Essays on Asset Pricing and Downside Risk

Bruno Cara Giovannetti

Submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2011
Abstract

Essays on Asset Pricing and Downside Risk
Bruno Cara Giovannetti

This dissertation contributes to the recent and diverse literature on the relation between downside risk and asset prices.

In chapter one, we use a famous quote among professional investors, "focus on the downside, and the upside will take care of itself", to motivate a representative consumer-investor who only cares about the downside. The consumption-based asset pricing model that emerges from this idea explains the main existing puzzles found within the asset pricing literature. These include the equity premium and the risk-free rate puzzles, the countercyclicality of the equity premium and the procyclicality of the risk-free rate. The model is parsimonious, requiring only three preference-related parameters: the time discount factor, the elasticity of intertemporal substitution, and the downside risk aversion. When we use the model to understand the relation between returns and consumption in the US, we find that the fitted parameter values are consistent with what is expected from the micro foundations.

In chapter two, we show that the model proposed in chapter one can also explain the financial puzzles in other developed countries. This is an important step in the empirical validation of the model. The estimated parameters are robust across highly capitalized countries and qualitatively close to the ones obtained for the US.
Moreover, the risk measure under the quantile utility model can better justify the differences in risk premia across countries when compared to the risk measure under the expected utility model.

In chapter three, we evaluate the effect of margin requirements on asset prices, an additional channel for the relation between downside risk and prices. We provide evidences of the existence of an aggregate margin-related premium in the economy. In particular, we show that (i) a margin-related factor is able to predict future excess returns of the S&P 500 and (ii) stocks with high betas on the margin-related factor pay on average higher returns compared those with low margin betas. These result are important not only to understand asset prices, but also the unconventional polices implemented by the Fed during the great recession of 2007-2010. Although data on margin requirements for the S&P 500 futures are publicly available, it is in general very hard to obtain information on margins for other assets. Given that, we also propose a nonparametric model for estimating margins as a function of the asset’s value at risk. This is theoretically justifiable and has good empirical results.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents</td>
<td>i</td>
</tr>
<tr>
<td>List of Tables</td>
<td>iv</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vi</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>viii</td>
</tr>
<tr>
<td>Preface</td>
<td>ix</td>
</tr>
<tr>
<td><strong>1 Asset Pricing under Quantile Utility Maximization</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Quantile utility maximization and asset pricing</td>
<td>8</td>
</tr>
<tr>
<td>1.2.1 Elements</td>
<td>8</td>
</tr>
<tr>
<td>1.2.2 Asset pricing</td>
<td>15</td>
</tr>
<tr>
<td>1.3 Dynamics, model solution and simulation</td>
<td>21</td>
</tr>
<tr>
<td>1.3.1 Dynamics 1: constant economic uncertainty</td>
<td>21</td>
</tr>
<tr>
<td>1.3.2 Dynamics 2: stochastic economic uncertainty</td>
<td>26</td>
</tr>
<tr>
<td>1.3.3 What is a reasonable value for $\tau$ ?</td>
<td>32</td>
</tr>
<tr>
<td>1.3.4 Comparing results</td>
<td>35</td>
</tr>
</tbody>
</table>
Contents

1.4 Model estimation ................................................................. 37
  1.4.1 A general estimation method ............................................ 38
  1.4.2 A simple two-step estimation procedure .............................. 45
  1.4.3 Empirical results ......................................................... 47
1.5 Conclusion ................................................................. 50

2 International Data and the Quantile Utility Asset Pricing Model ... 52
  2.1 Introduction ................................................................. 52
  2.2 International data ........................................................ 54
  2.3 Fitting the model to each country ....................................... 60
    2.3.1 Model estimation ...................................................... 63
    2.3.2 Updated data set ..................................................... 65
  2.4 Conclusion ................................................................. 70

3 An Empirical Evaluation of the Effects of Margins on Asset Prices . 71
  3.1 Introduction ................................................................. 71
  3.2 The theoretical model .................................................... 76
  3.3 Testable implications of the model ..................................... 79
    3.3.1 Predicting the market portfolio return ............................ 80
    3.3.2 The margin factor and the cross-section of expected returns ... 93
  3.4 Further empirical investigation ......................................... 97
    3.4.1 Omitting consumption ............................................... 98
    3.4.2 Modelling margins .................................................. 99
    3.4.3 Results with augmented sample .................................. 107
  3.5 Conclusion ................................................................. 113
Contents

Bibliography ................................................................. 114

Appendix ................................................................. 123
List of Tables

1.1 Configuration of the model parameters for simulation. 29
1.2 Simulated moments. 30
1.3 Estimates under no serial-correlation 48
1.4 Estimates under serial-correlation 49
2.5 International descriptive statistics 56
2.6 International parameters under lognormality 61
2.7 International risk aversion under expected utility 62
2.8 International estimated parameters 64
2.9 Updated international descriptive statistics 67
2.10 International estimated parameters - updated sample 69
3.11 1-month ahead regressions 86
3.12 2-month ahead regressions 86
3.13 4-month ahead regressions 87
3.14 12-month ahead regressions 87
3.15 Controlling for possible isolated effects of the ted spread and margins. 92
3.16 Cross-sectional regressions with 10 and 25 portfolios. 95
3.17 The effect of omitting consumption 99
List of Tables

3.18 Using fitted margins instead of true margins ........................................ 105
3.19 Omitting consumption and using fitted margins ................................. 106
3.20 Descriptive statistics ........................................................................ 107
3.21 Turbulent period: 12-month ahead regression ..................................... 112
## List of Figures

1.1 The quantile utility agent’s reasoning. ................................................. 9
1.2 A simple lottery and certainty equivalents. ........................................... 34
1.3 Relation between $\tau$ and $\gamma$ ....................................................... 35
2.4 Mean excess return X risk under expected utility. ............................... 57
2.5 Mean excess return X risk under expected utility (high capitalized countries). .................................................. 58
2.6 Mean excess return X risk under quantile utility. ............................... 59
2.7 Mean excess return X risk under quantile utility (high capitalized countries). .................................................. 60
2.8 International risk-return trade-off (expected utility) - updated sample .... 67
2.9 International risk-return trade-off (quantile utility) - updated sample ...... 68
3.10 The time-series for $x_t$ ................................................................. 83
3.11 The time-series for $m_t$ and $\psi_t$ .................................................... 84
3.12 The time-series for $x_t m_t \psi_t$ (level not identified) .......................... 85
3.13 The margin-related premium .......................................................... 90
3.14 10 portfolios sorted by margin beta ................................................. 94
3.15 Margin x fitted margin (in sample) ................................................... 103
3.16 Comparing in and out-of-sample fitted values ................................... 104
List of Figures

3.17 The factor for the augmented sample ($\tilde{m}_t\psi_t$) ........................................ 108
3.18 t-statistic vs. ted standard deviation (1-month ahead regressions) .......... 110
3.19 t-statistic vs. factor ($m_t\psi_t$) standard deviation (1-month ahead regressions) ................................................................. 111
Acknowledgments

I would like to thank my advisor Dennis Kristensen for invaluable support on this dissertation. This work would not have been possible without his guidance and suggestions along the way. I also thank Andrew Ang, Pierre-André Chiappori and Marcelo Moreira for constant and crucial advice. I am deeply indebted to them as well. Many other people have directly contributed to this work. Among them, I would like to highlight John Donaldson, Guilherme Martins, Marcos Nakaguma, Ricardo Reis, and Bernard Salanié.

I have to express my deepest gratitude to my parents Marcio and Salete, who are the foundation of all my accomplishments. Thank you for everything.

This dissertation is dedicated to my wife Priscilla. Thank you so much for all your love, support, company, and encouragement during this time. You are the responsible for this.
Preface

This dissertation contributes to the recent and diverse literature on the relation between downside risk and asset prices.

Distinct theoretical models have been developed to understand and formalize such a relation. For example, Bekaert, Hodrick and Marshall (1997), Epstein and Zin (2001), Barberis, Huang and Santos (2001), Ang, Bekart and Liu (2005) and Routledge and Zin (2010) address some standard financial and macroeconomic puzzling facts employing asymmetric preferences (agents somehow overweight bad outcomes when evaluating a risky situation). In a different framework, considering heterogeneous-risk-aversion agents facing margin constraints, Garleanu and Pedersen (2009) demonstrate that margin requirements, which are directly related to downside risk, may be relevant to determine prices (Brunnermeier and Pedersen 2008, Gromb and Vayanos 2010 and Geanakoplos 2010 present similar results). In addition, the disaster models constitute another example of the connection between asset prices and downside risk. Barro (2006) and Kelly (2009), in the same spirit as Reitz (1988), show that the mere potential for infrequent extremely bad events can have important effects on asset prices.¹

The dissertation consists of essentially two parts, both of which are related to the literature above. In the first part, developed in chapters one and two, we contribute to the research on asset pricing under asymmetric preferences. In the second part, developed

¹ As Donaldson and Mehra (2008) point out, asymmetric preferences models and disasters models can be seen as dual to one another: either agents in the model must effectively be very sensitive to bad outcomes, or it is the outcomes themselves that must be very bad.
in chapter three and co-authored with Guilherme B. Martins, we investigate the effect of margins on asset prices.

In chapter one, we use a famous quote among professional investors, "focus on the downside, and the upside will take care of itself", to motivate a representative consumer-investor who only cares about the downside. The consumption-based asset pricing model that emerges from this idea explains the main existing puzzles found within the asset pricing literature. These include the equity premium and the risk-free rate puzzles, the countercyclicality of the equity premium and the procyclicality of the risk-free rate. The model is quite parsimonious, requiring only three preference-related parameters: the time discount factor, the elasticity of intertemporal substitution, and the downside risk aversion. When we use the model to understand the relation between returns and consumption in the US, we find that the fitted parameter values are consistent with what is expected from the micro foundations.

The parsimony of the model is a relevant characteristic. The good empirical results from Barberis, Huang and Santos (2001) and Routledge and Zin (2010) indicate that the consideration of asymmetric preferences over good and bad outcomes is a promising path for theories on choices and, in particular, for a well-accepted resolution of the asset pricing puzzles. Nevertheless, the large number of preference-related parameters in these models (six and five, respectively), which is crucial for their success, is a delicate issue. First, it is not easy to translate the models into a comprehensive view of the whole process. Second, it is hard to assign precisely the corresponding importance of each parameter to the obtained results. Finally, and perhaps most problematic, matching data by augmenting the
parametric dimension is subject to the standard over-fitting critique. Given its parsimony, the model developed in chapter one addresses all these issues.

Chapter two takes chapter one’s model to international data. As Campbell (1999, 2003) shows, the standard financial puzzles are also present in other developed countries. Hence, we have an opportunity to submit our model to an additional test. Would it be successful if confronted with an international data set? Chapter two presents evidences of a positive answer to this question. By estimating the model for ten developed countries, we obtain reasonable estimates for the risk and intertemporal preferences in general. The estimated parameters are robust across highly capitalized countries and qualitatively close to the ones obtained for the US. We compare our results to Campbell’s (2003), who estimates the canonical expected utility model for the same countries. Moreover, we show that the risk measure under the quantile utility model can better justify the differences in risk premia across countries when compared to the risk measure under the expected utility model.

The second part of the dissertation is developed in chapter three. As mentioned above, a number of recent theoretical papers have been suggesting that margins can affect asset prices in periods where risk tolerant agents are credit constrained.

The relation between margin requirements and downside risk is straightforward. When an investor buys stocks on margin, some money is put up by him (initial margin), and the remainder is borrowed from the broker, with the purchased shares used as collateral. How does the broker define the maximum lending amount? According to the collateral evalu-
ated at a worst-case scenario. Therefore, the worse the worst-case scenario, the higher the initial margin requirement.

The result of margins affecting prices would be important not only to understand asset prices per se, but also the unconventional credit policy implemented by the Fed during the great recession of 2007-2010. The size and composition of the Fed’s balance sheet has suffered major changes in the past three years. In January 2007, the Fed carried no risk of default in its assets, holding basically US Treasury bills ($780 billion). During the crisis, however, a variety of asset were included in the balance sheet in significant amounts. For example, commercial papers ($350 billion), repurchase agreements ($150 billion), mortgage-backed securities ($1 trillion), Federal agency debt securities ($150 billion) and others ($100 billion). In December 2010, the total size of the balance sheet was almost $2.5 trillion.

As Geanakoplos (2010) points out, the negative effect of margins on prices, together with the fact that these elements feed back one each other, could justify such a radical change in the credit policy. According to him, during some periods, "the Fed must step around the banks and lend directly to investors, at more generous collateral levels than the private markets are willing to provide."

In addition, the margin premium may break the usual non arbitrage link between the Fed fund rate and the rate of returns of other assets, affecting the ability of the monetary authority to promote an expansionary policy. As we shall see in chapter three, the margin premium is the product of the margin requirement, the cost of margin, and the importance of the leveraged agents in aggregate consumption. The cost of margin is equal to the shadow
cost of capital, which can be measured by the difference between the uncollateralized and
the collateralized short term rates. The latter is closely related to the Fed fund rate, while
the former depends on the liquidity and credit condition in the interbank market. Hence,
during a financial crisis, when margin constraints are binding, a reduction in the Fed fund
rate may not translate into a fall on the rate of returns of other assets. The reason is that
the consequently higher shadow cost of capital steepens the margin-return relation, and this
increases the required return on assets with high margin requirements. Since in bad periods
margins are significantly higher across assets, the interest rate reduction can then have
small, zero, or even a positive effect on the required return of other assets in the economy.

Despite the importance of this result, empirical evidence is still scarce. Chapter
three contributes to fill this gap, finding empirical support for the existence of an aggre-
gate margin-related premium.

Our empirical findings are related to both the time-series and cross-section of returns.
In particular, we show that (i) a margin-related factor is able to predict the future excess
returns of the usual proxy for the market portfolio (S&P 500), and (ii) portfolios with high
betas on the margin factor pay on average higher returns in relation to those with low
margin betas.

Although data on margin requirements for the S&P 500 futures are publicly available,
it is in general very hard to obtain information on margins for other assets. Given that,
chapter three also proposes a nonparametric model for estimating margins as a function of
the asset’s value at risk. This is theoretically justifiable and has good empirical results.
Chapter 1
Asset Pricing under Quantile Utility Maximization

1.1 Introduction

A famous quote among professional investors is "Focus on the downside, and the upside will take care of itself". In this paper, we consider a representative consumer-investor who follows this advice. Surprisingly, the consumption-based asset pricing model that emerges from this idea explains the main existing puzzles found within the asset pricing literature. These include the equity premium and the risk-free rate puzzles, the countercyclicality of the equity premium and the procyclicality of the risk-free rate.

In the proposed model, the consumer-investor is concerned with the so-called downside risk. This is done by replacing the standard setting of expected utility optimizing agents with the concept of quantile utility. Under this framework, the agent summarizes a risky situation using a worst-case scenario which is a function of his downside risk aversion. The more downside risk averse the agent, the worse the worst-case scenario he considers. The $\tau$ quantile of a continuous random variable can be interpreted as the worst possible outcome that can occur with probability $1 - \tau$. Hence, instead of maximizing the expected value of his utility function, the agent maximizes a given $\tau$ quantile of it. As we will see, $\tau$ defines his downside risk aversion: the lower $\tau$, the higher the downside risk aversion.\(^2\)

---

\(^2\) A search of this sentence on the internet returns many results.

\(^3\) One could say that the agent’s objective function is given by the value at risk (VaR) of his utility. However,
This is a novel extension of the static decision-theoretical framework developed by Manski (1988) and Rostek (2010) for a dynamic asset pricing setting. In a standard economy with one risky and one risk-free asset, we can derive an arbitrage-free asset pricing model, where both main characteristics of the canonical expected utility consumption-based approach (Hansen and Singleton (1982), Mehra and Prescott (1985), hereinafter, the canonical model) are modified. The equity premium is no longer based on the covariance between the risky return and the consumption growth. Instead, it is a linear function of the risky return standard deviation. In addition, risk aversion and elasticity of intertemporal substitution (EIS), which are linked throughout a single parameter in the canonical model, are automatically disentangled in a simple way.

These two endogenous changes are the main drivers of the good empirical results. Since stock returns historically have a high standard deviation, the price of such a risk, i.e., the level of downside risk aversion, will not have to be high to match the empirical excess returns. Moreover, the attitude towards intertemporal substitution is not polluted by risk preferences.

To reproduce (i) the first and second moments of the risk-free return, the equity premium, and the consumption growth, (ii) the low covariance between risky return and consumption growth, (iii) the countercyclical risk premium, and (iv) the procyclical risk-free rate that we see in data, our model requires only three parameters related to preferences: a downside risk aversion ($\tau$) of about 0.43, an EIS ($\psi$) of about 0.5 and a time discount factor ($\beta$) of less than 1. A downside risk aversion of such a magnitude is reasonable in that it

---

since $\tau$ here is a free parameter defining preference towards risk, it is not restricted to being close to zero (as in standard VaR applications).
produces reasonable certainty equivalents for bets on continuously distributed random variables (stock indexes, for example). By comparing certainty equivalents under quantile and expected utility maximization, an agent with this level of downside risk aversion is analogous to an expected utility agent with a relative risk aversion coefficient of 3. According to Mehra and Prescott (1985) reasonable values for such a parameter would be between 1 and 10. An EIS of about 0.5 is also an acceptable value. In a recent work using micro-data, Engelhardt and Humar (2009) estimate the EIS to be 0.74, with a 95% confidence interval that ranges from 0.37 to 1.21. Using macrodata and separating stockholders from nonstockholders, Vissing-Jorgensen (2002) estimates the EIS around 0.4 and 0.9 for these respective groups.

To illustrate the main differences between the predictions of our framework and the predictions of the canonical model, we first derive equations in closed-form for the risky return, the risk-free rate, and the equity premium. These equations come from combining the Euler equations of the quantile agent with the standard assumption of joint lognormality of returns and consumption growth. In order to replicate the well-evidenced existence of predictability in future excess returns, we then allow for time-varying economic uncertainty in the aggregate economy dynamics. From this, a countercyclical risk premium and a procyclical risk-free rate are produced.

Taking the model to data, we first perform simulation exercises, matching the first and second moments of consumption growth, risk-free rate and excess returns. Then, to evaluate the model free of distributional assumptions, we propose a GMM-based estimation method for its parameters.
The derived Euler equations impose restrictions on the functional forms of the conditional \( \tau \) quantiles of consumption growth and excess return. They are well-defined functions of the period-by-period risk-free rate and of the other parameters related to preferences. However, as \( \tau \) in this framework is not given (it is the downside risk aversion to be estimated), the standard asymptotic results for quantile regressions as a GMM problem do not apply. Hence, we derive sufficient conditions for the parameters to be globally identified and for the proposed estimator to be consistent.

The fact that the model separates risk and time preferences allows us to estimate the EIS. This is a useful result of this paper. Under the standard technology for disentangling EIS and risk aversion (Epstein and Zin’s (1989) preferences), one has to use instrumental variables to estimate the EIS. This is what Hall (1988) and Campbell (2003) do for example. Such estimations were recently found to suffer from weak-instruments related issues\(^4\) and therefore are not reliable (see Neely, Roy, and Whiteman (2001) and Yogo (2004), for instance). However, the EIS estimation under our model does not require the use of any instrument.

We conclude the introduction by positioning this study in the related literature. The research in asset pricing can be separated according to the modifications proposed with respect to the canonical model. Such modifications are about (i) preferences, (ii) market and asset structure, and (iii) the endowment process. Group (i) could be further divided into two branches: (i.i) preferences inside and (i.ii) preferences outside the expected utility

\(^4\) To estimate the EIS under Epstein and Zin’s preferences one has to use instruments for consumption growth or returns. Since both of these variables are only weakly predictable, the instruments are weak.

The current study belongs to branch (i.ii), which was initiated by Epstein and Zin (1989) and Weil (1989). These authors use the recursive preferences of Kreps and Porteous (1978) as a way of separating time and risk preferences, something that is not possible under the canonical model. By disentangling risk aversion and EIS, they end up with a three-parameter model which is able to generate a reasonable level for the risk-free rate. However, since no innovation in the risk dimension is made, a high level of risk aversion is still necessary to fit the equity premium.

Epstein and Zin (1990, 2001) and Bekaert, Hodrick and Marshall (1997) investigate the use of Gul’s (1991) disappointment aversion preferences to explain the equity premium puzzle.\(^5\) According to these preferences, outcomes below the certainty equivalent are over-weighted relative to outcomes above it. Although such preferences are a one-parameter extension of the expected utility framework, these papers extend the canonical model in two parameters, since they also use the model of Epstein and Zin (1989) to disentangle risk aversion and EIS. However, they are able to fit the equity premium with only a slightly lower, still unreasonable, risk aversion level.\(^6\)

Going further, Routledge and Zin (2010) extend the disappointment aversion model in one additional dimension. They generalize Gul’s preferences by defining an outcome


\(^6\) Bonomo and Garcia (1993) show that it is crucial to combine Gul’s preferences with a joint process for consumption and dividends that follows a Markov switching model in order to match the first and second moments of risk-free and excess returns under reasonable parameter values. However, a model such as that would be in both groups (i.ii) and (iii) defined above.
as disappointing only when it is sufficiently far (defined by the new parameter) from the certainty equivalent. Since their model also separates risk aversion and EIS using Epstein and Zin (1989) preferences, they are a three-parameter extension of the expected utility model, resulting in a total of five preference-related parameters. Under this richer structure, the disappointment aversion-based framework is finally able to address the financial puzzles successfully.

An alternative way of considering the fact that people care asymmetrically about good and bad outcomes is provided by the prospect theory of Kahneman and Tversky (1979). Applying prospect theory to asset pricing, Barberis, Huang and Santos (2001) are also able to reproduce the financial data patterns under reasonable parameter values. In their model, the representative agent derives direct utility not only from consumption, but also from changes in the value of his financial wealth. Moreover, he is more sensitive to negative movements in his financial wealth than to positive movements. Besides that, such a sensitivity also is a function of the agent’s past portfolio experience: if he had losses in the past relative to a time-varying benchmark, he now is more sensitive to further losses. A functional form reflecting this mechanism is imposed by the researchers.

Barberis, Huang and Santos’s (2001) model also employs a large number of preference-related parameters; six, to be exact. The first two are the time discount factor and the relative risk aversion related to consumption. The third is the agent’s extra sensitivity to losses in his portfolio wealth. The forth defines how previous losses impact the third parameter. The fifth determines how the benchmark used by the agent to define gains and losses

---

evolves over time. The sixth controls the overall importance of utility from gains and losses in financial wealth relative to utility from consumption.

The good empirical results from Barberis, Huang and Santos (2001) and Routledge and Zin (2010) indicate that consideration of asymmetric preferences over good and bad outcomes is a promising path for theories on choices and, in particular, for a well-accepted resolution of the asset pricing puzzles. Nevertheless, the large number of preference-related parameters in these models, which is crucial for their success, is a delicate issue. First, it is not easy to translate the models into a comprehensive view of the whole process. Second, it is hard to assign precisely the corresponding importance of each parameter to the obtained results. Finally, and perhaps most problematic, matching data by augmenting the parametric dimension is subject to the standard over-fitting critique. According to this critique, the larger number of parameters may simply describe better the noise in the data, rather than the underlying economic relationships. In other words, these models could be providing spurious data-fitting.\footnote{This tense relationship between the augmentation of the expected utility framework with additional parameters and the over-fitting critique is raised, for instance, by Zin (2002). Based on that article, Watcher (2002) claims that "behavioral models leave room for multiple degrees of freedom in the utility function. Taken to an extreme, this approach could reduce structural modeling to a tautological, data-fitting exercise" and "I believe that parsimony lies at the root of what Zin refers to as reasonableness. A parsimonious model is a model in which the number of phenomena to be explained is much greater than the number of free parameters."}

The present paper addresses these issues. The quantile utility criterion comes from a loss-function that asymmetrically weighs good and bad outcomes, the well-known check loss-function. Hence, the derived model under this framework belongs to the class of models related to asymmetric preferences. Moreover, the model is quite parsimonious, requiring only three preference-related parameters: the time discount factor; the EIS; and the
downside risk aversion. Finally, it solves the main asset pricing puzzles addressed by Barberis, Huang and Santos (2001) and Routledge and Zin (2010).

The rest of this work is organized as follows. Section 1.2 presents the quantile utility agent in its general form and derives some basic results of asset pricing under quantile maximization. Section 1.3 solves the model under lognormality and simulates from it. Section 1.4 discusses how to estimate the model free of distributional assumptions and presents the results. Section 1.5 concludes.

### 1.2 Quantile utility maximization and asset pricing

In this section, we first present the elements of the quantile utility model, following Manski (1988) and Rostek (2010). Then, we apply this theoretical-decision framework to asset pricing.

#### 1.2.1 Elements

A general choice theory for quantile maximizing agents was developed recently. Rostek (2010) is the first study to axiomatize the quantile utility agent. Notwithstanding, the quantile maximization model for decision making under uncertainty was first proposed 22 years ago by Manski (1988).

The main idea is simple. An agent, when facing a situation where he has to choose among uncertain alternatives, picks the one that maximizes some given quantile of the utility distribution instead of its mean, as in the expected utility model. In this framework,
the agent cares about the worst outcome that can happen with a given probability. For instance, the given quantile can be the median of the utility distribution, or the 0.25 quantile. In the case of the 0.25 quantile for example, when evaluating an uncertain situation, he looks at the worst outcome that can occur with 75 percent probability (i.e., the chance of the realized scenario being better than the scenario he considers is 75 percent).

The quantile of concern is an intuitive measure of pessimism. If agent $A$ looks at the worst that may happen in 90 percent of the situations, i.e., quantile 0.10, and agent $B$ looks at the worst that may happen in 60 percent of the situations, i.e., quantile 0.40, we would naturally classify agent $B$ as more optimistic than agent $A$ : agent $A$ picks a more conservative scenario to summarize the lottery. Figure 1.1 illustrates this for a lottery that follows a normal distribution.

![Figure 1.1. The quantile utility agent’s reasoning.](image-url)
As we shall see below, the quantile of concern defines also the agent’s downside risk preference. Hence, downside risk preference is closely related to our standard notion of optimism-pessimism.

(i) Asymmetric preference

Because of the characteristics of his loss-function, we can say that the quantile agent cares asymmetrically about good and bad outcomes. This intuition comes from Manski (1988), based on the work of Wald (1937).

Assume that an agent has to evaluate an uncertain situation where $U$ is his utility level which can have different values in different states of the world. This uncertain situation is represented by the cumulative distribution function of $U$, denoted by $F_U$. According to the standard framework in decision theory introduced by Wald (1937), this agent should summarize (evaluate) $F_U$ using the criterion $\omega^*$ that minimizes the expected value of his loss-function, i.e., his risk-function.

A possible loss-function could be the square loss. In this case, he would summarize $F_U$ using

\[
\omega^* = \arg \min_{\omega \in \mathbb{R}} \int_{\mathbb{R}} (z - \omega)^2 \, dF_U(z)
\]

\[
= \int_{\mathbb{R}} z \, dF_U(z).
\]

Hence, he would use the expected utility criterion of von Neumann and Morgenstern (1944) and Savage (1954). This allows us to interpret the expected utility agent as someone who is evenly worried with underpredictions and overpredictions of his utility level in a risky
situation and uses squares ($L^2$ norm) to compute the distances between the utility level predictions and realizations.

