Essays on Banking and Portfolio Choice

Bo Larsson

DISSERTATIONS IN ECONOMICS 2005:4
To Malin
Contents

Acknowledgements vii

Introduction 3

I Banking and Optimal Reserves in an Equilibrium Model 11

II Optimal Rebalancing of Portfolio Weights under Time-varying Return Volatility 39

III Can Parameter Uncertainty Help Solve the Home Bias Puzzle? 73
Acknowledgments

What a long strange trip this has been, as The Grateful Dead ones wrote. I don’t think anyone that knew me in my early years could have predicted that I would write a thesis one day. Being 15 and having to choose which path of education to take, the choice was easy: only nerds and geeks took science to prepare for University. Better to take the combined Business and Language program where all the girls were and get a job after High school. But when I spent a year in Cazenovia NY being surrounded by friends and family that were going (or had been) to University I think that I started to reconsider. Therefore I want to thank them who sowed the seed of higher education in my stubborn mind, cheers family and friends in Caz. I also want to thank my boss when working at Handelsbanken, Pernilla Bergggren, who is the person that encouraged me to apply to the Ph.D. program.

I think that three ingredients are essential in order to success in the Ph.D. program: character, friends and theoretical skills. My commanding officers during the recruit and UN armed forces respectively, Per Dahlbom and Henrik Lettius, have helped me build my character by being exceptionally good role models (in case someone think I’m more of a character than having any, Per and Henrik aren’t to be blamed), which have help me through rough times.

When it comes to friends there are “drängarna” in Strängnäs whom I haven’t had time to see so much the last couple of years. My High school friends “E3-sacks” whom I don’t see very often but when we do, it rocks! Of course I mostly hang with other Ph.D. students and colleagues. The year I started on the program we were twelve new candidates that have had great fun over the years. In fact it seemed that the theories of Engle could be applied also on calendars, after immensely long periods without weekends, all of a sudden there could be arbitrarily many Saturdays in a row, often associated with the taking of exams. Being such good friends didn’t only mean a lot of party but also good support when times were hard, especially during the first year. Every year I also get to see fantastic views and have great adventures with Mats and Jesper when we go hiking, though Micke call it suffering, up north. When I’m particular fortunate my advisor in sports Lennart takes me to see Speedway featuring Tony Rickardsson and company, the gladiators of today. To see them race around and hit the boards in 80 mph really makes you realize that not being able to solve a model is not that big of a deal. Since Mackan said that one should drive MC during Ph.D. studies, I want to thank Putte and Nyqvist & Monell Trafikskola, Åkerman family, Anders, Kent and Dad for helping me get my MC drivers license.

My family also has a big part in this thesis; they are probably responsible for my interest for social sciences and economics. As long as I can remember there have always been discussions about politics and economics at home, especially my grandfather Vallis has been an inspiration through his strong engagement for current
affairs. My family has also been good practical support, mom do my taxes, dad has hooked me up with people in Banking to discuss my models with, my brother fix technical stuff and my sister have filled my spare time with the opportunity to ride magnificent horses.

I think I was well prepared for graduate studies by Björn Hansson who was my undergraduate thesis advisor and also gave a really challenging course. He has also helped me with data to this thesis. During the graduate program I have really been inspired by Timo Teräsvirta, both through courses and discussions of problems. Taking courses by John Campbell and Sune Karlsson has also been rewarding, both essay two and three were started as term papers in their respective courses. I also want to thank Dan both as friend and co-author. The final touch on the language in the thesis has been supplied by Christina, which resulted in many improvements.

When it comes to actually finishing this thesis I cannot thank my advisor Hans Wijkander enough. He has both been like a coach in creative writing and a demanding listener to assure that my thinking is straight and also improving my modeling. Through his suggestions I have really improved my skills in writing scientific papers.

Last I want to thank Malin for putting up with me through my ups, and downs. Listening to endless explanations of integrals and probability functions for which you probably couldn’t care less. For proof reading my articles, and maybe the most important thing living with me!

For those of you that I have kicked in the shins, tackled and/or managed to shoot in the face during lunch football and ball-hockey, remember that I have the perfect excuse being Rose Mary’s Baby…

Stockholm, August 2005

Bo Larsson

Funding from the department of Economics at Stockholm University, Sparbankernas forskningsstiftelse, Siamon and Wallenberg’s Jubileumsfond is gratefully acknowledged.
Introduction
Introduction

Throughout time, economists have often encountered empirical problems and/or behavior that cannot easily be explained with existing models and theory. As a social scientist, it is both challenging and inspiring to address problems where standard models have previously failed. Although this thesis might seem quite divided as the three essays are in rather different fields, their connection is, in fact, that they all discuss problems where little is known about the cause of economic agents actions. In the last paper, coauthored with Dan Nyberg, we discuss one of the more famous economic puzzles, termed “the home-bias puzzle”. The first two papers discuss questions of great importance within banking and finance that have received relatively little attention from researchers.

In chapter one, I address the question of why banks in reality are so “overcapitalized” in relation to regulatory requirements. Since banks profit from more leverage, they ought to seek as low reserves as possible. To strengthen international capital markets, banking is now regulated to ensure that sound banking is practiced, which puts a bound on how low the reserves can be. Most developed countries have adopted the reserve requirements of the Bank of International Settlements (BIS), which regulates tier one capital, i.e. short-term assets and cash, and tier two capital, i.e. banks’ long-term debt, BIS (2003). The reserve requirements stipulate that the ratio of tier one and tier two capital to a bank’s risk-adjusted loan portfolio should be at least eight percent, and that the considered tier two capital cannot exceed the tier one capital. This effectively means that the tier one ratio should be at least four percent. However, banks typically have a much larger tier one capital; in Sweden the market leading banks have as high tier one ratios as 7-8 percent. Jackson et. al. (1999) show that internationally, for G-10 banks, the average reserve ratio is 11.2 %.

Reserves in a bank cannot be directly compared to cash holdings in a regular company, but some similarities exist. For both banks and regular firms, cash holdings raise the endurance, i.e. the bank/firm can withstand larger losses. Accordingly, one explanation for carrying large reserves could be risk reduction over time. Ingvar Kamprad, the successful founder of IKEA, stated that having large amounts of capital is the best way of ensure long-term growth, even though analysts frequently complain about inefficient use of capital. IKEA is thus independent of capital markets when new projects are slow starters, thereby allowing it to pursue
long-term investments despite possibly very poor short-term outcomes.\textsuperscript{1} However, in this article, I study another possibility. How does systematic risk affect optimal banking strategies? In many earlier capital market articles, it is often assumed that systematic risk is absent, e.g. Diamond (1984) or Williamson (1986). But the financial literature has supplied massive support for the presence of systematic risk, and path breaking models such as both CAPM by Sharpe (1964) and APT by Ross (1976) build on the separation between idiosyncratic and systematic risk.

By introducing systematic risk in the banking model by Williamson (1986), I can show that there could be insurance motives for carrying reserves, independent of the banking regulations. In fact, banks have always carried reserves independent of regulations. With systematic risk, banks have a positive probability of default, which results in expected auditing costs for depositors. By holding cash reserves, banks can reduce this expected auditing cost. Auditing costs are unproductive and therefore, the expected return on lending, to be split between banks and depositors, can be increased with the use of reserves.

Swedish regional banks in general carry more than twice the reserves of market leading national banks. If conjecturing that small regional banks are less diversified since they cannot pool regional risk, this might be the reason why they carry more reserves than large banks. I show that more systematic risk in my model indeed optimally leads to larger reserve ratios.

