Some Markov-Switching Models for the Toronto Stock Exchange

Accolley, Delali

Le Mans Université

19 March 2021

Some Markov-Switching Models for the Toronto Stock Exchange *

Delali Accolley
accolleyd@aim.com
Laboratoire de recherche en Économie GAINS, Le Mans Université, Le Mans, France

March 19, 2021

Abstract

This research motivates the use of Markov chains in modeling financial time series. Then, it explains the returns and the volatility on the Toronto Stock Exchange (TSX) using some Markov-switching models. These models are: the conditional capital asset pricing model, the conditional Sharpe model, and the exponential autoregressive model with state-dependent heteroscedasticity. It also tests for cointégration between the TSX and some other major exchanges, relying on the first-order and the second-order Markov chains.

The asymmetry, the multiple peaks, or the fat tails in the distribution of the returns on the TSX and on the other exchanges indicates they could not be modelled as random realizations from a single normal distribution. The switching regressions turn out to have a greater explanatory power and provide further understanding of the TSX.

Keywords: Econometrics, Finance, Markov Chain.

JEL: G0, 016

Résumé

Cette recherché justifie l'utilisation des chaînes de Markov dans la modélisation des séries chronologiques financières. Ensuite, elle explique les rendements et la volatilité sur la Bourse de Toronto (TSX) en utilisant quelques modèles de Markov à changement de régime. Ces modèles sont : le modèle conditionnel d'évaluation des actifs financiers, le modèle conditionnel de Sharpe et le modèle autorégressif exponentiel avec une hétéroscedasdacité conditionnelle qui dépend du régime. Elle teste également la cointégration entre le TSX et d'autres bourses, en s'appuyant sur les chaînes de Markov de premier et de second ordre.

L'asymétrie, les nombreux pics ou l'épaisseur des queues de la distribution des rendements sur la TSX et sur les autres bourses indiquent qu'ils ne peuvent pas être modélisés comme étant des réalisations aléatoires provenant d'une seule distribution normale. Il s'avère que les régressions avec changement de régime ont un plus grand pouvoir explicatif et fournissent une meilleure compréhension du TSX.

Mots clés : Econométrie, Finance, chaîne de Markov.

JEL : G0, 016

*I am grateful to Messrs Jill Scullion and John Andrew from the TSX for having provided me with some information on the global industry classification standard.

Contents

Some Abbreviations and Acronyms

ADF	Augmented Dickey-Fuller
AEG	Augmented Engle-Granger
AIC	Akaike Information Criterion
AR	Auto-Regressive
ARCH	Auto-Regressive Conditional eteroskedasticity
ARMA	Auto-Regressive Moving Average
ASX	Australian Securities Exchange
BIC	Bayesian Information Criterion
Bovespa	Bolsa de Valores do Estado de São Paulo
BSE	Bombay Stock Exchange
CAC	Cotation Assistée en Continu
CDF	Cumulative Distribution Function
DAX	Deutscher Aktienindex
DCC	Dynamic Conditional Correlation
EM	Expectation-maximization
FTSE	Financial Times Stock Exchange
GARCH	Generalized Auto-Regressive Conditional eteroskedasticity
HMM	Hidden Markov Model
IBEX	Índice Bursátil Español
ISEQ	Irish Stock Exchange Quotient
NASDAQ	National Association of Securities Dealers Automated Quotations
NYSE	New York Stock Exchange
OLS	Ordinary Least Squares
PDF	Probability Density Function
SENSEX	Sensitive Index
SMI	Swiss Market Index
S&P	Standard & Poor's
TSX	Toronto Stock Exchange
UK	United Kingdom
US	United States

Non-Technical Summary

Motivation Financial time series are not normally distributed as often assumed. For example, on Toronto Stock Exchange (TSX), the likelihood of extreme negative or extreme positive returns is higher than in a normal distribution. Besides, whereas the graph of a normal distribution is bell-shaped, the graph of the actual distribution of the returns on the TSX is not symmetrical and has several peaks.

Objectives This research seeks to explain returns across the TSX by accommodating the asymmetry in their distributions and the higher likelihood of both extreme negative or extreme positive values. It also purports to study appropriately the existence of a long-run equilibrium relationship between stock prices on the TSX and stock prices on some other major exchanges.

Methodology I have assumed the returns are random realizations from a mixture of normal distributions, each distribution having its own mean and variance. Each of these means and variances are associated to one of the recurring states of a stock market (the bull or the bear markets).

Key Contributions I have modeled returns, volatility, and the state variable simultaneously across the sectors of the TSX. I have taken into account the recurring states of stock markets, while investigating the existence of long-run equilibrium relationship between the TSX and some other major exchanges.

Findings Materials (particularly, gold) and utilities are the defensive sectors of the TSX. During periods of low volatility, the financial, the energy, and the materials sectors account for more than half of the market returns. However, during periods of high volatility, the share of the energy sector drops considerably. Volatility on the TSX has also turned out to be asymmetric.

In bull markets which are periods of low volatility, the autocorrelation of returns could be high or low depending on whether the returns are low or extremely high. On the other hand, in bear markets which are periods of high volatility, the autocorrelation is low because returns are often negative and extremely low.

Modeling and identifying the recurring states of stock markets have revealed the existence of long-run equilibrium relationships between the stock prices on the TSX and the stock prices on some other major exchanges that include the New York Stock Exchange, the Tokyo Stock Exchange, and the London Stock Exchange. Whatever the method of estimation used, I have not found any evidence of equilibrium relationship between the stock prices on the TSX and the stock prices on either the NASDAQ or the Bolsa de Madrid.

It also turned out that dissociating the trends of stock markets from the turning points could improve considerably the explanatory power of the models.

Sommaire Non-Technique

Motivation Contrairement à ce qu'on suppose généralement, les séries chronologiques financières ne sont pas des variables qui suivent une distribution normale. Par exemple, sur la Bourse de Toronto (TSX), la probabilité d'observer des rendements négatifs ou positifs extrêmes est plus élevée que celle d'une distribution normale. Par ailleurs, alors que la représentation graphique d'une distribution normale a la forme d'une cloche, celle de la distribution des rendements du TSX n'est pas symétrique et a plusieurs pics.

Objectifs Cette recherché veut expliquer les rendements sur le TSX en tenant compte de leur distribution asymétrique et de la probabilité plus élevée d'observer des rendements positifs ou négatifs extrêmes. Elle vise également à étudier convenablement l'existence d'une relation d'équilibre de long terme entre le cours des actions sur le TSX et le cours des actions sur d'autres bourses.

Methodologie J'ai supposé que les rendements sont des réalisations aléatoires tirées d'un mélange de distributions normales ayant chacune sa propre moyenne et variance. Chacune de ces moyennes et variances est associée à un des états récurrents d'un marché boursier (marché haussier ou marché baissier).

Contributions majeures J'ai modélisé les rendements, la volatilité et la variable d'état simultanément à travers les secteurs du TSX. J'ai pris en compte les états récurrents des marchés boursiers en étudiant l'existence de relations d'équilibre de long terme entre le TSX et les autres bourses.

Résultats Les matériaux (particulièrement l'or) et les services publics sont les secteurs défensifs du TSX. Durant les périodes de faible volatilité, les secteurs de la finance, de l'énergie et des matériaux génèrent plus de ela moitié des rendements du marché. Toutefois, quand la volatilité est forte, la contribution du secteur de l'énergie baisse considérablement. La volatilité sur le TSX s'avère être asymétrique.

Dans les marches haussiers qui sont des périodes de faible volatilité, l'autocorrélation des rendements peut être élevée ou faible selon que les rendements sont faibles ou extrêmement élevés. Par contre, dans les marchés baissiers qui sont des périodes de haute volatilité, l'autocorrélation est faible parce que les rendements y sont souvent négatifs et extrêmement bas.

Le fait de modéliser et d'identifier les états récurrents des marches boursiers a révélé l'existence d'une relation d'équilibre de long terme entre les cours boursiers sur le TS et les cours des actions sur d'autres bourses, dont la Bourse de New York, la Bourse de Tokyo et la Bourse de Londres. Quelle que soit la méthode d'estimation utilisée, je n'ai trouvé aucune preuve de relation d'équilibre entre les cours des actions sur le TSX et les cours des actions sur le NASDAQ ou la Bourse de Madrid.

Il s'est avéré que le fait de dissocier les tendances des marchés boursiers de leurs tournants peut améliorer considérablement le pouvoir explicatif des modèles.

1 Introduction

Toronto Stock Exchange (TSX) is the largest stock market in Canada and also one of the ten most important in the world. Companies listed on the TSX can be classified into eleven sectors according to their main activities. These sectors are: consumer discretionary (*e.g.*, automobiles, consumer durables, hotels), consumer staples (*e.g.*, beverage, food, tobacco), energy (*e.g.*, oil and gas production, refining, or storage, coal), financial (*e.g.*, banks, capital markets, insurance), health care (*e.g.*, biotechnology, health care services, pharmaceuticals), industrial (*e.g.*, consulting services, security services, transportation), information technology (*e.g.*, electronic components or equipment, technology distributors), material (*e.g.*, aluminum, copper, gold), real estate (*e.g.*, equity real estate investment trusts, real estate development), telecommunication service (*e.g.*, advertising, broadcasting, entertainment), and utilities (*e.g.*, electricity, gas, water). For further details on this classification, see MSCI Barra and Standard & Poors (2018).

The indicators used to track the overall performance of the TSX and the performance of the sectors of this exchange are the various Standard & Poor's (S&P)/TSX indices and sub-indices. These indices are weighted averages of the trading prices of selected stocks. Their growth rates give the returns on the market and on a sector portfolio.

Two main trends characterize a stock exchange: the *bear* and the *bull* markets. A bear market is a less frequent period of financial turbulence where returns are generally low and highly volatile. On the other hand, a bull market is a period of widespread optimism and euphoria among investors. Returns across most sector are generally high and less volatile, during a bull market (Vendrame, Guermat, and Tucker, 2018). Each of these two trends, in its own way, interacts with the economic activity. For example, an incipient bear market that has started during a period of economic expansion, leads to a recession, which in turn causes a grizzly bear market. Hamilton and Lin (1996) find that the stock market volatility leads the economic activity by one month. They also find that, in turn, recession is the single and major cause of stock volatility. [1]

As Hamilton and Susmel (1994), Abdymomunov and Morley (2011), and Vendrame, Guermat, and Tucker (2018) point out, ignoring the state prevailing in the market leads to wrong estimates and forecasts when fitting models to financial time series. As a matter of fact, the parameters of models explaining returns and their volatility are not constant but instead are time-varying due to the alternation of various states on financial markets.

Therefore, this research aims at modeling returns on the TSX and their volatility conditionally on the state prevailing on this market. A way of doing this is through the use of a *Markov chain*. A Markov chain offers the possibility of modeling in a flexible and general way the probability of moving from an unobserved state to the other. For this reason, it can be used to explain and predict the alternation or the persistence of

[1] Hamilton and Lin (1996) measure the economic activity using the change in the natural logarithm of the industrial production index of the economy of the United States for the period ranging from January 1965 to June 1993. The stock market excess returns are the changes in the natural logarithm of the S&P 500 plus the dividend yields minus the monthly equivalent of the 3-month Treasury bill yield.

stock market trends. One can therefore induce time variation in the parameters of an econometric model by allowing its likelihood to depend on a Markov chain. Conditional models generated in this way are referred to as Markov-switching models.

Earlier research that have used Markov-switching models to explain returns on the TSX and their volatility include Van Norden and Schaller (1993). They deem that the existence of speculative bubbles in asset prices could explain their fluctuations. As a consequence, the states of stock markets (*i.e.*, crashes and booms) could stem from the apparent deviation of asset prices from their market fundamental price. The fundamental price of an asset is the sum of the current dividend it pays and all the expected future dividends discounted by the risk-free rate. Using data on the TSX for the period 1956-1989, Van Norden and Schaller calculate *inter alia* the *ex ante* and the *ex post* probability of a market crash. The *ex post* probability shows spikes that correspond to actual crashes. They also find that the *ex ante* probability rises before a crash, which suggests that deviations from the fundamentals could predict the states of stock markets.

In this research, I have used Markov chains to estimate: (1) the conditional capital asset pricing model (CAPM) for the sectors of the TSX, (2) the contribution to the market return of each of the sectors making up the TSX (the conditional Sharpe model), (3) the relation between the variance of the returns on the TSX and their autocorrelation using a conditional exponential autoregressive model, and (4) the conditional bivariate relationship between the stock prices on the TSX and the stock prices on some other major stock exchanges.

The rest of this paper is organized as follows. Section 2 motivates the use of the Markov-switching models when dealing with financial time series. The main reason is that the histogram or the graph of the nonparametric probability distribution of financial time series is not symmetrical and bell-shaped as the unconditional normal distribution used to represent them implies. Section 3 briefly introduces to Markov chains and Markov-switching models. The presentation of these two tools follows Hamilton (1994) and mostly Zucchini and MacDonald (2009). Sections 5 through 7 present the econometric models used to explain returns and their volatility as well as the empirical evidence. Section 8 concludes this research.

The particularity of some of the empirical investigations is the use of both the first-order and the second order Markov chains as well as the simultaneous estimations of the models using state-dependent multivariate normal distributions. The second-order Markov chain adds precision to the models, as it dissociates the two main trends of stock markets, which are the bear and the bull markets, from their turning points.

In Section 4, I have fitted mixtures of state-dependent multivariate normal models. Out of their component expected values and variance-covariance matrices, I have estimated the parameters of the conditional CAPM, simultaneously for all the sectors of the TSX. It emerges from this investigation that the consumer staples and the utilities sectors and the gold sub-industry offer a hedge against market downturns.

In Section 5, I have estimated the shares of each of the sectors of the TSX in the market return. It turns out that during periods of low volatility, the financial, the energy, and the materials sectors account for more than half of the market returns. On the other hand, during periods of high volatility, it is the financial sector followed by the consumer discretionary and the materials that generate most of the market returns. The importance of the contribution of most sectors depends on the state prevailing in the market. These investigations have also confirmed the existence of asymmetry in the volatility on the TSX.

In Section 6, I have investigated the relation between the variance of the daily ;returns and their autocorrelation. The exploratory analysis of the data shows that when the autocorrelation is high the volatility is low, but the autocorrelation is not always high when the volatility is low. This observation leads me to model the relation between the current and the lagged daily returns on the TSX using an exponential autoregressive process with state-dependent conditional heteroskedasticity. I turns out that during a bull markets (for the same level of volatility), when the returns are low their autocorrelation is high and when they are extremely high their autocorrelation is low. During bear markets where extreme negative returns are frequent, their autocorrelation is low.

In Section 7, I have investigated the existence of cointegration (*i.e.* a long-run equilibrium relationship) between the TSX and some other majors exchanges. Unlike some authors who have studied cointegration assuming structural breaks in the data, I have assumed that the bivariate relationships between the stock prices on the TSX and the stock prices on each of the other 15 exchanges in my sample depend instead on the recurring states of the stock markets. To estimate the parameters of these bivariate relations, I have fitted a state-dependent multivariate normal distribution to the data. The New York Stock Exchange, the Tokyo Stock Exchange, the London Stock Exchange, and some other exchanges have turned out to be cointegrated with the TSX. But, I have not found any evidence of cointegration between the TSX and either the NASDAQ or the Bolsa de Madrid.

2 Motivation

When fitting models to financial time series, it is often assumed that they are normally distributed. A random variable that is normally distributed has a bell-shaped *density curve*, which is symmetrical about its mean. A density curve is a graphical representation of a probability distribution. Furthermore, a normally distributed variable is completely described by its mean and its variance, which are assumed to be constant.

There are various ways of checking whether a financial time series is actually normally distributed. One could either compare its *kernel density curve* to that of a normal distribution or compute statistics describing the shape of its distribution. A kernel density curve is the graph of probability values estimated without assuming priorly a parametric statistical distribution. (I provide a note on kernel density estimation in Appendix B.1.) Two descriptors of the shape of a distribution are the *skewness* (S) and the *kurtosis* (K). For observations r_t ($t = 1, \ldots, T$) with mean $\bar{r}$, these descriptors are expressed as follows:

$$S = T^{\frac{1}{2}} \frac{\sum_{t=1}^{T} (r_t - \bar{r})^3}{\left[\sum_{t=1}^{T} (r_t - \bar{r})^2\right]^{\frac{3}{2}}}$$
$$K = T \frac{\sum_{t=1}^{T} (r_t - \bar{r})^4}{\left[\sum_{t=1}^{T} (r_t - \bar{r})^2\right]^{2}}.$$

Skewness measures the asymmetry of a distribution. A negative skewness indicates that the tail on the left-hand side (lhs) of a density curve is fatter than the one on its right-hand side (rhs). On the other hand, a positively skewed distribution has a longer tail on its rhs. As for the kurtosis, it indicates whether the tails on either side of a density curve are thinner or fatter than those of a normal distribution. The kurtosis of a normal distribution is 3. The excess kurtosis is therefore defined as the kurtosis minus 3. So, a distribution with thinner tails has a negative excess kurtosis and a distribution with fatter tails has a positive excess kurtosis. One can use simultaneously these two descriptors of the shape of a distribution to test for normality. A way of doing this is to perform Jarque-Bera test. The joint null hypothesis of this test is $S = 0$ and $K = 3$, and its alternative hypothesis is either $S \neq 0$ or $K \neq 3$. The statistic of Jarque-Bera (JB) test, which follows a χ^2 distribution with 2 degrees of freedom, is

$$JB = \frac{T}{6} \left[S^2 + \frac{(K-3)^3}{4} \right] \sim \chi^2(2).$$

Figure 2.1 compares the kernel density curves of returns across the TSX to those of normal distributions. It appears that assuming a normal distribution for returns across the TSX is not quite appropriate. The kernel density curves of the returns in sectors such as financial, industrial, information technology, and telecommunication service are slender than those of the normal distributions generated using their respective means and variances. Besides, unlike the density curve of a normal distribution, they are not bell-shaped and many of them have more than one peak.

Table 2.1 reports the estimates for the skewness, the kurtosis, and the Jarque-Bera stastitic across the TSX, *inter alia*. Returns on the TSX are negatively skewed, except for the gold sub-industry. This means that, with the exception of the gold sub-industry, the likelihood of extreme negative returns is higher than that of extreme positive returns across the TSX. The highest skewness is observed in the financial sector. All the returns across the TSX have a positive kurtosis. This means they have more outliers, *i.e.* extreme values, than a normally distributed variable. The returns in the financial and in the information technology sectors display the highest kurtosis. Finally, all the Jarque-Bera statistics are greater than their 5% critical value, which equals 5.991. Therefore, the null hypothesis that the returns are normally distributed cannot be accepted. In an earlier investigation, Episcopos (1996) conclude, using daily data ranging from July 30, 1990 to June 30, 1994, that returns across the TSX follow distributions that deviate from the normal distribution.

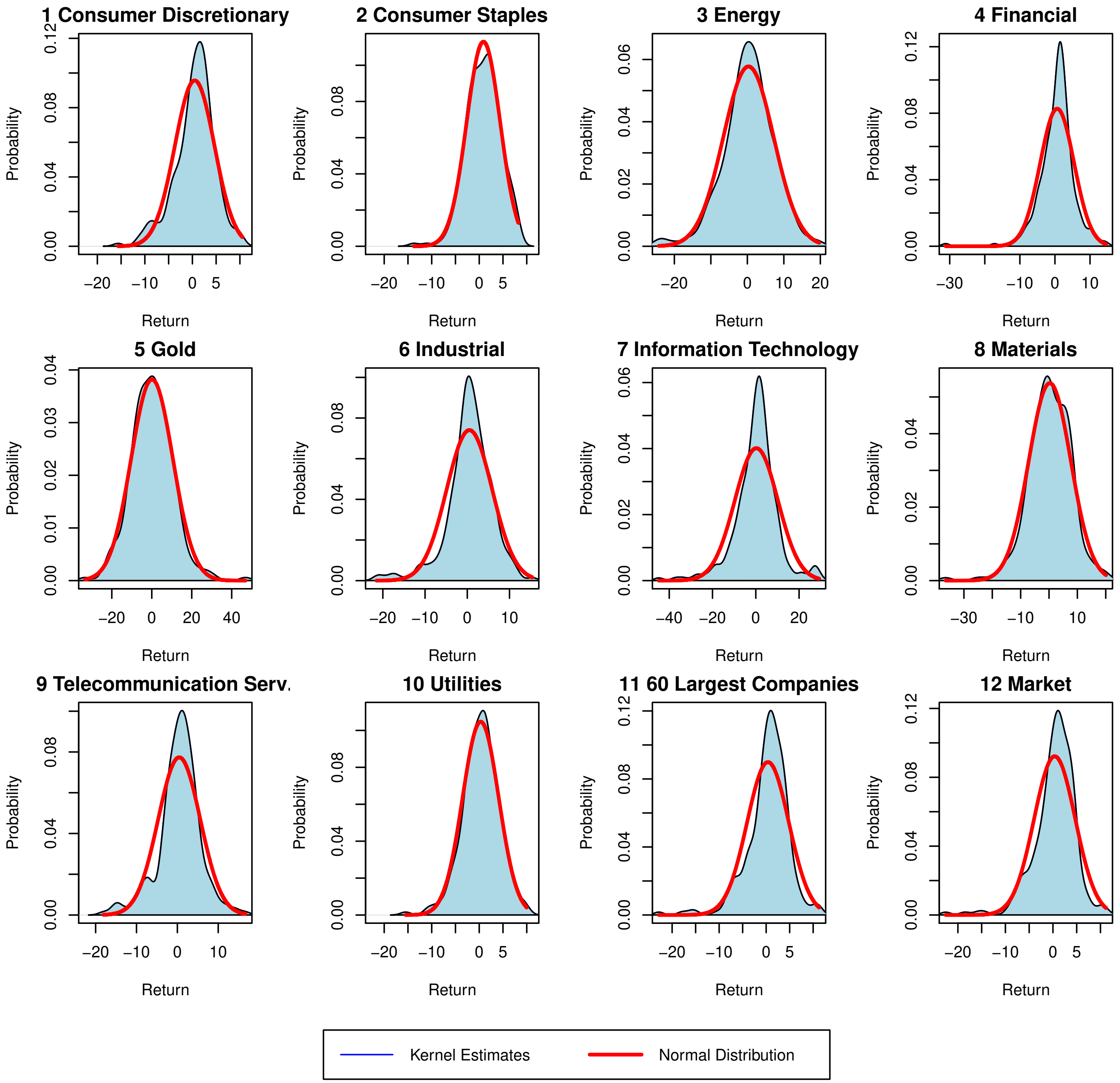

Figure 2.1: Comparison of the Kernel Density Curves of Returns across the TSX to those of Normal Distributions.

Table 2.1: Characteristics of Returns, TSX, 1998:M1-2017:M12.

