov chain, the determina-

tion of the states and of their values, the stability of the results with respect to the choice of the state space and the forecasting of dividend and fundamental value and risk.

In Chapter 4, we presented a dividend valuation model based on a multivariate Markov chain model. Our valuation setting extends the approaches by Gordon and Shapiro (1956), Hurley and Johnson (1994, 1998), Yao (1997), Ghezzi and Piccardi (2003), Barbu et al. (2017) by placing more emphasis on how forecasts are updated, accounting for possible dependencies between the state of the growth-dividend process of each stock. In our model, future prospects are dependent on a vector of states so that forecasts are updated whenever a stock moves from one state to another.

We proposed a linear system of equations for the first and second order price-dividend ratios that are attached to the vector of states. Moreover, we introduced a formula for the computation of the variances and covariances between stocks for portfolio selection and valuation purposes. Finally, we demonstrated the validity of the model with an application to dividend growth series from three US stocks with a long history of dividend payments and correlation between the series. The application shows how to practically implement the model and how our proposed multivariate model performs better than other dividend valuation models.

Chapter 5 introduced a new measure for price discovery using a multivariate Markov chain model. Our measure is based on the observation of the dependencies between price changes of cointegrated price series. It represents a way to capture the dynamics of price change and, therefore, the timing of incorporating new information, i.e. one of the main objectives of price discovery. Moreover, we can understand the price dependencies that exist between the price series. Also, along with the price leadership share, we proposed a concentration index that measures the concentration of price leadership and it is useful to make a comparison between different PLS outcomes.

An empirical application to six gold contracts across the world, such as spot, futures and ETF contracts, confirms results expected from the simple microstructure

analysis, as well as previous literature findings on price discovery. Results show that the price leadership share measure performs well under different conditions and produces reliable results as expected from the microstructure analysis. Moreover, it offers advantages over the existent measures, i.e. information share, component share, and informational leadership share, because it can process more than two price series simultaneously, thus producing a ranking of the price leadership and a model of the price change dependencies between series. Also, the price leadership share measure does not require us to include lagged observations in the model, in contrast to what happens with other measures. Hence, our model can benefit from a bigger sample for the analysis that proves to be a strength in the case of illiquid stocks with a small set of observations.

Our results highlight that the lower the sampling frequency, the better the outcome. Ideally, the optimal solution would be to model the price change directly at event time, i.e. quote time. However, this is not possible with a discrete-time homogeneous Markov chain, but it leaves open possibilities to model the price change dynamics with more complex models, e.g. semi-Markov models. Moreover, future research can consider modelling price changes with a high-order multivariate Markov chain to include some dependency from the past.

A Appendix: proofs

Proof of Proposition 3.3

For $k, j, i \in \mathbb{N}$ with $j > i$ we consider the following expectation:

$$\begin{aligned}
& \mathbb{E}_k \left[\prod_{h=1}^i \prod_{w=1}^j G(k+h)G(k+w) \right] = \mathbb{E}_k \left[\prod_{h=1}^i G^2(k+h) \prod_{w=i+1}^j G(k+w) \right] \\
& = \mathbb{E}_k \left[\prod_{h=1}^i G^2(k+h) \prod_{w=i+1}^{j-1} G(k+w) \mathbb{E}_{k+j-1}[G(k+j)] \right] \\
& \leq \mathbb{E}_k \left[\prod_{h=1}^i G^2(k+h) \prod_{w=i+1}^{j-1} G(k+w) \right] \bar{g},
\end{aligned}$$

where the last inequality follows from (3.9). If we proceed to compute the expectation by conditioning up to time $i+1$ and we use at each step (3.9) we get

$$\begin{aligned}
& \mathbb{E}_k \left[\prod_{h=1}^i \prod_{w=1}^j G(k+h)G(k+w) \right] \leq \mathbb{E}_k \left[\prod_{h=1}^i G^2(k+h) \right] (\bar{g})^{j-i} \\
& = \mathbb{E}_k \left[\prod_{h=1}^{i-1} G^2(k+h) \mathbb{E}_{k+i-1}[G^2(k+i)] \right] (\bar{g})^{j-i} \leq \mathbb{E}_k \left[\prod_{h=1}^{i-1} G^2(k+h) \right] \bar{g}^{(2)} (\bar{g})^{j-i},
\end{aligned} \tag{A.1}$$

where the last inequality follows from (3.11). If we proceed to compute the expectation by conditioning up to time 1 and we use at each step (3.11) we get

$$\mathbb{E}_k \left[\prod_{h=1}^i \prod_{w=1}^j G(k+h)G(k+w) \right] \leq (\bar{g}^{(2)})^i (\bar{g})^{j-i}. \tag{A.2}$$

Therefore

$$p^{(2)}(k) = \sum_{i=1}^{+\infty} \frac{\mathbb{E}_k[D^2(k+i)]}{r^{2i}} + 2 \sum_{i=1}^{+\infty} \sum_{j>i} \frac{\mathbb{E}_k[D(k+i)D(k+j)]}{r^{i+j}}. \quad (\text{A.3})$$

Consequently, we obtain

$$\begin{aligned} p^{(2)}(k) &= \sum_{i=1}^{+\infty} \frac{\mathbb{E}_k[\prod_{j=1}^i G^2(k+j)]}{r^{2i}} d^2(k) \\ &\quad + 2 \sum_{i=1}^{+\infty} \sum_{j>i} \frac{\mathbb{E}_k[\prod_{h=1}^i G(k+h) \prod_{w=1}^j G(k+w)]}{r^{i+j}} d^2(k) \\ &\leq \sum_{i=1}^{+\infty} \frac{(\bar{g}^{(2)})^i}{r^{2i}} d^2(k) + 2 \sum_{i=1}^{+\infty} \sum_{j>i} \frac{(\bar{g}^{(2)})^i (\bar{g})^{j-i}}{r^{i+j}} d^2(k). \end{aligned} \quad (\text{A.4})$$

From (A.4), A1 and A2, using the properties of geometric series, we obtain that $p^{(2)}(k) := \mathbb{E}_k[P^2(k)] < +\infty$.

Since the expected values in Formula (A.3) do depend only on $g(k)$, the second order moment can be expressed in the compact form $p^{(2)}(k) = \psi_2(g(k))d^2(k)$. Let us denote the second order price-dividend ratio by

$$\psi_2(g(k)) := \frac{p^{(2)}(k)}{d^2(k)}. \quad (\text{A.5})$$

Let $\bar{\psi}_2 = \max_i(\psi_2(g_i))$; then $0 \leq \mathbb{E}_k[P^2(k)] \leq \bar{\psi}_2 \mathbb{E}_k[D^2(k+i)]$, which is equivalent to

$$\frac{\mathbb{E}_k[P^2(k+i)]}{r^{2i}} \leq \bar{\psi}_2 \frac{\mathbb{E}_k[D^2(k+i)]}{r^{2i}}. \quad (\text{A.6})$$

Since $p^{(2)}(k) = \mathbb{E}_k[P^2(k)] < +\infty$, we have that $\lim_{i \rightarrow +\infty} \frac{\mathbb{E}_k[D^2(k+i)]}{r^{2i}} = 0$ and hence from (A.6) we get

$$\lim_{i \rightarrow +\infty} \frac{\mathbb{E}_k[P^2(k+i)]}{r^{2i}} = 0. \quad (\text{A.7})$$

It remains to prove that $\lim_{N \rightarrow +\infty} \sum_{i=1}^N \frac{\mathbb{E}_k[D(k+i)P(k+N)]}{r^{i+N}} = 0$. From the Cauchy-

Schwarz inequality we have that

$$\begin{aligned} \lim_{N \rightarrow +\infty} \sum_{i=1}^N \frac{\mathbb{E}_k[D(k+i)P(k+N)]}{r^{i+N}} &\leq \lim_{N \rightarrow +\infty} \sum_{i=1}^N \left(\frac{\mathbb{E}_k[D^2(k+i)]}{r^{i+N}} \right)^{\frac{1}{2}} \left(\frac{\mathbb{E}_k[P^2(k+N)]}{r^{2N}} \right)^{\frac{1}{2}} \\ &= \lim_{N \rightarrow +\infty} \left(\frac{\mathbb{E}_k[P^2(k+N)]}{r^{2N}} \right)^{\frac{1}{2}} \lim_{N \rightarrow +\infty} \sum_{i=1}^N \left(\frac{\mathbb{E}_k[D^2(k+i)]}{r^{i+N}} \right)^{\frac{1}{2}} = 0, \end{aligned}$$

where the last equality holds true because the first factor is zero using (A.7), while the second one is finite as it follows from the finiteness of (A.3).

Proof of Proposition 3.4

Let $k \in \mathbb{N}$ be the current time. At time k the dividend process and the growth dividend process assume two known values denoted by $D(k) = d(k) \in \mathbb{R}$ and $G(k) = g(k) \in E$, respectively. Let us consider first the case when $g(k) = g_1$. By combining Equations (A.5), (3.7) and (3.4) we get

$$\begin{aligned} \psi_2(g_1)d^2(k) &= \mathbb{E}_k \left[\frac{(G(k+1)d(k) + P(k+1))^2}{r^2} \right] \\ &= \mathbb{E}_k \left[\frac{G^2(k+1)d^2(k)}{r^2} \right] + \mathbb{E}_k \left[\frac{P^2(k+1)}{r^2} \right] + \mathbb{E}_k \left[\frac{2G(k+1)d(k)P(k+1)}{r^2} \right]. \end{aligned} \quad (\text{A.8})$$

Now, let us compute these three expectations:

$$\mathbb{E}_k[G^2(k+1)d^2(k)] = d^2(k)(p_{11}g_1^2 + p_{12}g_2^2); \quad (\text{A.9})$$

$$\begin{aligned} \mathbb{E}_k[P^2(k+1)] &= \mathbb{E}_k[\mathbb{E}_{k+1}[P^2(k+1)|G(k+1)]] = \mathbb{E}_k[\psi_2(g(k+1))d^2(k+1)] \\ &= \mathbb{E}_k[\psi_2(g(k+1))G^2(k+1)d^2(k)] = d^2(k)(p_{11}\psi_2(g_1)g_1^2 + p_{12}\psi_2(g_2)g_2^2); \end{aligned} \quad (\text{A.10})$$

$$\begin{aligned}
& \mathbb{E}_k[G(k+1)d(k)P(k+1)] = d(k)\mathbb{E}_k[\mathbb{E}_{k+1}[G(k+1)P(k+1)|G(k+1)]] \\
&= d(k)\mathbb{E}_k[G(k+1)\mathbb{E}_{k+1}[P(k+1)|G(k+1)]] = d(k)\mathbb{E}_k[G(k+1)\psi_1(g(k+1))d(k+1)] \\
&= d(k)\mathbb{E}_k[G(k+1)\psi_1(g(k+1))G(k+1)d(k)] = d^2(k)(p_{11}g_1\psi_1(g_1)g_1 + p_{12}g_2\psi_1(g_2)g_2) \\
&= d^2(k)(p_{11}g_1^2\psi_1(g_1) + p_{12}g_2^2\psi_1(g_2)). \tag{A.11}
\end{aligned}$$

A substitution of (A.9), (A.10) and (A.11) in (A.8) leads to

$$\begin{aligned}
\psi_2(g_1)d^2(k) &= \frac{1}{r^2} \left(d^2(k)(p_{11}g_1^2 + p_{12}g_2^2) \right. \\
&\quad \left. + d^2(k)(p_{11}\psi_2(g_1)g_1^2 + p_{12}\psi_2(g_2)g_2^2) + d^2(k)(p_{11}g_1^2\psi_1(g_1) + p_{12}g_2^2\psi_1(g_2)) \right).
\end{aligned}$$

Some computations yield

$$\psi_2(g_1)(r^2 - p_{11}g_1^2) - \psi_2(g_2)p_{12}g_2^2 = p_{11}g_1^2(1 + 2\psi_1(g_1)) + p_{12}g_2^2(1 + 2\psi_1(g_2)). \tag{A.12}$$

Symmetric arguments produce the second equation of the system (3.13). Concerning the uniqueness of the solution, it is sufficient to note that the matrix of the coefficients of the system has the form $A = \begin{bmatrix} r^2 - p_{11}g_1^2 & -p_{12}g_2^2 \\ -p_{21}g_1^2 & r^2 - p_{22}g_2^2 \end{bmatrix}$ and then

$$\det(A) = (r^2 - p_{11}g_1^2)(r^2 - p_{22}g_2^2) - (-p_{12}g_2^2)(-p_{21}g_1^2).$$

Using assumption A2, we have that

$$\begin{aligned}
\det(A) &> (p_{11}g_1^2 + p_{12}g_2^2 - p_{11}g_1^2)(p_{21}g_1^2 + p_{22}g_2^2 - p_{22}g_2^2) - (-p_{12}g_2^2)(-p_{21}g_1^2) \\
&= p_{12}g_2^2p_{21}g_1^2 - (-p_{12}g_2^2)(-p_{21}g_1^2) = 0.
\end{aligned}$$

The non negativity of the solution is due to the fact that $g_1, g_2 \geq 0$.