In chapter two, optimal long-term investment strategies given different return generating processes are analyzed. For a long time, the common investment advice has been to suggest a larger fraction of stocks in a portfolio if the investor has a long investment horizon. This advice has been given by both professionals as well as the popular press. A striking example of this is also the marketing of generation tailored mutual funds in the new Swedish pension system, where customers are supposed to pick their investments based on birth year. The reason for this investment behavior to be preferred is that stocks are less “risky” in the long run, where risk is often measured as standard deviation; see, for example, Siegel (1994). Campbell and Viceira (2002) point out that if return is independent and identically distributed, i.i.d., risk measured as standard deviation is inversely related to the square root of the investment horizon. If there is to be any reduction in risk with longer investment horizons, there must be some sort of predictability in returns, i.e. they cannot be i.i.d.

I restrict the analysis to wealth alone, i.e. investors do not have any labor income. Campbell, Cocco, Gomes and Maenhout (2001) show that if investors have labor income, this could lead to a lower demand for stocks as investors approach their retirement. However, this effect is unrelated to the relative riskiness of stocks, but is

\textsuperscript{1}Discussion in the documentary “Mannen som ville möblera om världen” on Swedish national television, SVT2 July 10, 2005.
a substitution effect between the risk-free asset and labor income. The marketing of “generation funds” in Sweden emphasized the old mantra of lower “risk” in stocks for long horizons.

Several authors have investigated portfolio choice decisions for long-term investors. But it seems that no one has considered the case where returns are distributed according to the popular GARCH-models first suggested by Nobel Laureate Robert Engle (1982). One reason for the popularity of GARCH models is that they capture the fact that return volatility clusters over time; see, for example, Mandelbrot (1963). Since these models are frequently used in practice, I investigate in what optimal portfolio rules for multi-period investors they result.

My approach is to first analyze the implications for long-term portfolio choice when only risk, measured as conditional variance, is time-varying. Earlier analyses of similar return structures have shown that for investors with infinite horizons, shocks to volatility affect portfolio choice such that there are some horizon effects; see Chacko and Viceira (2000). In contrast to that result, I find there to be no effects on portfolio weights when the investment horizon is altered. Naturally, shocks to volatility affect the optimal weights of the risky asset, but these have an equal effect on all investors, given the same level of risk aversion.

Merton (1973) showed that expected returns should be related to the level of risk through the coefficient of risk aversion. The intuition is that if you are risk averse, more risk should be compensated by higher expected returns. Moreover, the more risk averse you are, all else equal, you will require more compensation in terms of expected return to hold risky assets. Therefore, I also perform analyses where expected returns on risky assets depend on conditional volatility, which is my measure of risk. When returns are predictable, the optimal portfolio weights do change when the investment horizon is altered. These effects are small, however, and only present for about thirty periods, i.e. 2-3 years with monthly data. When there is a positive relationship between risk and expected return, investing in stocks is a hedge against low returns. This is because a lower than average return today raises the expected return tomorrow.

In the final chapter, I and my coauthor Dan Nyberg address international portfolio choice. A well-known puzzle in international finance is the equity home bias: in contrast to the prescriptions of standard portfolio theory, international diversification is not used as a means for decreasing the risk of a portfolio, e.g. French and Poterba (1991) and Lewis (1999). In this essay, we illustrate a mechanism where the exchange rate estimation risk causes an equity home bias.

Estimation risk is introduced into a standard mean-variance portfolio framework using return time-series of different lengths. We take the perspective of a domestic investor considering a foreign stock-market index investment. The investor cares about the local-currency return of the foreign investment as the investor’s consump-
The investment basket is denominated in local currency and observes historical data on stock index returns and the return on the exchange rate. The returns are assumed to be generated from a multivariate normal distribution and the investor uses a stochastic model to forecast future returns on the stock indices and the exchange rates.

The return history of the exchange rate is argued to be shorter than the available time series of equity index returns due to e.g. exchange rate shifts. A change in the exchange rate regime implies that the past time series on local currency returns of a foreign investment can no longer be expected to be informative about the future, and the sample needs to be truncated. To deal with this issue, we use a framework devised by Stambaugh (1997) to analyze investments whose histories differ in length. Stambaugh derives maximum likelihood and Bayesian predictive distribution mean-variance estimators of the combined sample. If the investors form expectations of future returns based on the ex post sample moments, then the present estimation risk is ignored. On the other hand, if investors take the estimation risk into account when calculating the next period’s expected return and covariances, Bayesian predictive mean-variance estimators should be used.

The impact of estimation risk on an optimal portfolio is illustrated with data from Sweden and the United States. Our results show the introduced estimation risk to mainly be associated with the exchange rate, and that explicitly accounting for the estimation risk causes the domestic investor to increase the fraction of domestic assets. While the introduction of exchange rate estimation risk is not powerful enough to explain the entire home bias observed in data, the results of this paper illustrate a potentially important mechanism that is often overlooked in discussions of international portfolio diversification.

My articles do not fully reveal the underlying mechanisms of economic agents behavior in the problems I study. However, I feel that the articles do shed some new light on these problems of great practical importance.
References


Essay I
Banking and Optimal Reserves in an Equilibrium Model*

1 Introduction

The direct effect of the reserve requirements on the banking system and their indirect effects on the real economy, are questions frequently investigated. The Bank of International Settlements’ regulations, which are adopted by most developed countries, state that the “cash” reserve ratio, i.e. cash and government backed assets to risk weighted capital, should be 4%. However, banks have much larger reserves. In Sweden, this tier 1 ratio is often as high as 7-8 % among the market leading banks. Smaller banks typically have even larger reserve ratios. Jackson et. al. (1999) find that G10 banks have reserve ratios of 11.2% on average, including secondary capital. This means that the reserve requirements by BIS rarely bind, thereby implying that banks hold reserves for other reasons than the BIS capital requirements.

In this paper, I investigate why banks should have reserves, and how large these should be. This is an important question due to the evidence that regulatory reserve requirements rarely bind. My set-up is to assume that banks cannot diversify the entire market risk, notwithstanding how large they grow. In fact, data shows that on average, banks’ credit losses are small, but occasionally there are sharp increases in these low levels, e.g. during the early nineties. My innovation of bringing in macroeconomic shocks that can severely hurt banks’ profitability and thereby their survival, allows me to analyze banking within a more realistic framework. Using this set-up, it is shown that banks can exist in equilibrium, they carry substantial reserves to reduce the probability of default and the destruction of value that occurs in defaults. This also explains why small regional banks have larger reserves than large banks: they are not able to diversify regional risk, which leaves them with more systematic risk in their debt portfolio. Moreover, banks are profitable in the sense that rich consumers have a higher return on reserves when starting a bank than when investing in a private portfolio of loans to entrepreneurs. For this result

*I am grateful for comments and suggestion by Hans Wijkander, Martin Holmén, and Ola Hammarlid. Financial support from Sparbankernas Forskningsstiftelse is gratefully acknowledged.
to come through, it is essential that banks have sufficiently large bargaining power both against borrowers and depositors.