Sector/Segment	Mean	Standard Deviation	Skewness	Excess Kurtosis	Jarque Bera
Consumer Discretionary	.488	4.170	-0.644	1.053	28.154
Consumer Staples	.884	3.532	-0.371	.760	9.849
Energy	.318	6.909	-0.440	1.227	26.112
Financial	.629	4.825	-1.266	7.997	5 156.879
Industrial	.502	5.388	-0.986	2.730	241.325
Information Technology	.267	9.949	-0.577	3.326	379.571
Materials	.272	7.421	-0.618	2.458	163.192
Gold	.114	10.443	0.386	1.630	49.084
Telecommunication Service	.458	5.156	-0.502	1.731	61.642
Utilities	.300	3.811	-0.400	1.126	20.595
The 60 Largest Companies	.386	4.438	-1.269	4.524	986.335
Market return	.370	4.328	-1.364	4.880	1 231.664

The two parameters describing a normal distribution, which are the mean and the variance, are assumed to be constant. Actually, the average value of stock returns vary depending on the general market sentiment, which can be bearish (*i.e.*, negative), bullish (*i.e.*, positive), mixed, or neutral. Besides, an empirical regularity characterizing financial time series referred to as *volatility clustering* proves wrong the assumption that the variances of stock returns are constant parameters. Volatility clustering is the observation that periods of unusually high volatility in financial time series are followed by quieter periods, as high (low) negative or positive returns tend to follow high (low) returns. As a consequence, the variance, a measure of volatility, is not a constant parameter but a random variable fluctuating around a constant mean.

To take into account volatility clustering, (generalized) autoregressive conditional heteroskedasticity (ARCH) models popularized by Engle (1982) and Bollerslev (1986) are used to estimate the variance of financial time series. But, as Hamilton and Susmel (1994) showed using weekly returns on the NYSE, these models impute a lot of persistence to stock volatility and give relatively poor forecasts. An alternative way of specifying these models is to consider high and low volatility periods as distinct states of nature and then assume that returns are random realizations from a state-dependent mixture of normally distributions, each having its own mean and variance.

3 The Markov-Switching Model

3.1 First-Order Markov Chain

A latent (*i.e.*, an unobserved) state variable S_t ($t = 1, \ldots, T$) that assumes only one of the discrete values $1, \ldots, m$, is said to be a Markov chain if it satisfies the following property

$$\Pr\left(S_{t+1} = j | S_t = i, \ldots, S_1 = g\right) = \Pr\left(S_{t+1} = j | S_t = i\right) = \gamma_{ij}. \tag{3.1}$$

According to (3.1), the probability of moving from a state i at time t to a state j at time $t+1$ does not depend on the past realizations of S_t. The conditional probabilities γ_{ij} can be compacted into an $m \times m$ matrix $\mathbf{\Gamma}$ referred to as transition probability matrix. Each row of $\mathbf{\Gamma}$ represents the probability of moving from a given state i to all the other possible states and consequently sums to unity.

$$\mathbf{\Gamma 1}' = \mathbf{1}', \tag{3.2}$$

where $\mathbf{1}'$ is the transpose of an $m \times 1$ vector of ones.

According to (3.2), $\mathbf{1}'$ and 1 are respectively the a right *eigenvector* and the corresponding *eigenvalue* of $\mathbf{\Gamma}$. Generally speaking, a nonzero column vector $\mathbf{v}$ is said to be a right eigenvector of a square matrix $\mathbf{\Gamma}$ and the scalar λ its eigenvalue, if $\mathbf{\Gamma v} = \lambda \mathbf{v}$. On the other hand, a nonzero row vector $\mathbf{u}$ is said to be the left eigenvector of $\mathbf{\Gamma}$, if $\mathbf{u\Gamma} = \lambda \mathbf{u}$. A right and a left eigenvectors differ from each other, unless $\mathbf{\Gamma}$ is a symmetric matrix. It thus follows that the left eigenvector corresponding to the eigenvalue 1 differs from $\mathbf{1}$, unless $\mathbf{\Gamma} = \mathbf{\Gamma}'$.

The left eigenvector of $\mathbf{\Gamma}$ corresponding the eigenvalue 1 is of particular interest because it defines the stationary distribution of S_t. To see that, let $\mathbf{u}_t$ denote the unconditional probability distribution of S_t,

$$\begin{aligned} \mathbf{u}_t &= [\Pr(S_t = 1), \ldots, \Pr(S_t = m)] \\ \mathbf{u}_{t+1} &= \mathbf{u}_t \mathbf{\Gamma}. \end{aligned} \tag{3.3}$$

The unconditional probability distribution $\mathbf{u}_t$ is stationary if, *inter alia*, its expectation $\mathrm{E}(\mathbf{u}_t)$ equals $\mathrm{E}(\mathbf{u}_{t+1}) = \mathbf{u}$, which implies that (3.3) becomes $\mathbf{u} = \mathbf{u\Gamma}$. (See Appendix B.3, for a definition of stationarity.) Thus, the left eigenvector of $\mathbf{\Gamma}$ associated to the eigenvalue 1 corresponds to the stationary distribution of the S_t. Since $\mathbf{u}$ is a vector of the unconditional probabilities of all the possible values that S_t can assume, one expects it to sum to unity. Putting together the condition for stationarity and the constraint $\mathbf{u1}' = 1$ gives the folling relation

$$\mathbf{u}\left(\mathrm{I}_m - \mathbf{\Gamma} + \mathbf{J}\right) = \mathbf{1}, \tag{3.4}$$

where $\mathbf{I}_m$ and $\mathbf{J}$ are respectively the identity matrix of size m and an $m \times m$ matrix of ones.

3.2 Higher-Order Markov Chain

When the probability of moving from a state i at time t to a state j at time $t+1$ depends also on some earler realizations of S_t, the latter latent variable is said to follow a higher-order Markov chain. As an example, a second-order Markov chain satisfies the following property

$$\Pr\left(S_{t+1} = j | S_t = i, \ldots, S_1 = g\right) = \Pr\left(S_{t+1} = j | S_t = i, S_{t-1} = h\right) = \gamma_{hij}. \tag{3.5}$$

The latent process described by (3.5) can be transformed into a first-order Markov chain by constructing a 1×2 vector from a combination of S_{t-1} and S_t. For a two-state Markov chain, there are four possible combinations of the values that S_{t-1} and S_t can assume, which are: (1) $[S_{t-1} = 1, S_t = 1]$, (2) $[S_{t-1} = 1, S_t = 2]$, (3) $[S_{t-1} = 2, S_t = 1]$, and (4) $[S_{t-1} = 2, S_t = 2]$. These four combinations are respectively labeled $S_t^* = 1, \ldots, 4$.

If the newlyy defined latent variable S_t^* is in state 1, which corresponds to the vector $[S_{t-1} = 1, S_t = 1]$, the next period, it will be impossible to move to state 3, which corresponds to the vector $[S_t = 2, S_{t+1} = 1]$, or to move to state 4 corresponding to $[S_t = 2, S_{t+1} = 2]$. The reason is that the second element of the vector corresponding to $S_t^* = 1$, which is $S_t = 1$, does not match the first element of the vectors associated to $S_{t+1}^* = 3$ or $S_{t+1}^* = 4$, which is $S_t = 2$. When $S_t^* = 1$, the only two options available are: either to remain in this state with probability c_1 or to move to state 2 with probability $1 - c_1$. The same reasoning applies to states S_t^* 2 through 4.

The transition probability matrix of the newly defined first-order Markov chain S_t^* resulting from the transformation of the second-order Markov chain S_t is therefore

$$\mathbf{\Gamma}^* = \begin{bmatrix} c_1 & 1-c_1 & 0 & 0 \\ 0 & 0 & c_2 & 1-c_2 \\ c_3 & 1-c_3 & 0 & 0 \\ 0 & 0 & c_4 & 1-c_4 \end{bmatrix}.$$

Hamilton and Lin (1996, p 578) show how to transform a third-order Markov chain into a first-order one. In this research, I only deal with first-order and second-order Markov chains.

3.3 Dependent Mixture

A time series R_t is said to be generated by an m-state Markov-switching model if, in addition to either (3.1) or (3.5), its conditional probability satisfies the following property

$$\begin{aligned} &\Pr\left(R_t | R_{t-1}, \ldots, R_1, \mathbf{X}_t, \ldots, \mathbf{X}_1, S_t, \ldots, S_1; \boldsymbol{\theta}\right) \\ &= \Pr\left(R_t | R_{t-1}, \ldots, R_{t-k}, \mathbf{X}_t, \ldots, \mathbf{X}_{t-k}, S_t, \ldots, S_{t-k}; \boldsymbol{\theta}\right), \\ &\equiv \Pr\left(R_t | \mathbf{Z}_t; \boldsymbol{\theta}\right) \end{aligned} \tag{3.6}$$

where $\mathbf{Z}_t = \left[R_{t-1}, \ldots, R_{t-k}, \mathbf{X}'_t, \ldots, \mathbf{X}'_{t-k}, S_t, \ldots, S_{t-k}\right]'$, $\mathbf{X}_t$ and $\boldsymbol{\theta}$ are respectively vectors of exogenous variables and parameters.

In (3.6), the distribution of R_t depends not only on its own past k realizations and the $k + 1$ most recent values assumed by some exogenous explanatory variables but also on the states of an unobserved Markov process. It is referred as state-dependent probability distribution because of it depends on the states of a Markov chain.

The state-dependent probability given by (3.6) is a general specification for all the models I will be dealing with. The vector of parameters $\boldsymbol{\theta}$ as well as the transition probabilities can be estimated either by directly maximizing the likelihood of the observations or by implementing the expectation-maximization (EM) algorithm (for more details, see Hamilton, 1990; Zucchini and MacDonald, 2009, among others).

4 The Conditional CAPM

The CAPM predicts a linear relationship between the expected *excess return* on a financial asset and the market excess return. An excess return (also known as risk premium) is the difference between the returns on a risky and a risk-free assets. In Appendix B.4, I present a derivation of the CAPM

$$\mathrm{E}(R_{it} - R_{ft}) = \beta_i \mathrm{E}(R_{mt} - R_{ft}), \tag{4.1}$$

where E is the expectation operator. The variables R_{ft}, R_{it}, and R_{mt} respectively denote the returns on the risk-free asset, on the risky asset i $(i = 1, 2 \dots)$, and on the market. The slope parameter β_i, which is the sensitivity of the excess return on an asset i to the excess market return, is referred to as systematic risk or (market) beta.

The CAPM has received little empirical support as an intercept term and other explanatory variables have turned out to be instrumental in explaining the excess return on an asset (see Fama and French, 2004, among others). [2]

Besides, the slope parameter β_i is not constant over time as the CAPM posits (Jagannathan and Wang, 1996; Abdymomunov and Morley, 2011; Vendrame, Guermat, and Tucker, 2018). For instance, cyclical stocks, as opposed to defensive stocks, tend to perform well when the market trends upwards and tend to perform poorly when it trends downwards. As a consequence, the beta of a portfolio of cyclical assets is expected to be higher when the market trends upwards than when it trends downwards.

There are various ways of modeling conditionally the CAPM. Henriksson and Merton (1981) propose the following deterministic approach to test for the ability of an investor to correctly forecast the sign of the excess market return,

$$R_{it} - R_{ft} = \alpha_i + \beta_{1i}(R_{mt} - R_{ft}) + \beta_{2i} \max(R_{mt} - R_{ft}, 0) + \varepsilon_{it}. \tag{4.2}$$

Model (4.2) distinguishes between two states, which are: (1) the up market where $R_{mt} - R_{ft} > 0$ and $\beta_i = \beta_{1i} + \beta_{2i}$, and (2) the down market where $R_{mt} - R_{ft} < 0$ and $\beta_i = \beta_{1i}$.

Pettengill, Sundaram, and Mathur (1995) and Lam (2001) also use a dummy a variable to distinguish between up markets and down markets. But, unlike Henriksson and Merton (1981), Pettengill, Sundaram, and Mathur (1995) use a two-pass approach to study instead the conditional relationship between the realized returns on risky portfolios and the market beta

$$R_{it} = \gamma_{ot} + \gamma_{1t}\delta_t\beta_i + \gamma_{2t}(1 - \delta_t)\beta_i + \varepsilon_{it}, \tag{4.3}$$

where $\delta_t = 1$, if $R_{mt} - R_{ft} > 0$, and $\delta = 0$, if $R_{mt} - R_{ft} < 0$. The explanatory variable in (4.3), which is β_i, results from a prior estimation of (4.1) for each individual asset using time series. The intercept terms γ_{0t} are the returns on the risk-free asset. The

[2] Some of these explanatory variables referred to as the CAPM anomalies are: the earnings-price ratio (*i.e.*, the capital gain and dividend on a stock relative to its market value), the debt-equity ratio (*i.e.*, the total liabilities of a company divided by the value of shareholders' equity), and the book-to-market ratio (*i.e.*, the ratio of the book value of a company's common equity to its market value).

parameters γ_{1t} and γ_{2t} are the market excess returns. Consequently, γ_{1t} is expected to be positive since it is associated to the dummy variable $\delta = 1$ and γ_{2t} is expected to be negative since it is associated with $\delta = 0$. To check for the conditional CAPM, Pettengill, Sundaram, and Mathur (1995) propose to perform some one-sided significance tests on the averages of the T cross-sectional estimates of γ_{1t} and γ_{2t}. If the averages γ_{1t} and γ_{2t} are respectively significantly positive and negative, this means there is actually a positive relationship between the returns and the beta of assets during periods of positive excess market return and a negative relationship during periods of negative excess market returns.

The problem with using either (4.2) or (4.3) to estimate the CAPM conditionally on the state prevailing in the market is the deterministic nature of the threshold used. For example, a one-off positive excess return does not mean that a market is trending upward or is in a bull state. Abdymomunov and Morley (2011) and Vendrame, Guermat, and Tucker (2018) defined the bear and the bull states by rather distinguishing between two states of the volatility in the market.

Abdymomunov and Morley (2011) estimate the CAPM distinguishing between the low and high volatility states of the market. The switch between these two states follows a Markov chain. They estimate the conditional CAPM in three main steps. The first main step of their three-pass approach consists in decoding the states of the market using the relation

$$R_{mt} - R_{ft} = \mu_{m,S_t} + \sigma_{m,S_t} z_{mt}, \tag{4.4}$$

where the market innovation, $\sigma_{m,S_t} z_{mt}$, follows a zero-mean state-dependent normal distribution They estimate the standard deviations of this distribution, σ_{m,S_t}, along with the expected values of market return, μ_{m,S_t}, using the maximum likelihood method. The second pass consists in estimating the conditional betas using the dynamic conditional correlation (DCC) model of Engle (2002). To explain the DCC model, let's consider the following two variables that folows a normal distribution

$$\begin{bmatrix} R_{mt} - R_{ft} \\ R_{it} - R_{ft} \end{bmatrix} = \begin{bmatrix} \mu_{mt} \\ \mu_{it} \end{bmatrix} + \begin{bmatrix} \sigma_{m,m,t} & \sigma_{m,i,t} \\ \sigma_{m,i,t} & \sigma_{i,i,t} \end{bmatrix} \begin{bmatrix} z_{mt} \\ z_{it} \end{bmatrix}, \ (i = 1, 2, \dots), \tag{4.5}$$

where the beta of asset i equals $\sigma_{m,i,t}/\sigma^2_{m,m,t}$, as shown in relation (B.26). Knowing the two conditional variances, $\sigma^2_{m,m,t}$ and $\sigma^2_{i,i,t}$, and the conditional correlation coefficient, one can estimate the dynamic covariance $\sigma_{m,i,t}$ using the following relation

$$\mathbf{\Sigma}_{it} = \mathbf{D}_{it}\mathbf{C}_{it}\mathbf{D}_{it},$$

where $\mathbf{\Sigma}_{it}$ denotes the conditional variance-covariance matrix defined in relation (4.5), the matrix $\mathbf{D}_{it}$ consists of the square root of the diagonal elements of $\mathbf{\Sigma}_{it}$, and the matrix $\mathbf{C}_{it}$ is the matrix of dynamic conditional correlations. To obtain the consecutive values of $\sigma^2_{m,m,t}$ and $\sigma^2_{i,i,t}$, Vendrame, Guermat, and Tucker (2018) assume they follows a generalized autoregressive conditional heteroskedasticity (GARCH) process (which is described in Section 5 and in Section 6). The estimator for the DCC proposed by Engle

(2002) is defined as follows.

$$\mathbf{Q}_{it} = (1 - a - b)\mathbf{C}_i + a\mathbf{e}_{t-1}\mathbf{e}'_{t-1} + b\mathbf{Q}_{i,t-1}$$
$$\mathbf{C}_{it} = \left(\text{diag}\mathbf{Q}_{it}\right)^{1/2} \mathbf{Q}_{it} \left(\text{diag}\mathbf{Q}_{it}\right)^{1/2}$$

where $\mathbf{Q}_{it}$ and $\mathbf{C}_i$ denote respectively the matrices of pseudo and unconditional covariances (correlations), and $\mathbf{e}_{t-1}$ is a vector of standardized residuals. The third pass of the approach of Vendrame, Guermat, and Tucker (2018) consists in estimating the bear and the bull risk premia using the individual fixed effects panel model. Using data on 25 portfolios of stocks over the sample period 1926-2015 and the sub-sample 1980-2015, they find (1) the bear risk premium to be negative and significant, and (2) the bull risk premium to be positive and significant.

Abdymomunov and Morley (2011) also estimate a Markov-switching CAPM. They assume volatility to be constant within each of the two or three possible states of the market. But, unlike Vendrame, Guermat, and Tucker (2018) who rely on a GARCH process to estimate several betas for a single asset, they assume that this parameter takes on two or three values that are switche by the very Markov chain that drives changes in the market volatility. They further assume that the news idiosyncratic to each asset (*i.e.*, the residuals of the CAPM) also follows a two-state Markov-switching process that is independent of that of the market. To estimate their models by the maximum likelihood method, they form portfolios using the returns of all stocks listed on the NYSE, the AMEX, and the NASDAQ over the sample period ranging from July 1963 to December 2010. They find supporting evidence for the Markov-switching conditional CAPM and, performing diagnostic checks on the residuals of their models, they also conclude that they capture the ARCH effects and the non-normalities oberved in the data.

I use a slightly different and simpler approach to estimate in one go the CAPM for ten of the sectors of the TSX. I assume that the market excess returns on the TSX and the excess returns of its ten sectors follow a multivariate normal distribution whose means and variance-covariance matrix depend on a single Markov process. This is a generalization of the model used by Vendrame, Guermat, and Tucker (2018). But, relying on the evidence from the diagnostic checks on residuals performed by Abdymomunov and Morley (2011), I assume that the volatility of each asset is constant within each of the states of the market. Earlier attempts to estimate the CAPM from state-dependent multivariate normal distributions include Tu (2010).

4.1 The Method of Estimation

Let's consider the column vector $\mathbf{Y}_t = [R_{mt} - R_{ft}, (\mathbf{R}_{st} - R_{ft}\mathbf{1})']'$ that lists the market excess return and the distribution of the excess returns across the sectors of the TSX, at time t. The joint probability of observing $\mathbf{Y}_t$ is described by the state-dependent multivariate normal distribution

$$
\begin{aligned}
p\left(\mathbf{Y}_t;\boldsymbol{\mu}_{S_t},\boldsymbol{\Sigma}_{S_t}\right) = (2\pi)^{-\frac{11}{2}}(\det\boldsymbol{\Sigma}_{S_t})^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\mathbf{Y}_t-\boldsymbol{\mu}_{S_t}\right)'\boldsymbol{\Sigma}_{S_t}^{-1}\left(\mathbf{Y}_t-\boldsymbol{\mu}_{S_t}\right)\right] \\
\text{with } \boldsymbol{\mu}_{S_t} = \begin{bmatrix}\boldsymbol{\mu}_{m,S_t}\\ \boldsymbol{\mu}_{s,S_t}\end{bmatrix} \text{ and } \boldsymbol{\Sigma}_{S_t} = \begin{bmatrix}\boldsymbol{\Sigma}_{m,m,S_t} & \boldsymbol{\Sigma}_{m,s,S_t}\\ \boldsymbol{\Sigma}_{m,s,S_t} & \boldsymbol{\Sigma}_{s,s,S_t}\end{bmatrix},
\end{aligned} \tag{4.6}
$$

where the operator det denotes the determinant and the variable S_t in subscript denotes either the first-order or the second-order Markov chain described by relation (3.1) or (3.5).

The conditional systematic risks, $\boldsymbol{\beta}_{S_t}$, and the measure of performance, $\boldsymbol{\alpha}_{S_t}$, of the sectors can be estimated simultaneously using the parameters of the state-dependent multivariate normal distribution in (4.6).

$$
\begin{aligned}
\boldsymbol{\beta}_{S_t} &= \boldsymbol{\Sigma}_{m,s,S_t}\boldsymbol{\Sigma}_{m,m,S_t}^{-1} \\
\boldsymbol{\alpha}_{S_t} &= \boldsymbol{\mu}_{s,S_t} - \left(\boldsymbol{\Sigma}_{m,s,S_t}\boldsymbol{\Sigma}_{m,m,S_t}^{-1}\right)\boldsymbol{\mu}_{m,S_t}
\end{aligned} \tag{4.7}
$$

The parameters $\boldsymbol{\mu}_{S_t}$ and $\boldsymbol{\Sigma}_{S_t}$ ($S_t = 1, 2, \dots$) of the state-dependent multivariate normal distributions are themselves estimated maximizing the likelihood of observing a sample $\mathbf{y}_t$ ($t = 1, 2, \dots, T$) of the realized excess returns. The likelihood of the first-order and the second-order Markov-switching models are

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= \prod_{t=1}^{T}\sum_{s_t=1}^{2}\gamma_{s_{t-1},s_t}p\left(\mathbf{y}_t|s_t;\boldsymbol{\theta}_{s_t}\right) \\
&= \mathbf{u}\prod_{t=1}^{T}\boldsymbol{\Gamma}\mathbf{P}\left(\mathbf{y}_t\right)\mathbf{1}' \\
\text{with } \boldsymbol{\theta} &= \left[\gamma_{11},\gamma_{12},\gamma_{21},\gamma_{22},\boldsymbol{\mu}_1',\boldsymbol{\mu}_2',\text{vech}(\boldsymbol{\Sigma}_1)',\text{vech}(\boldsymbol{\Sigma}_2)'\right] && \text{(4.8a)} \\
L(\boldsymbol{\theta}^*) &= \prod_{t=1}^{T}\sum_{s_t=1}^{2}\gamma^*_{s^*_{t-1},s^*_t}p\left(\mathbf{y}_t|s^*_t;\boldsymbol{\theta}^*_{s_t}\right) \\
&= \mathbf{u}^*\prod_{t=1}^{T}\boldsymbol{\Gamma}^*\mathbf{P}\left(\mathbf{y}_t\right)\mathbf{1}' \\
\text{with } \boldsymbol{\theta}^* &= \left[\gamma^*_{11},\gamma^*_{12},\dots,\gamma^*_{44},\boldsymbol{\mu}_1^{*\prime},\dots,\boldsymbol{\mu}_4^{*\prime},\text{vech}(\boldsymbol{\Sigma}^*_1)',\dots,\text{vech}(\boldsymbol{\Sigma}^*_4)\right], && \text{(4.8b)}
\end{aligned}
$$

where the variable s^*_t, as explained in Section 3, is a transformation of a second-order Markov chain into a first-order chain, the vector $\mathbf{u}$, the stationary distribution of s_t, is defined by relation (3.4), $\boldsymbol{\Gamma}$ is the matrix of transition probabilities, and $\mathbf{P}(\mathbf{y}_t)$ is a diagonal matrix whose j-th diagonal element corresponds to either $p(\mathbf{y}_t|s_t = j;\boldsymbol{\theta})$ or $p(\mathbf{y}_t|s^*_t = j;\boldsymbol{\theta}^*)$. The operator vech denotes the half-vectorization of the variance-covariance matrix (*i.e.* the transformation of this symmetric matrix into a column vector by stacking only its lower trinagular elements)

I have estimated the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ by maximizing directly the natural logarithm of the likelihood (log-likelihood, in short) of the Markov-switching models. To do that, I have used the base function *nlm*, which stands for non-linear minimization, of the software R (`www.r-project.org`). The models have multiple maxima. Therefore, for each of the two models, I have performed the numerical optimization hundred times, with different starting parameter values generated randomly, in order to select the estimates that give the highest log-likelihood. Since repeating this several times gives roughly the same estimates, I have concluded they are global maxima.