What if the decision maker was asymmetrically worried about under and overpredictions of his future utility level? We could describe a situation like that by the check loss-function of Koenker and Basset (1978). In this case, he would evaluate $F_U$ using

$$
\omega^* = \arg\min_{\omega \in \mathbb{R}} \int_{\mathbb{R}} (1 - \tau) |z - \omega| \times 1 [z < \omega] + \tau |z - \omega| \times 1 [z \geq \omega] dF_U(z)
$$

$$
= Q^\tau (U),
$$

where $Q^\tau (U)$ is the $\tau_{th}$ quantile of the random variable $U$ (if $F_U$ is continuous, $Q^\tau (U) = F_U^{-1}(\tau)$).

Therefore, a quantile maximizer can be described as someone who asymmetrically weighs underpredictions and overpredictions of his future utility level, in the ratio $(1 - \tau) / \tau$, and uses absolute values ($L_1$ norm) to compute the distances. In this case, the agent’s evaluation criterion is the $\tau_{th}$ quantile of his utility, that is, the worst possible utility level that may happen with probability $(1 - \tau)$. This is the optimal criterion to summarize $F_U$ given his asymmetric concern with the upper tails of utility distributions relative to their lower tails.

---

9 Such an agent could also compute distances under the $L^2$ norm. In this case, his criterion to evaluate $F_U$ would be the expectiles of Newey and Powell (1987)

$$
\omega^* (\tau) = E(U) + \left( \frac{2\tau - 1}{1 - \tau} \right) E [(U - \omega^* (\tau)) \times 1 [U < \omega^* (\tau)]] .
$$
(ii) Quantile agent definition

We now define the quantile agent in a more formal way. Let $S$ be a set of states of the world $s \in S$, and $\mathcal{X}$ be an arbitrary set of payoffs $x, y \in \mathcal{X}$. Then, the agent has to choose among simple acts $h : S \rightarrow \mathcal{X}$, which map from states to payoffs. Let $\mathcal{A}$ be the set of all such acts, and $E = 2^S$ be the set of all events. Define $\pi$ to be a probability measure on $E$, and $u$ a utility function over payoffs $u : \mathcal{X} \rightarrow \mathbb{R}$. For each act, $\pi$ induces a probability distribution over payoffs, referred to as a lottery. Given that, let $G, H$ denote the random variables (payoffs) induced by the acts $g, h \in \mathcal{A}$, respectively. Finally, define $F_G$ and $F_H$ as the lotteries induced by the acts $g$ and $h$, i.e., the cumulative distribution functions of $G$ and $H$, respectively.

A decision maker is defined as a $\tau$-quantile maximizer if there exists a unique $\tau \in [0, 1]$, a probability measure $\pi$ on $E$, and a utility function $u$, such that for all $g, h \in \mathcal{A}$,

$$g \succeq h \iff Q^\tau (u(G)) > Q^\tau (u(H)).$$

As always, we can think in terms of the lotteries:

$$F_G \succeq F_H \iff Q^\tau (u(G)) \geq Q^\tau (u(H)).$$

(iii) Downside risk aversion

For the standard expected utility agent, we may understand risk preferences using the following logic.

First we define riskiness. We say that the lottery $F_H$ is riskier than the lottery $F_G$ if $F_G$ second-order stochastic dominates\textsuperscript{10} (SSD) $F_H$ (see Rothschild and Stiglitz (1970)).

\textsuperscript{10} $F_G$ SSD $F_H$ if and only if
Then, we define $\Upsilon$ to be the class of all pairs of lotteries that SSD one another, i.e., $\Upsilon = \{(F_G, F_H) : F_G \text{ SSD } F_H\}$. It is natural to classify agent $A$ as more risk averse than agent $B$ if for all pairs of distributions in $\Upsilon$, whenever $B$ prefers a distribution which SSD the other, so does $A$. Finally, we show that this will be the case if and only if the utility function of agent $A$ is "more concave" than the utility function of agent $B$, i.e., $u_A(x) = \psi(u_B(x))$, where $\psi(\cdot)$ is an increasing concave function. Given that, we conclude that risk-aversion is described by the concavity of the utility function.

Manski (1988) and Rostek (2010) follow the same logic to attach the quantile maximizer’s attitude toward risk to the quantile he maximizes. The central point is that riskiness is characterized in a different way, the so-called downside risk: $F_H$ involves more downside risk than $F_G$ if $F_G$ crosses $F_H$ from below. We say that lottery $F_G$ crosses lottery $F_H$ from below if there exists $x, y \in \mathcal{X}$, such that $F_G(y) \leq F_H(y)$ for all $y < x$ and $F_G(y) \geq F_H(y)$ for all $y > x$. That is, downside risk is related to the probability of bad outcomes.\footnote{If $F_G$ and $F_H$ have the same mean, and $F_H$ has more downside risk than $F_G$, then $F_H$ has also more (second-order stochastic dominance) risk than $F_G$. However, under different means, this is not true.}

Just as above, considering the class of all pairs of lotteries with the single-crossing property, $\Phi = \{(F_G, F_H) : F_G \text{ crosses } F_H \text{ from below}\}$, we say that individual $A$ is more downside risk averse than individual $B$ if, for all pairs of distributions in $\Phi$, whenever $B$ prefers a distribution which crosses the other from below, so does $A$. Given that, we can show that agent $A$ is more downside risk averse than agent $B$ if and only if $\tau_A < \tau_B$, and

$$\int_{-\infty}^{x} [F_H(t) - F_G(t)] \, dt \geq 0, \text{ for any } x \in \mathcal{X}.$$
then $\tau$ can be defined as the downside risk aversion parameter in the decision model: the lower $\tau$, the more downside risk averse the agent.

But what role does the concavity of the utility function play under this framework? Because of the property of equivariance of quantiles to monotonic transformations, the answer to this question is "none", at least for static decision problems.

(iv) Equivariance of quantiles to monotonic transformations

A key aspect of the quantile utility model is that static decisions are invariant to any strictly increasing transformation of the utility function. This is described in Proposition 1 in Manski (1988).

If $m : \mathbb{R} \to \mathbb{R}$ is a strictly increasing function, and $X$ is a random variable, then

$$Q^\tau(m(X)) = m(Q^\tau(X)).$$

Hence, for lotteries $F_G$ and $F_H$,

$$F_G \succeq F_H \iff Q^\tau(u(G)) \geq Q^\tau(u(H))$$

$$\iff u^{-1}(Q^\tau(u(G))) \geq u^{-1}(Q^\tau(u(H)))$$

$$\iff Q^\tau(G) \geq Q^\tau(H),$$

where the second line follows from the fact that $u$ is a strictly increasing function.

---

12 The intuition under this result is that a strictly increasing transformation of the random variables doesn’t change the order of the values of their support.
Therefore, for static problems, the agent’s decision does not depend on $u$. Manski (1988) and Rostek (2010) refer to this as a robustness property: the choice is unaffected by misspecification of the utility function.

However, the utility function is relevant in intertemporal choices. When the utility function has more than one argument, it is not possible to use the equivariance property to get rid of $u$. In particular, under time-separability, the concavity of the utility function defines the preference towards intertemporal substitution as usual. This is going to play an important role in the asset pricing theory, allowing the downside risk aversion and the EIS to be disentangled. This idea is not in Manski (1988) or in Rostek (2010) and, to the best of our knowledge, is explored for the first time in the present study.

### 1.2.2 Asset pricing

We now apply the quantile maximization decision theory to the standard intertemporal problem of a consumer-investor agent. First, we define the consumption-investment problem and solve for the Euler equations that the agent must respect in equilibrium. Then we discuss the Law of One Price and the no-arbitrage condition under this framework.

The model to be considered has 2 periods. As Karni and Schmeidler (1991) show, once we depart from expected utility, one of the following three assumptions has to be relaxed: (i) time consistency; (ii) consequentialism; or, (iii) reduction of compounded lotteries. Assumptions (i) and (ii) are in the heart of the Principle of Optimality of dynamic programming (see Rust (2006), section 3.6). Therefore, to be able to solve a multiple-period problem outside of the expected utility framework by standard dynamic program-
ming, one must relax assumption (iii). However, by relaxing (iii), one would be including preferences about the time of resolution of the uncertainty in the model, just as in the recursive preferences of Kreps and Porteus (1978) and Epstein and Zin (1989). Since the central goal of this study is to develop a simple, parsimonious and stylized model to address the over-fitting critique within the asymmetric preferences literature, we restrict the model to a 2-period framework.

The economy has two assets, one risky and one risk-free. Define the value of the risky asset at \( t + 1 \) to be \( X_{t+1} = P_{t+1} + D_{t+1} \), where \( P_{t+1} \) is the price of the asset at \( t + 1 \) and \( D_{t+1} \) is the value of some cash flow the investor received between \( t \) and \( t + 1 \) (in the case of a stock, \( D \) is the dividend). Define \( X_{t+1}^f \) to be the value of the risk-free asset at \( t + 1 \) and \( P^f_t \) its price at \( t \). Let \( C_t \) be the agent’s consumption at \( t \), \( \xi \) and \( \xi^f \) be the quantity of the risky and risk-free assets he buys at \( t \) respectively, and \( W_t \) be his initial wealth. Then, under time-separability, he solves:

\[
\begin{align*}
\max_{\xi, \xi^f} & \quad Q_t^\tau(u(C_t) + \beta u(C_{t+1})) \\
\text{subject to} & \quad C_t = W_t - P_t \xi - P^f_t \xi^f \\
& \quad C_{t+1} = X_{t+1} \xi + X_{t+1}^f \xi^f
\end{align*}
\]

where \( \beta \) is the time discount factor, \( u \) is the utility function, \( Q_t^\tau(x) \) is the \( \tau^{th} \) quantile of the conditional distribution of the random variable \( x \) (conditional on the information set available at time \( t \)).

---

13 Indeed, according to Rust (2006), recursive preference is the only class of non-expected utility preferences that allows the use of standard dynamic programming (backward induction) to solve multi-period problems.
1.2 Quantile utility maximization and asset pricing

This agent derives utility only from consumption, as usual, and cares about the worst outcome (in terms of the utility for both periods) that may occur with probability \((1 - \tau)\). In other words, this agent follows the famous advice "Focus on the downside, and the upside will take care of itself". As discussed in sub-section 1.2.1, the higher his level of downside risk aversion, the lower \(\tau\).

A key feature of problem (1.2) is that downside risk aversion and elasticity of intertemporal substitution (EIS) are automatically disentangled. This is a direct consequence of the quantile’s equivariance for monotonic transformations. Note that, according to equation (1.1), we have

\[
Q_t^\tau \left( u(C_t) + \beta u(C_{t+1}) \right) = u(C_t) + \beta u(Q_t^\tau (C_{t+1})),
\]

since \(u\) is a strictly increasing function.

Hence, all uncertainty in problem (1.2) is resolved by parameter \(\tau\), since \(Q_t^\tau (C_{t+1})\) is deterministic at \(t\). The only role played by \(u\) is to discount consumption across time: depending on the concavity of \(u\), the agent will combine present consumption, \(C_t\), and the certainty equivalent of future consumption (which, for the quantile maximizer, is equal to \(Q_t^\tau (C_{t+1})\)). Hence, the concavity of \(u\) will only define the EIS, denoted by \(\psi\). Specializing \(u(c) = \frac{c^{1-\gamma}-1}{1-\gamma}\), we have \(\psi = \frac{1}{\gamma}\).\(^{14}\) Note that such an assumption for the functional form of \(u\) imposes no restriction on risk preference: it simply restricts the EIS to being constant.

\(^{14}\) Defining \(U(C_t, Q_t^\tau (C_{t+1})) = \frac{C_t^{1-\gamma}-1}{1-\gamma} + \beta \frac{(Q_t^\tau (C_{t+1}))^{1-\gamma}-1}{1-\gamma}\)

we have that
The EIS parameter, $\psi = \frac{1}{\gamma}$, defines the degree of substitutability-complementarity between consumption today, $C_t$, and the certainty equivalent of consumption tomorrow, $Q^*_t (C_{t+1})$. For $\psi \to 0$, $C_t$ and $Q^*_t (C_{t+1})$ become perfect complements, and we have the agent’s object function given by

$$ U (C_t, Q^*_t (C_{t+1})) = \min \{C_t, Q^*_t (C_{t+1})\} . $$

At the other extreme, for $\psi \to \infty$, $C_t$ and $Q^*_t (C_{t+1})$ become perfect substitutes, i.e., the agent maximizes

$$ U (C_t, Q^*_t (C_{t+1})) = C_t + \beta Q^*_t (C_{t+1}) . $$

For the intermediate case of $\psi = 1$, we end up with the Cobb-Douglas

$$ U (C_t, Q^*_t (C_{t+1})) = C_t (Q^*_t (C_{t+1}))^{\beta} . $$

With respect to the time discount factor $\beta$, its role is to determine the marginal rate of substitution between $C_t$ and $Q^*_t (C_{t+1})$. Therefore, $\psi$ defines the degree of substitutability-complementarity between $C_t$ and $Q^*_t (C_{t+1})$, and $\beta$ parameterizes such a relation.\(^{15}\)

What are the implications of the quantile maximization asset pricing model? With the following proposition, proved in the appendix, we initiate this analysis.

\[^{15}\text{On the empirical side, we will see that both parameters are also separately identified by our estimation method.}\]
Proposition 1 Suppose a consumer-investor solves problem (1.2) and \( u(c) = \frac{c^{1-\gamma} - 1}{1-\gamma} \).

Then, the Euler equations are given by

\[
P_t = \beta \left( Q_t^r \left( \frac{C_{t+1}}{C_t} \right) \right)^{-\gamma} Q_t^r (X_{t+1})
\]

(1.3)

\[
P_t^f = \beta \left( Q_t^r \left( \frac{C_{t+1}}{C_t} \right) \right)^{-\gamma} X_{t+1}^f
\]

(1.4)

From now to the end of section 1.3, we study the asset pricing implications of equations (1.3) and (1.4). The first step is to understand whether they respect the Law of One Price and the no-arbitrage condition. Then, we solve the model under the standard assumption of joint lognormality for returns and consumption growth, deriving closed-forms for the risky return, the risk-free rate and the equity premium in equilibrium.

Since we ignore transaction costs, any candidate for an equilibrium pricing system has to respect the Law of One Price: prices should be linear. That is, denoting \( \Xi_t = \left( \xi_t, \xi_t^f \right) \) to be a portfolio formed at \( t \), with price given by \( P_t^\Xi \), the pricing system has to imply \( P_t^\Xi = \xi_t P_t + \xi_t^f P_t^f \). Otherwise, \( P_t \) and \( P_t^f \) cannot be equilibrium prices because of arbitrage opportunities among the individual assets and the portfolio. Equations (1.3) and (1.4) respect this condition. Defining \( \eta_t = \beta \left( Q_t^r \left( \frac{C_{t+1}}{C_t} \right) \right)^{-\gamma} \), we have
where the second line follows from the quantile equivariance. Note that for a degenerate random variable \( x \), \( Q^\tau (x) = x \) for any \( \tau \in [0, 1] \), and this implies \( Q^\tau_t (X_{t+1}^f) = X_{t+1}^f \).

As is well-known, a linear pricing system does not completely rule out arbitrage opportunities. Hence, we need to impose two mild conditions to end up with an arbitrage-free model.

**Proposition 2** Suppose that (i) the risky asset payoff \( X_{t+1} \) is a continuous random variable and (ii) \( \tau \in (0, 1) \). Then, the pricing model given by equations (1.3) and (1.4) rules out arbitrage opportunities.

Both conditions of proposition 2 (proved in the appendix) are reasonable. The continuity of the risky asset payoff comes for free for stock prices. The second condition, more subtle, rules out two well-known agents in decision theory, the so-called MaxMin and MaxMax. The MaxMin agent (\( \tau = 0 \)) summarizes a lottery by looking at the very worst case scenario that may take place (that is, the worst case scenario that may occur with probability 1). On the other hand, the MaxMax (\( \tau = 1 \)) summarizes a lottery by looking at the very best case scenario that may take place (or, in other words, the worst case scenario that may
occur with probability 0. Since both agents represent extreme behaviors (the extremely pessimistic and the extremely optimistic), excluding them is not a restrictive assumption.

1.3 Dynamics, model solution and simulation

We now solve the model in closed-form, under joint lognormality of returns and consumption growth, with both constant and fluctuating economic uncertainty.

Although the solution under constant economic uncertainty is enough to match both the risk-free rate and the risk premium under reasonable levels for the preference-related parameters, it does not generate a time-varying risk premium. To improve the model in this direction, we allow stochastic volatility in the economy dynamics. The model is then simulated under this richer environment.

1.3.1 Dynamics 1: constant economic uncertainty

Assume

\[
\begin{align*}
g_{t+1} &= \mu_c + \eta_{t+1}, \quad \eta_{t+1} \sim iid \ N \left(0, \sigma_c^2\right) \\
r_{t+1} &= \mu_r + u_{t+1}, \quad u_{t+1} \sim iid \ N \left(0, \sigma_r^2\right)
\end{align*}
\]

where \( g_{t+1} = \log \left( C_{t+1}/C_t \right) \), \( r_{t+1} = \log \left( X_{t+1}/P_t \right) \) and \( Cov \left( \eta_{t+1}, u_{t+1} \right) = \sigma_{cr} \).

Under this framework, the closed-forms for the risky return, the risk-free rate and the equity premium are given by the following proposition.
Proposition 3 If returns and consumption growth are jointly lognormally distributed, following (1.5), and the pricing system is given by equations (1.3) and (1.4), then

\[ r_{t+1} = -\log(\beta) + \gamma \mu_c + \Phi^{-1}(\tau)(\gamma \sigma_c - \sigma_r) + u_{t+1} \]  
\[ r^f_{t+1} = -\log(\beta) + \gamma \mu_c + \gamma \sigma_c \Phi^{-1}(\tau) \]  
\[ E_t\left(r_{t+1} - r^f_{t+1}\right) = -\sigma_r \Phi^{-1}(\tau) \]

where \( r^f_{t+1} \) refers to the risk-free asset return and \( \Phi^{-1} \) is the inverse of the cumulative distribution function of a standard normal random variable.

To gain intuition on equations (1.7) and (1.8), it is useful to compare them to the analogous equations from the canonical expected utility model. As first derived by Hansen and Singleton (1983), it is well-known that under expected utility maximization and lognormality of returns and consumption growth we have

\[ r^f_{t+1} = -\log(\beta) + \gamma \mu_c - \frac{1}{2} \gamma^2 \sigma_c^2 \]  
\[ E_t\left(r_{t+1} - r^f_{t+1}\right) = -\frac{1}{2} \sigma_r^2 + \gamma \sigma_{cr}. \]

We first focus on the predictions for the risk-free return. First, in both models, the risk-free rate is linear in expected consumption growth with the slope equal to the inverse of the elasticity of intertemporal substitution. The lower the EIS (i.e., the higher the desire for consumption smoothing across time), the higher the risk-free rate. This effect is increasing
in the expected consumption growth, meaning that the agent will be less willing to save if he expects tomorrow’s consumption to be higher.

Second, also common to both models, the higher the rate at which the agent discounts future utility (the lower $\beta$), the higher the risk-free rate he requires in order to save.

Third, and this is a first novelty of the quantile approach, a higher variability of consumption growth may have either positive or negative effects on the level of the risk-free rate under the quantile model. If $\tau > 0.5$, a high standard deviation of consumption growth generates a high risk-free rate. If $\tau < 0.5$, a high standard deviation of consumption growth generates a low risk-free rate. The intuition for this is clear: if the agent is optimistic ($\tau > 0.5$), a higher variability is interpreted by him as a higher chance of getting a high level of consumption tomorrow and hence, he becomes less willing to save (higher risk-free rate). In the case of pessimism ($\tau < 0.5$), a higher variability is interpreted as a higher chance of getting a low level of consumption tomorrow, which leads the agent to save more (lower risk-free rate). The strength of this effect, as expected, is increasing in the desire of smoothing consumption across time ($\gamma$).

The separation of intertemporal and risk preferences under the quantile model becomes evident when we compare the third terms of equations (1.7) and (1.9). In equation (1.9), we have $\gamma^2$, where one $\gamma$ stands for the risk aversion and the other $\gamma$ is the inverse of the EIS. In equation (1.7), we have the product between the inverse of the EIS and a function of the downside risk aversion.

We now turn to the equity premium equation (1.8). The risk premium does not depend on the covariance between consumption and stock returns as in the canonical model but,
instead, on the standard deviation of the stock return.\textsuperscript{16} A higher standard deviation may require either a higher or a lower expected return, depending again on whether $\tau$ is greater or less than 0.5. The intuition is the same as above: under optimism ($\tau > 0.5$), a high variability is interpreted as a high chance of getting good returns which, therefore, increases prices (decreasing expected returns). Under pessimism ($\tau < 0.5$) a high variability means a high chance of getting bad returns which causes prices to decrease (increasing expected returns).

These differences imply a better performance of the quantile model when taken to data. Because risk and time preferences are now disentangled we have degrees of freedom to fit both the risk-free rate and the equity premium (just as in Epstein and Zin (1989)). Moreover, the source of risk has now changed. Under expected utility, the covariance between consumption and risky return is the source of risk. This is empirically low, generating the necessity of a high risk aversion to match the equity premium. However, under quantile utility, risk is determined by the standard deviation of the risky return. This value is high in data and, therefore, we attenuate the role of the downside risk aversion.

Yearly US data on consumption and returns ranging from 1889 to 2009 can be found on Professor Robert Shiller’s website.\textsuperscript{17} The risky and risk-free returns are from the S&P 500 and 1-year treasury bill, respectively. The series for per capita consumption are based on the NIPA and NBER series of consumption.

\textsuperscript{16} The variance term that shows up in equation (1.10) is simply a Jensen’s inequality adjustment (since the expression is about log returns). All that matters for the difference between the risky and the risk-free returns is the covariance term.

\textsuperscript{17} http://www.econ.yale.edu/~shiller/data.htm, as in November 2010.
According to this data set, the average real stock log return has exceeded the average treasury bills log return in about 5 percent per year in the post-war period. Stock log return has had a standard deviation about 17 percent per year, and the covariance between stock log return and per capita log consumption growth has been about 0.2 percent. Inserting these values into equation (1.10) and solving for $\gamma$, we have $\gamma = 32$. Hence, in order to fit these patterns of the data, the canonical model requires a risk aversion coefficient that is too high (equity premium puzzle).

But let us suppose one is willing to accept $\gamma = 32$. Then we run into the risk-free rate puzzle. The per capita log consumption growth series has presented annual mean and standard deviation of about 2.1 and 2.2 percent, respectively. The risk-free log return has been about 1.4 percent. Calibrating equation (1.9) with these values and solving for the time discount factor ($\beta$), we have an absurd $\beta = 1.59$ (it is unreasonable to assume that people prefer later utility).

Doing the same exercise using the quantile model equations, we first impose the left hand side of (1.8) to be 5 percent and the standard deviation of the risky log return to be 17 percent. Solving for $\tau$, we have $\tau = 0.38$. So, in order to fit the equity premium, the agent has to care about the worst that may happen with probability 62 percent. At a first glance, this does not seem to be a high degree of pessimism. We soon will return to this point.

To compute the time discount factor ($\beta$) necessary to fit the observable risk-free rate we should calibrate equation (1.7) with empirically acceptable values for the EIS. In a recent work using microdata, Engelhardt and Humar (2009) estimate the EIS to be 0.74, with

---

18 When estimating the model, we will be able to separably identify the discount factor and the EIS.
a 95% confidence interval ranging from 0.37 to 1.21. By differentiating between stockholders and nonstockholders and using macrodata, Vissing-Jorgensen (2002) estimates the EIS to be around 0.4 and 0.9, respectively. Given that, we use $\gamma = 1.5$ (i.e., EIS equal to 0.67).\(^{19}\)

Calibrating equation (1.7) with $r_f^{t+1} = 1\%$, $\mu_c = 1.9\%$, $\tau = 0.36$, $\sigma_c = 0.021$ and $\gamma = 1.5$, and solving for $\beta$, we have $\beta = 1.007$, which is much better than 1.46. By increasing $r_f^{t+1}$ to 2\%, we have $\beta = 0.997$, a qualitatively acceptable value (2\% is reasonable number for the average risk-free rate as well).

### 1.3.2 Dynamics 2: stochastic economic uncertainty

A limitation of the quantile model presented so far is that it does not generate a time-varying equity premium (or a time-varying risk-free rate). Because of that, the model cannot theoretically explain two well documented empirical facts: the existence of excess returns predictability and countercyclical risk premia.\(^{20}\) Since a significant part of the current literature on consumption-based asset pricing addresses matching time variation in expected returns, it is important to improve the quantile model in this direction.

One possible way of doing that is to incorporate fluctuating economic uncertainty into the model. Bansal and Yaron (2004) provide empirical evidence that justifies such

\(^{19}\) All of these estimates are obtained under the expected utility framework. Even though the EIS has nothing to do with risk, one could conjecture that if the true model is related to quantile maximization, such estimates might be biased, which would complicate the calibration of $\gamma$ under the quantile model. However, the forthcoming estimates for the EIS that I obtain under the quantile model (which, as we will see, are separably identified from the discount factor) are around these values as well.

\(^{20}\) See Fama and French (1989), Ludvigson and Ng (2007) and Cooper and Priestley (2009), for instance, on the countercyclicality of the risk premium.
a modification. Bansal, Khatchatrian and Yaron (2002) extensively document that a time-varying consumption volatility holds up quite well across different samples and economies. Therefore, we now assume the following dynamics for the real economy:

\[ g_{t+1} = \mu_c + \sigma_t \eta_{t+1} \]  
\[ r_{t+1} = \mu_{r,t} + \varphi \sigma_t u_{t+1} \]  
\[ \sigma^2_{t+1} = \alpha + \rho (\sigma^2_t - \alpha) + \sigma_v v_{t+1} \]

where \( \eta_{t+1}, v_{t+1} \) and \( u_{t+1} \) are now standard gaussian random variables and \( \text{Cov} (\eta_{t+1}, u_{t+1}) = \sigma_{cr} \).

The stochastic volatility fluctuates around \( \alpha \), and \( \rho \) represents how quickly it gets pulled toward its mean. The evidence in Bansal and Yaron (2004) and Bansal, Khatchatrian and Yaron (2002) are of slow-moving fluctuations in economic uncertainty, implying a \( \rho \) close to one. The conditional variances of consumption growth and return are now given by \( \sigma^2_t \) and \( \varphi^2 \sigma^2_t \), respectively, and the conditional covariance between consumption growth and return is now \( \varphi \sigma^2_t \sigma_{cr} \).

Solving for \( \mu_{r,t} \), the next proposition shows that returns and risk premium are now time-variant.
Proposition 4  Under the dynamics defined in equations (1.11), (1.12) and (1.13) and the Euler equations (1.3) and (1.4) we have:

\[ r_{t+1} = -\ln \beta + \gamma \mu_c + (\gamma - \varphi) \sigma_t \Phi^{-1} (\tau) + \varphi \sigma_t u_{t+1} \quad (1.14) \]

\[ r^f_{t+1} = -\ln \beta + \gamma \mu_c + \gamma \sigma_t \Phi^{-1} (\tau) \quad (1.15) \]

\[ E_t \left( r_{t+1} - r^f_{t+1} \right) = -\varphi \sigma_t \Phi^{-1} (\tau) \quad (1.16) \]

If \( \tau < 0.5 \) (the pessimistic agent, as discussed in the previous subsection), periods with higher economic uncertainty are periods with higher demand for saving, and hence, lower risk-free rate. This effect is increasing in the desire for consumption smoothing \( \gamma \), the inverse of the EIS. Moreover, more economic uncertainty raises the risk premium, and this effect is increasing in \( \varphi \) - the parameter that links economic uncertainty to return uncertainty. Therefore, the time-variation goes in the (theoretically-) intuitive direction.