Since I am interested in debt markets and the concept of banking in particular, it is essential to have a model that does not rely on pure risk sharing with simple contingent claims of the Arrow and Debreu (1954) type. Introducing asymmetric information is one way of creating an environment that needs more complicated contracts than contingent claims. Classical ways of modeling asymmetric information in a market are through moral hazard or adverse selection; see e.g. Mas-Colell, Whinston and Green (1995). Both these approaches primarily deal with ex ante informational problems. Townsend (1979) explored another type of asymmetric information, costly state verification, hereafter CSV, which models ex post informational problems. With CSV, the realization of a stochastic project is only known to the insider. Outsiders can, at a cost, monitor projects to reveal the true outcome. Several authors have shown that with CSV and ex post non-random audit, the efficient contract between an entrepreneur and an investor is a standard debt contract, e.g. Townsend (1979), Diamond (1984), Gale and Hellwig (1985), and Williamson (1986, -87a).

I follow Williamson (1986, 87a), where he derives a model with credit rationing and endogenous debt contracts in equilibrium; the 1986 paper also derives intermediation. With CSV, Williamson (1986, -87a) shows there to be one fundamental question for the lender: to audit and bear the associated cost or not to audit. Since there are two different “states” for auditing, there must be a threshold for possible signaled results for borrowers, bad results leading to audit or good results leading to no audit. Naturally, borrowers will never signal a result above the threshold since they would like to keep as much of the result as possible to themselves. The resulting efficient contract is then a standard debt contract, either the entrepreneur who received financing re-pays the loan with agreed on interest, or he is audited and the lender seizes his entire wealth. Further, private information is shown to give rise to intermediation if entrepreneurs’ projects require more capital than a single investor can supply. Every entrepreneur now needs to borrow from several investors, which leads to overlapping audits. Since each investor has an incentive to lie to the other investors in the same project, all investors need to audit themselves. To circumvent the problem with overlapping auditing, Williamson (1986) introduces delegated auditing through banks, which eliminates overlapping audit. To solve the monitoring problem of banks, the values of entrepreneurs’ projects are assumed to be independently distributed. All risk will then be pooled and the monitoring of banks is superfluous: therefore, repayments on bank deposits are larger than the expected return on lending directly to entrepreneurs, i.e. net of monitoring costs.

My model differs from that of Williamson (1986, -87a), in that I allow for heterogeneous consumers and correlation between entrepreneurs’ projects. A very small
share of my consumers has substantially larger funds than the normal consumer. This means that these rich consumers can be as efficient in their auditing as banks, since they can completely fund projects. Hence, small and large investors will have different expected returns when lending. Further, I only consider rationed markets, i.e. there are more entrepreneurs wanting to borrow than consumers wanting to lend. Rationing together with the different efficiency in auditing costs, for my two types of agents, can lead to an equilibrium with two lending rates as in Williamson (1987b).

The most important innovation by far in my model is that I allow for systematic shocks. Introducing systematic risk might seem trivial. However, if direct lending is not prohibited, it is important that the probability distribution for the value of the debt portfolio is derived from the probability distribution for entrepreneurs’ values, rather than assumed. That is, the probability distribution for the value of the loan portfolio must be consistent with the probability distribution of entrepreneurs’ projects. To compare bank lending with direct lending, these distributions of outcomes must also be computable. Vaisceck (1987, -91, 2002) assumes entrepreneurs’ projects to follow correlated Wiener processes. Thus, he derives a computable probability distribution for debt portfolio losses. His model also works well empirically and it is common within management of debt portfolios, e.g. JPMorgan Credit Metrics, CreditSuisse+, and Moody’s KMV products. Using this probability distribution for the values of banks’ loan portfolios, the intermediary is not just a delegated monitor as in Williamson (1986). Instead, banks supply the market with a new product, deposits with low verification costs.

In the spirit of Holmström and Tirole (1997), I incorporate the importance of banks’ balance sheets into my model, and I only consider rationed markets, which have been thoroughly investigated by Stiglitz and Weiss (1981). I also view this work as a first step towards improving business cycle models with credit markets, such as Bernanke and Gertler (1989) or Carlstrom and Fuerst (1997), by including correlated investments.

Unlike the previously mentioned authors, and the vast majority of theoretical credit market work within economics, there is a strong link to the financial and statistical literature. I base my credit model on micro foundations from the financial literature. This way, I obtain a more realistic model environment, which then strengthens the relevance of my findings. There is a huge amount of research, both theoretical and empirical, performed within finance where the common model feature is the notion of both systematic and idiosyncratic risk, and the pricing of the two, e.g. Ross’s Arbitrage Pricing (1976) or the famous CAPM by Sharpe (1964).

First, I thoroughly present the model in section 2, describing the optimal contract and the different ways for agents of interacting with each other. The solution to the problem derived by Williamson (1987a), consumers directly lending to entrepreneurs
with my one factor distribution parametrized, is shown in the first subsection. The next subsection is devoted to the derivation of the entrepreneurial value and the resulting debt portfolio distribution, as this is a novelty in credit contracts research. An empirical example shows that the derived debt portfolio distribution fits real data quite well. In section 3, I start by setting up the bank’s problem of choosing what deposit rates to offer and how large the reserves should be. Then, I solve the contract between the bank and the entrepreneurs to determine the distribution for banks’ values. When the lending side of banking has been solved, I use the result to solve for optimal deposit rates and reserve levels. Finally my concluding remarks are in section 4.

2 Model

Consider a two-period economy where decisions are made in the first period and consumption takes place in the second. There are two basic types of agents, entrepreneurs and consumers, both of which are risk neutral. Each entrepreneur has an investment project that needs $N$ units of capital to be started. The majority of consumers have an endowment of 1 unit of capital, but a limited fraction of consumers are very rich. An entrepreneur’s investment project cannot be split or sold and needs capital to be initiated. If a project is initiated in period one, it results in a stochastic outcome $\tilde{V} \sim LN(\theta,\sigma^2)$, i.e. the final value of entrepreneurs’ projects is log-normally distributed. I denote the probability distribution function by $\psi(\cdot)$ and the cumulative distribution function by $\Psi(\cdot)$. The realizations of entrepreneurs’ projects, $V_i$, are correlated with the linear correlation coefficient $\rho$, i.e. pooling of projects cannot remove all stochastic variation.

The market for capital arises since consumption takes place in the second period. Consumers need to “store” their capital, which can only be converted to consumption in period two. In Figure 1, I present a scheme over my market and the possible contract solutions. There can be two possible market solutions, direct lending symbolized with dashed lines and a solution with banks symbolized by solid lines. With a banking solution, either reserves are needed and banks are started by the rich who use their capital as reserves, or no reserves are needed and any consumer can start banks. If no reserves are needed, banking must be associated with zero profits, or there will be arbitrage opportunities. Returns on banks portfolios fluctuate, but the deposit rate is fixed; hence banks always need to be audited or there will be arbitrage.

In Williamson (1986), a small consumer can act as an intermediary due to \textit{i.i.d.} project returns, which allows her to pool all risk by growing large and offer depositors a fix risk-free return. This solution is removed when systematic risk is introduced. For exactly the same reason that prevents delegated monitoring in
Entrepreneurs (E)

Rich (B)

BANK

Consumers (C)

Reserves

Deposit (1)

Loan

Profit

Loan

Loan

R, R, R

R, R

R, R

Williamson (1986), a group of small investors cannot form a bank. They would have to rely on delegated monitoring and each of them would have an incentive to lie to the others and engage in side bets with entrepreneurs to reap private profits. The very rich, on the other hand, could reduce the risk in deposits by holding reserves, thereby removing some of the downside in banking. Hence, the solutions are as described in Figure 1

An entrepreneur that contracts with rich consumers or banks borrows all his funds from the same source and the agreed repayment including interest is $\overline{R}$, resulting in an expected repayment of $r$ to the lender. When entrepreneurs borrow from normal consumers with limited funds, they need to contract with $N$ consumers with an agreed repayment including interest of $\overline{R}_D$ yielding an expected repayment of $r_D$ to the lender. In case consumers with limited funds decide to use the bank as an intermediary, they lend their money to the bank; that is, the bank offers the gross deposit rate $D$, which results in an expected repayment from the bank of $d$. The expected gross returns $r$, $r_D$ and $d$ are lower than the agreed repayment, if there
is a positive probability of default. If, for example, the value of project \( j \) is \( V_j < \bar{R}_D \) when a consumer is lending directly, the lender must audit and pay the cost \( c \), implying that the expected return \( r_D \) is lower than \( \bar{R}_D \). Note that higher repayment levels lead to a higher probability of default and therefore, higher expected auditing cost.