The estimated parameters are assumed to be asymptotically normal. Then, one needs their standard errors to perform the tests of significance. A way of getting them is by inverting the Hessian matrix (*i.e.*, the matrix of the second derivatives of the log-likelihood with respect to the parameters). [3] The Hessian matrix is computed numerically at the maximum. Since some of the estimated parameters at the maximum can lie on or close to the boundary of the parameter space, inverting the Hessain matrix does not always yield finite numbers. To overcome this issue, I have performed some bootstrapping using the estimated parameters to sample the explained variables $\mathbf{r}_{st} - r_{ft}\mathbf{1}$ $(t = 1, \ldots, T)$ one thousand times from state-dependent normal distributions. Then, I have estimated new parameters using the simulated time series and the observed $\mathbf{r}_{mt}$. The standard errors of the parameters are then computed as the standard deviation of the bootstrap estimates. To perform the significance tests, I have computed z-statistics by dividing the global maxima by the bootstrap standard errors.

4.2 The Findings

I present some evidence from estimating the expected excess returns and their variance-covariance matrices using, in turn, the first-order Markov chain (*i.e.*, maximizing the likelihood given by relation (7.2a)) and the second-order Markov chain (*i.e.*, maximizing the likelihood given by relation (7.2b)). Note that both types of Markov chains have two states. The resulting estimates of the conditional CAPM are also presented. The data used to estimate the models are described in Appendix A.

4.2.1 The First-Order Markov-Switching CAPM

Figure 4.1 compares the kernel density curves of the excess returns across the TSX fo the mixture of the state-dependent multivariate normal distributions fitted to these data. The marginal distributions that are produced using the component expected values and standard deviations of the Markov-switching model are close to the actual distributions of the excess returns. However, the actual distributions in the financial, industrial, and information technology sectors are somewhat taller than the ones fitted to the data.

Relation (4.9) shows the estimates of the transition probabilities of the two states of the Markov chain and their stationary distribution. State 1 is the most likely one. It occurs 64% of the time and state 2 occurs 36% of the time. As it appears in Figure 4.2,

[3] $\mathrm{var}(\boldsymbol{\theta}) = \left\{ -\mathrm{E}\left[\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \right\}^{-1}$.

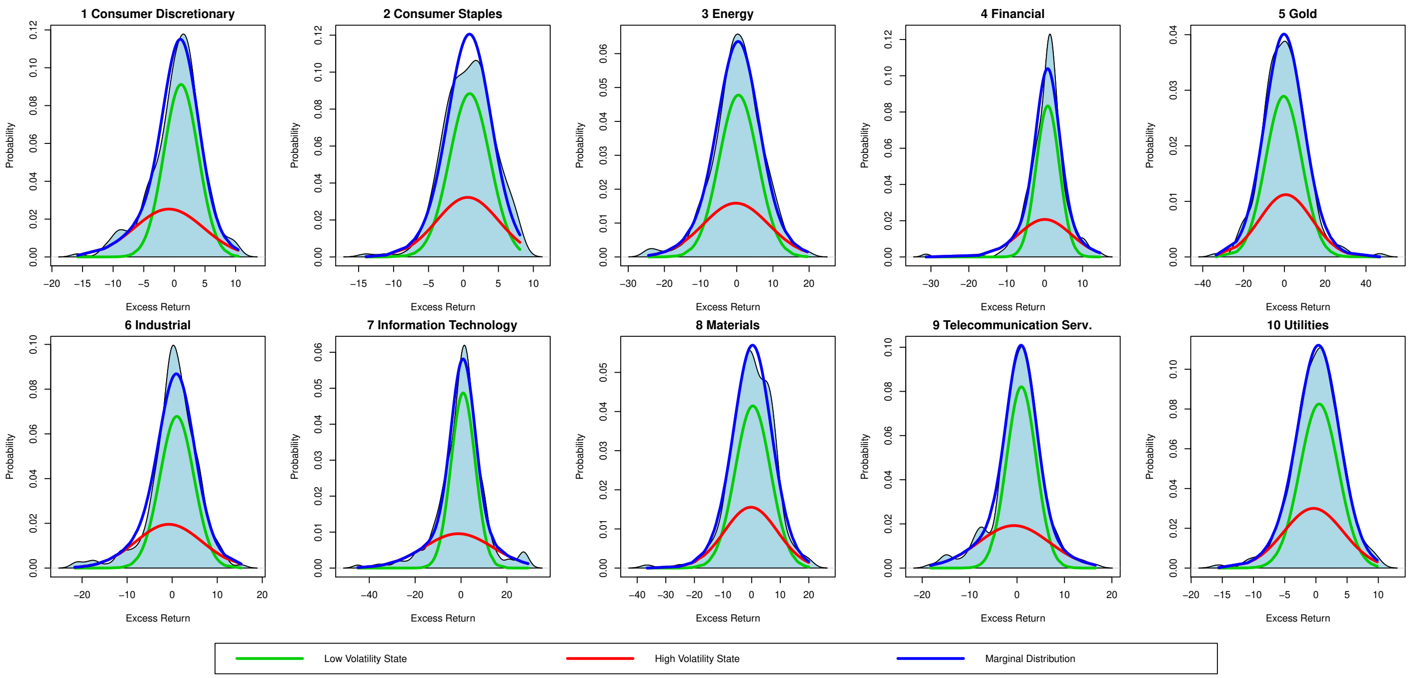

Figure 4.1: Kernel Density Curves of Returns across the TSX and their Stationary Markov-Dependent Mixture of Normal Distributions.

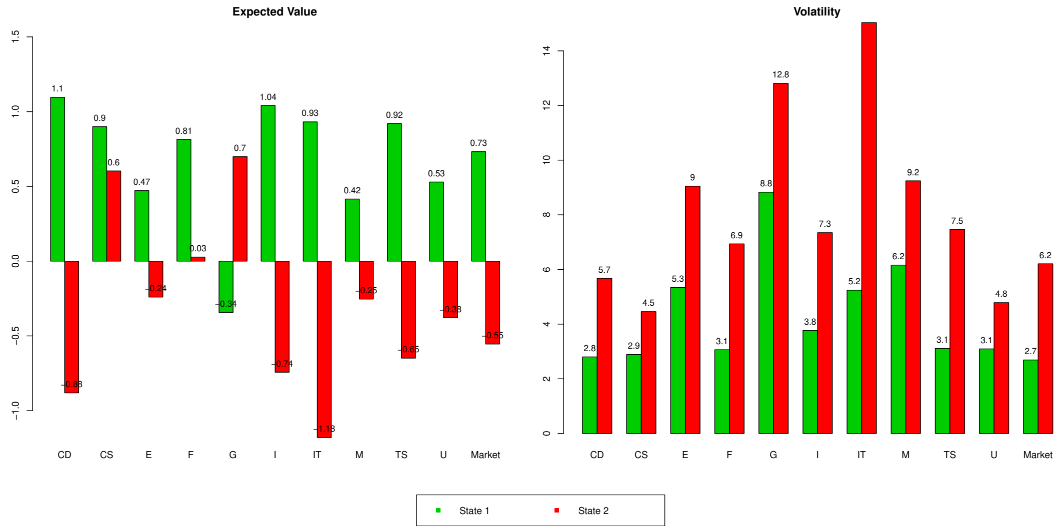

Figure 4.2: Expected Value and Volatility from a First-Order Markov-Dependent Mixture of Multivariate Normal Distributions Fitted to Excess Returns across the TSX, 1998:M1-2017:M12.

all the excess returns are less volatile in state 1 than in state 2. For example, the volatility (or standard deviation) of the market excess return is 2.69 in state 1 and 6.11 in state 2. On the other hand, the expected excess returns across the TSX are higher in state 1 than in state 2, except for the gold sub-industry. For example, in the market, the monthly expected excess returns are .73% in state 1 and -.55% in state 2 while, in the gold sub-industry, they are respectively -.34% and .7% in states 1 and 2. The expected excess returns are positive in state 1 and negative in state 2 in the other sectors, except in the consumer staples and the financial sectors where they are positive over the two states.

$$\begin{aligned} \tilde{\mathbf{\Gamma}} &= \begin{bmatrix} .953 & .047 \\ (43.70) & (2.17) \\ .084 & .916 \\ (2.14)2 & (3.25) \end{bmatrix} \\ \tilde{\mathbf{u}} &= \begin{bmatrix} .640 & .360 \\ (5.54) & (3.12) \end{bmatrix} \end{aligned} \tag{4.9}$$

State 1, the low volatility state, can be labeled as the bull market. The reason is that, in this state, the excepted excess returns are positive across the TSX, except for the gold sub-industry. Consequently, state 2, the high volatility state, can be labeled as the bear market. *Decoding* the states (*i.e.*, for each time period, identifying using the fitted model which of the two states is the most likely) reveals that the bear market include mainly the period ranging from November 2007 to June 2009, which indeed corresponds to the global financial crisis, and the early 2000s, which corresponds to the dot-com crash.

Excess returns are not necessarily positive during bull markets, evenn though their expected values are positive. Between 1998 and 2017, out of 156 months where the TSX is very likely to experience a bull market trend, the market excess returns are positive 101 times. In the gold sub-industry and the materials sector, positive excess returns occur respectively 76 and 82 times. The financial sector followed by the consumer discretionary are the ones that show more positive excess returns during bull markets, respectively 106 and 105 times. Conversely, positive excess returns occur during bear markets, where the market excess returns turn out to be negative only 40 times out of 83 months.

The estimates of the transition probability matrix reported in (4.9) indicate that both state 1 and state 2 are very persistent. The probability of leaving state 1 is only 4.7%. Therefore, state 1, which is identified as the bull market, is the most recurrent and the most persistent of the two states.

Note that all the estimates of the transition probabilities and the unconditional probabilities of the states reported in (4.9) are statistically significant since their z-values, which are displayed in parentheses, are greater than their 5% critical value, which is 1.64.

In Table 4.1, I have reported the estimates of the parameters of the first-order Markov-switching conditional CAPM. These parameters are the measure of performance of Jensen (*i.e.*, the intercept terms or the alphas) and the systematic risks (*i.e.*, the slope parameters or the betas). They are computed out of the component expected values and variance-covariance matrices, using relation (4.7). The adjusted coefficients of determination, $\bar{R}^2$, reported in the last column of this table have been computed after decoding the consecutive states that most likely prevailed on the TSX over the sample period.

All the betas are significantly positive, except those of the gold sub-industry and the utilities sector, in state 2. This means that, according to the first-order Markov-switching model, in the bear market, the overall excess return on the TSX does not significantly explain the excess returns in the gold sub-industry and in the utilities sector. The betas of the financial sector are almost the same, over the two states, .766 in the bull market and .75 in the bear market.

In bull markets, gold and more generally materials along with energy are high-risk assets. Their betas are greater than the market beta, *viz* these estimates, which are roughly equal to 1.5, are greater than 1. For this reason, they should offer the possibility of higher returns, but they deliver negative alphas, which lowers their theoretically appropriate excess returns. In bull markets, the sectors that tend to outperform the market are those having high betas and delivering, at the same time, positive alphas. These sectors are the consumer discretionary, the financial, the industrial, the information technology, and the telecommunication service sectors.

In the bear market, information technology is the only sector that has a beta greater than 1. Furthermore, its alpha is negative. Thus, investing in this sector is unattractive, since the market excess returns and, consequently, the theoretically required excess returns in this sector are very likely to be negative, in bear markets. In bear markets,

Table 4.1: CAPM, Estimates from a Two-State First-Order Markov-Switching Model, TSX, 1998:M1-2017:M12.

Sector State	Performance α		Systematic Risk β		$\bar{R}^2$
	1	2	1	2	
Consumer Discretionary	.705	-.517	.533	.657	.459
	(3.56)	(-1.36)	(9.33)	(6.63)	
Consumer Staples	.697	.703	.276	.180	.053
	(2.75)	(1.82)	(4.10)	(2.01)	
Energy	-.630	.211	1.505	.814	.409
	(-2.14)	(.47) 1	(8.70)	(4.29)	
Financial	.253	.443	.766	.750	.448
	(1.37)	(1.02)	(13.38)	(6.65)	
Industrial	.340	-.262	.958	.868	.522
	(1.44)	(-.56)	(15.45)	(7.24)	
Information Technology	.242	-.239	.942	1.697	.442
	(00)	(-.47)	(8.47)	(7.63)	
Materials	-.735	.204	1.571	.826	.373
	(-3.00)	(.49)	(14.35)	(4.41)	
Gold	-1.438	.935	1.496	.427	.108
	(-3.53)	(2.39)	(7.99)	(1.50)	
Telecommunication Service	.657	-.252	.360	.715	.298
	(2.40)	(-.59)	(4.89)	(5.41)	
Utilities	.241	-.317	.393	.111	.063
	(.91)	(-.82)	(5.64)	(1.12)	
Selection Criteria					
-Log-Likelihood					7 018.33
AIC					14 348.67
BIC					14 890.99

The numbers in parentheses are the z-statistics, which are computed dividing the maximum likelihood estimates by their respective bootstrap standard errors.

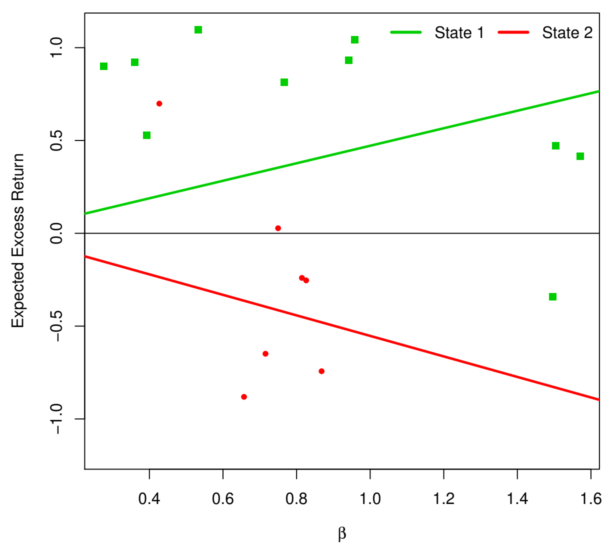

Figure 4.3: State-Dependent Security Market Line, TSX, 1998:M1-2017:M12.

the sectors where the excess returns are expected to be positive (the consumer staples sector and the gold sub-industry) have a low beta and a high alpha.

Figure 4.3 shows the security market lines, which relate the expected excess returns to the betas across the TSX, in both bull and bear markets. The slope parameters of these lines are estimated by ordinary least squares (OLS) regression through the origin. One can see that, in bull markets, the expected excess return tend to increase along with the beta, while the inverse trend is observed in bear markets. This reflects the fact that the slope of the security market line is an estimator of the market expected excess return, which is positive in bull markets and negative in bear markets. Two-sided t-tests performed on the slope parameters of the security market lines reveal that, respectively, they are not significantly different from .73% and -.55%, the actual estimates of the expected excess returns in bull and bear markets, reported in Figure 4.2.

The conditional CAPM switched by a two-state first-order Markov chain explains 52.8% of the variability observed in the excess returns of the industrial sector. This is the highest R^2 in the model. The adjusted coefficient of determination for this sector is the same as the one from an OLS regression (see Table B.1). The $\bar{R}^2$ for the the consumer staples and the financial sectors are slightly lower than those from the one-state OLS regressions. Thus, in terms of explanatory power, the traditional and the conditional CAPM make no difference, for some of the sectors of the TSX. However, as far as the performance of a portfolio is concerned, being able to properly decode the state prevailing in the market is rewarding since it contributes to a better asset allocation (Tu, 2010). Besides, the selection criteria (the AIC and the BIC) both suggest that, as a whole, the conditional CAPM in Table 4.1 is preferable to unconditional CAPM in Table B.1. The selection criteria from the conditional CAPM are the lowest . (For a brief detail on model selection criteria, see Subsection B.2.)

4.2.2 The Second-Order Markov-Switching CAPM

In (4.10) as in (4.9), the estimate of the unconditional probability of being in state 1 is 64%. Besides, the estimates of the conditional probability of remaining in state 1, which are produced using both the first-order and the second-order Markov chains, are almost the same, 95.8%. The estimates of the market excess return and its volatilty produced using these two types of two-state Markov chains are also similar. Therefore, as in the previous case, state 1 here can also be labeled as the bull market. Consequently, states 2 through 4 can be seen as breaking down the bear market into three: its start, its end, and its progress.

$$\tilde{\mathbf{\Gamma}} = \begin{bmatrix} .958 & .042 & .000 & .000 \\ (62.63) & (2.77) & & \\ .000 & .000 & .597 & .403 \\ & & (8.57) & (5.80) \\ .200 & .800 & .000 & .000 \\ (3.85) & (15.35) & & \\ .000 & .000 & .610 & .390 \\ & & (8.62) & (5.52) \end{bmatrix} \tag{4.10}$$

$$\tilde{\mathbf{u}} = \begin{bmatrix} .640 & .135 & .135 & .090 \\ (7.52) & (4.25) & (4.25) & (3.2)2 \end{bmatrix}$$

Figure 4.4 shows that, with the second-order Markov chain, state 1 is characterized by a low volatility in all sectors, with the exception of materials and, in particular, gold. Recall from Figure 4.2 that, according to the first-order Markov-switching model, excess returns in all sectors, with no exception, are less volatile in state 1. As with the first-order Markov-switching model, in state 1, the expected excess returns are positive across the TSX, excepting the gold sub-industry.

The estimate of the unconditional probabilities of the states indicates both states 2 and 3 are equally likely to occur . They occur 13.5% of the time and are more recurrent than state 4. Excess returns are more volatile in state 2 or 3, except for the information technology sector where state 4 is the most volatile. State 2 can be interpreted as an incipient bear market or the market top, state 3 as the market bottom, and state 4 as the progressing or ongoing bear market.

The estimated transition probability matrix in (4.10) indicates that the bull market (state 1) is very persistent. States 2 and 3 have no persistence since they are turning points. When a bear market starts (*i.e.*, when the TSX is in state 2), there is a 40.3% chance that it goes on. The probability the TSX exit an ongoing bear market is high (61%). When the market reaches a bottom, there is only a 20% probability that a bull market resume.

Table 4.2: CAPM, Estimates from a Two-State Second-Order Markov-Switching Model, TSX, 1998:M1-2017:M12.

	Performance α				Systematic Risk β				$\bar{R}^2$
Sector State	1	2	3	4	1	2	3	4	
Consumer Discretionary	.712	-1.463	.344	.577	.494	.739	.788	.395	.489
	(3.56)	(-4.19)	(.95)	(2.28)	(8.92)	(5.72)	(7.26)	(4.04)	
Consumer Staples	.666	-.498	1.208	1.823	.241	.335	.361	-.231	.147
	(2.95)	(-1.87)	(4.06)	(5.87)	(3.92)	(4.64)	(4.34)	(-2.25)	
Energy	-.660	.297	-1.080	.654	1.506	1.466	.804	.237	.470
	(-2.66)	(.79)	(-3.51)	(2.65)	(18.44)	(8.71)	(5.48)	(1.74)	
Financial	.272	1.239	-1.176	1.764	.759	.657	.998	.340	.530
	(1.47)	(4.15)	(-3.61)	(6.35)	(14.11)	(6.27)	(7.65)	(2.59)	
Industrial	.409	-.791	-.519	.469	.957	1.138	.953	.478	.555
	(1.76)	(-2.12)	(-1.37)	(1.49)	(15.07)	(8.12)	(7.62)	(2.93)	
Information Technology	.300	.655	-.026	-.827	.914	1.506	1.332	2.586	.483
	(.98)	(1.93)	(-.07)	(-1.04)	(8.40)	(6.24)	(7.53)	(6.28)	
Materials	-.855	.189	-.173	-.489	1.587	1.060	1.079	.037	.419
	(-4.33)	(.55)	(-.47)	(-1.98)	(16.10)	(8.81)	(7.18)	(.35)	
Gold	-1.642	.755	1.313	-2.556	1.525	.469	.760	-.307	.139
	(-5.88)	(3.55)	(4.89)	(-12.01)	(8.67)	(2.47)	(4.36)	(-1.92)	
Telecommunication Service	.707	-.177	.883 -	.892	.361	.712	.670	.695	.294
	(2.75)	(-.59)	(3.15)	(-2.50)	(5.65)	(5.46)	(5.95)	(2.93)	
Utilities	.299	-.633	-.868	-.227	.375	.590	.175	-.462	.203
	(1.16)	(-1.96)	(-3.13)	(-.67)	(5.74)	(4.67)	(4.29)	(-4.26)	
Selection Criteria									
-Log-Likelihood									6 841.73
AIC									14 307.45
BIC									15 392.11

The numbers in parentheses are the z-statistics, which are computed dividing the maximum likelihood estimates by their respective bootstrap standard errors.