As Bansal and Yaron (2004) claim, consumption and market volatilities are high during recessions. Given that, the risk premium in equation (1.16) is counter-cyclical.\(^{21}\) In addition, equation (1.15) implies a procyclical risk-free rate, in line with data as well.

**Simulation**

We now simulate from this model to better visualize its asset pricing implications. We simulate first the economic uncertainty from equation (1.13) and then feed equations (1.11), (1.14) and (1.15) with this series. As in Campbell and Cochrane (1999), Barberis, Huang and Santos (2001), Bansal and Yaron (2004), Bansal, Kiku and Yaron (2009) and many others, we assume that the decision interval of the agent is monthly but the targeted uncertainty.

\(^{21}\) The counter-cyclical feature of the risk premium in the long-run risk model of Bansal and Yaron (2004) also comes from the presence of the stochastic volatility in the risk-premium equation.
data to match are annual. Therefore, we simulate at the monthly frequency and aggregate to annual data.

The stochastic volatility structure added to the model is identical to the one considered in Bansal and Yaron (2004) and Bansal, Kiku and Yaron (2009), and we calibrate parameters $(\alpha, \rho, \sigma_v)$ with the same values of this last paper. \(^{22}\) With respect to $(\mu_c, \sigma_{cr})$, they are set in accordance the sample mean of the consumption growth and the sample covariance between consumption growth and risky return, respectively.

Given such values, we choose the free parameters $(\varphi, \beta, \tau, \gamma)$ seeking to match the first and second moments of the risk-free rate and excess return, and the second moment of consumption growth. Table 1.1 summarizes the parameters’ optimal choices.

<table>
<thead>
<tr>
<th>parameter for monthly simulation</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$ (mean of economic uncertainty)</td>
<td>0.0072 (^2)</td>
</tr>
<tr>
<td>$\sigma_v$ (standard deviation of log economic uncertainty)</td>
<td>$0.28 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\rho$ (log economic uncertainty persistence)</td>
<td>0.999</td>
</tr>
<tr>
<td>$\mu_c$ (mean consumption log growth)</td>
<td>0.0018</td>
</tr>
<tr>
<td>$\sigma_{cr}$ (covariance between $\eta$ and $u$)</td>
<td>0.5</td>
</tr>
<tr>
<td>$\varphi$ (adjustment of the log return standard deviation)</td>
<td>5.5</td>
</tr>
<tr>
<td>$\beta$ (discount factor)</td>
<td>0.9998</td>
</tr>
<tr>
<td>EIS (inverse of $\gamma$)</td>
<td>0.6</td>
</tr>
<tr>
<td>$\tau$ (downside risk aversion)</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Table 1.1. Configuration of the model parameters for simulation.

\(^{22}\) Equation (1.13) produces a small number (about 5%) of negative values for $\sigma_t^2$, as in Bansal and Yaron (2004) and Bansal, Kiku and Yaron (2009). Following them, I replace these negative values with the smallest positive value generated for $\sigma_t^2$. Obviously, one could model $\log(\sigma_t^2)$ to get rid of this technical problem (but, in this case, it wouldn’t be possible to follow their calibration).
1.3 Dynamics, model solution and simulation

The preference-related parameters \((\beta, \tau, \gamma)\) are close to those from the previous subsection. The time discount factor \((\beta)\) is slightly below one, the EIS of 0.6 implies \(\gamma = 1.66\), and the downside risk aversion is now even smaller with \(\tau = 0.45\).\(^{23}\)

Table 1.2 presents the impacts on the simulated moments of varying both the risk aversion and EIS. The other parameters are kept fixed in accordance with Table 1.1.

<table>
<thead>
<tr>
<th>(\tau)</th>
<th>EIS</th>
<th>(E(r_{-rf}))</th>
<th>(\sigma(r))</th>
<th>(E(rf))</th>
<th>(\sigma(rf))</th>
<th>(E(g))</th>
<th>(\sigma(g))</th>
<th>(\text{cov}(g,r))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.41</td>
<td>0.1</td>
<td>10.0</td>
<td>15.8</td>
<td>3.5</td>
<td>11.2</td>
<td>2.1</td>
<td>2.7</td>
<td>0.2</td>
</tr>
<tr>
<td>0.41</td>
<td>0.6</td>
<td>10.0</td>
<td>15.6</td>
<td>0.8</td>
<td>1.9</td>
<td>2.1</td>
<td>2.7</td>
<td>0.2</td>
</tr>
<tr>
<td>0.41</td>
<td>1.1</td>
<td>10.0</td>
<td>15.9</td>
<td>0.5</td>
<td>1.0</td>
<td>2.1</td>
<td>2.7</td>
<td>0.2</td>
</tr>
<tr>
<td>0.45</td>
<td>0.1</td>
<td>5.5</td>
<td>15.3</td>
<td>11.7</td>
<td>6.2</td>
<td>2.1</td>
<td>2.7</td>
<td>0.2</td>
</tr>
<tr>
<td><strong>0.45</strong></td>
<td><strong>0.6</strong></td>
<td><strong>5.5</strong></td>
<td><strong>15.3</strong></td>
<td><strong>2.1</strong></td>
<td><strong>1.1</strong></td>
<td><strong>2.1</strong></td>
<td><strong>2.7</strong></td>
<td><strong>0.2</strong></td>
</tr>
<tr>
<td>0.45</td>
<td>1.1</td>
<td>5.5</td>
<td>15.3</td>
<td>1.2</td>
<td>0.6</td>
<td>2.1</td>
<td>2.7</td>
<td>0.2</td>
</tr>
<tr>
<td>0.49</td>
<td>0.1</td>
<td>1.0</td>
<td>15.0</td>
<td>19.8</td>
<td>1.2</td>
<td>2.1</td>
<td>2.7</td>
<td>0.2</td>
</tr>
<tr>
<td>0.49</td>
<td>0.6</td>
<td>1.0</td>
<td>15.0</td>
<td>3.5</td>
<td>0.2</td>
<td>2.1</td>
<td>2.7</td>
<td>0.2</td>
</tr>
<tr>
<td>0.49</td>
<td>1.1</td>
<td>1.0</td>
<td>15.0</td>
<td>2.0</td>
<td>0.1</td>
<td>2.1</td>
<td>2.7</td>
<td>0.2</td>
</tr>
<tr>
<td>data</td>
<td></td>
<td>4.8</td>
<td>16.8</td>
<td>1.4</td>
<td>1.7</td>
<td>2.1</td>
<td>2.2</td>
<td>0.2</td>
</tr>
<tr>
<td>s.e.</td>
<td></td>
<td>(1.5)</td>
<td>(1.8)</td>
<td>(0.5)</td>
<td>(0.3)</td>
<td>(0.3)</td>
<td>(0.5)</td>
<td>(0.0)</td>
</tr>
</tbody>
</table>

*other parameters values: following Table 1.1*

Table 1.2. Simulated moments.

From Table 1.2 we see three effects: (i) higher values of downside risk aversion (i.e., lower values of \(\tau\)) increase the mean excess return; (ii) lower values for EIS increase the mean risk-risk free return and its volatility; and, (iii) decreasing \(\tau\) also impacts the mean and standard deviation of the risk-free rate, decreasing the former and increasing the latter.

---

\(^{23}\) Importantly, the quantile model does not need an EIS greater than one to produce good empirical results. This is relevant when compared to Bansal and Yaron (2004). For them, it is crucial for the good results to employ an EIS greater than one, more precisely, equal to 1.5 (and this value is not empirically reasonable, as discussed before.)
The theoretical reasons for the effects related to the first moments are the same as those under constant economic uncertainty. A higher downside risk aversion implies a higher price for the risk, and therefore, a higher risk premium, justifying effect (i). A higher complementarity between consumption at \( t \) and the certainty equivalent of consumption at \( t + 1 \) implies a higher desire to smooth consumption in time, and therefore, a higher risk-free rate to justify savings from \( t \) to \( t + 1 \), which explains effect (ii). Finally, a higher downside risk aversion leads to more savings from period \( t \) to period \( t + 1 \) for a given level of economic uncertainty at \( t \), lowering the risk-free rate and justifying (iii).

With respect to the effects related to the second moment of the risk-free rate, the theoretical explanations are the following. The effect in (ii) comes from the natural fact that the volatility of the risk-free rate is a function of the volatility of the economic uncertainty which is decreasing in the EIS (see equation (1.15)). This makes theoretical sense, since savings should respond more to economic uncertainty, the more the agent cares about smoothing consumption. The reasoning supporting the effect in (iii) follows the same line: the more downside risk averse the agent, the more savings should respond to economic uncertainty.

We therefore conclude that the quantile asset pricing model’s predictions are theoretically solid. In addition, when calibrated with empirically reasonable parameters and \( \tau = 0.45 \), the model is able to reproduce important patterns of financial and macroeconomic data. At this point, a natural question is: how reasonable is \( \tau = 0.45 \)?
1.3.3 What is a reasonable value for $\tau$?

Is $\tau = 0.45$ more reasonable than $\gamma = 35$ (the value obtained in sub-section 1.3.1 for the risk aversion under expected utility and lognormality) in terms of the implied attitude towards risk? Or, what is a reasonable range for $\tau$?

One way to evaluate $\tau$ is to compare the certainty equivalent implicit in a quantile model to the one implicit in a power utility model for risky situations with payoffs following continuous distributions, in accordance with Proposition 2.

Using certainty equivalents of simple bets to relate parameters from different models of behavior towards risk is a standard procedure in this literature. For instance, Epstein and Zin (1990) use such a strategy to compare the risk aversion levels in Yaari preferences with the risk aversion levels in the expected utility preferences (see their Tables 1 and 2). Bonomo and Garcia (1993), Epstein and Zin (2001), Routledge and Zin (2010), among others, do the same.

A simple and natural risky situation to use is the following. Suppose the agent wants to invest $1000 and the investment return follows the same distribution considered in (1.5). Therefore,

$$\ln (X_{t+1}) \sim N \left( \mu_r + \ln (1000), \sigma_r^2 \right),$$

where, as usual, $X_{t+1}$ is the value of the investment at $t + 1$.

For a one-year investment, the sample estimates for $\mu_r$ and $\sigma_r^2$ are about 0.08 and 0.03 respectively. The initial investment value is immaterial for the forthcoming conclusions.
We can first ask: what are the certainty equivalents for a quantile agent with $\tau = 0.45$ and for an expected power utility agent with $\gamma = 35$ for this uncertain outcome $X_{t+1}$?

For an expected utility agent with power utility, the certainty equivalent of a lottery with payoff $x$ is given by

$$CE_{EU} = \left[ E \left( x^{1-\gamma} \right) \right] ^{\frac{1}{1-\gamma}}.$$

For a $\tau$-quantile utility agent, the value of a lottery with payoff $x$ is equal to $Q^\tau \left[ u(x) \right]$. So, the certainty equivalent of such a lottery is the solution of $u(CE_{QU}) = Q^\tau \left[ u(x) \right]$. By quantile equivariance,

$$CE_{QU} = Q^\tau (x).$$

Figure 1.2 presents the histogram of the uncertain investment value at $t + 1$, which has mean and standard deviation around $1103$ and $212$, respectively. The vertical dashed lines are the certainty equivalents for the power utility agent with $\gamma = 35$ and for the quantile agent with $\tau = 0.45$ (they are around $643$ and $1057$, respectively).

A casual review of this figure suggests that the certainty equivalent of a power utility agent with $\gamma = 35$ is too small compared to what one would expect as reasonable. On the other hand, for a quantile agent with $\tau = 0.45$, his certainty equivalent looks much better. However, it is already well-known in the literature that $\gamma = 35$ generates extreme outcomes in an expected utility setting. So, one can argue that basically any alternative utility specification is going to behave more reasonably. Considering that, perhaps a clearer, more illustrative way to proceed would be to ask: which value of $\gamma$ would give the certainty...
Fig. 1.2. Histogram of uncertain payoff and certainty equivalents for $\gamma = 35$ and $\tau = 0.45$.

equivalent obtained with $\tau = 0.45$? The answer is $\gamma = 2.5$. In other words, in terms of certainty equivalents, a quantile utility agent with $\tau = 0.45$ would be analogous to an expected utility agent with $\gamma = 2.5$, a value which is commonly referred to as reasonable in the literature.

Pursuing this idea further, we can relate many values of $\tau$ to many values of $\gamma$ in terms of producing the same certainty equivalent for the bet defined above. Figure 1.3 presents this relationship.

Mehra and Prescott (1985) argue that acceptable values for $\gamma$ would be between 1 and 10. Hence, for the risky situation considered, the analogous interval for $\tau$ would be $[0.22, 0.48]$. 
1.3 Dynamics, model solution and simulation

1.3.4 Comparing results

So far we have compared our results only to those from the canonical model. This was done to illustrate the new features of the present approach with respect to the predictions for the risk-free rate and the equity premium.

In this sub-section we briefly compare the results obtained to those of Epstein and Zin (1989) and Weil (1989) (three parameters), Bonomo and Garcia (1993) (four parameters) and Routledge and Zin (2010) (five parameters), and Barberis, Huang and Santos (2001) (six parameters).

By using recursive preferences, Epstein and Zin (1989) and Weil (1989) disentangle risk aversion and EIS and still have the time discount rate - the same parameters we have here. By doing so, they are able to fit both the equity premium and the risk-free rate.
However, the extremely high risk aversion remains crucial. As Table 1 in Weil (1989) shows, in order to match the average of risk-free and excess returns, risk aversion and EIS have to be set at 45 and 0.1, respectively. If risk aversion is decreased to 1, the premium is as low as 0.45 percent, while the mean risk-free rate reaches 25 percent. Furthermore, nothing is said about second moments.

With one extra parameter compared to our model (the one that regulates the disappointment aversion), the model in Bonomo and Garcia (1993) under a joint random walk for consumption and dividend growth rates produces an average equity premium on the order of 2.5 percent with standard deviation about 12.8 percent. The risk-free rate averages about 4.5 percent. This is the best they are able to get using what they consider reasonable values for their parameters.

By adding one more parameter to the disappointment aversion model, Routledge and Zin (2010) are able to generate good results with this framework. By means of a countercyclical risk aversion (produced by an endogenous variation in the probability of disappointment), they produce a large equity premium (about 6 percent) and a risk-free rate with low volatility and mean. However, they still have difficulty with fitting the risky return volatility and maintaining the 6 percent equity premium at the same time.

Barberis, Huang and Santos (2001) assume a functional form for preferences based on prospect theory, which has 6 parameters. Their model succeeds in explaining the first and second moments of the risk-free rate, the equity premium and the consumption growth,

---

24 Comparable to the dynamics I use here.
and produces a time-varying risk premium (that comes from the impact of the agent’s past portfolio result on his sensitivity for future losses).

1.4 Model estimation

The previous section presented the quantile utility asset pricing model under the assumption of joint conditional lognormality of asset returns and consumption growth. This was useful for building intuition with respect to the model. However, it is well-known that the lognormality assumption is not consistent with all the properties of historical stock returns. For example, stock log returns show weak evidence of skewness and strong evidence of excess kurtosis, at least for short horizons. Hence, it is important to understand how the model performs if we relax the lognormality assumption.

In this section, we discuss how to estimate the model free of distributional assumptions. A GMM-based estimator is proposed, the identification of the parameters is analyzed, and sufficient conditions for consistency are established. Moreover, since the proposed estimator is defined over non-differentiable moments, its asymptotic distribution is derived.

In the appendix, we also estimate the model under the lognormality assumption. This complements the simulation exercise performed in section 1.3 by providing confidence bands to the parameters.
1.4 Model estimation

1.4.1 A general estimation method

The estimation of $\beta$, $\gamma$ and $\tau$ free of any distributional assumption will be performed by combining GMM and quantile regression’s elements. However, since $\tau$, the respective conditional quantile, also has to be estimated, the present problem is distinct from the standard quantile regression, where $\tau$ is taken as given.

In the case of the canonical expected utility model, the standard way of estimating the model free of distributional assumptions is by applying the GMM of Hansen (1982), as was first proposed by Hansen and Singleton (1982). This is straightforward since it is just a matter of transforming conditional into unconditional expectations. However, this is not the case if we want to estimate the quantile Euler equations (1.3) and (1.4). There is nothing analogous to the law of iterated expectations for quantiles. Moreover, equations (1.3) and (1.4) are not even moment conditions. But, as we see now, it is possible to overcome such difficulties in a simple fashion.

Let the vector $\theta_0 = (\tau_0, \beta_0, \psi_0)$ represent the populational values for the downside risk aversion, the time discount factor and the EIS, respectively. Define $Y_{t+1} = \left(\frac{C_{t+1}}{C_t}, R_{t+1}, R_{t+1}^f\right)$ and let $Y \equiv \{Y_t : \Omega \rightarrow \mathbb{R}_+ \times \mathbb{R}, t = 1, ..., T\}$ be a stochastic process defined on a complete probability space $(\Omega, \mathcal{F}, P)$, where $\mathcal{F} \equiv \{\mathcal{F}_t : t = 1, ..., T\}$ and $\mathcal{F}_t \equiv \sigma \{Y_s : s \leq t\}$. Define also $\varepsilon_{c,t+1}$ and $\varepsilon_{r,t+1}$ to be the random variables such that

$$\frac{C_{t+1}}{C_t} = Q^{\tau_0} \left(\frac{C_{t+1}}{C_t} | \mathcal{F}_t\right) + \varepsilon_{c,t+1}$$

(1.17)

and
\( R_{t+1} = Q^{\gamma_0} (R_{t+1} | \mathcal{F}_t) + \varepsilon_{r,t+1}. \) (1.18)

Given this structure, we first note that the asset pricing theory imposes functional forms on the conditional quantiles defined above. From Proposition 1, the risky and risk-free returns in equilibrium should respect the following two equations

\[
\beta_0 \left( Q^{\gamma_0} \left( \frac{C_{t+1}}{C_t} | \mathcal{F}_t \right) \right)^{-1/\psi_0} Q^{\gamma_0} (R_{t+1} | \mathcal{F}_t) = 1 \quad (1.19)
\]

\[
\beta_0 \left( Q^{\gamma_0} \left( \frac{C_{t+1}}{C_t} | \mathcal{F}_t \right) \right)^{-1/\psi_0} R^f_{t+1} = 1. \quad (1.20)
\]

where we now use the EIS parameter \( \psi_0 \) instead of its inverse \( \gamma_0 \).

By dividing equation (1.19) with equation (1.20) we get

\[
Q^{\gamma_0} (R_{t+1} | \mathcal{F}_t) = R^f_{t+1}. \quad (1.21)
\]

Rearranging equation (1.20), we have

\[
Q^{\gamma_0} \left( \frac{C_{t+1}}{C_t} | \mathcal{F}_t \right) = \left( \beta_0 R^f_{t+1} \right)^{\psi_0}. \quad (1.22)
\]

Hence, the theoretical model imposes that, in equilibrium, all the information that matters for the conditional quantiles of \( R_{t+1} \) and \( C_{t+1}/C_t \) is \( R^f_{t+1} \) (which is already known at \( t \), i.e., \( R^f_{t+1} \in \mathcal{F}_t \)). More than that, the model defines the whole functional form of such conditional quantiles.

Given (1.21) and (1.22), we can state the following proposition.
Proposition 5  Let $Z_t$ be an $m \times 1$ vector such that $Z_t \in \mathcal{F}_t$. Define
\[
g(Y_{t+1}, Z_t, \theta_0) = \begin{pmatrix} \tau - 1 \left[ \frac{C_{t+1}}{C_t} < \left( \beta \right) R_{t+1}^l \right] Z_t \\ \tau - 1 \left[ R_{t+1} < R_{t+1}^l \right] Z_t \end{pmatrix}
\]
where $1[\cdot]$ is the logical indicator function.

Then,
\[
E \left[ g \left( Y_{t+1}, Z_t, \theta_0 \right) \mid \mathcal{F}_t \right] = 0.
\] (1.23)

Therefore, we have $2m$ moment conditions and 3 parameters to be estimated. For $m \geq 2$ we may use Hansen’s (1982) GMM approach,
\[
\widehat{\theta} = \arg \min_{\theta \in \Theta \subseteq \mathbb{R}^3} \left( \frac{1}{T} \sum_{t=1}^{T} g_t \left( Y_{t+1}, Z_t, \theta \right) \right)' W_T \left( \frac{1}{T} \sum_{t=1}^{T} g \left( Y_{t+1}, Z_t, \theta \right) \right)
\] (1.24)
where $W_T$ is a general weighting matrix.

Even though the interpretation of a quantile regression as a GMM problem is standard, we cannot directly use the established asymptotic results (from Koenker and Basset (1978) and Powell (1984, 1896), for example). In quantile regressions, $\tau_0$ is a given number and not a parameter to be estimated. Hence, the fact that our central task is the estimation of $\tau_0$ places this econometric problem in a new environment.

We have to understand whether the GMM estimation of $\theta_0$ is indeed feasible. In other words, we have to understand whether $\theta_0$ is identified and derive the consistency and asymptotic distribution of $\widehat{\theta}$. Fortunately, as we see now, we can conclude under mild conditions that $\theta_0$ is globally identified and $\widehat{\theta}$ is consistent and asymptotically normal.

The following proposition presents sufficient conditions for consistency.
Proposition 6  Assume that (i) $V_{t+1} \equiv \left( R_{t+1}, C_{t+1}/C_t, R^f_{t+1}, Z_t \right)$ is strictly stationary and $\alpha$-mixing of size $-r/(r-1)$, with $r > 1$, (ii) $E \| Z_t \| < \infty$, where $\| \cdot \|$ denotes the $L_\infty$-norm, (iii) $\Theta \subseteq \mathbb{R}^3$ is a compact set (iv) $W_T \overset{p}{\to} W_0$, where $W_0$ is a positive definite matrix, (v) $C_{t+1}/C_t$ is a continuous random variable, (vi) $\left( 1, R^f_{t+1} \right)' \in Z_t$ and (vii) $\text{Var} \left( R^f_{t+1} \right) > 0$.

Then, $\tilde{\theta} \overset{p}{\to} \theta_0$, where $\tilde{\theta}$ is defined in equation (1.24).

Assumptions (i), (ii) and (iii) are technical and often present. Assumption (iv) is satisfied by a special choice for $W_T$, as Proposition 7 will show. Assumption (v) is standard in quantile regressions and natural for aggregate consumption growth. Assumption (vi) simply says that the instrument set should include a constant and the risk-free rate. Assumption (vii) is a standard rank condition which requires the explanatory variable to be non-degenerate. Assumptions (v), (vi) and (vii) are the crucial ones for global identification, as can be seen in the proof (in the appendix).

The proof of Proposition 6 shows that $E \left[ g \left( V_{t+1}, \theta \right) \right] = 0$ if and only if $\theta = \theta_0$. By combining this with the fact that $W_0$ is positive definite, we conclude that the populational object-function of our GMM estimator has a unique optimum at $\theta = \theta_0$, that is, $\theta_0$ is globally identified.\textsuperscript{25}

The global identification of the parameters can be seen as a fortunate achievement of the present model. In fact, according to Newey and McFadden (1994), "If $E \left[ g \left( z, \theta \right) \right] = 0 \Leftrightarrow \theta = \theta_0$, then the populational GMM object-function is uniquely minimized at $\theta = \theta_0$. However, as it is trivial to show, if $W_0$ is positive definite, one only needs $E \left( g \left( V_{t+1}, \theta \right) \right) = 0 \Leftrightarrow \theta = \theta_0$ to get the same result."

\textsuperscript{25} Lemma 2.3 in Newey and McFadden (1994) shows that if $W_0$ is positive semi-definite and $W_0 E \left( g \left( V_{t+1}, \theta \right) \right) = 0 \Leftrightarrow \theta = \theta_0$, then the populational GMM object-function is uniquely minimized at $\theta = \theta_0$. However, as it is trivial to show, if $W_0$ is positive definite, one only needs $E \left( g \left( V_{t+1}, \theta \right) \right) = 0 \Leftrightarrow \theta = \theta_0$ to get the same result.
is nonlinear in \( \theta \), then specifying primitive conditions for identification becomes quite difficult ... A practical solution to the problem of global GMM identification, that has often been adopted, is to simply assume identification. This practice is reasonable, given the difficulty of formulating primitive conditions, but it is important to check that it is not a vacuous assumption whenever possible, by showing identification in some special cases."

For instance, as Newey and McFadden (1994) points out, in the canonical model of Hansen and Singleton (1982) it is possible to derive global identification only under a particular form of the conditional distribution.

Proposition 7 now proposes a specific choice for \( W_T \).

**Proposition 7** Suppose that assumption (i) holds, assumption (ii) is strengthened to (ii') there exists some \( \delta > 0 \) such that \( E \| Z_t \|^{2r+2\delta} \) and additionally assume (viii) \( \tau_0 \in (0, 1) \), (ix) \( P(\varepsilon_{c,t+1} < 0, \varepsilon_{r,t+1} < 0 | Z_t) < \tau_0 \) and (x) \( E (Z_t Z_t') \) is nonsingular. Specialize \( W_T \) as

\[
W_T = \left( \frac{1}{T} \sum_{t=1}^{T} g \left( V_{t+1}, \tilde{\theta} \right) g \left( V_{t+1}, \tilde{\theta} \right)' \right)^{-1},
\]

where \( \tilde{\theta} \) is any estimator such that \( \tilde{\theta} \overset{p}{\rightarrow} \theta_0 \).

Then

\[
W_T \overset{p}{\rightarrow} \Sigma_0^{-1},
\]

where

\[
\Sigma_0 \equiv E \left[ g \left( V_{t+1}, \theta_0 \right) g \left( V_{t+1}, \theta_0 \right)' \right]
\]

is positive definite.

As usual, estimator \( \tilde{\theta} \) may be computed in a first step by \( \hat{\theta} \), with \( W_T \) as the identity matrix (according to Proposition 6). Assumption (viii) rules out the MaxMin and Max-
Max agents from the analysis, which had already been done to ensure no-arbitrage in the model. Hence, such agents are not only incompatible with no-arbitrage, but also may jeopardize the identification of the model. Assumption (ix) is also a mild one. First, note that under independence of $\varepsilon_{c,t+1}$ and $\varepsilon_{r,t+1}$, defined in equations (1.17) and (1.18), $P(\varepsilon_{c,t+1} \leq 0, \varepsilon_{r,t+1} \leq 0 | Z_t) = \tau_0^2$, and this is satisfied. Hence, this assumption is about $\varepsilon_{c,t+1}$ and $\varepsilon_{r,t+1}$ not being too positively correlated. But, note that in the extreme case of positive correlation, where $\varepsilon_{c,t+1} = \varepsilon_{r,t+1}$, we have $P(\varepsilon_{c,t+1} \leq 0, \varepsilon_{r,t+1} \leq 0 | Z_t) = \tau_0$. Therefore, imposing $P(\varepsilon_{c,t+1} \leq 0, \varepsilon_{r,t+1} \leq 0 | Z_t) < \tau_0$ is not restrictive at all. Assumption (x) is the usual rank condition on the instruments.

Note that Proposition 5 implies that $\{g(Y_{t+1}, Z_t, \theta_0), \mathcal{F}_t\}$ is a martingale difference sequence. Consequently, $g(Y_{t+1}, Z_t, \theta_0)$ is not serially correlated, and $\Sigma_0$ defined in proposition 7 is the asymptotic variance of the moment conditions.