The entrepreneurs’ problem is to offer lenders a repayment level, \( \bar{R}_D \), that yields the largest possible expected value net of the repayment, equation (1)

\[
\max_{\bar{R}_D} \int_{\bar{R}_D}^{\infty} (V - \bar{R}_D) \psi(V) \, dV.
\]

But the entrepreneurs’ offer must also satisfy lenders. Lenders’ expected utility from lending is the total expected repayment net of auditing cost, i.e. when borrowers default, lenders seize all their assets and bear the auditing cost \( c \), equation (2)

\[
\int_{0}^{\bar{R}_D} (V - c) \psi(V) \, dV + \bar{R}_D \left( 1 - \Psi(\bar{R}_D) \right) \geq r_D.
\]

Williamson (1986, -87a) shows that the constraint must hold with equality, or the value of the objective function can be raised by reducing \( \bar{R}_D \) without violating the constraint. If the lender is also assumed to have all the bargaining power, the solution is found by maximizing lenders’ expected utility (2).

Due to my use of a realistic probability distribution for project outcomes, I rely on numerical solutions to a large extent. This means that the problem must be parameterized. Some insights as to the admissible range of parameters are gained from the second-order condition of equation (2)

\[
-\psi(\bar{R}_D) - c\psi'(\bar{R}_D) < 0.
\]

By re-writing the second-order condition (3) using my distribution function for values of entrepreneurs’ projects \( \psi(\cdot) \), I obtain expression (4) describing the relationship between the parameters and the optimal choice of repayment level \( \bar{R}_D \)

\[
c \frac{\ln(\bar{R}_D - \theta)}{\bar{R}_D - \theta - c} < \sigma^2.
\]

It is immediately seen that in case the return on lending, i.e. \( \bar{R}_D - 1 \), is lower than \( \theta \), which is the mean return parameter for the distribution of entrepreneurs’ values, the condition for a maximum will be fulfilled for a wide range of plausible mean, variance and auditing cost parameters. Since the numerator is then negative and the denominator should be positive for reasonable parameters, the left-hand side
will be smaller than any chosen variance. This is also the case in reality, where the return on lending tends to be lower than the expected return on equity. If instead the return on lending is larger than the mean parameter $\theta$, the left-hand side of (4) will be positive and it is harder to find parameters that will fulfill the condition, since the variance of return is in general a very small number.

2.1 Parameterized solution without intermediaries

I use Swedish data to obtain reasonable numbers for the mean and standard deviation. Using a small sample from 1988-99 over twenty of the largest listed firms in Sweden, the estimated mean and standard deviation is .12 and .36, respectively. The average correlation is found to be .38. The mean and standard deviations are rounded to .1 and .3 to keep the calculation simple. Correlation is rounded down to .3, both to keep calculations simple and to account for non-traded firms in a bank’s portfolio. The solutions’ sensitivity to the choice of correlation will be analyzed later.

The expected utility for entrepreneurs and consumers is graphed in the left-hand panel of Figure 2. The auditing cost is set to .35, and the mean parameter and the variance are as estimated on the Swedish sample. With the estimated parameters for

![Graph of the expected utility for lender and entrepreneur with direct lending](image)

Figure 2: Graph of the expected utility of the lender and the entrepreneur, respectively, given direct lending. The left-hand panel shows the expected utility with parameters estimated on Swedish data, .1 and .3 for mean and standard deviation. In the right-hand panel, we replace the estimated risk and return parameters with a conjecture of these values. The mean parameter and the standard deviation on entrepreneurial projects are set to .2 and .1, respectively. The optimal repayment is 1.108, resulting in an expected return to the lender of 1.0357 and of 1.0346 to the entrepreneur.
investment projects, the expected return on a contract is negative, and the expected payback from lending one unit of capital is just over .8, which does not seem to be very attractive. We also see the effect of having an interest rate on lending that is higher than the mean parameter for entrepreneurs’ projects: the objective function (2) is almost flat around the optimal repayment level, which was shown in equation (4). Raising or reducing the auditing cost will not be enough to obtain a sufficient contract with a positive return. Moreover, if auditing costs are reduced, the interest rate becomes very large as the asymmetric information disappears and we approach pure risk-sharing.

Does this mean that no debt financing should exist? The answer is that in this two-period economy there is neither reputation nor collateral/cosignatories for which I must compensate. A simple way of adjusting for the lack of these empirical features is to alter the parameters of my probability distribution. The use of collateral and cosignatories remove much of the downside of a project: this effectively raises the mean and reduces the variance. Adjusting the mean parameter to .2 and reducing the standard deviation to .1 results in the right-hand graph in Figure 2. The result is obvious; there is a much more pronounced lump in the lenders’ expected utility function (2). The expected return on the efficient contract in the right-hand panel of Figure 2 is positive (.0357) at a repayment of (1.108), i.e. the expected auditing cost is (.0723).

To justify the high auditing costs chosen, the second-hand market for bankrupt firms’ assets can simply be considered. The market value of firms’ assets seldom amounts to the value of the firm just before bankruptcy; which is naturally why bankrupt firms are often reconstructed rather than defaulted. For example, debtors sometimes agree to convert debt into equity on the judgement that in time, more value can be extracted from the equity than from an immediate bankruptcy. My theoretical probability distribution for the value of entrepreneurs’ projects does not take this into account. Instead, the probability distribution is smooth over the default point; that is, I do not model the empirical feature that the value of firms’ assets in a bankruptcy are worth less than just before the default. To compensate for this, we can instead have a large auditing cost, which will create a discrete jump of a firm’s value at the defaulting point.

Williamson (1986) simply let banks’ auditing costs be $c/N$, as there is no overlapping audit. Clark (1988) studied banks in the U.S, finding evidence of decreasing returns to scale in banking. Cerasi and Daltung (2000) model this feature as depending on auditing costs that increase with scale. Hence, it could be argued that some non-linear auditing structure should be used. For simplicity, I adopt the Williamson (1986) method and let banks’ auditing costs be $c/N$, since a non-linear structure would only complicate the calculations. To be able to analyze markets with a positive probability for bank defaults, I need to decide on the auditing scheme for banks;
how much more expensive should an audit of a very large bank be relative to the audit cost, \( c \), of small entrepreneurs. A simple solution is to assume an auditing scheme with a fixed cost, equal to twice the overlapping audit cost for entrepreneurs, i.e. 2\( c \). The reasoning behind this assumption is as follows: essentially, a bank is just a large portfolio of loans, therefore it could be audited in two stages if it defaults. First, an “audit” is made where the loans in the bank are randomly distributed across the depositors. Second, the depositors audit the loan they just received in the first-stage division. As a result of the two-stage audit, the cost is doubled, i.e. 2\( c \).