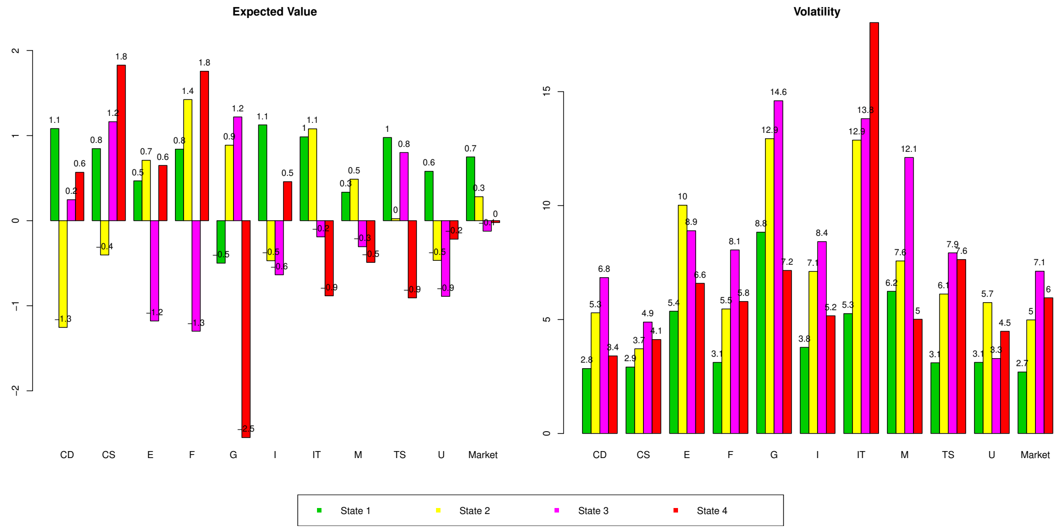

Figure 4.4: Expected Value and Volatility from a Second-Order Markov-Dependent Mixture of Multivariate Normal Distributions Fitted to Excess Returns across the TSX, 1998:M1-2017:M12.

Figure 4.4 shows that the lowest monthly expected excess return on the market, which is -.12%, is observed when the market reaches a bottom (*i.e.*, when the TSX is in state 3). At that moment, in the gold sub-industry, the expected return reaches its higher level, which 1.22%. This confirms that gold is indeed a defensive asset.

Table 4.2 shows the estimates of the parameters of the conditional CAPM switched by a two-state second-order Markov chain. The estimates of the betas for state 1 are close to those in Table 4.1. Thus, the energy and the materials sectors, and the gold sub-industry have a beta that is greater than 1. Their alphas are still the only ones to be negative. When the market reaches a high (*i.e.*, in state 2), the betas in the energy and the materials sectors are still greater than 1, but their alphas become positive, which raises their theoretically required returns above that of the market. Unlike the consumer discretionary, the consumer staples, and the industrial sectors, the gold sub-industry starts outperforming the market because of the rise in its alpha.

Whereas, in the conditional CAPM switched by the first-order Markov chain, the betas of the gold sub-industry and the utilities sector are positive but not statistically significant in bear markets, in the second-order Markov-switching model, these parameters become negative and statistically significant. The beta of the consumer staples, which is significantly positive but low in the former model, also becomes significantly negative in ongoing bear markets (*i.e.*, in state 4). Again, it transpires that the consumer discretionary, the gold, and the utilities are defensive assets. The beta of the materials is still positive but is no longer statistically significant, in this state.

According to the data, the conditional CAPM switched by the second-order Markov chain is more likely than the one switched by the first-order chain and, in addition, has the lowest AIC. As an example, the second-order Markov chain has raised the R^2 of the utilies sector from 7.4% to 22.6%. On the other hand, the BICS indicate that the first-order Markov-switching model fits better the data.

5 The Accounting of the Market Return

The purpose of this investigation is to measure how the TSX is exposed to the performance of each of its sectors. Following Sharpe (1992), I have estimated the following linearly constrained model

$$\begin{aligned} R_{mt} &= \mathbf{R}_{st}^{'}\mathbf{\Omega}_{S_t} + \varepsilon_t \\ \mathbf{\Omega}_{S_t}^{'}\mathbf{1} &= 1. \end{aligned} \tag{5.1}$$

The variable R_{mt} in (5.1) denotes the market return at time t, the column vector $\mathbf{R}_{st}$ denotes the returns at time t across the sectors of the TSX, and ε_t is the market news. The $N \times 1$ vector of parameters $\mathbf{\Omega}$ denotes the sensitivities of the TSX to the performance of its sectors. Since the second equation of (5.1) constrains these sensitivity parameters to sum to unity, $\mathbf{\Omega}$ can also be interpreted as the list of the shares of sectors in the market return. Relation (5.1) can be seen as resulting from a weighted average of the CAPM in (B.26) and (B.27).

My contribution to (5.1) is to allow heteroskedasticity in the market news (*i.e.*, $\varepsilon_t = z_t\sigma_t$) and a discrete shift in the state prevailing in the market. Both σ_t, the market volatility, and, $\mathbf{\Omega}$, the shares of sectors in the market return, are switched over time by a discrete Markov chain S_t. I have modeled the market volatility building on the GARCH processes of order 1 and 1 put forth by Bollerslev (1986) and Glosten, Jagannathan, and Runkle (1993).

The symmetric GARCH(1,1) process of Bollerslev posits that volatility, measured by the conditional variance of the market news, is time-varying and depends linearly on both the volatility and the square of the market news of the previous period.

$$\sigma_t^2 = \nu + \psi\sigma_{t-1}^2 + \eta\varepsilon_{t-1}^2 \tag{5.2}$$

The conditions $\nu, \psi, \eta > 0$ and $\psi + \eta < 1$ ensure the variance in (5.2) is positive and stationary. The unconditional variance of ε_t implied by (5.2) is

$$\begin{aligned} \sigma^2 &= \nu + \psi\sigma^2 + \eta\mathrm{E}(\varepsilon^2) \\ &= \frac{\nu}{1-\psi-\eta}, \end{aligned}$$

which means $\psi + \eta$ measures the persistence of volatility.

Glosten, Jagannathan, and Runkle model asymmetry by adding a dummy variable to (5.2) to distinguish states where the market news is negative from the states where it is positive.

$$\sigma_t^2 = \nu + \psi\sigma_{t-1}^2 + \eta\varepsilon_{t-1}^2 + \tau\varepsilon_{t-1}^2\mathrm{I}_{t-1}, \tag{5.3}$$

where the dummy variable I_{t-1} equals one when the market news at time $t-1$ is negative and zero otherwise. For $\tau > 0$, relation (5.3) implies that a decrease in stock prices tend to increase subsequent volatility by more than an increase in stock prices of the same magnitude would. This negative relationship between current return and subsequent volatility is termed a leverage effect. Relation (5.3) is named, after their authors, GJR-GARCH. To ensure the non-negativity and the stability of the variance, the restriction $\psi + \eta + \frac{1}{2}\tau < 1$ is imposed on its parameters (for more details, see Ling and McAleer, 2002).

Instead of using a dummy variable or a threshold to model the asymmetry in the response of the market to negative and positive news, I allow the parameters of the GARCH(1,1) process to depend on S_t, an underlying and unobserved discrete Markov chain, as Gray (1996) and Bauwens, Preminger, and Rombouts (2010), among others, have done.

$$
\begin{aligned}
\varepsilon_t &= z_t \sigma_t, \quad z_t \sim \mathcal{N}(0,1) \\
\sigma_t^2 &= \nu_{S_t} + \rho_{S_t} \sigma_{t-1}^2 + \phi_{S_t} \varepsilon_{t-1}^2
\end{aligned}
\tag{5.4}
$$

The conditions $\nu_{S_t}, \psi_{S_t}, \eta_{S_t} > 0$ and $\psi_{S_t} + \eta_{S_t} < 1$ are maintained to avoid explosive variances.

5.1 The Method of Estimation

I have estimated (5.1) and (5.4) by maximum likelihood assuming, in turn, that their parameters are switched by a two-state first-order Markov chain and a two-state second-order Markov chain. The first- and second-order Markov chains are respectively described by relations (3.1) and (3.5).

To compute the likelihood, I have assumed that the conditional distribution of R_{mt} given $\mathbf{R}_{st}$ and S_t is normal. Thus, in the case of a two-state first-order Markov chain, the density of an observation r_{mt} given $\mathbf{r}_{st}$ and s_{t-1} is

$$
\begin{aligned}
p\left(r_{mt} | \mathbf{r}_{st}, s_{t-1}; \boldsymbol{\theta}\right) &= \sum_{s_t=1}^{2} \gamma_{s_{t-1}, s_t} p\left(r_{mt} | \mathbf{r}_{st}, s_t; \boldsymbol{\theta}_{s_t}\right) \\
p\left(r_{mt} | \mathbf{r}_{st}, s_t; \boldsymbol{\theta}\right) &= \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[-\frac{\left(r_{mt} - \mathbf{r}_{st}^{'} \boldsymbol{\Omega}_{s_t}\right)^2}{2\sigma_t^2}\right],
\end{aligned}
\tag{5.5}
$$

with $\boldsymbol{\theta} = [\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \nu_1, \nu_2, \psi_1, \psi_2, \eta_1, \eta_2, \boldsymbol{\Omega}_1', \boldsymbol{\Omega}_2']$.

Similarly, in the case of a two-state second-order Markov chain, the density of an observation r_{mt} given $\mathbf{r}_{st}$, s_{t-1}, and s_{t-2} is

$$\begin{aligned} p\left(r_{mt}|\mathbf{r}_{st}, s^*_{t-1}; \boldsymbol{\theta}^*\right) &= \sum_{s_t=1}^{4} \gamma^*_{s^*_{t-1}, s^*_t} p\left(r_{mt}|\mathbf{r}_{st}, s^*_t; \boldsymbol{\theta}^*_{s_t}\right) \\ p\left(r_{mt}|\mathbf{r}_{st}, s^*_t; \boldsymbol{\theta}^*\right) &= \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[-\frac{\left(r_{mt} - \mathbf{r}^{'}_{st}\boldsymbol{\Omega}_{s^*_t}\right)^2}{2\sigma_t^2}\right], \end{aligned} \tag{5.6}$$

with $\boldsymbol{\theta}^* = [\gamma^*_{11}, \gamma^*_{12}, \ldots, \gamma^*_{44}, \nu^*_1, \ldots, \nu^*_4, \psi^*_1, \ldots, \psi^*_4, \eta^*_1, \ldots, \eta^*_4, \boldsymbol{\Omega}^{*\prime}_1, \ldots, \boldsymbol{\Omega}^{*\prime}_4]$. As shown in Section 3, the variable s^*_t, which is a first-order Markov chain, results from the combination of two consecutive realizations of a second-order Markov chain.

It follows from (5.5) that the likelihood of observations $r_{m1}, \ldots, r_{mT}$ is

$$\begin{aligned} L\left(\boldsymbol{\theta}\right) &= \prod_{t=1}^{T} \sum_{s_t=1}^{2} \gamma_{s_{t-1}, s_t} p\left(r_{mt}|\mathbf{r}_{st}, s_t; \boldsymbol{\theta}_{s_t}\right) \\ &= \mathbf{u} \prod_{t=1}^{T} \boldsymbol{\Gamma}\mathbf{P}\left(r_{mt}|\mathbf{r}_{st}\right)\mathbf{1}', \end{aligned} \tag{5.7}$$

where the vector $\mathbf{u}$, the stationary distribution of s_t, is defined by relation (3.4), $\boldsymbol{\Gamma}$ is the matrix of transition probabilities, and $\mathbf{P}\left(r_{mt}|\mathbf{r}_{st}\right)$ is a diagonal matrix whose j-th diagoal element corresponds to $p\left(r_{mt}|\mathbf{r}_{st}, s_t = j; \boldsymbol{\theta}\right)$. In the case of a second-order Markov chain, the likelihood of the observations is defined in a similar way. I have estimated the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ by maximizing directly the log-likelihoods, using the base function *nlm* of the software R. I have computed the z-statistics of the parameters after performing some bootstrapping. This has consisted in sampling r_{mt} $(t = 1, \ldots, T)$ one thousand times from the state-dependent normal distributions and using the observed $\mathbf{r}_{st}$ to produce new estimates.

5.2 The Findings

The dataset I have used to estimate (5.1) and (5.4) is described in Appendix A. Table B.2 displays the results of the augmented Dickey-Fuller (ADF) tests for stationarity. (The ADF test is explained in Appendix B.3.) These statistical tests indicate that all the returns are stationary, which could prevent from obtaining misleading results in estimating the models.

The exploratory analysis of the data reveals a high correlation among some of the explanatory variables. The correlation coefficient between the returns in the consumer discretionary and the industrial sectors is .71 (see Table A.1). For this reason, I have droped from (5.1) the returns of the industrial sector to estimate a restricted model. Table 5.1 presents some selection criteria from estimating both the unrestricted and restricted models assuming that heteroskedasticity either follows a Markov-switching GARCH(1,1) process or is statewise (*i.e.*, the variance is constant only within the states).

Table 5.1: Accounting for the Market Return, Some Selection Criteria from the Estimation of Two-State First-Order and Second-Oder Markov-Switching Models, TSX, 1998:M1-2017:M12.

	GARCH(1,1)		Statewise Heteroskedasticity	
Sector	Unrestricted	Restricted	Unrestricted	Restricted
First-Order Markov-Switching Model				
-Log-Likelihood	196.361	231.776	243.484	274.955
AIC	448.722	515.552	530.968	589.909
BIC	546.063	605.940	607.450	659.439
Number of parameters	28	26	22	20
Second-Order Markov-Switching Model				
-Log-Likelihood	139.880	189.659	166.005	233.840
AIC	387.761	479.32	420.011	547.680
BIC	575.490	653.140	572.975	686.739
Number of parameters	54	50	44	40

The two unrestricted Markov-switching GARCH models outperform the others. Their estimates are presented in Table 5.2.

All the models have been estimated without the returns of the health care sector. For this reason, the binding constraint in relation (5.1) has been relaxed and replaced with $\mathbf{\Omega}'_{S_t}\mathbf{1} < 1$. Therefore, the shares of the health care sector turn out to be the difference between one and the sum of the estimated shares, $\tilde{\mathbf{\Omega}}'_{S_t}\mathbf{1}$, with $\tilde{\mathbf{\Omega}}_{S_t}$ being now an $(N-1) \times m$ matrix. Also, recall from Appendix A that the financial sector includes the real estate.

5.2.1 The First-Order Markov-Switching Sharpe Model

The first block of Table 5.2 displays the estimates of the unrestricted Sharpe model produced using the two-state first-order Markov-switching GARCH. This models accounts for 90.3 % of the variability observed in the market return. The $\bar{R}^2$ computed after decoding the states is .899. According to this model, 67.5% of the time, the TSX is in state 1 and, the rest of the time, it is in state 2. These unconditional (stationary) probabilities are significantly greater than zero, since their z-statistics are greater than the 5% critical value, which is 1.64. The two states are very persistent. The probabilities that the TSX remain in state 1 and in state 2 are respectively 94% and 87.5%.

The parameters of the GARCH processes are significantly positive, except the intercepts. Volatility is very persistent over the two states. As a matter of fact, the sum of the parameters ψ and η is .998 in state 1 and .999 in state 2. The response of the variance of the market return to a change in the lagged squared residuals (the parameter η) is higher in state 2 than in state 1, which indicates asymmetry in the GARCH process.

In both states, the financial sector accounts for the highest share of the market returns, 27.6% in state 1 and 23.4% in state 2. In state 1, which is the low volatility state, the financial sector followed by the energy and the materials explain 67.1% of the market returns. In state 2, the financial sector followed by the consumer discretionary

Table 5.2: Estimates of the Contribution of the Sectors of the TSX to the Market Return Using Two-State First-Order and Second-Order Markov-Switching Models, 1998:M1-2017:M12.

	First-Order Markov		Second-Order Markov			
Sector State	1	2	1	2	3	4
Mean Equation						
Consumer Discretionary	.056	.160	.042	.138	.149	.092
	(3.64)	(4.85)	(7.71)	(6.71)	(70.71)	(14.12)
Consumer Staples	.028	.065	.034	.075	.130	.063
	(2.48)	(2.67)	(6.35)	(50.02)	(50.20)	(8.98)
Energy	.230	.106	.231	.085	.174	.104
	(33.70)	(5.70)	(48.42)	(76.92)	(52.38)	(13.22)
Financial	.276	.234	.280	.128	.037	.240
	(32.24)	(9.20)	(47.83)	(81.10)	(14.16)	(30.77)
Industrial	.091	.047	.094	.201	.070	.093
	(8.68)	(1.82)	(16.67)	(110.12)	(27.63)	(11.98)
Information Technology	.055	.102	.053	.161	.090	.099
	(10.46)	(7.66)	(12.57)	(152.44)	(25.58)	(14.40)
Materials	.171	.155	.168	.036	.181	.156
	(34.03)	(10.29)	(36.76)	(36.49)	(47.22)	(20.65)
Telecommunication Service	.035	.107	.037	.083	.052	.114
	(4.99)	(4.86)	(6.60)	(55.25)	(17.77)	(16.31)
Utilities	.046	.018	.047	.080	.107	.021
	(4.42)	(.72)	(8.48)	(55.95)	(42.73)	(2.87)
Health Care	.012	.006	.014	.013	.010	.018
Variance Equation						
Intercept ν	.004	.000	.005	.000	.000	.017
	(1.14)	(.53)	(2.02)	(5.24)	(42.20)	(2.06)
Lagged Variance ψ	.584	.455	.570	.004	.000	.371
	(8.17)	(2.35)	(7.95)	(29.31)	(24.29)	(3.11)
Lagged Squared Residuals η	.414	.544	.430	.000	.084	.629
	(5.13)	(2.71)	(6.00)	(3.93)	(2.67)	(5.26)
Markov Chain						
Stationary Distribution	.675	.325	.644	.033	.033	.291
	(7.05)	(3.39)	(5.54)	(3.53)	(3.53)	(2.68)
Transition Probability Matrix	$\begin{bmatrix} .940 & .060 \\ .125 & .875 \end{bmatrix}$		$\begin{bmatrix} .949 & .051 & .000 & .000 \\ .000 & .000 & .001 & .999 \\ .999 & .001 & .000 & .000 \\ .000 & .000 & .112 & .888 \end{bmatrix}$			

The numbers in parentheses are the z-statistics, which are computed dividing the maximum likelihood estimates by their respective bootstrap standard errors.

and the materials account for 54.9% of the market returns.

In state 1, health care followed by consumer staples, telecommunication service, and utilities account for the lowest share of the market returns, 12.1% altogether. In state 2, the high volatility state, the health care, the utilities, and the industrial sectors generate the lowest share (7.2% altogether) and represent a good hedge against market downturns.

5.2.2 The Second-Order Markov-Switching Sharpe Model

The second block of Table 5.2 displays the maximum likelihood estimates of the unrestricted Sharpe model with a GARCH process. As it appears in Table 5.1, both the log-likelihood and the AIC indicate that the unrestricted second-order Markov-switching Sharpe model provides a better fit for the data. However, after decoding the states, its determination coefficient, R^2, which turns out to be 85.3% indicates that its explanatory power is lower. The $\bar{R}^2$ that results from decoding the states is .825.

Some of the estimates produced assuming a second-order Markov process are similar to those from the first-order Markov-switching model. As a matter of fact, the estimates of the coefficients for state 1 and state 4 are respectively close to those for the low and the high volatility states of the first-order Markov process. Thus, state 1 represents, at least, three consecutive months of low volatility, while state 4 represents a minimum of three consecutive months of high volatility. In-between, state 2 is a turning point from the low to the high volatility state and state 3 is a period of recovery after episodes of high volatility.

The stationary distribution, which gives the unconditional probabilities of being in each of the four states, indicates that state 1 is the most frequent one. The unconditional profitability that the TSX be in any of the two low volatility states (*i.e.*, either state 1 or 3) is 67.7%. This estimate is closer to the unconditional probability of state 1 in the first-order Markov-switching Sharpe model, which equals 67.5%.

The estimates of the variance equation show evidence supporting the leverage effect. As a matter of fact, there is asymmetry in the GARCH process as the response of the variance of the market return to a change in the lagged squared residuals (the parameter η) is higher when the TSX enters the high volatility state (*i.e.*, state 4). In state 4 (which represents, at least, three consecutive months of high volatility), this response parameter is equal to .629, whereas in state 1 (which represents, at least, three consecutive months of low volatility) it equals .43.

Altogether, the financial, the energy, and the materials sectors account for 67.9% and 50% of the market returns, respectively in states 1 and 4. In state 2 (which is the turning point from the low to the high volatility state) the industrial sector followed by the information technology and the consumer discretionary explain half of the returns on the TSX whereas they only account for 19% in the low volatility state (state 1).

All the cyclical sectors of the TSX do not behave the same way. The energy sector tends to account for a greater share of the market return when the TSX enters any of the two low volatility states (*i.e.*, either state 1 or 3) than when it enters any of the two high volatility states. On the other hand, the shares of the information technology and the telecommunication service tend to be higher over any of the two high volatility states.

There are significant differences in the estimates of the contribution of most sectors to the market return, over the various states. An implication of this evidence is that, to match or outperform the TSX, the managers of index exchange traded funds should change their allocation of assets depending on the state prevailing or expected in the market. In this accounting exercise, I have used historical returns to estimate simultaneously the shares of the sectors and the state prevailing in the market. An alternative approach proposed by Kritzman, Page, and Turkington (2012) is to identify the states using rather *financial turbulence*, inflation, and economic growth time series data. The financial turbulence is a multivariate measure of distance computed as follows $d_t = (\mathbf{r}_t - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{r}_t - \boldsymbol{\mu})'$.

6 The Exponential Autoregressive Model

Stock returns are not white noises, (*i.e.* a serially uncorrelated random variable with a zero mean and a constant variance). Using both the daily and the weekly returns on the S&P composite index, LeBaron (1992) observes that some of the most volatile periods are also those with the lowest serial correlation. He also observes that the serial correlation of the daily returns on the Dow Jones is inversely related to its conditional variance. Then, he tests the hypothesis of a negative relationship between these two statistics by fitting a variant of the exponential autoregressive (AR) model of Ozaki (1980) to these data, The exponential AR (p) model relates the current return to its previous realization as follows:

$$r_t = \left[\varphi_1 + \varpi_1 \exp\left(-\upsilon r_{t-1}^2\right)\right] r_{t-1} + \cdots + \left[\varphi_p + \varpi_p \exp\left(-\upsilon r_{t-1}^2\right)\right] r_{t-p} + \varepsilon_t, \tag{6.1}$$

where the single scale parameter $\upsilon > 0$ is set so as to minimize the AIC (Haggan and Ozaki, 1981). The estimates of the exponential AR coefficients ($\varphi_i,\ \varpi_i,\ i = 1, ..., p$) depend on the value assigned to υ.

The exponential AR model can be seen as a state-dependent model or a smooth threshold model, since if $\varpi_1, \ldots, \varpi_p$ are all zero or if the absolute value of r_{t-1} is large, (6.1) becomes a linear autoregressive model with parameters $\varphi_1, \ldots, \varphi_p$.