We now turn to the asymptotic distribution of $\widehat{\theta}$. To address the nondifferentiability of $g(\cdot)$, we use the empirical processes theory approach presented in Andrews (1994) which, under some regularity conditions, replaces the differentiability of $g(\cdot)$ by the differentiability of $E[g(\cdot)]$. The next proposition derives the asymptotic distribution of the estimator.

**Proposition 8** Suppose all assumptions of Proposition 6 hold, where assumption (ii) is strengthened to (ii') of Proposition 7. Furthermore, assume that (xi) $f_{\varepsilon_{c,t+1}}(0 | Z_t)$ is bounded away from zero, and (xii) the matrix $G'_0 W_0 G_0$ is nonsingular, where $G_0 \equiv \nabla_{\theta} E(g(V_{t+1}, \theta_0))$ is a $2m \times 3$ matrix with entries
\[ G_{i1} = E(Z_{it}) \]
\[ G_{i2} = -\psi_0^\beta_{0}^{(\psi_0^{-1})} E(f_{x_{c,t+1}}(0|Z_t) (R_{t+1}^f)^{\psi_0} Z_t) \]
\[ G_{i3} = -\beta_0^{(\psi_0)} E(f_{x_{c,t+1}}(0|Z_t) (R_{t+1}^f)^{\psi_0} \log(\beta_0 R_{t+1}^f) Z_t) \]
\[ G_{j1} = E(Z_{jt}) \]
\[ G_{j2} = 0 \]
\[ G_{j3} = 0 \]

for \( i = 1, \ldots, m \) and \( j = m + 1, \ldots, 2m \).

Then

\[ \sqrt{T} \left( \hat{\theta} - \theta_0 \right) \xrightarrow{d} N \left( 0, (G_0' W_0 G_0)^{-1} G_0' W_0 \Sigma_0 W_0 G_0 (G_0' W_0 G_0)^{-1} \right) . \]

Assumption (xi) is standard in quantile regressions, and rules out having zero in the denominator. Assumption (xii) implies the existence of the term \((G_0' W_0 G_0)^{-1}\) in the asymptotic variance. Proposition 8 tells us that the usual GMM asymptotic distribution for differentiable moments conditions is valid for our nondifferentiable specific case as well. This implies that the optimal choice for \( W_T \) is the one that converges in probability to \( \Sigma_0^{-1} \), which is the weighting matrix defined in Proposition 7. The optimal weighting matrix simplifies the estimator’s asymptotic variance to \((G_0' \Sigma_0^{-1} G_0)^{-1}\).
1.4.2 A simple two-step estimation procedure

Functions such as (1.24) are difficult to optimize by the standard packages algorithms (*fminsearch*, in MATLAB, or *nlm* and *optim* in R, for instance): they are nonsmooth and highly nonconvex, with numerous local optima. However, as we have only 3 parameters with well defined theoretical bounds (such as \( \tau_0 \in [01, .99] \), \( \beta_0 \in [0.9, 1.1] \) and \( \psi_0 \in [0, 5] \)), the optimization is feasible using a grid search in our case.

Nevertheless, it is useful to note that \( \theta_0 \) can be consistently estimated in an even simpler manner, using a two-step procedure. Such an estimator is not going to be efficient, but this discussion builds intuition into the model and provides a rapid and simple technology for estimating, for instance, the EIS (the estimation of the EIS under Epstein and Zin (1989) preferences, the alternative technology of disentangling risk and time preferences, is much more involving).

In a first step, we estimate \( \tau_0 \). Equation (1.21) implies

\[
E \left[ \tau_0 - 1 \left[ R_{t+1} < R^f_{t+1} \right] \right] = 0.
\]

Hence, a consistent estimator of \( \tau_0 \) is

\[
\widetilde{\tau} = \frac{1}{T} \sum_{t=1}^{T} \left[ R_{t+1} < R^f_{t+1} \right],
\]

which is the relative number of observations in the sample such that \( R_{t+1} < R^f_{t+1} \). From standard arguments, its asymptotic distribution is given by

\[
\sqrt{T} (\widetilde{\tau} - \tau_0) \xrightarrow{d} N \left( 0, \tau_0 (1 - \tau_0) \right).
\]
Given $\tau$, we can now estimate $(\beta_0, \psi_0)$ by a standard linear quantile regression. This is the case since, by the equivariance property of quantiles, equation (1.22) implies

$$Q^{\tau_0}(g_{t+1}|F_t) = \lambda_0 + \psi_0 r_{t+1}^f,$$

where $g_{t+1} = \log(C_{t+1}/C_t)$, $r_{t+1}^f = \log(R_{t+1}^f)$ and $\lambda_0 = \psi_0 \log(\beta_0)$.

The only drawback of using $\tau$ instead of $\tau_0$ in equation (1.27) is the usual problem with standard errors of the second step. As is well-known, they have to be corrected because of the noise produced in the first-step estimation. However, in practice, this implies no additional computational cost for our two-step procedure. In standard quantile regressions, the coefficients’ asymptotic variance contains the unknown conditional distribution of the error term. Because of that it is common to compute standard errors by bootstrap. Hence, to address the two-step estimation issue, it is natural to incorporate the first step in the bootstrap procedure.\footnote{That is, from $S$ bootstrapped samples one estimates $S$ pairs $\left(\hat{\lambda}, \hat{\psi}\right)$ and computes their empirical variance matrix.}

From $\left(\hat{\lambda}, \hat{\psi}\right)$ one consistently computes $\hat{\beta} = \exp\left(\hat{\lambda}/\hat{\psi}\right)$. The standard error of $\hat{\beta}$ should be computed from the bootstrapped covariance matrix of $\left(\hat{\lambda}, \hat{\psi}\right)$ by the delta method. Accordingly,

$$\sqrt{T} \left(\hat{\beta} - \beta_0\right) \overset{d}{\to} N\left(0, \exp\left(\frac{2\lambda_0}{\psi_0} \left(\frac{1}{\psi_0^2} \sigma_\lambda^2 + \frac{\lambda_0^2}{\psi_0^2} \sigma_\psi^2 - 2 \frac{\lambda_0}{\psi_0^3} \sigma_{\lambda\psi}\right)\right)\right),$$

where $\sigma_\lambda^2$ is the asymptotic variance of $\hat{\lambda}$, $\sigma_\psi^2$ is the asymptotic variance of $\hat{\psi}$, and $\sigma_{\lambda\psi}$ is the asymptotic covariance between both estimators.
1.4.3 Empirical results

We now apply the estimation procedures discussed above to a monthly data set. Such data frequency is used to maintain the assumption that the decision interval of the agent is monthly, as in the simulation exercise. Per capita consumption is the sum of personal consumption expenditures on services (PCES, St. Louis Fed) and personal consumption expenditures on nondurable goods (PCEND, St. Louis Fed), divided by the total population (POP, St. Louis Fed). The risky return is the S&P 500 return including dividend payments, and the risk-free return is the 1-month risk-free rate series from Professor Fama located in the CRSP data base. All series are deflated by the consumer price index for all urban consumers (CPIAUCSL, St. Louis Fed). Since both consumption series start in January 1959 in the St. Louis Fed data base, the data set ranges from January 1959 to December 2009.

We define three distinct instrument vectors, $Z_t^{(1)} = \left( 1, R_{t+1}^f \right)$, $Z_t^{(2)} = \left( 1, R_{t+1}^f, R_t^f \right)$, and $Z_t^{(3)} = \left( 1, R_{t+1}^f, R_t^f, R_{t-1}^f \right)$, all three satisfying assumption (vi) in Proposition 6. We do not include lags of consumption growth and risky returns since they have very weak forecasting power over their future realizations (see Cochrane (2006), pp 268).

Columns 2, 3 and 4 of Table 1.3 present the estimates of $\theta_0$ under the general (one-step) estimation method. Standard errors are analytically computed using the asymptotic distribution derived in Proposition 8. The fifth column of Table 1.3 shows the result from the two-step procedure presented in the last sub-section. Standard errors are calculated by

\[ 27 \text{ We estimate } f_{c_t,t+1} (0 | Z_t) \text{ nonparametrically, following Powell (1986), using } \hat{e}_{c,t+1} = \frac{C_{t+1}}{C_t} - \left( \hat{\beta}_0 R_{t+1}^f \right)^\psi. \]
bootstrap according to the previous sub-section, addressing both issues of \( \tau_0 \) estimated in a previous step and of the unknown distribution in the asymptotic variance.

Table 1.4 reproduces Table 3, but allows for the presence of auto-correlation in the empirical moments. In columns 2, 3 and 4, \( W_T \) is computed by Newey and West’s (1987) estimator. In column 5, we employ overlapping block-bootstrap to compute the variance matrix of \( \left( \hat{\lambda}, \hat{\psi} \right) \).

<table>
<thead>
<tr>
<th></th>
<th>1-step procedure</th>
<th>2-step procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>block 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta )</td>
<td>1.002</td>
<td>1.002</td>
</tr>
<tr>
<td>(se)</td>
<td>(0.001)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>EIS</td>
<td>0.37</td>
<td>0.36</td>
</tr>
<tr>
<td>(se)</td>
<td>(0.07)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>( \tau )</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>(se)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>J stat.</td>
<td>5.1</td>
<td>7.2</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.02)</td>
<td>(0.07)</td>
</tr>
</tbody>
</table>

Columns 2, 3 and 4 present the 1-step estimates. \( Z^{(j)} \) contains up to the j-th lag of the risk-free rate. Column 5 presents the 2-step estimates. For all columns, no serial-correlation is assumed, justified by the fact that moments are martingale difference sequences according to proposition 5.

Table 1.3. Estimates under no serial-correlation
1.4 Model estimation

<table>
<thead>
<tr>
<th>block 1</th>
<th>$Z^{(1)}$</th>
<th>$Z^{(2)}$</th>
<th>$Z^{(3)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>1.002</td>
<td>1.002</td>
<td>1.002</td>
</tr>
<tr>
<td>(se)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>EIS</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>(se)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>$\tau$</td>
<td>0.45</td>
<td>0.45</td>
<td>0.44</td>
</tr>
<tr>
<td>(se)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>J stat.</td>
<td>3.8</td>
<td>4.9</td>
<td>9.7</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.05)</td>
<td>(0.18)</td>
<td>(0.08)</td>
</tr>
</tbody>
</table>

Columns 2, 3 and 4 present the 1-step estimates. $Z^{(j)}$ contains up to the j-th lag of the risk-free rate. Column 5 presents the 2-step estimates. Serial-correlation is allowed for all columns and asymptotic variance is estimated by Newey-West with 6 lags (columns 2, 3 and 4) and by overlapping block-bootstrap with 6 lags (column 5).

Table 1.4. Estimates under serial-correlation

The estimates from Tables 1.3 and 1.4 are very similar. This should be a consequence of the very low empirical serial correlation of consumption growth and returns. The estimates across columns in both tables are also very similar, which is evidence of the robustness of the estimation methods. In particular, the results from the one-step and the two-step procedures are very close to each other. This was expected since both procedures are consistent.

Although the time discount factor estimates are slightly above one, it is in general not possible to reject the hypothesis $\beta_0 < 1$. The estimates of the elasticity of intertemporal substitution go from 0.35 to 0.39 and are all significantly different from zero. The downside risk aversion is estimated ranging from 0.43 to 0.45 and are all significantly different from 0.5.
The EIS estimation under an alternative framework is a contribution of the present paper. As discussed in Guvenen (2006), most of the estimated Euler equations deliver extremely low values for such a parameter, often not significantly different from zero. However, macroeconomists calibrate their models using positive values for the EIS, generally between 0.5 and 1. Hence, the present results diminish this contradiction between the dynamic macroeconomics literature and the Euler-equations-based estimates for the EIS.

With respect to the model specification, the overidentifying restrictions test rejects the model at 5% only in the first column of Table 1.3. This is a remarkable result given the usual rejection of asset prices models by the J-test.

Since these results from estimation are qualitatively the same as those obtained under simulation (the time discount factor used in the simulation exercise was 0.9998, the EIS was 0.6 and the downside risk aversion was 0.45), we conclude that such values are robust.

1.5 Conclusion

We considered a framework where the representative agent makes his decision about consumption-investment looking at worst-case scenarios, which depend on his degree of pessimism. We used a well-known quote among professional investors to motivate this agent: "Focus on the downside, and the upside will take care of itself".

Under the quantile utility maximizer agent of Manski (1988) and Rostek (2010), we attached the agent’s degree of pessimism to a well defined parameter. As a consequence, we disentangled attitude towards risk and attitude towards intertemporal substitution in a novel way.
Two important results emerged. First, with only 3 preference-related parameters, the model was able to reproduce the historical averages and volatilities of the excess return, risk-free rate and consumption growth, the low covariance between stock return and consumption growth, the countercyclicality of the risk premium, and the procyclicality of the risk-free rate. Second, it was possible to estimate the EIS from an Euler equation in which such a parameter was separably identified. Related to the second result, a novel and simple two-step estimation procedure for the EIS was proposed.

The developed model was restricted to a single risky asset and a risk-free security. This was enough to address the proposed questions. Because of the nonlinearity of the quantile operator, the derived Euler equations cannot be directly generalized to more than one risky asset. Hence, a quantile asset pricing model with multiple risky assets, to study the cross-section of returns, is an interesting topic for future research.

A pure quantile maximizer agent is probably not a good representation for general behavior towards risk. Given that, the present model should be understood as a stylized and parsimonious study within the class of models that use asymmetric preferences over good and bad outcomes (as in prospect theory and disappointment aversion). As such, it makes an important contribution to the literature. Given its ability to explain the financial puzzles parsimoniously, it (i) offers a simpler view regarding the relationship between asymmetric preferences and financial data, and (ii) provides evidence that the good empirical results obtained by the studies employing asymmetric preferences are not due to over-fitting.
Chapter 2
International Data and the Quantile Utility Asset Pricing Model

2.1 Introduction

The asset pricing model proposed in chapter one considers an investor who only cares about downside risk. As we saw, the model can address the financial puzzles related to the first and second moments of risky and risk-free returns and consumption growth for the US.

According to Campbell (1999, 2003), the US stylized facts described in the previous chapter also apply to international data. For example, excess returns are high and covariances between risky returns and consumption are low (even negative). As a consequence, the canonical expected utility model also faces serious difficulties when confronted with such data, that is, the financial puzzles are a robust phenomenon in international data.

Given that, an important step in the empirical validation of the quantile utility model is to check whether it can explain the relation between returns and consumption in other developed countries. As we present in this chapter, this seems to be the case.

The fact of having data available for more than one country allows us to perform an interesting exercise. A major difference between the quantile model and the canonical expected utility approach is the idea of risk. Under expected utility, the agent is worried about smoothing consumption across states of the world. As a consequence, risk is given by the covariance between return and consumption. The quantile utility agent, however,
cares about how bad he is going to be when a bad state occurs. In this case, risk can be shown to be about the standard deviation of the investment return. Accordingly, we can examine which definition of risk can better explain the differences in excess returns across countries, considering that risk aversion levels do not vary much from country to country.

This is the first empirical analysis of this chapter. The conclusion is clear. On the one hand, since the covariance between risky return and consumption growth is negatively related to the mean excess return, the expected utility model cannot explain why different countries pay different mean excess returns without imposing a very large variation in the risk aversion levels across countries. On the other hand, the standard deviation of return has no clear relation to the mean excess return. Hence, the differences in risk premia may be easier justified by smaller differences in the levels of risk aversion.

We then compute the parameters of the quantile utility model for each country. We do that by first imposing joint lognormality of returns and consumption growth. Following, we estimate the model free of any distributional assumption. In both exercises we obtain reasonable values for the preference-related parameters, namely, the time discount factor, the elasticity of intertemporal substitution (EIS) and the downside risk aversion. The discount factor is in general below 1, the EIS between 0 and 1, and the downside risk aversion between 0.30 and 0.45. These results lead us to conclude that the success of the model in explaining the financial puzzles in the US also holds internationally.

The rest of the chapter is organized as follows. Section 2.2 presents the international data set and discusses the risk-return trade-off across countries, section 2.3 performs the main empirical analysis and section 2.4 concludes.
2.2 International data

Campbell (1999, 2003) reviews the behavior of stock prices in relation to consumption using international data. His goal is to see which features of the US experience apply more generally.

To construct an international quarterly data set, he uses Morgan Stanley Capital International (MSCI) stock market data covering the period since 1970. He combines the MSCI data with macroeconomic data on consumption, interest rates, and the price level from the International Financial Statistics (IFS) of the International Monetary Fund.


The main data set used here is the same one used in Campbell (2003). This has two reasons. The first is that we can directly compare the present results with Campbell’s (2003). The second is the implementation of the EuroZone (EZ) in 1999.

The euro was designed to help build a single market by easing travel of citizens and goods, eliminating exchange rate problems, providing price transparency, creating a single financial market, and providing a currency used internationally and protected against shocks by the large amount of internal trade within the EZ. Hence, since 1999, the investment in stock markets and treasury bonds was greatly facilitated across EZ members.
Because of that, imposing that investors of the Netherlands, for instance, can only invest in their home stock market and treasury bonds is not a reasonable assumption from 1999 on.

Nevertheless, we construct an updated version of Campbell (2003) data set for the countries outside the EZ (Australia, Canada, Japan, Sweden, Switzerland and the UK), with series up to 2009, and present the estimates using these data as well.

As Campbell (2003) highlights, different national stock markets are of very different sizes, both absolutely and in proportion to national GDP’s (capitalization ratio). In the UK and Switzerland, for example, the capitalization ratio was about 80% during the 90’s, whereas in Germany and Italy it was less than 20%. This is a relevant issue, because of the theoretical convention of treating the stock market as a claim to total consumption, or as a proxy for the aggregate wealth of an economy. Moreover, according to La Porta et al (1997), stock ownership also tends to be much more concentrated in the countries with low capitalization.

Because of that, empirical tests of asset pricing theories using aggregate data make much more sense in highly capitalized countries. With that in mind, in what follows, we draw our main conclusions from such countries’ results.

Table 2.5 reports summary statistics concerning returns and consumption for each country. The statistics are the mean and the standard deviation of the (annualized and real) risky and risk-free log returns and consumption log growth, and the covariance between risky log returns and consumption log growth. Countries are ordered according to their capitalization ratio.
Table 2.5. International descriptive statistics

<table>
<thead>
<tr>
<th>country</th>
<th>capitalization *</th>
<th>E(r)</th>
<th>σ(r)</th>
<th>E(rf)</th>
<th>σ(rf)</th>
<th>E(c)</th>
<th>σ(c)</th>
<th>cov(r,c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switzerland</td>
<td>88%</td>
<td>13.7%</td>
<td>21.8%</td>
<td>1.4%</td>
<td>1.5%</td>
<td>0.5%</td>
<td>2.1%</td>
<td>-0.12%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>80%</td>
<td>8.2%</td>
<td>21.2%</td>
<td>1.3%</td>
<td>3.0%</td>
<td>2.2%</td>
<td>2.5%</td>
<td>0.12%</td>
</tr>
<tr>
<td>Netherlands</td>
<td>46%</td>
<td>14.1%</td>
<td>17.2%</td>
<td>3.4%</td>
<td>1.6%</td>
<td>1.8%</td>
<td>2.6%</td>
<td>0.02%</td>
</tr>
<tr>
<td>Australia</td>
<td>42%</td>
<td>3.5%</td>
<td>22.7%</td>
<td>2.0%</td>
<td>2.5%</td>
<td>2.1%</td>
<td>2.1%</td>
<td>0.16%</td>
</tr>
<tr>
<td>Japan</td>
<td>40%</td>
<td>4.7%</td>
<td>21.9%</td>
<td>1.4%</td>
<td>2.3%</td>
<td>3.2%</td>
<td>2.6%</td>
<td>0.11%</td>
</tr>
<tr>
<td>Sweden</td>
<td>36%</td>
<td>10.6%</td>
<td>23.8%</td>
<td>2.0%</td>
<td>2.8%</td>
<td>1.0%</td>
<td>1.9%</td>
<td>0.03%</td>
</tr>
<tr>
<td>Canada</td>
<td>31%</td>
<td>5.4%</td>
<td>17.3%</td>
<td>2.7%</td>
<td>1.9%</td>
<td>2.1%</td>
<td>2.0%</td>
<td>0.19%</td>
</tr>
<tr>
<td>France</td>
<td>22%</td>
<td>9.0%</td>
<td>23.4%</td>
<td>2.7%</td>
<td>1.8%</td>
<td>1.2%</td>
<td>2.9%</td>
<td>-0.10%</td>
</tr>
<tr>
<td>Germany</td>
<td>17%</td>
<td>9.8%</td>
<td>20.1%</td>
<td>3.2%</td>
<td>1.2%</td>
<td>1.7%</td>
<td>2.4%</td>
<td>0.03%</td>
</tr>
<tr>
<td>Italy</td>
<td>9%</td>
<td>3.2%</td>
<td>27.0%</td>
<td>2.4%</td>
<td>2.9%</td>
<td>2.2%</td>
<td>1.7%</td>
<td>-0.03%</td>
</tr>
</tbody>
</table>

* size of stock market (market value) over GDP in 1993

data up to 1998 (Campbell 2003)

For all countries in Table 2.5, the mean risky return exceeds the mean risk-free rate as expected. However, the difference between them, that is, the mean excess return, is quite variable. It goes from 0.8% (Italy) to 13.7% (Switzerland). Can this difference in mean excess returns be justified by the difference in the risk levels across countries? For example, is Switzerland much riskier than Italy or Australia?

The risk-return trade-off across countries

The equilibrium expected excess return is in general given by the product between the size of risk and the risk price (risk aversion level). This is the case for the canonical expected utility model and the quantile utility model, as we see in equation (1.8) and (1.10) in chapter one. Hence, if the risk aversion levels don’t vary much across countries (or vary in a way unrelated to the risk levels), we should see a positive relation between risk and excess returns in a cross-section of countries.
Under the canonical expected utility model, risk is about the covariance between risky return and consumption growth. If we plot these covariances and the excess returns computed for each country, we have a rather frustrating evidence for the expected utility framework. As Figure 2.4 shows, there is a negative relation between risk under expected utility and mean excess return.

![Graph showing mean excess return vs. risk under expected utility](image-url)

**Fig. 2.4.** Mean excess return vs. risk under expected utility.

Although it is true that the countries’ risk aversion levels may be different, this would hardly justify the relation we see in Figure 2.4. Countries with higher risk would have to have extremely lower risk aversion to compensate for the lower mean excess return. Moreover, the fact that some countries (Switzerland and France) have negative covariance and high mean excess returns would remain puzzling.
If countries with low capitalization ratio (below 30%) are excluded from Figure 2.4, things get even worse. The negative relation between risk and mean excess return becomes stronger (the t-statistic of the regression line’s slope is $-6.7$). Figure 2.5 presents this plot.

Fig. 2.5. Mean excess return X risk under expected utility (countries with capitalization ratio greater than 30%).

Using the information in Table 2.5, we can submit the quantile asset pricing model to the same basic test. We construct the same plots but using the standard deviation of the risky return as the risk factor. Figures 2.6 shows that there is no strong relation between mean excess return and risk across countries.
2.2 International data

Because of Italy’s location in the plot, the relation is still slightly negative. However, Italy is the country with the lowest capitalization ratio (9%) and, as discussed above, may be not a good empirical counterpart of the theoretical model. Figure 2.7 considers only the countries with capitalization ratio above 30%.

Fig. 2.6. Mean excess return X risk under quantile utility.
2.3 Fitting the model to each country

We first calibrate equations (1.7) and (1.8) from chapter one with the values from Table 2.5, and solve for the preference related parameters $\psi$ (the EIS, the inverse of $\gamma$) and $\tau$. We...
first solve for $\tau$ in equation (1.8) and then, fixing $\beta = 0.999$, solve for $\psi$ in equation (1.7).

Table 2.6 presents the results.

<table>
<thead>
<tr>
<th>country</th>
<th>downside risk aversion ($\tau$)</th>
<th>EIS ($\psi$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switzerland</td>
<td>29%</td>
<td>-0.5</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>37%</td>
<td>1.2</td>
</tr>
<tr>
<td>Netherlands</td>
<td>27%</td>
<td>0.1</td>
</tr>
<tr>
<td>Australia</td>
<td>47%</td>
<td>1.0</td>
</tr>
<tr>
<td>Japan</td>
<td>44%</td>
<td>2.2</td>
</tr>
<tr>
<td>Sweden</td>
<td>36%</td>
<td>0.2</td>
</tr>
<tr>
<td>Canada</td>
<td>44%</td>
<td>0.7</td>
</tr>
<tr>
<td>France</td>
<td>39%</td>
<td>0.2</td>
</tr>
<tr>
<td>Germany</td>
<td>37%</td>
<td>0.3</td>
</tr>
<tr>
<td>Italy</td>
<td>49%</td>
<td>0.9</td>
</tr>
</tbody>
</table>

The EIS values were obtained by fixing the discount factor ($\beta$) at 0.999

Table 2.6. International parameters under lognormality

The estimates for the downside risk aversion range from 27% (Netherlands) to 49% (Italy). If we consider only the countries with capitalization ratio above 30%, the maximum $\tau$ is 47%.

Such values of $\tau$ belong to the acceptable interval 22% – 48% discussed in chapter one. This implies that in order to explain the excess returns across countries, one would need a quantile asset pricing model with reasonable levels of downside risk aversion.

This is in sharp contrast with the results from the canonical expected utility model. Table 4 in Campbell (2003) provides the values for each country’s risk aversion computed from the closed-form equation for the risk premium under expected utility. For Switzerland, France and Italy (the countries with negative covariance between risky return consumption growth) risk aversion levels are negative and, for all other countries, unreason-
ably high. Moreover, as suggested from the discussion in the previous section, they present a extremely large variation across countries. Table 2.7 reproduces those values.

<table>
<thead>
<tr>
<th>country</th>
<th>risk aversion (under expected utility)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switzerland</td>
<td>negative</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>186</td>
</tr>
<tr>
<td>Netherlands</td>
<td>850</td>
</tr>
<tr>
<td>Australia</td>
<td>58</td>
</tr>
<tr>
<td>Japan</td>
<td>83</td>
</tr>
<tr>
<td>Sweden</td>
<td>1713</td>
</tr>
<tr>
<td>Canada</td>
<td>59</td>
</tr>
<tr>
<td>France</td>
<td>negative</td>
</tr>
<tr>
<td>Germany</td>
<td>599</td>
</tr>
<tr>
<td>Italy</td>
<td>negative</td>
</tr>
</tbody>
</table>

from Table 4 in Campbell (2003)

Table 2.7. International risk aversion under expected utility

We now turn to the EIS. Although the EIS is a key parameter that plays a crucial role in policy and welfare assessments, the empirical literature involving it is far from a mature stage. On the one hand, the estimates that come from aggregate data and Epstein and Zin (1989) preferences (Hall 1988, Campbell and Mankiw (1989) and Campbell 2003, for example) usually find very small, most of the times not even significant, values, for both the US and other developed countries. On the other hand, macroeconomists usually calibrate their models using positive values for the EIS, generally between 0.5 and 1. Positive values are also found by some studies using microdata. For instance, Engelhardt and Humar (2009) estimate its 95% confidence interval to be 0.37 — 1.21 (for the US).