For technical purposes, I will assume infinitely large credit markets, i.e. there is an infinite number of entrepreneurs as well as consumers. This simplifies the calculation of actual losses, which is shown in the below subsection, where the distribution for debt portfolios is derived.

### 2.2 Production distribution

Entrepreneurial outcome is driven by both a common factor, as well as an individual-specific factor. Systematic risk from the common factor will then remain, even if banks grow infinitely large. Since I analyze debt, the portfolio distribution will not be some simple transformation of the underlying distribution with smaller variance, as for equity portfolios. The reason is that extremely poor project outcomes will not be pooled by extremely good project outcomes for debt financing.

Denote the value on the \( i \):th entrepreneurial project \( V_i \), and assume it to follow a logarithmic Wiener process resulting in the following stochastic differential equation (5). The distribution model is based on a continuous time Wiener process, though my model is a discrete two-period model. It only uses time periods \( T \) equal to 0 and 1

\[
dV_i = \delta_i V_i dt + \sigma_i V_i dW_i, \tag{5}
\]

where

\[
dt \cdot dW_i = 0, \quad (dt)^2 = 0, \quad (dW_i)^2 = dt, \quad dW_i \cdot dW_j = \rho dt \ \forall i \neq j. \tag{6}
\]

\( \rho \) is the ordinary linear coefficient of correlation between different firms’ asset values. For a more intuitive interpretation, the Wiener process \( W_i \) can be represented as,

\[
W_i = \sqrt{\rho} Y + \sqrt{1 - \rho} \varepsilon_i, \tag{7}
\]

where \( Y \) is a common systematic factor, a macroeconomic shock, and \( \varepsilon_i \) an idiosyncratic noise component.

Using a standard textbook on stochastic mathematics, e.g. Björk, (1998), the solution for returns on wealth \( v_i \), for a single firm is shown in equation (8), where
the standard notation for continuous return is used, \( v_i = \ln V_i (1) - \ln V_i (0) \)

\[
v_i = \delta_i - \frac{1}{2} \sigma_i^2 + \sigma_i \sqrt{\rho} Y + \sigma_i \sqrt{1 - \rho} \varepsilon_i. \tag{8}
\]

Here, \( Y \) and \( \varepsilon_i \) are \( i.i.d. \) standard normal, with cumulative distribution \( \Phi (0, 1) \) and probability distribution \( \phi (0, 1) \). The dependence between firms is now captured by having the common process \( Y \) instead of correlated processes \( W_i \). Hence, it is simple to find the default probability for a single firm, it is simply the probability that the firm’s return \( v_i \) is lower than the predetermined interest rate \( \ln (\overline{R}) \), equation (9)

\[
P \left( v_i < \ln \overline{R} \right) = \Phi \left( \frac{\ln \overline{R} - \delta_i + \frac{1}{2} \sigma_i^2}{\sigma_i} \right). \tag{9}
\]

Note that from here on, I have defined the individual default probability in (9) as \( \rho \). It should also be noted that the choice of borrowing rate for entrepreneurs directly affects the individual default probability. From equation (9), it is seen that the individual default probability is inversely related to the borrowing rate, equation (10)

\[
\left( \ln \overline{R} - \delta + \frac{1}{2} \sigma_i^2 \right) / \sigma_i = \Phi^{-1} (\rho). \tag{10}
\]

### 2.2.1 Default Fraction for a Debt Portfolio

When risk is not idiosyncratic as in Williamson (1986, -87a), the distribution for banks’ lending must be derived from the underlying distribution for entrepreneurs. This is because the defaulting projects cannot be fully pooled with successful projects, as entrepreneurs never repay more than the loan and the agreed on interest. Vasicek (1987, -91, 2002) has derived the probability distribution for the unconditional default fraction for debt portfolios with a one factor setup. The unconditional default fraction is the share of defaulting loans in the portfolio. I follow Schönbucher (2000) and use the conditional independence to show how the unconditional default fraction can be found. Given the systematic shock, entrepreneurs are independent and normally distributed. Due to this independency of defaults, it is simple to find the conditional probability of default for an individual entrepreneur \( p (y) = P \left( v_i < \ln \overline{R} | Y = y \right) \), shown in equation (11)

\[
p (y) = P \left( v_i < \ln \overline{R} | Y = y \right) = P \left( \delta_i - \frac{1}{2} \sigma_i^2 + \sigma_i \sqrt{\rho} Y + \sigma_i \sqrt{1 - \rho} \varepsilon_i < \ln \overline{R} | Y = y \right) = \Phi \left( \frac{\ln \overline{R} - \delta_i + \frac{1}{2} \sigma_i^2 - \sigma_i \sqrt{\rho} y}{\sigma_i \sqrt{1 - \rho}} \right). \tag{11}
\]
First, denote the share of defaulting firms, the default fraction, by \( X \). Then, I can find the distribution for these default fractions, \( P (X \leq x) \), of a bank by iterated expectations as follows

\[
P (X \leq x) = E_Y [p (X \leq x | Y)] = \int_{-\infty}^{\infty} P (X \leq x | Y) \phi (y) \, dy. \tag{12}
\]

To find the conditional probability for default fractions, I use the assumption that a bank’s portfolio is infinite. Then, the fraction of actual defaults given \( y \) converges almost surely to the individual conditional default probability (11)

\[
\lim_{i \to \infty} P (|X_i - p (y)| \leq \epsilon, \ \forall i \geq I) = 1, \ \forall \epsilon > 0. \tag{13}
\]

Since the fraction of defaults is equal to the conditional default probability for an entrepreneur, expression (12) can be written as follows

\[
\int_{-\infty}^{\infty} P (X = p (y) \leq x | Y) \phi (y) \, dy. \tag{14}
\]

In (14), we see that the conditional default probability is either smaller or larger than default fraction \( x \). This is used to split the support for the integral into two areas with an indicator function (15). It takes the value of one if the conditional default probability is lower than the default fraction \( x \), and zero otherwise.

\[
I_x = \begin{cases} 
1 & \text{if } p (y) \leq x \\
0 & \text{else}
\end{cases} \tag{15}
\]

We can now use the indicator function and substitute for the first term in the integral (14), which results in the cumulative distribution:

\[
\int_{-\infty}^{\infty} I_x \phi (y) \, dy = \int_{-y^*}^{\infty} \phi (y) \, dy = \Phi (y^*). \tag{16}
\]

The lower boundary \( y^* \) is found by inverting (11), the conditional individual default probability, which yields:

\[
y^* = \frac{1}{\sigma_i \sqrt{\rho}} \left( \sigma_i \sqrt{1 - \rho \Phi^{-1} (x)} - \ln \frac{\mathcal{H}}{\delta_i} - \delta_i + \frac{1}{2} \sigma_i^2 \right). \tag{17}
\]

That is, I find the boundary point where the business climate induces a default fraction of exactly \( x \).

The intuition when going from (12) to (16) is that I know that the better the business climate is, large values for \( Y \) in my case, the smaller is the fraction of
Figure 3: The left graph displays realized losses for SHB. In the right graph, the frequency of losses is plotted together with the theoretical loss-function as stated in equation (18). The histogram is standardized so that the “hole” between the third and fourth bar is filled when the total mass is calculated. The height of the bar is the mean of the height of bar three and four.

defaulting firms. It must be the case that there is some business climate $-y^*$ in the support for $Y$, where smaller/larger realizations yield a larger/smaller default fraction than $x$. In terms of the common shock, I then obtain a truncation point; lower values on the common shock result in a functional value for $p(X \leq x|Y)$ of zero, and do not affect the value of the integral.