Proof of Proposition 3.5

For $n = 1$ it is simple to compute the expectation

$$E^{(1)}p(d_k, g_a) = \sum_{j \in E} p_{aj}p(d_k g_j, g_j) = \sum_{j \in E} p_{aj}g_j p(d_k, g_j). \quad (\text{A.13})$$

Consequently, we have the matrix form

$$\begin{pmatrix} E^{(1)}p(d_k, g_1) \\ \vdots \\ E^{(1)}p(d_k, g_s) \end{pmatrix} = \mathbf{PI}_g \begin{pmatrix} p(d_k, g_1) \\ \vdots \\ p(d_k, g_s) \end{pmatrix},$$

and, taking into account (cf. (3.8)) that $p(d_k, g_a) = p(d(k), g(k)) = d(k)\psi_1(g(k)) = d_k\psi_1(g_a)$, we obtain (3.22) for $n = 1$. For $n = 2$ we have

$$\begin{aligned} E^{(2)}p(d_k, g_a) &= E_{(d(k), g_a)}[P(D(k+2), g(k+2))] \\ &= E_{(d(k), g_a)}[E_{(D(k+1), g(k+1))}[P(D(k+2), g(k+2))] \\ &= E_{(d(k), g_a)}[E^{(1)}P(D(k+1), g(k+1))] \\ &= \sum_{j \in E} p_{aj}E^{(1)}P(g_j d_k, g_j) = \sum_{j \in E} p_{aj} \sum_{h \in E} p_{jh}p(d_k g_j g_h, g_h) \\ &= \sum_{j \in E} \sum_{h \in E} p_{aj}p_{jh}g_j g_h p(d_k, g_h). \end{aligned} \quad (\text{A.14})$$

As we did previously for the case $n = 1$, taking into account that $(\mathbf{PI}_g)^2 = \mathbf{P}^2\mathbf{I}_{g^2}$ we obtain the matrix form given in (3.22) for $n = 2$. An iteration up to the n th step produces the result given in (3.21) and the corresponding matrix form (3.22).

Proof of Lemma 3.8

Note that we have

$$(\mathbf{I}_r^n - \mathbf{P}\mathbf{I}_g^n)^{-1} = \left(\mathbf{I}_r^n \cdot (\mathbf{I} - \mathbf{P}\mathbf{I}_g^n\mathbf{I}_r^{-n}) \right)^{-1} \quad (\text{A.15})$$

$$= \mathbf{I}_r^{-n} \sum_{s \geq 0} \mathbf{P}^s (\mathbf{I}_g^n)^s (\mathbf{I}_r^{-n})^s \quad (\text{A.16})$$

where we used the fact that matrices like \mathbf{I}_r^n or \mathbf{I}_g^n commute with any matrix. Taking into account Lemma 3.7, we have

$$\begin{aligned} \frac{\partial (\mathbf{I}_r^n - \mathbf{P}\mathbf{I}_g^n)^{-1}}{\partial p_{ij}} &= \mathbf{I}_r^{-n} \sum_{l \geq 0} \left(\frac{\partial \mathbf{P}^l}{\partial p_{ij}} \right) (\mathbf{I}_g^n)^l (\mathbf{I}_r^{-n})^l \\ &= \mathbf{I}_r^{-n} \sum_{l \geq 0} \sum_{k=1}^l \mathbf{P}^{k-1} \left(\frac{\partial \mathbf{P}}{\partial p_{ij}} \right) \mathbf{P}^{l-k} (\mathbf{I}_g^n)^l (\mathbf{I}_r^{-n})^l \\ &= \mathbf{I}_r^{-n} \sum_{k \geq 1} \mathbf{P}^{k-1} \left(\frac{\partial \mathbf{P}}{\partial p_{ij}} \right) \sum_{l \geq k} \mathbf{P}^{l-k} (\mathbf{I}_g^n)^l (\mathbf{I}_r^{-n})^l \\ &= \mathbf{I}_r^{-n} \sum_{k \geq 1} \mathbf{P}^{k-1} (\mathbf{I}_g^n)^{k-1} (\mathbf{I}_r^{-n})^{k-1} \left(\frac{\partial \mathbf{P}}{\partial p_{ij}} \right) \sum_{l \geq k} \mathbf{P}^{l-k} (\mathbf{I}_g^n)^{l-k} (\mathbf{I}_r^{-n})^{l-k} (\mathbf{I}_g^n) (\mathbf{I}_r^{-n}) \\ &= \mathbf{I}_g^n (\mathbf{I}_r^n - \mathbf{P}\mathbf{I}_g^n)^{-1} \cdot \frac{\partial \mathbf{P}}{\partial p_{ij}} \cdot (\mathbf{I}_r^n - \mathbf{P}\mathbf{I}_g^n)^{-1}. \end{aligned}$$

Proof of Theorem 3.10

First, note that we have

$$\widehat{\Psi}_1(m) = \Phi_1(\widehat{p}_{ij}(m)), i = 1, \dots, s, j = 1, \dots, s-1). \quad (\text{A.17})$$

Using the strong consistency of $\widehat{p}_{ij}(m)$, as m goes to infinity (cf. Proposition 3.6), and the continuous mapping theorem, we obtain the strong consistency of $\widehat{\Psi}_1(m)$.

Second, taking into account the expressions (3.25) and (A.17) for Ψ_1 and $\widehat{\Psi}_1(m)$,

respectively, together with the asymptotic normality of the vector

$(\sqrt{m}(\widehat{p}_{ij}(m) - p_{ij}))_{i=1,\dots,s,j=1,\dots,s-1}$ (cf. Proposition 3.6) with covariance matrix $\mathbf{\Gamma}$ defined as the restriction of $\widetilde{\mathbf{\Gamma}}$ given in (3.24) to $s(s-1) \times s(s-1)$, we obtain the asymptotic normality stated in (3.31) using the delta method because Ψ_1 is a differentiable function of \mathbf{P} .

The last point we have to deal with is to give an expression of the partial derivative matrix Φ'_1 . Note that we have

$$\Phi'_1 = \begin{pmatrix} \frac{\partial \Phi_1^1}{\partial p_{11}} & \cdots & \frac{\partial \Phi_1^1}{\partial p_{1(s-1)}} & \cdots & \frac{\partial \Phi_1^1}{\partial p_{s1}} & \cdots & \frac{\partial \Phi_1^1}{\partial p_{s(s-1)}} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \frac{\partial \Phi_1^s}{\partial p_{11}} & \cdots & \frac{\partial \Phi_1^s}{\partial p_{1(s-1)}} & \cdots & \frac{\partial \Phi_1^s}{\partial p_{s1}} & \cdots & \frac{\partial \Phi_1^s}{\partial p_{s(s-1)}} \end{pmatrix} \in \mathcal{M}_{s \times s(s-1)}. \quad (\text{A.18})$$

Note that in the computation of the derivative of Φ_1 with respect to its argument $(p_{ij}, i = 1, \dots, s, j = 1, \dots, s-1)$, this argument is ordered in the lexicographic order: $(p_{11}, \dots, p_{1(s-1)}, p_{21}, \dots, p_{2(s-1)}, \dots, p_{s1}, \dots, p_{s(s-1)}) \in \mathbb{R}^{s(s-1)}$.

An arbitrary column of this matrix Φ'_1 corresponding to $(i, j), i = 1, \dots, s, j = 1, \dots, s-1$, is given by

$$\begin{aligned} \frac{\partial \Phi_1}{\partial p_{ij}} &= \frac{\partial}{\partial p_{ij}} ((\mathbf{I}_r - \mathbf{P}\mathbf{I}_g)^{-1} \mathbf{P}\mathbf{g}) \\ &= \frac{\partial (\mathbf{I}_r - \mathbf{P}\mathbf{I}_g)^{-1}}{\partial p_{ij}} \mathbf{P}\mathbf{g} + (\mathbf{I}_r - \mathbf{P}\mathbf{I}_g)^{-1} \frac{\partial \mathbf{P}}{\partial p_{ij}} \mathbf{g}. \end{aligned} \quad (\text{A.19})$$

Using the expression of $\frac{\partial \mathbf{P}}{\partial p_{ij}}$ given in (3.27), together with the computation of $\frac{\partial (\mathbf{I}_r - \mathbf{P}\mathbf{I}_g)^{-1}}{\partial p_{ij}}$ given in Lemma 3.8 (for $n = 1$), allows us to completely compute $\frac{\partial \Phi_1}{\partial p_{ij}}$ and thus Φ'_1 .

Proof of Theorem 3.12

First, note that we have

$$\widehat{\Psi}_2(m) = \Phi_2(\widehat{p}_{ij}(m), i = 1, \dots, s, j = 1, \dots, s-1). \quad (\text{A.20})$$

Using the strong consistency of estimator $\widehat{\Psi}_1(m)$ (cf. Theorem 3.10), the strong consistency of $\widehat{p}_{ij}(m)$, as m goes to infinity (cf. Proposition 3.6), and the continuous mapping theorem, we obtain the strong consistency of $\widehat{\Psi}_2(m)$.

Second, taking into account the expressions (3.34) and (A.20) for Ψ_2 and $\widehat{\Psi}_2(m)$, respectively, together with the asymptotic normality of the vector $(\sqrt{m}(\widehat{p}_{ij}(m) - p_{ij}))_{i=1, \dots, s, j=1, \dots, s-1}$ (cf. Proposition 3.6) with covariance matrix $\mathbf{\Gamma}$ defined as the restriction of $\widetilde{\mathbf{\Gamma}}$ given in (3.24) to $s(s-1) \times s(s-1)$, we obtain the asymptotic normality stated in (3.36) using the delta method because Ψ_2 is a differentiable function of \mathbf{P} being Ψ_1 a differentiable function of \mathbf{P} .

