ESSAYS ON LIQUIDITY-PREFERENCE IN MARKETS WITH BORROWING RESTRICTIONS

Proefschrift voorgedragen
tot het behalen van de graad
van doctor in de Toegepaste Economische Wetenschappen
door
Fernando MIERZEJEWSKI
Committee

Prof. Dr. Marc Goovaerts (promotor)
*Katholieke Universiteit Leuven*

Prof. Dr. Cynthia Van Hulle (promotor)
*Katholieke Universiteit Leuven*

Prof. Dr. Jan Beirlant
*Katholieke Universiteit Leuven*

Prof. Dr. Jan Dhaene
*Katholieke Universiteit Leuven*

Prof. Dr. Roger Laeven
*Universiteit van Tilburg*

Prof. Dr. Steven Vanduffel
*Katholieke Universiteit Leuven*
## Contents

**Preface** vii

1 The Allocation of Economic Capital in Opaque Financial Conglomerates 1

1.1 Agency Costs and Moral-Hazard in Opaque Financial Institutions 6

1.2 The Optimal Capital Structure 9

1.3 Liquidity Restrictions in Capital Markets 15

1.4 Economic Capital as the Optimal Deductible 19

1.5 The Allocation of Capital within Multidivisional Corporations 22

1.6 The Distorted Demand for Capital 26

1.7 Risk-Based Allocation for Centralised Organisations 29

1.8 Optimal Decentralised Mechanism 33

1.9 Heavy-Tailed Probability Distributions 37

1.10 Conclusions 41

2 The Optimal Liquidity Principle with Restricted Borrowing 43

2.1 Liquidity-Preference in the Monetary Equilibrium 45
2.2 Preference for Liquidity as Behaviour towards Risk . . . . . . . . 49
2.3 The Optimal Liquidity Principle . . . . . . . . . . . . . . . . 55
2.4 The Aggregate Liquidity-Preference . . . . . . . . . . . . . . 62
2.5 The Monetary Equilibrium with the Optimal Liquidity Prin-
ciple . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66
2.6 The Gaussian Liquidity Principle . . . . . . . . . . . . . . . . 67
2.7 The Capital Market Line Extended . . . . . . . . . . . . . . . 72
2.8 Extended Macroeconomic Analysis . . . . . . . . . . . . . . . 76

3 A Model of Equilibrium in Markets of Cash Balances 83
3.1 Equilibrium in the Market of Cash Balances . . . . . . . . . . 84
3.2 Preference for Liquidity as the Optimal Retention . . . . . . 87
3.3 The Internal Rate of Return on Risk . . . . . . . . . . . . . . 90
3.4 Scaling the Equilibrium Rate of Interest . . . . . . . . . . . . 94
3.5 The Neutral Gaussian Demand for Cash Balances . . . . . . 95
3.6 The Neutral Gaussian Demand with a Regulatory Condition 100
3.7 The Risk-Structure of Interest Rates . . . . . . . . . . . . . . 104
3.8 Deviations from Equilibrium in Neutral Markets . . . . . . 108
3.9 Non-Neutral Markets of Cash Balances . . . . . . . . . . . . 113
3.10 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 117

A Utility Theory of Choice under Risk 119

B The Dual Theory of Choice 123

C An Axiomatic Characterisation of the Distorted Probability
List of Figures

1.1 The Optimal Capital Principle of Neutral Decision Makers . . 39
1.2 The Optimal Capital Line and the Allocation Rule of Merton
and Perold (1993) . . . . . . . . . . . . . . . . . . . . . . . 40

2.1 The Gaussian Liquidity Principle with Different Distortions . 69
2.2 The Gaussian Liquidity Principle with Different Combin-
tions of Mean Returns and Volatilities . . . . . . . . . . . . 71

3.1 Semi-Elasticity of the Demand for Cash Balances with Re-
spect to the Rate of Interest. . . . . . . . . . . . . . . . . . . 99
3.2 Confidence Rate and Semi-Elasticity of the Demand for Cash
Balances as a Function of the Confidence Ratio under Gaus-
sian Risks. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 102
3.3 Risk-Structure of Interest Rates as a Function of the Ex-
pected Return. . . . . . . . . . . . . . . . . . . . . . . . . . . 106
3.4 Risk-Structure of Interest Rates as a Function of Volatility. . 107
Preface

“...as it is impossible to hammer anything out of it for moral purposes, let us treat it aesthetically, and see if it will turn to account in that way.”

Thomas De Quincey — from On Murder Considered as One of the Fine Arts.

A model is presented in this thesis for the characterisation of the preference for liquidity in different economic contexts. It is intended as a preliminary step in the formulation of a general theory of liquidity.

The model is built on two fundamental assumptions. Firstly, individuals are supposed to face borrowing restrictions when trading capital and securities, in such a way that they cannot continuously modify the composition of their portfolios. Secondly, facing borrowing restrictions imply that the holders of security portfolios and insurance claims are exposed to imbalances (over the level of cash holdings) that cannot be rectified with the aid of short-term debt. The costs induced by such imbalances are accordingly characterised by the actuarial prices of the corresponding residual insurance claims.

The approach is connected to the model of deposit insurance of Robert Merton, which characterises the price of economic capital as the price of a warrant on the value of the underlying portfolio of securities. Assuming that individuals can trade capital and securities without restrictions, Merton concludes that the price of such guarantees must be equal to the price of a put option on the value of the portfolio of securities. Since this hypothesis is released in the model proposed in this thesis, the price of deposit insurance contracts is described by actuarial principles instead of put options — i.e. a tool from financial economics is replaced by a tool from insurance practice.
The characterisation of an *optimal liquidity principle* — representing the *optimal* cash balance allocated to a given portfolio of risky claims — is at the core of the model. This principle is related to the capital strategy that must be followed by any *rational* decision-maker. This is the same approach suggested by James Tobin, 1958, to characterise *liquidity-preference*.

Unlike the model of Tobin, however, where rational individuals maximise the *expected utility* the wealth of a certain portfolio combining cash holdings and securities induces on its holder, in the model proposed in this thesis rational individuals minimise the *total cost of insuring* the portfolio of securities. This plan is consistent with the alternative description of *deposit insurance* proposed above.

At the corporate level, the optimal liquidity principle represents the *optimal* level of cash reserves maintained by financial companies and their subsidiaries. It provides a basis to define both *centralised* and *decentralised* capital allocation rules, which are strictly *risk-based*.

More generally, the model can be used to describe the aggregate demand for *cash holdings* in the economy — i.e. it can be used to characterise the *preference for liquidity* of the economy. It thereby provides a theoretical setting for the analysis of the *monetary equilibrium*, on which basis, the traditional problems of macroeconomics can be reinterpreted.

The optimal liquidity principle can be alternatively used to analyse the effects of credit and monetary flows over the cost of capital. The *liquidity response* of aggregate credit and capital markets can be described on these terms.

The fundamentals and major conclusions of the model will be briefly discussed hereafter.

**Borrowing Restrictions**

According to the Modigliani and Miller (1958) proposition, *rational* investors demand no *cash provisions*, for they are costly and do not affect the market value of aggregate portfolios. The theorem is a consequence of the hypothesis that in *perfect* markets financial securities and cash balances can be traded at any moment and without restrictions in quantities. Then individuals can always remove the imbalances in their portfolios, for they
can always borrow and lend any amount of capital at a fixed level of the interest rate.

In real markets, however, financial firms and private investors face borrowing restrictions, which means that if the amount of funds invested on liabilities is raised up to certain levels, the cost of borrowing must increase as well. As a matter of fact, lenders charge premiums on loans that depend on the credit quality of their counterparts. Thus, if leverage ratios are drastically incremented in a certain market, lenders will likely increase their concerns about the capability to pay of borrowers and accordingly raise the price of credit.

Borrowing restrictions are likely to appear in markets with information asymmetries, which may arise from two different sources: agency costs between shareholders and managers, and the moral-hazard implicit in the contracts established by firms with their customers.

The problems caused by agency costs include mismanagement and underperformance resulting when stockholders cannot fully observe the actions taken by managers — or when due to institutional rigidities shareholders cannot act promptly to reverse undesired results. It can be also the case that managers behave poorly from the point of view of stockholders because they pursue strategies that maximise their own interest, which may differ from that of the firm. This situation may be especially severe in companies where incentives are not properly established.\(^1\)

Moral-hazard, on the other hand, is induced by the fact that the portfolios held by financial institutions are not observed by their customers, who are thus unable to effectively assess the probability that their deposits will be returned in due time — i.e. customers cannot assess the risk of default of their deposits. As stated by Merton (1997), financial firms are opaque institutions (see also Ross, 1989).

Averse-to-risk customers accordingly require some institutional guaranty that promised payments will be honoured with certainty. Such a guaranty can take the form of a cash holding (equal to a percentage of total liabilities) to be delivered in case of default — which explicitly accounts for moral-hazard — and can be reinforced by a warrant issued by another institution or some governmental division, whose capacity and willingness to pay are beyond question. In other words, customers agree to make deposits only if

---

\(^{1}\)The problem of agency costs has been mentioned, among others, by Fama, 1980, Barnea et al., 1981, Jensen, 1986, and Merton and Perold, 1993. See also Tobin, 1982b.
they are at least partially insured (see also Merton and Bodie, 1992).

Merton (1974, 1977, 1978) demonstrates that providing deposit insurance is equivalent to issue a put option on the value of the aggregate portfolio. The cost of deposit insurance can then be explicitly stated in terms of the volatility or standard deviation of the series of capital returns of the underlying portfolio.

However, the hypothesis of continuous trading is also required in the derivation of the option pricing formula (see e.g. Black and Scholes, 1973). In fact, if individuals can modify the composition of their portfolios at any moment, then cash holding strategies can be always replicated by issuing and exchanging options. In other words, hedging, capital cushions and deposit insurance are all perfect substitutes under such circumstances.

But financial companies do face borrowing restrictions in practice, which means that they cannot always execute transactions in the quantities they need — at least at a price they are willing to pay. On the other hand, non-standarised policies are normally traded in insurance markets, which are assigned different prices depending on the information owned by the insurer and insured parties (see e.g. Venter, 1991, Wang et al., 1997, and Goovaerts et al., 2005). Within this context, the proposition of restricted borrowing can be reestablished by claiming that hedging is not a perfect substitute for insurance.

A new framework for the pricing of deposit insurance that effectively incorporates the presence of borrowing restrictions is thus lacking. Providing such a framework is one of the major goals of this thesis — as has been already stated — for only in this way the effect of borrowing restrictions on economic decisions and hence, the role of liquidity in the economy, are possible to be characterised.

For this purpose, the problem of deposit insurance will be analysed with the tools and models of actuarial sciences. A broad class of mathematical tools is actually available in this field for the pricing of risky claims, which is connected in its roots to the same theory of choice under risk that is at the base of financial theory — as will be shown hereafter.

---

2 The Modigliani-Miller proposition can be actually obtained from a well known result from option pricing theory, the put-call parity. See Miller, 1998.
Actuarial Risk Principles

*Risks* will be regarded as *random variables* in the following, defined over some common *probability space* \((\Omega, \mathcal{F}, \mathbb{P})\).\(^3\) Every risk \(X\) will be then completely characterised by its corresponding *cumulative* or *tail* probability function, respectively defined as \(F_X(x) = P\{X \leq x\}\) and \(T_X(x) = P\{X > x\}\), \(\forall x\). The fundamental identity \(T_X(x) = 1 - F_X(x)\) relates tail and cumulative probability functions.

The *actuarial pricing* of risks is intimately related to the practice of insurance, which requires every risk to be exchanged by some fixed payment. In other words, as suggested by Goovaerts et al., 1984, the calculation of *insurance* or *actuarial* prices is based on the hypothesis that every risky claim can be compensated by some fixed *premium*.

Consequently, if \(X\) denotes the class of random variables representing risks, then a *risk principle* \(\Pi\) is defined as a *functional*:

\[
\Pi : X \rightarrow \mathbb{R}
\]

Alternatively, every *risk principle* can be defined over the set of probability distributions — for every risk is connected to a unique probability function. Within this context, it is said that the premium \(\Pi[X]\) represents the *fair* or *actuarial* price of insuring the risk \(X\).

For a risk principle to effectively represent the contracts established in insurance markets, it must be related to some (mathematical) *order* defined in the class of risks \(X\). Only then it is possible to assess which claim is *riskier* than the other among any pair of prospects belonging to \(X\). Besides, as long as a risk principle preserves a certain order in \(X\), it will systematically assign *higher* prices to *riskier* claims.

In the following, we will be interested in one particular order, the *first stochastic order*, defined in such a way that a certain risk is *lower* than another one (or equivalently, one risk is *stochastically dominated* by the other) if the tail probability function of the first risk is uniformly lower than the tail probability function of the second risk, i.e.:

\(^3\)The letters \(\Omega, \mathcal{F}\) and \(\mathbb{P}\) respectively denote the set of possible *elementary* events, the set of possible *combined* events (*sigma-algebra*) and the function assessing probabilities to elementary and combined events. See e.g. De Finetti, 1973.

\(^4\)The *first* stochastic order actually belongs to a broader class of stochastic orders,
In other words, a certain risk $X$ is lower than the risk $Y$ in the first stochastic order if the probability that $X$ surpasses some level $x$ is always lower than the probability that $Y$ surpasses the same level. This condition provides a clear intuition of the risk assumed when holding a certain prospect.

Alternatively, every order on the class $X$ can be related to the preferences of individuals about risks. The characterisation of the decisions of individuals in economic contexts can be formally established on these grounds. This problem has lead to the formulation of a number of theories of choice under risk. Among them is distinguished the (expected) utility theory of choice under risk, which is at the base of much of modern economic and financial economic theory.

An alternative theory of choice has been given increasing attention during the last twenty years in the actuarial and financial literature, baptised as the dual theory of choice under risk. The main departure of the dual theory from the utility theory is that in the dual theory preferences are characterised by a principle that modifies probabilities, instead of payments, as it is the case in utility theory.

The main features of the expected utility and the dual theories of choice are briefly examined in Appendixes A and B. We will next show that a mathematical principle can be formally established on these grounds to characterise the optimal proportion of cash (that should be) added to any portfolio of securities.

### The Optimal Liquidity Principle

As stated by Tobin, 1958, the problem of liquidity-preference corresponds to the determination of the optimal combination of risk and cash.

More precisely, Tobin analyses the behaviour of a representative decision-maker who holds some aggregate exposure $X$ and a cash balance that can be lent at the interest rate $r$. Assuming that the underlying risk $X$ is described by a Gaussian probability distribution and that both capital and defined in terms of a recursive series of statistical moments, see e.g. Bawa, 1978, Goovaerts et al., 1984, Wang and Young, 1998, and Bauerle and Muller, 2006.
securities can be traded without restrictions (in such a way that every combination of risk and cash can be attained by performing market operations), he demonstrates that the locus of efficient combinations of risk and cash is represented by a straight line in the plane of expected returns and standard deviations.

On the other hand, Tobin points out that if the preferences of the representative decision-maker regarding the different combined portfolios can be characterised by a certain utility function (measuring the level of satisfaction that he or she obtains from the wealth produced by portfolios), every level of utility determines an indifference curve in the plane of expected returns and standard deviations.

Hence the optimal portfolio containing risk and cash can be characterised by that combination in the line of efficient portfolios that maximises the expected utility (defined in Equation A.1, in Appendix A) of the representative decision-maker. A necessary condition to guarantee the existence of the optimal portfolio is that the representative decision-maker shows aversion-to-risk. On these grounds, Tobin regards liquidity preference as behaviour towards risk.

Within this context, the method of Tobin determines an optimal liquidity principle, which can be applied to describe the demand for liquidity under conditions of perfect competition, i.e. when capital and securities can be traded without restrictions. In a similar way, the distorted probability principle of Equation B.1 (in Appendix B) can be used to obtain an alternative liquidity principle, which explicitly accounts for the cost of bankruptcy and accordingly incorporates the condition of restricted borrowing.

Indeed, notice that if $X$ represents the percentage return of the underlying portfolio of securities, then every individual that maintains a proportion $\lambda$ of his or her total investment in cash, has to afford the following loss at the end of the investment period:

$$(X + \lambda)_- = -\min(0, X + \lambda)$$

---

5 The optimal portfolio is thus determined at the point where the rate of variation of the expected return with respect to the standard deviation is equal to the marginal utility of substituting a unit of expected return by an additional unit of standard deviation. See Equations (3.4), (3.5) and (3.7) in the paper of Tobin, 1958.

6 The model has been actually used to derive an expression for the money demand, see e.g. Holsntrom and Tirole, 2000, Lucas, 2000, and Choi and Oh, 2003.
Thus the loss is only realised when \( X < -\lambda \). Besides, the opportunity cost \( r \cdot \lambda \) has to be always paid, since the proportion of investment maintained as a balance could be alternatively lent to obtain the return \( r \).

The burden of bankruptcy can be alternatively transferred to some insurance company, but then the actuarial price of the claim has to be paid, which can be expressed, in particular, in terms of the distorted probability principle. Rational individuals will choose the cash proportion \( \lambda \) in order to minimise the total cost of insurance:

\[
\min_{\lambda} E_{\varphi}[(X + \lambda)_-] + r \cdot \lambda \tag{1}
\]

Therefore, the proportion \( \lambda \) determines a compromise between the cost of insurance on the one hand and the opportunity cost of capital on the other — because raising \( \lambda \) diminishes the insurance cost but at the same time increases the cost of keeping idle balances, and vice versa.

The solution to the optimisation problem of Equation 1 is determined at the point where the marginal reduction in the insurance price, equal to the default probability \( P\{X \leq -\lambda\} \), is equal to the marginal cost of capital \( r \), i.e.:\(^7\)

\[
T_{\varphi,-X}(\lambda^*) = P_{\varphi}\{-X > \lambda^*\} = P_{\varphi}\{X \leq -\lambda^*\} = r \iff \lambda^* = T_{\varphi,-X}^{-1}(r) \tag{2}
\]

The optimal proportion of cash is thereby characterised by an optimal exchange between a certain cash flow and a flow of probability. A well defined liquidity demand function is obtained in this way, that is always inversely related to the level of the interest rate — for the quantile function \( T_{\varphi,-X}^{-1} \) is always inversely related to its argument.

Notice that the optimisation problem of Equation 1 characterises a static decision — and not a dynamic decision, as the model of deposit insurance of Robert Merton does. This is a realistic representation of the problem faced by individuals who have to decide some level of reserves to be maintained during some fixed period of time, as a stock of capital to be used when borrowing becomes scarce or expensive.

\(^7\)This result is demonstrated in Sections 1.2 and 1.4 in this thesis. See also Sections 2.3 and 3.2.
The liquidity principle defined in Equations 1 and 2 is proposed in this thesis to characterise the optimal demand for liquidity in different economic contexts.

Firstly, at the corporate level, the principle provides a risk-based rule to determine the amount of capital to be maintained by financial institutions. Moreover, the rule can be applied to derive both a centralised and a decentralised mechanism for the allocation of capital among the subsidiaries of some financial conglomerate.

As stated in Chapter 1, a general theory of capital can be built on these grounds, which naturally extends the theoretical frameworks of the Modigliani-Miller proposition and the model of deposit insurance of Robert Merton.

Later in Chapter 2 the principle is used to characterise the money demand and hence, the monetary equilibrium of the economy. This model provides a theoretical setting to analyse the effect of monetary interventions in economies where individuals face borrowing restrictions.

Finally, in Chapter 3 the principle is applied to characterise the demand for cash balances in some market of short-term (interbank) loans. The market cost of capital — as determined by the market equilibrium — is then explicitly dependent on the statistical description of risks and the aggregate amounts of supplied and demanded balances. In fact, a risk structure of interest rates is defined on this basis.

At the Corporate Level

Recall that in the model of Robert Merton financial firms are obliged to insure their liabilities, for only in this way they can convincingly assure to credit sensitive customers that their deposits are free of default. As long as firms are able to continuously trade capital and securities, the price of the guarantee must be equal to the price of a put option on the value of net assets (equal to assets minus liabilities). Then the value of the firm does not depend on the level of capital, as predicted by the Modigliani-Miller proposition.

Since the amounts of capital and securities cannot be continuously modified in the presence of borrowing restrictions, an alternative approach should
be implemented to characterise the demand for capital under such circumstances. It is in fact natural to assume that companies will rely on cash holdings and insurance contracts when they cannot adjust their balances in financial markets. Then companies must replace an instrument from financial economics (hedging) by a tool from actuarial science (insurance).

As already stated, the problem of Equation 1 satisfactorily represents the trade-off faced by a decision-maker who has to decide between establishing an insurance contract (and paying the corresponding *actuarial premium*) on the one hand, and relying on borrowing and lending (at the interest rate $r$) in some market of loans on the other. This is the same problem faced by financial companies in markets with borrowing restrictions, as stated in the previous paragraph.

On these grounds, the optimal liquidity principle of Equation 2 is proposed in this thesis to characterise the demand for capital of opaque financial institutions that cannot hedge continuously.

An important result obtained in this setting is that the level of capital *does affect* the value of firms in the presence of borrowing restrictions — for the total price of deposit insurance, as determined by the objective function in Equation 1, depends on it. A theoretical explanation is thus provided of why financial firms and private investors demand capital in practice, and why the Modigliani-Miller proposition may fail to apply. Moreover, the optimal capital principle explicitly extends the option-based principle of Robert Merton.\footnote{Both principles are compared in Section 1.9 in this thesis.}

Within multidivisional corporations, the necessity of keeping divisional cash holdings induces a loss at the aggregate level. Indeed, if $X_1, \ldots, X_n$ and $\lambda_1, \ldots, \lambda_n$ respectively denote the risks and cash proportions maintained by divisions, $X$ and $\lambda$ respectively denote the risk and the cash proportion maintained at the aggregate level, and if $\omega_1, \ldots, \omega_n$ denote the proportions of funds invested in divisions (with respect to the total amount of funds invested by the conglomerate), then the following inequality holds:

$$E_\varphi [(X + \lambda)_-] \leq \sum_{i=1}^{n} \omega_i \cdot E_\varphi [(X_i + \lambda_i)_-]$$

Central managers are then obliged to minimise the sum of the insurance prices of the claims maintained by divisions, and *not* the insurance price of
the aggregate claim, as would be their first choice.\footnote{This issue is discussed in more details in Section 1.7 in this thesis.}

An optimal centralised allocation of capital is obtained in this way, which assigns the same cash proportions that central managers would choose if divisions were treated as stand-alone independent units, i.e. \( \lambda_i^* = T_{\varphi_i - X_i}(r) \), \( \forall i = 1, \ldots, n \). Within this context, choosing the stand-alone allocation implies that central managers recognise they cannot redistribute capital at their will inside the organisation.

However, information asymmetries and differences in the attitude toward risk among central and divisional managers may imply the aversion parameter \( \varphi \) of the central administration to be different from the aversion parameters \( \varphi_1, \ldots, \varphi_n \) of subsidiaries. Consequently, the cash proportions \( \lambda_i^* = T_{\varphi_i - X_i}(r) \), \( \forall i = 1, \ldots, n \) determined under the optimal centralised allocation may differ from the cash proportions \( \hat{\lambda}_1 = T_{\varphi_1 - X_1}(r), \ldots, \hat{\lambda}_n = T_{\varphi_n - X_n}(r) \) that divisional managers would choose if they were allowed to decide independently.

Although central administrations prefer to implement the optimal centralised allocation, it is also in their interest to measure the internal differences between the aversion parameters. For this purpose, an optimal decentralised mechanism is proposed in Chapter 1. First central managers determine the internal price of capital and let subsidiaries to choose their cash proportions. Divisions are only allowed to invest their reserves at a current account contracted with the central administration. By comparing the actual cash proportions chosen by subsidiaries with the amounts obtained when applying the liquidity principle of Equation 2, central managers can estimate the aversion parameters of divisional managers.\footnote{The centralised and decentralised mechanisms are respectively presented in Sections 1.7 and 1.8 in this thesis.}

The Monetary Equilibrium

Macroeconomics is the branch of economics that study aggregate variables — such as national output, unemployment and price (or inflation) levels, among others. There is agreement about what should be the major role of macroeconomic analysis, namely, to provide rules leading to price stability, low unemployment and rapid growth.
One of the fundamental relationships of macroeconomic analysis, already suggested by Keynes in his *General Theory* (1936, see the chapter devoted to the psychological and business incentives to liquidity; see also Keynes, 1937a, 1937b), is the one relating the total stock of money $M$ to the level of nominal output $Y$ and interest rates (see e.g. Equation (6) in Friedman, 1970):

$$M = Y \cdot \lambda(r) = P \cdot y \cdot \lambda(r)$$ (3)

where $P$ and $y$ respectively denote the level of prices and real output and $\lambda(\cdot)$ represents the preference for liquidity of the economy — i.e. the proportion of output that people prefer to maintain as cash holdings.

According to Equation 3, altering the money stock $M$ necessarily implies that at least some of the variables $P$, $y$ or $r$ must change until a new equilibrium is established. Therefore, if the rate of change of the level of prices (i.e. if the rate of inflation) and the rate of growth of real output were pegged to some predetermined levels, then the monetary authority would be always able to induce some preferred level of the interest rate by supplying the right amount of money to the economy.

The efficacy of the mechanism, however, depends on the flexibility of prices and the sensibility of the liquidity-preference function $\lambda(\cdot)$ with respect to the interest rate.

Indeed, if the liquidity-preference function is perfectly elastic with respect to the interest rate, then — even if prices are flexible — every variation in the amount of money will be completely absorbed by a change in $\lambda$, i.e. by a change in the amount of balances that people maintain in the economy. Liquidity-preference is said to be absolute in this situation (see e.g. Tobin, 1947, 1972). By contrast, if prices are flexible and the liquidity-preference function is perfectly inelastic, then any variation in the money stock must be followed by a change in the level of prices in the short-run and output adjustments in the long-run (Friedman, 1966, 1970, 1971).

Big controversy has arisen over this issue among economists, due to the consequences to the effectiveness of monetary policy in stimulating national output and reducing unemployment. Researchers and policy makers have been accordingly divided into two different schools: keynesians or supporters of fiscal interventions, and monetarists (see e.g. Modigliani, 1977, and Tobin, 1981, 1993).
On the one hand, keynesians, assuming that the preference for liquidity is absolute (or nearly so) claim that money plays no role in the determination of the monetary equilibrium and hence, that the only way of inducing the economy to attain full employment is to directly affect output by fiscal spending. This conclusion can be formally obtained in a context of general equilibrium with the help of the Hicksian IS-LM model (see Tobin, 1947, 1972 and 1982a, and also Romer, 1996, and Blanchard, 2005, for a presentation of the IS-LM model).

The variables that governments must take care of in this framework are national output and employment.

On the other hand, monetarists claim that monetary policy does indeed affect real output. In reaching this conclusion, they assume that the opposite hypotheses hold, namely, that prices are flexible and liquidity-preference is non-absolute — in such a way that variations in the stock of money are only partially absorbed by changes in the cash holdings demanded by the economy. Price adjustments are then expected to follow money stock variations in the short-run. As a consequence, production and spending can be encouraged in the short-run by increasing the amount of money in the economy. In the long-run, prices must return to their original levels, but at a higher level of real output.

For this mechanism to work efficiently, the growth rate of prices must be pegged to some fixed level. On these grounds, monetarists claim that the major concern of governments must be the rate of inflation (Friedman, 1968, 1970, 1971).

In consequence, a major issue behind the monetary controversy is the empirical assumption over the elasticity of the liquidity-preference function with respect to the interest rate.

Let us investigate how the monetary equilibrium is determined when the preference for liquidity of the economy is characterised by the optimal liquidity principle of Equation 2.

Notice in the first place that if the series of capital returns of national output is described by a Gaussian probability distribution with mean return $\mu$ and standard deviation $\sigma$ and people are neutral to risk (in such a way that the aversion parameter is equal to the identity function $\varphi(p) = 1$, $\forall p \in [0,1]$) then the liquidity principle explicitly depends on the risk-parameters:
\[ \lambda_{\mu,\sigma}(r) = \sigma \cdot \Phi^{-1}(1 - r) - \mu \] (4)

where \( \Phi \) denotes the cumulative probability distribution of a standard Gaussian random variable — whose mean return and standard deviation are respectively equal to zero and one.

Replacing Equation 4 into Equation 3 leads to the following alternative characterisation of the monetary equilibrium:

\[ M = Y \cdot \lambda_{\mu,\sigma}(r) = P y \cdot \lambda_{\mu,\sigma}(r) \] (5)

Within this setting, the monetary equilibrium not only depends on the level of prices, real output and interest rates, but also on the risk-parameters \( \mu \), \( \sigma \) describing the series of percentage returns of the nominal output. Accordingly, even if the rate of change of the level of prices is pegged to a fixed inflation target, the monetary authority cannot set the interest rate by simply controlling the money supply \( M \).

The conclusion is that neither fiscal nor monetary policy can be used alone to induce the economy to some predetermined equilibrium.

Additionally, since a theoretical expression is available for the liquidity principle in Equation 4, a more precise description of the evolution of the monetary equilibrium can be given. Particularly, the following expression is obtained for the semi-elasticity of the demand for cash holdings with respect to the interest rate:

\[ \eta\left(r, \frac{\mu}{\sigma}\right) = \frac{1}{\lambda_{\mu,\sigma}(r)} \frac{d\lambda_{\mu,\sigma}(r)}{dr} = \frac{-\sqrt{2\pi}}{\Phi^{-1}(1 - r) - \frac{\mu}{\sigma}} \exp\left(\frac{\left[\Phi^{-1}(1 - r)\right]^2}{2}\right) \] (6)

Hence the semi-elasticity function may well be equal to infinite in some cases, or in other words, the preference for liquidity of the economy may well become absolute under certain circumstances.\(^{11}\)

\(^{11}\)Actually \( \eta \to -\infty \) when \( \Phi^{-1}(1 - r) \to \mu/\sigma \). The states of the economy when this condition is satisfied are regarded as critical states in Section 2.6 in this thesis. In Section 2.8, the consequences of describing the preference for liquidity by the Gaussian liquidity principle are presented in more details.
Consequently, although the semi-elasticity function does indeed remain stable over a broad set of combinations of the variables $M$, $P$, $y$, $r$, $\mu$ and $\sigma$, which are involved in the alternative description of the monetary equilibrium presented in Equation 5, there are certain states of the economy when this function may converge to infinite.

In other words, although the economy can be well represented by the monetarist paradigm within a wide class of states, it can also evolve to states where the Keynesian fears are confirmed — and liquidity-preference becomes absolute. Therefore, even when monetary policy can sometimes effectively stimulate national output, it can become suddenly ineffective if certain paths are followed by the economy.

**Efficient Markets and Liquidity Crises**

A highly stylised model for the description of the flows of capital produced in some market of (interbank) loans is presented in Chapter 3.

In the model, alternative versions to Equations 3 and 5 are proposed, where the money stock $M$ is interpreted as the total supply of credit to the market. The total demand for balances, on the other hand, is expressed as a proportion $\lambda(r)$ of the total amount of funds $L$ invested on risk — or more precisely, invested on the aggregate or market portfolio containing the whole of securities maintained by firms and private investors — where $r$ denotes the opportunity cost of capital, i.e. the (average) return that borrowers are willing to pay.

The equilibrium in the market of loans is attained when aggregate outflows and inflows of capital are the same. The following relationship thereby describes the market equilibrium when the series of capital returns of the market portfolio follows a Gaussian probability distribution:

\[
M = L \cdot \lambda_{\mu,\sigma}(r)
\]  

Within this context, the discount factor $\lambda_{\mu,\sigma}(r)$ explicitly represents the rate at which a unit of investment in the market portfolio is exchange by a unit of capital, i.e. it represents the market price of risk. Accordingly, the interest rate $r$ represents the return accrued by a unit of capital invested on
the market portfolio.\footnote{Such rate is called the \textit{internal rate of return on risk} in Section 3.3. As explained in Section 3.7, in the Gaussian case a \textit{risk-structure} of interest rates is defined over the plane of mean returns and standard deviations.}

In \textit{Equation 7}, variations in the credit supply $M$ and the amount of funds $L$ spent on securities produce two kind of adjustments: adjustments in the price of capital $r$, and adjustments in the market price of risk that affect the risk-parameters $\mu$ and $\sigma$. In other words, the equilibrium in the market of balances simultaneously determines the equilibrium in two markets, the market of capital and the market of financial securities.

Two major consequences of the model must be emphasised. In the first place, notice from \textit{Equation 6} that in Gaussian markets the semi-elasticity function $\eta$ takes positive values under some combinations of the interest rate and the risk-parameters. This means that under certain circumstances individuals prefer to lend all their balances — and do not maintain cash reserves at all.

More precisely, we can state that during bearish trends, i.e. when the mean return of the market portfolio is lower than zero, people always demand cash holdings, and additionally, that the proportion of funds allocated to it (i.e. the preference for liquidity of the market) always increases with the magnitude of the mean capital loss. By contrast, during bullish trends, i.e. when the mean return of the market portfolio is greater than zero, people maintain reserves only in certain scenarios, specifically, when the condition $\Phi^{-1}(1-r) - \mu/\sigma > 0$ is satisfied. If instead $\Phi^{-1}(1-r) - \mu/\sigma \leq 0$, individuals prefer to exclusively rely on capital markets.

The second major consequence is related to the fact that the magnitude of the semi-elasticity of the demand for balances is equal to infinite in those states when $\Phi^{-1}(1-r) - \mu/\sigma = 0$. People are willing to substitute all their risky assets by cash holdings under such circumstances. These can be regarded as critical states of capital markets, which can be corresponded to episodes of liquidity crises. Most notably, such critical states can be produced in the middle of a bullish trend.

Although the possibility of the spontaneous appearance of liquidity crises in the model confirms a well known and documented fact, it contradicts one fundamental paradigm that has determined the economic policies of many countries during the last quarter of century, namely, the efficient market hypothesis.
Recall that the efficient market hypothesis, developed by Eugene Fama (1970, 1998), states that the prices at which securities are actually traded reflect all the available (and relevant) information. In other words, it claims that financial securities are always transacted at a fair price. As a consequence, it is impossible to beat or outperform markets, and hence, in particular, every kind of regulations and trading restrictions can only induce markets to inefficiently allocate resources.

Based on the efficient market hypothesis, many scholars have convinced themselves that if liquidity crises are observed, they still correspond to the most efficient state. Others have claimed that liquidity crises provide evidence that people do not behave rationally and hence, that deregulated markets do not always arrive to the most efficient equilibrium. This assumption is especially appealing, for according to another major economic principle, market prices must reflect economic fundamentals — and it is difficult to accept that sudden contractions of the credit supply are the reflection of economic fundamentals.

From the model of equilibrium described by Equation 7, a different conclusion is obtained. Indeed, notice in the first place that every state of the market, including those states related to liquidity crisis — when the semi-elasticity function $\eta$ is equal to infinite — represents the decisions taken by rational individuals, who pursue strategies that minimise the total cost of guaranteeing their underlying security portfolios (as stated in Equation 1). Within this context, liquidity crises provide no evidence of irrationality.

However, the fact that Equation 7 allows multiple combinations of transacted capital flows and interest rates, which are determined by the risk-parameters $\mu, \sigma$, implies that rationality and deregulation are not sufficient conditions to ensure that markets behave efficiently. Within this theoretical setting, governments and regulatory authorities must induce markets (through persuasion and mandatory statements) to attain those states that are compatible with some predetermined level of economic performance.

We thus arrive to one of the main consequences of the model of equilibrium described by Equation 7, namely, that the deregulation of financial markets is not necessarily compatible with financial stability and sustainable economic growth.
Chapter 1

The Allocation of Economic Capital in Opaque Financial Conglomerates

Financial intermediaries establish contractual liabilities with customers to attract funds that spend on financial securities and investment projects. They also maintain cash provisions, in the form of capital, in order to avoid negative balances and to take advantage of future investment opportunities.

Several types of capital are distinguished in the literature, depending on the purpose it serves to the institution and on the criterion employed to fix its level.

The cash capital, for example, represents a balance required to execute transactions, whereas the working capital additionally includes operational expenses (see e.g. Williams et al., 2002, and also Howells and Bain, 2005). On the other hand, the term regulatory capital is defined according to accounting standards (as in Basel, 2004), while equity is corresponded to the portion of reserves that is supplied by shareholders. Finally, many authors speak of economic and risk capital to refer to a surplus defined according to criteria based on economic or statistical considerations.

In lines with Merton (1974, 1977), the demand for capital will be corresponded to a demand for deposit insurance in this chapter.\(^1\) Accordingly,

\(^1\)The meaning and scope of this interpretation will be clarified later in Sections 1.1, 1.2 and 1.4.
the terms *economic* and *risk* capital will be indistinctly used to refer to the smallest amount required to insure the value of net assets against a loss in value relative to the risk-free investment of those net assets. In this context, the difference between *economic* and *equity* capitals represents a balance that is captured by managers in attention to some solvency requirement.

Three main components of the *capital structure* can thus be distinguished: a net liability contracted with *customers*, an amount of equity supplied by shareholders and a cash balance determined by managers. Since the economic capital is equal to the sum of the last two components, the problem of capital allocation can be roughly corresponded to the determination of the proportions of the portfolio of assets that are funded by means of *internal* and *external* debt.

Holding capital imposes an *opportunity* cost on firms because these funds could be alternatively employed on profitable investments. Such costs induce financial institutions to prefer (external) debt and accordingly demand *as less capital as possible*. In fact, in a seminal paper, Modigliani and Miller (1958) claim that, as long as borrowing and lending can be carried out at any moment and without quantity restrictions at a single interest rate, firms can adjust their balance sheets whenever is needed and hence cash provisions impose a cost without any benefit. Then *rational* decision-makers (who maximise value) should demand no capital at all.