LeBaron (1992) approximates the squared lagged return, r_{t-1}^2, in (6.1) with its conditional variance, σ_t^2. This establishes a direct connection between the conditional variance and the autocorrelation coefficients.

$$r_t = \delta + \left[\varphi + \varpi \exp\left(-\upsilon \sigma_t^2\right)\right] r_{t-1} + \varepsilon_t \tag{6.2}$$

For $\varpi > 0$, model (6.2) suggests an inverse relation between the conditional variance and the autocorrelation coefficients. LeBaron (1992) models the conditional variance of the returns using the GARCH process of Bollerslev (1986). His empirical investigations show evidence of changing correlations in the daily and the weekly returns on the S&P composite index, as well as in the daily returns on the value-weighted index from the

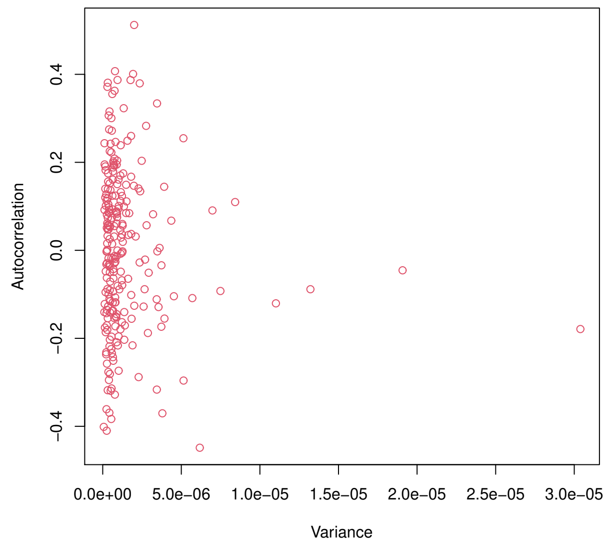

Figure 6.1: Daily Variance and Autocorrelation of Daily Returns, TSX, 1998M1-2017:M12.

Center for Research in Securities Prices and on the Dow index. The estimates of ϖ turn out to be significantly positive, for these index returns.

Episcopos (1996) undertakes the same investigation across the TSX. But, unlike LeBaron (1992), he takes into account the asymmetry in volatility using the exponential GARCH model of Nelson (1991). He finds a negative relationship between the autocorrelation and volatility across the TSX, except for the sector of paper and forest product and for the sector of transportation.

6.1 The Model

Figure 6.1 relates the variance of the daily returns on the TSX to their autocorrelation. These daily statistics are computed within consecutive monthly time frames defined over the period 1998-2017. The scatter plot shows that the highest autocorrelation coefficients are associated to the periods where volatility is the lowest. Conversely, the persistence (*i.e.* the autocorrelation) of the returns is low when volatility is high. However, the scatter plot reveals also an important cluster of low autocorrelation coefficients associated to the periods of low volatility. This latter observation suggests that the econometric model (6.1) originally proposed by Ozaki (1980), which is independent of the variance, might be more appropriate for the TSX than its variant proposed by LeBaron (1992), *i.e.* the model (6.2).

To introduce heteroskedasticity into (6.1), I propose the following variant

$$r_t = \left[\varphi + \varpi \exp\left(-\upsilon r_{t-1}^2\right)\right] r_{t-1} + z_t \sigma_{s_t}, \quad z_t \sim \mathcal{N}(0,1). \tag{6.3}$$

In (6.3), the innovations (or market news) and, consequently, the volatility are state-dependent The variable s_t is a first-order Markov chain that assumes the value 1 in the low volatility state (bull market) and 2 in the high volatility state (bear market). This formulation is constituent with the possibility of observing at times low persistence in the returns during a period of low volatility, as it appears in Figure 6.1. According to (6.3), the autocorrelation is low during a bull market, when the returns are extremely high.

In relation (6.3), the volatility is constant within each of the states of the market. To take into account asymmetry in the volatility, this assumption could be replaced with a Markov-switching GARCH process, which gives

$$\begin{aligned} r_t &= \left[\varphi + \varpi \exp\left(-\upsilon r_{t-1}^2\right)\right] r_{t-1} + z_t \sigma_t, \quad z_t \sim \mathcal{N}(0,1) \\ \sigma_t^2 &= \nu_{s_t} + \rho_{s_t} \sigma_{t-1}^2 + \phi_{s_t} \varepsilon_{t-1}^2. \end{aligned} \tag{6.4}$$

The likelihood of observing models (6.3) and (6.4) are defined as follows:

$$\begin{aligned} L(\boldsymbol{\vartheta}) &= \sum_{s_0=1}^{2} u_{s_0} \prod_{t=1}^{T} p\left(r_t | r_{t-1}, s_{t-1}; \boldsymbol{\vartheta}\right) \\ p\left(r_t | r_{t-1}, s_{t-1}; \boldsymbol{\vartheta}\right) &= \sum_{s_t=1}^{2} \frac{\gamma_{s_{t-1},s_t}}{\sqrt{2\pi\sigma_{s_t}^2}} \exp\left[-\frac{r_t - \varphi r_{t-1} - \varpi \exp\left(-\upsilon r_{t-1}^2\right) r_{t-1}}{2\sigma_{s_t}}\right]^2 \\ \text{with } \boldsymbol{\vartheta} &= \left[\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \varpi, \sigma_1, \sigma_2, \upsilon, \varphi\right] \end{aligned} \tag{6.5a}$$

$$\begin{aligned} L(\boldsymbol{\theta}) &= \sum_{s_0=1}^{2} u_{s_0} \prod_{t=1}^{T} p\left(r_t | r_{t-1}, s_{t-1}; \boldsymbol{\theta}\right) \\ p\left(r_t | r_{t-1}, s_{t-1}; \boldsymbol{\theta}\right) &= \sum_{s_t=1}^{2} \frac{\gamma_{s_{t-1},s_t}}{\sqrt{2\pi\sigma_{s_t}^2}} \exp\left[-\frac{r_t - \varphi r_{t-1} - \varpi \exp\left(-\upsilon r_{t-1}^2\right) r_{t-1}}{2\sigma_t}\right]^2 \\ \sigma_t^2 &= \nu_{s_t} + \rho_{s_t} \sigma_{t-1}^2 + \phi_{s_t} \varepsilon_{t-1}^2 \\ \text{with } \boldsymbol{\theta} &= \left[\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \nu_1, \nu_2, \varpi, \rho_1, \rho_2 . \sigma_1, \sigma_2, \upsilon, \varphi, \psi_1, \psi_2\right], \end{aligned} \tag{6.5b}$$

where γ_{s_{t-1},s_t} is the transition probability of the Markov chain defined in (3.1) and u_{s_0} its unconditional probability defined in (3.3).

6.2 The Findings

Table 6.1 reports the OLS estimates of (6.1) and the maximum likelihood estimates of (6.5). The value of the scale parameter υ that minimizes the AIC of the exponential AR model (6.1) is 18 860.3. This value is much lower than 693 174.1, the inverse of the

Table 6.1: A Simple and Two Markov-Switching Exponential AR Models of Order One Fitted to Daily Returns on the TSX, 1998-2017

	Simple	Markov-Switching			
	Model (6.1)	Model (6.3)		Model (6.4)	
Mean Equation					
φ	-.438	-.395		-.329	
	(-9.43)	(-11.01)		(-9.90)	
ϖ	.560	.478		.399	
	(9.90)	11.04		(10.02)	
υ	18 860.29	18 860.29		18 860.29	
Volatility					
σ	.001	.001	.002	.001	.174
ν				.000	.000
ρ				.961	.925
ψ				.026	.075
Markov Chain					
Stationary Distribution		.751	.249	.790	.210
Transition Probability		$\begin{bmatrix} .990 & .01 \\ .029 & .971 \end{bmatrix}$		$\begin{bmatrix} .972 & .028 \\ .894 & .106 \end{bmatrix}$	
Selection Criteria					
AIC	-53 794.41	-55 662.52		-56 162.38	
AIC	-53 774.83	-55 623.35		-56 084.03	

The numbers in parentheses are the t-statistics and z-statistics from the bootstrapping.

variance (also known as precision parameter) that is often used. I have also set υ at 18 860.3 while maximizing (6.5a) and (6.5b), to make the estimates comparable.

The OLS estimate of the parameters φ and ϖ are statistically significant (*i.e.*, the absolute value of their t-ratios are greater than their 5% critical value, which is 1.64). The estimates from the two Markov-switching models are also significantly positive. The parameter φ is the autocorrelation when the returns are either extremely high (generally during bull markets) or extremely low (generally during bear markets). The three estimates of this parameter are all negative. As for the three estimates of ϖ, they are all positive and greater than the absolute value of the estimates of φ, which means that when the returns are low or almost nil the autocorrelation $\varphi - \varpi$ is positive.

The selection criteria indicate that the two Markov-switching models provide a better fit to the data than the simple exponential AR model. They have the lowest AIC and BIC. Besides, the Markov-switching GARCH process turns out to be a better alternative to assuming that the volatility is constant within each of the two states of market. The Markov-switching GARCH process shows evidence of asymmetry in the volatility, since the parameter ψ, which is the response of the variance to the lagged squared residuals, is two times higher in the bear market (.075 versus .026).

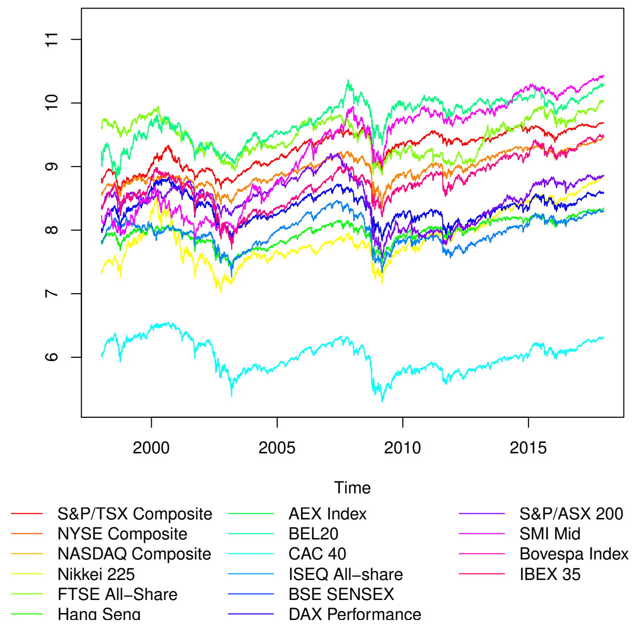

Figure 7.1: The Natural Logarithm of some Major Stock Exchange Benchmark Indices, 1998-2017 (daily)

7 The Cointegration of International Stock Markets

Figure 7.1 plots the natural logarithm of some stock market benchmarks. (These indices are described in Appendix A.) It reveals a synchronization in the movement of the stock price indices across the exchanges. This co-movement could be explained by the fact that the global prices of commodities, especially crude oil, and the state of the US economy have become some of the important determinants of asset prices (Kose, Claessens, and Terrones, 2011). The increasing expansion of the activities of companies abroad through trade, foreign portfolio investment and foreign direct investment, as well as their listings on exchanges in the host countries also play a role in this co-movement.

Most largest companies on the TSX also have a secondary listing on other exchanges and, the other way around, there are companies listed on other exchanges that have a secondary listing on the TSX. They are called interlisted companies. About one-fifth of the companies on the TSX are interlisted. Most of them also trade on other major exchanges, such as the New York Stock Exchange (NYSE) and the National Association of Securities Dealers Automated Quotations (NASDAQ), or on over-the-counter marketplaces.

Differences in the prices of interlisted stocks could create arbitrage opportunities, *i.e.* the possibility of making easy money by buying these stocks at a low price on one exchange and, simultaneously, selling them at a higher price on another exchange.

Advances in information technology systems enable the real-time access to stock prices across exchanges. To prevent arbitrage in the long-run, the changes in the prices of interlisted stocks would tend to synchronize and so would the prices of stocks or assets that are similar across exchanges. This would result in a cointegration, *i.e.* a long-run equilibrium relationship, between the stock price indices across exchanges.

There are two conditions for cointegration between exchanges. First, their stock price indices are all trended, but their first differences (*i.e.* the returns) are stationary. Second, there is, at least, a linear combination of their stock price indices that is stationary. A way of testing for cointegration between exchanges is to perform the ADF test for stationarity on their stock price indices and on the residuals obtained from regressing one of these indices on the others, as described in Appendix B.3. This way of testing for cointegration is referred to as the augmented Engle-Granger test (Engle and Granger, 1987). Table B.3 presents the results of the augmented Engle-Granger (AEG) test for bivariate cointegration between the TSX and 15 other major exchanges. Out of this sample of 15 exchanges, only 2 are found to be cointegrated with the TSX: the Hong Kong Stock Exchange and the Bombay Stock Exchange.

An alternative way of testing for cointegration is by implementing the Johansen (1988) procedure, as Narayan and Smyth (2004, 2005) and Khan (2011), among others, have done. Narayan and Smyth (2005) find no pairwise cointegration between stock prices in New Zealand and stock prices in either Australia or any of the G7 countries, which are Canada, France, Germany, Italy, Japan, the United Kingdom (UK) and the United States (US). On the other hand, they find some evidence of cointegration between the Australian stock market and the stock markets in the UK or in Canada (Narayan and Smyth, 2004). Khan (2011) tests for bivariate cointegration between the US and each of the stock prices in a sample of 22 developing and developed countries (which include Canada). He only finds cointegration between the US and the Netherlands.

Then, following Gregory and Hansen (1996), Narayan and Smyth (2004, 2005) and Khan (2011) allow for a structural change in the bivariate regressions and perform the test for cointegration on their residuals. By doing so, Narayan and Smyth end up finding that the New Zealand and the US stock markets are cointegrated. They also find cointegrating relations between the Australian stock market and stock markets in Canada, Japan, and Italy. The number of stock markets that are cointegrated with the US rise from 1 to 16, in the study of Khan.

7.1 The Model

Apparently, nothing in Figure 7.1 suggests a structure break in the stock market benchmarks or their linear combinations. Instead of assuming a structural break in the data as the test of Gregory and Hansen (1996) suggests, I propose another approach which consists in estimating switching regressions and identifying the recurring trends of stock markets.

1. I have assumed the natural logarithm of the stock price indices are generated by a mixture of state-dependent multivariate normal distributions. This dependent mixture is described by relations (7.1) and (7.2).

2. I have estimated the component means and variance-covariance matrices of the state-dependent mixture by maximum likelihood.

3. With each component mean and variance-covariance matrix, I have used (7.3) to estimate simultaneously the intercepts and the slopes of the linear model relating the natural logarithm of the stock price indices of the 15 other major exchanges (the dependent variables) to the natural logarithm of the S&P/TSX composite (the independent variable).

4. For each of the 3 973 trading days common to the 16 exchanges over the period 1998-2017, I have decoded the most likely state of the global financial market.

5. Using the intercept and the slope parameters corresponding to the most likely state of each time period, I have computed the fitted values and the residuals of the 15 bivariate relations. Then, I have performed the ADF test for unit root on these residuals.

$$\begin{aligned} p\left(\mathbf{y}_t;\boldsymbol{\mu}_{s_t},\boldsymbol{\Sigma}_{s_t}\right) &= (2\pi)^{-\frac{16}{2}}\left(\det\boldsymbol{\Sigma}_{s_t}\right)^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\mathbf{y}_t-\boldsymbol{\mu}_{s_t}\right)'\boldsymbol{\Sigma}_{s_t}^{-1}\left(\mathbf{y}_t-\boldsymbol{\mu}_{s_t}\right)\right] \\ \text{with } \boldsymbol{\mu}_{s_t} &= \begin{bmatrix}\boldsymbol{\mu}_{h,s_t}\\ \boldsymbol{\mu}_{f,s_t}\end{bmatrix} \text{ and } \boldsymbol{\Sigma}_{s_t} = \begin{bmatrix}\boldsymbol{\Sigma}_{h,h,s_t} & \boldsymbol{\Sigma}_{h,f,s_t}\\ \boldsymbol{\Sigma}_{h,f,s_t} & \boldsymbol{\Sigma}_{f,f,s_t}\end{bmatrix}, \end{aligned} \tag{7.1}$$

where the column vector $\mathbf{y}_t$ lists the natural logarithm of the stock price indices at time t, the vector $\boldsymbol{\mu}$ lists their expected values, the matrix $\boldsymbol{\Sigma}$ denotes their variance-covariance matrix, the subscipts h and f refer respectively to the home stock market (the TSX) and the foreign markets, and s_t is an unobserved state variable.

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{t=1}^{T}\sum_{s_t=1}^{2}\gamma_{s_{t-1},s_t}p\left(\mathbf{y}_t|s_t;\boldsymbol{\theta}_{s_t}\right) \\ &= \mathbf{u}\prod_{t=1}^{T}\boldsymbol{\Gamma}\mathbf{P}\left(\mathbf{y}_t\right)\mathbf{1}' \\ \text{with } \boldsymbol{\theta} &= \left[\gamma_{11},\gamma_{12},\gamma_{21},\gamma_{22},\boldsymbol{\mu}_1',\boldsymbol{\mu}_2',\text{vech}(\boldsymbol{\Sigma}_1)',\text{vech}(\boldsymbol{\Sigma}_2)'\right] \end{aligned} \tag{7.2a}$$

$$\begin{aligned} L(\boldsymbol{\theta}^*) &= \prod_{t=1}^{T}\sum_{s_t=1}^{2}\gamma^*_{s^*_{t-1},s^*_t}p\left(\mathbf{y}_t|s^*_t;\boldsymbol{\theta}^*_{s_t}\right) \\ &= \mathbf{u}^*\prod_{t=1}^{T}\boldsymbol{\Gamma}^*\mathbf{P}\left(\mathbf{y}_t\right)\mathbf{1}' \\ \text{with } \boldsymbol{\theta}^* &= \left[\gamma^*_{11},\gamma^*_{12},\ldots,\gamma^*_{44},\boldsymbol{\mu}_1^{*\prime},\ldots,\boldsymbol{\mu}_4^{*\prime},\text{vech}(\boldsymbol{\Sigma}^*_1)',\ldots,\text{vech}(\boldsymbol{\Sigma}^*_4)\right], \end{aligned} \tag{7.2b}$$

where the vector $\mathbf{u}$ is the stationary distribution of s_t, $\boldsymbol{\Gamma}$ is the matrix of transition probabilities, and $\mathbf{P}\left(r_{mt}|\mathbf{r}_{st}\right)$ is a diagonal matrix whose j-th diagoal element corresponds to $p\left(r_{mt}|\mathbf{r}_{st}, s_t = j;\boldsymbol{\theta}\right)$.

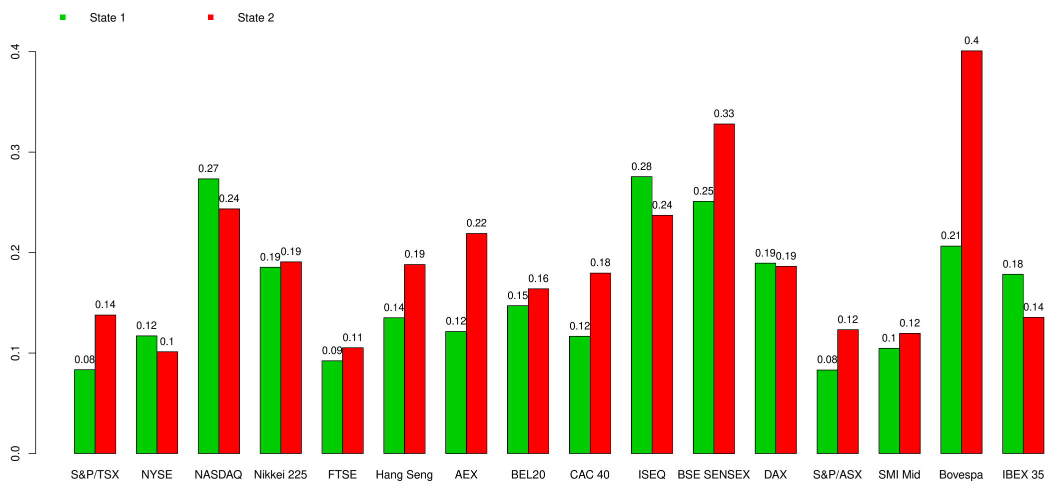

Figure 7.2: Standard Deviations from a First-Order Two-State Markov-Dependent Mixture of Multivariate Normal Distributions Fitted to some Major Stock Price Indices, 1998-2017 (daily)

$$\begin{aligned}\boldsymbol{\beta}_{s_t} &= \boldsymbol{\Sigma}_{h,f,s_t}\boldsymbol{\Sigma}^{-1}_{h,h,s_t}\\ \boldsymbol{\alpha}_{s_t} &= \boldsymbol{\mu}_{f,s_t} - \left(\boldsymbol{\Sigma}_{h,f,s_t}\boldsymbol{\Sigma}^{-1}_{h,h,s_t}\right)\boldsymbol{\mu}_{h,s_t}\end{aligned} \tag{7.3}$$

7.2 The Findings

This subsection presents the results of the AEG tests performed after estimating (7.3). In turn, I have assumed that the unobserved state variable is described by a two-state first-order and a two-state second-order Markov processes.

7.2.1 The First-Order Markov-Switching Cointegrating Relations

In state 1, the volatility is lower across the exchanges, with the exception of those in the US, Ireland, Germany, and Spain (see Figure 7.2). At the same time, the expected values of the stock price indices are higher across the exchanges, with the exception of the Euronext Amsterdam. Thus, state 1 can be labelled as a bull market. The estimates in (7.4) indicates that this state is the most recurrent and the most persistent.

$$\begin{aligned}\tilde{\boldsymbol{\Gamma}} &= \begin{bmatrix} .963 & .037 \\ .052 & .948 \end{bmatrix}\\ \tilde{\mathbf{u}} &= \begin{bmatrix} .586 & .414 \end{bmatrix}\end{aligned} \tag{7.4}$$

Table 7.1: First-Order Markov-Dependent Multivariate Normal Distributions: Bivariate Relationships between the Stock Price Indices of some Major Exchanges (Dependent Variables) and the S&P/TSX Composite (Explanatory Variable), AEG Test with Intercept but no Trend, 1998-2017 (daily)

	Intercept		Slope		AEG Test		R^2
Dependent Variable	1	2	1	2	Lags	τ-statistic	
NYSE Composite	-1.598	4.746	1.129	.443	3	-2.727	.879
	(-4.09)	(27.68)	(27.54)	(23.24)			
NASDAQ Composite	-9.265	1.339	1.831	.698	1	-1.587	.624
	(-11.41)	(4.25)	(21.47)	(20.06)			
N225	1.422	10.174	.856	-.085	2	-1.931	.236
	(3.71)	(48.95)	(21.24)	(-3.71)			
FTSE All-Share	.722	6.134	.775	.186	6	-2.207	.67 0
	(2.35)	(59.44)	(23.99)	(16.36)			
Hang Seng	-1.201	.493	.925	.992	2	-4.208	.906
	(3.95)	(.92)	(29.04)	(16.72)			
AEX	-.682	6.772	.705	-.081	1	-2.188	.117
	(-2.40)	(48.83)	(23.69)	(-5.29)			
BFL20	4.198	7.421	.406	.051	1	-1.970	.240
	(12.46)	(47.26)	(11.48)	(2.96)			
CAC 40	4.113	3.661	.449	.517	3	-2.274	.223
	(16.99)	(10.79)	(17.67)	(13.81)			
ISEQ All-Share	5.082	6.141	.360	.261	2	-1.237	063
	(12.97)	(25.16)	(8.73)	(9.66)			
BSESN	-4.397	-5.132	1.503	1.509	1	-3.533	.910
	(-6.07)	(-8.28)	(19.75)	(21.96)			
DAX Performance	-5.551	4.250	1.530	.468	2	-2.166	.733
	(-8.36)	(19.40)	(21.97)	(19.40)			
S&P/ASX 200	3.251	2.227	.557	.653	2	-3.204	.890
	(16.81)	(5.84)	(27.44)	(15.49)			
SMI MID	5.530	7.221	.359	.168	5	-2.198	.405
	(28.74)	(100.66)	(17.70)	(21.42)			
Bovespa Index	3.085	-10.623	.825	2.258	1	-3.359	.892
	(8.55)	(-11.93)	(21.71)	(22.82)			
IBEX 35	7.158	5.595	.219	.386	2	-2.524	.288
	(23.00)	(46.42)	(6.69)	(29.00)			
Selection Criteria							
Log-Likelihood							84 011.01
AIC							-167 410.00
BIC							-165 486.10

The numbers in parentheses are the z-statistics, which are computed dividing the maximum likelihood estimates by their respective bootstrap standard errors.