To find the density function for the default fraction, I can just use Liebnitz’s rule and differentiate the probability distribution (16) with respect to $x$, yielding (18)

$$
 f(x) = \frac{\sqrt{1-\rho}}{\sqrt{\pi}} \exp\left\{ \frac{1}{2} \left( \Phi^{-1}(x) \right)^2 - \frac{1}{2\pi \sigma_i \rho} \left( \sigma_i \sqrt{1-\rho} \Phi^{-1}(x) - (\ln R - \delta_i + \frac{1}{2} \sigma_i^2) \right)^2 \right\}. \tag{18}
$$

The density for default fractions is extremely complicated, involving the inverse of the normal distribution, but it can be numerically computed.

To see why this approach is common in practical debt management, I show an example using the credit losses for one major Swedish bank, Svenska Handelsbanken (SHB), Figure 3. The probability of individual default for any project is set to 1%, based on Carling et. al. (2004). I set the correlation to .3, based on the small sample estimate mentioned above. The left-hand panel shows credit losses as a time-series, and the right-hand panel is the resulting histogram for these losses, matched with the theoretical distribution.

Despite the small sample, the probability distribution does a good job matching
the data, though the sample is too small to perform a formal test.\footnote{A rule of thumb for a \( \chi^2 \)-test for the goodness-of-fit of the distribution, is that the expected number of objects in any bin is at least five. To circumvent the problem with very low number of expected observations in the “large” loss bins, bins two and three could be merged into one and the whole area to the right of the third bin be considered as the third bin. This way, there are only three bins, but they all have expected observations larger than five. The resulting test statistic under the null that SHB credit-losses are distributed according to the derived distribution is 2.74, which is \( \chi^2 \)-distributed with two degrees of freedom and a p-value of .25. Hence, the null cannot be rejected. This test procedure can be found in any standard literature for a basic statistics course, e.g. Aczel (1999).}

I assume equal sized exposures; in reality a loan portfolio consists of varying sizes. Vasichek (2002) shows that the distribution is robust for varying exposure. A necessary and sufficient condition is that

\[
\sum_{i=1}^{n} w_i^2 \to 0, \tag{19}
\]

where \( w_i \) is the portfolio weight \( i \). Condition (19) shows that as long as no exposure is too dominant in the portfolio, the result still holds. Schönbucher (2000) also shows that the exact distribution for a finite portfolio is almost identical to the limit distribution already for a portfolio of size 100. The infinite bank assumption is therefore not very restrictive when it comes to the derived distribution for default fractions.

### 2.2.2 Credit Losses

The standard way of solving for actual credit losses among practitioners and in the financial literature is to assume how large a share of a loan that can be recovered in a bankruptcy, henceforth called the recovery rate. I want to be slightly more general in this paper and have a varying recovery rate. This makes the portfolio loss fraction more consistent with the loss distribution for a single entrepreneur.

First, I study the bank’s actual gross cash-flow, the number of loans times the agreed repayment deducting losses made on defaulting borrowers, equation (20)

\[
\pi_m = m\bar{R} - \sum_{i \in P} R_i - V_i . \tag{20}
\]

\( m \) is the size of the bank in terms of the number of firms to which it lends and \( P \) the group of defaulting loans. I divide (20) with \( m \) to obtain the average gross cash-flow per loan in the portfolio, equation (21)

\[
\bar{\pi}_m = \bar{R} - \frac{\sum_{i \in P} R_i - V_i}{\bar{R}} . \tag{21}
\]
$\overline{V}$ denotes the average value on defaulting firms, $X$ denotes the default fraction as before, and the last part of the second term on the right-hand side is the unconditional average loss per loan, $l^c(\overline{R})$.

$\overline{V}$ is easily calculated; I take its log to obtain the unconditional average rate of the loss, defined as $\mu^c = \ln (\overline{V})$, which is calculated in (22)

$$\mu^c = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \left( \delta - \frac{1}{2} \sigma_i^2 + \sigma_i \sqrt{\rho Y} + \sigma_i \sqrt{1 - \rho \varepsilon} \right) \phi(\varepsilon) d\varepsilon \right) \phi(Y) dY. \quad (22)$$

The function $h(\overline{R}, Y)$ is defined as (23) below,

$$h(\overline{R}, Y) = \frac{\ln \overline{R} - \delta + \frac{1}{2} \sigma_i^2 - \sigma_i \sqrt{\rho Y}}{\sigma_i \sqrt{1 - \rho}}. \quad (23)$$

An explicit expression for the average mean recovery rate (22) cannot be obtained due to the normal kernel in the expectation. Using known central moments for the truncated normal, which can be found in e.g. Johnson et. al. (1994), (22) can be written as (24)

$$\mu^c = \int_{-\infty}^{\infty} \left\{ \left( \delta - \frac{1}{2} \sigma_i^2 + \sigma_i \sqrt{\rho Y} \right) \Phi(h(\overline{R}, Y)) - \sigma_i \sqrt{1 - \rho} \phi(h(\overline{R}, Y)) \right\} \phi(Y) dY. \quad (24)$$

Through these calculations, I only obtain the unconditional average loss rate, $l^c(\overline{R})$. It is not possible to obtain a simple expression for the conditional average loss rate. Instead, I assume the recovery rate to be equal to the unconditional average loss rate for all loss fractions; that is, the average credit loss is a function of the lending rate $\overline{R}$, but not the fraction of defaulting firms $X$.

3 Banking

In earlier papers with no verification costs of banks, such as e.g. Williamson (1986), it is sufficient to solve the lending side of banking. When introducing my systematic component in the economy, this implies that banks need to be monitored in the same way as entrepreneurs. Therefore, banks must solve two problems: first, lending to entrepreneurs and also monitoring them and second, offering a satisfactory deposit rate to consumers. All agents in the economy know the probability distribution for entrepreneurs’ values, but until banks’ lending rate, $\overline{R}$, is set, the probability distribution for banks’ values is unknown. If treating $\overline{R}$ as a parameter, I can still investigate the banks’ problem of offering a deposit rate to consumers and the role for reserves without first solving the lending rate.
3.1 Stating the problem

Banks will need to set a contract with consumers in the same way that they settle
the terms with entrepreneurs. This is done in the same way as entrepreneurs borrow
from consumers described in (1)-(2). The bank offers the depositors the repayment
$D$, and it holds reserves $\mathcal{R}$, maximizing its expected return on available reserves (27).
That is, banks maximize the return on invested capital. I let $b$ denote the bank’s
value per unit lended, $g(b)$ the probability distribution and $G(b)$ the cumulative
distribution for this value. Banks’ returns were defined in (21). By deducting the
auditing costs I obtain $b$ as a function of the realized loss-fraction $x$ of the stochastic
variable $X$ in (25)

$$b(x) = \overline{R} - \left( l^c (\overline{R}) + \frac{c}{N} \right) x.$$  
(25)

Since the value of the bank is determined by the realized loss-fraction $x$, the
probability distribution for banks’ values will also be determined by the probability
distribution of the loss-fraction, $f(x)$, defined in equation (18). I define the worst
possible value of banks by $\hat{b}$, which takes a positive value if the recovery rate for
banks is sufficiently large to cover the auditing costs. If credit losses are so large
that the limited liability constraint binds, $\hat{b}$ is zero. I then invert (25) so that $x$ is
expressed as a function of the parameters and $b$, this way I obtain my probability
function for bank return $g(b)$.

$$g(b) = \begin{cases} 
    a_1 f(x(b)) & \text{if } b > 0 \\
    a_2 & \text{for } b = 0 \\
    a_3 f(x(b)) & \text{for } b > 0
\end{cases}$$  
(26)

The constants $a_i$ are determined so as to make the probability mass integrate to 1.