More specifically, Merton (1997) states that the presence of *credit-sensitive* customers obliges *opaque* institutions (whose investment activities are not fully observed by outsiders) to rely on a third-party guarantor, who agrees to honour the outstanding liabilities when bankruptcy is declared.\(^2\) Then the market values of equity and debt can be respectively corresponded to the values of a *call* and a *put* option on the value of assets, with exercise price equal to the value of debt, which implies that the market value of the firm (or the market value of the assets’ portfolio) is independent of the capital structure, as predicted by the MM-proposition (see also Miller, 1998). A fundamental assumption for this mechanism to work is that capital and financial securities are *continuously* traded in competitive markets.

The correspondence of capital to deposit insurance implies that the hypothesis of *perfect hedging* can be reestablished by imposing that every single claim can be insured at a *unique* price. Under such circumstances, managers are indifferent between hedging and insurance and are certainly indifferent

---

\(^2\)The presence of *credit-sensitive* customers increases external controls and monitoring due to the *moral-hazard* implicit in the administration of deposits. See also Ross (1989).
about the amount of economic capital. However, hedging and insurance cannot be regarded as equivalent tools in practice. As a matter of fact, although competitive forces lead capital and securities to be transacted at a unique price, these tend to be weak in insurance markets, where non-standardised policies are transacted (see e.g. Venter, 1991, and Goovaerts et al., 2005).

Therefore, liquidity restrictions may arise from two different sources. In the first place, firms are not always able to trade continuously in capital and security markets. Then the moral-hazard arising due to the opacity of financial intermediaries induces the appearance of premiums over the market cost of capital, which should be established on an actuarial basis. Secondly, the buyers and sellers of insurance can maintain different perceptions about risks and can accordingly assign different prices to their corresponding guarantees.

Recent literature has dealt with these issues when analysing the capital structure. More precisely, capital decisions have been connected to the cost of external financing and hence, to the degree of financial flexibility enjoyed by firms — representing the firms’ accessibility to funding rearrangement at low cost.

Gamba and Triantis (2008), for example, argue that the value of financial flexibility actually depends on a series of variables, including the external cost of capital, the levels of personal and corporate tax rates, the firms’ maturity and growth potential, and the reversibility of capital. They are then able to show that firms can compensate for lower degrees of financial flexibility by raising their levels of reserves, an attitude that is reinforced in the presence of liquidity restrictions. However, firms can also increase their exposition to external debt if cash holdings are not favoured by the tax structure — as it is usually the case in practice.\footnote{This conclusion is connected to the original discussion of Modigliani and Miller on the subject, see e.g. Modigliani and Miller, 1963, and Miller, 1977. See also Hennessy and Whited, 2005 and 2007.}

One important determinant of financial decisions is not considered, however, in the model of Gamba and Triantis: the presence of agency costs in the relationship of managers and stockholders, which may lead to the misallocation of funds when managers dispose of free-cash-flow.\footnote{As stated by Jensen (1986), internal monitoring is likely to become more intense in firms that own high amounts of free-cash-flow.}

Recently, Kalcheva and Lins (2007) have shown that there is statistical...
Chapter 1: Economic Capital in Opaque Conglomerates

evidence suggesting that when shareholder protection is weak (and corporate agency costs are thus severe), increasing the level of cash holdings reduce the value of firms. Only when shareholder protection is strong (and agency costs are not severe), the value of firms is not related to cash holdings.

Within this context, the capital structure reflects a compromise between the internal and external costs endured when contracting debt, respectively determined by agency costs (in the relationship with shareholders) and moral-hazard (in the relationship with customers). In other words, firms must choose between reducing the amount of reserves in order to reduce internal controls and equity costs, but facing at the same time increasing external controls and credit costs, or raising the amount of reserves in order to face lower credit costs but higher equity costs.\(^5\)

Other models where agency costs and financial flexibility are shown to be determinants of the capital structure are those of Billet and Garfinkel, 2004, and Faulkender and Wang, 2006.

Billet and Garfinkel (2004) consider a model where firms raise funds from two different markets: a market of uninsured claims, where a risk premium is paid, determined by creditors, and a market of insured claims, where an insurance premium is paid, imposed by regulatory requirements. When both premiums differ, firms must show preference for the cheaper over the more expensive financing alternative. The extent to which such differences can be exploited, however, depends on a number of factors related to the costs incurred when accessing the corresponding markets — i.e. it depends on the degree of financial flexibility. When there is uncertainty regarding future premiums and interest rates, firms are also induced to demand cash holdings — or financial slack, as called by Billet and Garfinkel — as a means of a buffer against unexpected losses.

Faulkender and Wang (2006), on the other hand, examine the relation between market returns and the market value of cash holdings. They find that those firms that hold cash with higher valuations obtain higher capital returns — indicating the presence of frictions that may raise the cost of external financing. But they also find that the marginal benefit of cash

\(^5\)The level of the external cost of capital thereby represents the credit quality of borrowers, as specified by specialised institutions. Kisgen (2006) investigates the effect of credit ratings over capital decisions. In his analysis, he stresses the fact that credit ratings are actually made of a discrete number of categories, an issue that will be assigned a major importance in Section 1.3. The existence of discrete costs around rating changes leads Kisgen to conclude that firms close to a rating upgrade (or downgrade) tend to issue less debt — and to maintain more cash holdings.
holdings diminishes with the level of cash, which suggests the existence of some optimal level of reserves (see also Strebulaev, 2007).

The model of the capital structure that will be soon introduced proposes to measure the cost of equity by the expected return obtained in excess of the level of reserves. The expected value of the excess of loss over the level of reserves, on the other hand, is regarded as a measure of the cost of bankruptcy of the firm.

The costs of equity and bankruptcy, as defined in this chapter, can be thereby corresponded to the risk and insurance premiums determining the prices of uninsured and insured claims in the model of Billet and Garfinkel. More generally, they can be respectively associated with the cost of internal and external financing — and hence, they can be respectively connected to agency costs and moral-hazard.

It is possible to prove that an optimal capital structure exists under such circumstances. The optimal cash balance can be actually determined in order to maximise the market value of the firm, defined as the difference between the actuarial prices of equity and the default claim, plus the return offered by a non-risky bond.

A precise description of the conditions under which the capital structure matters is provided in this way, and the limits of the Modigliani and Miller invariance propositions are then clearly stated. This is the main contribution of this chapter.6

Within a multi-business environment, differences in expectations between central and divisional managers lead to discrepancies about the optimal levels maintained by subsidiaries. Inefficiencies then arise in the form of under- and over-investment, respectively corresponded to the cases when too much and too few capital is demanded. This is a most important issue in corporate finance, for in the former case favourable investment opportunities can be lost, while in the later case firms can damage their credit quality (as perceived by customers) if too much risk is assumed.

Since the optimal capital principle proposed in this chapter explicitly depends on risk and expectations, it proves to be a convenient tool to describe such interactions inside corporations. In fact, two optimal allocation principles can be defined on this basis, respectively related to centralised and

---

6The first part of the chapter is devoted to the optimal capital structure — Sections 1.1, 1.2, 1.3 and 1.4.
Chapter 1: Economic Capital in Opaque Conglomerates

decentralised administrations. This is the second main contribution of this chapter.\(^7\)

At the empirical level, the capital and allocation principles derived from the extended setting proposed in this chapter show important advantages over other principles found in the literature.\(^8\)

In the first place, they are founded on economic fundamentals. Secondly, since they are expressed in terms of the quantile function of underlying risk, they can be applied to any kind of probability distributions and hence, they are suitable both to finance and insurance applications. The quantile function is actually well-known by researchers and practitioners and it has been recommended by the Basel Committee on Banking Supervision (2004).

Finally, again at the theoretical level, the allocation rule proposed by Merton and Perold (1993) — which is based on the market price of deposit insurance with unrestricted hedging — can be obtained as a particular case of the optimal capital principle (Section 1.9). The most relevant facts of the classic theory of capital are extended in this way.

1.1 Agency Costs and Moral-Hazard in Opaque Financial Institutions

In the model of Modigliani and Miller (1958), the set of financial securities is divided into equivalent classes, in such a way that the returns offered by any two assets belonging to a certain class are proportional to each other — or equivalently, in such a way that every two assets belonging to the same class are perfect substitutes for each other.

Firms fund their assets’ portfolios with own capital \((K)\) or issue bonds offering a constant yield per unit of time (delivering some cash flow \(D\) at maturity). Then the market price of portfolios containing assets belonging to the same class can at most differ in some scaling factor \(\alpha\):

\[
V = A = K + D = \alpha \cdot r
\]

\(^7\)The optimal allocation principles are presented in Sections 1.5, 1.6, 1.7 and 1.8.

\(^8\)Albrecht, 2004, provides a concise survey on a broad family of capital principles currently used by financial institutions.
Both assets and bonds are supposed to be traded in *perfect* markets, which means that both the transactions of securities and bonds can be performed at any moment and without restrictions. Therefore, the proportions of capital and debt can be modified at any moment without affecting the value of the firm (which is always equal to $A = \alpha \cdot r$) and hence the market valori-sations of firms are always independent of the underlying capital structure.

In subsequent papers, Modigliani and Miller prove that the irrelevance of the capital structure is maintained in the presence of taxes and dividends delivered to shareholders (Modigliani and Miller, 1961). The critical assumption supporting these results is that managers and stockholders share expectations about the future payoffs of the net assets’ portfolio. Informa-tional asymmetries between managers and stockholders lead to the appearance of *agency costs* inside institutions. The cost of equity might depend on the level of capital under such circumstances.

Informational asymmetries are also present in the relationship between managers and customers. Indeed, as pointed out by Merton (1997), financial companies tend to be *opaque* institutions, for their investment activities are not fully observed by outsiders. As a consequence, it is difficult for customers to assess the risk assumed by intermediaries and for the later to effectively reflect their solvency state.

On the other hand, since the benefits of financial intermediation (such as economies of scale and reduced transaction costs) can be effectively trans-mitted to customers only if intermediaries can convincingly assure that their liabilities are free of default, both the availability and the price of capital are dependent on the perceptions of customers regarding the credit quality of the borrowing institutions. Such services are said to be *credit-sensitive* by Merton. In other words, because of the *moral-hazard* implicit in the relation-ship between managers and customers, firms are obliged to fix their capital structures in order to increase their *honourability* at the eyes of investors (see also Cummins and Sommer, 1996, and Myers and Read, 2001).

On these grounds, Merton states that the practice of financial interme-diation requires a third-party *guarantor* (whose willingness and capability to meet obligations are beyond question) to assure the safety of deposits. This role might be assumed by another financial company, as well as by some governmental division. The terms of the contract are the following. When bankruptcy is declared (i.e. when $X = A - D < 0$), the firm de-

---

9In fact, tax regimes tend to favour the use of debt, the only exception being some regimes of double taxation. See also Modigliani and Miller, 1963, and Miller, 1977.
faults its assets to the guarantor and the guarantor meets the bondholders’ claims. In this case, nothing is left to shareholders. A second contract is simultaneously established with shareholders, promising to pay the value of net assets when this balance is positive (i.e. when $X = A - D > 0$) in exchange of a certain amount of equity capital. Consequently, shareholders and bondholders respectively receive the payments $\max(0, A - D)$ and $D$ at the maturity date, while the guarantor has to afford the cost $\min(0, A - D)$.

In other words, while shareholders own the right to buy the portfolio of assets at the price $D$ at maturity, the firm owns the right to sell the portfolio of assets to the guarantor at the price $D$. Equivalently, we can say that shareholders are the owners of a European call option $C(A, D)$ on the value of assets with exercise price equal to the value of debt, while managers are the owners of a European put option $P(A, D)$ on the value of assets with exercise price equal to the value of debt. Bondholders, on the other hand, are the owners of the sure stream $D$.

Information asymmetries thereby impose on managers the obligation of combining three kind of securities: the firm itself (or the portfolio of assets held by the firm), a promise to pay at any event its outstanding liabilities, and a particular security guaranteeing the payment to creditors when bankruptcy is declared. Such a portfolio can be built by simultaneously hiring debt in the form of a zero coupon bond, selling a call option to shareholders and buying a put option to the guarantor (Merton, 1974 and 1977).

In fact, since the value of the call option $C(A, D)$ corresponds to a cash flow payed by shareholders in exchange of the surplus accrued by the portfolio of net assets, it can then be regarded as the market value of equity. In a similar way, the term $D \cdot e^{-rT} - P(A, D)$ can be regarded as the market value of the guaranty, for it represents the net stream received by the firm from the guarantor. From the Put-Call parity theorem (see Black and Scholes, 1973, Merton, 1974, and also Cummins and Sommer, 1996) we obtain that:

$$V = A = C(A, D) + D \cdot e^{-rT} - P(A, D)$$

(1.1)

Therefore, though both the price of equity and the price of the guarantee depend on the underlying capital structure, the market price of the firm is always equal to the value of assets, whatever the chosen funding strategy (as predicted by the MM-proposition, see Miller, 1998).
Consequently, even in the presence of informational asymmetries, the market valuation of financial institutions is not affected by their capital structures, as long as trading in assets and bonds takes place continuously in time.

However, this cannot be always said of real markets. Besides, if firms cannot hedge fully, hedging cannot be regarded as a perfect substitute for insurance. This means, specifically, that deposit insurance cannot be corresponded to an option claim and that Equation 1.1 cannot be invoked to support the MM-proposition. Instead, an actuarial approach should be introduced to determine the fair price of the claims \((A-D)_+ := \max(0, A-D)\) and \((A-D)_- := -\min(0, A-D)\) respectively held and afforded by shareholders and guarantors at the end of the investment period. This will be formally accomplished in Sections 1.2, 1.3 and 1.4.

### 1.2 The Optimal Capital Structure

Both asset and liability claims will be regarded as random variables in the following, while the market value of net assets (equal to the market value of the portfolio of assets minus liabilities) will be expressed as the product of the level of investment \(I\) and some random perturbation \(X\):

\[
A - D = I \cdot X \iff X := \frac{A - D}{I}
\]  

Then the level of investment can be regarded as the principal of the net portfolio. The level of capital will be also represented as a proportion of the level of investment:

\[
K = I \cdot k \iff k := \frac{K}{I}
\]  

The ratio \(k\) represents a capital-to-investment or a cash-to-risk ratio. It determines the proportion of internal financing of the firm. In this context, the determination of the capital structure involves choosing the proportions of internal and external financing when building solvent or equilibrated portfolios (which satisfy \(A - D > K\)).

Let us analyse the payments received at the maturity date by some firm
that maintains insurance and equity contracts in the terms stated in the previous section.

Recall, in the first place, that one agreement is celebrated between the firm and some guarantor, which obliges the later to honour the total capital loss $I \cdot (X + k)_- = I \cdot \min(0, X + k)$ in exchange of a certain premium paid by the former at the beginning of the investment period. Simultaneously, shareholders pay a certain price to managers at the beginning of the investment period, in exchange of receiving the random capital profit $I \cdot (X - k)_+ = I \cdot \max(0, X - k)$ at the end. The capital $K$ is invested in a banking account to obtain the risk-free return $r_0$ (in other words, it is converted to a risk-free zero-coupon bond with internal return $r_0$). The payments received by the firm at maturity are then given by:

$$
\begin{align*}
I \cdot (X - k) + r_0 \cdot K & \quad \text{if } X \geq k \\
I \cdot (X + k) + r_0 \cdot K & \quad \text{if } X \leq -k \\
0 & \quad \text{if } -k < X < k
\end{align*}
$$

Accordingly, when $X \geq k$ the firm can afford its debt and pay a surplus to shareholders, i.e. the firm is solvent in this case. By contrast, when $X \leq -k$ the capital $K$ is deliver to the guarantor who has to afford the residual loss $I \cdot (X + k)$. Shareholders receive nothing in this case. Finally, when $-k < X < k$ the firm cannot return the total amount of capital to shareholders, although the total debt attracted from customers can be honoured and the guaranty is not invoked. Stockholders might decide to sell their shares or to call for portfolio restructuring under such circumstances.

We have already pointed out that risk can be completely suppressed through hedging if cash and securities can be traded to any desired extent in the market. Indeed, under such circumstances, the prices (per unit of investment) of the contracts established with shareholders and guarantors are respectively given by the prices of a call and a put option on the value of the random capital return $X$ with exercise price equal to the cash-to-risk ratio $k$, in such a way that the put-call parity can be invoked to obtain (as in Equation 1.1):

$$
X = C(X, k) + k \cdot e^{-r_0 T} - P(X, k)
$$

where $T$ denotes the time to maturity. Thus, the value of the firm does not depend on the cash-to-risk ratio $k$ as long as continuos rebalancing of portfolios is allowed.
1.2 The Optimal Capital Structure

When continuous rebalancing is not possible due to liquidity restrictions, the prices of the contracts established with stockholders and guarantors should be determined on an actuarial basis. Accordingly, the price of equity and the cost of bankruptcy will be respectively corresponded to the terms \( E[(X - k)_{+}] \) and \( E[(X + k)_{-}] \) in the following, for these terms represent the fair or actuarial prices of the underlying exposures (see Goovaerts et al., 1984). Hence the value of the firm at the end of the investment period (as perceived by managers) is given by:

\[
V = (E[(X - k)_{+}] - E[(X + k)_{-}]) \cdot I + r_0 \cdot K
\]

Within this context, the cash-to-risk ratio \( k \) affects the net return on investment and hence the value of the firm.

Given any fixed level of investment \( I \), every rational manager must then choose the capital structure that maximises the firm’s value per unit of investment \( V/I \):

\[
\max_k E[(X - k)_{+}] - E[(X + k)_{-}] + r_0 \cdot k \tag{1.5}
\]

The solution to this maximisation problem can be determined by applying Lagrange optimisation. The first-order condition actually leads to:

\[
\frac{d}{dk} E[(X - k^*)_{+}] - \frac{d}{dk} E[(X + k^*)_{-}] + r_0 = 0
\]

On the other hand, the mathematical expectations of the corresponding excess return terms are defined as:

\[
E[(X - k)_{+}] = \int_k^\infty (x - k) \cdot dF_X(x) = \int_k^\infty (x - k) \cdot f_X(x)dx
\]

\[
E[(X + k)_{-}] = - \int_{-\infty}^{-k} (x + k) \cdot dF_X(x) = - \int_{-\infty}^{-k} (x + k) \cdot f_X(x)dx
\tag{1.6}
\]

\[\text{A similar approach is proposed by Froot et al., 1993. In this model, the fundamental motive for the existence of an optimal balance is that output variability is undesirable when investment presents diminishing marginal returns — i.e. when output is expressed as a concave function of the level of investment — because funding and investment plans can be affected in a costly way under such circumstances. See also Froot and Stein, 1998, and Froot, 2007.}\]
Chapter 1: Economic Capital in Opaque Conglomerates

where $F_X$ and $f_X$ respectively denote the cumulative and the density probability functions of the random variable $X$, which are defined as (see, for example, De Finetti, 1975):

$$F_X(x) = P\{X \leq x\} = \int_{-\infty}^{x} dF_X(x) \quad \text{with} \quad f_X(x) = \frac{dF_X}{dx}(x) > 0 \quad \forall \ x$$

Hence every random variable is uniquely determined by its corresponding probability distribution $F_X$. Besides, an equivalent characterisation is provided by the cumulative or tail (also known as survival) probability distribution $T_X(x) = 1 - F_X(x)$, $\forall \ x$, defined as:

$$T_X(x) = P\{X > x\} = \int_{x}^{\infty} dF_X(x) \quad \text{with} \quad \frac{dT_X}{dx}(x) < 0 \quad \forall \ x$$

As the derivative of the expected excess return requires to derive an integral operator with respect to a variable affecting its limits of integration, the Leibnitz’s rule can be applied:

$$\frac{\partial}{\partial z} \int_{u(z)}^{v(z)} h(z, x) dx = \int_{u(z)}^{v(z)} \frac{\partial h(z, x)}{\partial z} dx + h(z, v(z)) \cdot v'(z) - h(z, u(z)) \cdot u'(z)$$

Accordingly,

$$\frac{dE[(X-k)^+]}{dk} = -P\{X > k\} = -T_X(k)$$

$$\implies \frac{d^2E[(X-k)^+]}{dk^2} = -\frac{dT_X(k)}{dk}$$

(1.8)

and also, since $P\{X \leq -k\} = P\{-X < k\}$:

$$\frac{dE[(-X+k)^-]}{dk} = -P\{-X > k\} = -T_{-X}(k)$$

$$\implies \frac{d^2E[(-X+k)^-]}{dk^2} = -\frac{dT_{-X}(k)}{dk}$$

(1.9)
Therefore, the first-order condition leads to the following equality in terms of the negative and positive tail probability functions:

\[ T_{-X}(k^*) + r_0 = T_X(k^*) \quad (1.10) \]

or equivalently, in terms of the cumulative probability functions:

\[ 1 - F_{-X}(k^*) + r_0 = 1 - F_X(k^*) \]

Since the term \( E[(X - k)_{+}] \) represents the expected excess over capital, the term \( T_X(k) = -dE[(X - k)_{+}]/dk \) corresponds to the magnitude of the reduction in free-cash-flow produced when adding an additional unit of cash to the stock of capital — instead of investing it in the portfolio of assets. Similarly, the terms \( E[(X + k)_{-}] \) and \( T_{-X}(k) = -dE[(X + k)_{-}]/dk \) respectively represent the cost of bankruptcy and the marginal gain in value obtained due to the reduction in the cost of bankruptcy (and hence in the price of the guarantee) when an additional unit of investment is added to the stock of cash (see Equations 1.8 and 1.9).

Therefore, according to Equation 1.10, the optimal level of capital is determined at the point where the marginal gain equals the marginal loss in value due to allocating one additional unit of investment to the stock of reserves instead of spending it on assets. In other words, capital is demanded up to the point where the marginal return on capital (to the left-hand side of Equation 1.10) is equal to the marginal return on investment (to the right-hand side of Equation 1.10).

The existence of a solution to the optimisation problem of Equation 1.5 can be mathematically assured as long as the second derivative of the objective function is lower than zero, i.e. as long as (from Equations 1.8 and 1.9):

\[
\frac{d^2 E[(X - k^*)_{+}]}{dk^2} - \frac{d^2 E[(X + k^*)_{-}]}{dk^2} < 0 \iff \frac{dT_{-X}(k^*)}{dk} < \frac{dT_X(k^*)}{dk}
\]

and since \( T_X = 1 - F_X \), and also \( dF_X(x)/dx = f_X(x) = P\{X = x\} \forall x \), we finally obtain that the second-order condition can be expressed in terms of the mass probability densities:
Chapter 1: Economic Capital in Opaque Conglomerates

\[ P\{X = k^*\} < P\{X = -k^*\} \] (1.11)

We thus arrive to the (reasonable) conclusion that liquidity provisions provide a benefit to firms only when the probability of attaining a capital loss of a certain magnitude is greater than the probability of obtaining a capital gain of the same magnitude.

In particular, symmetric probability distributions (around the expected value) satisfy:

\[ P\{X - E[X] = x\} = P\{X - E[X] = -x\} \quad \forall x \]

Therefore, if \(E[X] < 0\) and if the density probability function \(f_X(x) = P\{X = x\}\) decreases with the absolute value \(|x|\) of the argument, then the following inequality holds:

\[ P\{X = -k\} = P\{X = E[X] - x\} = P\{X = E[X] + x\} > P\{X = k\} \]

\[ \forall k = x - E[X] \]

Hence, within the class of symmetric probability distributions, the objective function of the maximisation problem is concave when \(E[X] < 0\), which means that the value of the firm attains a maximum value at the point where the first order condition (stated in Equation 1.10) is satisfied. The existence of an optimum capital structure can then be mathematically assured in this case.

Similarly, if \(E[X] > 0\) and if the density probability function \(f_X(x) = P\{X = x\}\) decreases with the absolute value \(|x|\) of the argument, then:

\[ P\{X = -k\} < P\{X = E[X] - x\} = P\{X = E[X] + x\} = P\{X = k\} \]

\[ \forall k = x + E[X] \]

Hence, within the class of symmetric probability distributions, the objective function of the maximisation problem is convex when \(E[X] > 0\). This means that in this case the level of capital where the first order condition is satisfied actually implies that the value of the firm is minimised.
As a conclusion, we obtain that only when $E[X] < 0$ capital is beneficial to financial institutions, a result that can be interpreted as evidence that when $E[X] > 0$ firms do not fully internalise the cost of bankruptcy. This situation can be actually produced in markets where lenders cannot precisely state the risk assumed by their counterparts due to opacity and moral-hazard.

1.3 Liquidity Restrictions in Capital Markets

As demonstrated in the previous section, the capital structure may well affect the market value of financial institutions in the presence of liquidity restrictions.  

Liquidity restrictions arise, in the first place, because the portion of capital provided by stockholders is determined in a regular frequency (such as yearly, quarterly or monthly) and cannot be modified until the end of the period. Since the level of equity and the frequency of revisions are the result of negotiations between managers and shareholders, changing some agreement is costly and can reduce the market valorisation of the firm.

On the other hand, choosing some risk-based capital principle (such as the one defined by Equations 1.5 and 1.10) implies that the amount of economic capital must be subject to constant revisions if the riskiness of the series of the value of the net-assets’ portfolio is varying — i.e. if the series of capital P&L of the net-assests’ portfolio is non-stationary. This means that managers are obliged to rely on some market of cash balances (or inter-bank loans) in order to maintain a total level of capital that is consistent with the underlying exposure.

Albeit preferring external debt reduces the controls imposed by shareholders, this strategy also raises the costs associated to moral-hazard (on the side of customers) and bankruptcy.\footnote{This is especially true in highly leveraged firms, where managers have strong incentives to take risk. See Jensen, 1986. The role of bankruptcy costs in the determination of the capital structure has been already mentioned by Stiglitz, 1972 (see also Stiglitz, 1988).} The controls and monitoring established by customers and regulators are expected to increase in this situation. Consequently, when deciding their capital structures, firms have to face a trade off between paying high spreads because of opaqueness and signaling costs on the one hand, and sacrificing potential competitive advantages when maintaining idle balances on the other.\footnote{In the words of Stephen Ross (1989), firms and institutions are monitored and con-}
tion 1.10, the optimal cash-to-risk ratio $k^*$ must be determined at the point where the opportunity cost of capital, equal to the reduction of the excess of return $T_X(k^*) = -dE\left[(X - k^*)_+\right]/dk$, just offsets its marginal benefit, equal to the reduction in the cost of bankruptcy plus the risk-free interest rate, i.e. $T_X(k^*) + r_0 = -dE\left[(X + k^*)_-\right]/dk + r_0$.

Another kind of liquidity restrictions arises due to the fact that the opportunity cost of capital, that is perceived by managers as the reduction in the price of equity induced when certain level of capital is maintained, is not necessarily equal to the return $r$ they have to pay to borrow in the market of cash balances (i.e. in general $T_X(k^*) \neq r$).

Indeed, while the cost of equity reflects the agency costs between managers and stockholders, the market capital cost $r$ is determined according to the capacity and willingness to pay of the borrower and it then reflects the moral-hazard in the relationship with customers. It depends, in particular, on the capital structure of the borrower institution, i.e. (as it is normally established in the corporate finance literature, see Williams et al. 2002) on the leverage ratio of the borrower institution, which is normally defined as $K/D$ and also $D/I = (I - K)/I$, or equivalently, on its cash-to-risk ratio $k = K/I$.

As a matter of fact, to determine the price of loans, the creditors of opaque organisations rely on their own research and monitoring, as well as on the information published in the media and the risk categorisations provided by specialised (private and governmental) institutions.

The credit ratings observed in practice normally include a finite number of categories. Within each class, every firm is supposed to face the same risk of default and hence every firm is allowed to borrow at the same interest rate (as in the MM-proposition), in such a way that the more the concerns of creditors about the credit capacity of firms in a certain class, the higher the level of the corresponding cost of capital and vice-versa. This means that lenders cannot discriminate perfectly and that borrowers can remain in the same class as long as they do not drastically modify their capital structures.

In other words, as long as firms do not drastically vary their cash-to-risk ratios, they can regard the market capital cost as a constant.

\[\text{trolled through a complex set of implicit and explicit contractual relations. See also Fama, 1980.}\]
1.3 Liquidity Restrictions in Capital Markets

In order to explicitly introduce the cost of capital \( r \) in the model, let us consider a firm that belongs to a certain class determined by the capital cost \( r \) and maintains the cash-to-risk ratio \( k \). The cost of equity of this kind of firms is given by:

\[
E[(X - k)_+] = E[X_+] - r \cdot k \iff \frac{E[X_+] - E[(X - k)_+]}{k} = r \quad (1.12)
\]

Jensen (1986) has noticed that internal monitoring is more intense when positive balances are obtained at the end of the investment period and cash is at disposal in excess of what is required to fund every ongoing (solvent) investment project. In this case, it is said that a firm owns free-cash-flow. Accordingly, although the agency costs inside financial firms can be certainly reduced by diminishing the amount of equity, there is also a pressure to raise these costs since higher amounts of free-cash-flow are obtained in this case. In Equation 1.12, such effect is completely transferred to the moral-hazard effect, for reductions in the cash-to-risk ratio must be necessarily followed by increments in the market capital cost.

However, if (as previously suggested) borrowers are grouped in a finite number of categories, Equation 1.12 is likely to be violated and differences are expected between the internal and external estimations of the cost of capital:

\[
\Delta = \frac{E[X_+] - E[(X - k)_+]}{k} - r \neq 0
\]

Accordingly, capital is regarded as too expensive for those firms that obtain \( \Delta < 0 \), for in this case maintaining the cash-to-risk ratio \( k \) produces a loss that is lower than the alternative cost of borrowing the same balance in the market. These firms prefer to demand reserves instead of relying on external finance. Conversely, capital is cheap for those firms that obtain \( \Delta > 0 \), for they have to incur in a higher loss if they maintain some cash balance. These firms prefer to demand no capital at all.

The case \( \Delta > 0 \) actually represents the situation of firms that obtain gains over the level of capital and dispose of free-cash-flow. In the model of Jensen (1986), frictions and mismanagement are specially severe within firms disposing of high amounts of free-cash-flow, resulting from the competition, between managers and shareholders, to take control of the profits.
generated by the company. The level of capital demanded by such firms can be described by means of the second-order condition. Indeed, given any cash-to-risk ratio $k$, we expect the probability $P\{X = k\}$ to be higher for firms that obtain higher amounts of free-cash-flow, and that eventually this probability exceeds $P\{X = -k\}$ when a certain capital threshold is surpassed. As we have already noticed, such turning point is found at $k = E[X]$ in the case of symmetric probability distributions.

In any case, there will always be institutions with sufficiently high cash in excess for whom to maintain a single unit of capital induces a net loss in value. Those firms prefer to demand no capital at all and do not consequently internalise the cost of bankruptcy. Accordingly, not only agency costs are expected to be high inside firms with abundant free-cash-flow, also the incentives to provide capital cushions to guarantee their outstanding liabilities are non-existent in these firms, a situation that may aggravate the moral-hazard in the relationship between managers and customers.

In conclusion, short-term stickiness inherent in the equity contracts established with stockholders, as well as in the credit categorisations determined by lenders in the markets of cash balances, prevent financial institutions from continuously adjusting their capital structures. On the one hand, demanding additional equity from shareholders may well increase the agency costs inside the institution. But on the other hand, raising the amount of external debt may raise the bankruptcy costs faced by the institution and the moral-hazard inherent in their relationship with creditors.

More explicitly, diminishing the cash-to-risk ratio always leads the term $E[(X + k)\ldots]$ (equal to bankruptcy costs) to rise, and sometimes, when the magnitude of the variation in $k$ is high enough to produce the firm to transit from one risk category to another, also the level of the market capital cost $r$ may increase. Conversely, raising $k$ always reduces the actuarial price of the bankruptcy claim and sometimes also reduces the cost of capital.

An optimal capital principle will be derived in the next section, based on

---

13 Competition in product and factor markets should push utilities to a minimum level — eventually to zero. Then only those activities generating substantial economic rents are able to generate substantial amounts of free-cash-flow. Such activities are corresponded to product and factor markets where market forces are weak and where monitoring is more important than ever (Jensen, 1986).

14 Kisgen (2006) states that credit ratings directly affect the capital decisions of managers. This behaviour is explained on the basis of discrete costs and benefits arising when firms are located near to some change of rating — which lead these firms to issue less debt relative to equity than firms far to changes in rating. Thus Kisgen provides an alternative interpretation to the friction induced by credit ratings.
1.4 Economic Capital as the Optimal Deductible

The value, per unit of investment, of firms that can borrow at the interest rate \( r \) (whose cost of equity is given by Equation 1.12) is equal to:

\[
V = \frac{E[X_+] - E[(X + k)_-] - (r - r_0) \cdot k}{T}
\]

Maximising value is then equivalent to minimise the total burden of default, equal to the price of insurance (represented by the term \( E[(X + k)_-] \)) plus the net benefit obtained when investing capital at the interest rate \( r \) instead of maintaining it at the low (non-risky) rate \( r_0 \):

\[
\min_k E[(X + k)_-] + (r - r_0) \cdot k
\]  
(1.13)

This problem has been already used to derive a rule of capital allocation by Dhaene et al. (2003, 2008), Goovaerts et al. (2005) and also Laeven and Goovaerts (2004). They regard its solution as an optimal solvency margin, which establishes a compromise between the cost of capital on the one hand and solvency requirements on the other. When justifying the implementation of this rule, they emphasise that arbitrage opportunities are difficult to exploit in insurance markets.

The first and second order conditions of the optimisation problem of Equation 1.13 can be obtained by combining Equations 1.6 and 1.7:

\[
\frac{d}{dk} E[(X + k^*)_-] + r - r_0 = -T_{-X}(k^*) + r - r_0 = 0
\]

\[
\frac{d^2}{dk^2} E[(X + k^*)_-] = \frac{d^2}{dk^2} F_{-X}(k^*) = f_{-X}(k^*) = P \{X = -k^*\} > 0
\]

Therefore, as long as some capital loss is produced with non-zero probability, a range exists where the term \( E[(X + k)_-] \) is convex in \( k \), in such a way that if additionally the marginal benefit of adding the first unit of capital

an optimal compromise of bankruptcy costs and the market price of external debt.
is greater than the net investment premium (i.e. if additionally $T_{-X}(0) > r - r_0$), then a level of capital exists that minimises the criterion of Equation 1.13.

Under such circumstances, the optimal capital demand is determined by the quantile function of the probability distribution of the series of capital losses of the underlying risk:

$$k^* = T_{-X}^{-1}(r - r_0) = F_{-X}^{-1}(1 - r + r_0) \quad (1.15)$$

The optimal level of surplus is then expressed as the Value-at-Risk (or VaR) for the confidence probability level $\nu = r - r_0$.\(^{15}\)

The fact that the confidence level in the definition of VaR is replaced by a net premium in Equation 1.15 is a consequence of the first-order condition, which determines an exchange between a sure flow and a flow of probability (see Equation 1.14). Thus, the higher the liquidity premium $\nu$ (i.e. the more the free-cash-flow at disposal), the more expensive is to maintain a cash balance and hence the less capital is demanded. Conversely, the lower this premium, the cheaper the capital and hence the more the demanded quantity of this resource. The minimum and the maximum levels are respectively chosen when $\nu \geq 1$ and when $\nu \leq 0$.

From the actuarial viewpoint, the expected excess loss $E[(X + k)_-]$ represents the fair price of a special insuring contract (sometimes called layer) that obliges the insurer to pay to the policyholder the excess of the loss over the level $k$, when such a loss is produced (see Goovaerts, 1984, Dhaene et al., 2006, and also Appendix E in this thesis). In this context, the amount $k$ represents a guarantee provided by the policyholder in order to assure the insurer (up to some extent) that every reasonable care will be taken to reduce the underlying exposure. In other words, the guarantee $k$, which is known as the deductible or retention in the literature, is introduced in insurance contracts as a means of reducing the costs derived from moral-hazard.

Within this framework, the optimal level of capital defined in Equation 1.13 corresponds to the optimal deductible or optimal retention of the related insuring liability contract — and hence, it explicitly represents the moral-hazard induced by the opacity of the firm.

\(^{15}\)The reader not familiar with the concept of VaR can find a good survey in Hull (2000). The VaR has been recommended for the implementation of good risk management practices by the Basel Committee on Banking Supervision (2004).
Notice, however, that full-coverage is implicitly assumed in the model, because the actuarial prices of equity and insurance have been expressed in terms of mathematical expectations that consider unlimited losses and gains over the level of capital (see Equations 1.5 and 1.13). But insurance contracts always specify a maximum payment in practice and full-coverage does not actually exist in real markets. The question then arises of who does eventually bear the risk of deposits.

According to the terms of the guaranteeing contracts previously defined, we can say that risk-bearing is roughly distributed in the following way: any loss up to the retention level $k$ is payed by the firm (recall that shareholders only endure the equity component of the economic capital, $k^{\text{EQ}} \leq k$); losses that are higher than the retention level are payed by the guarantor or insurer, as long as these losses do not surpasses a maximum disaster level $M^{\text{DIS}}$; finally, some companies can look for additional protection by establishing a contract with some reinsurance institution that agrees to pay any loss greater than the disaster level $M^{\text{DIS}}$ but lower than some catastrophe level $M^{\text{CAT}}$.

Thus, in the case of catastrophic events, it is the society as a whole who has to afford the losses — through governmental divisions, private creditors, companies and householders. This explains why it is in the interest of regulators to define good practices and regulatory requirements that can induce financial intermediaries to seek for protection according to the risk borne.