The last point we have to deal with is to give an expression of the partial derivative matrix Φ'_2 . Note that we have

$$\Phi'_2 = \begin{pmatrix} \frac{\partial \Phi_2^1}{\partial p_{11}} & \dots & \frac{\partial \Phi_2^1}{\partial p_{1(s-1)}} & \dots & \frac{\partial \Phi_2^1}{\partial p_{s1}} & \dots & \frac{\partial \Phi_2^1}{\partial p_{s(s-1)}} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \frac{\partial \Phi_2^s}{\partial p_{11}} & \dots & \frac{\partial \Phi_2^s}{\partial p_{1(s-1)}} & \dots & \frac{\partial \Phi_2^s}{\partial p_{s1}} & \dots & \frac{\partial \Phi_2^s}{\partial p_{s(s-1)}} \end{pmatrix} \in \mathcal{M}_{s \times s(s-1)}. \quad (\text{A.21})$$

An arbitrary column of this matrix corresponding to $(i, j), i = 1, \dots, s, j = 1, \dots, s-1$, is given by

$$\begin{aligned} \frac{\partial \Phi_2}{\partial p_{ij}} &= \frac{\partial}{\partial p_{ij}} \left((\mathbf{I}_r^2 - \mathbf{P} \cdot \mathbf{I}_g^2)^{-1} \cdot \mathbf{P} \cdot (\mathbf{g} \diamond \mathbf{g} + 2\Psi_1 \diamond \mathbf{g} \diamond \mathbf{g}) \right) \\ &= \frac{\partial (\mathbf{I}_r^2 - \mathbf{P} \cdot \mathbf{I}_g^2)^{-1}}{\partial p_{ij}} \cdot \mathbf{P} \cdot (\mathbf{g} \diamond \mathbf{g} + 2\Psi_1 \diamond \mathbf{g} \diamond \mathbf{g}) \\ &\quad + (\mathbf{I}_r^2 - \mathbf{P} \cdot \mathbf{I}_g^2)^{-1} \cdot \frac{\partial \mathbf{P}}{\partial p_{ij}} \cdot (\mathbf{g} \diamond \mathbf{g} + 2\Psi_1 \diamond \mathbf{g} \diamond \mathbf{g}) \\ &\quad + (\mathbf{I}_r^2 - \mathbf{P} \cdot \mathbf{I}_g^2)^{-1} \cdot \mathbf{P} \cdot \left(\mathbf{g} \diamond \mathbf{g} + 2 \frac{\partial \Psi_1}{\partial p_{ij}} \diamond \mathbf{g} \diamond \mathbf{g} \right). \end{aligned} \quad (\text{A.22})$$

Using the computation of $\frac{\partial(\mathbf{I}_r^2 - \mathbf{P}\mathbf{I}_g^2)^{-1}}{\partial p_{ij}}$ given in Lemma 3.8 (for $n = 2$), the expression of $\frac{\partial \mathbf{P}}{\partial p_{ij}}$ given in (3.27), together with the computation of $\frac{\partial \Psi_1}{\partial p_{ij}}$ obtained in (A.19), we completely compute the value of $\frac{\partial \Phi_2}{\partial p_{ij}}$ and thus of Φ'_2 .

Proof of Theorem 3.13

First, note that we have

$$\begin{aligned} & \left(\widehat{E}^{(n)}(m)p(d_k, g_1), \dots, \widehat{E}^{(n)}(m)p(d_k, g_s) \right)^\top \\ &= \Theta(\widehat{p}_{ij}(m), i = 1, \dots, s, j = 1, \dots, s-1). \end{aligned} \quad (\text{A.23})$$

Using the strong consistency of $\widehat{p}_{ij}(m)$, as m goes to infinity (cf. Proposition 3.6), and the continuous mapping theorem, we obtain the strong consistency of estimator (3.38).

Second, taking into account the expressions (3.39) and (A.23) for the expected forecast fundamental prices and the corresponding estimator, together with the asymptotic normality of the vector $(\sqrt{m}(\widehat{p}_{ij}(m) - p_{ij}))_{i=1, \dots, s, j=1, \dots, s-1}$ (cf. Proposition 3.6) with covariance matrix $\mathbf{\Gamma}$ defined as the restriction of $\widetilde{\mathbf{\Gamma}}$ given in (3.24) to $s(s-1) \times s(s-1)$, we obtain the asymptotic normality stated in (3.41) using the delta method.

The last point we have to deal with is to give an expression of the partial derivative matrix Θ' . Note that we have

$$\Theta' = \begin{pmatrix} \frac{\partial \Theta^1}{\partial p_{11}} & \cdots & \frac{\partial \Theta^1}{\partial p_{1(s-1)}} & \cdots & \frac{\partial \Theta^1}{\partial p_{s1}} & \cdots & \frac{\partial \Theta^1}{\partial p_{s(s-1)}} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \frac{\partial \Theta^s}{\partial p_{11}} & \cdots & \frac{\partial \Theta^s}{\partial p_{1(s-1)}} & \cdots & \frac{\partial \Theta^s}{\partial p_{s1}} & \cdots & \frac{\partial \Theta^s}{\partial p_{s(s-1)}} \end{pmatrix} \in \mathcal{M}_{s \times s(s-1)}. \quad (\text{A.24})$$

An arbitrary column of this matrix corresponding to (i, j) , $i = 1, \dots, s$, $j = 1, \dots, s-$

1, is given by

$$\begin{aligned}\frac{\partial \Theta}{\partial p_{ij}} &= \frac{\partial}{\partial p_{ij}} (d_k \mathbf{P}^n \mathbf{I}_{\mathbf{g}^n} \Psi_1) \\ &= d_k \frac{\partial \mathbf{P}^n}{\partial p_{ij}} \mathbf{I}_{\mathbf{g}^n} \Psi_1 + d_k \mathbf{P}^n \mathbf{I}_{\mathbf{g}^n} \frac{\partial \Psi_1}{\partial p_{ij}}.\end{aligned}\tag{A.25}$$

Using the computation of $\frac{\partial \mathbf{P}^n}{\partial p_{ij}}$ given in Lemma 3.7, of $\frac{\partial \mathbf{P}}{\partial p_{ij}}$ given in (3.27) and of $\frac{\partial \Psi_1}{\partial p_{ij}}$ obtained in (A.19), we are able to compute the value of $\frac{\partial \Theta}{\partial p_{ij}}$ and thus of Θ' .

Proof of Theorem 4.1

To prove the finiteness of $p^{(\alpha)}(\mathbf{g}(k))$, let us consider first the tower property of conditional expectations

$$\mathbb{E}_{(k)} \left[\prod_{j=1}^i G^{(\alpha)}(k+j) \right] = \mathbb{E}_{(k)} \left[\prod_{j=1}^{i-1} G^{(\alpha)}(k+j) \mathbb{E}_{(k+i-1)} [G^{(\alpha)}(k+i)] \right].$$

Now, let us proceed to bound $\mathbb{E}_{(k+i-1)} [G^{(\alpha)}(k+i)]$ using Assumption 1:

$$\mathbb{E}_{(k+i-1)} [G^{(\alpha)}(k+i)] = \sum_{j=1}^m g_j^{(\alpha)} \cdot \mathbb{P}[G^{(\alpha)}(k+i) = j | F(k+i-1)],$$

where in general $F(s) = \sigma(G^{(1)}(a), G^{(2)}(a), \dots, G^{(\gamma)}(a), a \leq s)$ denotes the σ -algebra generated by the family of random variables $(G^{(1)}(a), G^{(2)}(a), \dots, G^{(\gamma)}(a), a \leq s)$.

Since $\forall n \in \mathbb{N}$

$$\mathbb{P}[G^{(\alpha)}(n+i) = j | (G^{(1)}(n) = \mathbf{e}^{(1)}, \dots, G^{(\gamma)}(n) = \mathbf{e}^{(\gamma)})] = \sum_{\beta=1}^{\gamma} \sum_{h=1}^m e_h^{(\beta)} \cdot \lambda_{\beta, \alpha} \cdot \mathbf{P}_{h,j}^{(\beta, \alpha)},$$

note that by Assumption 1,

$$\mathbb{E}_{(k+i-1)}[G^{(\alpha)}(k+i)] \leq \max_{\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(\gamma)}} \left(\sum_{j=1}^m \sum_{\beta=1}^{\gamma} \sum_{h=1}^m e_h^{(\beta)} \lambda_{\beta, \alpha} \mathbf{P}_{h,j}^{(\beta, \alpha)} g_j \right) = \bar{g}^{(\alpha;1)}.$$

Thus,

$$\mathbb{E}_{(k)} \left[\prod_{j=1}^i G^{(\alpha)}(k+j) \right] \leq \mathbb{E}_{(k)} \left[\prod_{j=1}^{i-1} G^{(\alpha)}(k+j) \right] \cdot \bar{g}^{(\alpha;1)}.$$

An iteration of previous arguments leads to

$$\mathbb{E}_{(k)} \left[\prod_{j=1}^i G^{(\alpha)}(k+j) \right] \leq (\bar{g}^{(\alpha;1)})^i,$$

which in turn gives

$$p^{(\alpha)}(k) \leq \sum_{i=1}^{+\infty} \frac{(g^{(\alpha;1)})^i}{r_{\alpha}^i} d^{(\alpha)}(k) = d^{(\alpha)}(k) \sum_{i=1}^{+\infty} \left(\frac{g^{(\alpha;1)}}{r_{\alpha}} \right)^i, \quad (\text{A.26})$$

which is a convergent geometric series being $\bar{g}^{(\alpha;1)} < r$ by Assumption 1.

As the second step in the proof, we prove the asymptotic condition

$$\lim_{i \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[\mathcal{P}^{(\alpha)}(k+i)]}{r_{\alpha}^i} = 0.$$

Relation (A.26) obviously implies finiteness of the price-dividend ration for the α -stock, i.e.

$$\psi_1^{(\alpha)}(\mathbf{g}(k)) = \frac{p^{(\alpha)}(\mathbf{g}(k))}{d^{(\alpha)}(k)} < +\infty.$$

Consequently denote by $\bar{\psi}_1^{(\alpha)} = \max_{\mathbf{g}(k) \in E^{\gamma}} \left(\psi_1^{(\alpha)}(\mathbf{g}(k)) \right)$, and observe that

$$p^{(\alpha)}(k) \leq \bar{\psi}_1^{(\alpha)} \cdot d^{(\alpha)}(k),$$

or equivalently

$$\mathbb{E}_{(k)}[\mathcal{P}^{(\alpha)}(k+i)] \leq \bar{\psi}_1^{(\alpha)} \cdot \mathbb{E}_{(k)}[D^{(\alpha)}(k+i)].$$

But, since $\sum_{t=1}^{+\infty} \frac{\mathbb{E}_{(k)}[D^{(\alpha)}(k+i)]}{r^t} < +\infty$,

$$\lim_{t \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[D^{(\alpha)}(k+t)]}{r_\alpha^t} = 0.$$

Hence,

$$\lim_{i \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[\mathcal{P}^{(\alpha)}(k+i)]}{r_\alpha^i} \leq \bar{\psi}_1^{(\alpha)} \cdot \lim_{i \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[D^{(\alpha)}(k+i)]}{r_\alpha^i} = 0.$$

Proof of Theorem 4.4

We first show that equation (4.16) holds true.

In formula (4.15) we established that

$$\begin{aligned} p^{(\alpha)}(k) &= \psi_1^{(\alpha)}(\mathbf{g}(k)) \cdot d^{(\alpha)}(k) = \frac{\mathbb{E}_{(k)}[D^{(\alpha)}(k+1) + \mathcal{P}^{(\alpha)}(k+1)]}{r_\alpha} = \\ &= \frac{\mathbb{E}_{(k)}[G^{(\alpha)}(k+1) \cdot d^{(\alpha)}(k) + \psi_1^{(\alpha)}(\mathbf{G}(k+1)) \cdot D^{(\alpha)}(k+1)]}{r_\alpha} = \\ &= \frac{\mathbb{E}_{(k)}[G^{(\alpha)}(k+1) \cdot d^{(\alpha)}(k)]}{r_\alpha} + \frac{\mathbb{E}_{(k)}[\psi_1^{(\alpha)}(\mathbf{G}(k+1)) \cdot D^{(\alpha)}(k+1)]}{r_\alpha} \end{aligned} \quad (\text{A.27})$$

Let us proceed to compute the expected values in Formula (A.27).

$$\begin{aligned} &\mathbb{E}_{(k)}[G^{(\alpha)}(k+1) \cdot d^{(\alpha)}(k)] \\ &= \sum_{j=1}^m g_j^{(\alpha)} \mathbb{P}[G^{(\alpha)}(k+1) = g_j^{(\alpha)} | (G^{(1)}(k) = \mathbf{e}^{(1)}, \dots, G^{(\gamma)}(k) = \mathbf{e}^{(\gamma)})] \cdot d^{(\alpha)}(k) \\ &= \sum_{j=1}^m \sum_{\beta=1}^{\gamma} \sum_{h=1}^m e_h^{(\beta)}(k) \lambda_{\beta,\alpha} P_{h,j}^{(\beta,\alpha)} g_j^{(\alpha)} \end{aligned} \quad (\text{A.28})$$

$$\begin{aligned}
& \mathbb{E}_{(k)} [\psi_1^{(\alpha)}(\mathbf{G}(k+1)) \cdot D^{(\alpha)}(k+1)] \\
&= \mathbb{E}_{(k)} [\psi_1^{(\alpha)}(\mathbf{G}(k+1)) \cdot G^{(\alpha)}(k+1) \cdot d^{(\alpha)}(k)] \\
&= \sum_{j_1, \dots, j_\gamma=1}^m \psi_1^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) \cdot g_{j_\alpha}^{(\alpha)} \cdot d^{(\alpha)}(k) \\
&\cdot \mathbb{P}[G(k+1) = (g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) | (G^{(1)}(k) = \mathbf{e}^{(1)}, \dots, G^{(\gamma)}(k) = \mathbf{e}^{(\gamma)})] \tag{A.29} \\
&= d^{(\alpha)}(k) \cdot \sum_{j_1, \dots, j_\gamma=1}^m \psi_1^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) \cdot g_{j_\alpha}^{(\alpha)} \\
&\cdot \prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w,f} \mathbf{P}_{c,j_f}^{(w,f)}
\end{aligned}$$

A substitution of (A.28) and (A.29) in (A.27) produces equation (4.16).