---

28 However, the standard procedure that has been employed in the EIS estimation since Hall (1988) was recently found to suffer from weak instruments. See Neely, Roy, and Whiteman (2001) and Campbell (2003).
The EIS values presented in Table 2.6 are in general consistent with those numbers, except for Switzerland (−0.5) and Japan (2.2). However, since we were not able to separably identify the two time-preference parameters, \( \beta \) and \( \psi \), it can be the case these results are not precise. As mentioned above, we first fixed \( \beta \) and then solved equation (1.7) for \( \psi \). In other words, we could only identify \( \psi \) as a function of \( \beta \), and the EIS values in Table 2.6 will not be accurate in case \( \beta \) is different from 0.999.

In next section, throughout the full estimation of the model, we overcome this issue and present the most reliable results for this important parameter.

### 2.3.1 Model estimation

We now estimate the model for each country using both the one- and the two-step procedures proposed in chapter one.\(^{29}\) As chapter one shows, under mild regularity condition, the one-step estimator globally identifies \((\beta, \psi, \tau)\), is consistent, and has a standard limiting distribution. Moreover, it does not depend on any distributional assumption. The two-step estimator, although not efficient, is a simple robustness check for the estimates, since it is also consistent. Table 2.8 reproduces the obtained results.

\(^{29}\) In the one-step procedure we use one lag of the risk-free rate as the instrument. As in chapter one, adding more lags changes the estimates very slightly.
We base our discussion on the one-step estimator results, given its efficiency and the fact that it allows us to perform specification tests. Note however that, as in chapter one, the two-step procedure produces very close estimates for the parameters.

With respect to the discount factor $\beta$, its 95% confidence interval contains values below 1 for all countries. In other words, it is not possible to reject at 95% of confidence that, in each country, the representative agent prefers earlier utility. More than that, for Switzerland, the United Kingdom and the Netherlands (the most capitalized countries), the estimate of $\beta$ is below 1.

The EIS estimates are reasonable values as well. Among the countries with capitalization ration above 30%, only Australia and Canada present insignificant estimates for...
the parameter. This would imply that, in these countries, present and future consumption are perfect complements. However, in Switzerland, the UK, the Netherlands, Japan and Sweden, the representative agent is willing to substitute consumption across time, as the positive EIS indicates.

Regarding the downside risk aversion of countries with capitalization ration above 30%, the estimates range from 0.31 (Netherlands) to 0.46 (Australia). These are reasonable values in accordance to chapter one. We can compare the estimates for $\tau$ with the values obtained for it in the previous sub-section. The differences are not higher than 3%, except for Switzerland, 7%, and Sweden, 6% (for these two countries the lognormality assumption is probably more binding).

The overidentifying restrictions tests do not reject the hypothesis that the model is correct at 5% for all countries other than Italy, the one with the lowest capitalization ratio.

As expected, the model works much better for the countries with not too low capitalization ratios (above 30%). Based on the results, it is fair to say that the theoretical model is empirically successful for these countries. As we see now, the conclusion is the same for the updated data set.

### 2.3.2 Updated data set

We construct an updated data set for the countries that do not belong to the Euro Zone. These are Australia, Canada, Japan, Sweden, Switzerland, and UK. The updated data set is also quarterly and ranges from 1970 to 2009. As a general rule, we use the same sources
used by Campbell (2003). All stock market and macroeconomic series are denominated in local currency units.

The source for stock market data is Morgan Stanley Capital International (MSCI). Quarterly returns are based on the quarterly MSCI National Gross Returns Indices, calculated with dividends reinvestment before withholding taxes have been paid.

The main source for the macroeconomic data is the International Monetary Fund’s International Financial Statistics (IFS). Interest rates quarterly series were constructed from monthly series, considering the last month of the quarter. We used treasury bill rates for Canada, Sweden, Switzerland and UK and money market rates for Australia and Japan. Regarding consumption, for each country a quarterly series on seasonally adjusted aggregate private consumption was obtained from IFS, and transformed into per capita consumption. Population, also from IFS, was only available in the annual frequency. Hence, we constructed quarterly population series by assuming that population grows at constant rates within the year. Returns and consumption growth were deflated using quarterly CPI for each country series also from IFS.

Table 2.9 presents the updated descriptive statistics.
2.3 Fitting the model to each country

<table>
<thead>
<tr>
<th>country</th>
<th>$E(r)$</th>
<th>$\sigma(r)$</th>
<th>$E(r_f)$</th>
<th>$\sigma(r_f)$</th>
<th>$E(c)$</th>
<th>$\sigma(c)$</th>
<th>$cov(r,c)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switzerland</td>
<td>7.7%</td>
<td>23.2%</td>
<td>1.2%</td>
<td>1.4%</td>
<td>0.7%</td>
<td>1.1%</td>
<td>0.02%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>5.3%</td>
<td>25.1%</td>
<td>1.7%</td>
<td>3.9%</td>
<td>2.1%</td>
<td>2.6%</td>
<td>-0.03%</td>
</tr>
<tr>
<td>Australia</td>
<td>4.9%</td>
<td>22.8%</td>
<td>2.7%</td>
<td>4.2%</td>
<td>1.8%</td>
<td>1.6%</td>
<td>-0.06%</td>
</tr>
<tr>
<td>Japan</td>
<td>3.5%</td>
<td>26.2%</td>
<td>1.2%</td>
<td>3.1%</td>
<td>1.7%</td>
<td>2.9%</td>
<td>0.30%</td>
</tr>
<tr>
<td>Sweden</td>
<td>8.7%</td>
<td>27.7%</td>
<td>2.1%</td>
<td>3.6%</td>
<td>2.3%</td>
<td>3.0%</td>
<td>0.30%</td>
</tr>
<tr>
<td>Canada</td>
<td>5.4%</td>
<td>17.8%</td>
<td>2.4%</td>
<td>2.9%</td>
<td>1.8%</td>
<td>2.2%</td>
<td>0.12%</td>
</tr>
</tbody>
</table>

Table 2.9. Updated international descriptive statistics

If we update the plots from section 2.2 using the information of Table 2.9, both expected and quantile utility models benefit. The strange negative relation between risk and return under expected utility disappears with the expanded sample, as Figure 2.8 shows below.

Fig. 2.8. Mean excess return X risk under expected utility (updated sample)
This is good for the model, since it can be the case that the differences in mean excess returns are now explained by smaller differences in the risk aversion levels. However, the expanded sample is also generous with the quantile utility model. The risk-return relation is now positive, according to Figure 2.9.

![Graph showing mean excess return vs risk under quantile utility](image)

**Fig. 2.9.** Mean excess return X risk under quantile utility (updated sample)

Finally, Table 2.10 presents the updated estimates for quantile utility model.
2.3 Fitting the model to each country

<table>
<thead>
<tr>
<th>Country</th>
<th>market value /GDP</th>
<th>β</th>
<th>ψ</th>
<th>τ</th>
<th>one-step</th>
<th>J-test stat.</th>
<th>β</th>
<th>ψ</th>
<th>τ</th>
<th>two-step</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switzerland</td>
<td>87%</td>
<td>0.998</td>
<td>0.54</td>
<td>0.38</td>
<td>0.3</td>
<td>0.998</td>
<td>0.53</td>
<td>0.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.12)</td>
<td>(0.04)</td>
<td></td>
<td>(0.0001)</td>
<td>(0.10)</td>
<td>(0.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>United Kingdom</td>
<td>79%</td>
<td>1.001</td>
<td>0.52</td>
<td>0.41</td>
<td>0.6</td>
<td>1.001</td>
<td>0.46</td>
<td>0.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.003)</td>
<td>(0.09)</td>
<td>(0.04)</td>
<td></td>
<td>(0.0002)</td>
<td>(0.09)</td>
<td>(0.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>42%</td>
<td>1.068</td>
<td>0.06</td>
<td>0.45</td>
<td>0.2</td>
<td>1.054</td>
<td>0.07</td>
<td>0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.098)</td>
<td>(0.07)</td>
<td>(0.04)</td>
<td></td>
<td>(0.0060)</td>
<td>(0.07)</td>
<td>(0.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>40%</td>
<td>1.001</td>
<td>0.56</td>
<td>0.39</td>
<td>0.1</td>
<td>1.002</td>
<td>0.59</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.003)</td>
<td>(0.13)</td>
<td>(0.04)</td>
<td></td>
<td>(0.0003)</td>
<td>(0.22)</td>
<td>(0.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>36%</td>
<td>1.004</td>
<td>0.16</td>
<td>0.43</td>
<td>0.1</td>
<td>1.005</td>
<td>0.22</td>
<td>0.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.029)</td>
<td>(0.24)</td>
<td>(0.04)</td>
<td></td>
<td>(0.0020)</td>
<td>(0.35)</td>
<td>(0.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>31%</td>
<td>1.115</td>
<td>0.03</td>
<td>0.42</td>
<td>1.9</td>
<td>1.144</td>
<td>0.01</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.44)</td>
<td>(0.10)</td>
<td>(0.40)</td>
<td></td>
<td>(0.058)</td>
<td>(0.08)</td>
<td>(0.04)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

J-test critical values: 6.63 (1%) and 3.84 (5%)

heteroskedasticity and serial correlation robust standard errors in parenthesis

Table 2.10. International estimated parameters - updated sample

We can compare the results from Tables 2.8 and 2.10. For Switzerland, the only significant change was an increase in the EIS estimate from 0.35 to 0.54. For the UK, the time discount factor is now below 1, the EIS went from 0.41 to 0.52 and the downside risk aversion was slightly reduced with \( \tau \) going from 0.36 to 0.41. Regarding Australia and Japan, changes were immaterial. For Sweden, major changes happened to the time preference. The time discount factor decreased from 1.004 to 0.974 and the EIS, which was 0.16 but not statistically significant, is now significant and large (0.94). In Canada, the change was due to a rise in the time discount factor, as well as in its standard error. Given the large standard error of 0.44, the estimation of \( \beta \) for this country seems now to be shaky. However, the issue of a small capitalization ratio should be playing a role one more time.
2.4 Conclusion

In this chapter, we provided evidence that the quantile utility asset pricing model can also explain the financial puzzles in other developed countries. This was an important step in the empirical validation of the model.

The estimated parameters were robust across highly capitalized countries and qualitatively close to the ones obtained for the US. Moreover, we showed that the risk measure under the quantile utility model can better justify the differences in risk premia across countries when compared to the risk measure under the expected utility model.
Chapter 3
An Empirical Evaluation of the Effects of Margins on Asset Prices

(co-authored with Guilherme B. Martins)

3.1 Introduction

Assuming that agents are explicitly worried about the downside is not the only way to relate downside risk to asset prices.

A number of recent theoretical works have shown that margin requirements impact asset prices. Some examples are Brunnermeier and Pedersen (2008), Garleanu and Pedersen (2009), Gromb and Vayanos (2010) and Geanakoplos (2010). The common key feature of these models is to consider leveraged positions which depend on margin requirements. As a consequence, if in some periods a significant fraction of agents are credit constrained, that is, it is harder to buy assets on margin, an additional premium may be required to hold them.

The relation between margin requirements and downside risk is simple and direct. When an investor buys stocks on margin, some money is put up by him (initial margin), and the remainder is borrowed from the broker, with the purchased shares used as collateral. How does the broker define the maximum lending amount? According to the collateral
evaluated at a worst-case scenario. Therefore, the worse the worst-case scenario, the higher the initial margin requirement.

As an example, suppose an investor wants to buy 100 shares at $10 each using a debt-financed purchase and the broker evaluates the 1% conditional quantile for such stock price at $7. In this case she accepts to lend $700 to the investor, keeping the 100 shares as collateral since, with 99% of chance, she will not lose any money. In other words, the initial margin is established at 30% and the investor has to self-finance 30% of the initial position. Were the 1% conditional quantile equal to $6, the margin would be set at 40%.

The analysis of the relevance of margins is important not only to understand asset prices. Some authors, such as Geanakoplos (2010) and Ashcraft, Garleanu, and Pedersen (2010), have been using these models to interpret the unconventional policies implemented by the Fed during the 2007-2010 financial crisis.

In the past three years, the size and composition of the Fed’s balance sheet has suffered major changes. In January 2007, the Fed carried no risk of default in its assets, holding basically US Treasury bills ($780 billion). During the crisis, however, a variety of asset were included in the balance sheet in significant amounts. For example, commercial papers ($350 billion), repurchase agreements ($150 billion), mortgage-backed securities ($1 trillion), Federal agency debt securities ($150 billion) and others ($100 billion). In December 2010, the total size of the balance sheet was almost $2.5 trillion.

As Geanakoplos (2010) points out, the negative effect of margins on prices, together with the fact that these elements feed back one each other, could justify such a radical change in the Fed’s policy. According to him, during some periods, "the Fed must step
around the banks and lend directly to investors, at more generous collateral levels than the private markets are willing to provide."

Moreover the extra margin premium may break the non arbitrage link between the Fed fund rate and the rate of returns of other assets in the economy. This would imply that restricting monetary policy to its traditional instruments would not always be efficient.

This idea is illustrated by Ashcraft, Garleanu, and Pedersen’s (2010) argument. For example, if a premium related to margin really exists, a reduction of the Fed fund rate during crises will not necessarily translates into a expansionary policy. As we shall see below, the margin premium is the product between the margin requirement, the cost of margin, and the importance of the leveraged agents in aggregate consumption. The cost of margin is equal to the shadow cost of capital for leveraged agents, which is the spread between the uncollateralized and the collateralized (Fed fund) risk-free rates. Given that, when the Fed fund rate is reduced in a crisis during which margin constraints bind, the consequent increasing in the shadow cost of capital steepens the margin-return relation and, thus, increases the required return on assets with high margin requirements. Since in bad periods margins are significantly higher across assets, the interest rate reduction can then have small, zero or, in the limit, even a positive effect on the aggregate required return in the economy.

In despite of the relevance of these theoretical results, there is still no empirical work supporting them, apart from some isolated examples for some individual assets that the theoretical papers mentioned above provide. This chapter contributes to fill this gap, by
looking for evidences on the existence of an aggregate premium for margins and by evaluating its importance.

Our empirical findings, which are related to both the time-series and cross-section of returns, are favorable to these models. In particular, we show that (i) a margin-related factor is able to predict the future excess returns of the usual proxy for the market portfolio (S&P 500), and (ii) portfolios with high betas on the margin factor pay on average higher returns in relation to those with low margin factor betas.

We construct the factor empirical counterpart using the margin requirements on the S&P 500 futures, the ted spread, and the consumption series for stockholders and non stockholders of Malloy, Moskowitz and Vissing-Jørgensen (2009). With that, we run predictability regressions and estimate a time series for the past aggregate premium related to margins, showing that it is economically meaningful.

Around the Black Monday (19 October, 1987), margins may have been responsible for as much as one third of the total fall in prices. Moreover, according to our results, a small increase in the ted spread from 0.7% (its sample average) to 1.7%, with margin requirements at their average level (4.2%), may depress the market portfolio value by 0.8% within a 1 month period. In a worse condition of capital constraints, considering the ted spread going from 0.7% to 3% and margins requirements fixed at 10%, the price fall would be about 4% during the month.

Motivated by the fact that data on margins are usually very hard to obtain, we also propose a nonparametric model for explaining margins from current and past values of the value at risk of the specific asset. This approach, as we argue, makes theoretical sense and
has good empirical performance. Indeed, our estimated model well predicts margin requirements even out of the sample. Moreover, reestimating our original forecasting regressions using the fitted values from the VaR-based model instead of the true margin requirements, we obtain qualitatively the same results. According to this evidence, one could potentially estimate the margin premium for any asset, even with no data on margins.

Using this idea, we then estimate the aggregate margin premium for periods where data on the margin on the S&P 500 futures are not available (before 1982). An interesting conclusion emerges from this exercise since, before 1982 (and after 2004), the margin factor seems to follow a different regime. During our main sample (1982 to 2004), once the margin factor spikes it returns fast to its original levels. Because of that, its predictive power on future returns appears for short horizons regressions (and disappears for longer horizons where other traditional factors play a more important role). However, for the years before 1982 and after 2004, the margin factor takes more time to revert after a spike. This brings the predictive power to longer horizons regressions.

This chapter is organized as follows. In Section 3.2 we provide a simple theoretical model that motivates our empirical work. Section 3.3 presents our data as well as our main empirical results. Section 3.4 discusses the model estimation under no data on consumption, proposes a model for explaining margins and extend our empirical results to turbulent periods.
3.2 The theoretical model

In this section we present a theoretical model, based on Garleanu and Pedersen (2009), which motivates our empirical work. We summarize their model and refer to their paper for a more detailed analysis, including how to solve for the general equilibrium.

The economy has two types of agents \( n \in \{a, b\} \). Agent \( a \) is the risk-averse type and \( b \) is the brave one, with a smaller risk aversion, equal to one. Both agents have CRRA preferences and maximize

\[
E_t \left( e^{-\rho(s-t)} \frac{C^{1-\gamma}}{1 - \gamma^n} \right) ds. \tag{3.28}
\]

There are several risky assets in the economy. The price of risky asset \( i \) follows a Geometric Brownian Motion process

\[
dP^i_t = \mu^i_t P^i_t dt + P^i_t \sigma^i_t dw_t. \tag{3.29}
\]

In addition to the risky assets there are two riskless money market assets, both in zero net supply. One represents borrowing and lending against collateral at the interest rate \( r^c_t \) and the other uncollateralized loans with interest rate \( r^u_t \).

The first type is available to all agents in the economy. For example, when one investor takes a long position in a risky asset she can borrow in the collateralized loan market. To do so she must make some collateral available to her broker. The amount of required collateral is determined by the haircut applied by the broker. The haircut is the margin requirement, denoted by \( m^c_t \), and determines how much of her own capital she must use to make the initial investment. Similarly, if she takes a short position, she must also deposit collateral as margin with her broker or at some exchange. In both cases, the margin
is computed as a fraction of the total position: if the agent invests a fraction $\theta_i^t$ of their wealth $W_i$ in the risk asset $i$, she must deposit $m_i^t|\theta_i^t|W_i$ as margin. Note again that she must deposit a positive margin whether she is long or short in the asset. Finally, the margin deposits are remunerated at $r^c_t$.

The uncollateralized loan market is a standard one. It is riskless as the collateralized loan. However, only type $b$ agents can contract uncollateralized loans and therefore, as we show below, when this agent is capital constrained, the two interest rates are different.

Every instant, each consumer can choose how much to consume ($C_t$), the fraction of her wealth she wants to invest in the risky assets, and in the uncollateralized loan market ($\eta^u_t$). Any residual wealth is invested in the collateralized loan market. The evolution of wealth is then given by

$$dW_t = \left[ W_t \left( r^c_t + \eta^u_t (r^u_t - r^c_t) + \sum_i \theta_i^t (\mu_i^t - r^c_i) \right) - C_t \right] dt + W_t \sum_i \theta_i^t \sigma_i^t dw_t. \quad (3.30)$$

Consumers take as given all prices and maximize (3.28) subject to (3.30) and, because of the margins requirement,

$$\sum_i m_i^t|\theta_i^t| + \eta^u_t \leq 1. \quad (3.31)$$

The Hamilton-Jacobi-Bellman equation for the type $b$ consumer is given by

$$0 = Max \left\{ e^{-\rho t} \frac{C_s^{1-\gamma}}{1-\gamma} + J_t - J_t W_t C_t \right\}$$

$$+ J_t W_t \left[ \left( r^c_t + \eta^u_t (r^u_t - r^c_t) + \sum_i \theta_i^t (\mu_i^t - r^c_i) \right) + \frac{J_W W_t}{J_t} \sum_i (\theta_i^t \sigma_i^t)^2 \right],$$

subject to (3.31).
3.2 The theoretical model

The solution to this problem yields, in the case agent $b$ is long in the risky asset $i$, two conditions:

$$ r^u_t - r^c_t = \psi_t, \quad (3.32) $$

$$ \mu^i_t - r^c_t = \gamma^t \beta^c_{b} + m^i_t \psi_t, \quad (3.33) $$

where $\psi_t$ is the shadow price of capital (i.e., the Lagrangian Multiplier associated with (3.31)), and $\beta^c_{b} \equiv \text{cov}_t \left( \frac{dC^b_t}{C^c_t}, \frac{dP^b_t}{P^c_t} \right)$.

A similar problem is solved by agents of type $a$, with the only difference that he cannot choose $\eta^a_t$. If we assume that his capital constraint is never binding, the solution to his portfolio choice problem is given by $\mu^i_t - r^c_t = \gamma^a \beta^c_{a}$. Then aggregating across consumers is straightforward and gives the main result from Garleanu and Pedersen (2009) that motivates our empirical work.

The risk premium of risky asset $i$, when only consumers of type $b$ can be capital constrained and are long in this asset, is given by a margin-based premium in addition to the standard consumption-based premium,

$$ \mu^i_t - r^c_t = \gamma^t \beta^c_{b} + x_t m^i_t \psi_t \quad (3.34) $$

where

$$ \gamma^{-1} = \frac{1}{\gamma^a} \frac{C^a_t}{C_t} + \frac{1}{\gamma^b} \frac{C^b_t}{C_t} \quad (3.35) $$

and

$$ x_t = \frac{C^b_t}{\gamma^b} \left( \frac{C^a_t}{\gamma^a} + \frac{C^b_t}{\gamma^b} \right). \quad (3.36) $$

Equation (3.34) is the main testable implication of the model. It states that the excess returns of any risky asset is composed of two terms. The first term is the standard risk
premium in the CAPM literature: the product of the price of risk, which is given by an average of the risk aversion of the different agents in the economy, and the covariance between aggregate consumption and the return of the asset. The second term is the novelty. Because some investors might be capital constrained and cannot deposit additional margins, they require an additional premium to hold such an asset in equilibrium.

This extra premium is a combination of three factors. First, $\psi_t$ measures how binding the capital constraint is. By equation (3.32), it is given by the difference of two interest rates $r^u_t - r^c_t$. The second factor $m^i_t$, is the margin requirement itself. The last term gives the importance of the constrained investor in the economy. As emphasized by Garleanu and Pedersen (2009), even though the consumption share of the type $b$ can be small, $x_t$ can still be large because it takes into account the differences in risk aversion. In the next sections, we empirically test the model and evaluate the importance of this extra premium.

### 3.3 Testable implications of the model

The theoretical model presented in the previous section has implications for both the time-series and the cross-section of expected returns. Indeed, equation (3.34) implies that (i) the margin-related factor should forecast future excess returns on a market portfolio and that (ii) stocks with higher margin requirements should, on average, earn higher returns (*ceteris paribus*). In this section we investigate both testable implications.
3.3 Testable implications of the model

3.3.1 Predicting the market portfolio return

Suppose that there is only one risky asset in the economy, the market portfolio. According to equation (3.34) we would have

\[ r_{t+h} - r^c_{t+h} = \delta_1 x_t m_t \psi_t + z_t' \delta_2 + e_{t+h}, \tag{3.37} \]

where \( r_{t+h} \) and \( r^c_{t+h} \) are the \( h \)-period ahead risky and collateralized risk-free returns respectively, \( e_{t+h} \) is an error term with zero conditional mean, \( z_t \) is a \( k \times 1 \) vector with standard risk factors other than the margin-related one, \( \delta_1 \) is the price of the margin factor and \( \delta_2 \) is a \( k \times 1 \) vector of parameters related to the prices of the other risk factors.

A main issue in the estimation of equation (3.37) is data availability. First, as Geanakoplos (2010) indicates, measures of aggregate margin are very hard to get historically. Second, data on \( x_t \), which measures the ratio of the aggregate consumption due to the brave investor (and not simply aggregate consumption), is also not readily available. Third, it may not be immediately clear which variable should well represent \( \psi_t \).

We circumvent these three issues. With respect to the first, the Chicago Mercantile Exchange has data on the margin requirements on the S&P 500 futures. This is the exact information one would need for the estimation of a margin premium for the aggregate economy under the standard assumption of proxing the market portfolio by the S&P 500 index. Since one can interchangeably trade spot and future contracts, margin requirements in future and spot markets should be tightly, if not perfectly, related. Hence, we use the CME margin requirements as \( m_t \) in our model.
Regarding $x_t$, the problem of disaggregating consumption among different groups of individuals is not new in the asset pricing literature. Since Mankiw and Zeldes (1989) a number of papers have been trying to come up with measures for the consumption of stockholders as a way to address the equity premium puzzle. Because stockholders’ consumption covariates more with returns, such studies are able to generate more reasonable risk aversion levels among other good results. Ait-Sahalia, Parker and Yogo (2004), for example, employ data on the consumption of luxury goods as a proxy for stockholder’s consumption. More recently, Malloy, Moskowitz and Vissing-Jørgensen (2009) use microlevel household consumption data to approximate this series. It is natural to use one of these series to construct a measure for $x_t$. We use the data of Malloy, Moskowitz and Vissing-Jørgensen (2009).

With respect to $\psi_t$, the shadow price of capital, equation (3.32) says that in equilibrium $\psi_t$ has to be equal to the spread between the uncollateralized and the collateralized risk-free rates. In other words, it is a measure of how binding the capital constraint is. The well-known ted spread is given by the difference between the interest rates on interbank loans (Libor) and American treasury bills and, because of that, it is a widely observed indicator of credit conditions in financial markets. Hence, it is a straightforward choice to represent $\psi_t$.

**Data construction**

Malloy, Moskowitz and Vissing-Jørgensen (2009) construct separate series for quarterly consumption growth rates for stockholders and nonstockholders, at monthly frequency,
using data from the (Consumer Expenditure Survey) CEX for the period March 1982 to November 2004. The CEX survey has responses from households where they indicate whether they hold a positive amount of "stocks, bonds, mutual funds and other such securities". About 77 percent of the households, on average for all periods, answer negatively to this question. However, besides than simply using the households with a positive response to constructing the stockholders' consumption, the authors employ a more careful approach. In order to mitigate response error, they supplement the CEX definition with a probit analysis designed to predict the probability that a household owns stocks. They then define a household as stockholder if it has both answered positively to the CEX question and its fitted probability of holding stocks is higher than half. We refer the reader to their paper for a detailed description of these series and their methodology.

We use their consumption growth rates to compute $x_t$ in accordance to equation (3.36), assuming that the brave and risk-averse agents in the theoretical model are, respectively, the stockholder and non stockholder agents from their paper. Since only growth rates are available from their original data, we are able to compute only the growth rates of $x_t$. Therefore, we can only identify $\delta_1$ up to a scale. However, this is enough to our goal. With that we can both test the existence of a margin-related factor, i.e., the significance of $\delta_1$, and estimate a time series for the size of the total margin premium, $\delta_1 x_t m_t \psi_t$.

Imposing $\gamma^B = 1$ and $\gamma^A = 10$, which are the values assumed by Garleanu and Pedersen (2009) in their analysis of the model’s predictions, we end up with the plot for $x_t$ presented in Figure 3.10.
3.3 Testable implications of the model

Fig. 3.10. The time-series for $x_t$ (level not identified)

With respect to $m_t$, we first compute the daily ratio between the margin requirements on S&P 500 futures for members of the Chicago Mercantile Exchange (available from April 1982) and the value of the underlying S&P 500 index multiplied by the size of the contract. This is the usual way of computing margins. Then we define $m_t$ as the monthly average of this series. The ted spread, $\psi_t$, is computed as the difference between the 3-month libor rate and the 3-month treasury bill. For the libor rate we use the Eurodollar 3-month deposit rate in the London market, collected by Federal Reserve Board of Governors. This series begins in January 1971 (because of the Eurodollar). Figure 3.11 plots $m_t$ and $\psi_t$. 
3.3 Testable implications of the model

Figure 3.12 plots the constructed times-series for the margin-related factor, $x_t \psi_t$, that ranges from April 1982 to November 2004. According to it, the margin-related factor spikes around periods of financial distress. Note also that when wandering around its low levels, it is rather persistent. However, once it spikes, it has in general a fast reversion. This characteristic, compatible to the idea of brief periods of binding capital constraints, will be important in the interpretation of the results below.
3.3 Testable implications of the model

Fig. 3.12. The time-series for $x_t m_t \psi_t$ (level not identified).