Table 7.1 displays the estimates of the intercept terms and the slope parameters. All the estimates of the slope parameters are significantly different from zero (*i.e.* the absolute value of their z-statistics are greater than 1.97, the critical value at the level of significance of 5% of a two-sided test). Comparing the coefficients of determination (R^2) computed after decoding the states to those from the OLS regression (in Table B.3) shows that, in most cases, the first-order Markov process has contributed to the improvement of the explanatory power of the bivariate models, especially in the case of Japan, the Netherlands, and France.

The asymptotic critical values for the ADF test performed on the residuals of a bivariate model are -3.34 and -3.04, respectively, at the levels of significance of 5% and 10% (Davidson and MacKinnon, 1993). In Table 7.1, the absolute value of the τ-statistics are greater than the critical value of 3.34, in three cases: the bivariate equations relating the stock prices on the Hong Kong Stockk Exchange, the Bombay Stock Exchange, and the Bovespa to the stock prices on the TSX. This means that, at the level of significance of 5%, the null hypothesis of unit root in the residuals can be rejected in these three cases and one can conclude that the stock prices in Canada and the stock prices in Hong Kong, India, and Brazil are cointegrated. One can also reject the null hypothesis of unit root in the residuals in the case of the Australian Security Exchange, but only at the level of significance of 10%. Narayan and Smyth (2004) also find a cointegration between the Canadian and the Australian stockk markets.

In conclusion, after distinguishing between two states of stock markets (the bull and the bear markets), two more exchanges have turned out to be cointegrated with the TSX.

7.2.2 The Second-Order Markov-Switching Cointegrating Relations

It appears in Figure 7.3 that the volatility of the benchmark indices is higher in state 4 across all the exchanges, with the exception of the NASDAQ. Thus, one can label state 4 as the bear market and, consequently, state 1 as the low volatility state (or the bull market). States 2 and 3 are the turning points respectively from the low to the high volatility states and from the high to the low volatility states.

On both Figures 7.2 and 7.3, the volatility of the stock prices is higher on the Bovespa, the Bombay Stock Exchange, and the Euronext Dublin. During bull markets, the volatility is lower on the Swiss Stock Exchange, the Australian Securities Exchange, and the London Stock Exchange.

The estimates in (7.5) indicate that the the high volatility state is persistent 30.6% of the time and transient 21.8% of the time. The sum of these two probabilities is higher than the estimate in (7.4).

$$
\begin{aligned}
\tilde{\mathbf{\Gamma}} &= \begin{bmatrix} .906 & .094 & .000 & .000 \\ .000 & .000 & .924 & .076 \\ .111 & .889 & .000 & .000 \\ .000 & .000 & .054 & .946 \end{bmatrix} \\
\tilde{\mathbf{u}} &= \begin{bmatrix} .258 & .218 & .218 & .306 \end{bmatrix}
\end{aligned}
\tag{7.5}
$$

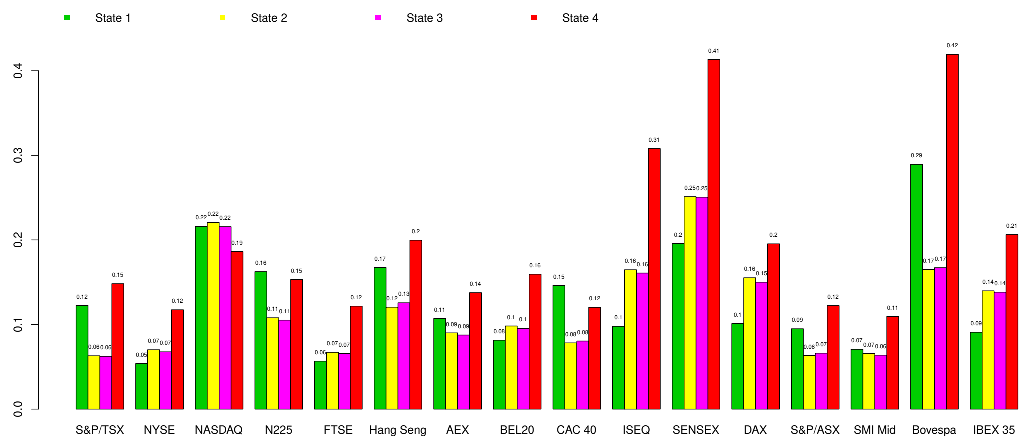

Figure 7.3: Standard Deviations from a Second-Order Two-State Markov-Dependent Mixture of Multivariate Normal Distributions Fitted to some Major Stock Price Indices, 1998-2017 (daily)

Comparing the selection criteria in Table 7.2 to those in Table 7.1 reveals that the two-state second-order Markov process helps provide a better fit to the data: its log-likelihood is higher and its information criteria (AIC and BIC) are lower. The determination coefficients computed after decoding the recurring states also indicate substantial improvements in the explanatory power of the bivariate models. The R^2 of the bivariate switching regressions that relate the stock prices in Canada to the stock prices in seven countries (Japan, the Netherlands, Belgium, France, Ireland, Switzerland, and Spain) have more than doubled. For example, in the case of Ireland, the R^2 rose from 6.3% in the model with the first-order Markov process to 50.3% in the one with the second-order Markov process.

In Table 7.2 as in Table 7.1, the slope parameter is negative and statistically significant over some states in the equation relating the stock prices on the Tokyo Stock Exchange to those on the TSX. Table 7.2 also indicates that the stock price indices on the Euronext Brussels and the Euronext Dublin are negatively related to those on the TSX, over some states, which might create some arbitrage opportunities between these exchanges.

In Table 7.2, the absolute value of the τ-statistics are greater than 3.34, the absolute value of their 5% critical value, in 11 cases. This time, the null hypothesis of unit root in the residuals of the 15 bivariate switching regressions cannot be rejected only in four cases: the equations relating the stock prices on the NASDAQ, the Euronext Dublin, the Bombay Stock Exchange, and the Bolsa de Madrid to the stock prices on the TSX.

While the TSX turns out to be cointegrated with the NYSE, I have not found any evidence of cointegration between the TSX and the NASDAQ. This nuances the result

Table 7.2: Second-Order Markov-Dependent Multivariate Normal Distributions: Bivariate Relationships between the Stock Price Indices of some Major Exchanges (Dependent Variables) and the S&P/TSX Composite (Explanatory Variable), AEG Test with Intercept but no Trend, 1998-2017 (daily))

	Intercept				Slope				AEG Test		R^2
Dependent Variable	1	2	3	4	1	2	3	4	Lags	τ-statistic	
NYSE Composite	7.096	2.147	2.119	3.251	.185	.741	.744	.604	5	-4.643	.927
	(51.12)	(6.57)	(6.90)	(11.99)	(11.89)	(21.77)	(23.24)	(20.66)			
NASDAQ Composite	2.451	-4.665	-4.637	1.304	.593	1.351	1.347	.689	12	-2.096	.691
	(6.25)	(-7.73)	(-8.50)	(4.62)	(13.53)	(21.51)	(23.73)	(22.66)			
N225	10.990	6.814	7.692	9.218	-.156	.303	.212	-.001	12	-4.301	.711
	(40.33)	32.55	(35.53)	(44.01)	(-5.17)	(13.92)	(9.43)	(-.06)			
FTSE All-Share	7.001	3.860	4.093	2.901	.103	.449	.425	.533	3	-4.543	.875
	(59.74)	(19.68)	(22.44)	(16.75)	(7.88)	(22.03)	(22.43)	(28.78)			
Hang Seng	2.116	1.281	1.386	.057	.815	.917	.905	1.045	12	-3.864	.894
	(3.77)	(3.15)	(3.68)	(.15)	(13.00)	(21.51)	(23.00)	(25.81)			
AEX	5.816	3.171	2.817	5.429	.056	.309	.346	.037	3	-4.851	.756
	(37.80)	(19.88)	(18.63)	(40.07)	(3.27)	(18.59)	(21.94)	(2.53)			
BFL20	9.001	6.590	6.315	4.466	-.113	.167	.196	.359	1	-4.049	.652
	(52.55)	(50.19)	(48.16)	(16.18)	(-5.92)	(12.20)	(14.31)	(12.02)			
CAC 40	5.872	7.233	7.554	6.324	.289	.130	.097	.200	3	-4.581	.622
	(18.95)	(84.30)	(95.36)	(49.32)	(8.35)	(14.54)	(11.67)	(14.56)			
ISEQ All-Share	7.932	11.343	11.371	13.365	.068	-.273	-.276	-.558	5	-3.209	.503
	(46.25)	(46.68)	(45.20)	(38.00)	(3.59)	(-10.69)	(-10.45)	(-14.65)			
BSESN	2.846	-7.406	-7.927	-10.333	.603	1.820	1.873	2.120	5	-2.561	.922
	(5.05)	(-8.78)	(-10.47)	(-13.87)	(9.56)	(20.65)	(23.70)	(26.58)			
DAX Performance	4.610	-2.561	-1.564	.816	.450	1.221	1.115	.835	12	-3.436	.855
	(16.86)	(-5.92)	(-3.74)	(2.65)	(14.65)	(27.03)	(25.62)	(25.31)			
S&P/ASX 200	4.235	3.922	3.820	2.086	.422	.492	.503	.673	3	-4.216	.941
	(18.05)	(19.00)	(18.45)	(7.48)	(16.03)	(22.84)	(23.30)	(22.45)			
SMI MID	7.547	7.779	8.377	6.550	.147	.132	.070	.230	3	-5.879	.831
	(69.00)	(62.63)	(66.05)	(37.71)	(12.14)	(10.15)	(5.23)	(12.31)			
Bovespa Index	-4.464	-2.477	-2.944	-9.368	1.551	1.401	1.448	2.156	12	-3.580	.934
	(-5.11)	(-3.94)	(-4.47)	(-11.38)	(15.78)	(21.34)	(21.02)	(24.40)			
IBEX 35	7.739	6.107	5.623	2.829	.158	.334	.387	.676	12	-2.908	.436
	(61.71)	(27.10)	(25.54)	(8.79)	(11.31)	(14.15)	(16.79)	(19.59)			
Selection Criteria											
Log-Likelihood											88 114.61
AIC											-175 005.2
BIC											-171 157.4

The numbers in parentheses are the z-statistics, which are computed dividing the maximum likelihood estimates by their respective bootstrap standard errors.

of Khan (2011), who finds evidence of cointegration between stock prices in Canada and the US. Khan (2011) use a single index, the S&P 500, to describe the US stock market.

Consistently with the result in Table 7.1 (from the first-order state-dependent model) or in Table B.3 (from the OLS regression), the stock prices in Canada are not cointegrated with the stock prices in Spain. On the other hand, the absence of cointegration between the TSX and the Bombay Stock Exchange comes as a surprise, since these two stock markets turned out to be cointegrated in the two previous investigations (see Table 7.1 and Table B.3). The stock prices in Canada are cointegrated with the stock prices in Ireland only at a level of significance of 10%.

8 Conclusion

I have introduced the Markov process in some models to explain stock prices, returns, and volatility on the TSX. This process has helped distinguished between the recurring states of stock markets: mainly, the bull and the bear markets. Doing so has helped improve the explanatory power of the model fitted to the financial data. These models are the conditional CAPM, the conditional Sharpe model, the exponential autoregressive model with state-dependent heteroskedasticity, the state-dependent cointegrating relations among international stock markets.

The state-dependent multivariate normal distributions from which I derived the conditional CAPM have enabled to characterized the sectors on the TSX. Consumer staples and gold have turned out to be the defensive stocks on the TSX, as their expected returns are positive in bear markets. Besides, the use of the second-order Markov chain in estimating the conditional CAPM has improved considerably the proportion of the variability in the excess returns explained by the market (*viz* the determination coefficient) in the sectors of consumer staples and utilities, and in the sub-industry of gold.

The conditional Sharpe models reveal that the contribution of each sector to the market return is variable, since some sectors of the TSX are cyclical and the others are defensive. It turns out from the first-order Markov-switching model that the financial sector accounts for the highest share of the market return in both the bull and the bear markets. The contribution of the sector of consumer staples is low in both states but higher in bear markets.

The exponential autoregressive model reveals that the returns on the TSX are nonlinear time series. This explains the non-linearity of their autocorrelation, which could be high or low in bull markets depending on whether the returns are low or extremely high. In bear markets (periods of high volatility), the autocorrelation is low because the returns are often negative and extremely low. Both the conditional Sharpe model and the exponential autoregressive model with state-dependent heteroskedasticity indicate asymmetry in the volatility of the market return. The response of the variance to the lagged squared residuals is higher in bear markets, which confirms the existence of a leverage effect.

The second-order Markov-switching regressions raised considerably the number of stock markets that are cointegrated with the TSX.

Appendices

A The Data

The investigations cover the period 1998-2017. The monthly S&P/TSX composite indices and the sub-indices relating to to the sectors of the TSX are from Statistics Canada (www.statcan.gc.ca). The health care sector is missing from this data set. Real estate is excluded form the data set, because it was an industry group of the financial sector, up to September 16, 2016.

The one-month treasury bill yields used as risk-free rates are also monthly data from Statistics Canada. Since the treasury bill yields are annual percentage rates, they have been converted into monthly rates as follows $R_{ft} = (1 + \tilde{R}_{ft})^{1/12} - 1$, where $\tilde{R}_{ft}$ designates the annual rate.

The daily benchmark indices of the 16 major stock exchanges are from Yahoo Finance (www.finance.yahoo.com). Public holidays might differ from one country to the other. For this reason, only the trading days that are common to the 15 countries where these exchanges are based are included in the sample. These benchmark indices are described below.

Ticker	**Description**	**Country**
NYA	New York Stock Exchange (NYSE) composite	US
IXIC	National Association of Securities Dealers Automated Quotations (NASDAQ) composite	US
N225	Nikkei 225 (or Nikkei stock average)	Japan
FTAS	Financial Times Stock Exchange (FTSE) all-share	UK
HSI	Hang Seng index	Hong Kong
AEX	Euronext Amsterdam index	Netherlands
BFX	Euronext Brussels BEL20	Belgium
FCHI	Euronext Paris CAC 40	France
ISEQ	Euronext Dublin ISEQ all-share	Ireland
GSPTSE	S&P/TSX composite index	Canada
BSESN	Bombay Stock Exchange sensitive index	India
GDAXI	Deutscher Aktienindex	Germany
AXJO	S&P/Australian Securities Exchange (S&P/ASX) 200	Australia
SSMI	Swiss market index midcap (SMI MID)	Switzerland
BVSP	Bolsa de Valores do Estado de São Paulo (Bovespa) index	Brazil
IBEX	Bolsa de Madrid IBEX 35	Spain

The returns on the benchmark indices have been computed taking the first difference of the natural logarithm of their closing values. Figure A.1 compares the distribution of returns on the TSX to those of its sectors while Figure A.2 compares the distribution of returns on the TSX to those of other major exchanges. In Appendix B.1, I have explained how one computes these distributions.

Table A.1 shows the correlation between returns across sectors on the TSX. The

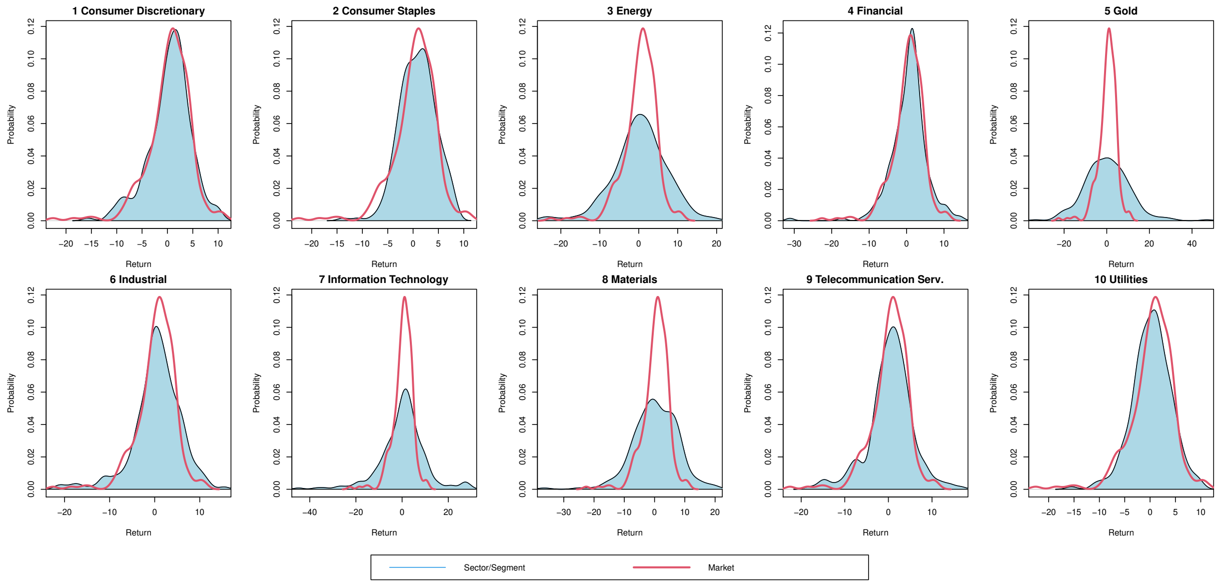

Figure A.1: Comparison of the Kernel Density Estimate of Returns on the TSX to those of Ten of its Sectors.

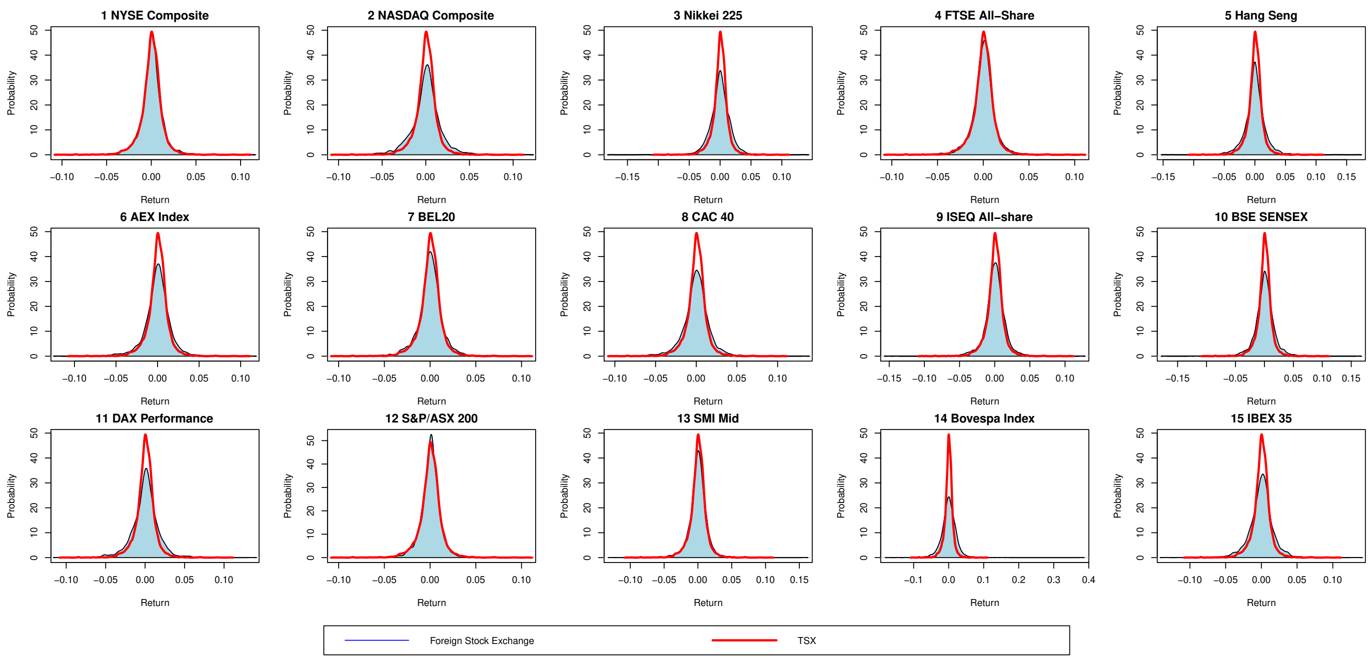

Figure A.2: Comparison of the Kernel Density Estimate of Returns on the TSX to those on some Foreign Stock Exchanges.

Table A.1: Correlation Coefficient between the Returns, TSX, Canada, 1998:M1-2016:M12

	CD	CS	E	F	G	I	IT	M	TS	U
CD	1.00									
CS	.43	1.00								
E	.26	.15	1.00							
F	.56	.33	.43	1.00						
G	-.06	.10	.36	.05	1.00					
I	.71	.29	.38	.58	- .01	1.00				
IT	.53	.10	.10	.33	- .07	.54	1.00			
M	.19	.15	.56	.26	.84	.32	.15	1.00		
TS	.49	.18	.11	.32	- .08	.37	.49	.10	1.00	
U	.18	.42	.34	.33	.18	.22	- .08	.26	.12	1.00

CD: Consumer Discretionary, CS: Consumer Staples, E: Energy, F: Financial, G: Gold, I: Industrial, IT: Information Technology, M: Materials, TS: Telecommunication Service, U: Utilities.

correlation between return in the gold sub-industry with returns in the consumer discretionary, the industrial, the information technology and the telecommunication service sector is weak. However, there is a high correlation between returns in the gold sub-industry and the materials sector (because gold is part of materials) and between returns in the sectors of consumer discretionary and industrial.

Table A.2 shows the first-order autocorrelation coefficient of both the returns and the squared returns. These coefficients measure the persistence of returns and that of their volatility.