From the economic point of view, the existence of an optimal level of capital implies that choosing a different level necessarily leads to over- or under-investment. Indeed, idle money, that could be assigned to profitable investments, is maintained in excess when more capital than the optimal is demanded. By contrast, when the stock of capital is lower than the optimal level, risk is taken in excess, a fact that might eventually increase the frequency of losses (as well as disaster and catastrophic events) and induce investors to raise their concerns about the credit quality of the firm. The price at which the firm can attract debt in the market might increase under such circumstances.

Therefore, independently of whether managers consider or not any of the optimisation problems established in Equations 1.5 or 1.13, their capital preferences should approach to the solutions of these problems, for only following this strategy the market value of firms is maximised (or the burden of bankruptcy is minimised). On these grounds, the capital principles
defined in Equations 1.10 and 1.15 can be regarded as decision variables, which provide a basis for the determination of the aggregate behaviour of markets or multidivisional corporations. Thus, in particular, such rules will be used later in Sections 1.7 and 1.8 to determine allocation mechanisms to be applied inside centralised and decentralised organisations.

Having defined a principle to determine the optimal cash balance, the problem of distributing it inside multidivisional organisations will be addressed in the rest of the chapter. As explained later in Section 1.5, the main issues to consider are informational asymmetries (producing agency costs) inside institutions and how the decentralisation of capital decisions can be used in the interest of the conglomerate as a whole.

Specifically, an optimal centralised allocation principle (depending on the beliefs of the central administration about the risks taken by subsidiaries) will be defined in Section 1.7, and an optimal decentralised mechanism will be proposed in Section 1.8, which leads to the centralised allocation (at the aggregate level) at the time that forces divisional managers to reveal their informational type.

But then some mathematical framework should be adopted allowing the insurance principles \( E[(X - k)_+] \) and \( E[(X + k)_+] \) to be distorted according to the preferences of decision-makers. As will be shown next in Section 1.6, the distorted-probability principle (which modify the probabilities in the expectation operator) provides a characterisation that leads to an explicit separation of the effects of risk and preferences in the capital structure. Such separation will allow for a clear representation of the informational asymmetries between central and divisional managers, as well as the costs derived from decentralisation.

1.5 The Allocation of Capital within Multidivisional Corporations

As established in the previous sections, the prices of securities that produce random outcomes are completely determined by risk. Then the insurance price of portfolios containing such instruments are also dependent on risk and hence, every capital allocation principle — distributing capital among the outstanding divisions of some multidivisional corporation — should be strictly based on it (see, among others, Albrecht, 2004, and Goovaerts et al., 2005).
1.5 The Allocation of Capital

In the following, we will respectively denote the series of random P&L produced by divisions and the stand-alone capital requirements at the divisional level as $X_1, \ldots, X_n$ and $K[X_1], \ldots, K[X_n]$, where the function $K[\cdot]$ is corresponded to some economic criterion. In particular, the amount of capital demanded at the corporate level is given by $K[X]$, where the random variable $X = X_1 + \ldots + X_n$ denotes the aggregated P&L.

It is a well-known theoretical fact that when the variations of the portfolios held by subsidiaries are not perfectly correlated, the losses suffered by some divisions can be at least partially compensated by the gains obtained in others — i.e. divisions can partially hedge or insure each other. It is then claimed that a benefit arises due to diversification, which implies that the capital required at the aggregate level is generally lower than the sum of the levels maintained by divisions when acting as independent units.

In fact, provided that the operator $K[\cdot]$ is convex, the Jensen’s inequality can be used to prove that (see Merton and Perold, 1993):

$$K[X] \leq \sum_{i=1}^{n} K[X_i]$$ (1.16)

Thus, in particular, some businesses that would be unprofitable on a stand-alone basis (due to high capital requirements) might be profitable within a firm holding businesses with offsetting risks. Under such conditions, the decentralisation of capital decisions might produce under-investment and hence full-allocation is frequently mentioned as a desirable property:

$$\sum_{i=1}^{n} K_i = K[X] \quad \text{with} \quad K_i \geq 0 \quad \forall \ i = 1, \ldots, n$$ (1.17)

where $K_1, \ldots, K_n$ denote the levels of capital determined according to some centralised criterion. In this context, the covariance allocation principle is introduced (see Albrecht, 2004):

$$K_i = E[X_i] + \frac{Cov(X, X_i)}{Var[X]} \cdot (K - E[X]) \quad \text{with} \quad X = \sum_{i=1}^{n} X_i$$

Full-allocation is obtained as a consequence of the following properties of
Chapter 1: Economic Capital in Opaque Conglomerates

the expectation and covariance operators:

\[ E[X] = \sum_{i=1}^{n} E[X_i] \quad \text{and} \quad Var[X] = \sum_{i=1}^{n} Cov(X, X_i) \]

Then the covariance principle satisfies the three properties just mentioned: it is risk-based, it explicitly introduces covariances and finally, it fully allocates capital.

Merton and Perold (1993) have pointed out, however, that full-allocation can also lead to under-investment under certain circumstances and consequently, that such strategy does not always correspond to the best practice. This claim can be understood by considering a specific approach of allocating capital.

Indeed, let us define the incremental capital requirement of some division as the amount of capital required at the corporate level \( K[X] \) minus the capital required by the conglomerate when the particular business unit is eliminated:

\[ I_{K}[X_i] = K[X] - K[X - X_i] \quad \text{with} \quad X = \sum_{i=1}^{n} X_i \]

Then the incremental risk capital represents the minimum cash balance required to support a certain business unit. Besides, it can be easily verified that for every \( n \geq 2 \) the following identity holds:

\[ (n - 1) \cdot X = \sum_{i=1}^{n} \sum_{j \neq i} X_j = \sum_{i=1}^{n} (X - X_i) \]

We can apply again the Jensen’s inequality (see the technical appendix in Merton and Perold, 1993) to obtain that:

\[ (n - 1) \cdot K[X] \leq \sum_{i=1}^{n} K[X - X_i] \implies \sum_{i=1}^{n} I_{K}[X_i] \leq K[X] \quad (1.18) \]
1.5 The Allocation of Capital

Hence the sum of the incremental capitals is lower than the capital required at the aggregate level and then full-allocation overstates the marginal capital requirements.

In other words, while the stand-alone allocation induces a loss in efficiency to the conglomerate (whenever a convex principle \( K[\cdot] \) is introduced), full-allocation can induce divisions to incur in a loss of efficiency, for the sum of the minimum additional (or incremental) capital requirements is lower than the level determined (according to the same principle) to the whole conglomerate. On these grounds, Merton and Perold recommend do not fully allocate capital.\(^\text{16}\)

Therefore, the main issue when specifying a capital allocation principle is how to simultaneously incorporate the preferences of central and divisional administrations.

Discrepancies inside institutions can be explained on the grounds of differing attitudes towards risk, as well as on differing expectations sustained by information and knowledge. Under such conditions, the distribution of capital reflects the competition between divisional and central managers on the one hand, and stockholders on the other, to take control of the funds generated by the company.

Denault (2001) shows that the theory of coalitional games can be used to build a model that explicitly characterises such competition. In the model, subsidiaries choose between working as stand-alone entities and forming holdings (or coalitions) with other divisions. In this setting, the sub-additivity property (Equation 1.16) implies that subsidiaries have incentives to form coalitions, since in this way they obtain a benefit due to the reduction of the aggregate cost of bankruptcy. However, the super-additivity property (which in particular satisfies the incremental allocation principle, see Equation 1.18) implies that the subsidiaries forming a certain coalition have incentives to abandon it, as long as in this case the aggregate capital requirement is higher than the sum of the optimal surpluses determined on a stand-alone basis. Within this theoretical framework, Denault proposes that any coherent allocation principle should satisfy full-allocation.

As will be shown later in Sections 1.7 and 1.8, the capital principle of

---

\(^\text{16}\)The principle can be naturally extended by considering the marginal-capital requirement, defined by Myers and Read (2001) as the variation in the capital requirement in response to a marginal increment in the exposition to risk, i.e. \( M_K[X_i] = dK[X]/dX_i \). But then full-allocation can be guaranteed only if some conditions on the valuation function \( K[X] \) are satisfied (see also Mildenhall, 2004, and Grundl and Schmeiser, 2007).
Equation 1.22 can be naturally adapted to characterise the divisional capital requirements in multi-businesses corporations. Thus, an optimal centralised allocation principle is defined in Section 1.7, in the sense that it minimises the sum of the insurance prices of the divisional excess losses. Later in Section 1.8, an optimal mechanism is described that is based on the stand-alone optimal levels of reserves, but that leads to the same aggregate requirements as in the centralised allocation. The instrument of this mechanism is the internal cost of capital.

We will first digress from this topic in the next section, in order to provide a mathematical description of the effect of expectations over actuarial prices, on which basis the allocation principles of Sections 1.7 and 1.8 will be built.

1.6 The Distorted Demand for Capital

Within the framework of the well-known utility-theory of choice under risk, preferences are uniquely corresponded to utility functions and economic decisions are based on the expected utility:

\[ E_u[X] = \int u(x) \cdot dF_X(x) \]

In fact, for any utility function \( u(\cdot) \), the expected utility operator \( E_u[\cdot] \) always provides an order of risks. On these grounds, the process of decision-making can be described. Although this framework has been already used for obtaining a demand for capital (or more generally, a demand for liquidity, see Tobin, 1958, Holmstrom and Tirole, 2000, and Choi and Oh, 2003), the formulas obtained in this way are not really tractable and the description in terms of preferences is dependent on the specification of the utility function.

Additionally, we would like to extend the capital principle defined in Equations 1.13 and 1.15 in such a way that its principal features are maintained. In this context, we prefer to distort probabilities instead of payments, as long as in this way the fundamental prescription given in Equation 1.15, relating the optimal level of capital to the quantile of the underlying exposure, is preserved. In other words, we would like the distorted capital to be equal to the quantile of some distorted probability distribution. According to this view, preferences must directly affect the perceptions of risk maintained by decision-makers.
1.6 The Distorted Demand for Capital

Yaari (1987) has demonstrated that preferences can be alternatively corresponded to a distortion function $\varphi : [0, 1] \rightarrow [0, 1]$ affecting the underlying probability distributions. On this basis, the distorted expectation operator is defined:

$$
E_{\varphi}[X] = \int x \, dF_{\varphi,X}(x) = \int T_{\varphi,X}(x) \, dx \quad \text{with} \quad T_{\varphi,X}(x) = \varphi(T_X(x)) \quad \forall x 
$$

where $F_{\varphi,X} = P_{\varphi}\{X \leq x\}$ and $T_{\varphi,X} = P_{\varphi}\{X > x\}$ respectively denote the distorted cumulative and the distorted tail probability functions representing the underlying risk (see Appendixes A, B and C).

Wang et al. (1997) prove that the risk-principle introduced in Equation 1.19 satisfy a set of good properties for the pricing of insurance claims. In particular, it preserves a class of stochastic orders, meaning that it strictly depends on risk — i.e. more risky claims are given higher prices and vice-versa (see Appendix D). Besides, risk-lover and averse-to-risk attitudes are respectively characterised by convex and concave distortion functions, for in these cases the expectation operator is respectively under- and over-estimated. Risk-neutrality is represented by the identity function $\varphi(x) = x$ $\forall x$ (see also Wang and Young, 1998).

Although the family of acceptable distortions (in the terms established by Yaari and Wang) is broad, we will consider in the following the particular class of proportional-hazards transforms, defined as $\varphi_\theta(p) = p^{1/\theta} \forall p \in [0, 1]$, in such a way that the corresponding risk-principle is now expressed as (see Wang, 1995):

$$
E_{\theta}[X] = \int x \, dF_{\theta,X}(x) = \int T_{\theta,X}(x) \, dx \quad \text{with} \quad T_{\theta,X}(x) = T_X(x)^{\frac{1}{\theta}} \quad \forall x 
$$

Consequently, the probability beliefs and hence the price of risk are amplified when $\theta > 1$ and then this range of the distortion parameter $\theta$ accounts for the behaviour of averse-to-risk individuals. Besides, the higher the magnitude of $\theta$, the more the price of risk is over-estimated and then the more the aversion-to-risk. Similarly, the price of risk is under-estimated when $\theta < 1$, so that in this way the behaviour of risk-lovers is characterised. Thus, the lower the magnitude of $\theta$, the lower the price of risk and then the more the preference for risk. Risk-neutral decision-makers are characterised
by $\theta = 1$, in which case the distorted probability principle is equal to the traditional expectation operator.

The capital principle defined in Equations 1.13 and 1.15 can be naturally extended to the risk pricing framework based on the distorted-probability principle defined in Equation 1.20:

$$\min_k E_\theta [(X + k)_-] + (r - r_0) \cdot k$$

(1.21)

with $E_\theta [(X + k)_-] = -\int_{-\infty}^{-k} (x + k) \cdot dF_{\theta,X}(x)$ (see Equation 1.6). The Leibnitz’s integral rule can be invoked once again to derive the first-order condition, as in the non-distorted case (see Equation 1.14), in such a way that the optimal demand for capital is given by the inverse function of the distorted tail-probability:

$$k_{\theta,X}(r - r_0) = T_{\theta,X}^{-1}(1 - r + r_0)$$

(1.22)

or equivalently:

$$k_{\theta,X}(r - r_0) = T_{-X}^{-1} \left( (r - r_0)^\theta \right) = F_{-X}^{-1}(1 - (r - r_0)^\theta)$$

Discrepancies relative to preferred cash-balances can thus be explained on the basis of the underlying risks, preferences and the opportunity cost of capital.

Notice that although in Equation 1.22 the parameter $\theta$ eventually affects the cost of capital, from the mathematical point of view it is not this variable but rather the underlying probability distribution what is distorted, a fact that has been stressed by choosing the notation $F_{\theta,X}$ and $T_{\theta,X}$. The reader should not confuse this setting, where expectations distort the risk perceptions of decision-makers, with that of heterogeneous estimations of the cost of capital.
1.7 Risk-Based Allocation for Centralised Organisations

Let $X$ and $K$ respectively denote the aggregate exposure and the level of capital maintained by some financial conglomerate, and let $X_1, \ldots, X_n$ and $K_1, \ldots, K_n$ respectively denote the series of capital returns of the portfolios held by its subsidiaries and their corresponding cash balances.

Consistently with the capital principle defined in Equations 1.21 and 1.22, we look for a rule based on the residual risks of subsidiaries. A criterion is thus required to compare the sum of the subsidiaries’ residual risks, equal to $\sum_{i=1}^{n} (I_i \cdot X_i + K_i)$, with the conglomerate’s residual exposure $(I \cdot X + K)_-$, where $I_1, \ldots, I_n$ and $I = \sum_{i=1}^{n} I_i$ respectively denote the levels of investment maintained by subsidiaries and the conglomerate, with $I \cdot X = \sum_{i=1}^{n} I_i \cdot X_i$.

The comparison of random variables can be carried out with the help of the concept of stochastic precedence. Indeed, a random variable $X$ is said to precede another random variable $Y$ (or equivalently, a random variable $Y$ is said to dominate $X$) in the first stochastic order if and only if at every point it accumulates less probability in the tails of the distribution function (see Goovaerts et al., 1984 and Appendix D):

$$X \leq Y \iff P\{X > k\} \leq P\{Y > k\} \quad \forall k \iff T_X(k) \leq T_Y(k) \quad \forall k \quad (1.23)$$

This is a convenient order indeed, for the optimal level of capital is expressed in Equations 1.15 and 1.22 in terms of the tail probability function.

Alternatively, it can be demonstrated that the following inequality holds with probability one for the first stochastic-order:

$$(I \cdot X + K)_- \leq \sum_{i=1}^{n} (I_i \cdot X_i + K_i)_-$$

where $\sum_{i=1}^{n} K_i \leq K$

Recall that, as stated by Denault, the property of super-additivity (which states that $\sum_{i=1}^{n} K_i \leq K$) implies that there are incentives for divisions to leave the conglomerate and to form independent firms. From the inequality above we obtain that a benefit exists, resulting from the reduction in the riskiness of the conglomerate’s portfolio with respect to the sum of the
divisional portfolios, that can be shared with divisions to put incentives on them to remain in the conglomerate.

It can be proved as well that the first stochastic order is preserved by the distorted-probability principle (defined in Equations 1.19 and 1.20), in such a way that, for any informational type \( \theta \):

\[
E_{\theta} \left[ (I \cdot X + K)_- \right] \leq \sum_{i=1}^{n} E_{\theta} \left[ (I_i \cdot X_i + K_i)_- \right]
\]

\[
\sum_{i=1}^{n} K_i \leq K
\]

Given some level of aggregate investment \( I > 0 \), define the following weights:

\[
\omega_i = \frac{I_i}{I} \quad \forall i \quad \Rightarrow \quad \sum_{i=1}^{n} \omega_i = 1
\]

Hence, letting \( k = K/I \) and \( k_i = K_i/I_i \), with \( i = 1, \ldots, n \), respectively denote the cash-to-risk ratios maintained by the conglomerate and subsidiaries, we obtain that:

\[
E_{\theta} \left[ (X + k)_- \right] \leq \sum_{i=1}^{n} \omega_i \cdot E_{\theta} \left[ (X_i + k_i)_- \right]
\]

\[
\sum_{i=1}^{n} \omega_i \cdot k_i \leq k
\]

An allocation principle can then be naturally introduced, which minimises the weighted sum of the expectations of the residual risks:

\[
\min_{k_1, \ldots, k_n} \sum_{i=1}^{n} \omega_i \cdot E_{\theta} \left[ (X_i + k_i)_- \right]
\]

\[
\sum_{i=1}^{n} \omega_i \cdot k_i = k
\]

Such allocation principle ensures the sum of the expected insured returns of subsidiaries is as close as possible to the lower bound \( E_{\theta} \left[ (X + k)_- \right] \).

As in the previous sections, the solution to the problem of Equation 1.24 can be obtained by applying Lagrange optimisation. In fact, the first-order conditions are now written as:
\[
\frac{\partial}{\partial k_i} \sum_{i=1}^{n} \omega_i \cdot E_\theta [(X_i + k_i^*)_-] + \gamma \cdot \omega_i = (-T_{\theta_i} - X_i (k_i^*) + \gamma) \cdot \omega_i = 0 \\
\forall \ i = 1, \ldots, n
\]
\[
\sum_{i=1}^{n} \omega_i \cdot k_i^* - k = 0
\]

(1.25)

where \( \gamma \) denotes the *Lagrange multiplier*. Combining both equations leads the Lagrange multiplier to be determined by:

\[
\sum_{i=1}^{n} \omega_i \cdot k_i^* = \sum_{i=1}^{n} \omega_i \cdot T_{\theta_i}^{-1} X_i (\gamma) = k
\]

Besides, since \( T_{\theta_i}^{-1} X_i = \omega_i \cdot T_{\theta_i}^{-1} X_i \forall \ \omega_i > 0 \), we obtain:

\[
\sum_{i=1}^{n} T_{\theta_i}^{-1} X_i (\gamma) = k \quad \text{with} \quad X = \sum_{i=1}^{n} \omega_i \cdot X_i
\]

The question then arises of whether there is some dependence structure for which the sum of the quantiles is expressed as the quantile of the sum of the individual exposures, for in this case it would be possible to obtain a close expression for the multiplier \( \gamma \).

In fact, as demonstrated by Dhaene et al., 2002, the property of the sum of the quantiles mathematically characterises the *comonotonic* dependence structure, where *comonotonicity* represents a case of extreme dependence, when no diversification effect can be attained by pooling risks together. In fact, given a random vector \((X_1, \ldots, X_n)\) with marginal cumulative distribution functions \((F_{X_1}, \ldots, F_{X_n})\), the *comonotonic* random vector \((X_1^c, \ldots, X_n^c)\) is mathematically defined in such a way that if \(U\) denotes the random variable uniformly distributed in the interval \([0, 1]\), such that \(F_U(u) = u \forall \ u \in [0, 1]\), \(F_U(u) = 0 \forall \ u < 0\) and \(F_U(u) = 1 \forall \ u > 1\), the following identity holds in distributions:

\[
(X_1^c, \ldots, X_n^c) = \left(F_{X_1}^{-1}(U), \ldots, F_{X_n}^{-1}(U)\right)
\]

In this way, the realisation of a single event (related to the uniform random variable \(U\)) simultaneously determines all the components of any *comono-
monic random vector. Moreover, since the functions \( F_{X_1}, \ldots, F_{X_n} \) are all non-decreasing, all the components of the vector \( (X_1^c, \ldots, X_n^c) \) move in the same direction. Hence, as already stated, the quantile function of the sum of the components of any comonotonic random vector is equal to the sum of the component quantile functions, a fact that supports the use of the comonotonic dependence structure to characterise the aggregate demand for cash balances.

Let \( T_{\theta,X^c} \) denote the tail-probability of the comonotonic sum \( X^c = \omega_1 \cdot X_1^c + \ldots + \omega_n \cdot X_n^c \), where \( (X_1^c, \ldots, X_n^c) \) represents the comonotonic random vector with the same marginal distributions as \( (X_1, \ldots, X_n) \). Then, from the first-order conditions written in Equation 1.25 we obtain that:

\[
T_{\theta,-X^c}(\gamma) = \sum_{i=1}^n T_{\theta,-\omega_i \cdot X_i}(\gamma) = k
\]

with \( X^c = \sum_{i=1}^n \omega_i \cdot X_i^c \)

where \( T_{\theta,-X^c} = \left( \sum_{i=1}^n T_{\theta,-\omega_i \cdot X_i} \right)^{-1} \) denotes the distribution function of the comonotonic sum. Hence the Lagrange multiplier is equal to:

\[
\gamma = T_{\theta,-X^c}(k) \tag{1.26}
\]

while the optimal allocation rule is given by:

\[
k^*_i = T_{\theta,-X_i}(\gamma) = T_{\theta,-X_i}(T_{\theta,-X^c}(k)) \quad \forall \ i = 1, \ldots, n \tag{1.27}
\]

Thus, from the mathematical point of view, the optimal levels of capital are corresponded to the projections over the subspaces determined by the individual risks. This interpretation is consistent with the fact that the problem of Equation 1.24 actually minimises a distance, as stated by Laeven and Goovaerts, 2004, and also Goovaerts et al., 2005.

Notice that the optimal allocation principle determined in Equation 1.27 depends on the level of capital \( k \) chosen for the conglomerate. When central managers choose the level that minimises the criterion of Equation 1.21, the optimal level of capital for the conglomerate depends on the net return \( r - r_0 \) as in Equation 1.22:
Replacing the optimal level $k_{\theta,X^c}$ in Equation 1.26 leads the Lagrange multiplier to be given by:

$$\gamma = T_{\theta,-X^c}^{-1}(r - r_0) = r - r_0$$

Thus the Lagrange multiplier is equal to the net return on capital — a fact that is consistent with the interpretation of the Lagrange multiplier as a shadow price. Replacing $\gamma$ in Equation 1.27 leads to the optimal allocation rule:

$$k_{\theta,X_i} = T_{\theta,-X_i}^{-1}(r - r_0) \quad \forall \ i = 1, \ldots, n$$

Accordingly, the optimal levels of capital correspond to the levels that central managers would choose for the divisional risks on the grounds of their net return on capital $r - r_0$ and their expectations (represented by the informational type $\theta$). On these grounds, it is regarded as a centralised allocation.

## 1.8 Optimal Decentralised Mechanism

Subsidiaries that are allowed to choose their capital structures on their own interest solve the problem of Equation 1.21:

$$\min_{k_i} E_{\theta_i} [(X_i + k_i)_{-}] + (r - r_0) \cdot k_i$$

The stand-alone allocation of capital is thus determined according to Equation 1.22:

$$k_{\theta_i,X_i} = T_{\theta_i,-X_i}^{-1}(r - r_0) \quad \forall \ i = 1, \ldots, n$$

Comparing Equations 1.30 and 1.32, we notice that the centralised and stand-alone allocations differ due to differing expectations between central
and divisional managers, i.e. due to differences in the parameters $\theta$ and $\theta_i$. A loss of efficiency is then produced at the corporate level. Indeed, under-investment occurs at the corporate level when $\sum_{i=1}^{n} \omega_i \cdot k_{\theta_i, X_i} > k_{\theta, X_c}$ and over-investment when $\sum_{i=1}^{n} \omega_i \cdot k_{\theta_i, X_i} < k_{\theta, X_c}$. Similarly, preferring the centralised allocation respectively leads to under- and over-investment at the divisional level when $k_{\theta, X_i} > \omega_i \cdot k_{\theta_i, X_i}$ and when $k_{\theta, X_i} < \omega_i \cdot k_{\theta_i, X_i}$.

Therefore, deciding between the centralised and the stand-alone allocations of capital is just a matter of preferences. Those central administrations that are confident in their own estimations will prefer to rely on the centralised allocation. Others will try to incorporate, at least partially, the view of divisional managers.

Notice, however, that the allocation principle of Equation 1.32 allows central managers to extract privately held information from subsidiaries. Moreover, it can be defined in order to collect the same aggregate amount of capital recommended by the centralised allocation.

Indeed, let $k_1, \ldots, k_n$ denote the surpluses prefer by divisional managers, in such a way that, from Equation 1.32 we obtain:

$$r - r_0 = T_{\theta_i, -X_i}(k_i) = T_{-X_i}(k_i)^{\frac{1}{\omega_i}} \implies \theta_i = \frac{\log T_{-X_i}(k_i)}{\log (r - r_0)} \quad \forall \ i = 1, \ldots, n$$

(1.33)

Hence, given any risk $X_i$ and a level of capital $k_i$, the liquidity premium $r - r_0$ and the divisional informational type must be related to each other. Accordingly, as long as the central administration force subsidiaries to invest their balances at some internal return $\rho$ (instead of the risk-free interest rate $r_0$), the later are forced to act on the interest of the conglomerate at the time that their types are revealed.

Such a mechanism can be explicitly defined in the form of an optimal contract, as suggested by Diamond and Verrecchia (1982):

$$\min_{k, \rho} \ E_{\theta} [(X + k)_{-}] + (r - r_0) \cdot k$$

subject to

$$k_i = \arg \min_{k_i} \ E_{\theta_i} [(X_i + k_i)_{-}] + (r - \rho) \cdot k_i \quad \forall \ i = 1, \ldots, n$$

$$\sum_{i=1}^{n} \omega_i \cdot k_i = k$$
1.8 Optimal Decentralised Mechanism

As long as the optimal stand-alone allocations are given by Equation 1.32, the contract can be equivalently established in the following way:

\[
\begin{align*}
\min_{k, \rho} & \quad E_\theta [(X + k) - ] + (r - r_0) \cdot k \\
\text{subject to} & \quad \sum_{i=1}^n \omega_i \cdot k_{\theta_i, X_i} (r - \rho) = k
\end{align*}
\] (1.34)

Letting \( \gamma \) denote the Lagrange multiplier, we obtain that the optimal level of capital at the corporate level (\( k^* \)) and the optimal internal return on capital (\( \rho^* \)) must be determined in order to satisfy the first-order conditions:

\[
\frac{d}{dk} E_\theta [(X + k^*) - ] + (r - r_0) - \gamma = -T_{\theta, X} (k^*) + (r - r_0) - \gamma = 0
\]

\[
\gamma \cdot \sum_{i=1}^n \omega_i \cdot k'_{\theta_i, X_i} (r - \rho^*) \cdot (-1) = 0 \implies \gamma = 0
\]

\[
\sum_{i=1}^n \omega_i \cdot k_{\theta_i, X_i} (r - \rho^*) = \sum_{i=1}^n \omega_i \cdot T_{\theta_i, -X_i}^{-1} (r - \rho^*) = \sum_{i=1}^n T_{\theta_i, -\omega_i, X_i}^{-1} (r - \rho^*) = k^*
\]

Therefore, if \( T_{\theta_1, \ldots, \theta_n, X^c} \) denotes the tail-probability function of the comonotonic sum with marginal distributions \( (T_{\theta_1, -X_1}, \ldots, T_{\theta_n, -X_n}) \), i.e. \( T_{\theta_1, \ldots, \theta_n, X^c} = \left( \sum_{i=1}^n T_{\theta_i, -\omega_i, X_i}^{-1} \right)^{-1} \), we obtain:

\[
\rho^* = r - T_{\theta_1, \ldots, \theta_n, X^c} (k^*) = r - T_{\theta_1, \ldots, \theta_n, X^c} \left( T_{\theta, -X^c}^{-1} (r - r_0) \right)
\]

with:

\[
k^* = T_{\theta, -X^c}^{-1} (r - r_0)
\] (1.35)

But the parameters \( \theta_1, \ldots, \theta_n \) are actually not observed by central managers, which means that this allocation cannot be implemented in practice. As a solution, we can assume that first central managers have estimations \( \hat{\theta}_1, \ldots, \hat{\theta}_n \) of the types of divisional managers — which can be determined on the basis of experience or previous allocations. Then the optimal internal return on capital should be defined as:

\[
\rho^* = r - T_{\hat{\theta}_1, \ldots, \hat{\theta}_n, X^c} \left( T_{\theta, -X^c}^{-1} (r - r_0) \right)
\] (1.36)
Replacing \( r_0 = \rho^* \) in Equation 1.32, we obtain that the decentralised optimal allocation of capital is attained at the levels:

\[
k_i^* = T_{\hat{\theta}_i,-X_i}(r - \rho^*) = T_{\hat{\theta}_i,-X_i} \left( T_{\hat{\theta}_1,...,\hat{\theta}_n,X_c} \left( T_{\hat{\theta}_i,-X_c}(r - r_0) \right) \right)
\]

\( \forall \ i = 1, \ldots, n \) (1.37)

Notice that the optimal internal cost of capital is equal to the external cost of capital only when central managers access to the private information of subsidiaries (and as long as they agree to incorporating it into decision-making), at the time that the proper comonotonic portfolio characterises the aggregated exposure. Then the liquidity premium \( \nu^* = r - \rho^* \), equal to the difference between the external and the optimal internal cost of capital, explicitly measures the disagreement between central and divisional managers.

The optimal mechanism can then be implemented in the following way. In Stage 1, based on their experience and information, central managers determine estimations of the informational types of subsidiaries, \( \hat{\theta}_1, \ldots, \hat{\theta}_n \). In case that only poor information is available, they can assume that divisional managers are risk-neutrals and accordingly choose \( \hat{\theta}_1 = \cdots = \hat{\theta}_n = 1 \).

In Stage 2, divisions are asked to freely decide their levels of reserves. The conditions of the contract are such that although subsidiaries can invest in the market to obtain the return \( r \), they can only invest their cash reserves in a special agreement with the central administration to obtain the return \( \rho^* \). In the final Stage 3, central managers use the allocations of Stage 2 to renew their estimations of the types of subsidiaries from Equation 1.33. In this way, divisional managers reveal their types.

In the model proposed by Stoughton and Zechner (2007), divisional managers are characterised by their attitude-towards-risk and the effort they expend to acquire information. Divisional cash flows are increased when more information is accessed. An informational parameter is then supposed to augment the expected value of the investment in a multiplicative fashion. Risk is described by volatility and decisions are affected by the net cost of capital. As in the model of Merton and Perold (1993), the economic capital is proportional to volatility. The internal cost of capital (fixed by central managers) is used as a tool to give incentives to subsidiaries to act in the general interest of the firm. Thus, capital allocation is justified as part of a general mechanism that stimulates the exchange of information between
central and divisional managers inside the institution. As a result, sometimes subsidiaries voluntarily reduce their exposure to risk and assign their cash excesses to less productive investments. So distortions can be present in the model in the form of under- and over-investment.

The interpretation of Stoughton and Zechner naturally applies to the mechanism stated above. Besides, the alternative expressions provided for the external and internal costs of capital can be easily implemented and are naturally corresponded to economic concepts. Finally, as demonstrated in the following section, in the particular case when risks are described by the Gaussian probability distribution, the optimal capital principle of Equation 1.22 extends the capital rule proposed by Merton and Perold (1993). This result is consistent with the fact that this principle has been established as an extension of the model of Merton and Perold, applicable to the case when decision makers cannot hedge fully and are obliged to rely on insurance and reinsurance to carry on with their business activities.

### 1.9 Heavy-Tailed Probability Distributions

An appealing feature of the capital principle defined in this chapter is its adaptability to any family of probability distributions. As a consequence, the capital requirements of different types of risks can be described on the same basis and hence, the model can be also implemented in institutions that hold securities described by probability distributions that present heavy tails — i.e. that uniformly assign more probability than the Gaussian distribution to the event of surpassing any given threshold.

This is particularly the case of insurance companies, that simultaneously deal with highly standardised policies, such as car and fire insurance, as well as some individual contracts involving high payments depending on events of low probability. This is also the case of some financial conglomerates that hold standard financial securities (transacted in highly liquid markets), as well as non-liquid derivatives and claims contingent on disaster and catastrophic events.

Explicit analytic expressions are obtained for a wide class of well-known probability distributions. In Table 1.1, some of these expressions are pre-

---

17Thus, while the external capital cost is related to the actuarial price of excess-cash-flow, the internal cost of capital is related to an optimal agreement between central and divisional managers.
Table 1.1: The Optimal Capital Principle under Different Risk Parametrisations

<table>
<thead>
<tr>
<th>Tail Probability</th>
<th>Optimal Capital Principle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>$T(x) = 1 - \Phi \left( \frac{x - \mu}{\sigma} \right)$ $\forall x$ $k(\nu) = \sigma \cdot \Phi^{-1} \left( 1 - \frac{\nu}{\nu_0} \right) - \mu$</td>
</tr>
<tr>
<td></td>
<td>with $\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{x} \exp \left( -\frac{y^2}{2} \right) dy$ $\forall x$ $\nu = r - r_0$</td>
</tr>
<tr>
<td>Log-Normal</td>
<td>$T(x) = 1 - \Phi \left( \frac{\mu + \ln(x)}{\sigma} \right)$ $\forall x$ $k(\nu) = \exp \left( \sigma \cdot \Phi^{-1} \left( 1 - \frac{\nu}{\nu_0} \right) - \mu \right)$</td>
</tr>
<tr>
<td>Exponential</td>
<td>$T(x) = \exp \left( \frac{C - x}{\beta} \right)$ $\forall x \geq 0$ $k(\nu) = C - (\theta \cdot \beta) \cdot \ln(\nu)$</td>
</tr>
<tr>
<td>Paretian</td>
<td>$T(x) = \left( \frac{1}{\beta} \right)^\frac{1}{\alpha}$ $\forall 0 &lt; x &lt; B$ $k(\nu) = B \cdot \nu^{\alpha}$</td>
</tr>
</tbody>
</table>

Presented in terms of the liquidity premium $\nu = r - r_0$.

In Figure 1.1, the optimal capital requirements are depicted for some of the risk classes presented in Table 1.1. Notice that, given any level of the liquidity premium, the optimal cash balance under Paretian risks is always higher than the cash balance demanded under Exponential risks, which in turn is always higher than the optimal capital under Gaussian risks. In fact, Paretian tails are uniformly greater than Exponential, which in turn are greater than Gaussian tails. This means that, according to the first-stochastic order, Gaussian risks are dominated by Exponential risks, which in turn are dominated by Paretian risks (see Equation 1.23 and the related discussion in Section 1.6). Therefore, the optimal capital principle consistently assigns higher surpluses to riskier claims and hence, it is strictly risk-based.

As depicted in Figure 1.2, the Gaussian capital principle (obtained when the underlying exposure follows a Gaussian probability distribution) follows a straight line in the plane of cash-to-risk ratios and standard deviations (used as estimators of volatilities). This is also the case of the allocation rule of Merton and Perold, because (as stated in the technical appendix of Merton and Perold, 1993) the formula of risk capital (per unit invested on net assets) can be approximated by:
1.9 Heavy-Tailed Probability Distributions

\[ k^{MP} = 0.4 \cdot \sigma \sqrt{T} \]

where \( T \) and \( \sigma \) respectively denote the time to maturity and the standard deviation of the series of \( P&L \) of the net assets' portfolio. In Figure 1.2, the value \( T = 1 \) has been assumed.

Therefore, while the capital principle of Merton and Perold intersects the capital axe at the origin and has a constant slope (equal to 0.4), the optimal principle intersects at the inverse of the mean return \( \mu \) of net assets and its slope \( \Phi^{-1}(1 - \nu^\theta) \) depends on the liquidity premium \( \nu \) and the informational type \( \theta \).

**Fig. 1.1: The Optimal Capital Principle of Neutral Decision Makers**

Lower liquidity premiums imply that firms hold less cash in excess and hence that they are willing to exchange more capital by any unit of incremented volatility (in particular, the slope tends to +\( \infty \) when \( \nu \to 0 \)). Averse-to-risk and risk-lover decision-makers (respectively characterised by \( \theta > 1 \) and \( \theta < 1 \), see Section 1.6) respectively under- and over-estimates the premium for liquidity. Moreover, since \( \Phi^{-1}(1 - \nu^\theta) < 0 \) when \( \nu > 0.5 \), the slope of the capital line turns negative in this case.

Accordingly, and contrary to the common intuition, the capital requirements may decrease with the level of volatility if the liquidity premium is sufficiently high (see the graph to the right lower corner in Figure 1.2). Eventually, when the volatility surpasses a certain level (depending on the
expected return $\mu$ and the premium $\nu > 0.5$) firms prefer to lend all their balances. Indeed, at low volatilities firms that obtain capital profits (with $\mu > 0$) prefer to lend all their balances and do not maintain reserves at all. In fact, they only demand capital when:

$$\sigma \cdot \Phi^{-1}(1 - \nu^\theta) - \mu > 0 \iff \sigma > \frac{\mu}{\Phi^{-1}(1 - \nu^\theta)}$$

This result is consistent with the behaviour of firms that own free-cash-flow (see Section 1.1). On the other hand, firms that obtain capital losses (with $\mu < 0$), choose a level of reserves equal to the mean loss of net assets, i.e. $k = -\mu$, when $\sigma = 0$. 