Banks’ objective function is then as follows in (27)

$$\max_{D, \mathcal{R}} \frac{1}{\mathcal{R}} \int_{D-\mathcal{R}}^{\mathcal{R}} (b + \mathcal{R} - D) g(b) \, db.$$  
(27)

$D$ and $\mathcal{R}$ are as defined above, agreed repayment on deposits and reserves held. The
bank’s offer needs to be competitive to be accepted by consumers. The expected
utility to consumers must be at least as high as if they were lending directly to
entrepreneurs, i.e. $r_D$. Hence, the banks’ maximization is constrained by consumers’
expected utility, shown in equation (28)

$$\int_{\hat{b}}^{D-\mathcal{R}} bg(b) \, db + D (1 - G(D - \mathcal{R})) - \gamma G(D - \mathcal{R}) \geq r_D.$$  
(28)

The monitoring costs of banks, $\gamma$, were set to $2c$ in section 2.1.
I let the bargaining power in the deposit market be with banks.\footnote{If consumers have the bargaining power, banking will not be profitable unless the monitoring costs of banks are extremely low. Naturally, there exists some solution where consumers and banks split the gains from banking somewhere in the middle, but I feel little to be gained from such an exercise, which is why I assume banks to have the bargaining power.} This assumption means that the solution method of Williamson (1986, -87a), optimizing depositors’ expected utility, is no longer valid. Instead, there is a regular constrained maximization problem for which the standard solution method is to form a Lagrangian. In addition to the requirement that depositors are satisfied, there are also restrictions on the choice of controls, $D$ and $\mathcal{R}$. An offered deposit rate will be lower than the maximum return $D \leq \mathcal{R}$. Reserves are also constrained both from above and below; to prevent looting, reserves must be positive, and since banks have the bargaining power, no pure transfers to depositors should take place, i.e. $0 \leq \mathcal{R} \leq D - b$. Instead of explicitly including the lower bound, I follow Chow (1997) and let the first-order conditions be smaller than zero, if the constraints bind. The multiplier for the upper bound on deposits is $\lambda_D$, and for reserves $\lambda_R$. For simplicity, I assume the net return on reserves to be zero. My Lagrangian is then defined by (29)

$$
\mathcal{L} = \frac{1}{\mathcal{R}} \int_{D - \mathcal{R}}^{\mathcal{R}} (b + \mathcal{R} - D) g(b) \, db + \lambda_D (\mathcal{R} - D) + \lambda_R (D - \frac{\mathcal{R}}{b} - \mathcal{R}) + \lambda_c \left( \int_{D - \mathcal{R}}^{\mathcal{R}} b g(b) \, db + D (1 - G(D - \mathcal{R})) - \gamma G(D - \mathcal{R}) - r_D \right)
$$

(29)

Williamson (1986, -87a) observed that the constraint for the lenders’ expected value must bind; otherwise the offered compensation could be reduced and the value of the objective function increased. This also holds true for my bank’s optimization problem. There is no point in offering consumers a combination of reserves and a deposit rate resulting in a larger expected value than their alternative, i.e. directly lending to entrepreneurs earning $r_D$.

$$
\lambda_c > 0 \Rightarrow \int_{D - \mathcal{R}}^{\mathcal{R}} b g(b) \, db + D (1 - G(D - \mathcal{R})) - \gamma G(D - \mathcal{R}) - r_D = 0
$$

(30)

First-order conditions, w.r.t. $D$ and $\mathcal{R}$, are,

$$
\text{w.r.t. } D \Rightarrow \frac{1}{\mathcal{R}} \int_{D - \mathcal{R}}^{\mathcal{R}} g(b) \, db - \lambda_D + \lambda_R + \lambda_c (1 - \mathcal{R} g(D - \mathcal{R}) - G(D - \mathcal{R}) - \gamma g(D - \mathcal{R})) = 0
$$

(31)
\[ w.r.t. \quad \mathcal{R} \Rightarrow -\frac{1}{12} \int_{D-\mathcal{R}}^{\mathcal{R}} (b - D) g(b) \, db - \lambda_{R} + \lambda_{e} g(D - \mathcal{R}) (\mathcal{R} + \gamma) = 0. \quad (32) \]

I use the fact that the restriction on the minimum expected utility to depositors is binding, this yields equation (30). \( \mathcal{R} \) set to zero and \( D \) to \( \mathcal{R} \) can be ruled out except for very special parameterizations since the first and third terms in (30) are zero, which yields \( b - r_{D} = 0 \). This means that the scrap value of banks is so large per loan that the expected return on direct lending requires recovery rates that are extremely high. Further, \( D \) at the upper bound and \( \mathcal{R} \) at the lower bound can also be ruled out since the first term in (30) is then banks’ expected return on lending \( r \) and \( G(D - \mathcal{R}) \) = 1, which yields \( r - \gamma = r_{D} \). Once more, the calculated return on direct lending and the auditing cost of banks and parameters stipulated in sections 2 and 2.1 are not compatible with this condition. Due to my small variance in the production distribution, the expected return on banking would be at such a high level that the probability for any firm of having such a successful project is only 0.02\%. Hence, \( r = r_{D} + \gamma \) would not be a plausible average return on banks’ portfolios, even if the cost for banks to audit entrepreneurs were zero.

Suppose banks set the deposit rate to \( D^{*} = r_{D} \) and that they insure depositors against expected monitoring costs by keeping reserves large enough to always meet its obligations \( \mathcal{R}^{*} = r_{D} - b \), the first term in (30) is then zero and \( G(D - \mathcal{R}) = 0 \), and (30) is satisfied. This will only be a solution if auditing cost of banks is sufficiently large to outweigh the increased return from bank leverage. To find the optimal controls, I must explicitly solve the system of equations (30)-(31). To solve this system, banks’ lending rate, \( \mathcal{R} \), which has been treated as a parameter up until now, must be determined.

### 3.2 Finding the optimal lending rate \( \bar{R} \)

\( \bar{R} \) is determined in the same manner as \( \bar{R}_{D} \) in section 2. Entrepreneurs offer banks the repayment \( \bar{R} \) that maximizes their own expected utility, equation (33).

\[
\max_{\bar{R}} \int_{\bar{R}}^{\infty} (V - \bar{R}) \psi(V) \, dV \quad (33)
\]

The maximization is constrained in that banks’ expected utility should be at least as large as the endogenous current market rate, \( r \). Since every loan granted by a bank is now part of a portfolio, it is the contribution of the loan’s to the portfolio’s expected return that is of importance. Hence, the constraint is that the average
Figure 4: The left-hand graph shows bank lending when using the Williamson assumption of i.i.d. projects, i.e. no systematic risk. The optimal point is a repayment of 1.1925, yielding an expected utility for the bank of 1.0831, and for the entrepreneur of .7497. The right-hand graph shows bank lending with systematic risk, the coefficient of correlation is set to .3. Banks’ expected utility falls by 25% when systematic risk is introduced. The expected utility is now 1.0626 and the optimal repayment is 1.1389. With this borrowing contract, the expected utility for an entrepreneur is .9444.

return on a loan should be at least as large as \( r \) in equation (34).