Even though this factor has a clear theoretical meaning given by the theoretical model and empirically confirmed by the relation between the spikes and events in the figure above, we should still control equation (3.37) for the traditional risk factors if we want to estimate $\delta_1$. Therefore we construct the $z_t$ vector with the following standard variables: dividend yield, volatility (mean of squared daily returns), lag of the return and dividend-earning ratio.

Results

Tables 3.11, 3.12, 3.13 and 3.14 present the estimates of model (3.37) for $h = 1, 2, 4$ and 12-months ahead.
### 3.3 Testable implications of the model

**Table 3.11. 1-month ahead regressions**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>x m ted</strong></td>
<td>2.55***</td>
<td>2.14***</td>
<td>2.71***</td>
<td>2.63***</td>
<td>2.55***</td>
<td>2.33***</td>
<td>1.9**</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.010)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.007)</td>
<td>(0.040)</td>
</tr>
<tr>
<td><strong>dy</strong></td>
<td>8.74</td>
<td></td>
<td></td>
<td></td>
<td>8.01</td>
<td>13.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
<td></td>
<td></td>
<td></td>
<td>(0.29)</td>
<td>(0.11)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
<td>(0.31)</td>
<td>(0.27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>lag</strong></td>
<td>0.11</td>
<td></td>
<td>-0.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.76)</td>
<td></td>
<td>(0.57)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>d-e</strong></td>
<td></td>
<td>-5.59</td>
<td>-18.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.71)</td>
<td>(0.25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>constant</strong></td>
<td>1.15</td>
<td>-6.64</td>
<td>2.71</td>
<td>0.54</td>
<td>-2.96</td>
<td>-4.52</td>
<td>-21.3</td>
</tr>
<tr>
<td></td>
<td>(0.77)</td>
<td>(0.43)</td>
<td>(0.50)</td>
<td>(0.91)</td>
<td>(0.81)</td>
<td>(0.59)</td>
<td>(0.21)</td>
</tr>
<tr>
<td><strong>R2-adj</strong></td>
<td>1.7%</td>
<td>1.9%</td>
<td>2.0%</td>
<td>1.4%</td>
<td>1.4%</td>
<td>2.1%</td>
<td>1.8%</td>
</tr>
</tbody>
</table>

*White adjusted p-values between brackets*

**Table 3.12. 2-month ahead regressions**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>x m ted</strong></td>
<td>1.94***</td>
<td>1.47**</td>
<td>1.91***</td>
<td>2.02***</td>
<td>1.9***</td>
<td>1.42**</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.021)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td>(0.031)</td>
<td>(0.170)</td>
</tr>
<tr>
<td><strong>dy</strong></td>
<td>9.91</td>
<td></td>
<td></td>
<td></td>
<td>10.14</td>
<td>15.31</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td></td>
<td></td>
<td></td>
<td>(0.10)</td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td><strong>vol</strong></td>
<td>2.1</td>
<td></td>
<td></td>
<td></td>
<td>2.88</td>
<td>2.85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.69)</td>
<td></td>
<td></td>
<td></td>
<td>(0.62)</td>
<td>(0.62)</td>
<td></td>
</tr>
<tr>
<td><strong>lag</strong></td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td>-0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.70)</td>
<td></td>
<td></td>
<td></td>
<td>(0.62)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>d-e</strong></td>
<td></td>
<td>-5.43</td>
<td>-20.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.66)</td>
<td>(0.13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>constant</strong></td>
<td>2.84</td>
<td>-5.99</td>
<td>2.51</td>
<td>2.25</td>
<td>-1.16</td>
<td>-6.66</td>
<td>-25.38</td>
</tr>
<tr>
<td></td>
<td>(0.38)</td>
<td>(0.39)</td>
<td>(0.44)</td>
<td>(0.56)</td>
<td>(0.91)</td>
<td>(0.34)</td>
<td>(0.08)</td>
</tr>
<tr>
<td><strong>R2-adj</strong></td>
<td>2.0%</td>
<td>3.1%</td>
<td>1.7%</td>
<td>1.7%</td>
<td>1.7%</td>
<td>2.8%</td>
<td>3.1%</td>
</tr>
</tbody>
</table>

*Newey-West adjusted (order equal to 1) p-values between brackets*
### 3.3 Testable implications of the model

#### Table 3.13. 4-month ahead regressions

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>x m ted</td>
<td>1.15*</td>
<td>0.63</td>
<td>1.11</td>
<td>1.17</td>
<td>1.15*</td>
<td>0.55</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.093)</td>
<td>(0.291)</td>
<td>(0.108)</td>
<td>(0.106)</td>
<td>(0.094)</td>
<td>(0.349)</td>
<td>(0.986)</td>
</tr>
<tr>
<td>dy</td>
<td>11.17**</td>
<td>(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vol</td>
<td>2.9</td>
<td>3.79</td>
<td>3.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.34)</td>
<td>(0.27)</td>
<td>(0.33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lag</td>
<td>0.02</td>
<td>-0.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.91)</td>
<td>(0.31)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d-e</td>
<td>-5.15</td>
<td>-22.9*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.68)</td>
<td>(0.09)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>4.82</td>
<td>-5.13</td>
<td>4.36</td>
<td>4.68</td>
<td>1.03</td>
<td>-6</td>
<td>-27.3*</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.43)</td>
<td>(0.15)</td>
<td>(0.12)</td>
<td>(0.91)</td>
<td>(0.35)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>R2-adj</td>
<td>1.4%</td>
<td>4.7%</td>
<td>1.2%</td>
<td>1.0%</td>
<td>1.2%</td>
<td>4.8%</td>
<td>7.0%</td>
</tr>
</tbody>
</table>

Newey-West adjusted (order equal to 3) p-values between brackets

#### Table 3.14. 12-month ahead regressions

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>x m ted</td>
<td>0.52</td>
<td>0.04</td>
<td>0.51</td>
<td>0.58</td>
<td>0.52</td>
<td>0.015</td>
<td>-0.28</td>
</tr>
<tr>
<td></td>
<td>(0.320)</td>
<td>(0.937)</td>
<td>(0.327)</td>
<td>(0.299)</td>
<td>(0.321)</td>
<td>(0.973)</td>
<td>(0.550)</td>
</tr>
<tr>
<td>dy</td>
<td>10.25</td>
<td>10.34*</td>
<td>14.04**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.09)</td>
<td>(0.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vol</td>
<td>0.28</td>
<td>1.08</td>
<td>0.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.84)</td>
<td>(0.51)</td>
<td>(0.64)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lag</td>
<td>0.08</td>
<td>-0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.70)</td>
<td>(0.37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d-e</td>
<td>0.95</td>
<td>-12.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.94)</td>
<td>(0.28)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>5.9*</td>
<td>-3.24</td>
<td>5.85*</td>
<td>5.44</td>
<td>6.6</td>
<td>-3.48</td>
<td>-15.17</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.35)</td>
<td>(0.06)</td>
<td>(0.14)</td>
<td>(0.47)</td>
<td>(0.63)</td>
<td>(0.22)</td>
</tr>
<tr>
<td>R2-adj</td>
<td>0.7%</td>
<td>10.1%</td>
<td>0.4%</td>
<td>0.7%</td>
<td>0.4%</td>
<td>9.9%</td>
<td>12.0%</td>
</tr>
</tbody>
</table>

Newey-West adjusted (order equal to 11) p-values between brackets
3.3 Testable implications of the model

In all tables above, column 1 presents the estimate of $\delta_1$ with no control, columns 2 to 5 add controls individually, column 6 adds both dividend-yield and volatility at the same time, and column 7 is the usually called "kitchen sink" regression.

The results indicate that the margin-related factor is highly significant for predicting excess returns at very short horizons, even after controlling for other risk factors. However, as $h$ increases its predictive power decreases. In fact, at 4-month ahead, the margin factor has possibly no role left in explaining expected future excess returns.

The fact that we see high forecasting power in the short-run regressions is probably due to the characteristics of the margin factor mentioned in the previous subsection. Its statistical significance should be coming from a limited number of points in the sample, the ones where it significantly spikes. As Figure 3.3 shows, they end relatively fast and, because of that, predictability shows up for the short run regressions (once it spikes prices go down but return soon). This is in sharp contrast to the dividend yield factor for instance, which tracks longer period movements and therefore do not predict short horizons returns.

It is fair to say that while the standard risk factors are more related to business cycles, the margin-based factor has more to do with financial sharp conditions.

However, if the margin-factor is significant at the 1-month horizon, one could expect it also to be significant at longer horizons. Indeed, if prices fall and rise within the next month, this variation is included within the next 3 months for example. So, why does $\hat{\delta}_1$ decrease and lose significance as $h$ increases? A simple explanation for that can be illustrated by the following stylized example.

Suppose that for a general factor $w$ we have
3.3 Testable implications of the model

\[ r_{t,t+1} = bw_t + \varepsilon_{t+1}, \ b > 0 \]  
(3.38)

\[ w_{t+1} = \begin{cases} \bar{w}, \text{w.p. } \pi \\ 0, \text{w.p. } 1 - \pi \end{cases} \]  
(3.39)

where \( r_{t,t+1} \) is the return from \( t \) to \( t+1 \), \( \varepsilon_t \) follows a white noise process and \( \pi \in [0, 1] \).

According to equation (3.38) the factor \( w \) has predictive power over future returns. Equation (3.39) tells us that at each period the factor can assume either a low (normal periods) or a high (capital constrained periods) value. For low \( \pi \), we would have a factor that spikes from time to time and, when it does, has a good chance of returning fast. According to Figure 3.12, this seems to be a fair representation of the margin factor, at least for this period in the sample.\(^{30}\)

A model like this would imply

\[ E_t(r_{t,t+1}) = bw_t \]

\[ E_t(r_{t,t+1} + r_{t+1,t+2}) = bw_t + b\pi \bar{w} \]

\[ \vdots \]

\[ E_t\left( \sum_{j=1}^{h} r_{t+j-1,t+j} \right) = bw_t + (h - 1) b\pi \bar{w} . \]

and, therefore,

\[ E_t(r_{t+h}) = 12 \frac{b}{h} w_t + 12 \frac{(h - 1) b\pi}{h} \bar{w} , \]  
(3.40)

\(^{30}\) A process for \( w_t \) with 3 levels instead of 2 would be more appropriate, but nothing would change in our argument.
where \( r_{t+h} \) is the \( h \)-month ahead return expressed per year (as in the regressions above).

Hence, as \( h \) increases, the factor coefficient in the regressions decreases and, with noisy data and other factors playing a role, eventually loses significance. This is probably what is happening in the regressions above.

In addition to being statistically significant for short horizons, price movements related to margins are economically meaningful, as Figure 3.13 illustrates.

![Fig. 3.13. The margin-related premium](image)

Figure 3.13 plots the time-series of \( \hat{\delta}_1 x_t m_t \psi_t \) for \( \hat{\delta}_1 = 2.33 \) (the value on the sixth column of Table 3.11). The difference between Figure 3.12 and Figure 3.13 is that the values on the \( y \)-axis are meaningful in the later (we are able to identify the whole premium). The \( y \)-axis in Figure 3.13 is in terms of simple returns for a 1-month period. For instance, in the months of August, September and October of 1982, around the event known as Black
Monday, prices would have plunged 2.2%, 3.9% and 4.1%, respectively, only because of margin risk. To get a sense of perspective, these numbers added represent almost one third of the total fall in stock prices (37%) during this period.

The other times of high premium were around the Fed (Volcker) fight against inflation, the first savings and loans crisis, the Iraq invasion of Kuwait, and the years of the Russian default, the LTCM downturn and the dot-com crash. In these periods, prices fell between 1% and 2.5% just because of margins.

Another useful exercise is to understand the impact in prices of changes in the credit market conditions. To compute it, we first fix the levels for \( x_t, m_t, \psi_t \) at their averages, namely, 0.92, 4.2% and 0.7% and, using the same sixth column of Table 3.11, we conclude that a 1% raise in margins (from 0.7% to 1.7%) would drop stock prices by 0.8% within one month. In a worse condition for capital constraints, considering the ted spread going from 0.7% to 3% and margins requirements fixed in 10%, the price decreasing would be about 4% during the month.

*What is driving the predictability?*

One may be fairly suspicious that the predictability results presented above are simply due to the ted spread, and have nothing to do with margins. Since the ted spread is generally high during periods of financial distress, it could predict future returns by itself. On the other hand, the same could be said about the margin requirements. Margins are higher in bad periods, and this could be enough to produce predictability of future returns.
To investigate this issue, we re-estimate column 1 from Tables 3.11, 3.12, 3.13 and 3.14, controlling for the ted spread and margin requirement alone. According to the theory from section 3.2, an increase in the ted spread should have no effect on future returns if margins are set to zero. Analogously, an increase in margins should induce no additional premium if there is no cost of buying on margin (zero ted spread). Hence, the coefficients of such controls should be not significant.

Table 3.15 shows that this is exactly what happens. The ted spread and margin requirement alone are not significant. Although the p-values of the margin factor under this augmented specification are larger than the original p-values (the ones from Tables 3.11, 3.12, 3.13 and 3.14, the correct specification according to equation 3.37), we still have empirical evidence of a margin premium at 10% of significance for 1- and 2-month ahead.

<table>
<thead>
<tr>
<th></th>
<th>h=1</th>
<th>h=2</th>
<th>h=4</th>
<th>h=12</th>
</tr>
</thead>
<tbody>
<tr>
<td>x m ted</td>
<td>4.20*</td>
<td>3.26*</td>
<td>2.41</td>
<td>2.36</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.099)</td>
<td>(0.172)</td>
<td>(0.155)</td>
</tr>
<tr>
<td>ted</td>
<td>-4.41</td>
<td>-2.54</td>
<td>-1.94</td>
<td>-6.02</td>
</tr>
<tr>
<td></td>
<td>(0.72)</td>
<td>(0.79)</td>
<td>(0.84)</td>
<td>(0.41)</td>
</tr>
<tr>
<td>m</td>
<td>-3.69</td>
<td>-3.55</td>
<td>-3.65</td>
<td>-3.51</td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
<td>(0.16)</td>
<td>(0.11)</td>
<td>(0.16)</td>
</tr>
<tr>
<td>constant</td>
<td>15.09</td>
<td>-15.79</td>
<td>17.93</td>
<td>19.65</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td>(0.13)</td>
<td>(0.06)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>R2-adj</td>
<td>1.6%</td>
<td>2.7%</td>
<td>3.9%</td>
<td>6.8%</td>
</tr>
</tbody>
</table>

White adjusted p-values between brackets

Table 3.15. Controlling for possible isolated effects of the ted spread and margins.

Hence, an increase in the ted spread will affect future returns only if margins requirements are positive. This is exactly the channel predicted by the theoretical model.
3.3.2 The margin factor and the cross-section of expected returns

As equation (3.34) indicates, in a setting with many risky assets, we can think about a common risk factor related to margins given by $x_t \psi_t$. In this case, each risky asset would load on this factor according to its average margin requirement.

Given that, as an additional test of the margin-CAPM, we can examine how average returns of different portfolios relate to their exposures to the process $x_t \psi_t$. The theory suggests that portfolios that covary more with $x_t \psi_t$ are the ones with higher average margin requirements and, because of that, should earn higher average returns.

To investigate that we first construct portfolios formed on the basis of stocks’ exposures to $x_t \psi_t$. In December of each year, we estimate a pre-ranking beta related to $x_t \psi_t$ (the margin beta) for every NYSE, AMEX and NASDAQ stock with share code 10 and 11 in the CRSP (Center for Research in Security Prices of the University of Chicago) database, using two to five years (as available) of prior monthly returns. We then form ten equally weight portfolios based on these pre-ranking margin betas and compute their returns for the next twelve months. We repeat this process for each year from 1984 to 2004. The result is monthly returns on ten margin-related beta-sorted portfolios. Figure 3.14 plots the ten portfolio’s average excess return against their post-ranking margin beta, estimated by regressing their monthly returns on $x_t \psi_t$. 
Since the ted spread is high in periods of financial distress, the margin betas are negative for all portfolios. In accordance to Figure 3.14, portfolios that suffer higher losses when $x_t \psi_t$ increases (portfolios with more negative margin betas or, given the theory, higher average margin requirements) present, on average, higher expect returns. This is in favor of the margin-CAPM.

A natural concern refers to the fact that since the ted spread is higher in bad financial periods, the price of the risk factor $x_t \psi_t$ may be closely related to the price of the standard CAPM risk. In fact, stocks that present higher losses in bad periods should pay on average higher expected returns under the standard CAPM. Given that, it is important to check whether the margin-related risk is still priced after controlling for the CAPM risk. To do that, we can estimate the usual cross-sectional regression.
\[
\sum_{t=1}^{T} r_{t,i} / T = \hat{\beta}_i^m \lambda^m + \hat{\beta}_i \lambda + u_i, \tag{3.41}
\]

where \(\hat{\beta}_i^m\) and \(\hat{\beta}_i\) are, respectively, the time-series estimates of the post-ranking margin beta and of the standard CAPM beta (obtained from regressing each of the portfolios on the market return). As usual, we do not include a free constant in the regression, since it would imply a paradoxical risk-free rate that has a nonzero excess return relative to itself. Moreover, since the sample is not too long, omitting an intercept and thus imposing a (theoretically valid) restriction of the model delivers more power.

Table 3.16 presents the results from the cross-sectional regression (3.41). Besides using the ten portfolios sorted on the margin-related beta, we also include the 25 Fama-French portfolios sorted by size and book-to-market to the analysis.\(^{31}\)

<table>
<thead>
<tr>
<th></th>
<th>10 portfolios sorted by margin beta</th>
<th>25 portfolios sorted by size and book-to-market</th>
</tr>
</thead>
<tbody>
<tr>
<td>margin beta</td>
<td>-0.003*** (0.0001)</td>
<td>-0.002*** (0.0003)</td>
</tr>
<tr>
<td></td>
<td>-0.013*** (0.0013)</td>
<td>-0.006*** (0.0021)</td>
</tr>
<tr>
<td>CAPM beta</td>
<td>0.009** (0.0008)</td>
<td>0.002** (0.0010)</td>
</tr>
<tr>
<td></td>
<td>0.009*** (0.0008)</td>
<td>0.005*** (0.0013)</td>
</tr>
</tbody>
</table>

Table 3.16. Cross-sectional regressions with 10 and 25 portfolios.

As we can see, the risk related to the factor \(x_i \psi_i\) is distinct from the standard CAPM risk. Although the price of the margin risk decreases when the regression is controlled for the CAPM risk, it still remains strongly significant. Note that the price of the margin risk

\(^{31}\) The regressions with the Fama-French portfolios use monthly observations from 1982 to 2004. For the 10 portfolios sorted by margin beta, we lost the first two years of the sample in order to estimate the first pre-ranking betas and, hence, we use monthly returns from 1984 to 2004.
appears with a negative sign because $\beta^m_i$ is also negative. Hence, a larger $\beta^m_i$ in absolute terms (i.e., a larger average margin requirement) implies a larger risk premium.

An issue with the regressions above is that the regressors in the second step, that is, the betas, are just estimates. A simple way to account for the effect of generated regressors in this application is to map the whole estimation into GMM.

Following the methodology presented in Cochrane (2005, pg. 241), for each portfolio $i = 1, \ldots, N$, we have

$$g_i (b_i) = \begin{bmatrix} E (r_{it} - \alpha_i - \beta^m_i x_t \psi_t - \beta^m_i r_{t}^{market}) \\ E [ (r_{it} - \alpha_i - \beta^m_i x_t \psi_t - \beta^m_i r_{t}^{market}) x_t \psi_t ] \\ E (r_{it} - \beta^m_i \lambda^m - \beta_i \lambda) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

where $b_i = (\alpha_i, \beta^m_i, \beta_i, \lambda^m, \lambda)$. Hence, there are $4N$ moment conditions to estimate $3N + 2$ parameters.

If we combine the last $N$ moments $E (r_{it} - \beta^m_i \lambda^m - \beta_i \lambda), i = 1, \ldots, N$, using a $2 \times N$ weighting matrix given by

$$\theta = \begin{bmatrix} \beta^m_1 & \beta^m_2 & \cdots & \beta^m_N \\ \beta_1 & \beta_2 & \cdots & \beta_N \end{bmatrix},$$

we will end up estimating the analogous to the OLS cross-sectional estimates for $\lambda^m$ and $\lambda$ presented in Table 3.16.\(^{32}\) The only difference is that we here take the uncertainty about the generated betas into account. In this case, we end up with $3N + 2$ moment conditions, the same number of parameters.

\(^{32}\) To see this, define $\theta_i = (\beta^m_i, \beta_i)'$ and $\Lambda = (\lambda^m, \lambda)'$. The combined moment conditions is given by $\theta E (R_t - \theta' \Lambda) = 0$ which can be solved as $\Lambda = (\theta' \theta)^{-1} \theta E (R_t)$. 
By performing such an exactly identified GMM estimation, we obtain values close to the ones from table 3.16. For the ten portfolios sorted by the margin-related beta, the estimates for $\lambda^m$ and $\lambda$ are, respectively, $-0.0021$ and $0.0018$, both significant at 1%. Using the 25 Fama-French portfolios, we obtain for $\lambda^m$ and $\lambda$ the values of $-0.004$ and $0.003$, respectively, both significant at 1% as well.

Given that, the results of this section support the prediction of equation (3.34) that high margin betas stocks earn comparatively higher average returns.

### 3.4 Further empirical investigation

The previous section provided empirical support for the existence of a risk premium related to margin requirements. In this section, we produce some additional results regarding the predictability of future returns.

Figure 3.11 shows that two interesting periods for the ted spread were left out of the analysis. The years between 1971 and 1982 and after 2004 are periods where our measure of capital constraint presented considerable larger mean and variance and, given that, it would be nice to test the theoretical model using these data as well. However, to do that, one would have to deal with the fact that there are no data for $x_t$ before 1982 and after 2004, and no data for margins before 1982.

To overcome these issues, we first show that data on $x_t$ is not essential to estimate equation (3.37). Since the variation in $x_t$ is small relatively to the variation in $m_t$ and $\psi_t$, omitting $x_t$ creates only a small bias. Moreover, we argue and empirically show that such a bias is negative, what, in case of not rejecting the model, does not harm the conclusion.
3.4 Further empirical investigation

After that, we propose a way of modeling margins as a function of past returns and values at risk. As we discuss, this is theoretically reasonable. Moreover, we empirically show that the proposed model is able to predict the important movements in margin, even out of the sample.

3.4.1 Omitting consumption

Suppose we didn’t have data on \( x_t \). How would this change the estimation results obtained in the previous section? Very little.

The central point is that what drives the spikes in Figure 3.13 are \( m_t \) and \( \psi_t \), with \( x_t \) almost playing no role. This happens since \( x_t \) accounts for a very small part in the variation of the factor \( x_t m_t \psi_t \). To see this we can compare the standard deviations of \( \log (x_t m_t \psi_t) \) and \( \log (m_t \psi_t) \).\(^{33}\) While the log of the full factor has standard error equal to 0.835, the standard error is equal to 0.834 for the log of the factor with \( x_t \) omitted.

To empirically understand the size and sign of the bias from omitting \( x_t \) in model (3.37), we rewrite it as

\[
\begin{align*}
  r_{t+h} - r_{t+h}^c &= \delta_1 m_t \psi_t + z_t^c \delta_2 + \delta_1 (x_t - 1) m_t \psi_t + e_{t+h}
\end{align*}
\]

and compare the estimate of \( \delta_1 \) from the restricted regression considering the full model (hence, the same estimate as above) and the regression omitting \( \delta_1 (x_t - 1) m_t \psi_t \). Table 3.17 brings both estimates of \( \delta_1 \), for \( h = 1, 2, 4 \) and 12.

\(^{33}\) Comparing the standard deviations without taking logs would be misleading since \( x_t \) is always below 1. This would depress the variance of \( x_t m_t \psi_t \) per se. With additivity from logs this effect vanishes.
Table 3.17. The effect of omitting consumption

Table 3.17 indicates that the small bias from omitting $x_t$ should be negative. The intuition for this is simple. First consider $x_t$ fixed. Since $x_t - 1 < 0$, when $m_t$ and/or $\psi_t$ increases, the omitted term $(x_t - 1)m_t\psi_t$ decreases. Since $x_t$ is very stable compared to $m_t$ and $\psi_t$, this argument is robust for letting $x_t$ to vary. Indeed, the estimated correlation between $m_t\psi_t$ and $(x_t - 1)m_t\psi_t$ is $-0.97$.

Based on this, we conclude that the lack of data on $x_t$ does not preclude the estimation of the model.

### 3.4.2 Modelling margins

Data on margins are rare. Therefore, a good model for predicting it from other observable variables should be useful. We show that a nonparametric model with past returns and values at risk can do this job.

In Brunnermeier and Pedersen (2008) model, for example, "Speculators finance their trades through collateralized borrowing from financiers who set the margins to control their
value-at-risk", where "... each financier ensures that the margin is large enough to cover the position’s value-at-risk ".

Indeed, as discussed in the introduction, the relation between margin requirements and a worst-case scenario is natural. Given that the VaR is the most common approach used by market practitioners to measure worst-case scenarios, margins and VaR should be related.

Based on this, we assume the margin requirement for an asset to be a function of its current and past 1% VaR and returns. We augment the reasoning above with past VaR since it may take some time for the margin settler to adjust it. The inclusion of past returns are justified since we are using margins as a ratio of some principal investment. If, instead of the 1% VaR we use the 2.5% or the 5%, we end up with, qualitatively, the same results.

Define $g : \mathbb{R}^{k_v+k_r+2} \to \mathbb{R}$ to be an unknown function. We then have

$$m_t = g (VaR_t, ..., VaR_{t-k_v}, r_t, ..., r_{t-k_r}) + u_t,$$

where $u_t$ is an error term.

To estimate (3.42), we have to first estimate the VaR series. There are a number of alternative ways to model the VaR of returns. Basically, they differ in how the distribution of the returns is estimated. One approach is to assume a given distribution for returns (in general, lognormality) and then to model the return’s conditional variance so that we can compute any given conditional quantile from there. This is the case of the RiskMetrics methodology. Another approach is to not impose any return distribution and directly model
the conditional quantile. We use this last approach because it is more general, robust and allows for a broad class of variables to have impact in our proxy for margins requirements.

We employ the CAViaR of Engle and Manganelli (2001). The CAViaR, which stands for conditional auto-regressive value at risk, extends a standard quantile regression by allowing a dynamic specification for the conditional quantile. The reason is that since volatilities of stock market returns cluster over time, the VaR, which is closely linked to the variance of the distribution, may exhibit a similar behavior.

Let $Q_{r_{t+1}}(\tau|I_t)$ be the $\tau_{th}$ conditional quantile of a return $r_{t+1}$ conditional on the information $I_t$ available at time $t$. Under the CAViaR specification,

$$Q_{r_{t+1}}(\tau|I_t) = X_t^\prime \theta_1(\tau) + \theta_2(\tau) Q_{r_t}(\tau|I_{t-1}),$$

(3.43)

where $X_t$ is a vector that contains the relevant conditioning information for estimating the $\tau_{th}$ conditional quantile. The conditional 1% VaR computed at $t$ for the S&P 500 returns at $t+1$ is then defined as

$$VaR_{t,1\%}(r_{t+1}) = -Q_{r_{t+1}}(0.01|I_t).$$

The recursive specification in equation (3.43) does not allow one to use standard quantile regressions methods for the estimation of $\theta_1$ and $\theta_2$. Because of that, we use the quasi-bayesian estimator of Chernozhukov and Hong (2003). We provide in the appendix a brief explanation of the estimation method and refer the reader to their article for further details.