Fig. 1.2: The Optimal Capital Line and the Allocation Rule of Merton and Perold (1993)
In conclusion, the optimal capital principle determines an \textit{optimal capital line} in the Gaussian case, which relates the optimal proportion of reserves in terms of the mean return and the volatility of the underlying risk, as well as the premium for liquidity offered in the market. The Merton’s principle is obtained as a particular case of the optimal capital line when risks are Gaussian with $\mu = 0$ and $\Phi^{-1}(1 - \nu^{\beta}) = 0.4$, i.e. when $\nu^{\beta} = 1 - \Phi(0.4) \approx 34.25\%$ (see the graph to the left lower corner in Figure 1.2).

Recall that the optimal capital principle has been obtained as an extension of the option-based Merton’s principle — both are based on the net assets’ claim, whose payments at maturity are given in \textit{Equation 1.4}. The optimal capital principle defined in this chapter is thus \textit{meaningful} from the economic point of view.

\section*{1.10 Conclusions}

Firms that continuously trade capital and securities demand no cash reserves, for they can fit their balances at any moment through borrowing and lending (Modigliani and Miller, 1958). In fact, as proved by Merton (1974, 1997), although the market prices of equity and deposit insurance (which are the main components of the capital structure) actually depend on the level of reserves, the value of firms does not depend on it (see also Miller, 1998). This result is a consequence of the fact that continuous hedging suppresses risk.

However, when firms face restrictions on liquidity (in other words, when borrowing and lending may change the price of capital if the transacted amounts break on through certain thresholds) the capital structure is determined by the agency costs and the moral-hazard implicit in the contracts that managers respectively establish with stockholders and customers. Hence the market prices of equity and deposit insurance should be determined on an actuarial basis.

As demonstrated in this chapter, an \textit{optimal} cash balance thereby exists, which leads to an optimal \textit{capital principle} that is consistent with economic fundamentals and is easy to implement for a wide class of probability distributions. Moreover, since the level of capital is explicitly related to the \textit{deductible} or \textit{retention} of the corresponding insurance contract, it explicitly represents the moral-hazard on the side of customers.
In particular, when the underlying risk follows a Gaussian probability distribution, an optimal capital line is obtained relating the optimal proportion of capital to the standard deviation. This principle naturally extends the capital allocation rule proposed by Merton and Perold (1993).

Within multidivisional corporations, differences in skills, information and aversion-to-risk between central and divisional managers implies that central and local administrations do not always agree on the optimal allocation of capital.

The capital principle defined in this chapter can be extended in such a way that a single informational type modify expectations. Differences in expectations can be explicitly measured in this way. Two different allocation principles can be defined on this basis, which are applicable to centralised and decentralised organisations.

In fact, an optimal mechanism can be established for the distribution of capital, which allows subsidiaries to independently choose their capital structures, at the time that leads the conglomerate to collect the same aggregate balance as under the centralised allocation. The implementation of the mechanism forces subsidiaries to reveal their types. In this context, the optimal mechanism can be used as a tool to stimulate the exchange of information between central and divisional administrations.
The primary role of money is to allow the exchange of goods and services in the economy. The transactions motive for holding money is usually justified on these grounds, which claims that the demand for money is in proportion to the volume of transactions, which in turn is considered as proportional to the level of income.

Individuals that hold portfolios containing assets and liabilities with different maturities are obliged to maintain some stock of cash in order to fulfil their outstanding balances. They are accordingly said to demand money for precautionary motives.

Finally, the presence of unknown capital profits and losses (P&L) in the balance sheets of the pursuers of investment projects producing random outcomes causes them to additionally demand cash balances for speculative motives.

A matter of fact, individuals that expect to obtain capital profits prefer to buy securities instead of keeping cash provisions, for in this way they assure to themselves a sure gain. Plenty of cheap credit is likely to be found in markets where such a mood predominate.

By contrast, credit is likely to become scarce and expensive in markets
where most of the public believe their assets will produce capital losses in the near future — people prefer to reduce the exposition to risk in their portfolios and to raise their stocks of reserves in this case, as a means of protection against unexpected shortfalls and bankruptcy.

Within this context, the well-known Keynes’s liquidity preference proposition is enunciated, according to which the demand for cash balances is positively affected by the level of income and negatively affected by the return offered by a certain class of money substitutes (see Keynes, 1937a and 1937b, and also Howells and Bain, 2005).

Two important models at the core of economic theory are connected to the liquidity-preference proposition.

In the first place, the Capital Asset Pricing Model (CAPM), originally developed by William Sharpe (1964, 1966) and John Lintner (1965), establishes the price at which some risky asset must be exchanged under conditions of equilibrium. The derivation of the model depends on the assumption that the expected return and the volatility of every efficient portfolio combining risk and cash must be related to each other according to a linear schedule. The collection of such portfolios is known as the capital market line. The optimal combination of risk and cash, which is determined at the tangency point of intersection between the capital market line and the curve representing the preferences of the decision-maker, ultimately determines the preference for liquidity.\footnote{The liquidity-preference function is first derived in this way by Tobin, 1958. See specially Sharpe, 1964, and also Section 2.2.}

A more explicit role is played by the liquidity-preference function in macroeconomic analysis. Recall that the monetary equilibrium of the economy is determined in such a way that the total demand for cash holdings is equal to the total stock of money supplied by the central bank. Within this context, the liquidity preference function, which explicitly measures the proportion of nominal income that is spent on cash holdings, corresponds to a property of the economy that determines the extent to which monetary interventions affect economic and financial conditions — as described by the level of prices $P$, the real output $y$ and the interest rate $r$.

An alternative theoretical setting will be proposed in this chapter for the characterisation of the preference for liquidity of the economy. The main departures from the classical setting is that in the alternative model national income is regarded as a random variable and people are supposed to face
restrictions when looking for funding in financial markets.

The alternative model is based on the approach of James Tobin, 1958. The money demand is accordingly corresponded to the maximisation of the expected value of some portfolio that contains cash holdings and a mutual fund delivering random payments. The main bibliographical references supporting the model are, on the one hand, Tobin, 1958, and Sharpe, 1964, for the characterisation of liquidity-preference and the CAPM, and Keynes, 1937a, and Friedman, 1970, for the description of the monetary equilibrium and the monetary mechanism (also Tobin, 1947, can be considered in this respect).

2.1 Liquidity-Preference in the Monetary Equilibrium

The liquidity-preference proposition can be represented by the following functional expression (see Equation (6) in Friedman, 1970):

\[ L(r) = Y \cdot \lambda(r) = P \cdot y \cdot \lambda(r) \quad \text{with} \quad \frac{d\lambda(r)}{dr} < 0 \quad (2.1) \]

where \( L(r) \) represents the aggregate cash balance demanded by the economy and the liquidity preference function \( \lambda(r) \) expresses the ratio between demanded cash balances and nominal income. The inverse ratio \( v(r) = 1/\lambda(r) \) is known as the velocity of money. The level of prices \( P \) establishes the connection between real and nominal incomes, respectively denoted as \( y \) and \( Y \), with \( Y = P \cdot y \). Recall that nominal magnitudes represent flows expressed in monetary units, while real quantities are expressed in terms of the goods and services that money can purchase.

The aggregate money supply, on the other hand, refers to the total amount or stock of money held by the public in the economy. It is traditionally related to a class of narrow money denoted as \( M1 \), which mostly contains currency held by non-banking institutions and householders. Other monetary aggregates have been proposed as well, such as \( M2 \), which includes small-denomination time deposits and retail mutual funds, and \( M3 \), which adds mutual funds, repurchase agreements and large-denomination time deposits (see Edwards and Sinzdak, 1997, and also Howells and Bain, 2005).
Letting $M$ denote the total stock of money supplied by the monetary authority, we obtain from Equation 2.1 that at equilibrium the following equation must necessarily hold:

$$M = Y \cdot \lambda(r) = P \cdot \lambda(r) \quad \text{with} \quad \frac{d\lambda(r)}{dr} < 0 \quad (2.2)$$

Therefore, any change in the nominal quantity of money $M$ induces a variation in any of the variables determining the money demand, $P$, $y$ or $r$, in order to reestablish the monetary equilibrium. Since the level of real income $y$ is expressed in terms of goods and services, it is normally assumed to depend on economic fundamentals and hence, it is normally regarded as a stable variable in the short-run. Short-term fluctuations are then expected to mostly affect the level of prices and interest rates.

On these grounds, if the level of prices and real output were pegged to some determined paths of variation (respectively corresponded to some determined rates of inflation and growth), the monetary authority would be able to provide, in principle, the amount of money that is consistent (in the sense that Equation 2.2 is satisfied) with some target level of the interest rate.

As a matter of fact, central banks set the short-term interest rate at which institutions exchange overnight balances in primary markets, as explained in details by Edwards and Sinzdak, 1997. Other financial institutions have to rely instead on secondary markets, where (except for minimum requirements and regulations) financial conditions are exclusively established according to market competition. The transmission mechanism of monetary policy is supposed to affect interest and exchange rates in the first place, together with asset prices and investors’ confidence. These factors later influence both the domestic and external demand, which eventually determine the level of prices and other relevant macroeconomic variables, such as output and employment. Thus the rate fixed by the monetary authority in primary markets ultimately affects the rate at which firms, investors and householders obtain credit to fund investment projects and consumption (see e.g. Romer, 1996).

The efficacy of the monetary mechanism depends, however, on how much of the response of the economy is performed through adjustments in the level of prices $P$, and how much is performed by modifying the demand for balances.
Indeed, assuming that the demand for money is *perfectly elastic* (in other words, assuming that individuals are satisfied at a single level of the interest rate) implies that the amount of money can vary while both the levels of nominal income and interest rates remain unchanged. Under such circumstances, expansions and contractions of the money supply must be respectively followed by increments and reductions of the same magnitude in the stock of cash, in such a way that the monetary mechanism proves to be *useless* for dealing with short-run fluctuations. The preference for liquidity is said to be *absolute* in this situation.²

By contrast, if liquidity-preference is *non-absolute*, every change in the money stock affects (at least partially) the level of nominal income — in such a way that every monetary expansion and every monetary contraction respectively stimulates and contracts the level of nominal output in the short-run.³

The main difficulty faced by monetary authorities when applying the monetary mechanism in practice is the lack of a well-established functional expression characterising the preference for liquidity of the economy.

In this respect, many researchers give it for granted that the *rate of variation* of the demand for cash balances with respect to the interest rate is a constant.⁴ Accordingly, *log-log* and *semi-log* functions (such that \( \lambda(r) = A \cdot r^{-\eta} \) and \( \lambda(r) = B \cdot e^{-\epsilon \cdot r} \) respectively, where \( A \) and \( B \) are constants) are normally used in empirical investigations of the money demand.⁵

²Keynes (1937a) and his disciples emphasise that the money demand must be *perfectly elastic* when individuals share expectations about the level of the interest rate. In this situation, *firmly* convinced investors will absorb any increment or reduction of the stock of money without changing their perceptions about the level of interest rates. Variations in the amount of money must be totally transmitted to the demand for balances in this case (see also Tobin, 1947).

³Any monetary expansion then leads to a new equilibrium involving higher prices for the same quantity, the higher this response the more inelastic the money demand. In short-terms, production is encouraged until prices are reestablished to their original levels. In the long-run, new producers enter the market and existing plants are expanded. Throughout the process, it may take time for output to adjust, but no time for prices to do so. See Friedman, 1968, 1970.

⁴Later in Sections 2.6 and 2.8, this rate will be corresponded to the *semi-elasticity* of the demand for balances, equal to the percentage change in cash holdings due to a point of variation of the interest rate.

⁵Such functional expressions can be justified on the grounds of a model of general equilibrium, where people allocate their funds to cash holdings and consumption. The money demand is derived in this framework by maximising the utility of a representative agent. See Lucas, 2000, and Holmstrom and Tirole, 2000.
Although these specifications lead to a satisfactory description of the money demand for the most of the recorded paths of monetary aggregates and interest rates, there are times when their predictions have failed to anticipate the actual liquidity needs of the economy. Multiple revisions of the model have intended to explain these results, but there is still no agreement on the subject, and there is still no alternative theoretical setting that can simultaneously incorporate all the scenarios observed during the last forty years.\footnote{The classic reference on this issue is Goldfeld, 1976, who examines the failures of the model of money demand occurred during the 1970s. Duca, 2000, and also Teles and Zhou, 2005, investigate whether a stable specification can be obtained if alternative monetary aggregates are considered. Choi and Oh, 2003, on the other hand, propose a model of utility maximisation that incorporates output uncertainty. Calza and Sousa, 2003, examine the effects of additional variables, such as the degree of aggregation of national income.}

In the theoretical setting that will be soon presented, it is assumed that the level of income $Y$ is a \textit{random} variable and hence, that individuals do not know with certainty the level that this variable will take in the future. However, they can observe the series of percentage income returns and estimate its parameters with respect to some class of \textit{probability distributions}. It is possible to prove, within this framework, that an \textit{optimal liquidity principle} exists, which explicitly depends on the \textit{riskiness} of national income (see Sections 2.3 and 2.4). This implies, in the first place, that the stock of money determined by the central bank is not corresponded to a unique level of the interest rate (as stated in Section 2.5), and in the second place, that the liquidity-preference is not necessarily a \textit{linear} function of the interest rate (as shown in Section 2.6). The consequences of these results to macroeconomic analysis are presented in Section 2.8.

First in Section 2.2, the model of liquidity-preference of James Tobin will be presented, which is taken as a reference (and a comparison basis) for the construction of the optimal liquidity principle in Section 2.3. The basic idea is that the preference for liquidity is determined by an optimal combination of a certain \textit{risky} fund and some \textit{non-risky} security.
2.2 Preference for Liquidity as Behaviour towards Risk

The theory of liquidity-preference, as stated by Tobin, 1958, is exclusively concerned with the problem of building efficient portfolios combining two different kind of financial products: some risky aggregate exposure (delivering some random payment at the maturity date) and a certain non-risky security (which provides some cash flow that is known with certainty at any moment before the instrument expires).

Non-risky securities are related to a class of monetary assets, with no risk of default, which offer some fixed return delivered at the maturity date of the instrument. Cash holdings and non-risky bonds belong to this class. The class of risky assets, on the other hand, contains individual securities as well as diversified portfolios and mutual funds. Every portfolio is supposed to be efficient, in the sense that it maximises the utility attained by its holder — in other words, every portfolio is built after the utility maximisation approach of Harry Markowitz, 1952. In this setting, portfolio decisions are taken before the level of cash reserves is decided.

In order to formally establish a problem leading to the optimal stock of cash, both the notion of risk and the preferences of individuals are characterised in mathematical terms.

In the model of Tobin, risks are uniquely corresponded to probability distributions. More specifically, individuals are supposed to assess the riskiness of investments based on the empirical frequencies of the price movements of the alternative securities. The series of price returns are additionally supposed to follow Gaussian probability distributions. Therefore, every risk is completely characterised by a unique pair of expected return and volatility, in such a way that if the market participants share their expectations about the future performance of securities, every portfolio is represented by a single pair of expected return and volatility in the market. Under such conditions, it is possible to prove that the mean return and the volatility of every assets backed portfolio (as we will refer in the following to every portfolio that combines some risky asset with a certain cash balance) must be related to each other according to a linear schedule.

Indeed, consider some portfolio that allocates a proportion \( \lambda \) of wealth to some non-risky bond offering the return \( r_0 \) per unit of investment, and the rest to a certain mutual fund providing the aggregate percentage profit and
loss $X$. Then the *capital return* of the *assets backed* portfolio at maturity, per unit of wealth, is determined by the random variable $Z = (1 - \lambda) \cdot X + \lambda \cdot r_0$, whereas its *expected return* is given by:

$$\mu_Z = E[Z] = (1 - \lambda) \cdot \mu_X + \lambda \cdot r_0$$  \hfill (2.3)

Besides, the volatility of the *assets backed* portfolio $Z$ can be expressed in terms of the volatility $\sigma_X$ of the risky fund:

$$\sigma_Z = \sqrt{E[(Z - \mu_Z)^2]} = (1 - \lambda) \cdot \sigma_X$$

Solving for $\lambda$ in both equations, a linear relationship is established between the expected return and the volatility of every *assets backed* portfolio:

$$\mu_Z = r_0 + \frac{\mu_X - r_0}{\sigma_X} \cdot \sigma_Z$$  \hfill (2.4)

This relationship determines the set of *efficient* portfolios, in the sense that for any combination contained in the set of investment opportunities and outside the line, it is always possible to build a new fund providing the same expected return and a lower risk, or the same risk but a higher return. Accordingly, only increasing the borne risk is possible to raise the expected return of the portfolio.\(^7\)

The locus of portfolios satisfying *Equation 2.4* in the plane $(\mu_Z, \sigma_Z)$ is known as the *capital market line* (Sharpe, 1964). The slope of the curve is regarded as the *market price of risk*, for it determines the rate at which a unit of expected return is exchanged by a unit of risk in the market:

$$S_X = \frac{\mu_X - r_0}{\sigma_X}$$  \hfill (2.5)

The coefficient $S_X$ is also known as the *Sharpe ratio*. Since the expected return and the volatility of the price returns of securities are *observable* variables — which can be estimated based on historical figures — the Sharpe ratio can be empirically determined (Sharpe, 1966).

\(^7\)Tobin considers a class of *pure* cash instruments, with $r_0 = 0$. See *Equation (3.4)* in the paper of Tobin, 1958.
2.2 Preference for Liquidity as Behaviour towards Risk

Regarding the preferences of decision-makers, in the model of Tobin they are represented by utility functions depending on the return of the portfolio $Z$, which satisfy the axioms of Von Neumann and Morgenstern (1944). Therefore, given any level of expected utility:

$$E[U(Z)] = \int U(z) \, dF_Z(z)$$  \hspace{1cm} (2.6)

an indifference curve is determined in the plane $(\mu_Z, \sigma_Z)$, containing all the portfolios that provide the expected utility $E[U(Z)]$, characterised in such a way that for a certain function $\varphi$:

$$\varphi(\sigma_Z, \mu_Z) = E[U(Z)]$$

Notice that the indifference curves of risk-lover decision-makers must show negative slopes, for such individuals are always willing to accept a lower expected return as long as there is a chance to obtain additional profits. Averse-to-risk decision-makers, on the other hand, do not accept to increase the assumed risk unless they are compensated by a greater expected return and consequently, their indifference curves have positive slopes. Besides, as long as more is regarded as better, the indifference curves located to the upper left corner of the plane are related to higher utilities. An implicit relationship is ultimately determined, which expresses the expected return of every hedged portfolio in terms of its volatility, i.e. $\mu_Z = \mu(\sigma_Z)$.

Within this theoretical setting, every rational decision-maker must choose, among those portfolios contained in the market capital line (Equation 2.4), the combination of risk and cash that maximises her or his expected utility. Such combination is determined at the point of tangency between the line of efficient portfolios (which represents the frontier of the set of investment opportunities) and the indifference curve representing the individual’s preferences (see also Sharpe, 1964).

---

8This means, in particular, that the preferences of averse-to-risk individuals are represented by concave utility functions (which satisfy $d^2U(z)/dz^2 < 0$, $\forall z$), whereas the preferences of risk-lovers are represented by convex utility functions (which satisfy $d^2U(z)/dz^2 > 0$, $\forall z$). In this context, averse-to-risk individuals must receive a greater expected return than risk-lovers in compensation for every additional unit of risk. Moreover, imposing that more is always regarded as better, implies that the marginal utility must by positive over the whole range (such that $dU(z)/dz > 0$, $\forall z$), both for averse-to-risk and risk-lover individuals.
In other words, the optimal portfolio $Z$ is determined at the point where the slope of the tangent to the indifference curve is equal to the slope of the capital market line:

$$\frac{d\mu(\sigma_Z)}{d\sigma_Z} = \frac{\mu_X - r_0}{\sigma_X}$$

In this way, an expression for the liquidity preference schedule in terms of the risk-free interest rate, $\lambda = \lambda(r_0)$, can be obtained. However, a tangency point of intersection between the capital market line and some indifference curve will only occur if the later has a positive slope, i.e. if $d\mu(\sigma_Z)/d\sigma_Z > 0$. As already stated, this is only true in the case of averse-to-risk individuals, for risk-lovers are precisely characterised by indifference curves with negative slopes. On these grounds, Tobin (1958) regards liquidity-preference as behaviour towards risk.

In conclusion, the utility maximisation approach provides a well established theoretical setting to derive the demand for cash holdings as a function of the interest rate. A model for the pricing of financial securities under conditions of equilibrium is built on this basis, which is known as the Capital Asset Pricing Model in the literature (abbreviated as CAPM, see Sharpe, 1964 and 1966, and also Lintner, 1965).

Two conditions must be necessarily satisfied: decision-makers must show aversion-to-risk, and the transactions of cash and securities must be carried out under conditions of perfect competition, usually enunciated in the following terms (see e.g. Sharpe, 1964). (PM1) The series of capital returns of the security prices follow Gaussian probability distributions. Hence only two measures completely describe risks: the expected return and the volatility of the series of capital returns, respectively corresponded to the mean return and the standard deviation of the underlying series of capital profit and losses. (PM2) Lending and borrowing are allowed at any moment for a common risk-free interest rate, at least up to some desired extent. (PM3) At any point of time, investors share expectations concerning the future performance of securities and thus portfolios.

---

9Tobin actually distinguishes between two kind of averse-to-risk individuals: those characterised by convex indifference curves, which satisfy $d^2\mu(\sigma_Z)/d\sigma^2_Z > 0$, and plungers, who are characterised by concave indifference curves, which satisfy $d^2\mu(\sigma_Z)/d\sigma^2_Z < 0$. The former include both cash and risk in their hedged portfolios, while the later maintain all their wealth in cash. See Figures 3.1, 3.2 and 3.3 in the paper of Tobin, 1958.

10Some recent contributions where this approach is adopted are those of Holmstrom and Tirole (2000), Lucas (2000) and Choi and Oh (2003).
2.2 Preference for Liquidity as Behaviour towards Risk

Some consequences of the utility maximisation approach, however, are not fully convincing economically speaking. For example, drastic state transitions are sometimes observed in capital markets, manifested as drastic variations in the level of the interest rate and the risk-parameters $\mu$ and $\sigma$, which ultimately induce adjustments in the market prices of securities, as deduced from Equation 2.5. These transitions can be only the consequence of sudden changes in the expectations of individuals, for these (apart from historical information) are the only determinants of their estimations of the risk-parameters. But drastic expectations and price adjustments are difficult to explain on the grounds of financial and economic theory.\(^{11}\)

On these grounds, many researchers have questioned the hypothesis (PM1) of perfect competition. They have pointed out instead that drastic state transitions do not necessarily require of drastic adjustments in the underlying risk-parameters if the series of capital returns are statistically modelled by means of heavy tailed probability distributions. As a matter of fact, heavy tailed distributions assign greater probability to big price movements than the Gaussian — whilst the Gaussian assigns greater probability to small price movements. Hence, big price movements do not necessarily correspond to structural adjustments when risks are modelled by heavy tailed probability distributions.\(^{12}\) Unfortunately, such models have not been satisfactorily integrated with economic theory and accordingly, the paradigm of perfect markets has predominated.

Finally, notice that the hypothesis (PM2) and (PM3) of perfect competition are crucial to guaranteeing the existence of the market equilibrium in the \textit{CAPM}.

Indeed, recall that every assets backed portfolio $Z$ is related to a single proportion of cash $\lambda$. Hence assuming that every portfolio in the capital market line can be attained by performing the appropriate transactions of securities and cash balances, necessarily implies that such transactions can be performed at any moment and without restrictions in capital markets. Conversely, if hedging were only possible up to some extent, some portfolios satisfying Equation 2.4 would require of lending or borrowing operations involving amounts that are not available in the market.

\(^{11}\)It is difficult to accept, in particular, that such adjustments reflect the behaviour of efficient markets, as every market is assumed to be according to the efficient markets hypothesis (in the terms that is formulated by Fama, 1970 and 1998).

\(^{12}\)Mandelbrot (1963) is among those that first attempted to introduce heavy tailed distributions for the statistical characterisation of the movements of stock prices. Merton (1976) and Cox and Ross (1976) reformulate the Black and Scholes’ option pricing model, in order to consider stochastic processes with jumps.
The hypothesis (PM3), on the other hand, implies that individuals agree on the estimations of the risk-parameters $\mu$, $\sigma$ and hence, on the market prices of securities. Only if this hypothesis is satisfied the transactions of assets can be performed at a unique price in the market, in such a way that the market is found at equilibrium (Sharpe, 1964).

The alternative model of equilibrium that will be presented in the following sections is built on a framework of imperfect competition — where hypotheses (PM1), (PM2) and (PM3) are not satisfied. Consequently, the set of risks will be corresponded to a general class of probability distributions. It will be additionally assumed that the transactions in the markets of assets and cash holdings are only possible if the quantities involved do not surpass certain limits and that individuals do not necessarily agree on their expectations about the future performance of securities.

The influence of liquidity restrictions in the funding strategies followed by decision-makers can be described by considering that the only substitute to borrowing in capital markets, apart from cash holdings, is deposit insurance. This approach is suggested by Robert Merton (1974, 1977) for the pricing of liability guarantees. Merton, however, assumes that individuals can trade securities and cash balances at their will — in other words, he assumes that investors can hedge continuously — concluding that the price of every guarantee must be equal to the price of a put option on the value of the underlying claim.

The Merton’s model of deposit insurance can be naturally extended if the price of the guarantee is related to the actuarial price of the underlying residual exposure — equal to the excess of loss over the level of reserves. As demonstrated later in Section 2.3, then an optimal surplus exists, which ensures that the value of the assets backed portfolio is maximised. The optimal liquidity principle thus obtained can be naturally aggregated to account for the preference for liquidity of markets, economic sectors and the economy as a whole, as shown in Section 2.4. On these grounds, the monetary equilibrium of the economy can be characterised, see Section 2.5.

As shown in Sections 2.6 and 2.7, an alternative to the CAPM’s characterisation of market equilibrium is obtained applying this model to the

---

13It can be actually demonstrated that the expected value of the excess of loss satisfy a set of basic mathematical properties and hence, that it can be regarded as a fair insurance price, see Goovaerts et al., 1984. In Dhaene et al., 2003, and Goovaerts et al., 2005, this principle is used as a tool for allocating capital inside financial institutions. In fact, a model of economic capital can be formally established on this basis, which I present in details in Mierzejewski, 2006a, 2006b, 2008a and 2008b.
particular case of Gaussian risks and homogeneous expectations. The model also leads to an extended approach to the monetary equilibrium (see Section 2.8). Market vulnerabilities, manifested as peaks in the sensibility of the liquidity-preference function with respect to the interest rate, are then consistent with some market scenarios, and can then be regarded as natural transitions in markets that face borrowing restrictions.

2.3 The Optimal Liquidity Principle

Let the parameter \( \theta \) denote the state of information of some firm or individual investor holding a mutual fund whose percentage return is represented by the random variable \( X = \Delta Y/Y \), where \( Y \) denotes the level of income of the fund. Because of the precautionary motive, a guarantee \( L \) is maintained until maturity in order to avoid bankruptcy, whose magnitude, on account of the transactions motive, is expressed as a proportion \( \lambda \) of the level of income, i.e. \( L = Y \cdot \lambda \). In the following, this surplus will be treated as an additional liability that induces the cost \( r_0 \cdot L \).

The total payment per unit of investment, delivered by the assets backed portfolio (which combines the risky fund \( X \) and the guarantee \( L \)) at maturity, is then equal to the claim \( Z = X - \lambda - r_0 \cdot \lambda \), in such a way that the expected return \( Y \cdot \mu_{\theta,Z} \) of the assets backed portfolio is given by:

\[
Y \cdot \mu_{\theta,Z} = Y \cdot E_{\theta}[Z] = (Y \cdot \mu_{\theta,X} - L) - r_0 \cdot L = Y \cdot [ (\mu_{\theta,X} - \lambda) - r_0 \cdot \lambda ]
\]

Financing decisions are thereby affected by the percentage return on income:

\[
\mu_{\theta,Z} = E_{\theta}[Z] = (\mu_{\theta,X} - \lambda) - r_0 \cdot \lambda 
\]  

(2.7)

Comparing Equations 2.3 and 2.7 we notice that the rules determining the assets backed portfolio in the derivation of the CAPM in Section 2.2 are different from the prescription considered in the alternative model.

Indeed, as established in Equation 2.3, in the former setting the proportion \( \lambda \) simultaneously determines the amount of funds allocated to risk and cash holdings. As long as any combination of assets and cash balances
can be attained in the market, any proportion $\lambda$ is corresponded to some portfolio that can be built by performing the appropriate transactions.

In order to incorporate the possibility that some combinations cannot be attained due to liquidity restrictions, in Equation 2.7 the exposition to risk is fixed, although individuals can modify their cash holdings by borrowing or lending at the interest rate $r_0$. Assuming this setting makes sense if the portfolio $X$ is regarded as a non-standardised fund that cannot be continuously transacted in the market. The holders of such portfolios are obliged to perform a complete reallocation of resources if they want to change their exposition to risk — in other words, they are obliged to implement again the Markowitz’s procedure to find a new optimal portfolio. Such adjustments are seen as structural changes by creditors, which might lead to increments or reductions in the market price of the fund. These price returns, in turn, might eventually lead to changes in the premiums (over the risk-free interest rate) the holders of the fund have to pay to borrow in the markets of cash balances.\(^{14}\)

Maximising the expected return of the assets backed portfolio as defined in Equation 2.7 actually leads to the trivial solution $\lambda = 0$, because in this case demanding cash holdings only produces an additional loss. To obtain this result, the hypothesis is implicit that individuals are indifferent between holding positive or negative balances.

However, if individuals face liquidity restrictions, transacting positive and negative balances might induce to some net profit or loss.

In fact, the return obtained under such circumstances can be explicitly measured in terms of the expected values of the claims $(X - \lambda)_+ = max(0, X - \lambda)$ and $(X + \lambda)_- = -min(0, X + \lambda)$, which respectively represent the surplus and the excess of loss with respect to the current stock of cash. Then the expected return (per unit of income) of the assets backed portfolio should be written as:

$$\mu_{\theta,Z} = E_\theta [(X - \lambda)_+] - E_\theta [(X + \lambda)_-] - r_0 \cdot \lambda = \Delta(\lambda) - r_0 \cdot \lambda$$  \quad (2.8)

\(^{14}\)According to Billet and Garfinkel, 2004, such premiums depend explicitly on the difference between the costs of internal and external financing, and thereby reflect the degree of financial flexibility of the institution. Thus, institutions with greater flexibility have access to cheaper funding sources, have greater market values and carry less cash holdings. Kashyap and Stein, 2000, analyse the effects of monetary policy over financial decisions under such circumstances.
2.3 The Optimal Liquidity Principle

The term $\Delta(\lambda) := E_\theta[(X - \lambda)_+] - E_\theta[(X + \lambda)_-]$ represents the economic margin obtained because of financial intermediation, while $E_\theta[(X + \lambda)_-]$ accounts for the cost of assuming bankruptcy, a role that can be adopted by the own investor, an insurance company or some governmental division.\(^{15}\)

From the actuarial point of view, the terms $E_\theta[(X - \lambda)_+]$ and $E_\theta[(X + \lambda)_-]$ represent the fair or actuarial prices of the corresponding claims. This means that these terms represent the prices at which the underlying exposures $(X - \lambda)_+$ and $(X + \lambda)_-$ should be transacted in some insurance market free of arbitrage (see Goovaerts et al., 1984, Venter, 1991, and Wang et al., 1997).

Within this context, the expected return $\mu_{\theta,Z}$ represents the fair price of the portfolio $Z$ when capital and insurance markets are found at equilibrium (see Mierzejewski, 2008b). Hence, as implied by Equation 2.8, the market value of the assets backed portfolio certainly depends on the proportion of funds $\lambda$ invested on cash reserves. We can thereby postulate that rational decision-makers choose the proportion $\lambda$ in order to maximise the expected return $\mu_{\theta,Z}$ — for in this way they maximise the market valorisation of their portfolios — but the question then arises of under which conditions the existence of such an optimal proportion can be assured.

In order to give an answer to this question, it is necessary, in the first place, to provide an explicit expression for the distorted expectation operator $E_\theta[\cdot]$. For this purpose, let us consider the proportional hazards transformation,\(^{16}\) introduced by Wang (1995) as an insurance principle:

$$E_\theta[X] = \int_x dF_{\theta,X}(x) = \int T_{\theta,X}(x) \ dx \quad \text{with} \quad T_{\theta,X}(x) := T_X(x)^{1/\theta} \quad \forall x \quad (2.9)$$

\(^{15}\)Froot et al., 1993, propose a similar model to characterise the optimal demand for capital, which is also based on the expected values of the positive part of the surpluses of the underlying portfolio. Unlike the model presented in this chapter, however, Froot et al. propose to add some random perturbation to the income of the portfolio, and do not multiply, as suggested in this chapter. Besides, they simultaneously maximise over the level of capital and the level of investment. See also Froot and Stein, 1998, and Froot, 2007.

\(^{16}\)So called since it is obtained by imposing a safety margin to the hazard rate $h_X(x) := d\ln T_X(x)/dx$ in a multiplicative fashion: $h_{\theta,X}(x) = (1/\theta) \cdot h_X(x)$, with $\theta > 0$. Other distortions can be used instead. In the general case, a distortion function is defined over the unit interval, and an axiomatic description is provided for the distorted price (see Wang et al., 1997 and Wang & Young, 1998). Averse-to-risk and risk-lover investors are then respectively characterised by concave and convex transformations. All the analysis that follows is maintained in the same terms under this general setting (see also Mierzejewski, 2006b).
The distorted cumulative and distorted tail probability distribution functions appear in Equation 2.9, respectively defined as $F_{\theta,X}(x) = P_{\theta}\{X \leq x\}$ and $T_{\theta,X}(x) = P_{\theta}\{X > x\}$, with $F_{\theta,X}(x) = 1 - T_{\theta,X}(x)$, $\forall x$. Whenever $\theta > 1$ the expected value of risk is overestimated, and underestimated when $\theta < 1$, in this way respectively accounting for the behaviour of risk-averse and risk-lover investors.

Applying Lagrange optimisation, leads the optimal proportion $\lambda^*$, which maximises the criterion of Equation 2.8, to be characterised by the first-order condition, determined at the point where the derivative of $\mu_{\theta,Z}$ with respect to $\lambda$ is equal to zero:

$$\frac{dE_{\theta}[X - \lambda]_+}{d\lambda} - \frac{dE_{\theta}[X + \lambda]_-}{d\lambda} - r_0 = T_{\theta,X}(\lambda^*) + F_{\theta,X}(-\lambda^*) - r_0 = 0$$

Since $F_{\theta,X}(-\lambda) = P_{\theta}\{X \leq -\lambda\} = P_{\theta}\{-X > \lambda\} = T_{\theta,-X}(\lambda)$, $\forall \lambda$, the following equivalent characterisation is obtained:

$$T_{\theta,-X}(\lambda^*) - T_{\theta,X}(\lambda^*) = r_0$$  \hspace{1cm} (2.10)

The rational liquidity demand is thus determined in such a way that the marginal gain minus the marginal loss on capital (i.e. the instantaneous benefit of liquidity) equals the marginal return of the sure investment. Within this context, the optimal proportion of cash is corresponded to an optimal exchange of a sure return and a flow of probability, and it is the mass accumulated in the tails of the distribution what matters. No explicit relationship is obtained for the cash demand, but some numerical procedure could be implemented to find the solution.

The existence of some optimal proportion $\lambda^*$ can be mathematically assured as long as, for any proportion level below the optimal, i.e. for any $\lambda < \lambda^*$, the expected income per unit of investment, equal to the term

\[ \frac{d}{dy} \int_{u(y)}^{v(y)} H(y,x)dx = \int_{v(y)}^{u(y)} \frac{\partial H(y,x)}{\partial y}dx + H(y,v(y)) \cdot \frac{dv(y)}{dy} - H(y,u(y)) \cdot \frac{du(y)}{dy} \]

the relationship is obtained by noticing that, from Equation 2.9, the following expressions are respectively obtained for the expected surplus and the expected excess of loss: $E_0[(X - \lambda)_+] = \int_0^\lambda (x - \lambda) dF_{\theta,X}(x)$ and $E_0[(X + \lambda)_-] = -\int_{-\infty}^{\lambda} (x + \lambda) dF_{\theta,X}(x)$.
\[ \Delta(\lambda) - r_0 \cdot \lambda \] is an increasing and concave function on the liquidity preference coefficient \( \lambda \). This requirement actually corresponds to the second-order condition of Lagrange optimisation (see Froot et al., 1993). In other words, an optimal proportion of cash exists as long as the following inequalities are simultaneously satisfied:

\[
\frac{d\Delta(\lambda)}{d\lambda} - r_0 > 0 \iff T_{\theta,-X}(\lambda) - T_{\theta,X}(\lambda) > r_0 \quad \forall \lambda < \lambda^* \\
\frac{d^2\Delta(\lambda)}{d\lambda^2} < 0 \iff \frac{d}{d\lambda} T_{\theta,-X}(\lambda) - \frac{d}{d\lambda} T_{\theta,X}(\lambda) < 0 \quad \forall \lambda < \lambda^*
\]

The first inequality implies that, for any given liquidity preference ratio \( \lambda \) lower than the optimal level \( \lambda^* \), the marginal loss due to financial intermediation is greater than the total cost of the guaranty and accordingly, that there are incentives to maintain some cash surplus. The second condition ensures concavity. In fact, recalling that \( T_{\theta,X} = 1 - F_{\theta,X} \), this condition can be written in terms of the density probability distribution \( f_{\theta,X}(x) := \frac{dF_{\theta,X}(x)}{dx} = P_{\theta}\{X = x\} \):

\[
P_{\theta}\{X = \lambda\} < P_{\theta}\{-\lambda\} \quad \forall \lambda < \lambda^*
\]

The second-order condition thereby implies that an optimal liquidity ratio \( \lambda^* \) exists as long as the probability of obtaining a certain capital gain is always lower than the probability of obtaining a capital loss of the same magnitude.