The existence and uniqueness of the solution of the linear system of equations obtained from (4.16) with $\mathbf{g}(k) \in E^\gamma$ is a consequence of the fact that Theorem 4.1 assumes the convergence of $p^{(\alpha)}(k)$, which together with $d^{(\alpha)}(k) \in \mathbb{R}$ implies that $\psi_1^{(\alpha)}(\mathbf{g}(k))$ should exist being equal to $\frac{p^{(\alpha)}(k)}{d^{(\alpha)}(k)}$, and should be unique because the series expressing $p^{(\alpha)}(k)$ converge to a unique value.

Proof of Theorem 4.7

To prove the finiteness of the second-order of the price process we first observe that from relation $D^{(\alpha)}(k+1) = G^{(\alpha)}(k) \cdot D^{(\alpha)}(k)$ it follow immediately that

$$\begin{aligned}
p_2^{(\alpha)}(\mathbf{G}(k)) &= \sum_{i=1}^{+\infty} \frac{\mathbb{E}_{(k)} [\prod_{j=1}^i (G^{(\alpha)}(k+j))^2] (d^{(\alpha)}(k))^2}{r_\alpha^{2i}} \\
&+ 2 \sum_{i=1}^{+\infty} \sum_{j>i} \frac{\mathbb{E}_{(k)} [\prod_{h=1}^i G^{(\alpha)}(k+h) \prod_{w=1}^j G^{(\alpha)}(k+w)] (d^{(\alpha)}(k))^2}{r_\alpha^{i+j}}.
\end{aligned}$$

Let us consider expected value $\mathbb{E}_{(k)} \left[\prod_{j=1}^i (G^{(\alpha)}(k+j))^2 \right]$ and apply the tower property of conditional expectation to get

$$\mathbb{E}_{(k)} \left[\prod_{j=1}^i (G^{(\alpha)}(k+j))^2 \right] = \mathbb{E}_{(k)} \left[\prod_{j=1}^{i-1} (G^{(\alpha)}(k+j))^2 \mathbb{E}_{(k+i-1)} \left[(G^{(\alpha)}(k+i))^2 \right] \right].$$

The expectation

$$\mathbb{E}_{(k+i-1)} \left[(G^{(\alpha)}(k+i))^2 \right] = \sum_{j_\alpha=1}^m (g_{j_\alpha}^{(\alpha)})^2 \cdot \mathbb{P}(G^{(\alpha)}(k+i) = j_\alpha | F(k+i-1)).$$

Since $\forall n \in \mathbb{N}$

$$\mathbb{P}(G^{(\alpha)}(k+i) = j_\alpha | G^{(1)}(n) = \mathbf{e}^{(1)}, \dots, G^{(\gamma)}(n) = \mathbf{e}^{(\gamma)}) = \sum_{\beta=1}^{\gamma} \sum_{h=1}^m e_h^{(\beta)} \cdot \lambda_{\beta\alpha} \cdot P_{h,j_\alpha}^{(\beta,\alpha)},$$

we have

$$\mathbb{E}_{(k+i-1)} \left[(G^{(\alpha)}(k+i))^2 \right] \leq \max_{\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(\gamma)}} \sum_{j_\alpha=1}^m \sum_{\beta=1}^{\gamma} \sum_{h=1}^m e_h^{(\beta)} \cdot \lambda_{\beta\alpha} \cdot P_{h,j_\alpha}^{(\beta,\alpha)} \cdot (g_{j_\alpha}^{(\alpha)})^2 = \bar{g}^{(\alpha;2)}.$$

Thus

$$\mathbb{E}_{(k)} \left[\prod_{j=1}^i (G^{(\alpha)}(k+j))^2 \right] \leq \mathbb{E}_{(k)} \left[\prod_{j=1}^{i-1} (G^{(\alpha)}(k+j))^2 \right] \bar{g}^{(\alpha;2)},$$

and by iteration

$$\mathbb{E}_{(k)} \left[\prod_{j=1}^i (G^{(\alpha)}(k+j))^2 \right] \leq (\bar{g}^{(\alpha;2)})^i.$$

Similar computations can be executed to prove

$$\mathbb{E}_{(k)} \left[\prod_{h=1}^i G^{(\alpha)}(k+h) \prod_{w=1}^j G^{(\alpha)}(k+w) \right] \leq (\bar{g}^{(\alpha;2)})^i (\bar{g}^{(\alpha;1)})^{j-1}.$$

These bounds can be applied as follows:

$$p_2^{(\alpha)}(k) \leq \sum_{i=1}^{+\infty} \frac{(\bar{g}^{(\alpha;2)})^i}{r_\alpha^{2i}} (d^{(\alpha)}(k))^2 + 2 \sum_{i=1}^{+\infty} \sum_{j>i} \frac{(\bar{g}^{(\alpha;2)})^i (\bar{g}^{(\alpha;1)})^{j-1}}{r_\alpha^{i+j}} (d^{(\alpha)}(k))^2, \quad (\text{A.30})$$

which is a convergent series being $\bar{g}^{(\alpha;1)} < r_\alpha$ and $\bar{g}^{(\alpha;2)} < r_\alpha^2$ by Assumptions 1 and

2, respectively.

To prove the asymptotic conditions (4.17a) and (4.17b), we denote by

$$\frac{p_2^{(\alpha)}(\mathbf{g}(k))}{(d^{(\alpha)}(k))^2} =: \psi_2^{(\alpha)}(\mathbf{g}(k)) < +\infty,$$

being $p_2^{(\alpha)}(\mathbf{g}(k)) < +\infty$.

Set $\psi_2^{(\alpha)} := \max_{\mathbf{G}(k) \in E^\gamma} \psi_2^{(\alpha)}(\mathbf{G}(k))$ and observe that

$$\mathbb{E}_{(k)}[(\mathcal{P}^{(\alpha)}(k+N))^2] \leq \psi_2^{(\alpha)} \mathbb{E}_{(k)}[(D^{(\alpha)}(k+N))^2],$$

implies

$$\frac{\mathbb{E}_{(k)}[(\mathcal{P}^{(\alpha)}(k+N))^2]}{r_\alpha^{2N}} \leq \psi_2^{(\alpha)} \frac{\mathbb{E}_{(k)}[(D^{(\alpha)}(k+N))^2]}{r_\alpha^{2N}}.$$

Accordingly

$$\lim_{N \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[(\mathcal{P}^{(\alpha)}(k+N))^2]}{r_\alpha^{2N}} \leq \psi_2^{(\alpha)} \lim_{N \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[(D^{(\alpha)}(k+N))^2]}{r_\alpha^{2N}}.$$

Now, from equation (A.30) we have

$$\sum_{i=1}^{+\infty} \frac{\mathbb{E}_{(k)}[(D^{(\alpha)}(k+i))^2]}{r_\alpha^{2N}} < +\infty,$$

and consequently

$$\lim_{N \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[(D^{(\alpha)}(k+N))^2]}{r_\alpha^{2N}} = 0. \tag{A.31}$$

It remains to prove the second asymptotic condition, i.e.

$$\lim_{N \rightarrow +\infty} \sum_{i=1}^N \frac{\mathbb{E}_{(k)}[D^{(\alpha)}(k+i)\mathcal{P}^{(\alpha)}(k+N)]}{r_\alpha^{i+N}} = 0.$$

To achieve the result, we apply the Cauchy-Schwartz inequality to get

$$\begin{aligned} & \lim_{N \rightarrow +\infty} \sum_{i=1}^N \frac{\mathbb{E}_{(k)}[D^{(\alpha)}(k+i)\mathcal{P}^{(\alpha)}(k+N)]}{r_\alpha^{i+N}} \\ & \leq \lim_{N \rightarrow +\infty} \left(\frac{\mathbb{E}_{(k)}[(\mathcal{P}^{(\alpha)}(k+N))^2]}{r_\alpha^{2N}} \right)^{\frac{1}{2}} \lim_{N \rightarrow +\infty} \sum_{i=1}^N \left(\frac{\mathbb{E}_{(k)}[(D^{(\alpha)}(k+i))^2]}{r_\alpha^{2i}} \right)^{\frac{1}{2}}. \end{aligned}$$

At this point, it is sufficient to note that

$$\lim_{N \rightarrow +\infty} \left(\frac{\mathbb{E}_{(k)}[(\mathcal{P}^{(\alpha)}(k+N))^2]}{r_\alpha^{2N}} \right)^{\frac{1}{2}} = 0,$$

from equation (A.31) and that due to the finiteness of $p_2^{(\alpha)}(k)$ we have

$$\lim_{N \rightarrow +\infty} \sum_{i=1}^N \left(\frac{\mathbb{E}_{(k)}[(D^{(\alpha)}(k+i))^2]}{r_\alpha^{2i}} \right)^{\frac{1}{2}} < +\infty.$$

Proof of Theorem 4.10

First of all, we establish the validity of equation (4.18). As we have established that

$$p_2^{(\alpha)}(k) = \psi_2^{(\alpha)}(\mathbf{g}(k))(d^{(\alpha)}(k))^2 = \frac{\mathbb{E}_{(\gamma)}[(G^{(\alpha)}(k+1)d^{(\alpha)}(k) + \mathcal{P}^{(\alpha)}(k+1))^2]}{r_\alpha^2},$$

by developing the square we obtain three expectations that need to be evaluated, i.e.

$$\begin{aligned} p_2^{(\alpha)}(k) &= \frac{1}{r_\alpha^2} \mathbb{E}_{(\gamma)}[(G^{(\alpha)}(k+1)d^{(\alpha)}(k))^2] + \frac{1}{r_\alpha^2} \mathbb{E}_{(\gamma)}[(\mathcal{P}^{(\alpha)}(k+1))^2] \\ &+ \frac{2}{r_\alpha^2} \mathbb{E}_{(\gamma)}[G^{(\alpha)}(k+1)d^{(\alpha)}(k)\mathcal{P}^{(\alpha)}(k+1)]. \end{aligned} \quad (\text{A.32})$$