To define the variables that belong to $X_t$, we follow Cenesizoglu and Timmermann (2008), who consider whether a range of economic variables are helpful in predicting dif-
3.4 Further empirical investigation

Different quantiles of the S&P 500 returns. According to their study, the most significant variables to predict the 1% conditional quantile are: earnings-price ratio, book to market ratio, default yield spread, default return spread, cross sectional premium, stock variance, dividend payout ratio, net equity expansion, term spread and inflation. As in Cenesizoglu and Timmermann (2008), we obtain these variables from the updated data set constructed by Goyal and Welch (2007).

In addition to these variables, we include the absolute value of the lag of the return. Engle and Manganelli (2001) provide a clear justification for using this variable: "... we would expect the VaR to increase as $r_{t-1}$ becomes very negative, as one bad day makes the probability of the next somewhat greater. It might be that very good days also increase VaR as would be the case for volatility models. Hence VaR could depend symmetrically upon $|r_{t-1}|$.

We estimate model (3.42) under four different specifications for function $g$. The first is the general function as written there, the second assumes full linearity, the third imposes linearity inside $g$ (single index model) and the forth is the additive nonparametric model,

$$m_t = g_{v,0}(VaR_t) + ... + g_{v,k_v}(VaR_{t-k_v})$$
$$+ g_{r,0}(r_t) + ... + g_{r,k_r}(r_{t-k_r}) + u_t,$$

(3.44)

where $g_{v,0}, ..., g_{v,k_v}, g_{r,0}, ..., g_{r,k_r}$ are unknown functions from $\mathbb{R}$ to $\mathbb{R}$.

We impose for all of them $k_v = k_r = 6$. It seems unreasonable that higher lag orders may be useful in the model. For a number of criteria both inside and outside the sample
(correlation, mean squared error and mean absolute error), the last specification is the one that best performs and, therefore, we focus on it.

We use in the estimation of (3.44) the method of penalized splines, firstly proposed by O’Sullivan (1986), refined by Eiders and Marx (1996), and made popular through the book by Ruppert, Wand, and Carroll (2003). We present the method in the appendix.\textsuperscript{34}

First, using data from April 1982 (when data on margin begins) to November 2008 (when data for $X_t$, from Goyal and Welch, for the VaR estimation ends), we estimate $\hat{m}_t$, the fitted margin under model (3.44). Figure 3.15 plots $m_t$ and $\hat{m}_t$.

![Fig. 3.15. Margin x fitted margin (in sample).](image)

The figure above shows that the spike in margin that occurred about October 1987 was perfectly fitted by the model. The timing of the spike from the last financial crisis was

\textsuperscript{34} The other nonparametrics specifications for $g$ (full nonparametric and single index) were estimated by local linear regression and Ichimura’s method respectively.
perfectly fitted too, although the model over-predicted its size. In general, the trend of the margin is well reproduced (the only exception is the period between 2003 and 2007). The main issue with the model may be that it predicts more variability for margins than there exists indeed. However, this should not affect estimation if it is random. The correlation between both series is 65%.

One may fairly wonder how much of these good results are due to over-fitting of the nonparametric model. Very little, if anything. To show this, we estimate the model using data up to December 2001 (75% of the sample) and analyzes the results from January 2002 to November 2008 (25% of the sample).

Figure 3.16 plots the margin and both the in-sample (from the previous regression) and out-of-sample fitted margins.

![Fig. 3.16. Comparing in and out-of-sample fitted values.](image-url)
According to Figure 3.16, there is no significant loss from predicting margins out of the sample. The correlations between fitted and actual margins for this period are 60% and 50% in and out of the sample, showing a small decrease. It is fair to say, however, that the size of the raise in margins in the last financial crisis was better predicted out of the sample.

This indicates that model (3.44) along with a CAViaR estimation for the value at risk can well explain the important movements in margins. To confirm this, we reestimate model (3.37) using now the fitted margins instead of the true ones and compare the results.

<table>
<thead>
<tr>
<th>horizon</th>
<th>true margin</th>
<th>fitted margin</th>
</tr>
</thead>
<tbody>
<tr>
<td>h=1</td>
<td>2.55***</td>
<td>2.38***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>h=2</td>
<td>1.94***</td>
<td>2.02***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>h=4</td>
<td>1.15*</td>
<td>1.24*</td>
</tr>
<tr>
<td></td>
<td>(0.093)</td>
<td>(0.088)</td>
</tr>
<tr>
<td>h=12</td>
<td>0.52</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>(0.320)</td>
<td>(0.260)</td>
</tr>
</tbody>
</table>

*Newey-West adjusted p-values between brackets*

Table 3.18. Using fitted margins instead of true margins

We can also consider both omissions at the same time. That is, we rewrite

\[
 r_{t+h} - r_{t+h}^c = \delta_1 x_t m_t \psi_t + z_t' \delta_2 + e_{t+h} \\
 = \delta_1 x_t (\hat{m}_t + u_t) \psi_t + z_t' \delta_2 + e_{t+h} \\
 = \delta_1 x_t \hat{m}_t \psi_t + z_t' \delta_2 + \delta_1 x_t u_t \psi_t + e_{t+h} \\
 = \delta_1 \hat{m}_t \psi_t + z_t' \delta_2 + \delta_1 (x_t - 1) \hat{m}_t \psi_t + \delta_1 x_t u_t \psi_t + e_{t+h}
\]
and compare the original estimates of $\delta_1$ (with both no control and controlling for dy and vol) with the ones obtained when running

$$r_{t+h} - r_{t+h}^c = \delta_1 \tilde{m}_t \psi_t + z_t' \delta_2 + \tilde{e}_{t+h}. \quad (3.45)$$

Table 3.19 present the results.

<table>
<thead>
<tr>
<th>_horizon</th>
<th>$\delta_1$ estimate (no controls)</th>
<th>$\delta_1$ estimate (controlling for dy and vol)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>h=1</td>
<td>2.55***</td>
<td>2.21***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>h=2</td>
<td>1.94***</td>
<td>1.88***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>h=4</td>
<td>1.15*</td>
<td>1.17*</td>
</tr>
<tr>
<td></td>
<td>(0.093)</td>
<td>(0.076)</td>
</tr>
<tr>
<td>h=12</td>
<td>0.52</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>(0.320)</td>
<td>(0.242)</td>
</tr>
</tbody>
</table>

*Newey-West adjusted p-values between brackets*

Table 3.19. Omitting consumption and using fitted margins

Column 1 in Table 3.19 brings the original estimates for $\delta_1$ using no control in the regression. Column 3 brings the original estimates of $\delta_1$ including in the regressions the dividend yield and the volatility. Columns 2 and 4 are their counterparts omitting consumption and using fitted margins as in equation (3.45).

Individual results from both previous sections still hold. The estimates for $\delta_1$ under equation (3.45) are close to the original ones, mainly for $h > 1$. For 1-month ahead, the already discussed negative bias is more evident, but still not large.

We now use such evidences to justify the augmentation of our sample.
3.4 Further empirical investigation

### 3.4.3 Results with augmented sample

Table 3.20 presents the average and standard deviation of our main variables.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P 500 margin</td>
<td>-</td>
<td>4.08%</td>
<td>6.23%</td>
<td>-</td>
<td>1.62%</td>
<td>2.38%</td>
</tr>
<tr>
<td>hed spread</td>
<td>1.86%</td>
<td>0.69%</td>
<td>0.88%</td>
<td>1.05%</td>
<td>0.57%</td>
<td>0.79%</td>
</tr>
<tr>
<td>consumption ratio</td>
<td>-</td>
<td>0.92</td>
<td>-</td>
<td>-</td>
<td>0.08</td>
<td>-</td>
</tr>
<tr>
<td>margin factor</td>
<td>-</td>
<td>2.55</td>
<td>-</td>
<td>-</td>
<td>2.80</td>
<td>-</td>
</tr>
</tbody>
</table>

* until April 2010

Table 3.20. Descriptive statistics

Two special periods were not considered in the estimation in section 3.3, namely, the years between 1971 and 1981 and after 2004. They were not included in our main sample since there are no data for margins and consumption before 1982 and no data for consumption after 2004. However, as we argued previously, estimating equation (3.45) should deliver reasonable estimates for \( \delta_1 \). Therefore, we use this fact to investigate the model using the whole period between January 1971 and April 2010.

We define two distinct series for \( \hat{m}_t \). The first is simply the fitted values from the nonparametric model. The second uses these fitted values only for the periods where there are no margin data. The following results are qualitatively the same for both series, and we present the ones that uses the later definition.

The included periods are special since, as we can see in Table 3.20, they have higher average and standard deviation for hed spread and margins. However, this is not the whole story, as Figure 3.17 shows.
According to Figure 3.17, for the years to the left of the first dashed line and to the right of the second dashed line, the factor seems to follow a different stochastic process compared to the one between dashed lines. Crucially, it spends more time at high than at low levels. Once it spikes, it takes a while to return.

Hence, besides the included periods being more volatile with respect to the margin factor variables, they also have a different regime for the margin factor. In what follows, we refer to these periods as turbulent periods. Such a difference in regimes within the sample have interesting consequences for the empirical analysis of the theoretical model, as we show now.

When we run equation (3.45) using the whole sample from 1971 to 2010, the estimated coefficient of the interaction $\hat{m}_t \psi_t$ is non-significant for all horizons with any control
(and no control). This would be enough evidence for an econometrician to reject the theoretical model. However, as we argue now, this conclusion would be misleading.

In a nutshell, between 1982 and 2004, we have a margin factor that spikes and returns fast to its original level. As presented and discussed in section 3.3.1, this leads to forecasting regressions with a positive and significant coefficient for the margin factor at short horizons, but a coefficient statistically equal to zero at longer horizons. However, on the other hand, for the periods before 1982 and after 2004 the spikes take more time to return. The immediate effect of that is the forecasting power moving from short to longer horizons. As a consequence, when we run regressions mixing these two distinct regimes for the margin factor, we wrongly conclude that there is no premium for margin. The explanation for that is the following. Considering the full sample, the included periods (before 1982 and after 2004) are important enough to eliminate the forecasting power at the short run, but not important enough not to create a forecasting power at the long run.

To see this, we first run rolling-windows regressions for $h = 1$ using the full sample. We estimate equation (3.45), with dividend yield and volatility included, using overlapping windows with 60 months each. We then plot the t-statistics obtained in each window, relating them to two variables that measure volatility in the margin factor (which, as discussed above, is related to the length of the spikes): the standard deviations of the ted spread and of the factor ($\hat{m}_t \psi_t$). Figures 3.18 and 3.19 present the results.

---

35 Using longer windows is not appropriate since the included periods have 10 and 5 years each.
Fig. 3.18. t-statistic vs. ted standard deviation (1-month ahead regressions).
3.4 Further empirical investigation

Fig. 3.19. t-statistic vs. factor ($m_t\psi_t$) standard deviation (1-month ahead regressions).

In both figures above the vertical dashed lines indicate the 95% significance levels. The horizontal dashed lines indicate the third quartile of the variables in the y-axis. From the 407 sub-samples, in 93 of them we conclude for the presence of a positive margin premium (t-statistic higher than 1.96). In 10 sub-samples we have a negative and significant t-statistic and we see this as a small sample issue (since it has no theoretical reasoning and the sub-samples are small indeed).

Figure 3.18 uses the standard deviation of the ted spread as a measure of turbulence in the credit market. The points above (below) the horizontal dashed line come from turbulent (non turbulent) sub-samples. Out of the 93 sub-samples with premium for margin, only 1 belongs to a turbulent period. Figure 3.19 defines turbulent periods as the ones with
high standard deviation of the product $\hat{m}_t \psi_t$, our factor in this section. According to this criterion, no turbulent period with margin premium was left.

In other words, turbulent periods (which are characterized by the margin factor having longer spikes) do not produce predictability at short horizons. A direct conclusion is that the margin premium for short horizons disappeared with the augmented sample because the included periods are turbulent ones.

To complete the story, we have now to show that in such turbulent periods the forecasting power of the margin factor has moved to the long horizons regressions. Accordingly, we run a 12-month ahead regression for the months between 1971 and 1980 (returns go until 1981 then). Unfortunately, to run the same regression for the recent financial crisis period however, we would have to wait some additional years (as Figure 3.17 shows, the margin factor returned to its low level only in the end of 2009). Table 3.21 presents the results.

<table>
<thead>
<tr>
<th>h = 12, period: 1971 - 1980</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>m fitted X ted</td>
<td>0.69**</td>
<td>0.79**</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.041)</td>
</tr>
<tr>
<td>dy</td>
<td>29.27*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td></td>
</tr>
<tr>
<td>vol</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>-6.19</td>
<td>-49.7</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>R2-adj</td>
<td>5.2%</td>
<td>24.2%</td>
</tr>
</tbody>
</table>

Newey-West adjusted (order equal to 11) p-values between brackets.

Table 3.21. Turbulent period: 12-month ahead regression
As expected, the predictive power of the margin factor shows up for longer horizons during this period. Given that, we conclude that the horizon at which the margin premium shows up may depend on the regime of the margin factor. During most of the time, when the margin factor spikes it returns fast and, hence, the premium shows up at short horizons. However, during turbulent periods, the spikes are longer and the premium takes more time to show up. Interestingly, a regression with both kinds of periods mixed together may lead to a wrong rejection of the model. This last point may be a useful observation for other models’ tests.

3.5 Conclusion

We evaluated the effect of margins on asset prices. This is an additional channel for the relation between downside risk and prices. Our main contribution was to provide evidence of the existence of an aggregate margin-related premium. This fact have important theoretical consequences that are not limited to the understanding of asset prices. For instance, it affects monetary policy efficiency during some periods.

Besides discussing the choice of data to estimate the margin-based models, we also provided results that may be useful to deal with problems of data availability. Indeed, margin requirements data are not easily available. To deal with this issue, we proposed a nonparametric model for explaining margins from current and past values of the value at risk of the specific asset. We argued that this model has good theoretical and empirical properties. This isolated result should be useful in many applications that use margin requirements.
Bibliography


Rates and Haircuts.” NBER Macroeconomics Annual, forthcoming.


European Economic Review 49 (3): 531-560.

risks model for asset prices.”, working paper


Appendix

Appendix to Chapter 1

Proof of Proposition 1:

Substituting the restrictions into the object function, the problem is given by

$$\max_{\xi \in \mathbb{R}} Q_t^\tau \left( u \left( W_t - P_t \xi - P_t^f \xi^f \right) + \beta u \left( X_{t+1} \xi + X_{t+1}^f \xi^f \right) \right)$$

By the quantile equivariance, this is equivalent to

$$\max_{\xi \in \mathbb{R}} u \left( W_t - P_t \xi - P_t^f \xi^f \right) + \beta u \left( \xi Q_t^\tau (X_{t+1}) + X_{t+1}^f \xi^f \right)$$

and the first order conditions are

$$\xi : u' (C_t) P_t = \beta u' (Q_t^\tau (C_{t+1})) Q_t^\tau (X_{t+1})$$

$$\xi^f : u' (C_t) P_t^f = \beta u' (Q_t^\tau (C_{t+1})) X_{t+1}^f$$

which implies

$$P_t = \frac{\beta u' (Q_t^\tau (C_{t+1}))}{u' (C_t)} Q_t^\tau (X_{t+1})$$

$$P_t^f = \frac{\beta u' (Q_t^\tau (C_{t+1}))}{u' (C_t)} X_{t+1}^f$$

Specializing \( u(c) = \frac{c^{1-\gamma} - 1}{1-\gamma} \),

$$P_t = \beta \left( Q_t^\tau \left( \frac{C_{t+1}}{C_t} \right) \right)^{-\gamma} Q_t^\tau (X_{t+1})$$

$$P_t^f = \beta \left( Q_t^\tau \left( \frac{C_{t+1}}{C_t} \right) \right)^{-\gamma} X_{t+1}^f$$

CQFD. □

Proof of Proposition 2:
The risky asset and risk-free asset prices are given, respectively, by

\[ P_t = \eta_t Q_t^r (X_{t+1}) \]  \hspace{1cm} (6.46)
\[ P_{t}^{f} = \eta_t X_{t+1}^{f} \]

where \( \eta_t \equiv \beta \left( Q_t^r \left( \frac{C_{t+1}}{C_t} \right) \right)^{-\gamma} \).

An arbitrage opportunity occurs if and only if it is possible to construct \( \Xi_t = \left( \xi_t, \xi_t^{f} \right) \) such that

\[ \xi_t P_t + \xi_t^{f} P_{t}^{f} = 0 \]  \hspace{1cm} (6.47)
\[ \xi_t X_{t+1} + \xi_t^{f} X_{t+1}^{f} \geq 0 \]

with the second equation holding as an inequality for at least one point in the support of \( X_{t+1} \).

Substituting (6.46) into the first equation of (6.47),

\[ \xi_t \eta_t Q_t^r (X_{t+1}) + \xi_t^{f} \eta_t X_{t+1}^{f} = 0 \]

\[ \Rightarrow \xi_t^{f} X_{t+1}^{f} = -\xi_t Q_t^r (X_{t+1}) \]

which, into the second equation of (6.47) gives the necessary and sufficient condition for arbitrage,

\[ \xi_t (X_{t+1} - Q_t^r (X_{t+1})) \geq 0 \]

with inequality for at least one point in the support of \( X_{t+1} \).

Therefore, all we need to rule out arbitrage is to impose

\[ Q_t^r (X_{t+1}) \in (\min \{ \text{supp} \ (X_{t+1}) \}, \max \{ \text{supp} \ (X_{t+1}) \}) \]
If $X_{t+1}$ is a continuous random variable, this is implied by imposing $\tau \in (0, 1)$.

CQFD. ■

**Proof of Proposition 3:**

First, note that if $\ln(x) \sim N(\mu, \sigma^2)$ then $Q_\tau(x) = \exp(\mu + \sigma\Phi^{-1}(\tau))$. This holds since

$$F_X(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

$$\Rightarrow F_X^{-1}(\tau) = \exp(\mu + \sigma\Phi^{-1}(\tau))$$

According to (1.5),

$$\log(C_{t+1}/C_t) | t \sim N(\mu_c, \sigma_c^2)$$

$$\log(R_{t+1}) | t \sim N(\mu_r, \sigma_r^2)$$

Therefore,

$$Q^*_t(C_{t+1}/C_t) = \exp(\mu_c + \sigma_c\Phi^{-1}(\tau))$$

$$Q^*_t(R_{t+1}) = \exp(\mu_r + \sigma_r\Phi^{-1}(\tau))$$

Dividing both sides of (1.3) and (1.4) by $P_t$, and using the quantile equivariance property,

$$1 = \beta \left( Q^*_t \left( \frac{C_{t+1}}{C_t} \right) \right)^{-\gamma} Q^*_t(R_{t+1})$$

$$1 = \beta \left( Q^*_t \left( \frac{C_{t+1}}{C_t} \right) \right)^{-\gamma} R^f_{t+1}$$

where $R_{t+1} = \frac{X_{t+1}}{P_t}$. 
Substituting (6.48) into (6.49) and taking logs from both sides,

\[ \log (\beta) - \gamma \mu_c - \gamma \sigma_c \Phi^{-1}(\tau) + \mu_r + \sigma_r \Phi^{-1}(\tau) = 0 \]

Hence, since \( E_t(r_{t+1}) = \mu_r \),

\[ E_t(r_{t+1}) = -\log (\beta) + \gamma \mu_c + \Phi^{-1}(\tau) (\gamma \sigma_c - \sigma_r) \]

For the risk-free rate, using (6.50) and (6.48) in the same way,

\[ r_{t+1}^f = -\log (\beta) + \gamma \mu_c + \Phi^{-1}(\tau) \gamma \sigma_c \]

Therefore,

\[ E_t(r_{t+1} - r_{t+1}^f) = -\sigma_r \Phi^{-1}(\tau) \]

CQFD. \( \Box \)

**Proof of Proposition 4:**

As in the proof of Proposition 3, we use the fact that if \( \ln (x) \sim N(\mu, \sigma^2) \) then \( Q^r(x) = \exp (\mu + \sigma \Phi^{-1}(\tau)) \). Given that,

\[ Q^r_t(C_{t+1}/C_t) = \exp (\mu_c + \sigma_c \Phi^{-1}(\tau)) \quad (6.51) \]

\[ Q^r_t(R_{t+1}) = \exp (\mu_r + \varphi \sigma_r \Phi^{-1}(\tau)) \]

Hence, using (6.49),

\[ \ln \beta - \gamma \mu_c - \gamma \sigma_c \Phi^{-1}(\tau) + \mu_r + \varphi \sigma_r \Phi^{-1}(\tau) = 0 \]

and, since \( E_t(r_{t+1}) = \mu_r \), we have

\[ E_t(r_{t+1}) = -\ln \beta + \gamma \mu_c + (\gamma - \varphi) \sigma_r \Phi^{-1}(\tau) \]
For the risk-free rate, using (6.50) and the conditional quantile for consumption growth,

\[ r_{t+1}^f = -\ln \beta + \gamma \mu_c + \gamma \sigma_c \Phi^{-1}(\tau) \]

Therefore,

\[ E_t \left( r_{t+1} - r_{t+1}^f \right) = -\varphi \sigma_t \Phi^{-1}(\tau) \]

CQFD. ■

Proof of Proposition 5:

\[
E \left[ \left( \tau_0 - 1 \left[ C_{t+1}/C_t < \left( \beta_0 R_{t+1}^f \right)^{v_0} \right] \right) Z_t \mid \mathcal{F}_t \right] \\
= \left( \tau_0 - E \left[ 1 \left[ C_{t+1}/C_t < \left( \beta_0 R_{t+1}^f \right)^{v_0} \right] \mid \mathcal{F}_t \right] \right) Z_t \\
= \left( \tau_0 - \Pr \left( \varepsilon_{c,t+1} < 0 \mid \mathcal{F}_t \right) \right) Z_t \\
= 0, \text{ since } Q^{\tau_0} \left( \varepsilon_{c,t+1} \mid \mathcal{F}_t \right) = 0.
\]

Using the same steps, we also get

\[
E \left[ \left( \tau_0 - 1 \left[ R_{t+1} < R_{t+1}^f \right] \right) Z_t \mid \mathcal{F}_t \right] = 0.
\]

CQFD. ■

Proof of Proposition 6:

We verify the conditions of Theorem 2.6 of Newey and McFadden (1994) - NM below. First, note that the theorem requires \( V_{t+1} \equiv \left( R_{t+1}, C_{t+1}/C_t, R_{t+1}^f, Z_{t} \right) \) to be iid. However, as the authors point out on page 2133, the iid assumption may be replaced by strictly stationarity and ergodicity. According to Proposition 3.44 in White (2001),
strictly stationarity and $\alpha$-mixing implies ergodicity, so assumption (i) ensures $V_{t+1} \equiv \left(Y_{t+1}, Z_t, R^f_{t+1}\right)$ to be strictly stationary and ergodic.

(NM 2.6.i) This is the condition that ensures global identification (see lemma 2.3 in NM). However, if instead of $W_0$ being positive semi-definite one imposes $W_0$ to be positive definite, NF2.6.i can be trivially exchanged for $E[g(V_{t+1}, \theta)] = 0$ if and only if $\theta = \theta_0$. By assumption (iv) $W_0 > 0$ (a choice for $W_T$ that satisfies this will be provided). So, we have to show that $E[g(V_{t+1}, \theta)] = 0$ if and only if $\theta = \theta_0$.

The fact that $E[g(V_{t+1}, \theta_0)] = 0$ was already derived in the body of the text. We are left to show that $E[g(V_{t+1}, \theta)] = 0 \Rightarrow \theta = \theta_0$.

First, considering the second set of moment conditions,

\[
E \left[ \left( \tau - 1 \left[ R_{t+1} \leq R^f_{t+1} \right] \right) Z_t \right] \\
= E \left[ \left( \tau - E \left[ 1 \left[ R_{t+1} \leq R^f_{t+1} \right] \mid Z_t \right] \right) Z_t \right] \\
= E \left[ \left( \tau - F_{R_{t+1}|Z_t} \left( R^f_{t+1} \mid Z_t \right) \right) Z_t \right] \\
= E \left[ (\tau - \tau_0) Z_t \right], \text{ since } R^f_{t+1} \in Z_t \text{ by assumption (vi)} \\
= 0 \Rightarrow \tau = \tau_0, \text{ since } 1 \in Z_t \text{ by assumption (vi)}
\]

Hence, $\tau_0$ is identified. We now consider the first set of moment conditions (with $\tau_0$ already identified):
\begin{align*}
E \left[ \left( \tau_0 - 1 \left[ \frac{C_{t+1}}{C_t} < \left( \beta R_{t+1}^f \right)^\psi \right] \right) Z_t \right]
&= E \left[ \left( \tau_0 - E \left[ \left[ \frac{C_{t+1}}{C_t} < \left( \beta R_{t+1}^f \right)^\psi \right] \mid Z_t \right] \right) Z_t \right] \\
&= E \left[ \left( \tau_0 - F_{(C_{t+1}/C_t)|Z_t} \left( \left( \beta R_{t+1}^f \right)^\psi \mid Z_t \right) \right) Z_t \right] \\
&= 0 \Rightarrow F_{(C_{t+1}/C_t)|Z_t} \left( \left( \beta R_{t+1}^f \right)^\psi \mid Z_t \right) = \tau_0, \text{ since } 1 \in Z_t \text{ by assumption (vi)}
\end{align*}

By assumption (v), \( F_{(C_{t+1}/C_t)|Z_t} \) is a continuous strictly increasing function within its support. By assumption (vi), \( R_{t+1}^f \in Z_t \) and hence

\[ F_{(C_{t+1}/C_t)|Z_t} \left( \left( \beta R_{t+1}^f \right)^\psi_0 \mid Z_t \right) = \tau_0. \]

Therefore, we must have

\[ \left( \beta_0 R_{t+1}^f \right)^\psi_0 = \left( \beta R_{t+1}^f \right)^\psi, \]

which holds if either

\[ (\beta, \psi) = (\beta_0, \psi_0) \]

or

\[ R_{t+1}^f = \frac{\psi_0 \log (\beta_0) - \psi \log (\beta)}{\psi - \psi_0} \text{ at every } t. \]

By assumption (vii), \( R_{t+1}^f \) is a non-degenerate random variable. Hence,

\[ (\beta, \psi) = (\beta_0, \psi_0) \text{ a.s.} \]
Therefore, we conclude

\[ E \left[ g \left( V_{t+1}, \theta \right) \right] = 0 \implies \theta = \theta_0 \text{ a.s.} \]

(NM 2.6.ii) Assumption (iii) ensures \( \theta_0 \) as an interior point of \( \Theta \).

(NM 2.6.iii) This is satisfied because \( g \left( V_{t+1}, \theta \right) \) is discontinuous only when \( \frac{C_{t+1}}{C_t} = \left( \beta R_{t+1}^f \right)^{1/\gamma} \) and \( R_{t+1} = R_{t+1}^f \). By assumption (v), these two cases have probability zero.