One additional condition has to be satisfied, however, for the optimal cash balance to be determined by Equation 2.10. Indeed, recall that for the market price of the assets backed portfolio to be characterised by the expected return \( \mu_{\theta,Z} \) defined in Equation 2.8, individuals must be able to sell their surpluses at the price \( E_\theta[(X - \lambda)_+] \), in such a way that the benefit they have to resign (in average) for holding the proportion of capital \( \lambda \) is equal to:

\[
r_{\theta,X}(\lambda) = \frac{E_\theta[X_+] - E_\theta[(X - \lambda)_+]}{\lambda} \\
\iff E_\theta[(X - \lambda)_+] = E_\theta[X_+] - r_{\theta,X}(\lambda) \cdot \lambda
\]

Combining Equations 2.8 and 2.11:
\[ \mu_{\theta,Z} = E_\theta [X_+] - E_\theta [(X + \lambda)_+] - (r_0 + r_{\theta,X}(\lambda)) \cdot \lambda \]  

(2.12)

In this context, the return \( r_{\theta,X}(\lambda) \) can be interpreted as an extra premium paid for keeping the balance \( L = Y \cdot \lambda \) as a cash stock, instead of investing it in the mutual fund \( X \).

Equivalently, we can say that the total cost of capital for the holders of the assets backed portfolio is equal to:

\[ r(\lambda) = r_0 + r_{\theta,X}(\lambda) \]  

(2.13)

Since the risk-free interest rate \( r_0 \) does not depend on the cash proportion \( \lambda \), deriving Equation 2.11 with respect to \( \lambda \) and rearranging terms, we obtain that the marginal change of the cost of capital with respect to the proportion of cash can be explicitly calculated:

\[ \frac{dr(\lambda)}{d\lambda} = -\frac{1}{\lambda} \cdot \left( r_{\theta,X}(\lambda) + \frac{dE_\theta [(X - \lambda)_+]}{d\lambda} \right) \]  

(2.14)

Under such circumstances, maximising the expressions of Equations 2.8 and 2.12 lead to the same optimal cash balance. Therefore, only if the cost of capital is determined according to Equations 2.13 and 2.14, the optimal cash balance is characterised in order to satisfy Equation 2.10.

Notice that Equations 2.13 and 2.14 determine the cost of capital as perceived by the holders of the assets backed portfolio. But if we assume that individuals borrow the cash balance \( L \) in some open market of capital, then the cost of capital must reflect the perceptions of lenders.

As a matter of fact, debt can be implemented by issuing a bond promising to pay a certain interest rate \( r \) at maturity. As long as the market regards this deposit as riskier than the risk-free security, the issuers of the bond have to offer some return higher than the risk-free interest rate in order to make it attractive to investors. Hence the condition \( r > r_0 \) must be satisfied. On the other hand, the bond issuers are not willing to pay a premium greater than \( r_{\theta,X}(\lambda) \), for then the alternative of providing these funds themselves (whose cost is measured by the premium \( r_{\theta,X}(\lambda) \)) would be cheaper. Hence, also the condition \( r \leq r_0 + r_{\theta,X}(\lambda) \) must hold.
Provided that the previous conditions are satisfied, the cost of capital $r$ must be determined by the credit quality of borrowers. Consequently, it can be only affected by events that change the perception of investors about the willingness and capability to pay of the bond issuers. It can then be assumed as constant in practice, as long as the issuers of bonds do not drastically change their capital structures — i.e. as long as the proportion of reserves $\lambda$ is not drastically modified.

Replacing the return $r$ in the place of $(r_0 + r_{\theta,X}(\lambda))$ in Equation 2.12, the following expression is obtained for the expected percentage income:

$$\mu_{\theta,Z} = E_{\theta}[X_+] - E_{\theta}[(X + \lambda)_-] - r \cdot \lambda$$  \hspace{1cm} (2.15)

Applying Lagrange optimisation, we obtain that individuals attract funds until the marginal return on risk equals the total cost of holding capital:

$$- \frac{d}{d\lambda} E_{\theta}[(X + \lambda^*)_-] - r = T_{\theta,-X}(\lambda^*) - r = \left[ T_{-X}(\lambda^*) \right]^{1/\theta} - r = 0$$

Equivalently, it can be said that investors stop demanding money at the level at which the marginal expected gain in solvency equals its opportunity cost. The optimal liquidity principle is thereby given by:

$$\lambda_{\theta,X}(r) = T_{\theta,-X}^{-1}(r) = T_{-X}^{-1} \left( r^\theta \right)$$  \hspace{1cm} (2.16)

From this expression, the optimal demand for cash balances always follows a non-increasing and (as long as the underlying probability distribution is continuous) continuous path, whatever the kind of risks and distortions, because the tail probability function, and hence its inverse, are always non-increasing functions of their arguments. The minimum and maximum levels of surplus are respectively demanded when $r \geq 1$ and $r \leq 0$. Besides, averse-to-risk and risk-lover individuals (respectively characterised by $\theta > 1$ and $\theta < 1$) systematically demand higher and lower amounts of cash holdings — for they respectively under- and over-estimates the cost of capital.\footnote{Consequently, one of the main advantages of the actuarial-based liquidity principle defined in Equation 2.16 is its functionality. Such result crucially depends on the choice of the distorted probability insurance principle of Equation 2.9. Indeed, if the expected utility operator (defined in Equation 2.6) were used instead to evaluate the hedged portfolio’s}
2.4 The Aggregate Liquidity-Preference

The aggregate liquidity-preference of some industry or economic sector will be now characterised, where each firm can borrow at a single interest rate $r$. As stated in the previous section, such rate depends on the credit quality of borrowers and is supposed to remain unchanged as long as firms do not drastically alter their capital structures. In other words, firms are supposed to remain in the same credit class (i.e. the return at which firms can borrow in the market is supposed to remain the same) as long as the levels of income and reserves in their portfolios are kept more or less invariant.

Let us additionally assume that firms hold securities and combinations of securities (or are involved in venture projects) producing capital returns represented by the random variables $X_1, \ldots, X_n$. The levels of income and the liquidity preference functions corresponding to each of the funds will be respectively denoted as $Y_1, \ldots, Y_n$ and $\lambda_1(r), \ldots, \lambda_n(r)$. The total surplus accumulated in the industry must then be equal to:

$$Y \cdot \lambda(r) = \sum_{i=1}^{n} Y_i \cdot \lambda_i(r) \quad \text{with} \quad Y = \sum_{i=1}^{n} Y_i$$

where $Y$ and $\lambda(r)$ respectively denote the level of income and the preference for liquidity accumulated in the industry. Dividing by $Y$ we obtain that:

$$\lambda(r) = \sum_{i=1}^{n} \omega_i \cdot \lambda_i(r) \quad \text{with} \quad \omega_i = \frac{Y_i}{Y} \quad \forall i \quad \text{and} \quad \sum_{i=1}^{n} \omega_i = 1 \quad (2.17)$$

Therefore, at any level of the interest rate, the liquidity-preference of the industry is equal to the sum of the liquidity-preferences of the different firms weighted by their relative magnitudes in terms of the levels of income.

return of Equation 2.15, then the first-order condition would lead to:

$$\int_{-\infty}^{-\lambda} u'(x + \lambda) \, df_X(x) - u(0) \cdot f_X(-\lambda) - r = 0$$

where $u'$ denotes the first derivative of the utility function. Then no explicit expression would be obtained for the optimal liquidity principle — except for some restricted class of utility functions.
Notice that the level of aggregation plays no role in Equation 2.17. Indeed, the random variables \( X_1, \ldots, X_n \) could be assumed to represent the capital \( P&L \) of the totality of firms belonging to the class, as well as the aggregates of some predetermined groups or clusters. Then the liquidity-preferences of the different economic sectors could be summed up in order to obtain the preference for liquidity of the economy as a whole. Alternatively, the function \( \lambda(r) \) could be represented in terms of the incomes and the cash balances demanded by each of the individuals participating in it, from householders and small companies, to big holdings and rich private investors. Although the functional specification and the evolution of \( \lambda(r) \) are certainly expected to depend on the level of aggregation, there is no formal difference in applying any of the alternative representations. They are all equivalent characterisations of the same property of the economy.

Actually, from the mathematical point of view, Equation 2.17 can be treated as an invariance condition leading to a certain set of functional specifications. Imposing that the different liquidity-preference functions \( \lambda_1(r), \ldots, \lambda_n(r) \) follow the same functional expression and that this expression is always preserved at different levels of aggregation (whatever the number of components or the relative magnitudes of incomes), necessarily leads to accept only a limited set of functions.

Let me illustrate the meaning of this claim by examining the case when individuals choose their balances according to the liquidity principle of Equation 2.16 and share expectations about the probability distributions describing risks — i.e. they agree on the informational type \( \theta \). Then the aggregate surplus must be equal to the sum of the distorted quantiles of the individual exposures:

\[
\lambda(r) = \sum_{i=1}^{n} \omega_i \cdot T_{\theta, -X_i}^{-1}(r) \quad \text{with} \quad \omega_i = \frac{Y_i}{Y}, \quad \sum_{i=1}^{n} \omega_i = 1
\]

Now define:

\[
\lambda_i = \omega_i \cdot T_{\theta, -X_i}^{-1}(r)
\]

\[
\Leftrightarrow \quad r = T_{\theta, -X_i} \left( \frac{\lambda_i}{\omega_i} \right) = P_\theta \left\{ -X_i > \frac{\lambda_i}{\omega_i} \right\} = P_\theta \left\{ -\omega_i \cdot X_i > \lambda_i \right\}
\]

\[
\Rightarrow \quad \lambda_i = T_{\theta, -\omega_i \cdot X_i}^{-1}(r)
\]
Chapter 2: The Optimal Liquidity Principle

Hence the contributions of firms and individuals to the aggregate liquidity-preference can be equivalently expressed as the optimal principles corresponded to the weighted capital returns \( \omega_1 \cdot X_1, \ldots, \omega_n \cdot X_n \):

\[
\lambda(r) = \sum_{i=1}^{n} T^{-1}_{\theta_i, -\omega_i \cdot X_i}(r)
\]

Therefore, for the aggregate liquidity-preference to be expressed as the quantile of the aggregate capital \( P&L \), we must necessarily impose the sum of the quantiles of the underlying risks to be equal to the quantile of the aggregate exposure.

In fact, as demonstrated by Dhaene et al. (2002), the property of the sum of the quantiles mathematically characterises the *comonotonic dependence structure*. A random vector \((X^c_1, \ldots, X^c_n)\) is said to be *comonotonic* if a random variable \( \zeta \) exists, as well as a set of *non-decreasing* functions \( h_1, \ldots, h_n \), such that the realisation of any joint event is entirely determined by \( \zeta \), i.e.:

\[
(X^c_1, \ldots, X^c_n) = (h_1(\zeta), \ldots, h_n(\zeta))
\]

Hence the realisation of any joint event is uniquely related to some event contingent on the single exposure \( \zeta \). Besides, since the functions \( h_1, \ldots, h_n \) are all *non-decreasing*, all the components of the random vector \((X^c_1, \ldots, X^c_n)\) move in the same direction. On these grounds, it is said that *comonotonicity* characterises an *extreme* case of dependence, when no benefit can be obtained from diversification.

Let \((X^c_1, \ldots, X^c_n)\) denote the random vector described by the same marginal probability distributions as \((\omega_1 \cdot X_1, \ldots, \omega_n \cdot X_n)\) and let \( X^c = X^c_1 + \ldots + X^c_n = \omega_1 \cdot X_1 + \ldots + \omega_n \cdot X_n \) denote the *comonotonic aggregate* (or *comonotonic sum*) of the individual capital returns. Then the quantile \( T^{-1}_{\theta, -X^c} \) of the comonotonic sum is equal to the sum of the quantiles of the weighted exposures \((\omega_1 \cdot X_1, \ldots, \omega_n \cdot X_n)\), in such a way that the preference for liquidity of the economy can be written as:

\[
\lambda(r) = T^{-1}_{\theta, -X^c}(r) = \sum_{i=1}^{n} T^{-1}_{\theta_i, -\omega_i \cdot X_i}(r) \quad \text{with} \quad X^c = \sum_{i=1}^{n} \omega_i \cdot X_i \quad (2.18)
\]
2.4 The Aggregate Liquidity-Preference

The comonotonic aggregate $X^c$ thereby characterises the preference for liquidity in economies where individuals rely on the optimal liquidity principle of Equation 2.16.

When differing expectations are allowed in the economy, the aggregate money demand is given by:

$$
\lambda(r) = T_{\theta_1,\ldots,\theta_n,-X^c}^{-1}(r) = \sum_{i=1}^{n} T_{\theta_i,-\omega_i,X_i}^{-1}(r) = \sum_{i=1}^{n} T_{-\omega_i,X_i}^{-1}(r^{\theta_i}) \quad (2.19)
$$

where $\theta_1,\ldots,\theta_n$ denote the different informational types and $T_{\theta_1,\ldots,\theta_n,-X^c} = \left(\sum_{i=1}^{n} T_{\theta_i,-\omega_i,X_i}^{-1}\right)^{-1}$ denotes the distribution function of the comonotonic sum when the marginal distributions are given by $T_{\theta_1,-\omega_1,X_1},\ldots,T_{\theta_n,-\omega_n,X_n}$.

Comparing Equations 2.18 and 2.19, we observe that there is no formal difference between assuming equal and different expectations: in both cases, the aggregate liquidity-preference is determined by the quantile function of the probability distribution of the sum of the underlying exposures. Moreover, as long as the proportions $\omega_1,\ldots,\omega_n$ and the riskiness of the capital returns $X_1,\ldots,X_n$ remain constant, the instability of both functional expressions are dependent alone on the instability of the expectations firmly maintained by individuals, and not on whether individuals agree or not on these expectations. Hence the difference between the homogeneous and the non-homogeneous expectations settings is not relevant in explaining the instability of the money demand of the economy.\(^{19}\)

Endowed with an expression for the aggregate liquidity-preference of the economy, we can now proceed to characterise the monetary equilibrium when individuals determine their cash holdings according to the optimal liquidity principle defined in Equation 2.16.

\(^{19}\)This conclusion contradicts the Keynes’s argument, that the money demand of the economy must be absolute (and so, that monetary policy is useless) in the case of homogeneous expectations (see Keynes, 1937a and 1937b). As explained later in Section 2.8, the preference for liquidity can indeed be absolute under certain circumstances, but as a consequence of the riskiness of national income.
2.5 The Monetary Equilibrium with the Optimal Liquidity Principle

Replacing Equations 2.16 and 2.18 into Equation 2.2, we obtain that in the case of homogeneous expectations the monetary equilibrium is determined by the following equation:

\[
M = Y \cdot \lambda(r) = Y \cdot T_{\theta, -X^c}^{-1}(r) = Y \cdot T_{-X^c}^{-1} \left( r^{\theta} \right) \tag{2.20}
\]

where \( M \) denote the total stock of money in the economy. Hence both the riskiness of national income (determined by the random variable \( X^c \)) and the market expectations (characterised by the informational type \( \theta \)) explicitly affect the monetary equilibrium.

This means, in particular, that the monetary policy choose by the central bank (characterised by the money supply \( M \)) is not corresponded to a unique level of the interest rate, as obtained from Equation 2.2. In fact, given any money stock \( M \), multiple interest rates can satisfy Equation 2.20, depending on the probability distribution describing the riskiness of national income and the informational type \( \theta \) corresponded to the market expectations.

The influence of expectations over the monetary equilibrium can be actually more precisely described.

Indeed, notice, on the one hand, that since the cost of capital is underestimated in averse-to-risk economies (characterised by \( \theta > 1 \)), the interest rate attained at equilibrium in this case is always greater than the levels attained in risk neutral and risk-lover economies (respectively characterised by \( \theta = 1 \) and \( \theta < 1 \)) for the same money supply \( M \) and the same aggregate exposure \( X^c \).

On the other hand, the level of interest rates attained at equilibrium in risk-lover economies is always lower than the levels attained in risk neutral and averse-to-risk economies, because risk-lover individuals systematically over-estimate the cost of capital. As a consequence, in economies where both the riskiness of the percentage return of national output and the monetary policy implemented by the central bank remain constant, changes in expectations must be necessarily followed by adjustments in the rate of interest.
When individuals maintain different expectations about risks, the equilibrium interest rate depends on the particular combination of the informational parameters $\theta_1, \ldots, \theta_n$.

The other determinant of the monetary equilibrium in Equation 2.20 is the riskiness of national income. In the particular case when some analytical expression is available for the probability function describing the random variable $X^c$, such dependence can be investigated in terms of the underlying risk-parameters.

A careful examination of the model under the different families of probability distributions found in the statistical literature is out of the scope of this thesis. Instead, the case of the Gaussian distribution will be analysed in the following sections. In this way, the classic theoretical framework supporting the CAMP and the classical analysis of the monetary equilibrium will be naturally extended.

Indeed, as established in Section 2.7, an extended version of the capital market line is obtained when the Gaussian quantile function is replaced in Equation 2.20. Later in Section 2.8, an extended theoretical framework for the conduction of monetary policy will be presented, based on the fact that the slope of the money demand (or the semi-elasticity of the preference for liquidity) explicitly depends on the mean return and the volatility of the aggregate exposure $X^c$ when the Gaussian liquidity principle is introduced.

### 2.6 The Gaussian Liquidity Principle

In the particular case when the aggregate percentage return $X$ is represented by a Gaussian probability distribution with mean return $\mu$ and volatility $\sigma$, the optimal liquidity principle takes the form (see Equation 2.16):

$$\lambda_{\mu,\sigma}(r) = \sigma \cdot \Phi^{-1}(1 - r) - \mu \quad (2.21)$$

where $\Phi$ denotes the cumulative probability distribution of a standard Gaussian random variable, whose mean and volatility are respectively equal to zero and one (see e.g. De Finetti, 1975, and also Dhaene et al., 2002):

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp \left( -\frac{y^2}{2} \right) dy \quad \forall x$$
As depicted in Figures 2.1 and 2.2, the Gaussian liquidity principle always follows a decreasing and continuous path, independently of the levels of the risk parameters $\mu$ and $\sigma$ and the informational type $\theta$.

This implies that the derived demand for cash holdings $L(r) = Y \cdot \lambda_{\mu,\sigma}(r)$ always follows a decreasing and continuous path — for every fixed level of income $Y$ — and consequently, that the derived money demand $L(r)$ is well defined.

As depicted in Figure 2.1, the dependence of the Gaussian liquidity principle on the informational type $\theta$ follows indeed the patterns described in Section 2.3 for general probability distributions. Accordingly, given fixed levels of expected return and volatility, and at any level of the interest rate, averse-to-risk individuals (characterised by $\theta > 1$) always demand higher cash balances than neutral or risk-lover individuals. By contrast, risk-lovers individuals (characterised by $\theta < 1$) always prefer to maintain lower surpluses than neutral and averse-to-risk individuals. Besides, ceteris paribus, the size of the cash stock always increases with the informational type $\theta$.

Regarding the dependence of the Gaussian liquidity principle on income uncertainty, notice in the first place, that given any fixed level of volatility, raising the mean return always implies that the demand curve is moved to the left (see the upper graph of Figure 2.2). More specifically, when $\mu < 0$ and when $\mu > 0$ the cash requirement $\lambda_{\mu,\sigma}(r)$ per unit of income respectively increases and decreases with the magnitude of the mean return. Consequently, the amount of reserves always rises with the magnitude of realised expected losses (when $\mu < 0$). By contrast, when positive returns are obtained (and $\mu > 0$), at least part of the losses are cancelled by realised profits, in such a way that the higher the magnitude of the expected capital gain, the lower the required cash balance and vice versa.

Secondly, as depicted in the lower graph of Figure 2.2, for every fixed level of expected return, the slope of the Gaussian liquidity principle always increases with volatility, which means that the higher the variability of the underlying series of percentage returns, the more sensible are individuals to the interest rate and vice versa. This result makes economic sense, as long as the parameter $\sigma$ measures the variability (and hence the riskiness) of income. Moreover, as a consequence of the symmetry of the Gaussian distribution, all the demand curves intersect at the point $r = 0.5$ in the lower graph of Figure 2.2. At this level, there is an equal chance of obtaining a capital gain or a capital loss, no matter the level of volatility, and hence the same balance is demanded, equal to the expected value of the fund.
In fact, the sensibility of the Gaussian liquidity principle with respect to the interest rate can be explicitly measured by the semi-elasticity of the Gaussian liquidity-preference function with respect to the interest rate, equal to the percentage variation in the proportion of reserves with respect to the interest rate. From Equation 2.21, the following expression is obtained for this coefficient:

\[
\eta(r, \mu, \sigma) = \frac{1}{\lambda_{\mu,\sigma}(r)} \cdot \frac{d\lambda_{\mu,\sigma}(r)}{dr} = -\sqrt{2\pi} \cdot \Phi^{-1}(1 - r) - \frac{\mu}{\sigma} \cdot \exp \left( \frac{[\Phi^{-1}(1 - r)]^2}{2} \right)
\]

(2.22)

The sign of the variation thereby depends on the interest rate and the risk parameters \( \mu \) and \( \sigma \), in such a way that:

\[
\eta(r, \frac{\mu}{\sigma}) < 0 \iff \Phi^{-1}(1 - r) - \frac{\mu}{\sigma} > 0 \iff \lambda_{\mu,\sigma}(r) > 0
\]

\[
\eta(r, \frac{\mu}{\sigma}) > 0 \iff \Phi^{-1}(1 - r) - \frac{\mu}{\sigma} < 0 \iff \lambda_{\mu,\sigma}(r) < 0
\]

Therefore, a curve with negative slope (as every demand curve should look like, according to classic economic analysis) is always obtained, no matter the risk and informational parameters (as depicted indeed in Figures 2.1 and 2.2).
Regarding the magnitude of the semi-elasticity, it can be easily verified from Equation 2.22 that:

\[
\left| \Phi^{-1}(1 - r) - \frac{\mu}{\sigma} \right| \uparrow +\infty \quad \text{and} \quad \left| \Phi^{-1}(1 - r) \right| < +\infty \implies \left| \eta(r, \frac{\mu}{\sigma}) \right| \downarrow 0 \\
\left| \Phi^{-1}(1 - r) - \frac{\mu}{\sigma} \right| \downarrow 0 \quad \text{or} \quad \left| \Phi^{-1}(1 - r) \right| \uparrow +\infty \implies \left| \eta(r, \frac{\mu}{\sigma}) \right| \uparrow +\infty
\]

Hence, the semi-elasticity of the Gaussian liquidity principle is actually undefined when \( \Phi^{-1}(1 - r) = \mu/\sigma \), since it converges to magnitudes with opposite signs depending on whether the term \( \Phi^{-1}(1 - r) - \mu/\sigma \) approaches to zero from the right or from the left:

\[
\eta(r, \frac{\mu}{\sigma}) \downarrow -\infty \quad \text{when} \quad \Phi^{-1}(1 - r) - \frac{\mu}{\sigma} \downarrow 0 \\
\eta(r, \frac{\mu}{\sigma}) \uparrow +\infty \quad \text{when} \quad \Phi^{-1}(1 - r) - \frac{\mu}{\sigma} \uparrow 0
\]

On these grounds, we can say that the condition \( \Phi^{-1}(1 - r) = \mu/\sigma \) determines a critical point, since at this point the sign of the liquidity principle is undefined and its magnitude converges to infinite. Liquidity-preference becomes absolute under such circumstances.

Let us finally verify whether the sum of income returns preserves the Gaussian liquidity principle. Indeed, consider a series of Gaussian exposures \( X_1, \ldots, X_n \) with means \( \mu_1, \ldots, \mu_n \) and volatilities \( \sigma_1, \ldots, \sigma_n \). Let the individual and aggregate income levels be respectively denoted by \( Y_1, \ldots, Y_n \) and \( Y \), with \( Y = Y_1 + \cdots + Y_n \).

Replacing the liquidity principles \( \lambda_1(r), \ldots, \lambda_n(r) \) according to Equation 2.21, we obtain that the optimal individual cash balances are given by:

\[
L_i(r) = Y_i \cdot \lambda_i(r) = Y_i \cdot \left[ \sigma_i \Phi(1 - r) - \mu_i \right] \quad \forall i = 1, \ldots, n
\]

and summing up the individual cash contributions, the following expression is obtained for the aggregate cash balance:

\[
L(r) = \sum_{i=1}^{n} L_i(r) = Y \cdot \sum_{i=1}^{n} \omega_i \cdot \left[ \sigma_i \Phi(1 - r) - \mu_i \right] \quad \text{with} \quad \omega_i = \frac{Y_i}{Y} \quad \forall i
\]
2.6 The Gaussian Liquidity Principle

Fig. 2.2: The Gaussian Liquidity Principle with Different Combinations of Mean Returns and Volatilities

Hence the aggregate Gaussian liquidity principle is equal to the optimal liquidity principle related to a Gaussian exposure whose mean and volatility are respectively given by the weighted average means and volatilities:

\[ \mu = \sum_{i=1}^{n} \omega_i \cdot \mu_i \quad \text{and} \quad \sigma = \sum_{i=1}^{n} \omega_i \cdot \sigma_i \]  \hspace{1cm} (2.24)

Dhaene et al. (2002) actually demonstrate that the comonotonic sum of Gaussian random variables is also a Gaussian random variable, whose mean and volatility are defined as in Equation 2.24. Then the aggregation of the Gaussian liquidity principle according to Equations 2.21 and 2.24 is
consistent with the aggregation condition established in Equation 2.17.

2.7 The Capital Market Line Extended

Let us consider some industry whose aggregate capital return is represented by a Gaussian random variable $X$ with mean return $\mu$ and volatility $\sigma$. As stated in Sections 2.4 and 2.6, the optimal cash balance demanded at the aggregate level must then be given by:

$$L_{\mu,\sigma}(r) = Y \cdot \lambda_{\mu,\sigma}(r)$$

where $Y$, $r$ and $\lambda_{\mu,\sigma}$ respectively denote the level of income produced by the industry, the cost of capital and the aggregate Gaussian liquidity principle (as determined by Equation 2.21).

Let us additionally assume that when covering their short-term imbalances firms and investors rely on some secondary market of capital, and let $M$ denote the total cash balance supplied by lenders. Since the aggregate balance demanded by the industry must be necessarily equal to the total money stock $M$, the level of the interest rate must adjust at equilibrium in order to satisfy the condition $M = L_{\mu,\sigma}(r)$, i.e.:

$$M = Y \cdot \lambda_{\mu,\sigma}(r) \iff m := \frac{M}{Y} = \sigma \cdot \Phi^{-1}(1-r) - \mu$$

where $m = M/Y$ denotes the relative money supply or the relative stock of cash in the industry. Rearranging terms leads to the equation:

$$\mu = -m + \Phi^{-1}(1-r) \cdot \sigma \quad (2.25)$$

Therefore, the equilibrium in the market of cash balances implies that the mean return and the volatility of the underlying fund must be related to each other according to a linear schedule.

---

20 Later in Section 2.8 the monetary equilibrium will be investigated, where $Y$ represents the national output, $\mu$ and $\sigma$ respectively represents the mean return and the volatility of the series of capital returns of $Y$, and $M$ represents the stock of money determined by the central bank.
Recall that the balance $L_{\mu,\sigma}(r)$ has been defined as the sum of the surpluses preferred by individuals who seek to maximise the expected return of their assets backed portfolios (as determined by Equation 2.15). In other words, the demand function $L_{\mu,\sigma}(r)$ corresponds to the sum of the stocks of reserves maintained by individuals that build efficient portfolios — i.e. that choose efficient combinations of risk and cash holdings.

On these grounds, Equation 2.25 can be regarded as an alternative relationship to the capital market line (abbreviated CML in the following) presented in Equation 2.4. It will then be known as the capital market line extended (abbreviated CML-extended) in the following. Some important discrepancies regarding the interpretation of the variables and parameters that appear in Equations 2.4 and 2.25 should be pointed out, however.

In fact, notice in the first place that while in Equation 2.4 the risk-parameters $(\mu_Z,\sigma_Z)$ of the assets backed portfolio are related to each other, the risk-parameters $(\mu,\sigma)$ of the underlying fund, which in Equation 2.4 are denoted as $(\mu_X,\sigma_X)$, are related to each other in Equation 2.25. This fact reflects a fundamental disagreement between both theoretical frameworks. Indeed, while the rate of interest and the risk-parameters of the underlying series of capital returns are regarded as exogenous variables in the CML, which are supposed to remain unchanged at least in short-terms, the same variables are endogenously determined in the CML-extended, in order to equalise the incoming and outgoing cash flows. In other words, the CML and the CML-extended are respectively corresponded to static and dynamic approaches to the market equilibrium.

Secondly, recall that in the CML the cost of capital is corresponded to the return $r_0$ offered by a class of risk-free securities, which is supposed to remain unaltered as long as individuals are price takers and their preferences — as well as the market conditions in general — are more or less stable. In the CML-extended, on the other hand, the cost of capital is equal to the risk-free interest rate plus some liquidity premium, established by creditors to compensate for the additional risk of default (see Equation 2.13). Since such premium explicitly depends on the risk-parameters $\mu,\sigma$ when the underlying exposure follows a Gaussian probability distribution, from now on we will denote it as $r_{\mu,\sigma}$.

Having explained the meaning of the variables involved in the CML and the CML-extended, let us now analyse how the line is determined in both settings. We must then analyse the slope of the line and its intercept with the mean return’s axis.
We already know that the intercept of the CML is equal to the risk-free interest rate \( r_0 \). This is consistent with the fact that in the CAMP individuals can always lend their balances to some free-of-default counterpart to obtain the return \( r_0 \), independently of the size of the loan.

The intercept of the CML-extended with the mean return axis, on the other hand, is equal to the additive inverse of the relative stock \( m \). This result is connected to the fact that in the extended model the level of the interest rate may actually be affected when the amount of demanded balances surpasses certain levels. In this context, the relative stock of money (equal to the level of reserves per unit of income) plays the role of a compensation for the expected capital loss of the nominal income.

In fact, in the particular case when \( \sigma = 0 \), we obtain that:

\[
\mu + m = 0
\]

This equation explicitly establishes a balance of expected return and cash reserves. It reflects the fact that when \( \mu < 0 \) the relative stock \( m \) can be used to pay back at least some of the realised losses. When \( \mu > 0 \), by contrast, a pressure exists to sell every outstanding cash balance.

Regarding the slope of the line of efficient portfolios, recall that it determines the rate at which a unit of volatility is exchanged by a unit of expected return in the market, or equivalently, it determines the market price of risk. Besides, the slope of the CML is equal to the Sharpe ratio (see Equations 2.4 and 2.5).

The slope of the CML-extended, on the other hand, is equal to the term \( R = \Phi^{-1}(1 - r) \), which means that the market price of risk and the equilibrium interest rate are simultaneously determined in the extended model:

\[
R = \Phi^{-1}(1 - r) \iff r = 1 - \Phi(R)
\]  

(2.26)

On these grounds, every risk can be interpreted as a zero coupon or discount bond, which promises a discounted payment at some maturity date. In this context, the level of income \( Y \) represents the face or nominal value of the investment, while the interest rate \( r \) represents the internal return earned by the holder of the instrument. As with zero coupon bonds, the market price of risk and its internal return are inversely related to each other (see
Additionally, from Equation 2.25 we obtain that at equilibrium the market price of risk \( R \) must be simultaneously determined in terms of the risk-parameters of the underlying portfolio and the relative stock \( m \):

\[
R = \frac{\mu + m}{\sigma} = \frac{\mu + M/Y}{\sigma}
\]  

(2.27)

Comparing Equation 2.27 to Equation 2.5 we see that the coefficient \( R \) can be indeed regarded as an extended measure of risk to the Sharpe ratio. However, while the Sharpe ratio is expressed as a reward over the level of the risk-free interest rate \( r_0 \), the discount factor \( R \) is expressed as a reward over the relative stock \( m = M/Y \).

These differences are consistent with the different roles that cash holdings play in both models. Indeed, while in the CAPM individuals can always attract deposits if they offer the interest rate \( r_0 \) (no matter the size of the deposits), in the extended model the relative stock represents a guarantee maintained in order to compensate for the average capital return \( \mu \).

From Equations 2.26 and 2.27 we conclude that the market price of risk \( R \) is actually determined by the equilibrium of two different markets.

Thus, on the one hand, as established in Equation 2.26, the market price of risk is related to the return (\( r \)) at which short-term loans are offered to the firms in the class. On the other hand, as established in Equation 2.27, the market price of risk determines a reward (\( m \)) over the level of expected return per unit of volatility. In other words, the market price of risk (and hence the market interest rate) determines the exchange rate of capital for risk that implies the markets of cash balances and securities to be at equilibrium.

Accordingly, if the market is found in a certain state of equilibrium, changing the relative stock \( m \) necessarily implies that all or some of the variables \( r, \mu, \) and \( \sigma \) must vary until a new equilibrium is attained. While changes in \( m \) and \( r \) respectively correspond to quantity and price adjustments, changes in \( \mu \) and \( \sigma \) should be more properly interpreted as structural adjustments, for they involve changes in the composition of the underlying portfolio \( X \) or in the expectations of individuals.

In conclusion, unlike the CML and the CAPM, the CML-extended (as
Chapter 2: The Optimal Liquidity Principle

defined by *Equations* 2.25, 2.26 and 2.27) provides a theoretical framework that is intimately connected to the monetary equilibrium.

In fact, since the total stock of money demanded by the economy is obtained in the extended model by summing up the aggregate balances demanded by the different industries and economic sectors, *Equation* 2.25 can be used as well to characterise the monetary equilibrium of the economy.

For this purpose, the involved variables should be defined accordingly: thus, the money stock \( M \) should be corresponded to some monetary aggregate controlled by the central bank; the level of income \( Y \) should be related to the output obtained at the national level, and finally, the risk-parameters \( \mu, \sigma \) should be related to the series of capital returns of the national income. Such a model will be next analysed in *Section* 2.8.

### 2.8 Extended Macroeconomic Analysis

Consider some economy that produces the income percentage return \( X = \Delta Y/Y \), where \( Y \) denotes the level of national income. Let us additionally assume that the return \( X \) is distributed as a Gaussian random variable with mean \( \mu \) and volatility \( \sigma \). Since in this case the function \( \lambda_{\mu,\sigma}(r) \) (defined in *Equation* 2.21) determines the *optimal* aggregate cash holding \( L_{\mu,\sigma}(r) = Y \cdot \lambda_{\mu,\sigma}(r) \) demanded at the aggregate level,\(^{21}\) the level of the interest rate \( r \) and the risk parameters \( \mu, \sigma \) must be related to each other at equilibrium, in such a way that the optimal aggregate cash balance is equal to the total stock of money \( M \) supplied by the monetary authority:

\[
M = Y \cdot \lambda_{\mu,\sigma}(r) = P \cdot y \cdot [\sigma \Phi^{-1}(1-r) - \mu] \tag{2.28}
\]

where, as in *Equation* 2.2, the variables \( P \) and \( y \) respectively denote the level of prices and the level of real income. Accordingly, variations in the amount of money \( M \) must be followed by changes in any of the variables \( P, y, r, \mu \) and \( \sigma \) in order to reestablish the monetary equilibrium.

Therefore, the main difference between the *classic* and the *alternative* theoretical settings describing the monetary equilibrium (respectively char-

\(^{21}\)This must be the balance demanded by the economy if it efficiently allocates resources, for only in this way the expected output of the economy, as defined in *Equation* 2.15, is maximised.
acterised by Equations 2.2 and 2.28) is that national income is regarded as a random variable in the alternative setting. Then the risk-parameters $\mu$ and $\sigma$ (which describe the riskiness of the series of capital returns of national income) explicitly affect the preference for liquidity of the economy and are thereby determinants of the monetary equilibrium.

On these grounds, the alternative model of equilibrium can be regarded as an extended model.

Let us now investigate how the monetary equilibrium is established in the short-run in the extended model. More precisely, we would like to know how the level of the interest rate $r$ adjusts in the short-run in response to variations in the money stock $M$, assuming that the risk-parameters $\mu, \sigma$ remain unchanged. Applying differences to Equation 2.28 we actually obtain that:

$$\frac{\Delta M}{M} = \pi + \xi + \frac{\Delta \lambda_{\mu,\sigma}(r)}{\lambda_{\mu,\sigma}(r)} \quad \text{with} \quad \pi := \frac{\Delta P}{P} \quad \text{and} \quad \xi := \frac{\Delta y}{y}$$

where $\pi$ denotes the rate of inflation, equal to the percentage variation in the level of prices, and $\xi$ denotes the growth rate of the economy, equal to the percentage variation in the level of real output. The equation above can be equivalently expressed in terms of the semi-elasticity $\eta(r, \mu/\sigma)$ of the Gaussian liquidity principle with respect to the interest rate (see Equations 2.21 and 2.22):

$$\frac{\Delta M}{M} - \xi = \pi + \eta \left( r, \frac{\mu}{\sigma} \right) \Delta r \quad \text{with} \quad \eta \left( r, \frac{\mu}{\sigma} \right) = \frac{1}{\lambda_{\mu,\sigma}(r)} \frac{d\lambda_{\mu,\sigma}(r)}{dr} \quad (2.29)$$

Hence the monetary policy chosen by the central bank can be related to some monetary trend that assures a certain path $\xi$ of economic growth (consistent with the rate of growth of productivity in the economy) together with some predetermined (and preferably low) level of inflation $\pi$ (see Friedman, 1968 and 1970, Romer, 1996, Edwards and Sinzdak, 1997, and also, Howells and Bain, 2005).