According to formula (A.28), we have

$$\frac{1}{r_\alpha^2} \mathbb{E}_{(\gamma)}[(G^{(\alpha)}(k+1)d^{(\alpha)}(k))^2] = \frac{1}{r_\alpha^2} \sum_{j_\alpha=1}^m \sum_{\beta=1}^{\gamma} \sum_{h=1}^m e_h^{(\beta)}(k) \lambda_{\beta,\alpha} P_{h,j_\alpha}^{(\beta,\alpha)} (g_{j_\alpha}^{(\alpha)})^2 (d^{(\alpha)}(k))^2 \quad (\text{A.33})$$

$$\begin{aligned}
\frac{1}{r_\alpha^2} \mathbb{E}_{(\gamma)} [(\mathcal{P}^{(\alpha)}(k+1))^2] &= \frac{1}{r_\alpha^2} \mathbb{E}_{(\gamma)} [\psi_2^{(\alpha)}(\mathbf{g}(k))(D^{(\alpha)}(k+1))^2] \\
&= \frac{1}{r_\alpha^2} \mathbb{E}_{(\gamma)} [\psi_2^{(\alpha)}(\mathbf{g}(k))(G^{(\alpha)}(k+1)d^{(\alpha)}(k))^2] \\
&= \frac{1}{r_\alpha^2} \sum_{j_1, \dots, j_\gamma=1}^m \psi_2^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)})(g_{j_\alpha}^{(\alpha)})^2 \\
&\quad \cdot \mathbb{P} \left[\mathbf{G}^{(\alpha)}(k+1) = (g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) \mid G^{(1)}(k) = \mathbf{e}^{(1)}, \dots, G^{(\gamma)}(k) = \mathbf{e}^{(\gamma)} \right] (d^{(\alpha)}(k))^2 \\
&= \frac{(d^{(\alpha)}(k))^2}{r_\alpha^2} \sum_{j_1, \dots, j_\gamma=1}^m \psi_2^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)})(g_{j_\alpha}^{(\alpha)})^2 \left(\prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w,f} \mathbf{P}_{c,j_f}^{(w,f)} \right).
\end{aligned} \tag{A.34}$$

$$\begin{aligned}
\frac{2}{r_\alpha^2} \mathbb{E}_{(\gamma)} [G^{(\alpha)}(k+1)d^{(\alpha)}(k)\mathcal{P}^{(\alpha)}(k+1)] \\
&= \frac{2}{r_\alpha^2} \mathbb{E}_{(\gamma)} [G^{(\alpha)}(k+1)d^{(\alpha)}(k)\psi_1^{(\alpha)}(\mathbf{g}(k))D^{(\alpha)}(k+1)] \\
&= \frac{2}{r_\alpha^2} \mathbb{E}_{(\gamma)} [G^{(\alpha)}(k+1)d^{(\alpha)}(k)\psi_1^{(\alpha)}(\mathbf{g}(k))G^{(\alpha)}(k+1)d^{(\alpha)}(k)] \\
&= \frac{2(d^{(\alpha)}(k))^2}{r_\alpha^2} \mathbb{E}_{(\gamma)} \left[(G^{(\alpha)}(k+1))^2 \psi_1^{(\alpha)}(\mathbf{g}(k)) \right] \\
&= \frac{2(d^{(\alpha)}(k))^2}{r_\alpha^2} \sum_{j_1, \dots, j_\gamma=1}^m \psi_1^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)})(g_{j_\alpha}^{(\alpha)})^2 \\
&\quad \cdot \mathbb{P} \left[\mathbf{G}^{(\alpha)}(k+1) = (g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) \mid G^{(1)}(k) = \mathbf{e}^{(1)}, \dots, G^{(\gamma)}(k) = \mathbf{e}^{(\gamma)} \right] \\
&= \frac{2(d^{(\alpha)}(k))^2}{r_\alpha^2} \sum_{j_1, \dots, j_\gamma=1}^m \psi_1^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)})(g_{j_\alpha}^{(\alpha)})^2 \left(\prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w,f} \mathbf{P}_{c,j_f}^{(w,f)} \right).
\end{aligned} \tag{A.35}$$

A substitution of (A.33), (A.34), and (A.35) into (A.32) leads to

$$\begin{aligned}
\psi_2^{(\alpha)}(\mathbf{g}(k))(d^{(\alpha)}(k))^2 &= \frac{2(d^{(\alpha)}(k))^2}{r_\alpha^2} \sum_{j_\alpha=1}^m \sum_{\beta=1}^{\gamma} \sum_{h=1}^m e_h^{(\beta)}(k) \lambda_{\beta,\alpha} P_{h,j_\alpha}^{(\beta,\alpha)} (g_{j_\alpha}^{(\alpha)})^2 \\
&+ \frac{(d^{(\alpha)}(k))^2}{r_\alpha^2} \sum_{j_1, \dots, j_\gamma=1}^m \psi_2^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)})(g_{j_\alpha}^{(\alpha)})^2 \left(\prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w,f} \mathbf{P}_{c,j_f}^{(w,f)} \right) \\
&+ \frac{2(d^{(\alpha)}(k))^2}{r_\alpha^2} \sum_{j_1, \dots, j_\gamma=1}^m \psi_1^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)})(g_{j_\alpha}^{(\alpha)})^2 \left(\prod_{f=1}^{\gamma} \sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{w,f} \mathbf{P}_{c,j_f}^{(w,f)} \right).
\end{aligned}$$

Some algebraic computations yield equation (4.18).

The existence and uniqueness of the solution of the linear system of equations obtained from equation (4.18) with $\mathbf{g}(k) \in E^\gamma$ is a consequence of Theorem 4.7, which

affirms that $p_2^{(\alpha)}(k) < +\infty$ and since $(d^{(\alpha)}(k))^2 \in \mathbb{R}$ and $\psi_2^{(\alpha)}(\mathbf{g}(k)) = \frac{p_2^{(\alpha)}(k)}{(d^{(\alpha)}(k))^2}$, also $\psi_2^{(\alpha)}(\mathbf{g}(k))$ should be unique $\forall \mathbf{g}(k) \in E^\gamma$ and exists as it is the ratio of two finite quantities.

Proof of Theorem 4.11

To prove the finiteness of $p_2^{(\alpha,\beta)}(\mathbf{g}(k))$ we first represent it in terms of the growth-dividend process of α and β stocks. To do this, we observe that

$$D^{(\alpha)}(k+i) = \prod_{j=1}^i G^{(\alpha)}(k+j) \cdot d^{(\alpha)}(k) \quad (\text{A.36})$$

$$D^{(\beta)}(k+i) = \prod_{j=1}^i G^{(\beta)}(k+j) \cdot d^{(\beta)}(k). \quad (\text{A.37})$$

A substitution of equations (A.36) and (A.37) in (4.10) gives

$$\begin{aligned} p_2^{(\alpha,\beta)}(\mathbf{g}(k)) &= \sum_{i=1}^{+\infty} \frac{\mathbb{E}_{(k)} \left[\prod_{j=1}^i G^{(\alpha)}(k+j) G^{(\beta)}(k+j) \right] d^{(\alpha)}(k) d^{(\beta)}(k)}{(r_\alpha r_\beta)^i} \\ &+ \sum_{i=1}^{+\infty} \sum_{j>i} \frac{\mathbb{E}_{(k)} \left[\prod_{h=1}^i G^{(\alpha)}(k+h) G^{(\beta)}(k+h) \prod_{w=i+1}^j G^{(\beta)}(k+w) \right] d^{(\alpha)}(k) d^{(\beta)}(k)}{r_\alpha^i r_\beta^j} \\ &+ \sum_{i=1}^{+\infty} \sum_{j>i} \frac{\mathbb{E}_{(k)} \left[\prod_{h=1}^i G^{(\alpha)}(k+h) G^{(\beta)}(k+h) \prod_{w=i+1}^j G^{(\alpha)}(k+w) \right] d^{(\alpha)}(k) d^{(\beta)}(k)}{r_\beta^i r_\alpha^j}. \end{aligned} \quad (\text{A.38})$$

Now, let us proceed to bound the three expectations in Formula A.38.

Let us start from $\mathbb{E}_{(k)} \left[\prod_{j=1}^i G^{(\alpha)}(k+j) G^{(\beta)}(k+j) \right]$.

An application of the tower property of conditional expectations gives

$$\begin{aligned}
& \mathbb{E}_{(k)} \left[\prod_{j=1}^i G^{(\alpha)}(k+j) G^{(\beta)}(k+j) \right] \\
&= \mathbb{E}_{(k)} \left[\prod_{j=1}^{i-1} G^{(\alpha)}(k+j) G^{(\beta)}(k+j) \mathbb{E}_{(k+i-1)} \left[G^{(\alpha)}(k+j) G^{(\beta)}(k+j) \right] \right] \\
&= \mathbb{E}_{(k)} \left[\prod_{j=1}^{i-1} G^{(\alpha)}(k+j) G^{(\beta)}(k+j) \sum_{j_\alpha=1}^m \sum_{j_\beta=1}^m g_{j_\alpha}^{(\alpha)} g_{j_\beta}^{(\beta)} \mathbb{P} \left(G^{(\alpha)}(k+i) = j_\alpha, \right. \right. \\
&\quad \left. \left. G^{(\beta)}(k+i) = j_\beta \mid F(k+i-1) \right) \right].
\end{aligned}$$

Since $\forall n \in \mathbb{N}$

$$\begin{aligned}
& \mathbb{P} \left(G^{(\alpha)}(n+1) = j_\alpha, G^{(\beta)}(n+1) = j_\beta \mid G^{(1)}(n) = \mathbf{e}^{(1)}, \dots, G^{(\gamma)}(n) = \mathbf{e}^{(\gamma)} \right) \\
&= \left(\sum_{z=1}^{\gamma} \sum_{h=1}^m e_h^{(z)} \cdot \lambda_{z\alpha} \cdot P_{h,j_\alpha}^{(z,\alpha)} \right) \left(\sum_{z=1}^{\gamma} \sum_{h=1}^m e_h^{(z)} \cdot \lambda_{z\beta} \cdot P_{h,j_\beta}^{(z,\beta)} \right),
\end{aligned}$$

it results that

$$\begin{aligned}
\mathbb{E}_{(k)} \left[G^{(\alpha)}(k+i) G^{(\beta)}(k+i) \right] &\leq \left(\max_{\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(\gamma)}} \sum_{h=1}^m \sum_{j_\alpha=1}^{\gamma} \sum_{h=1}^m e_h^{(z)} \cdot \lambda_{z\alpha} \cdot P_{h,j_\alpha}^{(z,\alpha)} \cdot g_{j_\alpha}^{(\alpha)} \right) \\
&\cdot \left(\max_{\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(\gamma)}} \sum_{h=1}^m \sum_{j_\beta=1}^{\gamma} \sum_{h=1}^m e_h^{(z)} \cdot \lambda_{z\beta} \cdot P_{h,j_\beta}^{(z,\beta)} \cdot g_{j_\beta}^{(\beta)} \right) = \bar{g}^{(\alpha;1)} \cdot \bar{g}^{(\beta;1)}.
\end{aligned}$$

In this way we obtained the following bound

$$\mathbb{E}_{(k)} \left[\prod_{j=1}^i G^{(\alpha)}(k+j) G^{(\beta)}(k+j) \right] \leq \mathbb{E}_{(k)} \left[\prod_{j=1}^{i-1} G^{(\alpha)}(k+j) G^{(\beta)}(k+j) \right] \cdot \bar{g}^{(\alpha;1)} \cdot \bar{g}^{(\beta;1)}.$$

An iteration of previous arguments leads to

$$\mathbb{E}_{(k)} \left[\prod_{j=1}^i G^{(\alpha)}(k+j) G^{(\beta)}(k+j) \right] \leq (\bar{g}^{(\alpha;1)})^i \cdot (\bar{g}^{(\beta;1)})^i. \quad (\text{A.39})$$

Consequently, the first addendum on the right hand side of Equation (A.39)

converges because

$$\begin{aligned} & \sum_{i=1}^{+\infty} \frac{\mathbb{E}_{(k)} \left[\prod_{j=1}^i G^{(\alpha)}(k+j) G^{(\beta)}(k+j) \right] d^{(\alpha)}(k) d^{(\beta)}(k)}{(r_{\alpha} r_{\beta})^i} \\ & \leq \sum_{i=1}^{+\infty} \frac{(\bar{g}^{(\alpha;1)})^i \cdot (\bar{g}^{(\beta;1)})^i \cdot d^{(\alpha)}(k) d^{(\beta)}(k)}{(r_{\alpha} r_{\beta})^i} < +\infty \end{aligned}$$

due to Assumption 1 for both stocks α and β .