B Some Basic Concepts

In this section, I introduce some concepts in econometric and finance used throughout this research.

B.1 Kernel Density Estimation

The kernel density estimation is the computation without any prior assumption of the probability density function (PDF) of a random variable. Let F denote the unknown cumulative distribution function (CDF) of a real-valued random variable R. The value of the function F at a point r in the support of R is

$$F(r) = \Pr(R \leq r) .$$

Given a time series R_t $(t = 1, \ldots, T)$, the empirical estimator of the value of F at the point r is the count of the observations that are less than or equal to r, which is divided

Table A.2: The First-Order Autocorrelation of the Returns and their Squares, TSX, Canada, 1998:M1-2017:M12

Sector/Segment	Returns	Squared Returns
Consumer Discretionary	.197	.118
Consumer Staples	.003	.022
Energy	.098	.391
Financial	.235	.061
Industrial	.191	.229
Information Technology	.092	.296
Materials	-.037	.225
Gold	-.131	.165
Telecommunication Service	.247	.250
Utilities	..017	.155
60 Largest Companies	.211	.155
Market	.223	.175

by T, the size of the sample.

$$\hat{F}(r) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{1}(R_t \leq r), \tag{B.1}$$

where $\mathbf{1}$ is an indicator function that takes the value 1 if $R_t \leq r$ or 0, otherwise.

The PDF, f, of R is defined as the first derivative of F with respect to r.

$$f(r) == \lim_{h\to 0}\frac{F(r+h)-F(r-h)}{2h}$$

Given the latter definition of a derivative and relation (B.1) defining the empirical CDF, the estimator of f is

$$\begin{aligned}\hat{f}(r) &= \frac{\hat{F}(r+h)-\hat{F}(r-h)}{2h}\\ &= \frac{1}{2hT}\sum_{t=1}^{T}\mathbf{1}(r-h \leq R_t \leq r+h),\end{aligned} \tag{B.2}$$

where h is a small positive increment referred to as smoothing parameter or bandwidth. The bandwidth is determined so as to minimize the mean squared error of $\hat{f}(r)$, $\mathrm{E}\left[\hat{f}(r)-f(r)\right]^2$.

In relation (B.2), a constant weight, which is $1/2$, is put on $\hat{f}$ whenever $r-h \leq R_t \leq r+h$ or $-1 \leq (R_t - r)/h \leq 1$. This kernel corresponds to a uniform PDF defined over the interval $[-1,\ 1]$. For this reason, relation (B.2) is called a uniform kernel density estimator.

Other non-negative and symmetric weighting functions can be used in (B.2) as long as they integrate to one, hence the following general expression.

$$\hat{f}(r) = \frac{1}{hT}\sum_{t=1}^{T} k\left(\frac{R_t - r}{h}\right)$$

In this paper, I have used the standard normal distribution as kernel to estimate the nonparametric densities, *i.e.*

$$k\left(\frac{R_t - r}{h}\right) = \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{R_t - r}{h}\right)\right]^2 \quad -\infty < \frac{R_t - r}{h} < \infty.$$

For more details on kernel density estimation, see Li and Racine, 2007; Fan and Yao, 2008, among others.

B.2 The Linear Regression Model

Let's consider the multiple linear regression model

$$Y_t = \mathbf{X}_t'\boldsymbol{\beta} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right) \tag{B.3}$$

where Y_t is the explained variable, $\mathbf{X}_t$ is a $K \times 1$ vector of explanatory variables, and ε_t is an error term. [4] The error term, also known as innovation, is normally distributed with a zero mean and constant variance, σ_ε^2.

Given y_t and $\mathbf{x}_t$, a sample of T observations of Y_t and $\mathbf{X}_t$, the unknown parameter vector $\boldsymbol{\beta}$ can be estimated either by minimizing the sum of the squares of the errors, which is known as the ordinary least squares (OLS) estimation, or by maximizing the likelihood of observing y_t given $\mathbf{x}_t$, which is known as maximum likelihood estimation. The estimator of the $K \times 1$ vector $\boldsymbol{\beta}$ produced by either of these methods is

$$\hat{\boldsymbol{\beta}} = \left(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\sum_{t=1}^{T}\mathbf{x}_t y_t. \tag{B.4}$$

Note that, for the $K \times K$ matrix $\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'$ in (B.4) to be invertible, no explanatory variable should be an exact linear combination of the others.

Given (B.4), the regression residual is $e_t = y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}$. The maximum likelihood estimator of the variance of ε_t is

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{T}\sum_{t=1}^{T} e_t^2.$$

This latter estimator is *biased*, *i.e.* its expectation differs from the true parameter, σ_ε^2. The OLS estimator of this latter parameter is unbiased and is generally preferred

$$s^2 = \frac{1}{T-K}\sum_{t=1}^{T} e_t^2, \tag{B.5}$$

[4] To allow for an intercept term, one of elements of $\mathbf{X}_t$, say its first element, is set to one all the time.

where K is the number of explanatory variables including the intercept term. The square rood of (B.5) is referred to as the standard error of the regression.

The covariance of the estimator $\hat{\boldsymbol{\beta}}$ is

$$\begin{aligned}\operatorname{cov}\left(\hat{\boldsymbol{\beta}}\right) &= s^2 \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'\right)^{-1} \\ &= \frac{1}{T-K} \sum_{t=1}^{T} e_t^2 \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'\right)^{-1} . \end{aligned} \tag{B.6}$$

The square root of the diagonal elements of (B.6) gives the standard errors of the estimator $\hat{\boldsymbol{\beta}}$. For further details on linear regression, see Greene (2000); Verbeek (2008); Ruppert (2011), among others.

B.2.1 The Significance Tests

The error term, ε_t, is assumed to be normally distributed. Thus, the estimator $\hat{\boldsymbol{\beta}}$ is also a random variable that is normally distributed with mean $\boldsymbol{\beta}$ and covariance matrix $\sigma_\varepsilon^2 \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'\right)^{-1}$. Both $\boldsymbol{\beta}$ and σ_ε^2 are unknown. One can use their estimates to make inference about them. For instance, one can be interested in knowing whether the true value of either a single parameter β_k or a subset of $\boldsymbol{\beta}$ is zero, given their estimates. This is called a statistical significance test.

The significance test consists of two hypotheses: H_0, the null hypothesis, versus H_A, the alternative hypothesis. For a single parameter β_k, one might perform a one-sided or a two-sided test.

One-sided test $\mathrm{H}_0: \ \beta_k = \beta_k^0$ versus either $\mathrm{H}_A\colon \ \beta_k < \beta_k^0$ or $\mathrm{H}_A\colon \ \beta_k > \beta_k^0$,

Two-sided test $\mathrm{H}_0: \ \beta_k = \beta_k^0$ versus $\mathrm{H}_A\colon \ \beta_k \neq \beta_k^0$,

where β_k^0 is the researcher's a priori about β_k. Generally, β_k^0 is set to zero, for the purpose of knowing whether a variable should be kept in a model or removed.

The standard normal distribution would be used to test for these hypotheses if one knew σ_ε^2. To overcome this issue, one replaces σ_ε^2 by its empirical estimate given by (B.5) to compute the test statistic

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{\text{s.e.}\left(\hat{\beta}_k\right)} \tag{B.7}$$

where $\text{s.e.}\left(\hat{\beta}_k\right)$, the kth diagonal element of (B.6), is the standard error of $\hat{\beta}_k$. The test statistic t_k in (B.7) is called a t-ratio. It follows a Student's t-distribution with $T-K$ *degrees of freedom*, as it is the ratio of a standardized normally distributed variable, $(\hat{\beta}_k - \beta_k^0)/\sigma_{\beta_k}$, and a variable following a Chi-square distribution, $(T-K)\text{s.e.}\left(\hat{\beta}_k\right)/\sigma_{\beta_k}$.

In principle, K independent observations are enough to estimate the K unknown parameters in the regression model (B.3). Once the fitted values $\mathbf{x}_t'\hat{\boldsymbol{\beta}}$ are computed, the remaining free $T-K$ observations called degrees of freedom can be used to estimate the standard error.

While performing a test, there is a risk of making two wrong decisions: either rejecting the null hypothesis whereas it is true, which is called a type I error, or accepting the null hypothesis whereas it is untrue, which is called a type II error (Casella and Berger, 2002, chap 8). One controls the type I error probability by setting a significance level for the test. The significance level is generally set at 5% and corresponds to the probability of making a type I error.

The lower and upper tails of the t-CDF associated with a significance level depends on $T-K$, the degrees of freedom. These tails are the critical values of the test. For a one-sided test with a 5% significance level, the critical value associated with H_A: $\beta_k > \beta_k^0$ is $t_{5\%}(T-K)$ such that $\Pr\left(t_k > t_{5\%}(T-K)\right) = .05$. The null hypothesis is rejected when $t_k > t_{5\%}(T-K)$. The length of the monthly returns used in this research is 239. Consequently, the upper tail of the test, $t_{5\%}(T-K)$, is about 1.65. Moreover, since the t-distribution is symmetrical about zero like the standard normal distribution, its lower and upper tails are opposite numbers. Thus, the lower tail of the test is about -1.65. When the alternative hypothesis is rather H_A: $\beta_k < \beta_k^0$, the null hypothesis will be rejected when $t_k < -1.65$.

When performing a two-sided t-test with a 5% significance level, the null hypothesis is accepted when $-t_{2.5\%}(T-K) \leq t_k \leq t_{2.5\%}(T-K)$, with $|t_{2.5\%}(T-K)| \approx 1.97$ for $T = 239$.

B.2.2 The Model Selection Criteria

There are several statistics that can be used to compare alternative models in order to find out the one that provides a good fit to a data set. The most popular of these statistics is the *coefficient of determination*, R^2. The coefficient of determination indicates the share of the total variation in the dependent variable explained by a model.

$$\begin{aligned} R^2 &= \frac{\sum_{t=1}^{T}\left(\hat{y}_t - \bar{y}\right)^2}{\sum_{t=1}^{T}\left(y_t - \bar{y}\right)^2} \\ &= 1 - \frac{\sum_{t=1}^{T} e^2}{\sum_{t=1}^{T}\left(y_t - \bar{y}\right)^2}, \end{aligned}$$

where $\hat{y}_t$ equals $\mathbf{x}_t'\hat{\boldsymbol{\beta}}$, the fitted value of y_t, and $\bar{y}$ denotes the mean of y_t. In case a model has no intercept term, the mean of the dependent variable, $\bar{y}$, is set at zero in the coefficient of determination (see Barten, 1987; Eisenhauer, 2003, among others).

The coefficient of determination can be used to test for the joint significance of all the sensitivity parameters in the vector $\boldsymbol{\beta}$, which excludes the intercept. The test hypotheses are

$$\mathrm{H}_0: \; Y_t = \alpha + \varepsilon_t,$$

$\mathrm{H}_A: \ Y_t = \mathbf{X}_t'\boldsymbol{\beta} + \varepsilon_t$ (model (B.3)),

and its statistic is

$$F = \frac{T-K}{K-1}\frac{R^2}{1-R^2},$$

which follows an F-distribution with $K-1$ and $T-K$ degrees of freedom, $F(K-1, T-K)$ in short.

The coefficient of determination tends to increase along with K, the number of explanatory variables. But, when K increases, the degrees of freedom, $T-K$, decreases and less observations become available to estimate accurately the parameters.

To encourage parsimony in the specification of models, other selection criteria imposes penalties that increase along with the number of explanatory variables. Such selection criteria include the adjusted R^2 or $\bar{R}^2$, the Akaike information criterion (AIC), and the Schwartz Bayesian information criterion (BIC).

$$\bar{R}^2 = 1 - \frac{T-1}{T-K}\frac{\sum_{t=1}^{T} e^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2} \tag{B.8a}$$

$$AIC(K) = \ln\left(\frac{1}{T}\sum_{t=1}^{T} e_t^2\right) + 2\frac{K}{T} \tag{B.8b}$$

$$BIC(K) = \ln\left(\frac{1}{T}\sum_{t=1}^{T} e_t^2\right) + \frac{K}{T}\ln(T) \tag{B.8c}$$

Unlike R^2, it appears in (B.8a) that $\bar{R}^2$ may fall when K increases and the residual sum of squares does not fall in the same proportion as the loss of degrees of freedom. One would prefer either a model with a higher $\bar{R}^2$ or a model with a lower AIC and BIC. Comparing (B.8b) to (B.8c) indicates that even though both criteria impose a penalty that increases with K, the penalty imposed by the BIC is larger for $\ln(T) > 2$ (*i.e.*, for $T \geq 8$).

Alternative specifications of the AIC and BIC are:

$$AIC(K) = -2\ln\left[L\left(\boldsymbol{\beta}\right)\right] + 2K \tag{B.9a}$$

$$BIC(K) = -2\ln\left[L\left(\boldsymbol{\beta}\right)\right] + K\ln(T), \tag{B.9b}$$

where L denotes the likelihood of the observations. I have computed the AIC and BIC of the Markov-switching models using the latter specifications.

B.2.3 An Application

The following linear relation between asset and market expected returns, referred to as CAPM, is considered in Section 4 (see its derivation in Appendix B.4).

$$\mathrm{E}(R_{it} - R_{ft}) = \beta_{m,i}\mathrm{E}(R_{mt} - R_{ft}), \tag{B.10}$$

To estimate it by OLS, one can assume that investors make errors in forecasting and, as a consequence, the realized returns are the sum of the expected and unexpected returns.

$$r_{it} = \mathrm{E}(r_{it}) + u_{it}$$
$$r_{mt} = \mathrm{E}(r_{mt}) + u_{mt}$$

The forecast errors u_{it} and u_{mt} are neither contemporaneously nor serially correlated. Plugging them into (B.10) gives

$$r_{it} - r_f = \beta_i(r_{mt} - r_f) + u_t, \tag{B.11}$$

where $u_t = u_{it} - \beta_i u_{mt}$.

Table B.1 reports the OLS estimates of (B.11) for the sectors of the TSX. The one-month treasury bill is used as the risk-free asset and the growth rate of the S&P/TSX composite is used as the market returns.

One tests for the statistical significance of each estimate by comparing its t-ratio in parentheses to the critical value, $t_{2.5\%}(226) = 1.97$. Except for the consumer staples, these tests confirm that the intercept terms are not significantly different from zero, as the CAPM suggests. All the slope parameter (*i.e.* the beta coefficients) are statistically significant. In the information technology sector, the excess returns are highly sensitive to the fluctuations in the market excess returns, unlike in the utilities and the consumer staples sectors. While the market explains about half of the fluctuations in the excess returns in the industrial sector, movements in the excess returns in the consumer staples and in the utilities sectors or in the gold sub-industy are mostly idiosyncratic and thus unpredictable.

B.3 Stationarity and Cointegration

B.3.1 Stationarity

There are a weak and a strong form of stationary. A time series R_t $(t = 1, \dots, T)$ is said to be strongly stationary if its joint unconditional probability at time t is unaffected by any arbitrary change of time origin.

$$\Pr(R_t, R_{t-1}, \dots, R_1) = \Pr(R_{t+k}, R_{t-1+k}, \dots, R_{1+k})$$
$$\prod_{\tau=1}^{t} \Pr(R_\tau) = . \prod_{\tau=1}^{t} \Pr(R_{\tau+k})$$

A time series R_t $(t = 1, \dots, T)$ is said to be weakly stationary or covariance stationary if

1. $\mathrm{E}(R_t) = \mathrm{E}(R_{t-1}) = \mu < \infty$,
2. $\mathrm{var}(R_t) = \mathrm{E}(R_t - \mu)^2 = \gamma_0 < \infty$,
3. $\mathrm{cov}(R_t, R_{t-k}) = \mathrm{E}(R_t - \mu)(R_{t-k} - \mu) = \gamma_k < \infty$, $k = 1, 2, \dots$,

Table B.1: CAPM, OLS Regressions, TSX, Canada, 1998:M1-2017:M12

Sector/Segment	Intercept	Market β	R^2
Consumer Discretionary	.218	.644	.444
	(1.08)	(13.82)	
Consumer Staples	.737	.206	.060
	(3.32)	(4.01)	
Energy	-.047	.982	.376
	(-.13)	(12.02)	
Financial	.329	.751	.452
	(1.42)	(14.04)	
Industrial	.160	.901	.522
	(.66)	(16.15)	
Information Technology	-.244	1.507	.427
	(-.50)	(13.37)	
Materials	-.100	1.007	.342
	(-.26)	(11.18)	
Gold	-.164	.671	.073
	(-0.25)	(4.46)	
Telecommunication Service	.190	.637	.283
	(.67)	(9.73)	
Utilities	..156	.194	.045
	(.65)	(3.48)	
60 Largest Companies	.013	1.014	.977
	(.30)	(101.22)	
Selection Criteria			
AIC			14 970.93
BIC			15 092.29

$t_{2.5\%}(226) = 1.97$

where E, var, and cov denote respectively the mathematical expectation, the variance, and the covariance operators. The first and the second conditions for weak stationarity require the mean and the variance of the time series to be a finite number and constant. The third condition requires the autocovariances to depend only on the time interval between two observations.

Strong stationarity and weak stationarity are equivalent, when a time series is normally and identically distributed. The reason is that normally distributed variables are completely described by their means and variance-covariance matrix. Under other distributional assumptions, a weakly stationary time series might not be strongly stationary. Besides, as obtaining information about the statistical distribution of a time series might be difficult or joint probability distribution might not be easy to deal with, the weak form of stationarity is the most commonly referred to.

A time series that fluctuates around a constant mean is an example of a stationary variable. This process is referred to as mean reversion. A general process called autoregressive moving average (ARMA) models can also be used to illustrate stationarity.

$$R_t = \delta + \rho_1 R_{t-1} + \cdots + \rho_p R_{t-p} + \varepsilon_t + \eta_1 \varepsilon_{t-1} + \cdots + \eta_q \varepsilon_{t-q},$$

where the innovation ε_t is normally distributed with a mean of zero and a constant variance σ_ε^2, $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$. The above $ARMA(p,q)$ model decomposes R_t into a linear combination of its p most recent values and a linear combination of the $q+1$ latest innovations. To simplify the illustration, I have set p at one and q at zero. This is called an autoregressive model of order 1.

$$R_t = \delta + \rho R_{t-1} + \varepsilon_t \tag{B.12}$$

Given (B.12), it follows from the three conditions required for weak stationarity that the mean, the variance, and the covariances of R_t are

$$\begin{aligned} \mathrm{E}\left(R_t\right) &= \frac{\delta}{1-\rho} \\ \operatorname{var}\left(R_t\right) &= \frac{\sigma_\varepsilon^2}{1-\rho^2} \\ \operatorname{cov}\left(R_t, R_{t-k}\right) &= \rho^k \frac{\sigma_\varepsilon^2}{1-\rho^2}. \end{aligned} \tag{B.13}$$

The results in (B.13) indicate that, for $\rho = 1$, the mean and the variance of R_t are not finite numbers. Note that if ρ equals 1, R_t is said to have a unit root or to follow a random walk. Besides, for $\rho < -1$ or $\rho > 1$, the variance of R_t is negative, which is counterfactual, and its autocovariances tend to infinity, *i.e.* they become extremely high, as k increases. Therefore, R_t is weakly stationary only if $-1 < \rho < 1$. It follows that one can test for stationarity on a time series using the OLS estimate of ρ and its *standard error*. This is called Dickey-Fuller unit root test (Dickey and Fuller, 1979).

An alternative and more convenient specification of (B.12) used to test for unit root is obtained by subtracting R_{t-1} from both sides of the equation.

$$\Delta R_t = \delta + \pi R_{t-1} + \varepsilon_t, \tag{B.14}$$

where Δ denotes the first-difference operator, *i.e.* $\Delta R_t = R_t - R_{t-1}$, and $\pi = \rho - 1$. The null hypothesis H_0 and the alternative hypothesis H_A of the Dickey-Fuller unit root test are

$$\mathrm{H}_0: \ \pi = 0 \text{ (a unit root)},$$

$$\mathrm{H}_A: \ \pi < 0 \text{ (stationarity)}.$$

For a 5% significance level, *i.e.* for a 5% probability of rejecting by mistake the null hypothesis of unit root, the critical value of this one-sided test is

$$\tau_{5\%}(T-3) \text{ such that } \Pr\left(\frac{\hat{\pi}}{\text{s.e.}(\hat{\pi})} > \tau_{5\%}(T-3)\right) = 5\%$$
$$\text{s.e.}(\hat{\pi}) = \sqrt{\left(\frac{1}{T-3}\sum_{t=2}^{T} e_t^2\right)\left(\sum_{t=2}^{T} R_{t-1}^2\right)^{-1}},$$

where $\hat{\pi}$, s.e.$(\hat{\pi})$, and e_t denote respectively the OLS estimate of π, the estimated standard error of $\hat{\pi}$, and the residuals from estimating (B.14) by OLS. [5]

Even though, the Dickey-Fuller test and the t-test are analogous, in general, the critical values of the former are smaller than those of the latter. As a consequence, one would too often reject the null hypothesis of unit root using the t distribution.

In a higher order autoregressive model, unit root occurs when $\sum_i \rho_i = 1$, *i.e.* when the autoregressive parameters sum to unity. To estimate $\sum_i \rho_i - 1$ in one go, the autoregressive model can be specified as follows through linear transformations

$$\Delta R_t = \delta + \pi R_{t-1} + \cdots + \varrho_{p-2}\Delta R_{t-p+2} + \varrho_{p-1}\Delta R_{t-p+1} + \varepsilon_t. \tag{B.15}$$

where $\pi = \rho_1 + \cdots + \rho_p - 1$, $\varrho_{p-1} = -\left(\rho_{p-1} + \rho_p\right)$, and $\varrho_{p-1} = -\rho_p$.

Unit root tests performed on (B.15) are called augmented Dickey-Fuller (ADF) tests. A time trend can be introduced into (B.14) and (B.15), to test for unit root. When the unit root hypothesis is not rejected, the test is repeated on the first difference of the time series.

Table B.2 reports the τ statistics, *i.e.* the ratios $\hat{\hat{\pi}}/\text{s.e.}(\hat{\pi})$, from ADF tests performed on the natural logarithm of the stock prices and the treasury bill rate described in Appendix A. The statistics are computed with the software R using the package 'urca' (Pfaff, Zivot, Stigler, and Pfaff, 2016). The package 'urca', which stands for unit root and cointegration analysis, is able to select the number of lags to include in the ADF tests using either the lowest AIC or the lowest BIC. I have selected the order of the autoregressive processes to use for the tests using the BIC. The ADF tests indicate the stock prices and the treasury bill rate are not stationary, but their first differences are all stationary. Such time series are said to be integrated of order one, $I(1)$.

---

[5] The degree of freedom of the critical value, which equals $T-3$, corresponds to the number of observations used in the OLS regression, $T-1$, minus the number of estimated parameters, 2.