Within this context, the levels of inflation and interest rates are respectively corresponded to the instrument and the target of monetary policy.
Accordingly, when inflation is above its target level, the central bank must react by reducing the amount of money $M$. As long as $\eta(r, \mu/\sigma) < 0$, such policy has the effect of raising the level of interest rates and cooling the economy, which are conditions that ultimately reduce inflation. Conversely, when inflation is below its target, the central bank must take actions conducting to lowering interest rates, i.e. it must increase the amount of money $M$. This usually has the effect of accelerating the economy and raising inflation.

During the process, individuals are informed about what the central bank considers the target inflation rate. In this way, the efficiency of the mechanism is increased — eventually leading to increased economic stability.

However (as already stated in Section 2.1), the efficacy of the mechanism depends on the magnitude of the semi-elasticity $\eta(r, \mu/\sigma)$.

Indeed, notice from Equation 2.29 that the portion of the variation of the money supply that is explained by inflation decreases with the magnitude of $\eta(r, \mu/\sigma)$. In other words, given some fixed rate of economic growth $\xi$, the lower the term $|\eta(r, \mu/\sigma)|$, the more monetary interventions are transmitted to inflation — and hence the more effective is monetary policy. In the limit when $|\eta(r, \mu/\sigma)| \to 0$ the whole effect is transmitted to the level of prices:

$$\frac{\Delta M}{M} - \xi = \pi \quad \text{with} \quad \eta \left( r, \frac{\mu}{\sigma} \right) = 0$$

Monetary policy performs at its best under such circumstances.

By contrast, the greater the term $|\eta(r, \mu/\sigma)|$ in Equation 2.29, the less the variations in the money stock are explained by means of changes in the liquidity preference of individuals and hence, the less effective is monetary policy to induce the desired inflation rate. In the limit when $|\eta(r, \mu/\sigma)| \to \infty$, variations in the amount of money have no effect on interest rates and hence, monetary policy is useless under such circumstances — recall that liquidity preference is absolute in this case.

The magnitude of the semi-elasticity can be precisely determined in the...
case of the Gaussian liquidity principle.

In fact, as stated in Section 2.6, when the series of income returns follows a Gaussian probability distribution, low and high semi-elasticities are corresponded to specific states of the market characterised by the level of the interest rate and the risk-parameters $\mu, \sigma$.

Thus, on the one hand, as established in Equation 2.23, if $|\Phi^{-1}(1 - r)| < +\infty$, i.e. if $r > 0$ and $r < 1$, then:

$$\left| \Phi^{-1}(1 - r) - \frac{\mu}{\sigma} \right| \rightarrow +\infty \implies \left| \eta \left( r, \frac{\mu}{\sigma} \right) \right| \rightarrow 0$$

Therefore, as long as $0 < r < 1$, the magnitude of the semi-elasticity is diminished both when the magnitude of the income’s expected return is increased (no matter the sign of the expected return) and when the volatility of income is reduced. Accordingly, the monetary mechanism is more effective in economies that produce higher expected returns (both when positive and negatives returns are obtained) and show lower variability.

On the other hand, from Equation 2.22 we obtain that the magnitude of the semi-elasticity converges to infinite when the level of the interest rate converges to zero or one:

$$r \downarrow 0 \text{ or } r \uparrow 1 \implies |\Phi^{-1}(1 - r)| \rightarrow +\infty \implies \left| \eta \left( r, \frac{\mu}{\sigma} \right) \right| \rightarrow +\infty$$

The same result is obtained when $|\Phi^{-1}(1 - r) - (\mu/\sigma)| \rightarrow 0$, i.e.:

$$\Phi^{-1}(1 - r) = \frac{\mu}{\sigma} \implies r = 1 - \Phi \left( \frac{\mu}{\sigma} \right) \implies \left| \eta \left( r, \frac{\mu}{\sigma} \right) \right| = +\infty$$

Then the magnitude of the semi-elasticity is equal to infinite when the interest rate attains any of the values $r = 0$, $r = 1$ or $r = 1 - \Phi(\mu/\sigma)$. Consequently, in any of these states the preference for liquidity of the economy is absolute and hence, monetary policy is useless for dealing with price and output fluctuations. On these grounds, these interest rates values are corresponded to critical states of the economy.
Other complications may arise when implementing the monetary mechanism due to the dependence of the cost of capital on the market expectations and the riskiness of national output.

Indeed, recall that the market interest rate $r$ must lie in the interval determined by the risk-free interest rate $r_0$ and the liquidity premium $r_{\theta,X}$ (where the liquidity premium depends on the benefit lost from maintaining cash holdings instead of investing on risk, see Equations 2.11 and 2.13 and the related discussion in Section 2.3), in such a way that:

$$r_0 \leq r \leq r_0 + r_{\theta,X}$$

In this context, the returns $r$ and $r_0 + r_{\theta,X}$ denote the cost of capital as perceived by lenders and borrowers respectively. Accordingly, individuals prefer to maintain cash holdings and do not rely on capital markets to fit their balances when $r > r_0 + r_{\theta,X}$, because in this case the cost impose by lenders is too expensive for them.

As a consequence, if the premium $r_{\theta,X}$ is diminished (i.e. if the income surplus over the level of reserves is reduced, see Equation 2.11) until the borrowers’ perceptions of the cost of capital is under the lenders’ estimations of it (i.e. until $r_0 + r_{\theta,X} < r$), people will be induced to modify their funding strategies, moving from external to internal financing — i.e. moving from debt to capital. By contrast, if the premium $r_{\theta,X}$ is increased, then the profit that is obtained from relying on capital markets instead of keeping cash holdings (equal to $r_0 + r_{\theta,X} - r$) will be augmented, and hence the incentives to replace capital by debt will be incremented.

In other words, in the extended model the monetary equilibrium can be affected by changes in the expectations of individuals and in the riskiness of national output, which in the case of Gaussian risks are reflected in the risk-parameters $\mu, \sigma$. Such adjustments are manifested as fluctuations in the amount of funds demanded at the aggregate level.\(^{23}\)

Finally, recall that in the extended model the riskiness of national income is expressed in terms of the riskiness of the outputs produced by the portfolios held by individuals at different aggregation levels (as stated in Equations 2.17, 2.18 and 2.19). Consequently, the variability of income at

\(^{23}\)Recall that in the model creditors are regarded as price takers, who set the price of loans based on the credit class the borrower belongs to according to the market, see Equation 2.15 and the related discussion.
the economic level might be induced by a single industry or economic sector — in such a way that, in particular, the volatility and the mean return of national income might be determined by a single industry or economic sector. Hence, the possibility of contagion naturally arises in the model.24

24 Some recent studies emphasise the role of aggregation in explaining macroeconomic and financial stability. Thus, for example, Calza and Sousa (2003) postulate that considering aggregation effects it is possible to explain why the money demand has been more stable in the euro area than in other large economies. The fact that Germany has a large weight in the M3 aggregate for the euro area and that the money demand has been historically stable in that country contributes to support such hypothesis. In other words, the stability of the German economy is supposed to be shared by the rest of the economies in the block, as a positive externality.
Chapter 3

A Model of Equilibrium in Markets of Cash Balances

Companies and investors holding financial securities that deliver random outcomes are exposed to unknown balances equal to the difference between the market values of outstanding assets and liabilities. Positive balances can be lent at the overnight interest rate in some market of interbank loans. Short-term debt can be hired in these markets as well, when a net lost is suffered. However, premiums over the risk-free interest rate are normally charged by lenders, intended to reflect the default and market risk implicit in the loans, as well as the involved transaction costs (see e.g. Howells and Bain, 2005).

Although the availability of loans is dependent on the liquidity state of the market, there is no consensus about how to measure this property.

After Modigliani and Miller (1958), many scholars have neglected any liquidity constraints, claiming that the only factors preventing investors from attracting all the funds required to carry out solvent projects, other than market imperfections, are unnecessary regulations and aversion to risk (see e.g. Miller, 1998).

Practitioners, on the other hand, are daily exposed to credit restrictions, and have to deal with changing scenarios, going from periods of high optimism and cheap money, to severe short-falls that sometimes occur all in a sudden and without apparent reasons in the middle of expansionary cycles.
Chapter 3: Equilibrium in Markets of Cash Balances

Providing a theoretical framework for the characterisation of liquidity, capable of accounting for the behaviour of markets under normal circumstances as well as in times of crises, is the main purpose of this chapter.

3.1 Equilibrium in the Market of Cash Balances

Let us start by considering some market of cash balances where individuals can lend and borrow at a single interest rate $r$. The funds in this market are provided by creditors, in such a way that the total capital or credit supply cannot surpass a given stock $M$ — which can be regarded as a limited guarantee that every outstanding liability will be honoured.

At equilibrium, the stock $M$ must be equal to the aggregate demand for cash holdings, which will be expressed as an inversely related function of the interest rate. The following equation must then hold at equilibrium:

$$M = L \cdot \lambda(r) \quad \text{with} \quad \frac{d\lambda(r)}{dr} < 0 \quad (3.1)$$

where $L$ and $\lambda(r)$ respectively denote the amount of funds spent on the transactions of securities and the proportion of funds maintained in the form of reserves.\(^1\) In a more general context, the function $\lambda(r)$ represents the preference for liquidity of the market, while the equilibrium interest rate can be related to the cost of contracting non-risky debt plus premiums and transaction costs.\(^2\)

Variations in the credit supply $M$ must then be compensated by quantity and price adjustments, respectively performed through changes in the variables $L$ and $r$. Indeed, applying differences to Equation 3.1, we obtain that at equilibrium:

\(^1\)Notice that the level of the interest rate (and hence the price of balances) is determined in Equation 3.1 on the grounds of a certain class of money substitutes (whose aggregate supply is denoted by $M$) and a particular class of securities (whose aggregate demand is denoted by $L$), independently from the quantities and prices determining the equilibrium in other markets of financial securities. This means that the model described by Equation 3.1 corresponds to a model of partial equilibrium.

3.1 Equilibrium in the Market of Cash Balances

\[ \Delta M = \Delta L \cdot \lambda (r) + L \cdot \Delta \lambda (r) \implies \frac{\Delta M}{M} = \frac{\Delta L}{L} + \frac{1}{\lambda (r)} \cdot \frac{d\lambda (r)}{dr} \cdot \Delta r \] (3.2)

The factor of the interest rate deviation in Equation 3.2 is known as the \textit{semi-elasticity} of the preference for liquidity:

\[ \eta (r) := \frac{1}{\lambda (r)} \cdot \frac{d\lambda (r)}{dr} = \frac{d \ln (\lambda (r))}{dr} \] (3.3)

where \( \ln (\cdot) \) denotes the \textit{natural logarithm} function. The \textit{semi-elasticity} function \( \eta (r) \) thereby determines the percentage variation in the liquidity-preference function \( \lambda (r) \) induced by a point of variation of the rate of interest.

Consequently, the lower the magnitude of the semi-elasticity, the higher the interest rate adjustment that is consistent with given changes in the aggregate cash flows \( M \) and \( L \). In the limit when \( \eta (r) \to 0 \) (which corresponds to the case when the demand for cash holdings is inelastic), no adjustment is required in the interest rate to conduct the market to a new equilibrium. In fact, in this case every variation in the credit supply must be compensated by another movement, equal in sign and magnitude, affecting the amount of funds spent on transactions, i.e. the condition \( \Delta M/M = \Delta L/L \) must be satisfied. The market of cash balances is the most stable in this scenario — because the equilibrium interest rate is the most stable in this case.

By contrast, even big variations in the aggregate cash flows \( M \) and \( L \) can produce but small movements in the interest rate in markets with high semi-elasticities, or equivalently, even small movements in the interest rate can produce big adjustments in the balances transacted at equilibrium. In the limit when \( |\eta (r)| \to \infty \) (which corresponds to the case when the demand for cash balances is perfectly elastic), any movement of the interest rate will be followed by an increment of infinite magnitude in the level of required cash holdings. In other words, any variation in the supply of loans is completely absorbed by adjustments in the preference for liquidity of investors. Then liquidity-preference is said to be absolute. The market of cash balances is the most unstable under such circumstances — because the response of the market to a certain variation of the credit supply is the most uncertain in this case.

Within this context, the equilibrium in the market of cash balances (as
Chapter 3: Equilibrium in Markets of Cash Balances

described by Equations 3.1 and 3.2) and hence, the liquidity state of the market, are completely determined by the functions $\lambda(r)$ and $\eta(r)$. In fact, more liquid markets are corresponded to lower levels of the liquidity-preference function $\lambda(r)$, for in this case individuals prefer to maintain fewer reserves and to invest more on risk. Besides, the semi-elasticity $\eta(r)$ measures the percentage points of liquidity that are gained by reducing the interest rate in one point. A precise description of the liquidity states of markets thereby requires of a theoretical framework leading to some explicit characterisation of these fundamental properties.

The characterisation of the preference for liquidity that will be presented hereafter follows the classical approach of James Tobin, 1958, according to which individuals maintain cash holdings as behaviour towards risk. More precisely, Tobin proposes to relate the preferred balances to the combinations of risk and cash holdings — or some financial substitute to the later, such as a bond or deposit offering some fixed return at maturity — that maximise the utility of the decision-maker. A fundamental hypothesis of the model is that individuals can borrow and lend any amount of funds at a single interest rate, or equivalently, that individuals face no liquidity restrictions (see also Sharpe, 1964).

Unlike the model of Tobin, in the alternative model presented in this chapter the presence of liquidity restrictions is regarded as an inherent quality of markets. Individuals may be prevented from exclusively relying on inter-bank loans under such circumstances — in order to avoid undesired imbalances — and be persuaded to look for additional protection by keeping cash stocks and insuring deposits. As long as the prices of insurance contracts are expressed in actuarial terms, it is possible to prove that the value of the net portfolio (which contains the underlying risk plus cash holdings) explicitly depends on the size of the guarantee. The optimal amount of equity must then be found at the level where firms are indifferent between assuming the risk of default by themselves and transferring it to some insurance institution.

In fact, it is possible to prove that the optimal demand for balances explicitly depends on the riskiness of the underlying exposure under such circumstances. This implies that both the functions $\lambda(r)$ and $\eta(r)$ are explicitly dependent on risk. Hence the market equilibrium (as determined by Equations 3.1 and 3.2) simultaneously determines the equilibrium in two markets: the market of cash balances on the one hand, which is characterised by the cash stock $M$ and the interest rate $r$, and the market of securities on the other, which is characterised by the aggregate amount of
funds $L$ spent on transactions and the riskiness of the aggregate claim $X$.

Such theoretical framework allows for a precise description of *liquidity states* in terms of the market variables $M$, $L$ and $r$, and the risk-parameters that statistically describe the random variable $X$. State transitions are meaningful in this setting, as long as any of the explanatory variables is modified. In particular, episodes of *liquidity crises* — which are corresponded to scenarios when the functions $\lambda(r)$ and $\eta(r)$ are drastically increased during short intervals of time — must be corresponded to some specific scenarios.

### 3.2 Preference for Liquidity as the Optimal Retention

Let us assume that every firm and investor in the market holds a portfolio that combines some cash balance $K$ and a certain mutual fund delivering the net payment $X$ per unit of investment. Let the guarantee $K$ be expressed as a proportion $\lambda$ of the volume $L$ spent on transactions, such that $K = L \cdot \lambda$. Then the total *profit and loss (P&L)* accrued by the mutual fund and the net portfolio are respectively given by $L \cdot X$ and $L \cdot (X + \lambda)$.

Within a class of firms facing the same cost of capital $r$, the optimal balance will be determined in such a way that the total cost of bankruptcy is minimised:

$$
\min_{\lambda} E_{\theta} \left[ (X + \lambda)_- \right] + r \cdot \lambda
$$

where $(X + \lambda)_- := -\min(0, X + \lambda)$, and where the parameter $\theta$ accounts for the preferences of the decision-maker — dependent on information and aversion-to-risk. As I have previously demonstrated (see Mierzejewski, 2006 and 2008), this strategy maximises the value of the portfolios held by *opaque* institutions, who face informational asymmetries in their relations with customers and stockholders. This strategy is thereby compatible with *rational* decision-making.\(^3\)

\(^3\)Fama (1980) and Jensen (1986) describe the problems arising inside organisations in the presence of agency costs between managers and stockholders. See also Froot et al., 1993. Ross (1989) and Merton (1997), on the other hand, analyse the consequences of the moral-hazard induced by the *opaqueness* of financial institutions — due to the fact that the investment and funding strategies followed by these companies are not fully observed by customers.
From the actuarial point of view, the term \( E_\theta [(X + \lambda)_-] \) represents the *excess of loss* suffered by the issuer of an insuring policy called *layer*, which gives the holder the right to demand from the insurer the amount \((X + \lambda)\) in the case that \(X < -\lambda\). This term represents the *fair or actuarial* price of insurance (Goovaerts et al., 1984). Then the terms of the contract impose the obligation to make the payment \(\lambda\) at *any event*, which means (in the actuarial nomenclature) that \(\lambda\) represents the *retention or deductible* of the corresponding policy. The term \(r \cdot \lambda\), on the other hand, corresponds to the opportunity cost of maintaining \(\lambda\) as a cash stock instead of lending it in the market of cash balances at the interest rate \(r\) (see also Dhaene et al., 2003, and Goovaerts et al., 2005).

On these grounds, the *optimal* level of surplus, as determined by *Equation 3.4*, can be corresponded to the *optimal retention-level*.

Notice that the optimisation criterion of *Equation 3.4* determines a compromise between two conflicting objectives. On the one hand, by augmenting the level of capital, the firm can reduce the excess of loss to be paid in case of bankruptcy and thus reduce the price of the insurance contract. But in this way the (opportunity) cost of holding idle balances is incremented. Conversely, the cost of the guarantee can be reduced by diminishing the level of reserves, but only if at the same time the cost of insuring the outstanding liabilities is incremented.

Hence the *optimal deductible* characterises the surplus at which decision-makers are indifferent between insuring their liabilities and assuming the cost of insolvency by themselves. Allowing different estimations of \(\theta\) due to market imperfections provides a justification to the appearance of a demand for insurance (see Venter, 1991, and Wang et al., 1997).

In order to obtain a solution for the optimisation problem of *Equation 3.4*, an expression for the expectation operator \(E_\theta\) is required. In lines with Wang (1995), the following risk-principle will be used:

\[
E_\theta[X] = \int x \, dF_{\theta,X}(x) = \int T_{\theta,X}(x) \, dx \quad \text{with} \quad T_{\theta,X}(x) := T_X(x)^{\frac{1}{\theta}} \quad (3.5)
\]

where \(F_{\theta,X}\) and \(T_{\theta,X}\) respectively denote the *distorted cumulative* and the *distorted tail* probability functions describing the underlying risk, respectively defined as \(F_{\theta,X}(\lambda) = P_\theta \{X \leq \lambda\} = \int_{-\infty}^{\lambda} dF_{\theta,X}(x)\) and \(T_{\theta,X}(\lambda) = P_\theta \{X > \lambda\} = \int_{\lambda}^{+\infty} dF_{\theta,X}(x)\), such that \(T_{\theta,X}(\lambda) = 1 - F_{\theta,X}(\lambda) \forall \lambda\).
3.2 Preference for Liquidity as the Optimal Retention

Consequently, the more the aversion-to-risk of decision-makers (i.e. the higher the magnitude of \( \theta \)), the higher the price of risk. Similarly, the less the aversion-to-risk of decision-makers (i.e. the lower the magnitude of \( \theta \)), the lower the price of risk. Risk-neutral decision-makers are characterised by \( \theta = 1 \), in which case the distorted probability principle is equal to the traditional expectation operator.

As long as the expectation operator is defined according to Equation 3.5, the excess of loss can be written as:

\[
E_{\theta}[(X + \lambda)_-] = -\int_{-\infty}^{-\lambda} (x + \lambda) \cdot dF_{\theta,X}(x)
\]

Then by applying the Leibniz integral rule for the differentiation of a definite integral whose limits are functions of the differential variable, the following expression for the optimal cash balance is obtained by applying Lagrange optimisation:

\[
\frac{dE_{\theta}[(X + \lambda)_-]}{d\lambda} + r = -T_{\theta,X}(\lambda) + r = 0 \tag{3.6}
\]

Hence decision-makers demand reserves up to the point where its opportunity cost equals the marginal return of risk and the optimal demand for capital is given by the inverse function of the tail-probability:

\[
\lambda(r) = T_{\theta,X}^{-1}(r) = F_{\theta,X}^{-1}(1 - r) \tag{3.7}
\]

or equivalently:

\[
\lambda(r) = T_{\theta,X}^{-1}(r^{\theta}) = F_{\theta,X}^{-1}(1 - r^{\theta})
\]

According to this specification, the demand for cash balances always follows a non-increasing and continuous path — as long as the probability function

\[
\frac{\partial}{\partial z} \int_{u(z)}^{v(z)} \varphi(x, z) \, dx = \int_{u(z)}^{v(z)} \frac{\partial \varphi(x, z)}{\partial z} \, dx + \varphi(v(z), z) \cdot \frac{\partial v(z)}{\partial z} - \varphi(u(z), z) \cdot \frac{\partial u(z)}{\partial z}
\]

\footnote{According to this rule, for every functional \( \varphi(x, z) \):}
describing risk is continuous. Besides, the minimum and maximum levels of surplus are respectively demanded when \( r \geq 1 \) and \( r \leq 0 \). Discrepancies relative to preferred cash-balances can thus be explained on the basis of the underlying risks, expectations and the opportunity cost of capital.

A main hypothesis that will be maintained in Section 3.3, when characterising the aggregate demand for cash holdings and the equilibrium in the market of balances, is that investors can lend and borrow at a single interest rate \( r \). Thus, the model describes the situation of markets where lenders cannot fully observe the composition of the portfolios held by borrowers (due to the opacity of financial decisions, as suggested by Merton, 1997) and cannot efficiently discriminate through the price of liabilities. Under such circumstances, creditors prefer to behave as price-takers.

### 3.3 The Internal Rate of Return on Risk

Let \( X_1, \ldots, X_n \) denote the aggregate exposures held by individuals (firms, private investors and financial intermediaries) in some market of cash balances, and let \( \theta_1, \ldots, \theta_n \) represent their informational types.

The question is under what circumstances the level of reserves demanded at the aggregate level is equal to the sum of the cash balances preferred by individuals. Given that the optimal surpluses are expressed in terms of the quantiles of the underlying probability distributions (as stated in Equation 3.7), we would specifically like to know whether the inverse probability function of the sum \( X = X_1 + \ldots + X_n \) is equal to the sum of the quantiles of the marginal distributions.

As demonstrated by Dhaene et al. (2002), the property of the sum of the quantiles mathematically characterises the comonotonic dependence structure, where comonotonicity represents a case of extreme dependence, when no diversification effect can be attained by pooling risks together, for all of them move in the same direction.

In fact, given a random vector \( (X_1, \ldots, X_n) \) with marginal cumulative distribution functions \( (F_{X_1}, \ldots, F_{X_n}) \), the comonotonic random vector \( (X_1^c, \ldots, X_n^c) \) is mathematically defined in such a way that if \( U \) denotes the random variable uniformly distributed in the interval \([0, 1]\), such that \( F_U(u) = u \forall u \in [0, 1] \), \( F_U(u) = 0 \forall u < 0 \) and \( F_U(u) = 1 \forall u > 1 \), the following identity holds in distributions:
3.3 The Internal Rate of Return on Risk

\( (X_1^c, \ldots, X_n^c) = \left( F_{X_1}^{-1}(U), \ldots, F_{X_n}^{-1}(U) \right) \)

Equivalently, the vector \( (X_1^c, \ldots, X_n^c) \) is said to be comonotonic if a random variable \( Z \) exists, as well as a set of non-decreasing functions \( g_1, \ldots, g_n \), such that:

\( (X_1^c, \ldots, X_n^c) = (g_1(Z), \ldots, g_n(Z)) \)

In this way, the realisation of any single event (which can be either related to the uniform random variable \( U \) or to the risk \( Z \)) simultaneously determines all the components of any comonotonic random vector. Moreover, since both the series of functions \( \left( F_{X_1}^{-1}, \ldots, F_{X_n}^{-1} \right) \) and \( (g_1, \ldots, g_n) \) are non-decreasing, all the components of the vector \( (X_1^c, \ldots, X_n^c) \) move in the same direction.

As already stated, the quantile function of the sum of the components of any comonotonic random vector is equal to the sum of the component quantile functions, a fact that supports the use of the comonotonic dependence structure to characterise the aggregate demand for cash balances.

Within this setting, if loans are supplied by creditors at a single interest rate \( r \), the following expression is obtained for the aggregate preference for liquidity:

\[
\lambda(r) = T_{\theta,-X}^{-1}(r) = \sum_{i=1}^{n} T_{\theta_i,-X_i}^{-1}(r) \tag{3.8}
\]

where the process of capital \( P&L \) of the market portfolio and the aggregated effect of the distortions introduced by investors are respectively described by the comonotonic sum \( X = X_1^c + \cdots + X_n^c \) and the informational parameter \( \theta \), and where \( T_{\theta,-X} = \left( \sum_{i=1}^{n} T_{\theta_i,-X_i}^{-1} \right)^{-1} \) denotes the distribution function of the comonotonic sum when the marginal returns are described either by the cumulative or by the tail probability functions, respectively denoted as \( (F_{\theta_1,-X_1}, \ldots, F_{\theta_n,-X_n}) \) and \( (T_{\theta_1,-X_1}, \ldots, T_{\theta_n,-X_n}) \). Thus precautionary industries rely on the most pessimistic case, when the failure in any single firm spreads all over the market.

Having characterised the aggregate demand for reserves, the determination of the equilibrium in the market of balances is straightforward.
Indeed, letting $M$ and $L$ respectively denote the aggregate stock of cash and the total amount of funds spent on securities, and replacing Equation 3.8 in Equation 3.1, the following relationship is obtained:

$$M = T_{\theta,-X}(r) \cdot L \Leftrightarrow m := \frac{M}{L} = T_{\theta,-X}(r) \Leftrightarrow r = T_{\theta,-X}(m) \quad (3.9)$$

The discount factor $T_{\theta,-X}(r)$ can then be regarded as the market-price of risk, for it corresponds to the rate at which a unit of investment on risk is exchanged for a unit of cash in the market. Hence the cash-to-risk ratio $m = M/L$ determines the return $r = T_{\theta,-X}(m)$ obtained when investing one unit of cash on the underlying risk, which can be accordingly regarded as the Internal Rate of Return on Risk (IRRR).

Within this context, the IRRR can be interpreted as the return received when holding a given uncertain claim instead of investing on a non-risky zero-coupon bond with a predetermined maturity, in the same way the Internal Rate of Return (IRR) of a zero-coupon bond represents the opportunity cost of receiving a cash-flow at some future date instead of today (see e.g. Hull, 2000). Both coefficients can then be considered as alternative measures of liquidity.

Besides, since the IRRR is defined as the probability that the current cash stock suffices to cover the imbalance of the aggregate or market portfolio, this rate reflects the facilities that individuals face when looking for cash holdings in the market, and hence, the liquidity state of the market. Indeed, notice that lower IRRRs are related to markets where the percentage loss of the aggregate portfolio surpasses the cash-to-risk ratio less frequently. This implies that investors face less difficulties to adjust their end-of-day balances in markets with lower internal returns, or equivalently, that market with lower internal returns are more liquid. Conversely, higher IRRRs are corresponded to markets where net losses are suffered more frequently, which means that higher internal returns are corresponded to more illiquid markets.

It must be noticed, in addition, that the variables involved in the equilibrium of Equation 3.9 are of different nature. As a matter of fact, while the cash-to-risk ratio $m = M/L$ is affected by lenders and borrowers, who respectively control the supply and the demand for cash balances, the discount factor $T_{\theta,-X}(r)$ provides a measure of the market’s response. In other words, while the variables $M$ and $L$ are exogenously determined, the in-
formational parameter $\theta$ and the statistical description of the aggregated portfolio $X$ are intrinsic characteristics of the market and so they are *endogenously* determined.

On these grounds, $M$ and $L$ can be regarded as *control* variables of the market, whereas $\theta$ and $X$ represent *state* variables. The only variable involved in the equilibrium which is not directly observable is the IRRR, although its value is uniquely determined once the rest of the parameters are fixed.\(^5\)

Changes in the money stock $M$ can then be compensated by adjustments on the IRRR — i.e. on the market price of risk — but also on the volumes of funds $L$ spent on securities, as well as on the market preferences and the riskiness of the market portfolio. Moreover, the magnitude of the involved adjustments can be precisely determined from the equation of differences stated in *Equation 3.2*. Such analysis will be carried out later in *Section 3.8*.

Before in *Sections 3.5*, 3.6 and 3.7 the equilibrium will be presented in the particular case when the riskiness of the market portfolio is described by a *Gaussian* probability distribution — which implies that risk is completely described by the *mean return* and the *volatility* of the corresponding series of percentage $P&P&L$. This characterisation will support the analysis of *Section 3.8*.

The aim of following this schedule is twofold. In the first place, much of the classic financial and economic theory is founded on this setting, thus the Gaussian approach can be naturally integrated into the existing literature. In the second place, an explicit functional expression is available for the semi-elasticity $\eta(r)$ in the Gaussian case, which allows to precisely assess the *instability* of markets.

First an explanation must be given for a major pitfall of the model of equilibrium stated in *Equation 3.9*. Indeed, since the equilibrium interest rate is expressed as a probability value in *Equation 3.9*, it must necessarily lie in the interval $[0,1]$. This fact is hardly sustained on the grounds of the empirical evidence, for interest rates greater than one have been actually observed in the past. A *scaling* factor will be then introduced in *Section 3.4*, which will be related to the *attitude-towards-risk* of the decision-maker.

\(^5\)This interpretation of the equilibrium interest rate stresses the fact that the model presented in this chapter corresponds to a model of *partial* equilibrium, as already stated in *Section 3.1*. 
Chapter 3: Equilibrium in Markets of Cash Balances

3.4 Scaling the Equilibrium Rate of Interest

As established in Equation 3.6, the optimal demand for cash balances is determined by an optimal exchange between a flow of probability and a flow of cash delivered with certainty. Accordingly, in Equation 3.9 the equilibrium interest rate or IRRR is given by the tail probability of the underlying exposure and hence the set of interest rates that are accepted by the model is contained in the interval $[0, 1]$, since $0 \leq r = T_{\theta-X}(m) \leq 1$. Such prediction does not fit to the experience, for interest rates higher than one are possible to find in real markets.

A meaning can be given to this result, however, if a scaling factor $\psi$ is introduced in the definition of the IRRR, in such a way that, from Equation 3.9:

$$r = \psi \cdot T_{\theta-X}(m) = \psi \cdot T_{-X}(m)^\frac{1}{\theta}, \quad \text{with} \quad \psi \geq 1$$ (3.10)

In this way, values of the interest rate outside the range $[0, 1]$ can be obtained by multiplying the tail probability by the appropriate scaling factor. We will now show that such a transformation is meaningful within the proposed theoretical framework.

In fact, it has been demonstrated by Wang that the distorted probability principle is scale-invariant, for the effect of scaling the underlying random variable is the same as scaling the expectation operator (see the Property 4.3 in Wang, 1995):

$$E_{\theta}[\psi \cdot X] = \int (\psi \cdot x) \, dF_{\theta,X}(x) = \psi \cdot \int x \, dF_{\theta,X}(x) = \psi \cdot E_{\theta}[X] \quad \forall \ psi > 0$$

Therefore, amplifying the amount $(X + k)_-$ to be paid in case of bankruptcy leads the optimisation problem established in Equation 3.4 to be written as:

$$\min_{\lambda} \psi \cdot E_{\theta} [(X + \lambda)_-] + r \cdot \lambda \quad \iff \quad \min_{\lambda} E_{\theta} [(X + \lambda)_-] + \frac{r}{\psi} \cdot \lambda$$ (3.11)

It can be easily shown that establishing the first-order condition of this problem leads to Equation 3.10, in such a way that the optimal amount of capital is given by:
3.5 The Neutral Gaussian Demand for Cash Balances

\[ \lambda(r) = \frac{T^{-1}_{\theta, -X}}{\psi} = \frac{T^{-1}_{\theta, -X}}{\psi^\theta} \]  

(3.12)

Hence the effect of scaling the expected loss is the same as amplifying the cost of capital by the multiplicative inverse of the scaling factor.

Notice that in Equation 3.11 the factor \( \psi \) increases the importance attached to the cost of default (or the cost of deposit insurance), or equivalently, it reduces the importance attached to the cost of maintaining idle balances, when deciding the optimal cash stock. On these grounds, the cases \( \psi > 1 \) and \( \psi < 1 \) will be respectively corresponded to averse-to-risk and risk-lover attitudes towards risk — for the burden of bankruptcy is magnified at the time that the cost of reserves is neglected in the former case, whereas in the later, the cost of bankruptcy is neglected and the cost of reserves is magnified.\(^6\)

Like the informational parameter \( \theta \), the scaling coefficient \( \psi \) can thus be interpreted as a risk-aversion corrector. In this context, interest rates greater than one are possible to be found at equilibrium in averse-to-risk markets.

3.5 The Neutral Gaussian Demand for Cash Balances

Let us consider a market where individual risks are distributed as Gaussians with means \( \mu_1, \ldots, \mu_n \) and volatilities \( \sigma_1, \ldots, \sigma_n \), while their contributions to the aggregate exposure are given by the coefficients \( \omega_1, \ldots, \omega_n \), with \( 0 \leq \omega_i \leq 1 \) \( \forall i \), such that \( L_i = \omega_i \cdot L \) and \( L = L_1 + \cdots + L_n \), or equivalently, \( \omega_i = L_i/L \). Under such conditions, the series of capital \( P&L \) of the market portfolio, equal to the comonotonic sum of individual exposures, is also a Gaussian random variable, whose mean and volatility are given by their corresponding weighted averages (as demonstrated by Dhaene et al., 2002).

Recall that neutral markets are characterised by \( \theta = \psi = 1 \). Consec-

---

\(^6\) Alternatively, the higher the magnitude of \( \psi \), the lower the importance given to the opportunity cost of capital with respect to the cost of bankruptcy when deciding the level of economic capital and so the higher the aversion-to-risk of the decision-maker. Conversely, the lower the magnitude of \( \psi \), the lower the aversion-to-risk of the decision-maker.
Chapter 3: Equilibrium in Markets of Cash Balances

Consequently, since the individual returns \(-X_1, \ldots, -X_n\) are also distributed as Gaussian, with the same volatilities but mean returns with opposite sign, i.e. \(-\mu_1, \ldots, -\mu_n\), we obtain from Equations 3.7 and 3.8 that the preference for liquidity of the market is equal to:

\[
\lambda_{\mu,\sigma}(r) = \sigma \cdot \Phi^{-1}(1 - r) - \mu
\]  

with:

\[
\mu = \sum_{i=1}^{n} \omega_i \cdot \mu_i \quad \& \quad \sigma = \sum_{i=1}^{n} \omega_i \cdot \sigma_i
\]

where \(\Phi\) denotes the cumulative distribution function of a standard-Gaussian distribution, whose mean and volatility are respectively equal to zero and one and which will be denoted as \(X^{\text{std}}\). According to this notation, the individual and the market exposures are respectively represented by the random variables \(-X_i = \sigma_i \cdot X^{\text{std}} - \mu_i\), with \(i = 1, \ldots, n\), and \(-X = \sigma \cdot X^{\text{std}} - \mu\).

From Equations 3.3 and 3.13, an explicit expression is obtained for the semi-elasticity of the demand for cash balances with respect to the interest rate. In fact, letting \(\phi\) denote the density probability function of the standard Gaussian distribution:

\[
\phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp \left( -\frac{x^2}{2} \right) \quad \text{with} \quad \phi(x) = \frac{d\Phi(x)}{dx}
\]

we have that for any level of probability \(p = \Phi(x)\):

\[
\phi(x) = \frac{d\Phi(x)}{dx} \quad \Rightarrow \quad \frac{d\Phi^{-1}(p)}{dp} = \frac{1}{\phi(\Phi^{-1}(p))}
\]

Hence, replacing \(p = 1 - r\):

\[
\frac{d\lambda_{\mu,\sigma}(r)}{dr} = \sigma \cdot \frac{d\Phi^{-1}(1-r)}{dr} = -\sqrt{2\pi} \sigma \cdot \exp \left( \frac{\left[\Phi^{-1}(1-r)\right]^2}{2} \right)
\]
3.5 The Neutral Gaussian Demand for Cash Balances

in such a way that:

\[ \eta(r, \frac{\mu}{\sigma}) = -\frac{1}{\lambda_{\mu,\sigma}(r)} \cdot \frac{d\lambda_{\mu,\sigma}(r)}{dr} = \]
\[ -\sqrt{2\pi} \exp \left( \frac{\left[\Phi^{-1}(1-r)\right]^2}{2} \right) \cdot \left[ \Phi^{-1}(1-r) - \frac{\mu}{\sigma} \right]^{-1} \]  

(3.14)

Then the semi-elasticity of the optimal cash demand of every Gaussian risk is a function of the rate of interest and the mean-to-volatility ratio \( \mu/\sigma \).