As second step of the proof, we proceed to prove the convergence of the second addendum of the r.h.s. of equation (A.38). To this end, we have

$$\begin{aligned} & \mathbb{E}_{(k)} \left[\prod_{h=1}^i G^{(\alpha)}(k+h) G^{(\beta)}(k+h) \prod_{w=i+1}^j G^{(\beta)}(k+w) \right] \\ & = \mathbb{E}_{(k)} \left[\prod_{h=1}^i G^{(\alpha)}(k+h) G^{(\beta)}(k+h) \prod_{w=i+1}^{j-1} G^{(\beta)}(k+w) \mathbb{E}_{(k+j-1)} [G^{(\beta)}(k+j)] \right]. \end{aligned}$$

Within proof of Theorem 4.1 we obtained that for a given stock it results

$$\mathbb{E}_{(k+j-1)} [G^{(\beta)}(k+j)] \leq \bar{g}^{(\beta;1)}$$

and by iteration, we get

$$\begin{aligned} & \mathbb{E}_{(k)} \left[\prod_{h=1}^i G^{(\alpha)}(k+h) G^{(\beta)}(k+h) \prod_{w=i+1}^j G^{(\beta)}(k+w) \right] \\ & \leq \mathbb{E}_{(k)} \left[\prod_{h=1}^i G^{(\alpha)}(k+h) G^{(\beta)}(k+h) \right] (\bar{g}^{(\beta;1)})^{j-i} \\ & \leq (\bar{g}^{(\alpha;1)})^i \cdot (\bar{g}^{(\beta;1)})^i \cdot (\bar{g}^{(\beta;1)})^{j-i} = (\bar{g}^{(\alpha;1)})^i \cdot (\bar{g}^{(\beta;1)})^j. \end{aligned}$$

Consequently, for both the stocks α and β , we get

$$\begin{aligned}
& \sum_{i=1}^{+\infty} \sum_{j>i} \frac{\mathbb{E}_{(k)} \left[\prod_{h=1}^i G^{(\alpha)}(k+h) G^{(\beta)}(k+h) \prod_{w=i+1}^j G^{(\beta)}(k+w) \right] d^{(\alpha)}(k) d^{(\beta)}(k)}{r_{\alpha}^i r_{\beta}^j} \\
& \leq \sum_{i=1}^{+\infty} \sum_{j>i} \frac{(\bar{g}^{(\alpha;1)})^i (\bar{g}^{(\beta;1)})^j d^{(\alpha)}(k) d^{(\beta)}(k)}{r_{\alpha}^i r_{\beta}^j} \\
& \leq \sum_{i=1}^{+\infty} \left(\frac{\bar{g}^{(\alpha;1)}}{r_{\alpha}} \right)^i \sum_{j=1}^{+\infty} \left(\frac{\bar{g}^{(\beta;1)}}{r_{\beta}} \right)^j d^{(\alpha)}(k) d^{(\beta)}(k),
\end{aligned} \tag{A.40}$$

where the last inequality in (A.40) is due to the non-negativity of $\bar{g}^{(\alpha;1)}$ and $\bar{g}^{(\beta;1)}$. It is simple now to realise that (A.40) converges due to properties of geometric series and Assumption 1 for both the stocks.

The third addendum of the r.h.s. of equation (A.38) can be proved to be convergent using similar computations as those that demonstrated the convergence of the second term, since they are symmetrical with respect to the stocks α and β .

It remains to prove the validity of the asymptotic conditions (4.19a), (4.19b), and (4.19c).

From Formula (A.38), we have that

$$p_2^{(\alpha,\beta)}(\mathbf{g}(k)) = \psi_2^{(\alpha,\beta)}(\mathbf{g}(k)) d^{(\alpha)}(k) d^{(\beta)}(k) < +\infty$$

where

$$\begin{aligned}
\psi_2^{(\alpha,\beta)}(\mathbf{g}(k)) &= \sum_{i=1}^{+\infty} \frac{\mathbb{E}_{(k)} \left[\prod_{j=1}^i G^{(\alpha)}(k+j) G^{(\beta)}(k+j) \right]}{(r_{\alpha} r_{\beta})^i} \\
&+ \sum_{i=1}^{+\infty} \sum_{j>i} \frac{\mathbb{E}_{(k)} \left[\prod_{h=1}^i G^{(\alpha)}(k+h) G^{(\beta)}(k+h) \prod_{w=i+1}^j G^{(\beta)}(k+w) \right]}{r_{\alpha}^i r_{\beta}^j} \\
&+ \sum_{i=1}^{+\infty} \sum_{j>i} \frac{\mathbb{E}_{(k)} \left[\prod_{h=1}^i G^{(\alpha)}(k+h) G^{(\beta)}(k+h) \prod_{w=i+1}^j G^{(\alpha)}(k+w) \right]}{r_{\beta}^i r_{\alpha}^j}.
\end{aligned}$$

Let denote by

$$\psi_2^{(\alpha,\beta)} := \max_{\mathbf{g}(k) \in \mathbb{E}^\gamma} \left(\psi_2^{(\alpha,\beta)}(\mathbf{g}(k)) \right),$$

and observe that

$$0 \leq \frac{\mathbb{E}_{(k)}[\mathcal{P}^{(\alpha)}(k+N)\mathcal{P}^{(\beta)}(k+N)]}{(r_\alpha r_\beta)^N} \leq \psi_2^{(\alpha,\beta)} \frac{\mathbb{E}_{(k)}[D^{(\alpha)}(k+N)D^{(\beta)}(k+N)]}{(r_\alpha r_\beta)^N},$$

from which we get

$$\lim_{N \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[\mathcal{P}^{(\alpha)}(k+N)\mathcal{P}^{(\beta)}(k+N)]}{(r_\alpha r_\beta)^N} \leq \psi_2^{(\alpha,\beta)} \lim_{N \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[D^{(\alpha)}(k+N)D^{(\beta)}(k+N)]}{(r_\alpha r_\beta)^N}. \quad (\text{A.41})$$

Since $\mathbb{E}_{(k)}\left[\prod_{j=1}^N G^{(\alpha)}(k+j)G^{(\beta)}(k+j)\right] \leq (\bar{g}^{(\alpha;1)})^N \cdot (\bar{g}^{(\beta;1)})^N$, then

$$\begin{aligned} & \frac{\mathbb{E}_{(k)}[D^{(\alpha)}(k+N)D^{(\beta)}(k+N)]}{d^{(\alpha)}(k)d^{(\beta)}(k)} \\ &= \mathbb{E}_{(k)}\left[\prod_{j=1}^N G^{(\alpha)}(k+j)G^{(\beta)}(k+j)\right] \leq (\bar{g}^{(\alpha;1)})^N \cdot (\bar{g}^{(\beta;1)})^N. \end{aligned}$$

But from Assumption 1, we know that

$$\begin{aligned} & \sum_{N=1}^{+\infty} \frac{(\bar{g}^{(\alpha;1)})^N \cdot (\bar{g}^{(\beta;1)})^N}{(r_\alpha r_\beta)^N} < +\infty \\ \implies & \sum_{N=1}^{+\infty} \frac{\mathbb{E}_{(k)}[D^{(\alpha)}(k+N)D^{(\beta)}(k+N)]}{(r_\alpha r_\beta)^N} \cdot \frac{1}{d^{(\alpha)}(k)d^{(\beta)}(k)} < +\infty \end{aligned}$$

which implies that

$$\lim_{N \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[D^{(\alpha)}(k+N)D^{(\beta)}(k+N)]}{(r_\alpha r_\beta)^N} = 0,$$

and then

$$\lim_{N \rightarrow +\infty} \frac{\mathbb{E}_{(k)}[\mathcal{P}^{(\alpha)}(k+N)\mathcal{P}^{(\beta)}(k+N)]}{(r_\alpha r_\beta)^N} = 0.$$

It remains to prove that

$$\lim_{N \rightarrow +\infty} \sum_{i=1}^N \frac{\mathbb{E}_{(k)} [D^{(\alpha)}(k+i) \mathcal{P}^{(\beta)}(k+N)]}{r_\alpha^i r_\beta^N} = \lim_{N \rightarrow +\infty} \sum_{i=1}^N \frac{\mathbb{E}_{(k)} [D^{(\beta)}(k+i) \mathcal{P}^{(\alpha)}(k+N)]}{r_\beta^i r_\alpha^N} = 0.$$

To this end, we proceed to verify only the first limit because the second one is symmetrical with respect to stocks α and β .

An application of the Cauchy-Schwartz inequality gives

$$\begin{aligned} & \lim_{N \rightarrow +\infty} \sum_{i=1}^N \mathbb{E}_{(k)} \left[\frac{D^{(\alpha)}(k+i) \mathcal{P}^{(\beta)}(k+N)}{r_\alpha^i r_\beta^N} \right] \\ & \leq \lim_{N \rightarrow +\infty} \sum_{i=1}^N \left(\frac{\mathbb{E}_{(k)} [(D^{(\alpha)}(k+i))^2]}{r_\alpha^{2i}} \right)^{\frac{1}{2}} \left(\frac{\mathbb{E}_{(k)} [(\mathcal{P}^{(\beta)}(k+N))^2]}{r_\beta^{2N}} \right)^{\frac{1}{2}} \\ & = \lim_{N \rightarrow +\infty} \sum_{i=1}^N \left(\frac{\mathbb{E}_{(k)} [(D^{(\alpha)}(k+i))^2]}{r_\alpha^{2i}} \right)^{\frac{1}{2}} \lim_{N \rightarrow +\infty} \left(\frac{\mathbb{E}_{(k)} [(\mathcal{P}^{(\beta)}(k+N))^2]}{r_\beta^{2N}} \right)^{\frac{1}{2}}. \end{aligned}$$

Now, it is sufficient to remark that

$$\lim_{N \rightarrow +\infty} \left(\frac{\mathbb{E}_{(k)} [(\mathcal{P}^{(\beta)}(k+N))^2]}{r_\beta^{2N}} \right)^{\frac{1}{2}} = 0$$

directly from Theorem 4.7 and that, again from Theorem 4.7, we know that $p_2^{(\alpha, \beta)}(\mathbf{g}(k)) < +\infty$, which in turn implies that

$$\lim_{N \rightarrow +\infty} \sum_{i=1}^N \left(\frac{\mathbb{E}_{(k)} [(D^{(\alpha)}(k+i))^2]}{r_\alpha^{2i}} \right)^{\frac{1}{2}} < +\infty.$$

Proof of Theorem 4.14

We start by establishing the validity of equation (4.21). In formula (4.20) we get

$$\psi_2^{(\alpha; \beta)}(\mathbf{g}(k)) \cdot d^{(\alpha)}(k) \cdot d^{(\beta)}(k) = p_2^{(\alpha; \beta)}(\mathbf{g}(k)).$$

Using the definition of $p_2^{(\alpha;\beta)}(\mathbf{g}(k))$, we have

$$\begin{aligned}
\psi_2^{(\alpha;\beta)}(\mathbf{g}(k)) \cdot d^{(\alpha)}(k) \cdot d^{(\beta)}(k) &= \mathbb{E}_{(k)} \left[\left(\frac{D^{(\alpha)}(k+1) + \mathcal{P}^{(\alpha)}(k+1)}{r_\alpha} \right) \left(\frac{D^{(\beta)}(k+1) + \mathcal{P}^{(\beta)}(k+1)}{r_\beta} \right) \right] \\
&= \mathbb{E}_{(k)} \left[\left(\frac{G^{(\alpha)}(k+1)d^{(\alpha)}(k) + \mathcal{P}^{(\alpha)}(k+1)}{r_\alpha} \right) \left(\frac{G^{(\beta)}(k+1)d^{(\beta)}(k) + \mathcal{P}^{(\beta)}(k+1)}{r_\beta} \right) \right] \\
&= \frac{1}{r_\alpha r_\beta} \left(\mathbb{E}_{(k)} [G^{(\alpha)}(k+1)G^{(\beta)}(k+1)d^{(\alpha)}(k)d^{(\beta)}(k)] + \mathbb{E}_{(k)} [G^{(\alpha)}(k+1)\mathcal{P}^{(\beta)}(k+1)d^{(\alpha)}(k)] \right. \\
&\quad \left. + \mathbb{E}_{(k)} [\mathcal{P}^{(\alpha)}(k+1)G^{(\beta)}(k+1)d^{(\beta)}(k)] + \mathbb{E}_{(k)} [\mathcal{P}^{(\alpha)}(k+1)\mathcal{P}^{(\beta)}(k+1)] \right). \quad (\text{A.42})
\end{aligned}$$

Now, let us compute these for expectations in formula (A.42).