(NM 2.6.iv) Note that since for any value of \( \theta \) we have

\[
\begin{align*}
&\left\| (\tau - 1 \left[ \frac{C_{t+1}}{C_t} < \left( \beta R_{t+1}^f \right)^{\psi} \right] \right\| Z_t\| \leq \| Z_t \|
&\left\| (\tau - 1 \left[ R_{t+1} < R_{t+1}^f \right] \right\| Z_t\| \leq \| Z_t \|
\end{align*}
\]

we ensure \( E \left( \sup_{\theta \in \Theta} \| g \left( V_{t+1}, \theta \right) \| \right) < \infty \) by assumption (ii).

Therefore, we conclude that \( \hat{\theta} \overset{p}{\to} \theta_0 \), by Theorem 2.6 of Newey and McFadden (1994), CQFD. \( \blacksquare \)

**Proof of Proposition 7**: By Lemma 1 below, specializing \( \theta = \tilde{\theta} \), we have \( W_T \overset{p}{\to} \Sigma_0^{-1} \)

where

\[
\Sigma_0 \equiv E \left[ g \left( V_{t+1}, \theta_0 \right) g \left( V_{t+1}, \theta_0 \right)' \right].
\]

Now, we prove that \( \Sigma_0 \) is a positive definite matrix (since every positive definite matrix is invertible and its inverse is also positive definite, we then are done: \( \Sigma_0^{-1} \) exists and is positive definite.) First, note that
\[ \Sigma_0 = E[E(A_{t+1}|Z_t) \otimes Z_t'] \]

where \( A_{t+1} \) is a 2 \( \times \) 2 matrix with entries

\[
A_{11} = \left( \tau_0 - 1 \left[ \frac{C_{t+1}}{C_t} < Q_t^r \left( \frac{C_{t+1}}{C_t} \right) \right] \right)^2
\]
\[
A_{12} = A_{21} = \left( \tau_0 - 1 \left[ \frac{C_{t+1}}{C_t} < Q_t^r \left( \frac{C_{t+1}}{C_t} \right) \right] \right) (\tau_0 - 1 [R_{t+1} < Q_t^r (R_{t+1})])
\]
\[
A_{22} = (\tau_0 - 1 [R_{t+1} < Q_t^r (R_{t+1})])^2
\]

under the theoretical model.

We now compute \( E(A_{11}|Z_t) \), \( E(A_{12}|Z_t) \), \( E(A_{21}|Z_t) \) and \( E(A_{22}|Z_t) \).

\[
E(A_{11}|Z_t) = E(A_{22}|Z_t) = \tau_0^2 (1 - \tau_0) + (\tau_0 - 1)^2 \tau_0 = \tau_0 (1 - \tau_0)
\]

and
\[ E(A_{12}|Z_t) = E(A_{21}|Z_t) \]
\[ = E \left[ \left( \tau_0 - 1 \left[ \frac{C_{t+1}}{C_t} < Q_t^{r_0} \left( \frac{C_{t+1}}{C_t} \right) \right] \right) | Z_t \right] E \left[ (\tau_0 - 1) [R_{t+1} < Q_t^{r_0} (R_{t+1})] | Z_t \right] \]
\[ + Cov \left[ \tau_0 - 1 \left[ \frac{C_{t+1}}{C_t} < Q_t^{r_0} \left( \frac{C_{t+1}}{C_t} \right) \right], \tau_0 - 1 [R_{t+1} < Q_t^{r_0} (R_{t+1})] | Z_t \right] \]
\[ = (\tau_0 (1 - \tau_0) + (\tau_0 - 1) \tau_0) (\tau_0 (1 - \tau_0) + (\tau_0 - 1) \tau_0) \]
\[ + Cov \left[ \tau_0 - 1 \left[ \frac{C_{t+1}}{C_t} < Q_t^{r_0} \left( \frac{C_{t+1}}{C_t} \right) \right], \tau_0 - 1 [R_{t+1} < Q_t^{r_0} (R_{t+1})] | Z_t \right] \]
\[ = Cov \left[ \tau_0 - 1 \left[ \frac{C_{t+1}}{C_t} < Q_t^{r_0} \left( \frac{C_{t+1}}{C_t} \right) \right], \tau_0 - 1 [R_{t+1} < Q_t^{r_0} (R_{t+1})] | Z_t \right] \]
\[ = P(\varepsilon_{c,t+1} < 0, \varepsilon_{r,t+1} < 0 | Z_t) - P(\varepsilon_{c,t+1} < 0 | Z_t) P(\varepsilon_{r,t+1} < 0 | Z_t) \]
\[ = \varphi_t - \tau_0^2, \text{ for } \varphi_t \equiv P(\varepsilon_{c,t+1} < 0, \varepsilon_{r,t+1} < 0 | Z_t). \]

Therefore, \( E(A_{t+1}|Z_t) \) is positive definite if both the following conditions hold,

\[ \tau_0 (1 - \tau_0) > 0 \]
\[ \tau_0^2 (1 - \tau_0)^2 - (\varphi_t - \tau_0^2)^2 > 0. \]

The first condition is ensured by assumption (viii). The second condition can be simplified further,
\[
\tau_0^2 (1 - \tau_0)^2 > (\varphi_t - \tau_0^2)^2 \\
[\tau_0 (1 - \tau_0)]^2 > (\varphi_t - \tau_0^2)^2 \\
(\tau_0 - \tau_0^2)^2 > (\varphi_t - \tau_0^2)^2 \\
\tau_0 - \tau_0^2 > \varphi_t - \tau_0^2 \\
\varphi_t < \tau_0,
\]

which is assumption (ix).

With respect to \(Z_tZ_t^t\) we can also show that it is positive definite. In fact, for any \(\lambda \in \mathbb{R}^m\),

\[
\lambda'Z_tZ_t^t\lambda = (Z_t^t\lambda)^2 \geq 0,
\]

holding with inequality only if \(Z_t^t\lambda = 0\). But, given assumption (x), \(Z_t^t\lambda = 0\) only if \(\lambda = 0\).

Therefore, since both \(E[A_{t+1}|Z_t]\) and \(Z_tZ_t^t\) are positive definite, \(E[A_{t+1}|Z_t] \otimes Z_tZ_t^t\) is positive definite and \(\Sigma_0\) is also positive definite, CQFD. ■

**Proof of Proposition 8:**

First, an observation:

Even though \(g(V_{t+1}, \theta)\) is not differentiable in \(\theta\), \(E[g(V_{t+1}, \theta)]\) is. In fact, for

\[
g_1(V_{t+1}, \theta) \equiv \left( \tau - 1 \left[ C_{t+1}/C_t < \left( \beta R_{t+1}^f \right)^{\psi} \right] \right) Z_t
\]

and

\[
g_2(V_{t+1}, \theta) \equiv \left( \tau - 1 \left[ R_{t+1} < R_{t+1}^f \right] \right) Z_t
\]
we have:

$$\partial_r E \left[ g_1 (V_{t+1}, \theta) \right] = E (Z_t)$$

$$\partial_\beta E \left[ g_1 (V_{t+1}, \theta) \right] = -\partial_\beta E \left[ 1 \left[ C_{t+1} / C_t < \left( \beta R_{t+1}^f \right)^\psi \right] Z_t \right]$$

$$= -\partial_\beta E \left[ E \left[ 1 \left[ C_{t+1} / C_t < \left( \beta R_{t+1}^f \right)^\psi \right] | Z_t \right] Z_t \right]$$

$$= -\partial_\beta E \left[ F_{(C_{t+1} / C_t)} | Z_t \right] \left( \left( \beta R_{t+1}^f \right)^\psi \left[ Z_t \right] \right)$$

$$= -E \left[ \partial_\beta F_{(C_{t+1} / C_t)} | Z_t \right] \left( \left( \beta R_{t+1}^f \right)^\psi \left[ Z_t \right] \right)$$

$$= -E \left[ f_{(C_{t+1} / C_t)} | Z_t \right] \left( \left( \beta R_{t+1}^f \right)^\psi \left[ Z_t \right] \right) \psi (\psi - 1) \left( R_{t+1}^f \right)^\psi \left[ Z_t \right]$$

$$= -E \left[ f_{z_{c,t+1}} (0 | Z_t) \psi_0 \beta_0 (\psi - 1) \left( R_{t+1}^f \right)^\psi \left[ Z_t \right] \right], \text{ for } \theta = \theta_0.$$

$$\partial_\psi E \left[ g_1 (V_{t+1}, \theta) \right] = -E \left[ \partial_\psi F_{C_{t+1} / C_t} \left( \left( \beta R_{t+1}^f \right)^\psi \left[ Z_t \right] \right) \right]$$

$$= -E \left[ f_{z_{c,t+1}} (0 | Z_t) \left( \beta_0 R_{t+1}^f \right)^\psi_0 \log \left( \beta_0 R_{t+1}^f \right) \left[ Z_t \right] \right], \text{ for } \theta = \theta_0.$$

$$\partial_r E \left[ g_2 (V_{t+1}, \theta) \right] = E (Z_t)$$

$$\partial_\beta E \left[ g_2 (V_{t+1}, \theta) \right] = 0$$

$$\partial_\psi E \left[ g_2 (V_{t+1}, \theta) \right] = 0$$
Given that, define $G_0 = \nabla_\theta E [g (V_{t+1}, \theta_0)]$, where $\nabla_\theta E [g (V_{t+1}, \theta_0)]$ is the $2m \times 3$ matrix derived above. 

(end of observation)

We now check conditions (i) to (v) from Theorem 7.2 of Newey and McFadden (1994) to establish the asymptotic normality of our estimator.

(NF.7.2.i) $E [g (V_{t+1}, \theta_0)] = 0$ is shown in the body of the text.

(NF.7.2.ii) The fact that $E [g (V_{t+1}, \theta)]$ is differentiable at $\theta_0$ was shown in the observation in the beginning of the proof. $G_0' W_0 G_0$ is nonsingular by assumption (xii).

(NF.7.2.iii) Assumption (iii) ensures $\theta_0$ as an interior point of $\Theta$.

(NF.7.2.iv) According to proposition 5 $\{g (V_{t+1}, \theta_0), \mathcal{F}_t\}$ is a martingale difference sequence. Given that, we check the conditions of Corollary 5.26 in White’s (2001). We have

$$E \| g (V_{t+1}, \theta_0) \|^{2+2\delta} \leq E \| Z_t \|^{2+2\delta}$$

$$\leq \max \left\{ 1, E \| Z_t \|^{2r+2\delta} \right\}, \text{ where } r > 2$$

$$\leq \infty \text{ by assumption (ii').}$$

Moreover, applying Lemma 1 below for $\theta = \theta_0$ we have

$$\frac{1}{T} \sum_{t=1}^{T} g (V_{t+1}, \theta_0) g (V_{t+1}, \theta_0)' \overset{p}{\to} \Sigma_0,$$

where $\Sigma_0 \equiv E [g (V_{t+1}, \theta_0) g (V_{t+1}, \theta_0)']$.

Therefore, according to Corollary 5.26 in White(2001),
Appendix

\[ \sqrt{T} \left( \frac{1}{T} g(V_{t+1}, \theta_0) \right) \xrightarrow{d} N(0, \Sigma_0). \]

(NF.7.2.v) Andrews (1994) shows that empirical processes defined from moment conditions as \( g(V_{t+1}, \theta_0) \) are stochastically equicontinuous \((g(V_{t+1}, \theta_0) \) fits in what he calls type I class of real functions - note that even though \( g_1(V_{t+1}, \theta_0) \) has a nonlinear function of the parameters inside the indicator function,

\[
g_1(V_{t+1}, \theta) = \left( \tau - 1 \left[ C_{t+1}/C_t < \left( \beta R^f_{t+1} \right)^\psi \right] \right) Z_t
\]

this can be written as,

\[
g_1(V_{t+1}, \theta) = \left( \tau - 1 \left[ \log C_{t+1}/C_t < \psi \log \beta + \psi \log R^f_{t+1} \right] \right) Z_t
\]
given that the log is a strictly increasing function and \( C_{t+1}/C_t, \beta, R^f_{t+1} > 0 \).

Therefore, by Theorem 7.2 of Newey and McFadden (1994), we conclude that

\[
\sqrt{T} \left( \hat{\theta} - \theta_0 \right) \xrightarrow{d} N \left( 0, (G_0' W_0 G_0)^{-1} \right. \left. \hat{\Sigma}_0 \right) \]

CQFD. ■

**Lemma 1**: Define \( \Sigma(\theta) = E \left[ g(V_{t+1}, \theta) g(V_{t+1}, \theta)' \right] \). Then,

\[
\frac{1}{T} \sum_{t=1}^{T} g(V_{t+1}, \theta) g(V_{t+1}, \theta)' \xrightarrow{p} \Sigma(\theta).
\]

**Proof of Lemma 1**: First, note that \( g(V_{t+1}, \theta) \) is an \( \mathcal{F}_{t+1} \) measurable function which is strictly stationary and \( \alpha \)-mixing what implies that \( g(V_{t+1}, \theta) g(V_{t+1}, \theta)' \) is also strictly stationary and \( \alpha \)-mixing of the same size (Theorem 3.49 of White (2001)).
Now, all we need is to apply a Law of Large Numbers for $\alpha$-mixing sequences (Corollary 3.48 of White (2001)). The conditions of White’s corollary are (a) $\{g(V_{t+1}, \theta) g(V_{t+1}, \theta)\}'$ has to be an $\alpha$-mixing sequence of size $-r/(r - 1)$, $r > 1$ and (b) $E \|g(V_{t+1}, \theta) g(V_{t+1}, \theta)\|^ {r+\delta} < \infty$ for some $\delta > 0$, where $\|\cdot\|$ denotes the $L_\infty$-norm. Condition (a) is directly satisfied by assumption (i). For condition (b), note that

$$\|g(V_{t+1}, \theta) g(V_{t+1}, \theta)\|$$

$$\equiv |g_{i_0}(V_{t+1}, \theta) g_{j_0}(V_{t+1}, \theta)|, \text{ where } (i_0, j_0) = \arg\max_{i \geq 1, j \leq \dim(g)} |g_i(V_{t+1}, \theta) g_j(V_{t+1}, \theta)|$$

$$= |g_{i_0}(V_{t+1}, \theta) \|g_{j_0}(V_{t+1}, \theta)|$$

$$\leq C^2 \|g(V_{t+1}, \theta)\|^2,$$  

by norm equivalence, for some positive constant $C$.

and hence

$$E \|g(V_{t+1}, \theta) g(V_{t+1}, \theta)\|^{r+\delta} \leq C^2 \max\left\{1, E \|g(V_{t+1}, \theta)\|^{2r+2\delta}\right\}$$

by Cauchy-Schwarz. So, we would need some assumption such as "there exist some $\delta > 0$ such that $E \|g(V_{t+1}, \theta)\|^{2r+2\delta} < \infty$". However, note that

$$\left\| \left( \tau - 1 \left[ \frac{C_{t+1}}{C_t} < \left( \beta R_{t+1}^f \right)^\psi \right] \right) Z_t \right\| \leq \|Z_t\|$$

and, therefore, it is enough to assume that there exist some $\delta > 0$ such that $E \|Z_t\|^{2r+2\delta} < \infty$, which is our assumption (ii’) CQFD. ■
Model estimation under lognormality

The solved model under lognormality and stochastic economic uncertainty is given by

\[
\begin{align*}
    r_{t+1} &= -\ln \beta_0 + \frac{1}{\psi_0} \mu_c + \left( \frac{1}{\psi_0} - \varphi_0 \right) \sigma_t \Phi^{-1} (\tau_0) + \varphi_0 \sigma_t u_{t+1} \\
    r^f_{t+1} &= -\ln \beta_0 + \frac{1}{\psi_0} \mu_c + \frac{1}{\psi_0} \sigma_t \Phi^{-1} (\tau_0) \\
    g_{t+1} &= \mu_c + \sigma_r \eta_{t+1} \\
    \sigma^2_{t+1} &= \alpha_0 + \rho_0 (\sigma^2_t - \alpha_0) + \sigma_v v_{t+1}
\end{align*}
\]

in accordance to sub-section 1.3.2.

A possible estimator for the parameters is the simulated method of moments (SMM) of McFadden (1986), Pakes and Pollard (1987), and Duffie and Singleton (1993), the last one in the context of time-series as we have here.

Analogous to sub-section 1.3.2, we focus only on the estimation of \( \theta_0 = (\beta_0, \psi_0, \tau_0) \), fixing the dynamics parameters using the values in Table 1.1.

Define \( m_t \) to be a vector of empirical observations on variables whose moments are of interest: the risk-free rate, the excess return, and the consumption growth. Such a vector should contain the moments to be matched by the estimator. In our case, \( p = 6 \) and

\[
m_t = \left( r_t - r^f_t, \left( r_t - r^f_t \right)^2, r^f_t, \left( r^f_t \right)^2, g_t, \sigma^2_t \right).
\]

Define \( m_t(\theta) \) to be a vector with the synthetic counterpart of \( m_t \), whose elements are computed on the basis of artificial data generated by the model using parameter
values $\theta$. The number of observations in the artificial time series is given by $\kappa T$, where $T$ is the sample size and $\kappa$ is a positive integer.

The SMM estimator of $\theta_0$ is defined as

$$
\hat{\theta}_{SMM} = \arg \min_{\theta \in \Theta \subseteq \mathbb{R}^3} \left( \frac{1}{T} \sum_{t=1}^{T} m_t - \frac{1}{\kappa T} \sum_{t=1}^{\kappa T} m_t(\theta) \right) \left( \frac{1}{T} \sum_{t=1}^{T} m_t - \frac{1}{\kappa T} \sum_{t=1}^{\kappa T} m_t(\theta) \right)'
$$

where, to allow for a direct comparison with the simulation results from section 3.2, each moment is equally weighted.

Under the regularity conditions of Duffie and Singleton (1993),

$$
\sqrt{T} \left( \hat{\theta}_{SMM} - \theta_0 \right) \xrightarrow{d} N \left( 0, (1 + 1/\kappa) (D_0' D_0)^{-1} D_0' \Omega_0 D_0 (D_0' D_0)^{-1} \right),
$$

where,

$$
D_0 = E \left( \partial_0 m_t (\theta) | \theta = \theta_0 \right)
$$

and

$$
\Omega_0 = \sum_{j=-\infty}^{\infty} E \left( (m_t - E [m_t]) (m_{t-j} - E [m_{t-j}])' \right).
$$

As usual, $\Omega_T$ can be obtained by the Newey-West estimator. With respect to $D_0$, since there is no analytical solution for the differentiation, the derivatives are numerically computed, and the expectation approximated by the average over the $\kappa T$ simulated points.
Under this framework, by drawing monthly observations, aggregating them to yearly, and constructing $m_t$ from the same data used in section 1.4.3 (also yearly-aggregated), we end up with the following estimates:

<table>
<thead>
<tr>
<th># of draws:</th>
<th>$12 \times 10^3$</th>
<th>$12 \times 10^4$</th>
<th>$12 \times 10^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
</tr>
<tr>
<td>(se)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>EIS</td>
<td>0.61</td>
<td>0.59</td>
<td>0.61</td>
</tr>
<tr>
<td>(se)</td>
<td>(0.25)</td>
<td>(0.17)</td>
<td>(0.19)</td>
</tr>
<tr>
<td>$\tau$</td>
<td>0.47</td>
<td>0.46</td>
<td>0.46</td>
</tr>
<tr>
<td>(se)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
</tbody>
</table>

As in the simulation exercise we assume that the decision interval of the agent is monthly but the targeted data to match are annual. Therefore, we simulate at the monthly frequency but match the yearly moments.

These results are in line with the calibrated values in section 1.3.2.

Appendix to Chapter 3

CAViaR Estimation Method

This is the Chernozhukov and Hong (2003) quasi-bayesian estimator. Basically, to estimate $\{\theta_1, \theta_2\}$ in

$$Q_{\tau_{t+1}}(\tau|I_t) = X\theta_1(\tau) + \theta_2Q_{\tau_t}(\tau|I_{t-1}),$$
we draw a long series \( \{\theta_{1i}, \theta_{2i}\}_{i=1}^{S} \) using MCMC (Metropolis-Hastings) from the pseudo density function, the so-called quasi-posterior,

\[
P_n(\theta_1, \theta_2) = \frac{e^{L_n(\theta_1, \theta_2)}}{\int e^{L_n(\theta_1, \theta_2)} d\theta_1 d\theta_2},
\]

where \( \pi(\theta_1, \theta_2) \) is a flat prior, \( L_n(\theta_1, \theta_2) \) is a pseudo likelihood function,

\[
L_n(\theta_1, \theta_2) = -\sum_{t=s}^{N} w_t(\tau) \rho_{\tau}(r_{t+1} - Q_{r+1}(\tau|I_t)),
\]

and

\[
w_t(\tau) = \frac{1}{\tau(1-\tau)} \text{ and } \rho_{\tau}(u) = (\tau - 1 \cdot (u < 0)) u.
\]

Then, as shown in Chernozhukov and Hong (2003),

\[
\hat{\theta}_1 = \frac{1}{S} \sum_{i=1}^{S} \theta_{1i} \overset{p}{\rightarrow} \theta_1 \text{ and } \hat{\theta}_2 = \frac{1}{S} \sum_{i=1}^{S} \theta_{2i} \overset{p}{\rightarrow} \theta_2.
\]

Because of the recursive specification in equation (3.43), one needs to compute initial conditions to initialize equation (6.57). We set \( s = 100 \) and use the unconditional quantile to evaluate \( Q_{r_{99}}(\tau|I_{98}) \).
Nonparametric Additive Model for Margins

To illustrate the idea underneath the estimation method suppose first that there is only one explanatory variable, that is, \( m_t = g(y_t) + u_t \), where the only assumption about \( g \) is smoothness. Then, to estimate \( g \) one can use some basis function with "good approximation" properties, parametrizing \( g \) as

\[
g(y, \beta) = \sum_{j=1}^{P+K} B_j(y) \beta_j
\]

where \( P \geq 1 \) is an integer, \( \beta = (\beta_0, \beta_1, \ldots, \beta_{P+K})' \) is a vector of regression coefficients, and \( B_0(y), \ldots, B_{P+K}(y) \) is the basis. A common choice for the basis is the so-called cubic spline, where \( K = 2 \),

\[
B_j(y) = |y - \kappa_j|^3, \text{ for } j = 1, ..., P
\]

\[
b_{P+1}(y) = 1
\]

\[
b_{P+2}(y) = y
\]

and \( \kappa = \{\kappa_j : j = 1, ..., P\} \) is a set of points in the range of \( y \), called knots.

Defining \( x_t = (|y_t - \kappa_1|^3, |y_t - \kappa_2|^3, \ldots, |y_t - \kappa_P|^3, 1, y_t)' \) and \( X = (x_1 x_2 \ldots x_T) \), we have

\[
m_t = X\beta + u
\]

where \( u = (u_1, u_2, \ldots, u_T)' \).

In principle, if \( P \) was chosen to be large enough to approximate \( g \) well, this model could be fitted by minimizing \( u'u \). However, if \( P \) is too large, the estimation is going
to over-fit the data, that is, it will begin to fit the noise in the data, which will cause, in
the limit, the perfect interpolation of the data points. An option would be to choose \(P\) by
cross-validation methods, where the knots would be selected from a set of candidate knots
in a way similar to stepwise regression. However, as one can suspect, this can easily turn
out to be computationally too expensive.

A better solution is to use penalized splines (P-splines). O’Sullivan (1986, 1988) and
Eiders and Marx (1996) idea is to set \(P\) intentionally large (it can even be a knot at each
unique value of the support, which, in this case, would be called smoothing splines) and
control over-fitting by using least-squares estimation with a roughness penalty. The penalty
is on the integral of the square of a specified derivative, usually the second. In this case,
one would minimize

\[
\sum \left( \frac{g(y)}{e} \right) ^2 + \lambda \int_{y_{\text{min}}}^{y_{\text{max}}} \left[ \frac{d^2 g(y)}{dy^2} \right] ^2 dy
\]

(6.60)

The first term in (6.60) is the traditional sum of the square of the residuals. The sec-
ond term is the roughness penalty, which increases as the cubic splines get rougher, that is,
when their slope change very rapidly – the integrated second derivative of the regression
function is a measure for it. Therefore, \(\lambda\) defines the degree of smoothing: the larger \(\lambda\), the
larger the smoothness of the estimator (the estimator’s bias increases and its variance de-
creases). For \(\lambda = 0\), one tends to the perfect interpolation of all data points as \(P\) increases;
by the other hand, for \(\lambda \rightarrow \infty\), one has the linear least squares estimator. The advantage
of the P-splines is that now one has to choose a single parameter value to determine the
smoothness of the estimator, in contrast to having to define the number and location of the knots. Cross-validation is now computationally much cheaper.

It is important to note that object (6.60) is equal to \( u' u + \lambda \beta' \Lambda \beta \), where \( \Lambda \) is a matrix of known numbers. In the case of the cubic splines,

\[
\Lambda = \begin{pmatrix}
12 & 12 - \frac{\kappa_2}{2} - \frac{\kappa_1}{2} + \kappa_1 \kappa_2 & \cdots & 12 - \frac{\kappa_p}{2} - \frac{\kappa_1}{2} + \kappa_1 \kappa_P & 0 & 0 \\
12 - \frac{\kappa_2}{2} - \frac{\kappa_1}{2} + \kappa_1 \kappa_2 & 12 & \cdots & 12 - \frac{\kappa_p}{2} - \frac{\kappa_2}{2} + \kappa_2 \kappa_P & 0 & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
12 - \frac{\kappa_2}{2} - \frac{\kappa_1}{2} + \kappa_1 \kappa_P & 12 - \frac{\kappa_p}{2} - \frac{\kappa_2}{2} + \kappa_2 \kappa_P & \cdots & 12 & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0
\end{pmatrix}
\]

Minimizing (6.60), one gets the following closed-form estimator of \( g(y) \),

\[
\hat{g}(y) = X (X'X + \lambda \Lambda)^{-1} X'm
\]  

(6.61)

where \( X \) and \( \Lambda \) were defined above and \( m = (m_1, m_2, \ldots, m_T)' \).

The estimator in (6.61) was obtained for a single explanatory variable. It is straightforward to generalize it for the case of the complete model

\[
m_t = g_{e,0}(VaR_t) + \ldots + g_{e,k_v}(VaR_{t-k_v}) \\
+ g_{r,0}(r_t) + \ldots + g_{r,k_r}(r_{t-k_r}) + u_t.
\]  

(6.62)

In this case the function to be minimized is
Appendix

\[ u'u + \lambda_1 \int_{V a R_t^{\min}}^{V a R_t^{\max}} \left[ g''_{v,0}(V a R_t) \right]^2 dV a R_t + \ldots \] \hspace{1cm} (6.63)

\[ + \lambda_{k_v + 1} \int_{V a R_t^{\min}}^{V a R_{t-k_v}^{\max}} \left[ g''_{v,k_v}(V a R_{t-k_v}) \right]^2 dV a R_{t-k_v} \]

\[ + \lambda_{k_v + 2} \int_{r_t^{\min}}^{r_t^{\max}} \left[ g''_{r,0}(r_t) \right]^2 dr_t + \ldots \]

\[ + \lambda_{k_v + k_r + 2} \int_{r_{t-k_r}^{\min}}^{r_{t-k_r}^{\max}} \left[ g''_{r,k_r}(r_{t-k_r}) \right]^2 dr_{t-k_r} \]

where \( u = (u_{1+q}, u_{2+q}, \ldots, u_T) \) from equation (6.62) and \( q = \max (k_v, k_r) \).

Analogously to the univariate case, we approximate each unknown smooth function in (6.63) using a cubic spline basis and, therefore, we can write \( m_t = X \beta + u \) and minimize \( u'u \) with respect to \( \beta \).