Two different states of the market are determined by the sign of the semi-elasticity function:

\[ \eta(r, \frac{\mu}{\sigma}) \leq 0 \iff \Phi^{-1}(1-r) - \frac{\mu}{\sigma} \geq 0 \iff r \leq 1 - \Phi \left( \frac{\mu}{\sigma} \right) \]
\[ \eta(r, \frac{\mu}{\sigma}) > 0 \iff \Phi^{-1}(1-r) - \frac{\mu}{\sigma} < 0 \iff r > 1 - \Phi \left( \frac{\mu}{\sigma} \right) \]  

(3.15)

Then a critical rate \( r_{\mu,\sigma} \) exists, where the transition from one state to another is produced, which is dependent on the cumulative probability \( \Phi \) of the standard-Gaussian random variable \( X^{\text{std}} \):

\[ r_{\mu,\sigma} := 1 - \Phi \left( \frac{\mu}{\sigma} \right) = P \{ X^{\text{std}} > \frac{\mu}{\sigma} \} = P \{ \sigma \cdot X^{\text{std}} - \mu > 0 \} \]  

(3.16)

Notice that the interest rate \( r = r_{\mu,\sigma} \) also determines a change in the sign of the liquidity-preference function:

\[ \lambda_{\mu,\sigma}(r) \geq 0 \iff r \leq r_{\mu,\sigma} = P \{ \sigma \cdot X^{\text{std}} - \mu > 0 \} \]
\[ \lambda_{\mu,\sigma}(r) < 0 \iff r > r_{\mu,\sigma} = P \{ \sigma \cdot X^{\text{std}} - \mu > 0 \} \]  

(3.17)

Since the random variable \( \sigma \cdot X^{\text{std}} - \mu \) represents the (still unknown) loss of the underlying fund, we then obtain that firms only maintain reserves when the opportunity cost of capital \( r \) is lower than the probability of suffering capital losses. Indeed, under such circumstances, the amount spent in an additional unit of capital is lower than the marginal reduction on the excess
of loss (which is given by the tail probability of the underlying exposure, see Equation 3.6) and thus a net reduction in the total bankruptcy costs can be attained by attracting an additional unit of cash. When \( r > r_{\mu,\sigma} \), on the other hand, the opportunity cost of capital is high enough to make lending more attractive than guaranteeing the risk \( \sigma \cdot X_{std} - \mu \).

Regarding the limiting behaviour of the semi-elasticity function, from Equation 3.14 we have that:

\[
\begin{align*}
r \downarrow 0 & \implies \eta \downarrow -\infty \quad \text{and} \quad r \uparrow 1 & \implies \eta \uparrow +\infty
\end{align*}
\]

which means that two asymptotes are produced at the axis \( r = 0 \) and \( r = 1 \). Another asymptote is produced at the point \( r = r_{\mu,\sigma} \) due to the fact that:

\[
\begin{align*}
r \uparrow r_{\mu,\sigma} & \implies \eta \downarrow -\infty \quad \text{and} \quad r \downarrow r_{\mu,\sigma} & \implies \eta \uparrow +\infty
\end{align*}
\]

which is located to the left of the axis \( r = 0.5 \) when the mean-to-volatility ratio is greater than zero, and to the right of the axis \( r = 0.5 \) when this ratio is less than zero.

Accordingly, as depicted in the upper graph of Figure 3.1, the Gaussian semi-elasticity is negative and shows an inverse U-shape when \( r < r_{\mu,\sigma} \), but it is positive and U-shaped when \( r > r_{\mu,\sigma} \). Besides, the greater the magnitude of the mean-to-volatility ratio, the more the asymptote \( r = r_{\mu,\sigma} \) approaches to the axis \( r = 0 \) when \( \mu/\sigma > 0 \), and the more the asymptote approaches to the axis \( r = 1 \) when \( \mu/\sigma < 0 \).

The dependence of the semi-elasticity on the mean-to-volatility ratio can be also precisely described. Indeed, from Equation 3.14:

\[
\begin{align*}
\frac{\mu}{\sigma} \downarrow -\infty & \implies \eta \downarrow 0 \quad \text{and} \quad \frac{\mu}{\sigma} \uparrow +\infty & \implies \eta \downarrow 0
\end{align*}
\]

as well as:

\[
\begin{align*}
\frac{\mu}{\sigma} \uparrow \Phi^{-1}(1-r) & \implies \eta \downarrow -\infty \quad \text{and} \quad \frac{\mu}{\sigma} \downarrow \Phi^{-1}(1-r) & \implies \eta \uparrow +\infty
\end{align*}
\]
3.5 The Neutral Gaussian Demand for Cash Balances

Fig. 3.1: Semi-Elasticity of the Demand for Cash Balances with Respect to the Rate of Interest.

Consequently, as depicted in the lower graph of Figure 3.1, the semi-elasticity of a Gaussian risk follows the path of an hyperbole with asymptote at the level:

\[
\frac{\mu}{\sigma} = \Phi^{-1}(1 - r)
\]

Thus, for example, for the interest rate levels \(r = 1\%, r = 5\%, r = 50\%\) and \(r = 84\%\) presented in Figure 3.1, the asymptotes are respectively located at the axis \(\mu/\sigma = 2.33\), \(\mu/\sigma = 1.67\), \(\mu/\sigma = 0\) and \(\mu/\sigma = -1\).

The fact that the semi-elasticity function shows an asymptote at some
critical state (characterised by the critical rate \( r_{\mu, \sigma} \) or the critical mean-to-volatility ratio \( \mu/\sigma = \Phi^{-1}(1 - r) \)) implies that the preference for liquidity of the market may become absolute under certain circumstances — and hence, that the market may attain the most unstable state, as explained in Section 3.1. The relevance of this result will be clarified later, in Section 3.8.

### 3.6 The Neutral Gaussian Demand with a Regulatory Condition

Most of financial institutions nowadays determine their minimum reserves (or capital) requirements based on the quantile function of the probability distribution describing the underlying risk, as stated in Equation 3.7. This principle is also known as Value-at-Risk (or VaR) in the risk management literature (see Hull, 2000). It has been adopted as an international standard by risk managers and regulators, a fact that is reflected in (and also induced by) the recommendations of the Basel Committee on Banking Supervision (2006).

Implementing the VaR model requires, in particular, to specify some probability level \( p \) where the quantile function is evaluated, which is related to the market interest rate \( r \) in Equation 3.7, in such a way that \( p = 1 - r \). As a general rule, raising the probability level implies that the probability of default is reduced (and vice-versa) and accordingly, this variable is interpreted as a confidence level. The levels chosen by institutions in practice normally range from 95\% to 99.9\%. Since there is no theoretically founded rule for determining the confidence level, the Basel Committee has proposed to specify a common and relatively conservative confidence level as a minimum standard — arbitrarily fixed at the level of 99\% in Basel, 1995.

The equivalence of the roles of probabilities and interest rates in the model of liquidity presented in this paper (as specified in Equations 3.6 and 3.7) implies that regulatory conditions depending on the confidence level actually correspond to restrictions imposed on the price of capital. As price restrictions, such conditions are likely to induce fluctuations in quantities in response to variations in the market price of capital or the riskiness implicit in the market portfolio. This conclusion supports some criticism recently attracted by the VaR model, which claims that although its implementation can certainly encourage better risk management practices, it can also menace the stability of financial markets, because the capital limits established by the model can vary with economic conditions (see e.g. Benford
At the light of basic economic theory, the alternative to this regulatory policy is straightforward: instead of exclusively depending on the price of balances, both the price and the amount of funds demanded at equilibrium should be considered. In fact, according to orthodox economic theory, every control or regulation should be avoided, for only in this way the most efficient allocation of resources is obtained. However, as will be soon formally stated, efficiency (or, more properly, deregulation) and stability are properties that do not always come together.

Indeed, let us start by considering some market where a regulatory condition is imposed on quantities, in such a way that firms and investors are obliged to fix their levels of reserves in order to ensure that their equity-to-risk ratios \( K/L \) are not lower than a certain fixed proportion \( \xi \). From Equation 3.13, such restriction is satisfied at the aggregate level as long as the following condition holds:

\[
\lambda_{\mu,\sigma}(r) \geq \xi \iff r \leq 1 - \Phi\left(\frac{\mu + \xi}{\sigma}\right) \tag{3.18}
\]

Then, as expected, establishing controls on quantities is equivalent to set price levels. In fact, the price limit is determined by a confidence interest rate that explicitly depends on the confidence ratio \( \xi \):

\[
r_{\xi} := 1 - \Phi\left(\frac{\mu + \xi}{\sigma}\right) = P\left\{\sigma \cdot X^{std} - \mu > \xi\right\} \tag{3.19}
\]

Consequently, every regulatory condition can be equivalently established by setting a minimum level \( \xi \) on the equity-to-risk ratios, by setting a maximum interest rate \( r_{\xi} \), or by fixing a confidence probability level \( p_{\xi} = 1 - r_{\xi} \).

Notice that given any fixed pair of the mean return \( \mu \) and the confidence ratio \( \xi \), the confidence rate \( r_{\xi} \) (as defined in Equation 3.19) is directly related to the volatility \( \sigma \), whereas for any fixed pair of the volatility and the confidence ratio, the confidence rate is inversely related to the mean return. Besides, as depicted in the upper graph of Figure 3.2, ceteris paribus, the confidence rate always follows a decreasing path on the confidence ratio.

The effect of implementing a regulatory condition can be assessed by comparing its related probability characterisation, as stated in Equations
Fig. 3.2: Confidence Rate and Semi-Elasticity of the Demand for Cash Balances as a Function of the Confidence Ratio under Gaussian Risks.

3.18 and 3.19, with that of solvent portfolios, determined by the condition $\lambda_{\mu,\sigma} > 0$, as in Equation 3.17. Solvency requirements are then directly related to the confidence rate $r_\xi$ and the critical rate $r_{\mu,\sigma}$ (defined in Equation 3.16):

$$\sigma \cdot X^{std} - \mu \geq \xi \geq 0 \iff r_\xi \leq r_{\mu,\sigma} \text{ with } r_\xi \to r_{\mu,\sigma} \text{ as } \xi \to 0$$

In other words, imposing a regulatory condition implies that an upper limit must be set to the cost of balances, lower than required for just assuring the solvency of the market portfolio.
Notice also that the class of risks satisfying the regulatory condition (i.e. the class of risks satisfying $\sigma \cdot X^{\text{std}} - \mu \geq \xi$) becomes smaller as the confidence ratio is incremented. Consequently, if the main interest of the regulatory authority were the level of activity of the market (measured in terms of the amount of portfolios simultaneously held by firms and investors), this coefficient should be set at a level as low as possible, preferable at the level zero. Or alternatively, the confidence rate $r_\xi$ should be fixed at a level as high as possible, preferable at the level $r_{\mu, \sigma}$ (or still equivalently, the confidence probability level $p_\xi$ should be fixed as a level as low as possible).

Additionally, those individuals that estimate higher returns on capital than the confidence rate $r_\xi$ would prefer to maintain lower equity-to-risk ratios than the confidence level $\xi$. They are thus obliged to incur in a loss of efficiency as long as they have to implement capital structures that are different from their preferred levels. However, as a compensation for this efficiency loss, regulations also induce individuals to take actions in order to reduce the probability of default of their portfolios. The benefit deduced from these actions are shared by the whole market.

In fact, since market adjustments, manifested as changes in the risk-parameters $\mu$ and $\sigma$, might induce solvent portfolios to go on bankruptcy, regulators must be not only concerned with activity and efficiency, but also with the stability of the market.

Recall that market stability can be assessed on the grounds of the semi-elasticity of the preference for liquidity function with respect to the interest rate. Assuming that most of investors and financial institutions determine their capital structures in order to fit the regulatory requirement of Equation 3.18 with exactitude, in such a way that the interest rate of equilibrium is very close to the confidence rate $r_\xi$, we obtain that the value of the semi-elasticity function evaluated at the confidence rate, denoted as $\eta_\xi$ in the following, can be used as an indicator of the stability of the market. Replacing Equation 3.19 in Equation 3.14, we actually obtain that:

$$\eta_\xi := \eta \left( r_\xi, \frac{\mu}{\sigma} \right) = -\frac{\sqrt{2\pi}}{\xi} \cdot \frac{\sigma}{\xi} \cdot \exp \left[ \frac{(\mu + \xi)^2}{2\sigma^2} \right]$$

or equivalently:

$$\eta_\xi = -\frac{\sigma}{\xi} \cdot \left[ \phi \left( \frac{\mu + \xi}{\sigma} \right) \right]^{-1} = \frac{-\sigma}{\xi \cdot P \{ \sigma \cdot X^{\text{std}} - \mu = \xi \}}$$
Therefore, the magnitude of the coefficient $\eta_\xi$ increases with volatility, but it diminishes with the confidence ratio and with the probability that the market portfolio satisfy the regulatory condition with exactitude. Hence the stability of the market diminishes with volatility, but it increases with the confidence ratio and with the probability that the market portfolio satisfy the regulatory condition with exactitude.

As depicted in the lower graph of Figure 3.2, the semi-elasticity $\eta_\xi$ follows an inverse U-shaped path as a function of $\xi$. This means that the magnitude of the semi-elasticity $|\eta_\xi|$ attains a minimum value at some confidence ratio $\xi$ and accordingly, that the market attains some state of maximum stability at some confidence ratio. In conclusion, a certain regulatory policy exists which induces maximum stability to the market.

So far, we have analysed the situation of stationary markets, i.e. markets with constant expected return and volatility. Additionally, the supply of cash balances has been assumed to be unbounded, in such a way that any amount of funds can be borrowed at the market interest rate. Under such circumstances, the total volume of reserves $K$ is completely determined by the demand for balances.

In the following section, the hypothesis of constant risk-parameters and unbounded capital supply will be released. We will thus obtain that for every aggregate stock $M$ and every volume $L$ demanded for transactions, a risk-structure of interest rates in the plane of expected returns and volatilities exists, where each rate represents the maximum marginal return (due to the reduction of bankruptcy costs) obtained by attracting an additional unit of capital. Such a return is connected to the confidence rate defined in Equation 3.19, although it is defined in a different context and will be accordingly given a different economic interpretation.

### 3.7 The Risk-Structure of Interest Rates

From Equations 3.9 and 3.13, the levels of the interest rate $r$, the cash-to-risk ratio $m$ and the risk-parameters $\mu$ and $\sigma$ compatible with the equilibrium in a Gaussian market of cash balances are determined by the following equation:

$$m = \sigma \cdot \Phi^{-1}(1 - r) - \mu$$
or equivalently:

\[
\frac{\mu + m}{\sigma} = \Phi^{-1}(1 - r) \iff \quad r = 1 - \Phi\left(\frac{\mu + m}{\sigma}\right)
\]  

(3.21)

For every fixed cash-to-risk ratio \( m \), a risk-structure of interest rates on the plane of mean returns and volatilities is thereby determined by Equation 3.21, where each rate represents the interest accrued by a unit of cash invested on the corresponding Gaussian risk at equilibrium, and where the discount factor \( \lambda_{\mu,\sigma}(r) \) represents the market price of risk — whereas, as already stated in Section 3.3, each rate in the term-structure of interest rates represents the return asked for receiving a cash-flow today instead of at some future date and the corresponding discount factor gives the price of a zero-coupon bond (see Equation 3.9 and the related discussion).

Thus, in particular, the discount factor \( \Phi^{-1}(1 - r) \) represents the market price of a standard Gaussian risk \( X_{std} \) (whose cumulative probability distribution is equal to the function \( \Phi \)) that offers the return \( r \). Therefore, as long as Gaussian risks with different mean returns and volatilities earn the same interest \( r \), the corresponding cash-to-risk ratios that are compatible with equilibrium are those that satisfy Equation 3.21. In other words, adjustments in the cash-to-risk ratios allow the guarantees of Gaussian risks with different parametrisations to be transacted at the same price, in such a way that lenders are indifferent about whom to supply their funds.

As depicted in Figure 3.3, given any cash-to-risk ratio \( m \) and any volatility level \( \sigma \), the expected return and the internal rate of return on risk are inversely related. Recall that the market price of risk is diminished when the IRRR is incremented and vice-versa — in the same way the price of a zero-coupon bond decreases when the IRR is incremented and vice-versa. Therefore, since the curve is moved to the left when the ratio \( m \) is raised, within a class of securities showing the same variability, the market price of risk (or equivalently, the level of the IRRR) can be maintained after a reduction of the mean return only if the cash-to-risk ratio is incremented.

The dependence of the risk-structure on volatility is shown in Figure 3.4. As depicted in the upper graph of Figure 3.4, within a class of securities offering the same positive expected return, the IRRR rises with volatility, at the time that for every level of volatility, the IRRR diminishes with the cash-to-risk ratio. Consequently, the market price of risk can be inflated both by incrementing liquidity and by controlling the variability of income.
— relations that are compatible with economic intuition.

In the lower graph of Figure 3.4, the risk-structure is shown for a class of funds offering a negative expected return equal in magnitude to the return used in the upper graph. Notice that the curve presents the same pattern in both situations, as long as \( m \) is greater in magnitude than the expected loss. However, the relationship between IRRRs and volatilities is reverted when \( m \) is lower than the expected loss, i.e. when the credit supply does not suffices to honour all outstanding liabilities, in such a way that higher IRRRs are observed for lower volatilities. In other words (and contrary to common intuition) when \( m < \mu \) the market price of risk increases with the level of volatility. Besides, \( IRRR \to +\infty \) when \( \sigma \to 0 \), i.e. the market price of risk converges to zero when \( \sigma \to 0 \). The asymptote separating both
3.7 The Risk-Structure of Interest Rates

**Fig. 3.4: Risk-Structure of Interest Rates as a Function of Volatility.**

Trends at the axis $r = 0.5$ is a consequence of the fact that both $\sigma \uparrow +\infty$ when $r \uparrow 0.5$ and $\sigma \uparrow +\infty$ when $r \downarrow 0.5$.

The *risk-structure* of interest rates presented in this section can be considered as an extension of the version presented by Robert Merton, 1974. In fact, as in the model of Merton, the curve provides a basis for the pricing of *risky coupon bonds*, which explicitly depends on the return offered by a certain class of money substitutes and the *probability of default* of the instrument. Unlike the model of Merton, however, borrowing restrictions are regarded as an *intrinsic* quality of markets, in such a way that balances and securities *cannot* be transacted to any desired extent. This means that hedging cannot be continuously implemented, a condition that is at the base of the Merton’s model.
Consequently, the interest rate $r$ in Equation 3.21 is related to debt that can be risky in terms of default (whereas Merton, 1974, considers free of default debt). In fact, the terms of liquidity, determined by the cash stock $M$ and the amount of funds $L$ spent on securities, explicitly affects equilibrium in Equation 3.21 through the cash-to-risk ratio $m = M/L$. On these grounds, we can claim that the internal return on risk (as determined by Equation 3.21) simultaneously accounts for liquidity and market risk.

Besides, since the internal return (and hence the market price) of risk only depends on the cash-to-risk ratio and the risk-parameters $\mu$ and $\sigma$, and not on the particular features of the instrument, the risk-structure derived from Equation 3.21 provides a common basis for the pricing of a wide class of financial instruments, including bonds, stocks and financial derivatives.\footnote{The only requirement is the risk-parameters $\mu$ and $\sigma$ to be calculated in the same base: i.e. they must represent the mean return and the standard deviation of the series of P&L of the market valorisations of the instruments.}

It is worth noticing, finally, that although the risk-structure follows a continuous path when the underlying risk is described by a Gaussian probability distribution (more generally, when the underlying risk follows a continuous probability distribution), the rate of variation of the curve with respect to the risk-parameters is not linear (in fact, the curve can degenerate under certain circumstances, see Figures 3.3 and 3.4). This means that state transitions, manifested as variations in the risk-parameters, can sometimes imply severe adjustments in the rate of interest attained at equilibrium.

In the next section, we will investigate how this situation affects the market equilibrium (as determined by Equations 3.1 and 3.2), or more precisely, how the market can be affected by monetary interventions, implemented both by the lenders of balances or by the buyers of securities (who respectively control the variables $M$ and $L$).

### 3.8 Deviations from Equilibrium in Neutral Markets

We already know that in order to preserve the equilibrium in the market of cash balances (which is specified in Equation 3.1) variations in the aggregate stock of cash $M$ and the funds spent on securities $L$ must be necessarily followed by interest rate adjustments, in such a way that, as deduced from Equations 3.2 and 3.3:
\[
\frac{\Delta M}{M} - \frac{\Delta L}{L} = \eta(r) \cdot \Delta r
\] (3.22)

Therefore, as long as \( \eta(r) < 0 \), increments in the logarithmic (or relative) supply of balances and the logarithmic (or relative) amount of funds invested on securities must be respectively followed by reductions and increments in the rate of interest at equilibrium, i.e.:

\[
\Delta \ln M = \frac{\Delta M}{M} \uparrow \Rightarrow \Delta r \downarrow \\
\Delta \ln L = \frac{\Delta L}{L} \uparrow \Rightarrow \Delta r \uparrow
\]

Both relationships make economic sense, for the interest rate is indeed expected to decrease as the availability of credit is raised, while by contrast, the pressure established by an increasing demand for cash balances is expected to induce the cost of capital to rise.

Accordingly, regimes inducing \( \Delta \ln M > 0 \) and \( \Delta \ln L < 0 \) are corresponded to expansions of the market of cash balances, whereas \( \Delta \ln M < 0 \) and \( \Delta \ln L > 0 \) are corresponded to market contractions.

Let us consider the particular case when the supply of balances is modified by a certain amount \( \Delta M \) while the funds spent on securities remain unaltered, i.e. \( \Delta L = 0 \). The variations in the supply of reserves and the interest rate must then be related to each other according to Equation 3.22, in such a way that:

\[
\Delta M = M \cdot \eta(r) \cdot \Delta r
\]

Therefore, as long as \( \eta(r) < 0 \), the semi-elasticity represents the stock reduction, per unit of reserves, required to raise the equilibrium interest rate in one percentage point. Similarly, when the supply of reserves remains constant, i.e. when \( \Delta M = 0 \), variations in the interest rate and the funds invested on securities must be related to each other as:

\[
\Delta L = -L \cdot \eta(r) \cdot \Delta r
\]

Hence, the semi-elasticity can be alternatively interpreted as the percentage increment in the funds invested in securities required to raise the equilibrium
interest rate in one percentage point.

On these grounds, we will define the \textit{transmission rate} $\tau(r)$ as the stock reduction, or investment increase, per million of cash units, that is required to produce a movement of one \textit{basis point} in the equilibrium interest rate, where one \textit{percentage} point is equal to 100 \textit{basis points}:

\[
\tau(r) = -(1 \text{ million}) \cdot \eta(r) \cdot \frac{1}{100} \quad \Rightarrow \quad \tau(r) = -10^4 \cdot \eta(r) \tag{3.23}
\]

Then the function $\tau(r)$ represents the credit \textit{supply increment}, per million invested on guarantees, required to diminish the equilibrium interest rate in one basis point, or alternatively, it represents the credit \textit{demand reduction}, per million invested on securities, required to diminish the equilibrium interest rate in one basis point.

Since the transmission rate is equal to the semi-elasticity function multiplied by a negative constant, the graphic representations of both coefficients have to be the same, except for a scaling factor and the inversion of the sign.

Accordingly, based on the paths depicted in Figure 3.1, in a Gaussian setting we expect the transmission rate $\tau(r)$ to be positive and \textit{U-shaped} to the left of the \textit{critical} interest rate $r_{\mu,\sigma}$ (defined in Equation 3.16) and negative and \textit{inverse U-shaped} to the right of the critical rate, in such a way that:

\[
\begin{align*}
\tau(r) &\to +\infty \quad \text{when} \quad r \to 0 \\
\tau(r) &\to +\infty \quad \text{when} \quad r \to r_{\mu,\sigma} \quad \text{with} \quad r < r_{\mu,\sigma} = 1 - \Phi \left( \frac{\mu}{\sigma} \right) \\
\tau(r) &\to -\infty \quad \text{when} \quad r \to r_{\mu,\sigma} \quad \text{with} \quad r > r_{\mu,\sigma} = 1 - \Phi \left( \frac{\mu}{\sigma} \right) \\
\tau(r) &\to -\infty \quad \text{when} \quad r \to 1
\end{align*}
\]

Therefore, the model predicts that some interest rate exists in the interval $(0, r_{\mu,\sigma})$ for which a \textit{minimum} cash variation is needed to produce a movement of one basis point in the cost of capital. However, the required cash variation tends to infinite when $r \to 0$ as well as when $r \to r_{\mu,\sigma}$. This means that when the cost of capital approaches to the critic values $r = 0$ or $r = r_{\mu,\sigma}$ only an \textit{infinite} increment of the credit supply can induce the

\*Among practitioners, it is customary to express interest rates in terms of \textit{basis} instead of \textit{percentage} points. See e.g. Hull, 2000.
interest rate to decrease in one basis point — or equivalently, only a reduction \textit{infinite} in magnitude, in the cash balance demanded for transactions, can induce the interest rate to raise one basis point.

These results are inverted when \( r > r_{\mu,\sigma} \), for in this case the semi-elasticity function changes its sign, i.e. \( \eta(r) > 0 \) when \( r > r_{\mu,\sigma} \). In particular, when \( r = r_{\mu,\sigma} \) and when \( r = 1 \) only an \textit{infinite} reduction of the credit supply can induce the interest rate to decrease in one basis point. The required cash balance variation is then \textit{undetermined} around the level \( r = r_{\mu,\sigma} \).

Regarding the path of the transmission rate as a function of the mean-to-volatility ratio, still on the grounds of Figure 3.1, we expect to obtain a \textit{hyperbole} with a positive branch to the left of the critic mean-to-volatility ratio and a negative branch to the right of it:

\[
\tau(r) \to 0 \quad \text{when} \quad \frac{\mu}{\sigma} \to -\infty \\
\tau(r) \to +\infty \quad \text{when} \quad \frac{\mu}{\sigma} \to 1 - \Phi^{-1}(r) \quad \text{with} \quad \frac{\mu}{\sigma} < 1 - \Phi^{-1}(r)
\]

\[
\tau(r) \to -\infty \quad \text{when} \quad \frac{\mu}{\sigma} \to 1 - \Phi^{-1}(r) \quad \text{with} \quad \frac{\mu}{\sigma} > 1 - \Phi^{-1}(r)
\]

\[
\tau(r) \to 0 \quad \text{when} \quad \frac{\mu}{\sigma} \to +\infty
\]

Hence the magnitude of the cash balance variation required to produce the interest rate to move one basis point converges to zero both when the mean-to-volatility ratio converges to zero and when it converges to +\( \infty \). This magnitude is however \textit{undetermined} around the critic ratio \( 1 - \Phi^{-1}(r) \), a point at which a \textit{phase transition} is produced in the market — characterised by a change in the sign of the semi-elasticity of the market’s liquidity-preference.

The predictions of the model so far stated for Gaussian markets can be readily verified on the grounds of the results depicted in Table 3.1. Indeed, inspecting the columns of Table 3.1 we observe, on the one hand, that for every level of the interest rate, two different regimes are distinguished: first \( \tau(r) \) is \textit{positive} and \textit{increasing} as a function of the mean-to-volatility ratio, but later it is \textit{negative} and \textit{decreasing in magnitude} as a function of the mean-to-volatility ratio. We already know that a phase transition is produced at the \textit{critical} ratio \( \mu/\sigma = 1 - \Phi^{-1}(r) \) (which appears between brackets below each interest rate in Table 3.1), leading the transmission rate \( \tau(r) \) from positive to negative values.

On the other hand, looking at the rows of Table 3.1, we notice that given
any fixed level of the mean-to-volatility ratio, again positive and negative phases are distinguished. In the first regime, when \( \tau(r) > 0 \), \( \tau(r) \) shows a decreasing followed by an increasing dependence on the interest rate, while in the second regime, when \( \tau(r) < 0 \), \( \tau(r) \) shows an increasing followed by a decreasing dependence on the interest rate. The transition is produced at the critical rate \( r_{\mu,\sigma} = 1 - \Phi(\mu/\sigma) \) (which appears between brackets after every mean-to-volatility ratio in Table 3.1).

Table 3.1: Net Balances required per Million of Invested Cash Units to induce the Interest Rate to vary one Basis Point, given different Interest Rates and Mean-to-Volatility Ratios.

<table>
<thead>
<tr>
<th>( \mu/\sigma = -2.00 )</th>
<th>( \mu/\sigma = -1.00 )</th>
<th>( \mu/\sigma = -0.60 )</th>
<th>( \mu/\sigma = -0.25 )</th>
<th>( \mu/\sigma = -0.10 )</th>
<th>( \mu/\sigma = 0.01 )</th>
<th>( \mu/\sigma = 0.10 )</th>
<th>( \mu/\sigma = 0.25 )</th>
<th>( \mu/\sigma = 0.60 )</th>
<th>( \mu/\sigma = 1.00 )</th>
<th>( \mu/\sigma = 2.00 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r = 5% )</td>
<td>( r = 10% )</td>
<td>( r = 15% )</td>
<td>( r = 20% )</td>
<td>( r = 30% )</td>
<td>( r = 1% )</td>
<td>( r = 5% )</td>
<td>( r = 10% )</td>
<td>( r = 15% )</td>
<td>( r = 20% )</td>
<td>( r = 30% )</td>
</tr>
<tr>
<td>(2.33)</td>
<td>(1.64)</td>
<td>(1.04)</td>
<td>(0.84)</td>
<td>(0.52)</td>
<td>(0.33)</td>
<td>(0.64)</td>
<td>(1.04)</td>
<td>(0.84)</td>
<td>(0.52)</td>
<td>(0.33)</td>
</tr>
<tr>
<td>26,602</td>
<td>36,660</td>
<td>43,192</td>
<td>51,170</td>
<td>55,569</td>
<td>58,591</td>
<td>59,308</td>
<td>62,763</td>
<td>69,512</td>
<td>92,797</td>
<td>150,359</td>
</tr>
<tr>
<td>17,364</td>
<td>24,975</td>
<td>30,284</td>
<td>37,205</td>
<td>41,244</td>
<td>44,118</td>
<td>44,812</td>
<td>48,225</td>
<td>55,238</td>
<td>83,604</td>
<td>150,359</td>
</tr>
<tr>
<td>14,125</td>
<td>21,061</td>
<td>26,209</td>
<td>33,340</td>
<td>37,740</td>
<td>40,986</td>
<td>41,785</td>
<td>45,801</td>
<td>54,536</td>
<td>98,272</td>
<td>1,177,193</td>
</tr>
<tr>
<td>12,570</td>
<td>19,395</td>
<td>24,777</td>
<td>32,721</td>
<td>37,934</td>
<td>41,943</td>
<td>42,951</td>
<td>48,164</td>
<td>60,375</td>
<td>147,831</td>
<td>-225,530</td>
</tr>
<tr>
<td>11,393</td>
<td>18,867</td>
<td>25,579</td>
<td>37,140</td>
<td>46,062</td>
<td>53,819</td>
<td>55,912</td>
<td>67,769</td>
<td>104,814</td>
<td>-380,439</td>
<td>-60,473</td>
</tr>
</tbody>
</table>

Two main conclusions of the model of equilibrium presented in this paper are thereby illustrated in Table 3.1 — concerning the state and the evolution of the markets of cash balances.

Firstly, regarding the equilibrium states, the amount of funds required to induce a certain movement in the equilibrium interest rate widely varies with the market conditions, i.e. with the levels of expected return and volatility. In other words — and contrary to common beliefs — the amount
of funds required to produce a given expansion (a regime characterised by decreasing interest rates) or a given contraction (characterised by increasing interest rates) significantly differ from one state of equilibrium to another. Thus, although in certain states this amount can be low, let us say in the order of thousand of cash units per million invested, in others it can raise until surpasses the million of cash units per million invested.

Secondly, there are certain combinations of the interest rate and the mean-to-volatility ratio, characterised by the condition \( r = 1 - \Phi(\mu/\sigma) \), for which the semi-elasticity function is undetermined (for, as depicted in Figure 3.1, the semi-elasticity function respectively converges to \(-\infty\) and \(+\infty\) depending on whether the interest rate approaches the critical rate \( r_{\mu,\sigma} = 1 - \Phi(\mu/\sigma) \) from the left or from the right). This implies that, under certain circumstances, small oscillations of the interest rate can lead to big adjustments in the cash-to-risk ratio \( m = M/L \), i.e. in the investment decisions taken by the lenders of balances and investors, and hence to induce instability in the market of cash balances.

Such combinations of the parameters can thus be corresponded to critical states. When markets enter in any of such states — which can be corresponded to times of liquidity crises — no variation in the aggregates \( M \) or \( L \) is high enough to produce the necessary adjustment in the interest rate to push the market to a new equilibrium.

In this context, liquidity shortfalls are interpreted in the model as purely physical adjustments. Only structural adjustments, performed through the mean return or the volatility of the market portfolio, can make the market to return to equilibrium under such circumstances.

### 3.9 Non-Neutral Markets of Cash Balances

As stated in the previous section, state transitions in the markets of balances are characterised by the semi-elasticity function \( \eta(r) \). More precisely, this function determines how the equilibrium interest rate is affected by changes in the amount of funds invested on securities and cash holdings, or equivalently, how much these funds must be adjusted in order to obtain a certain movement of the interest rate attained at equilibrium — as determined by the transmission rate defined in Equation 3.23.

We will now examine how the conclusions found in the previous section
Chapter 3: Equilibrium in Markets of Cash Balances

are extended to markets where the underlying risk is described by some general probability distribution, not necessarily Gaussian, and where individuals respectively distort probabilities and payments according to some informational type $\theta$ and some scaling factor $\psi$ (as specified in Equations 3.5 and 3.10).

Let us start by noticing that if $\lambda(r) = T_{-X}^{-1}(r)$ denotes the preference for liquidity of some neutral market whose aggregate exposure is represented by the random variable $X$ (obtained by replacing the values $\theta = 1$ and $\psi = 1$ in Equation 3.12), then the semi-elasticity function is given from Equation 3.3 as:

$$
\eta_{-X}(r) = \frac{1}{\lambda(r)} \frac{d\lambda(r)}{dr} = \frac{d\ln \lambda(r)}{dr} = \frac{d\ln T_{-X}^{-1}(r)}{dr} \quad (3.24)
$$

An alternative expression can be obtained that explicitly depends on the tail probability function $T_{-X}$ (and not on the quantile function $T_{-X}^{-1}$), since for every $\lambda = T_{-X}^{-1}(r)$:

$$
T_{-X}(\lambda) = T_{-X}(T_{-X}^{-1}(r)) = r \implies \frac{d}{d\lambda} T_{-X}(\lambda) \cdot \frac{d}{dr} T_{-X}^{-1}(r) = 1
$$
in such a way that:

$$
\frac{d}{dr} T_{-X}^{-1}(r) = \left[ \frac{d}{d\lambda} T_{-X}(\lambda) \right]^{-1} \quad \text{with} \quad \lambda = T_{-X}^{-1}(r)
$$

$$
\implies T_{-X}^{-1}(r) \cdot \frac{1}{T_{-X}(r)} \frac{d}{dr} T_{-X}^{-1}(r) = \left[ T_{-X}(\lambda) \cdot \frac{1}{T_{-X}(\lambda)} \frac{d}{d\lambda} T_{-X}(\lambda) \right]^{-1}
$$

Hence,

$$
T_{-X}^{-1}(r) \cdot \frac{d}{dr} \ln T_{-X}^{-1}(r) = \left[ T_{-X}(\lambda) \cdot \frac{d}{d\lambda} T_{-X}(\lambda) \right]^{-1} \quad \text{with} \quad \lambda = T_{-X}^{-1}(r)
$$
which means that the semi-elasticity can be also defined as:

\[
  r \cdot \eta_{-X}(r) = \left[ \lambda \cdot \frac{d \ln T_{-X}(\lambda)}{d\lambda} \right]^{-1} \quad \text{with} \quad \lambda = T_{-X}^{-1}(r) \tag{3.25}
\]

In fact, the derivative of the logarithm of the tail probability function is a well known measure in actuarial research, which is known as the hazard-rate (see Wang, 1995):

\[
  h_{-X}(\lambda) := -\frac{d \ln T_{-X}(\lambda)}{d\lambda} \tag{3.26}
\]

The hazard-rate \( h_{-X} \) thereby measures the reduction in the probability of default as the proportion of reserves \( \lambda \) (with respect to the cash flow \( L \) invested on securities) is increased. It can be alternatively interpreted as a failure-rate, for it measures the rate at which the frequency of defaults in the market is diminished as the proportion of reserves is increased — or in other words, it measures the reduction in the number of insolvent portfolios resulting from increasing the aggregate cash balance.

The greater the hazard-rate, the more the probability of default is reduced per decimal point increase in the proportion of reserves (or the more the amount of firms and investors that are rescued from bankruptcy per additional decimal point in the proportion of reserves) and hence the more efficient is the cash guarantee in fulfilling its role. In the limit when \( h_{-X} \to +\infty \), a very small increment in cash stocks leads to a big reduction in the probability of bankruptcy. On these grounds, the role of cash holdings in reducing the risk of portfolios is explicitly stated.

The efficiency of cash guarantees can be equivalently established in terms of the semi-elasticity of the preference for liquidity. Indeed, from Equations 3.25 and 3.26, we obtain that:

\[
  r \cdot \eta_{-X}(r) = \frac{-1}{\lambda \cdot h_{-X}(\lambda)} \quad \text{with} \quad \lambda = T_{-X}^{-1}(r) \tag{3.27}
\]

Therefore, given any fixed pair \( (r, \lambda) \) of the interest rate and the liquidity-preference, more efficient markets of balances, which are corresponded to higher hazard-rates according to the previous analysis, are also characterised
by lower semi-elasticity values. This conclusion is in line with the definition of the semi-elasticity function, for the greater the semi-elasticity, the more sensible is the demand for cash balances with respect to the interest rate and hence, the greater the proportion of reserves that is purely required to satisfy the (increasing) liquidity-preference of decision-makers.