$$\begin{aligned}
&\mathbb{E}_{(k)} [G^{(\alpha)}(k+1)G^{(\beta)}(k+1)d^{(\alpha)}(k)d^{(\beta)}(k)] \\
&= \sum_{j_\alpha=1}^m \sum_{j_\beta=1}^m g_{j_\alpha}^{(\alpha)} g_{j_\beta}^{(\beta)} \mathbb{P}[G^{(\alpha)}(k+1) = j_\alpha, G^{(\beta)}(k+1) = j_\beta | G^{(1)}(k) = \mathbf{e}^{(1)}, \dots, G^{(\gamma)}(k) = \mathbf{e}^{(\gamma)}] \\
&= \sum_{j_\alpha=1}^m \sum_{j_\beta=1}^m g_{j_\alpha}^{(\alpha)} g_{j_\beta}^{(\beta)} \prod_{f \in \{\alpha, \beta\}} \left(\sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{wf} P_{c,j_f}^{(w,f)} \right) d^{(\alpha)}(k) d^{(\beta)}(k).
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}_{(k)} [G^{(\alpha)}(k+1)\mathcal{P}^{(\beta)}(k+1)d^{(\alpha)}(k)] \\
&= \mathbb{E}_{(k)} [G^{(\alpha)}(k+1)\psi_1^{(\beta)}(\mathbf{G}(k+1))D^{(\beta)}(k+1)]d^{(\alpha)}(k) \\
&= \mathbb{E}_{(k)} [G^{(\alpha)}(k+1)\psi_1^{(\beta)}(\mathbf{G}(k+1))G^{(\beta)}(k+1)]d^{(\alpha)}(k)d^{(\beta)}(k) \\
&= \sum_{j_1, \dots, j_\gamma=1}^m g_{j_\alpha}^{(\alpha)} g_{j_\beta}^{(\beta)} \psi_1^{(\beta)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) \cdot \\
&\quad \mathbb{P}[\mathbf{G}(k+1) = (g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) | G^{(1)}(k) = \mathbf{e}^{(1)}, \dots, G^{(\gamma)}(k) = \mathbf{e}^{(\gamma)}] d^{(\alpha)}(k) d^{(\beta)}(k) \\
&= \sum_{j_1, \dots, j_\gamma=1}^m g_{j_\alpha}^{(\alpha)} g_{j_\beta}^{(\beta)} \psi_1^{(\beta)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) \prod_{f=1}^{\gamma} \left(\sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{wf} P_{c,j_f}^{(w,f)} \right) d^{(\alpha)}(k) d^{(\beta)}(k).
\end{aligned}$$

The same calculation can be applied to obtain

$$\begin{aligned} & \mathbb{E}_{(k)} [\mathcal{P}^{(\alpha)}(k+1)G^{(\beta)}(k+1)d^{(\beta)}(k)] \\ &= \sum_{j_1, \dots, j_\gamma=1}^m g_{j_\alpha}^{(\alpha)} g_{j_\beta}^{(\beta)} \psi_1^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) \prod_{f=1}^{\gamma} \left(\sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{wf} P_{c,j_f}^{(w,f)} \right) d^{(\alpha)}(k) d^{(\beta)}(k). \end{aligned}$$

The fourth term of equation (A.42) can be evaluated as follows:

$$\begin{aligned} & \mathbb{E}_{(k)} [\mathcal{P}^{(\alpha)}(k+1)\mathcal{P}^{(\beta)}(k+1)] \\ &= \mathbb{E}_{(k)} [\psi_1^{(\alpha)}(\mathbf{G}(k+1)D^{(\alpha)}(k+1)\psi_1^{(\beta)}(\mathbf{G}(k+1)D^{(\beta)}(k+1))] \\ &= \mathbb{E}_{(k)} [\psi_1^{(\alpha)}(\mathbf{G}(k+1)G^{(\alpha)}(k+1)d^{(\alpha)}(k)\psi_1^{(\beta)}(\mathbf{G}(k+1)G^{(\beta)}(k+1)d^{(\beta)}(k))] \\ &= \mathbb{E}_{(k)} [\psi_1^{(\alpha)}(\mathbf{G}(k+1)\psi_1^{(\beta)}(\mathbf{G}(k+1)G^{(\alpha)}(k+1)G^{(\beta)}(k+1))]d^{(\alpha)}(k)d^{(\beta)}(k) \\ &= \sum_{j_1, \dots, j_\gamma=1}^m g_{j_\alpha}^{(\alpha)} g_{j_\beta}^{(\beta)} \psi_1^{(\alpha)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) \psi_1^{(\beta)}(g_{j_1}^{(1)}, \dots, g_{j_\gamma}^{(\gamma)}) d^{(\alpha)}(k) d^{(\beta)}(k) \cdot \\ & \quad \prod_{f=1}^{\gamma} \left(\sum_{w=1}^{\gamma} \sum_{c=1}^m e_c^{(w)}(k) \lambda_{wf} P_{c,j_f}^{(w,f)} \right). \end{aligned}$$

A substitution of these four expectations into formula (A.42) and little algebraic operations give equation (4.21).

Bibliography

- Agosto, A. and Moretto, E., 2015, 'Variance matters (in stochastic dividend discount models)', *Annals of Finance*, vol. 11, no. 2, pp. 283–295.
- Agosto, A., Mainini, A., and Moretto, E., 2018, 'Stochastic dividend discount model: Covariance of random stock prices', *Journal of Economics and Finance*. Online first article available at <https://doi.org/10.1007/s12197-018-9455-9>.
- Anderson, T.W. and Goodman, L.A., 1957, 'Statistical inference about Markov chains', *The Annals of Mathematical Statistics*, vol. 28, no. 1, pp. 89–110.
- Baillie, R.T., Booth, G.G., Tse, Y., and Zobotina, T., 2002, 'Price discovery and common factor models', *Journal of Financial Markets*, vol. 5, no. 3, pp. 309–321.
- Barbu, V.S., D'Amico, G., and De Blasis, R., 2017, 'Novel advancements in the Markov chain stock model: Analysis and inference', *Annals of Finance*, vol. 13, no. 2, pp. 125–152.
- Barsky, R.B. and De Long, J.B., 1993, 'Why does the stock market fluctuate?', *The Quarterly Journal of Economics*, vol. 108, no. 2, pp. 291–311.
- Basu, S., 1977, 'Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis', *The Journal of Finance*, vol. 32, no. 3, pp. 663–682.
- Berchtold, A. and Raftery, A.E., 2002, 'The mixture transition distribution model for high-order Markov chains and non-Gaussian time series', *Statistical Science*, vol. 17, no. 3, pp. 328–356.

- Beveridge, S. and Nelson, C.R., 1981, 'A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle', *Journal of Monetary Economics*, vol. 7, no. 2, pp. 151–174.
- Billingsley, P., 1961a, *Statistical inference for Markov processes*. University of Chicago Press, Chicago, IL.
- Billingsley, P., 1961b, 'Statistical methods in Markov chains', *The Annals of Mathematical Statistics*, vol. 32, no. 1, pp. 12–40.
- Black, F., 1972, 'Capital market equilibrium with restricted borrowing', *The Journal of Business*, vol. 45, no. 3, pp. 444–455.
- Blanchard, O.J. and Watson, M.W., 1982, 'Bubbles, rational expectations and financial markets', in Wachtel, P. (ed.), *Crises in the Economic and Financial Structure*. D.C. Heath and Company, Lexington, MA.
- Blume, M.E. and Friend, I., 1973, 'A new look at the capital asset pricing model', *The Journal of Finance*, vol. 28, no. 1, pp. 19–34.
- Boehmer, E., Fong, K.Y., and Wu, J., 2013, 'Algorithmic trading and changes in firms' equity capital'. Available at SSRN <https://ssrn.com/abstract=2348730>.
- Bohl, M.T., Salm, C.A., and Schuppli, M., 2011, 'Price discovery and investor structure in stock index futures', *Journal of Futures Markets*, vol. 31, no. 3, pp. 282–306.
- Booth, G.G., Lin, J.C., Martikainen, T., and Tse, Y., 2002, 'Trading and pricing in upstairs and downstairs stock markets', *The Review of Financial Studies*, vol. 15, no. 4, pp. 1111–1135.
- Brémaud, P., 1999, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, vol. 31. Springer Science & Business Media, Berlin, DE.

- Brooks, R. and Helms, B., 1990, 'An n-stage, fractional period, quarterly dividend discount model', *Financial Review*, vol. 25, no. 4, pp. 651–657.
- Campbell, J.Y. and Shiller, R.J., 1987, 'Cointegration and tests of present value models', *Journal of Political Economy*, vol. 95, no. 5, pp. 1062–1088.
- Campbell, J.Y., Lo, A.W., and MacKinlay, A.C., 1997, *The econometrics of financial markets*, vol. 2. Princeton University Press, Princeton, NJ.
- Chan, K., 1992, 'A further analysis of the lead–lag relationship between the cash market and stock index futures market', *The Review of Financial Studies*, vol. 5, no. 1, pp. 123–152.
- Ching, W.K. and Ng, M.K., 2006, *Markov chains: Models, algorithms and applications*. Springer US, New York.
- Ching, W.K., Fung, E.S., and Ng, M.K., 2002, 'A multivariate Markov chain model for categorical data sequences and its applications in demand predictions', *IMA Journal of Management Mathematics*, vol. 13, no. 3, pp. 187–199.
- Chordia, T., Roll, R., and Subrahmanyam, A., 2005, 'Evidence on the speed of convergence to market efficiency', *Journal of Financial Economics*, vol. 76, no. 2, pp. 271–292.
- Chordia, T., Roll, R., and Subrahmanyam, A., 2008, 'Liquidity and market efficiency', *Journal of Financial Economics*, vol. 87, no. 2, pp. 249–268.
- Chu, Q.C., Hsieh, W.I.G., and Tse, Y., 1999, 'Price discovery on the S&P 500 index markets: An analysis of spot index, index futures, and SPDRs', *International Review of Financial Analysis*, vol. 8, no. 1, pp. 21–34.
- Cochrane, J.H., 2009, *Asset pricing: Revised edition*. Princeton University Press, Princeton, NJ.
- Cootner, P.H., 1962, 'Stock prices: Random vs. systematic changes', *Industrial Management Review*, vol. 3, no. 2, pp. 24–45.

- Cootner, P.H., 1964, *The random character of stock market prices*. MIT Press, Cambridge, MA.
- Cowles, A. and Jones, H.E., 1937, ‘Some a posteriori probabilities in stock market action’, *Econometrica*, vol. 5, no. 3, pp. 280–294.
- D’Amico, G., 2017, ‘Stochastic dividend discount model: Risk and return’, *Markov Processes and Related Fields*, vol. 23, no. 3, pp. 349–376.
- D’Amico, G., 2013, ‘A semi-Markov approach to the stock valuation problem’, *Annals of Finance*, vol. 9, no. 4, pp. 589–610.
- D’Amico, G. and De Blasis, R., 2018, ‘A multivariate Markov chain stock model’. Available at SSRN: <http://ssrn.com/abstract=3296253>.
- D’Amico, G., Janssen, J., and Manca, R., 2006, ‘Homogeneous semi-Markov reliability models for credit risk management’, *Decisions in Economics and Finance*, vol. 28, no. 2, pp. 79–93.
- D’Amico, G., Janssen, J., and Manca, R., 2016, ‘Downward migration credit risk problem: A non-homogeneous backward semi-Markov reliability approach’, *Journal of the Operational Research Society*, vol. 67, no. 3, pp. 393–401.
- Damodaran, A., 2012, *Investment valuation: Tools and techniques for determining the value of any asset*. John Wiley & Sons, New York.
- De Blasis, R., 2018, ‘A new measure of price discovery in financial markets’. Available at SSRN: <http://ssrn.com/abstract=3297869>.
- Donaldson, R.G. and Kamstra, M., 1996, ‘A new dividend forecasting procedure that rejects bubbles in asset prices: The case of 1929’s stock crash’, *The Review of Financial Studies*, vol. 9, no. 2, pp. 333–383.
- Fama, E.F., 1963, ‘Mandelbrot and the stable Paretian hypothesis’, *The Journal of Business*, vol. 36, no. 4, pp. 420–429.