Table B.2: Statistics from ADF Unit Root Tests with Intercept but no Trend

	Level		First-Difference	
Variable	Lags	τ-statistics	Lags	τ-statistics
Monthly Data				
S&P/TSX Composite	1	-1.570	1	-8.977
Consumer Discretionary	1	-.110	1	-9.689
Consumer Staples	1	-0.311	1	-9.742
Energy	1	-2.786	1	-9.596
Financial	1	-1.410	1	-8.948
Industrial	1	-0.691	1	-10.207
Information Technology	1	-1.725	1	-10.706
Material	1	-1.718	1	-10.842
Gold	1	-1.861	1	-12.608
Telecommunication Service	1	-.765	1	-9.483
Utility	2	-1.342	1	-8.957
60 Largest companies	1	-1.438	1	-9.130
Treasury Bills	3	-1.764	2	-6.099
Daily Data				
S&P/TSX Composite	1	-1.437	1	-45.474
NYSE Composite	1	-1.314	1	-46.349
NASDAQ Composite	1	-.618	1	-44.589
N225	1	-1.449	1	-46.658
FTSE All-Share	2	-1.477	1	-47.773
Hang Seng	1	-1.771	1	-46.967
AEX	1	-1.825	1	-46.702
BFL20	2	-1.797	1	-45.667
CAC 40	2	-2.499	1	-47.904
ISEQ All-Share	2	-1.234	1	-46.610
BSESN	1	-.676	1	-44.869
DAX Performance	1	-1.018	1	-45.884
S&P/ASX 200	1	-1.688	1	-46.796
SMI MID	1	-2.060	5	-29.449
Bovespa Index	1	-1.373	1	-46.662
IBEX 35	2	-2.568	1	-47.170
5% Critical Value		-2.88		-2.88

B.3.2 Cointegration

A linear combination of non-stationary time series can be stationary. In this case, these variables are said to be cointegrated. One can test for cointegration or long-run equilibrium relationship between variables by performing the ADF test described earlier on their regression residuals.

Table B.3 shows the OLS estimates of the bivariate relationship between the benchmark indices of some major stock exchanges and the S&P/TSX Composite as well as the statistics from the ADF test performed on their residuals. Out of the 15 exchanges, only two are cointegrated with the TSX: the Hong Kong Stock Exchange and the Bombay Stock Exchange. As a matter of fact, the ADF test statistics from the regression of the Hang Seng on the S&P/TSX Composite and that from the regression of the Bombay Stock Exchange sensitive index on the S&P/TSX Composite are both greater in absolute value than their 5% asymptotic critical value.

B.4 The Derivation of the CAPM

The CAPM relates the performance of an asset to that of the stock market. There are several ways of deriving the CAPM. In this section, following Breeden (1979), I have derived this relation from a consumption optimization problem.

Let C_t denote the consumption at time t of a representative investor endowed with a utility $U(C_t)$. The utility function U is increasing and concave, *i.e.* its first derivative is positive, $\mathrm{d}U(C_t)/\mathrm{d}C_t \geq 0$, and its second derivative is negative, $\mathrm{d}^2U(C_t)/\mathrm{d}C_t^2 < 0$.

The representative investor is infinitely-lived and seeks to maximize the discounted value of his expected lifetime utility $\mathrm{E}_t \sum_{t=0}^{\infty} U(C_t)/(1+\rho)^t$, where the parameter ρ denotes his subjective discount rate, also known as time preference rate, and the operator E_t denotes the expectation conditional on the information available at time t. Each time period, he faces the following budget constraint

$$C_t + \mathbf{P}_t'\mathbf{A}_t = (\mathbf{P}_t + \mathbf{D}_t)'\,\mathbf{A}_{t-1},$$

where $\mathbf{A}$, $\mathbf{D}$, and $\mathbf{P}$ are respectively column vectors that list the amount of each of the N financial assets held by the representative investor, the dividends paid, and their prices. The budget constraint states that the representative investor's wealth consists of the market value of the N financial assets he owns plus the dividends he receives for holding them. Out of this wealth, he finances his consumption and acquires new assets.

B.4.1 The Optimization Problem

The representative investor's optimization program can be written as follows

$$\begin{aligned} V(\mathbf{A}_{t-1}, z_t) = \max_{C_t, \mathbf{A}_t} \quad & U(C_t) + \frac{1}{1+\rho}\mathrm{E}_t V(\mathbf{A}_t, z_{t+1}) \\ & + \lambda_t \left[(\mathbf{P}_t + \mathbf{D}_t)'\,\mathbf{A}_{t-1} - C_t - \mathbf{P}_t'\mathbf{A}_t\right]. \end{aligned} \tag{B.16}$$

The first line of (B.16) splits $\mathrm{E}_t \sum_{t=0}^{\infty} U(C_t)/(1+\rho)^t$, the representative investor's discounted lifetime utility, into his instantaneous utility and the discounted value of all

Table B.3: Bivariate Relationships between the Benchmarks of some Major Exchanges(Dependent Variables) and the S&P/TSX Composite (Explanatory Variable): Augmented Engle-Granger Test (with Intercept but no Trend)

	OLS Regression			ADF Test on Residuals	
Dependent Variable	Intercept	Slope	R^2	Lags	τ-statistics
NYSE Composite	1.709	.781	.828	3	-1.717
	(32.56)	(138.06)			
NASDAQ Composite	-2.085	1.076	.550	3	.451
	(-14.55)	(69.69)			
N225	6.087	.366	.141	1	-1.811
	(45.63)	(25.48)			
FTSE All-Share	2.879	.548	.630	4	-1.921
	(46.53)	(82.23)			
Hang Seng	-.674	1.123	.907	2	-4.227
	(-12.70)	(196.31)			
AEX	5.635	.042	.002	2	-1.801
	(43.35)	(3.01)			
BFL20	4.324	.394	.235	2	-1.819
	(41.36)	(34.96)			
CAC 40	5.889	.265	.135	3	-2.165
	(59.56)	(24.90)			
ISEQ All-Share	6.715	.192	.025	2	-1.178
	(38.42)	(10.19)			
BSESN	-15.345	2.652	.895	1	-3.570
	(-114.56)	(183.70)			
DAX Performance	-1.154	1.068	.686	3	-1.527
	(-10.86)	(93.21)			
S&P/ASX 200	.402	.857	.889	3	-2.908
	(9.02)	(178.43)			
SMI MID	4.845	.432	.415	5	-2.115
	(64.21)	(53.13)			
	(-78.17)	(153.13)			
IBEX 35	5.562	.389	.287	2	-2.430
	(61.68)	(40.02)			
5% Critical Value	$t_{5\%}(3971) = 2.13$				$\tau_{5\%} = -3.34$

his future utilities. The function V is referred to as value function. For more details on value function, see among others, Sargent (1987) or Stockey, Lucas Jr, and Prescott (1989).

The value function V depends on variables describing both the financial state of the representative investor, $\mathbf{A}_{t-1}$, and the level of the technology used to produce the good he consumes, z_t. [6] The variable z_t is the only source of uncertainty in the model. The variable λ_t is the Lagrange multiplier.

The first-order conditions from the optimization program are

$$C_t: \quad U'(C_t) = \lambda_t \tag{B.17a}$$

$$\mathbf{A}_{t-1}: \quad \nabla V(\mathbf{A}_{t-1}, z_t) = \lambda_t (\mathbf{P}_t + \mathbf{D}_t) \tag{B.17b}$$

$$\mathbf{A}_t: \quad \frac{1}{1+\rho} \mathrm{E}_t \nabla V(\mathbf{A}_t, z_{t+1}) = \lambda_t \mathbf{P}_t, \tag{B.17c}$$

where the symbol ∇ denotes the gradient vector. Thus, ∇V is the gradient vector of V, *i.e.* the column vector consisting of its partial derivatives.

One can get rid of the gradient vector of V in (B.17c) by plugging in the first lead of (B.17b),

$$\frac{1}{1+\rho} \mathrm{E}_t \lambda_{t+1} (\mathbf{P}_{t+1} + \mathbf{D}_{t+1}) = \lambda_t \mathbf{P}_t.$$

Relation (B.17a) and its first lead also can be plugged into (B.17c) to get rid of the Lagrange multiplier

$$\frac{1}{1+\rho} \mathrm{E}_t U'(C_{t+1}) (\mathbf{P}_{t+1} + \mathbf{D}_{t+1}) = U'(C_t) \mathbf{P}_t. \tag{B.18}$$

Relation (B.18), referred to as Euler equation or pricing equation, governs the representative investor's choice between consumption and investment in financial assets. For a given asset $i = 1, \ldots, N$, this relation can be written as follows

$$\frac{1}{1+\rho} \mathrm{E}_t \left[\frac{U'(C_{t+1})}{U'(C_t)} (1 + R_{i,t+1}) \right] = 1 \tag{B.19}$$

where $R_{i,t+1}$, the return on asset i at time $t+1$, equals $(P_{i,t+1} + D_{i,t+1})/P_t$. It follows from (B.19) that if an asset is risk-free, its gross return will equal

$$1 + R_{f,t+1} = \frac{1+\rho}{\mathrm{E}_t \left[\frac{U'(C_{t+1})}{U'(C_t)} \right]}. \tag{B.20}$$

Given the covariance between two variables, say x and y, equals $\mathrm{E}(xy) - \mathrm{E}(x)\mathrm{E}(y)$, the lhs element of the pricing equation (B.19) can be written as follows

$$\frac{1}{1+\rho} \mathrm{E}_t \left[\frac{U'(C_{t+1})}{U'(C_t)} (1 + R_{i,t+1}) \right] = \frac{1}{1+\rho} \mathrm{cov}_t \left[\frac{U'(C_{t+1})}{U'(C_t)}, (1 + R_{i,t+1}) \right]$$

[6] Following Mehra and Prescott (1985), one can assume that z_t is expected to grow at a rate γ_{S_t} that is subject to a Markov chain, *viz* technological change undergoes occasional shifts driven by the unobserved variable S_t describing the state of the economy.

$$+\frac{1}{1+\rho}\mathrm{E}_t\left[\frac{U'(C_{t+1})}{U'(C_t)}\right]\mathrm{E}_t(1+R_{i,t+1})$$

The above relation along with (B.19) and (B.20) imply that

$$\mathrm{E}_t(R_{i,t+1})-R_{f,t+1}=-\frac{\mathrm{cov}_t\left[U'(C_{t+1}),R_{i,t+1}\right]}{\mathrm{E}_t\left[U'(C_{t+1})\right]}. \tag{B.21}$$

According to Stein's lemma, if two variables, say x and y, are jointly normally disturbed, then $\mathrm{cov}\left[g(x),y\right]=\mathrm{E}\left[g'(x)\right]\mathrm{cov}(x,y)$. Applying this result to (B.21) gives

$$\mathrm{E}_t(R_{i,t+1})-R_{f,t+1}=-\frac{\mathrm{E}_t\left[U''(C_{t+1})\right]}{\mathrm{E}_t\left[U'(C_{t+1})\right]}\mathrm{cov}_t\left(C_{t+1},R_{i,t+1}\right). \tag{B.22}$$

Let's now define the variable $R_{c,t+1}=C_{t+1}/C_t-1$. This growth rate will be referred to as consumption return. Since the representative investor's wealth consists only of financial assets, growth in his consumption depends on the average return of his portfolio. Replacing C_{t+1} in (B.22) with its expression in terms of $R_{c,t+1}$ and then rearranging gives

$$\mathrm{E}_t(R_{i,t+1})-R_{f,t+1}=-\frac{\mathrm{E}_t\left[U''(C_{t+1})\right]}{\mathrm{E}_t\left[U'(C_{t+1})\right]}C_t\mathrm{cov}_t\left(R_{c,t+1},R_{i,t+1}\right). \tag{B.23}$$

For $R_{i,t+1}=R_{c,t+1}$, relation (B.23) becomes

$$\frac{\mathrm{E}_t(R_{c,t+1})-R_{f,t+1}}{\mathrm{var}_t\left(R_{c,t+1}\right)}=-\frac{\mathrm{E}_t\left[U''(C_{t+1})\right]}{\mathrm{E}_t\left[U'(C_{t+1})\right]}C_t, \tag{B.24}$$

which means the risk-adjusted excess return on consumption equals the representative investor's relative risk aversion. Finally plugging this latter result into (B.23), gives the consumption CAPM.

$$\begin{aligned}\mathrm{E}_t(R_{i,t+1})-R_{f,t+1}&=\beta_{c,i}\left[\mathrm{E}_t(R_{c,t+1})-R_{f,t+1}\right],\\ \text{with } \beta_{c,i}&=\frac{\mathrm{cov}_t\left(R_{c,t+1},R_{i,t+1}\right)}{\mathrm{var}_t\left(R_{c,t+1}\right)}.\end{aligned} \tag{B.25}$$

The slope parameter $\beta_{c,i}$, called systematic risk or consumption beta of asset i, measures the sensitivity of returns on this asset to variations in the returns on consumption.

B.4.2 The Consumption versus the Traditional CAPM

The formulation of the consumption CAPM differs from that of the traditional CAPM in the way the systematic risk is measured. According to the traditional CAPM, the appropriate measure of the risk of an asset is rather the covariance of its return with the market return (proxied as the return on a stock market index). Mankiw and Shapiro (1984) test empirically both the consumption CAPM and the traditional CAPM hypotheses using a cross-section of 464 stocks of companies listed continuously on the NYSE between 1959 and 1982. They estimate the following model

$$R_i=a_0+a_1\beta_{m,i}+a_1\beta_{c,i}+u_i,$$

where u_i denotes the error term. While the consumption CAPM appears preferable on theoretical grounds, it received little empirical support as the coefficient on the market beta, $\beta_{m,i}$ turned out to be larger and more statistically significant. Some possible reasons for the poor performance of the consumption CAPM are: consumption data suffer from measurement errors and many consumers do not actively take part in the stock market.

To estimate the CAPM consistently with the evidence of Mankiw and Shapiro (1984), I will replace the return on consumption R_{ct} with the market return R_{mt} in (B.25). To do this, I assume that all investors populating the economy are identical and have the same return on consumption. Besides, firms are perfectly competitive and produce a single good. This gives the traditional CAPM

$$\begin{aligned} \mathrm{E}(R_{i,t+1}) - R_{f,t+1} &= \beta_{m,i}\left[\mathrm{E}(R_{m,t+1}) - R_{f,t+1}\right], \\ \text{with } \beta_{m,i} &= \frac{\operatorname{cov}\left(R_{m,t+1}, R_{i,t+1}\right)}{\operatorname{var}_t\left(R_{m,t+1}\right)}. \end{aligned} \tag{B.26}$$

Note that the market beta can be written as follows

$$\beta_{m,i} = \frac{\operatorname{cov}(R_{i,t+1}, R_{m,t+1})}{\sum_{i=1}^{N} \omega_i \operatorname{cov}(R_{i,t+1}, R_{m,t+1})}, \tag{B.27}$$

where ω_i is the weight of asset i in the market portfolio. Relation (B.27) indicates that the market beta of asset i is its covariance risk relative to the weighted average covariance risk of all assets.

B.4.3 The Market Excess Return and the Risk Aversion

Consider now relation (B.24). Assuming that the representative investor's relative risk aversion is constant will imply that the market excess return is also constant. One can achieve constant relative risk aversion using a utility of the type

$$\mathcal{U}\left(C_t\right) = \frac{C_t^{1-e} - 1}{1-e},$$

where $e = -C_t\mathcal{U}''\left(C_t\right)/\mathcal{U}'\left(C_t\right) > 0$ is referred to as coefficient of relative risk aversion. [7]

Jagannathan and Wang (1996), Abdymomunov and Morley (2011), and Vendrame, Guermat, and Tucker (2018) attribute the empirical failure of the CAPM to the hypothesis of constant relative risk aversion that constrains the excess market return to remain unchanged over time.

[7] This utility embeds several well-known functions as special cases. For $e = 0$, the utility is linear and, for $e = 1$, it is logarithmic.

References

Abdymomunov, A., and J. Morley (2011): "Time variation of CAPM betas across market volatility regimes," *Applied Financial Economics*, 21(19), 1463–1478.

Barten, A. P. (1987): "The coefficient of determination for regression without a constant term," in *The Practice of Econometrics*, pp. 181–189. Springer.

Bauwens, L., A. Preminger, and J. V. Rombouts (2010): "Theory and inference for a Markov switching GARCH model," *The Econometrics Journal*, 13(2), 218–244.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, 31(3), 307–327.

Breeden, D. (1979): "An lntertemporal Asset Pricing Model with Stochastic Investment and Consumption Opportunities," *Journal of Financial Economics*, 6, 273–296.

Casella, G., and R. L. Berger (2002): *Statistical inference*, vol. 2. Duxbury Pacific Grove, CA.

Davidson, R., and J. G. MacKinnon (1993): *Estimation and inference in econometrics*. Oxford University Press.

Dickey, D. A., and W. A. Fuller (1979): "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American statistical association*, 74(366a), 427–431.

Eisenhauer, J. G. (2003): "Regression through the origin," *Teaching statistics*, 25(3), 76–80.

Engle, R. (2002): "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models," *Journal of Business & Economic Statistics*, 20(3), 339–350.

Engle, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica: Journal of the Econometric Society*, pp. 987–1007.

Engle, R. F., and C. W. Granger (1987): "Co-integration and error correction: representation, estimation, and testing," *Econometrica: journal of the Econometric Society*, pp. 251–276.

Episcopos, A. (1996): "Stock return volatility and time-varying betas in the Toronto Stock Exchange," *Quarterly Journal of Business and Economics*, pp. 28–38.

Fama, E. F., and K. R. French (2004): "The capital asset pricing model: Theory and evidence," *The Journal of Economic Perspectives*, 18(3), 25–46.

Fan, J., and Q. Yao (2008): *Nonlinear time series: nonparametric and parametric methods*. Springer Science & Business Media.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993): "On the relation between the expected value and the volatility of the nominal excess return on stocks," *The journal of finance*, 48(5), 1779–1801.

Gray, S. F. (1996): "Modeling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics*, 42(1), 27–62.

GREENE, W. H. (2000): "Econometric analysis (International edition)," .

GREGORY, A. W., AND B. E. HANSEN (1996): "Residual-based tests for cointegration in models with regime shifts," *Journal of Econometrics*, 70, 99–126.

HAGGAN, V., AND T. OZAKI (1981): "Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model," *Biometrika*, 68(1), 189–196.

HAMILTON, J. D. (1990): "Analysis of time series subject to changes in regime," *Journal of econometrics*, 45(1-2), 39–70.

HAMILTON, J. D. (1994): *Time series analysis*, vol. 2. Princeton university press Princeton.

HAMILTON, J. D., AND G. LIN (1996): "Stock market volatility and the business cycle," *Journal of Applied Econometrics*, 11(5), 573–593.

HAMILTON, J. D., AND R. SUSMEL (1994): "Autoregressive conditional heteroskedasticity and changes in regime," *Journal of econometrics*, 64(1-2), 307–333.

HENRIKSSON, R. D., AND R. C. MERTON (1981): "On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills," *Journal of business*, pp. 513–533.

JAGANNATHAN, R., AND Z. WANG (1996): "The conditional CAPM and the cross-section of expected returns," *The Journal of finance*, 51(1), 3–53.

JOHANSEN, S. (1988): "Statistical analysis of cointegration vectors," *Journal of economic dynamics and control*, 12(2), 231–254.

KHAN, T. A. (2011): "Cointegration of international stock markets: An investigation of diversification opportunities," *Undergraduate Economic Review*, 8(1), 7.

KOSE, M. A., S. CLAESSENS, AND M. E. TERRONES (2011): "Financial Cycles: What? How? When?," .

KRITZMAN, M., S. PAGE, AND D. TURKINGTON (2012): "Regime shifts: Implications for dynamic strategies (corrected)," *Financial Analysts Journal*, 68(3), 22–39.

LAM, K. S. (2001): "The conditional relation between beta and returns in the Hong Kong stock market," *Applied Financial Economics*, 11(6), 669–680.

LEBARON, B. (1992): "Some relations between volatility and serial correlations in stock market returns," *Journal of Business*, pp. 199–219.

LI, Q., AND J. S. RACINE (2007): *Nonparametric econometrics: theory and practice*. Princeton University Press.

LING, S., AND M. MCALEER (2002): "Stationarity and the existence of moments of a family of GARCH processes," *Journal of Econometrics*, 106(1), 109–117.

MANKIW, N. G., AND M. D. SHAPIRO (1984): "Risk and return: Consumption versus market beta," .

MEHRA, R., AND E. C. PRESCOTT (1985): "The equity premium: A puzzle," *Journal of monetary Economics*, 15(2), 145–161.

MSCI BARRA, AND STANDARD & POORS (2018): "Global industry classification standard (gics)," Discussion paper, Technical report, Standard & Poors.

NARAYAN, P. K., AND R. SMYTH (2004): "Modelling the linkages between the Australian and G7 stock markets: common stochastic trends and regime shifts," *Applied Financial Economics*, 14(14), 991–1004.

——— (2005): "Cointegration Of Stock Markets Between New Zealand, Australia And The G7 Economies: Searching For Co-Movement Under Structural Change," *Australian Economic Papers*, 44(3), 231–247.

NELSON, D. B. (1991): "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica: Journal of the Econometric Society*, pp. 347–370.

OZAKI, T. (1980): "Non-linear time series models for non-linear random vibrations," *Journal of Applied Probability*, pp. 84–93.

PETTENGILL, G. N., S. SUNDARAM, AND I. MATHUR (1995): "The conditional relation between beta and returns," *Journal of Financial and quantitative Analysis*, 30(1), 101–116.

PFAFF, B., E. ZIVOT, M. STIGLER, AND M. B. PFAFF (2016): "Package 'urca'," *Unit root and cointegration tests for time series data. R package version*, pp. 1–2.

RUPPERT, D. (2011): *Statistics and data analysis for financial engineering*, vol. 13. Springer.

SARGENT, T. J. (1987): *Dynamic Macroeconomic Theory*.

SHARPE, W. F. (1992): "Asset allocation: Management style and performance measurement," *Journal of portfolio Management*, 18(2), 7–19.

STOCKEY, N. L., R. E. LUCAS JR, AND E. C. PRESCOTT (1989): *Recursive Methods in Economic Dynamics*.

TU, J. (2010): "Is regime switching in stock returns important in portfolio decisions?," *Management Science*, 56(7), 1198–1215.

VAN NORDEN, S., AND H. SCHALLER (1993): "The predictability of stock market regime: evidence from the Toronto Stock Exchange," *The Review of Economics and Statistics*, pp. 505–510.

VENDRAME, V., C. GUERMAT, AND J. TUCKER (2018): "A conditional regime switching CAPM," *International Review of Financial Analysis*, 56, 1–11.

VERBEEK, M. (2008): *A guide to modern econometrics*. John Wiley & Sons.

ZUCCHINI, W., AND I. L. MACDONALD (2009): *Hidden Markov models for time series: an introduction using R*. CRC press.