As already stated in Sections 3.1 and 3.8, in the limit when \( \eta_{-X} \to -\infty \), individuals are satisfied at a single level of the interest rate (liquidity-preference is absolute in this case) and hence, variations in the relative stock \( m = M/L \) are entirely absorbed by adjustments in the preference for liquidity of the market. The market is the most unstable at this point.

From Equation 3.27, the condition \( \eta_{-X} \to -\infty \) implies that \( h_{-X} \to 0 \). Under such circumstances, increasing the proportion of reserves \( \lambda \) produces no effect on the probability of default \( T_{-X}(\lambda) \), i.e. cash holdings become useless for reducing the riskiness of the market portfolio. The demand for balances cannot be satisfied in this situation, for firms and investors will try to liquidate all their risky assets — in order to substitute them for cash holdings.

Let us now assume that individuals distort the probability distribution according to the parameters \( \theta \) and \( \psi \), as in Equation 3.12. Then,

\[
\lambda = T^{-1}_{\theta_{-X}} \left( \frac{r}{\psi} \right) \quad \Rightarrow \quad r = \psi \cdot T_{\theta_{-X}}(\lambda) = \psi \cdot T_{-X}(\lambda)^{\frac{1}{\theta}}
\]

Define the distorted tail probability function \( T_{-X}^{*} \) in the following way:

\[
T_{-X}^{*}(\lambda) := \psi \cdot T_{-X}(\lambda)^{\frac{1}{\theta}}
\]

in such a way that:

\[
\ln T_{-X}^{*}(\lambda) = \ln \psi + \frac{1}{\theta} \cdot \ln T_{-X}(\lambda) \quad \Rightarrow \quad \frac{d \ln T_{-X}^{*}(\lambda)}{d \lambda} = \frac{1}{\theta} \cdot \frac{d \ln T_{-X}(\lambda)}{d \lambda}
\]

Therefore,

\[
h_{-X}(\lambda) = \frac{1}{\theta} \cdot h_{-X}(\lambda) \quad \forall \lambda \quad \forall \theta > 0 \quad \forall \psi \geq 1 \quad (3.28)
\]
Replacing Equation 3.28 into Equation 3.27 we additionally obtain that:

$$\eta^*_X(r) = \theta \cdot \eta_X(r) \quad \forall r \quad \forall \theta > 0 \quad \forall \psi \geq 1 \quad (3.29)$$

Hence market transitions are not affected by the scaling factor $\psi$, because neither the semi-elasticity nor the hazard-rate is affected by this parameter. The informational type $\theta$, on the other hand, has the effect of scaling the semi-elasticity function.\textsuperscript{9}

### 3.10 Conclusions

A theoretical framework has been presented in this chapter to describe the equilibrium in a market of cash balances (or inter-bank loans) with opaque intermediaries faced to liquidity restrictions.

First the optimal demand for balances is characterised as the optimal level of reserves maintained to avoid the losses endured when holding portfolios with random outcomes. The problem is formulated in actuarial terms, in such a way that its solution is expressed as the quantile function (or Value-at-Risk) of the probability distribution representing the underlying risk.

Within this framework, the equilibrium in the market of cash balances determines the amount of funds transacted in the market, as well as the rate at which a unit of capital is exchange by a unit of risk (or the Internal Rate of Return on Risk, as it is called in Section 3.3, abbreviated IRRR), i.e. it determines the market price of risk.

Therefore, the equilibrium in the market of cash balances actually implies that two markets must be simultaneously found at equilibrium: the market of balances, characterised by the equilibrium interest rate, and the market of securities, which in a Gaussian setting is characterised by the mean return and the volatility of the aggregate portfolio. This result makes sense from the actuarial point of view, because actuarial prices are strictly risk-based.

As a conclusion, variations in the supply of balances or the funds spent

\textsuperscript{9}Strictly speaking, the function $T^*_X$ is not a probability function, since as long as the scaling factor is greater than one, some liquidity-preference coefficient $\lambda$ may exist such that $T^*_X(\lambda) > 1$. However, the distorted probability $T^*_X$ is still a well defined mathematical measure (in the sense of Lebesgue, see e.g. Nielsen, 1997), in such a way that the distorted probability principle (see Equation 3.5) can be defined on its basis.
on securities not only imply adjustments in the equilibrium interest rate, as stated in classical economic analysis, they also induce adjustments in the mean return and the volatility of the aggregate portfolio.\(^\text{10}\)

A precise description of the liquidity state of the market of balances is thereby obtained, in terms of the amount of funds allocated to cash reserves and securities, the equilibrium interest rate and the risk-parametrisation of the aggregate portfolio.

The market stability, on the other hand, is measured by the semi-elasticity of the liquidity-preference function with respect to the interest rate — equal to the logarithmic variation of the preference for liquidity with respect to the interest rate. This function explicitly depends on the risk-parameters of the underlying exposure, i.e. on the liquidity state.

One important consequence of the model presented in this chapter is the existence of multiple equilibria — or multiple liquidity states that are compatible with equilibrium in the market of cash balances. This implies, in particular, that it is not always possible to implement monetary interventions that guarantee market stability — guaranteeing convergence of prices and allocations to some equilibrium state.

In fact, even if such interventions were possible, the existence of unstable equilibria implies that the market can be pushed away to some different set of allocations and prices once the convergence process terminates.

As a matter of fact, in a Gaussian setting, instability can be corresponded to some specific combinations of the internal return on risk and the mean-to-volatility ratio of the aggregate portfolio — which imply that the magnitude of the semi-elasticity function converges to infinite. Such states can then be naturally corresponded to times of liquidity crises, a phenomenon that in the model is interpreted as a purely physical adjustment.

\(^{10}\) As long as the market price of securities is strictly risk-based, this result is compatible with basic economic theory, according to which every change in the quantities transacted in the market must be necessarily followed by an adjustment in the price attained at equilibrium.
Appendix A

Utility Theory of Choice under Risk

In its crudest form, the utility principle states that acts or decisions are judged by individuals according to their consequences, which in turn are related to the comfort or happiness they bring to decision-makers. The best known essay on utilitarianism is that of John Stuart Mill, published in 1861 (Mill, 1861), where the author introduces the principle for the description of the ethical behaviour of individuals.¹

The notion of utility as a quantitatively measure of satisfaction, however, is formalised much later by Von Neumann and Morgenstern (1944). They justify this choice claiming that only if consequences can be numerically assessed, then the preferences of individuals about them can be described from a mathematical and economic perspective. Their ultimately goal is to provide a satisfactory treatment of the question of rational behaviour, which helps to deal with more specific economic problems.

Within this context, every utility function is corresponded to some real function \( u : \mathbf{X} \rightarrow \mathbb{R} \), defined over the space of consequences \( \mathbf{X} \). Every utility function thereby determines an order of consequences, in such a way that for every pair of consequences \( X, Y \in \mathbf{X} \), \( X \) is said to be preferred by \( Y \) if the utility (or satisfaction) that it induces is lower than the utility induced by \( Y \), i.e. if \( u(X) < u(Y) \). The notation \( X < Y \) is then introduced. In other

¹Early formulations of the principle can be found in the works of David Hume and specially Jeremy Bentham during the second half of the 18th. and the first half of the 19th. century.
words, every set of preferences (which ultimately characterises the behaviour of some decision-maker) can be corresponded to some utility function.

In a world with incomplete knowledge, consequences are represented by random variables, which are completely described by probability distributions. Hence every set of consequences is determined by a certain set of events and their corresponding probabilities.2

As stated in the expected utility theorem, the utility of any lottery can be calculated as a linear combination of the utilities of the underlying events, provided that the order of consequences induced by utilities satisfies the axioms of completeness, transitivity, convexity and independence.

The axiom of completeness imposes that every pair of prospects can always be compared, i.e. \( \forall X, Y \in X, X \leq Y \) or \( X \geq Y \), whereas the axiom of transitivity implies that if \( X \leq Y \) and \( Y \leq Z \) then \( X \leq Z \). The axiom of convexity, on the other hand, requires that if \( X \leq Y \leq Z \), then a level of probability \( p \) exists, such that the combination \( p \cdot X + (1 - p) \cdot Y \) is preferred to \( Z \). Finally, the axiom of independence implies that if \( X \) is equally preferred to \( Y \), then for every level of probability \( p \) and every prospect \( Z \), the combination \( p \cdot X + (1 - p) \cdot Z \) is equally preferred to the combination \( p \cdot Y + (1 - p) \cdot Z \).

Within this framework, Von Neumann and Morgenstern propose to relate rational behaviour to the maximisation of expected utility:

\[
E[u(X)] = \int u(x) \cdot dF_X(x) \quad (A.1)
\]

The fact that if utilities can be expressed in numerical terms then they can be naturally connected to monetary rewards has supported the use of the utility maximisation approach in a number of economic constructs. In particular, three of the major tools of financial economics, namely the Portfolio Allocation method of Harry Markowitz (1952), the characterisation of Liquidity Preference of James Tobin (1958), and the Capital Asset Pricing Model of William Sharpe (1964), are all obtained from the maximisation of

---

2In classic utility theory, individuals are assumed to choose among lotteries or risky prospects, as in Von Neumann and Morgenstern, 1944, Friedman and Savage, 1948, Arrow, 1951, Aumann, 1962 and Fishburn, 1968. In a more general context, De Finetti, 1975, states that the notion of probability provides a range of gradations of possibilities, which is dependent on the state of information and which permit a limitation of expectations. See also Tversky, 1967.
some utility function.

In expected utility theory, the *attitude towards risk* or *aversion to risk* of decision makers is explicitly determined by the utility function. Indeed, recall that *risk aversion* is defined as the reluctance to exchange some uncertain prospect for another non-risky, though possibly lower, payment. Two *measures* of aversion to risk are thus introduced in classic theory, the *absolute* and *relative* aversion coefficients, respectively defined as:

\[
\alpha^{\text{ABS}} = -\frac{u''(x)}{u'(x)} \quad \text{and} \quad \alpha^{\text{REL}} = -\frac{x \cdot u''(x)}{u'(x)}
\]

Both measures of risk aversion depend on the *curvature* of the utility function \(u(\cdot)\) (see Pratt, 1964, Diamond and Stiglitz, 1974, and also Ross, 1981).

Accordingly, the behaviour of *averse-to-risk* and *risk-lover* individuals is respectively characterised by *concave* and *convex* utility functions.³ The behaviour of *risk-neutral* individuals, on the other hand, is represented by the *identity* \(u(x) = x, \forall x\).

Much of the criticism attracted by the expected utility theory has focused on the fact that some of its predictions enter in contradiction with actual observed choices, see e.g. Allais, 1953. Kahneman and Tversky (1979) present experimental evidence on the subject and propose an alternative model, called *prospect theory*, to describe the process of decision making under risk (see also Loomes and Sugden, 1982).

Another approach is proposed by Quiggin (1982, 1991), who suggests to substitute the expectation operator for a *weighted sum of utilities* — or *anticipated utility* — in such a way that the *distortion* of (subjective) probabilities is permitted. The theory of anticipated utility is built on the base of an alternative set of axioms, including *weakened* versions of the axioms of *convexity* and *independence*. Machina (1982, 1987) actually demonstrates that the basic concepts and results of expected utility theory can be maintained if the axiom of independence is eliminated — and the weaker assumption of smoothness of preferences over alternative probability distributions is introduced.

³A real valued function \(u(\cdot)\) is said to be *concave* if the following inequality holds for every \(p \in [0, 1]\):

\[
\forall x, y \quad u(p \cdot x + (1 - p) \cdot y) \leq p \cdot u(x) + (1 - p) \cdot u(y)
\]

When the inequality is inverted, the function \(u(\cdot)\) is said to be *convex*. 
Quiggin (1982) demonstrates that the first stochastic order is preserved by every utility function that satisfies the axioms of the theory of anticipated utility. This means that every two risky prospects can be always ordered by means of some anticipated utility function and hence, that the preferences of every individual can be represented by some anticipated utility function. Roell (1987) proves in addition that every anticipated utility function admits an integral representation.

In fact, an alternative theory of choice under risk can be built on this basis. Yaari (1987) demonstrates that this theory can be formally established on the grounds of a set of axioms that are explicitly dependent on the probability distributions describing risks (see also Roell, 1987). Unlike expected utility theory — where independence is required with respect to probability mixtures of risky prospects — in the alternative theory independence is required with respect to direct mixing of payments of risky prospects. On these grounds, Yaari baptises this theory as the dual theory of choice under risk.
Appendix B

The Dual Theory of Choice

Let us consider a class $X$ of probability distributions — representing some set of consequences or states of the world. We will introduce a dual order $\leq^*$ over the class $X$, which satisfies the axioms of completeness and transitivity, as formulated in Appendix A. The axioms of convexity and independence, however, will be replaced by the axioms of continuity and dual independence.

The axiom of continuity imposes a technical condition\(^1\) implying that if $X \leq^* Z$ and $Y \leq^* Z$, then $Z$ is preferred in the dual order to any combination $p \cdot X + (1 - p) \cdot Y$, for all $p \in [0, 1]$. The axiom of dual independence, on the other hand, requires independence in the space of quantiles of the probability distributions describing risks, in such a way that if $X$ is preferred to $Y$, then for every $p \in [0, 1]$ the following inequality must hold:

\[
(p \cdot F_X^{-1} + (1 - p) \cdot F_Z^{-1})^{-1} \geq (p \cdot F_Y^{-1} + (1 - p) \cdot F_Z^{-1})^{-1}
\]

where $F_X^{-1}$, $F_Y^{-1}$ and $F_Z^{-1}$ denote the quantile or inverse functions of their corresponding probability distributions.\(^2\)

These axioms are sufficient to guarantee the existence of a real-valued and

---

\(^1\)Putted in mathematical terms, the continuity axiom requires that the sets $\{p : p \cdot X + (1 - p) \cdot Y \leq Z\}$ and $\{p : p \cdot X + (1 - p) \cdot Y \leq Z\}$ are closed with respect to the $L_1$-norm. See Yaari, 1987, and also Roell, 1987.

\(^2\)Formally, the quantile function is defined as:

\[
F_X^{-1}(p) = \sup_{x \leq p} \{ x : F(x) = q \} \quad \forall p \in [0, 1]
\]
linear function representing the dual order \( \leq^* \) (in the sense of Herstein and Milnor, 1953, see Roell, 1987). It is possible to prove, moreover, that the dual order can be given an integral representation if an additional axiom is introduced.

Indeed, let us impose the dual order to satisfy the axiom of \textit{monotonicity}, which requires that the \textit{first stochastic order} must be preserved (i.e. \( X \leq^* Y \) must imply that \( T_X(x) \leq T_Y(x) \) \( \forall x \), where \( T_X \) and \( T_Y \) denote the corresponding \textit{tail probability} functions). Then for any pair of risky prospects \( X, Y \):

\[
X \leq^* Y \iff E_\phi[X] \leq E_\phi[Y]
\]

with \( E_\phi \) denoting the integral operator:

\[
E_\phi[X] = \int x \cdot dF_{\phi,X}(x) = \int T_{\phi,X}(x) \cdot dx \quad \text{with} \quad T_{\phi,X}(x) = \phi(T_X(x)) \quad \forall x
\]  

(B.1)

where \( \phi \) denotes a continuous and non-decreasing function defined on the unit interval, and where \( F_{\phi,X}(x) = P_{\phi}\{X \leq x\} \) and \( T_{\phi,X}(x) = P_{\phi}\{X > x\} \) respectively denote the \textit{distorted} cumulative and tail probability functions.

The operator \( E_\phi \) defined in \textit{Equation} B.1 can then be regarded as an alternative \textit{measure} of risk to the operator \( E_u \) defined in \textit{Equation} A.1, which distorts \textit{probabilities} instead of \textit{payments}. It will be known as the \textit{distorted probability principle} in the following.

Wang et al. (1997) propose an axiomatic characterisation for the \textit{distorted probability principle} (see Appendix C). Besides, as demonstrated by Wang (1996), the \textit{distorted probability principle} can be formally considered as an \textit{actuarial risk principle} that preserves the \textit{first stochastic order} (see Wang and Young, 1998, and also Appendix D). Wang (2002) proves in addition that the principle can be applied to assess the price of both financial and insurance claims and hence, that it can be used as a \textit{universal rule} for the pricing of both financial and insurance risks.

Since the distorted probability principle induces a well defined order over the class of probability distributions, it can be used to represent the \textit{preferences} of individuals towards risky prospects. An alternative theory of choice
under risk, known as the dual theory of choice, can be built on these grounds (as formally demonstrated by Yaari, 1987; see also Roell, 1987, and Wang and Young, 1998).3

The notion of aversion to risk is naturally extended in the dual theory of choice. Indeed, recall that in expected utility theory concave and convex utility functions respectively characterise the behaviour of averse-to-risk and risk-lover individuals. Here aversion to risk is defined as aversion towards some mean preserving increase in risk. Yaari (1987) demonstrates that utility and distortion functions are actually related to each other as \( \varphi = u^{-1} \). He thereby concludes that in the dual theory of choice the behaviour of averse-to-risk and risk-lover individuals is respectively characterised by convex and concave distortion functions. Moreover, the behaviour of risk-neutral individuals is represented by the identity function \( \varphi(p) = p \forall p \) (see also Roell, 1987).

---

3The dual theory of choice rationalises many of the paradoxes arising under the expected utility theory (such as the Allais' paradox, Allais, 1953). However, the dual theory of choice has its own paradoxes, which are rationalised under expected utility theory. See Yaari, 1987.
Appendix C

An Axiomatic Characterisation of the Distorted Probability Principle

Risk taking is a subjective experience that in financial and insurance markets people internalise through practice. As a matter of fact, individuals are able to estimate the frequencies of the profit and losses produced by financial securities and insurance claims and accordingly assign prices to them. Every pricing rule should thereby be consistent with some set of intuitive rules about risk handling. On these grounds, a risk principle will be next defined in such a way that a set of basic properties or axioms is satisfied.

Let us consider a risk represented by the random variable $X$ — accounting for the total profit or loss of a certain portfolio at some point of time — defined over some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and characterised by its cumulative or tail probability distribution function, respectively defined as $F_X(x) = P\{X \leq x\}$ and $T_X(x) = P\{X > x\}$. Every risk principle is corresponded to some scalar function $\Pi$ defined over the class of random variables $X$ (as in Goovaerts et al., 1984):

$$\Pi : X \rightarrow \mathbb{R}$$

Following Wang et al., 1997, we will consider a special family of scalar
functions satisfying the following axioms: state independence, monotonicity, comonotonic additivity and continuity.

According to the conditional state independence hypothesis, risk’s prices only depend on the corresponding probability distribution functions, not on the particular state of the world. Monotonicity imposes the very intuitive condition that more risk must always imply a higher price, i.e.:

\[ X(\omega) \leq Y(\omega) \quad \forall \omega \in \Omega \quad \implies \quad \Pi(X) \leq \Pi(Y) \]

Comonotonic additivity imposes additivity for comonotonic risks, i.e.:

\[ \Pi[X + Y] = \Pi[X] + \Pi[Y] \quad \text{if } X, Y \text{ are comonotonic} \]

Comonotonicity defines a dependence structure allowing risks to be expressed through their distribution functions as \( X = F_X^{-1}(U) \), where \( U \) denotes a uniformly distributed random variable. In this way, an extreme form of dependence is obtained, since the realisation of (the external event represented by) the random variable \( U \) completely determines the random vector \((X_1, \ldots, X_n) = (F_{X_1}^{-1}(U), \ldots, F_{X_n}^{-1}(U))\).

Finally, the continuity property claims the price of any layer \((X - k)_+ := \max(X - k, 0)\), approaches the price of the underlying risk when the size of the deductible \(k\) tends to zero:

\[ \lim_{k \to 0} \Pi[(X - k)_+] = \Pi[X] \]

Wang et al. (1997) show that any risk principle satisfying the previous axioms is represented as the mathematical expectation with respect to a transformed or distorted probability distribution (see also Schmeidler, 1986). In fact, according to the Choquet integral representation, for non-negative random variables, the principle is expressed in the following way:

\[ E_{\varphi}[X] = \int_{0}^{\infty} x \cdot dF_{\varphi,X}(x) = \int_{0}^{\infty} T_{\varphi,X}(x) \cdot dx \quad \text{with } T_{\varphi,X}(x) = \varphi(T_X(x)) \quad \forall x \]

where the distortion is represented by a continuous and strictly increasing
function $\varphi : [0, 1] \to \mathbb{R}$, such that $\varphi(0) = 0$ and $\varphi(1) = 1$. On this grounds, the principle is known as the \emph{distorted probability principle}. The traditional expectation operator is obtained when the \emph{identity} is used as distortion.

Wang and Young, 1998, establish the properties:

- $\varphi$ concave $\implies \varphi(y) \geq y \ \forall \ y \in [0, 1] \implies \mathbb{E}_{\varphi}[X] \geq \mathbb{E}[X]$  
- $\varphi$ convex $\implies \varphi(y) \leq y \ \forall \ y \in [0, 1] \implies \mathbb{E}_{\varphi}[X] \leq \mathbb{E}[X]$  

Accordingly, decision makers characterised by \emph{concave} and \emph{convex} distortions respectively overestimates and underestimates the price of risk. Hence, the behaviour of \emph{averse-to-risk} and \emph{risk-lover} investors is respectively characterised by \emph{concave} and \emph{convex} transformations.

The principle is extended in the following way to the case of \emph{real-valued} random variables (see Wang et al., 1997):

$$
\mathbb{E}_{\varphi}[X] = \int_{-\infty}^{0} [T_{\varphi,X}(x) - 1] \cdot dx + \int_{0}^{\infty} T_{\varphi,X}(x) \cdot dx
$$

Applying a change of variable, we obtain:

$$
\mathbb{E}_{\varphi}[X] + \mathbb{E}_{\varphi}[X_-] = \mathbb{E}_{\varphi}[X_+]
$$

The right-hand-side of the equation shows the cost of holding a portfolio containing an insured version of the fund, while the left-hand-side equals the price of the fund plus the value of a guarantee to pay the loss incurred by it — whenever occurs. Both portfolios are related to the same contingent claim, $X_+ = \text{Max} \{X, 0\}$, and both should then be assigned the same market price. Therefore, the condition is consistent with the \emph{non-arbitrage principle}. 

Appendix D

Ordering Risks with Distorted Probabilities

Wang and Young, 1998, introduce an order of risks that is defined on the grounds of the class of distortions $\Phi^*_n$:

\[
X <^*_n Y \iff E_{\varphi}[X] < E_{\varphi}[Y] \quad \forall \varphi \in \Phi^*_n
\]

with

\[
\Phi^*_n = \{ \varphi \mid (-1)^{k+1}\varphi^{(k)} \geq 0 \quad \forall \ k = 1, \ldots, n-1, \phi^{(n-1)} \text{ exists and } (-1)^n\varphi^{(n-1)} \text{ is non increasing} \}
\]

Therefore, some securities will be regarded as riskier than others if they are given higher premiums for a whole class of distortion functions. Equivalently, it can be said that some securities are riskier than others if they are assigned higher prices among a whole class of investors, for distortion functions completely characterise economic decisions in this setting.

The order defined in Equation D.1 is only partial, for it is not possible to compare every pair of securities on its basis. Besides, the order becomes weaker as the exponent $n$ is incremented, i.e.:

\[
X <^*_{n-1} Y \implies X <^*_n Y \quad \forall \ n \geq 1
\]

This conclusion is consistent with the fact that more conditions are imposed.
on the distortion functions of the class $\Phi^*_n$ as the exponent $n$ increases, such that $\Phi^*_n \subset \Phi^*_{n-1}$ $\forall$ $n \geq 1$.

An equivalent characterisation can be given based on the $k$th-dual-moment:

$$H_k[X] = \int_0^\infty \left[ 1 - (1 - T_X(x))^k \right] dx$$

where, as usual, the tail probability function is defined as $T_X(x) = P\{X > x\}$ and where:

$$T_{(n),X}^{-1}(q) = \int_0^q T_{(n-1),X}^{-1}(p) dp \quad \forall \ n \geq 2 \text{ and } 0 \leq q \leq 1$$

$$T_{(1),X}^{-1}(q) = T_X^{-1}(q) \quad \forall \ 0 \leq q \leq 1$$

In a similar way:

$$T_{(n),X}(z) = \int_z^\infty T_{(n-1),X}(x) dx \quad \forall \ n \geq 2$$

$$T_{(1),X} = T_X$$

We then obtain that the order defined in Equation D.1 is equivalently characterised by the set of inequalities:

$$X <^*_n Y \iff H_k[X] < H_k[Y] \ \forall \ k = 1, \ldots, n-1 \text{ and } T_{(n),X}^{-1} \leq T_{(n),Y}^{-1} \quad (D.2)$$

Let us now consider the following alternative order:

$$X <_n Y \iff E[X^k] < E[Y^k] \ \forall \ k = 1, \ldots, n-1 \text{ and } T_{(n),X} \leq T_{(n),Y} \quad (D.3)$$

The stochastic orders of Equations D.2 and D.3 are regarded as dual to each other, since they are defined on mirror spaces (i.e. the set of quantiles and the set of realisations of the random variable respectively) related by the inversion of the tail probability distribution.
Accordingly, it is said that some random variable \( X \) precedes another random variable \( Y \) according to the \( n \)-th stochastic-order if Equation D.3 is satisfied. Similarly, a random variable \( X \) is said to precede another random variable \( Y \) according to the \( n \)-th dual-stochastic-order if Equation D.2 is satisfied (see e.g. Goovaerts et al., 1984).

Wang et al., 1998, prove that the stochastic order can be defined over the class of utility functions \( \Phi_n \):

\[
X <_n Y \Leftrightarrow E[u(X)] < E[u(Y)] \quad \forall u \in \Phi_n
\]

with

\[
\Phi_n = \{ u \mid (-1)^{k+1}u^{(k)} \geq 0 \quad \forall \ k = 1, \ldots, n-1, \}
\]

\[
u^{(n-1)} \text{ exists and } (-1)^{n}u^{(n-1)} \text{ is non increasing}
\]

Consequently, the orders of Equations D.1 and D.4 are regarded as dual to each other.
Appendix E

The Distorted Probability Principle as the Limit of Series of Layers

The practice of insurance involves the exchange of some risky claim — which can be demanded during some fixed period of time — for a fixed payment made at the moment the corresponding policy is established.

Layers are basic insurance contracts giving the holders the right to demand the committed payments when the magnitude of the underlying claim is higher than a certain level \( d \) called deductible or retention. Deductibles can be interpreted as a guarantee provided by policyholders that they will take every reasonable care to avoid suffering the corresponding losses — i.e. they compensate for the moral-hazard implicit in the agreement. The responsibility took by insurers, however, is normally bounded to a maximal coverage or intervention equal to, let us say, \( C - d \).

The stop-loss or excess-of-loss premium of a certain layer, denoted as \( \Lambda_X(d,C) \) in the following, is equal to the residual loss paid by the insurer, i.e.:

\[
\Lambda_X(d,C) = \begin{cases} 
0 & \text{if } X > -d \\
X + d & \text{if } -C < X \leq -d \\
C - d & \text{if } X \leq -C
\end{cases} \quad (E.1)
\]
where \( X \) denotes the magnitude of the loss produced by the underlying risk. Full coverage is obtained when \( C \to \infty \).

The holders of the layer \( \Lambda_X(d,C_1) \) can always expand its coverage by establishing a new reinsuring contract \( \Lambda_X(C_1,C_2) \), with \( C_1 < C_2 \). Hence every insurance contract can be replicated by a series of layers — at least up to a limited extent. It will be demonstrated in the following that such decomposition can be formally established in mathematical terms.

The decomposition ultimately leads to an alternative characterisation of the distorted probability principle defined in Appendix C, which has been already proposed by Dhaene at al., 2006. The presentation proposed in this appendix follows closely the paper of Dhaene at al.

**E.1 The Distorted Probability Principle in the Class of Bounded Risks**

Next the proposition will be demonstrated, that any non-positive risk of bounded support, which contains a finite amount of points summing non-zero mass probability, can be fully insured by accumulating a series of layers.

In fact, let \( X < 0 \) denote such a random variable, in such a way that a series of maximum interventions \( C_0, C_1, \ldots, C_J \in [0,C] \) exists, with \( d = C_0 < C_1 < \ldots < C_J = C \), \( P\{-X = C_j\} > 0 \forall j \) and \( \sum_{j=0}^{J} P\{-X = C_j\} = 1 \), then the tail-probability function of \( X \) can be expressed as:

\[
T_X(x) = P\{-X > x\} = \sum_{j=1}^{J} P\{-X > C_{j-1}\} \cdot I\{C_{j-1} \leq x < C_j\} \quad \forall x > 0
\]

(E.2)

where \( I(C_{j-1} \leq x < C_j) \) denotes the indicator function:

\[
I\{C_{j-1} \leq x < C_j\} = \begin{cases} 
1 & \text{if } C_{j-1} \leq x < C_j \\
0 & \text{otherwise}
\end{cases}
\]

The identity above can be easily obtained by noticing that:
E.1 Bounded Risks

\[ C_{j-1} \leq x < C_j \implies P\{-X > x\} = P\{-X > C_{j-1}\} \]

because the random variable \( X \) accumulates the probability in the points \( C_0, \ldots, C_J \).

Notice that if the whole probability is accumulated in the set \( \{C_0, \ldots, C_J\} \), then the layers \( \Lambda_X(C_{j-1}, C_j), j \in \{1, \ldots, J\} \), with retention \( C_{j-1} \) and maximum coverage \( (C_j - C_{j-1}) \), take the following particular form:

\[
\Lambda_X(C_{j-1}, C_j) = \begin{cases} 
0 & \text{if } 0 \leq -X < C_j \\
C_j - C_{j-1} & \text{if } -X \geq C_j
\end{cases}
\]

which corresponds to a two-point distribution function with probabilities:

\[
P\{\Lambda_X(C_{j-1}, C_j) = (C_j - C_{j-1})\} = P\{-X \geq C_j\} = P\{-X > C_{j-1}\}
\]

\[
P\{\Lambda_X(C_{j-1}, C_j) = 0\} = 1 - P\{-X > C_{j-1}\}
\]

in such a way that for every \( x > 0 \):

\[
P\{\Lambda_X(C_{j-1}, C_j) > x\} = \begin{cases} 
P\{-X > C_{j-1}\} & \text{if } 0 \leq x < C_j - C_{j-1} \\
0 & \text{if } x \geq C_j - C_{j-1}
\end{cases}
\]

As a conclusion, the tail-probability function \( T_j \) of any layer \( \Lambda_X(C_{j-1}, C_j) \) is given by:

\[
T_j(x) = P\{\Lambda_X(C_{j-1}, C_j) > x\} = P\{-X > C_{j-1}\} \cdot I\{C_{j-1} \leq (x + C_{j-1}) < C_j\}
\]

(E.3)

From Equations E.2 and E.3, the tail-probability of every bounded risk of finite-support can be expressed as the sum of the tail-probabilities of the layers based on the points with non-zero probability. Hence the following identity can be written in terms of probabilities:
In this context, bounded risks of finite-support can be always replicated by sums of insuring contracts. From Equation E.4, every risk-premium defined over the class of layers can be extended to the class of bounded risks with finite support.

In order to prove this proposition, let us first notice that the following identity is satisfied in distributions:

\[
\Lambda_X(C_j - 1, C_j) = (C_j - C_j - 1) \cdot B_j
\]

where \(B_j\) denotes a Bernoulli random variable, receiving only the values zero and one, such that \(P\{B_j = 1\} = P\{-X > C_j - 1\} = 1 - P\{B_j = 0\}\), whose expectation is given by \(E\{B_j\} = \text{one} \cdot P\{B_j = 1\} + \text{zero} \cdot P\{B_j = 0\} = P\{-X > C_j - 1\}\).

Accordingly, the expectations of layers can be expressed in terms of the expectations of Bernoulli risks:

\[
E[\Lambda_X(C_j - 1, C_j)] = (C_j - C_j - 1) \cdot E[B_j]
\]

More generally, every risk-principle \(H[X]\) satisfying the scale-invariance property (i.e. satisfying \(H[\gamma \cdot X] = \gamma \cdot H[X], \forall \gamma > 0\)) fulfils the equation above:

\[
H[\Lambda_X(C_j - 1, C_j)] = (C_j - C_j - 1) \cdot H[B_j]
\]

The premium is then uniquely determined up to a distortion function \(\varphi : [0, 1] \to [0, 1]\), in such a way that:

\[
H_{\varphi}[B_j] = \varphi(P\{-X > C_j - 1\})
\]

The only conditions imposed on \(\varphi\) are, in the first place, to allow that every sequence of probabilities is transformed into a new set of probabilities, so
that $\varphi(0) = 0$ and $\varphi(1) = 1$, and secondly, to ensure that tail-probabilities are not diminished, so that distortions must be non-decreasing functions and hence $p \leq q \implies \varphi(p) \leq \varphi(q)$. These conditions are sufficient to ensure that the first stochastic order is preserved (see Appendix D).

Therefore,

$$H_{\varphi}[\Lambda_X(C_{j-1}, C_j)] = (C_j - C_{j-1}) \cdot \varphi(P\{X > C_{j-1}\})$$

$$= E_{\varphi}[\Lambda_X(C_{j-1}, C_j)] \quad (E.5)$$

On these grounds, the risk-principle $H_{\varphi}$ can be interpreted as the expectation with respect to some distorted probability distribution. It can thereby be related to the distorted probability principle defined in Appendix C.

### E.2 Extension to Unbounded Risks

The distorted probability expectation belongs to a class of non-additive operators, which preserves addition only in the case of comonotonic risks (see Dhaene et al., 2006, and also Schmeidler, 1986).

A certain series of random variables $X_1, \ldots, X_J$ is said to be comonotonic if and only if, a series of non-decreasing functions $\alpha_1, \ldots, \alpha_J$ exists, such that for any risk $X$ the following equality holds in distributions: $(X_1, \ldots, X_J) = (\alpha_1(X), \ldots, \alpha_J(X))$. Hence, comonotonicity characterises a case of extreme dependence, when no diversification can be attained by pooling exposures together, for all of them move in the same direction (see Dhaene et al., 2002).

It is easy to verify that the layers $\Lambda_X(C_0, C_1), \ldots, \Lambda_X(C_{J-1}, C_J)$ are pairwise comonotonic random variables, in such a way that from Equations E.4 and E.5 the following relationship holds for any bounded risk $X < 0$:

$$E_{\varphi}[X] = \sum_{i=1}^{J} E_{\varphi}[\Lambda_X(C_{j-1}, C_j)]$$

$$= \sum_{j=1}^{J} (C_j - C_{j-1}) \cdot \varphi(P\{-X > C_{j-1}\}) \quad (E.6)$$

An actuarial intuition can be given for this result, for in order to avoid that
customers choose to split their covering (distributing it to various insurance companies while assuming part of it by themselves) insurers indeed prefer to assign risk-premiums satisfying:

\[ E_{\varphi}[X] \leq \sum_{j=1}^{J} E_{\varphi}[A_X(C_{j-1}, C_j)] \]

However, while offering prices satisfying:

\[ \sum_{j=1}^{J} E_{\varphi}[A_X(C_{j-1}, C_j)] > E_{\varphi}[X] \]

may be attractive for insurers in the case that diversification is possible to some extent, since under such conditions a gain can be attained by holding the whole portfolio, such a benefit disappears when risks are comonotonic. Accordingly, additivity for comonotonic risks is a necessary requirement for insurance prices.

In order to extend the distorted risk-principle to the case of a non-positive, bounded and continuous or at least piecewise continuous random variable \( X \in [-C, 0] \), whose probability is not necessarily accumulated in a finite set of points, let us notice that the tail-probability of such a random variable can be approximated by the following piecewise constant tail-function:

\[
T_{J}(x) = \sum_{j=1}^{2^J} T_X\left(\frac{j}{2^J} \cdot C\right) \cdot I \left\{ \frac{j-1}{2^J} \cdot C \leq x < \frac{j}{2^J} \cdot C \right\} \quad \forall \ x \geq 0
\]

Convergence to the integral is guaranteed when the following property concerning the limiting behaviour of the principle is satisfied (Dhaene et al., 2006):

\[
\lim_{C \to \infty} E_{\varphi}[\min(-X, C)] = E_{\varphi}[X] \quad \text{(E.7)}
\]

It is then possible to define the distorted probability principle for any positive random variable, such that if \( T_{\varphi,X}(x) = \varphi(T_X(x)) \) is defined:
In this way, a risk-premium is formally characterised in terms of the premiums offered by a family of layers (or Bernoulli risks).
Bibliography


Doctoral Dissertations from the Faculty of Business and Economics

From August 1, 1971


44. Niet toegekend.


46. DEGADT, Jan. Specificatie van een econometrisch model voor vervuilingenproblemen met proeven van toepassing op de waterverontreiniging in Belgie. Leuven, s.n., 1982. 2 V.


49. VAN HULLE, Cynthia (08/83). Shareholders’ unanimity and optimal corporate decision making in imperfect capital markets. s.l., s.n., 1983. 147 pp. + appendix.


61. LUYTEN, Robert (04/87). A systems-based approach for multi-echelon production/inventory systems. s.l., s.n., 1987. 3V.

63. VAN CAYSEELE, Patrick (20/05/87). Regulation and international innovative activities in the pharmaceutical industry. s.l., s.n., 1987. XI, 169 pp.


158. SKUDELNY, Frauke (21/05/2002). Essays on The Economic Consequences of the European Monetary Union. Leuven, K.U. Leuven,