- Fama, E.F., 1965a, 'The behavior of stock-market prices', *The Journal of Business*, vol. 38, no. 1, pp. 34–105.
- Fama, E.F., 1965b, 'Random walks in stock market prices', *Financial Analysts Journal*, vol. 21, pp. 55–59.
- Fama, E.F. and Blume, M.E., 1966, 'Filter rules and stock-market trading', *The Journal of Business*, vol. 39, no. 1, pp. 226–241.
- Fama, E.F. and French, K.R., 1988, 'Permanent and temporary components of stock prices', *Journal of Political Economy*, vol. 96, no. 2, pp. 246–273.
- Fama, E.F. and French, K.R., 1992, 'The cross-section of expected stock returns', *The Journal of Finance*, vol. 47, no. 2, pp. 427–465.
- Fama, E.F. and French, K.R., 1993, 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics*, vol. 33, no. 1, pp. 3–56.
- Fama, E.F. and MacBeth, J.D., 1973, 'Risk, return, and equilibrium: Empirical tests', *Journal of Political Economy*, vol. 81, no. 3, pp. 607–636.
- Fleming, J., Ostdiek, B., and Whaley, R.E., 1996, 'Trading costs and the relative rates of price discovery in stock, futures, and option markets', *Journal of Futures Markets*, vol. 16, no. 4, pp. 353–387.
- Frino, A., Walter, T., and West, A., 2000, 'The lead–lag relationship between equities and stock index futures markets around information releases', *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, vol. 20, no. 5, pp. 467–487.
- Fuller, R.J. and Hsia, C.C., 1984, 'A simplified common stock valuation model', *Financial Analysts Journal*, vol. 40, no. 5, pp. 49–56.
- Garbade, K.D. and Silber, W.L., 1979, 'Dominant and satellite markets: A study of dually-traded securities', *The Review of Economics and Statistics*, vol. 61, no. 3, pp. 455–460.

- Garbade, K.D. and Silber, W.L., 1983, 'Price movements and price discovery in futures and cash markets', *The Review of Economics and Statistics*, vol. 65, no. 2, pp. 289–297.
- Garbade, K.D., Pomrenze, J.L., and Silber, W.L., 1979, 'On the information content of prices', *The American Economic Review*, vol. 69, no. 1, pp. 50–59.
- Ghezzi, L.L. and Piccardi, C., 2003, 'Stock valuation along a Markov chain', *Applied Mathematics and Computation*, vol. 141, no. 2-3, pp. 385–393.
- Goettler, R.L., Parlour, C.A., and Rajan, U., 2009, 'Informed traders and limit order markets', *Journal of Financial Economics*, vol. 93, no. 1, pp. 67–87.
- Gonzalo, J. and Granger, C., 1995, 'Estimation of common long-memory components in cointegrated systems', *Journal of Business & Economic Statistics*, vol. 13, no. 1, pp. 27–35.
- Gordon, M.J., 1959, 'Dividends, earnings, and stock prices', *The Review of Economics and Statistics*, vol. 41, no. 2, pp. 99–105.
- Gordon, M.J., 1962, *The investment, financing, and valuation of the corporation*. RD Irwin, Homewood, IL.
- Gordon, M.J. and Shapiro, E., 1956, 'Capital equipment analysis: The required rate of profit', *Management Science*, vol. 3, no. 1, pp. 102–110.
- Grammig, J., Melvin, M., and Schlag, C., 2005, 'Internationally cross-listed stock prices during overlapping trading hours: Price discovery and exchange rate effects', *Journal of Empirical Finance*, vol. 12, no. 1, pp. 139–164.
- Grünbichler, A., Longstaff, F.A., and Schwartz, E.S., 1994, 'Electronic screen trading and the transmission of information: An empirical examination', *Journal of Financial Intermediation*, vol. 3, no. 2, pp. 166–187.

- Gutiérrez, M.J. and Vázquez, J., 2004, 'Switching equilibria: The present value model for stock prices revisited', *Journal of Economic Dynamics and Control*, vol. 28, no. 11, pp. 2297–2325.
- Harris, F., McInish, T.H., and Wood, R.A., 2002, 'Security price adjustment across exchanges: An investigation of common factor components for Dow stocks', *Journal of Financial Markets*, vol. 5, no. 3, pp. 277–308.
- Harrison, J.M. and Kreps, D.M., 1978, 'Speculative investor behavior in a stock market with heterogeneous expectations', *The Quarterly Journal of Economics*, vol. 92, no. 2, pp. 323–336.
- Hasbrouck, J., 1995, 'One security, many markets: Determining the contributions to price discovery', *The Journal of Finance*, vol. 50, no. 4, pp. 1175–1199.
- Hauptfleisch, M., Putniņš, T.J., and Lucey, B., 2016, 'Who sets the price of gold? London or New York', *Journal of Futures Markets*, vol. 36, no. 6, pp. 564–586.
- Hendershott, T. and Jones, C.M., 2005, 'Island goes dark: Transparency, fragmentation, and regulation', *The Review of Financial Studies*, vol. 18, no. 3, pp. 743–793.
- Hou, K. and Moskowitz, T.J., 2005, 'Market frictions, price delay, and the cross-section of expected returns', *The Review of Financial Studies*, vol. 18, no. 3, pp. 981–1020.
- Huang, R.D., 2002, 'The quality of ECN and NASDAQ market maker quotes', *The Journal of Finance*, vol. 57, no. 3, pp. 1285–1319.
- Hurley, W.J. and Johnson, L.D., 1994, 'A realistic dividend valuation model', *Financial Analysts Journal*, vol. 50, no. 4, pp. 50–54.
- Hurley, W.J. and Johnson, L.D., 1998, 'Generalized Markov dividend discount models', *The Journal of Portfolio Management*, vol. 25, no. 1, pp. 27–31.

- Jarrow, R.A., Lando, D., and Turnbull, S.M., 1997, 'A Markov model for the term structure of credit risk spreads', *The Review of Financial Studies*, vol. 10, no. 2, pp. 481–523.
- Jensen, M.C., Black, F., and Scholes, M.S., 1972, 'The capital asset pricing model: Some empirical tests', in Jensen, M.C. (ed.), *Studies in the theory of capital markets*. Praeger, New York.
- Johansen, S., 1991, 'Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models', *Econometrica*, vol. 59, no. 6, pp. 1551–1580.
- Kamstra, M., 2003, 'Pricing firms on the basis of fundamentals', *Economic Review-Federal Reserve Bank of Atlanta*, vol. 88, no. 1, pp. 49–66.
- Kraft, D., 1988, 'A software package for sequential quadratic programming', in *Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt*. DFVLR.
- Lehmann, B.N., 2002, 'Some desiderata for the measurement of price discovery across markets', *Journal of Financial Markets*, vol. 5, no. 3, pp. 259–276.
- Lintner, J., 1965, 'Security prices, risk, and maximal gains from diversification', *The Journal of Finance*, vol. 20, no. 4, pp. 587–615.
- Lo, A.W. and MacKinlay, A.C., 1988, 'Stock market prices do not follow random walks: Evidence from a simple specification test', *The Review of Financial Studies*, vol. 1, no. 1, pp. 41–66.
- Malkiel, B.G., 1963, 'Equity yields, growth, and the structure of share prices', *The American Economic Review*, vol. 53, no. 5, pp. 1004–1031.
- Markov, A.A., 1906, 'Extension of the law of large numbers to dependent quantities', *Izv. Fiz.-Matem. Obsch. Kazan Univ.(2nd Ser)*, vol. 15, pp. 135–156.

- Markowitz, H., 1952, 'Portfolio selection', *The Journal of Finance*, vol. 7, no. 1, pp. 77–91.
- Marshall, B.R., Nguyen, N.H., and Visaltanachoti, N., 2013, 'ETF arbitrage: Intraday evidence', *Journal of Banking & Finance*, vol. 37, no. 9, pp. 3486–3498.
- Molodovsky, N., May, C., and Chottiner, S., 1965, 'Common stock valuation: Principles, tables and application', *Financial Analysts Journal*, vol. 21, no. 2, pp. 104–123.
- Nicolau, J., 2014, 'A new model for multivariate Markov chains', *Scandinavian Journal of Statistics*, vol. 41, no. 4, pp. 1124–1135.
- O'Hara, M. and Ye, M., 2011, 'Is market fragmentation harming market quality?', *Journal of Financial Economics*, vol. 100, no. 3, pp. 459–474.
- Osborne, M.F., 1959, 'Brownian motion in the stock market', *Operations Research*, vol. 7, no. 2, pp. 145–173.
- Poterba, J.M. and Summers, L.H., 1988, 'Mean reversion in stock prices: Evidence and implications', *Journal of Financial Economics*, vol. 22, no. 1, pp. 27–59.
- Putniņš, T.J., 2013, 'What do price discovery metrics really measure?', *Journal of Empirical Finance*, vol. 23, pp. 68–83.
- Quan, J., 1992, 'Two-step testing procedure for price discovery role of futures prices', *Journal of Futures Markets*, vol. 12, no. 2, pp. 139–149.
- Raftery, A.E., 1985, 'A model for high-order Markov chains', *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 47, no. 3, pp. 528–539.
- Roll, R., 1977, 'A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory', *Journal of Financial Economics*, vol. 4, no. 2, pp. 129–176.

- Rösch, D.M., Subrahmanyam, A., and Van Dijk, M.A., 2016, 'The dynamics of market efficiency', *The Review of Financial Studies*, vol. 30, no. 4, pp. 1151–1187.
- Rosenberg, J.V. and Traub, L.G., 2009, 'Price discovery in the foreign currency futures and spot market', *The Journal of Derivatives*, vol. 17, no. 2, pp. 7–25.
- Sadek, A. and Limnios, N., 2002, 'Asymptotic properties for maximum likelihood estimators for reliability and failure rates of Markov chains', *Communications in Statistics - Theory and Methods*, vol. 31, no. 10, pp. 1837–1861.
- Samuelson, P.A., 1973, 'Proof that properly discounted present values of assets vibrate randomly', *The Bell Journal of Economics and Management Science*, vol. 4, no. 2, pp. 369–374.
- Sapp, S.G., 2002, 'Price leadership in the spot foreign exchange market', *Journal of Financial and Quantitative Analysis*, vol. 37, no. 3, pp. 425–448.
- Shannon, C.E., 1948, 'A mathematical theory of communication', *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423.
- Sharpe, W.F., 1964, 'Capital asset prices: A theory of market equilibrium under conditions of risk', *The Journal of Finance*, vol. 19, no. 3, pp. 425–442.
- Siu, T.K., Ching, W.K., Fung, S.E., and Ng, M.K., 2005, 'On a multivariate Markov chain model for credit risk measurement', *Quantitative Finance*, vol. 5, no. 6, pp. 543–556.
- Sorensen, E.H. and Williamson, D.A., 1985, 'Some evidence on the value of dividend discount models', *Financial Analysts Journal*, vol. 41, no. 6, pp. 60–69.
- Theil, H., 1967, *Economics and information theory*. North-Holland Pub. Co., Amsterdam, NL.
- Theissen, E., 2002, 'Price discovery in floor and screen trading systems', *Journal of Empirical Finance*, vol. 9, no. 4, pp. 455–474.

- Tse, Y.K., 1995, 'Lead-lag relationship between spot index and futures price of the Nikkei Stock Average', *Journal of Forecasting*, vol. 14, no. 7, pp. 553–563.
- Van der Vaart, A.W., 1998, *Asymptotic statistics (Cambridge series in statistical and probabilistic mathematics)*. Cambridge University Press, Cambridge, UK.
- Vasileiou, A. and Vassiliou, P.C., 2006, 'An inhomogeneous semi-Markov model for the term structure of credit risk spreads', *Advances in Applied Probability*, vol. 38, no. 1, pp. 171–198.
- Werner, I.M. and Kleidon, A.W., 1996, 'UK and US trading of British cross-listed stocks: An intraday analysis of market integration', *The Review of Financial Studies*, vol. 9, no. 2, pp. 619–664.
- West, K.D., 1988, 'Dividend innovations and stock price volatility', *Econometrica*, vol. 56, no. 1, pp. 37–61.
- Williams, J.B., 1938, *The theory of investment value*. Harvard University Press, Cambridge, MA.
- Yan, B. and Zivot, E., 2010, 'A structural analysis of price discovery measures', *Journal of Financial Markets*, vol. 13, no. 1, pp. 1–19.
- Yao, Y.F., 1997, 'A trinomial dividend valuation model', *The Journal of Portfolio Management*, vol. 23, no. 4, pp. 99–103.