\[
\overline{\Pi} - \left( V(\overline{\Pi}) + \frac{c}{N} \right) \int_0^1 x f(x, \overline{\Pi}) \, dx \geq r
\]  (34)

Just as in section 2, I use the observation by Williamson (1986, -87a) that if the bargaining power is with the lender, the solution is found by maximizing the constraint (34). To show how the Williamson (1986) result is affected by correlated project returns, I calculate two solutions; one where banks are able to diversify all systematic risk, and one with correlated project returns where the coefficient of correlation is set to .3. The auditing cost is \( c/N \) per project, where \( N \) is set to two. That is, we assume that every entrepreneur needs to borrow from two investors to have sufficient capital. Even with this choice of very small entrepreneurs, much of the hump around optimum is flattened. In Figure 4, we see that the bank’s utility function over varying repayment levels has a much less pronounced hump than the graph of direct lending with twice the auditing cost; right-hand graph in Figure 2.
Without systematic risk, left-hand graph in Figure 4, we see that the optimal repayment level is much higher than when banks are exposed to risk, as in the right-hand graph in Figure 4. When the banks can diversify all risk, they set an optimal repayment level of 1.1925. This results in an expected net return of 8.31% to banks, but entrepreneurs’ utility is reduced to .7497. In my setting with systematic risk through correlated returns on entrepreneurs’ projects, the expected net return per loan is reduced by 25% to 6.26%. The reason is that the systematic risk increases the expected auditing cost, which makes banks’ preferred repayment level much lower, 1.1389, to compensate for the increased default probability among borrowers.

If I compare the resulting optimum of bank lending to direct lending, there is a higher repayment level 1.1389, which naturally leads to more bankruptcies and lower expected utility for entrepreneurs. Banks’ expected net return per granted loan is almost twice (6.26%) that of the direct lending consumer (3.57%), with double overlapping auditing. Entrepreneurs are roughly 9% worse off with (.9444) borrowing from banks than directly borrowing from small consumers. Since entrepreneurs are worse off, I cannot say that the introduction of banks is Pareto improving.

![Figure 5: Plot of loss distributions when the correlation between entrepreneurs’ projects outcome varies between .01 and .5. The individual probability of default is set to 5%.](image)

To check if the choice of correlation, $\rho = .3$, is crucial for the optimal lending rate, some other coefficients of correlation are tried. Reducing the correlation to .1 only marginally raises the expected payback per loan, the sixth decimal increase marginally. The same is true when we increase the correlation to .5, but the expected payback is now naturally reduced. That is, the impact on the lending rate is
Figure 6: Graph of banks’ expected utility in terms of percentage return on invested capital. The return is graphed over default exposure, i.e. how secure is the bank. When exposure is zero, the bank is fully insured against credit losses and depositors need not monitor. At the optimum point, default exposure is .002, i.e. the probability for a bank to default is .2 %. At the optimal point, the reserve ratio is 15.5 % and the expected utility per unit of reserves is 1.1646.

obtained when going from idiosyncratic risk to just introducing marginal amounts of systematic risk. To some extent, this result is due to the fact that the recovery rate is the same across states for a given lending rate. But mainly because small amounts of systematic risk results in substantial increase in tail probabilities, in Figure 5 it is shown how the loss structure changes with different correlations. It is clear that the probability for large losses, the right tail, is not changed much when the correlation is altered.

3.3 Optimal controls \(D\) and \(R\)

When I have solved for the optimal lending rate \(\overline{R}\), and the resulting recovery rates, I can solve for optimal reserves and deposit rates. The solutions are graphed in Figure 6. On the x-axis, I have the probability of default for the bank. It can be seen that with the chosen parameters, the probability of default in optimum is rather low, .2 %. If monitoring costs are raised, the “optimal” default probability will be even lower. The optimal reserve ratio is 15.5% and the expected utility on one unit of reserves is 1.1646. Rich consumers greatly enhance their return on capital if forming a bank. By using their capital as reserves, they can earn a 16 %
Figure 7: Sensitivity check of the choice of value for $\rho$. The left-hand graph shows how the expected gross return to the bank is changed when $\rho$ is reduced to .2. The optimal reserve level is then 13.7 %, the resulting probability of default is .25 %, and net return is 17.98%. In the right-hand graph, the correlation is raised to .5, which results in a solution where banks fully insure depositors by holding reserves sufficiently large to cover the maximum possible credit loss. Reserves are then 19.2 % and their rate of return 13.97 %.

return, which can be compared with directly investing in entrepreneurs’ projects, earning a 6 % return.

Even if the default probability is low, .2 %, this still means that depositors would occasionally need to audit banks, which reduces their expected return. The reason why it is not optimal to fully insure depositors and totally eliminate the non-productive auditing cost is that banks benefit from leverage. When banks’ reserves are reduced, they can lend to more entrepreneurs, yielding a lower return per loan but a greater base of loans generating a positive return.

I perform a sensitivity analysis by changing the correlation in the same way as was done in section 3.2. Unlike the sensitivity analysis for the lending rate, which only gave marginal effects of altered correlation, the change of systematic risk has a large impact on the determination of reserves and deposit rates. Naturally, this is due to the increased concavity imposed on the deposit problem through the larger monitoring cost, when double auditing is needed. When the probability of large default fractions in a bank’s portfolio is large, the cost for depositors in terms of non-productive auditing costs is increased. To compensate the depositors for this expected cost, banks’ optimal choice is to raise reserves neutralizing some of the
expected audit states. The right-hand panel of Figure 7 shows the expected utility for a bank when the systematic risk is increased by setting the correlation to .5. With this large amount of systematic risk it is, in fact, optimal for banks to fully insure depositors. Banks carry reserves equal to the difference between consumers’ outside option, lending directly, and the minimum payback on the bank’s loan portfolio.

When the correlation is instead reduced to .2, I obtain the opposite result showed in the left-hand panel of Figure 7, i.e. banks reduce their reserves to increase their leverage. This can be done due to the lower default risk for the bank, which reduces the expected auditing costs for depositors. The optimal choice for reserves with a correlation of .2 is 13.7 % and results in a default probability of .25 %. With the lower reserves, the return on investment to the bank is 17.98 %.

The sensitivity analysis captures the actual behavior among banks in Sweden. Small provincial banks keep much larger reserves than the large banks servicing all of Sweden. It can be argued that by mainly granting loans in one province, it is not possible to obtain geographical diversification which, in turn, implies a larger correlation in the debt portfolio. That is, my model actually captures some of the behavior among banks in the Swedish market.

Another implication of my model is that in the described economy, available financing through banks is directly related to the wealth of bankers. If banks’ total reserves are large, many loans are granted. This could point at a possible link to the real economy through the credit market in a dynamic model. A negative real shock to the economy erodes banks’ reserves, which contracts bank lending, leading to fewer firms that can generate real output in the next period.

4 Conclusion

I use mathematical methods to derive a distribution for a bank’s debt portfolio in a two-period model to study under what conditions financial intermediation will be established. Assuming the value of borrowers’ projects to be log-normal and correlated, the direct lending problem is derived. I show that for an efficient debt contract to arise, with log-normality, it is important that the mean-variance ratio is large, i.e. the project needs to be rather safe. When empirically estimated parameters are used, the resulting debt contract has a negative interest rate. The auditing cost must be set so low that the contract is almost pure risk-sharing, i.e. lenders always audit. This could be the reason why collateral and cosignatories are so frequently used in banking. An implication of this is that policies reducing the value of banks’ collateral could make credit markets work less efficiently, driving out debt financing which is mostly used by poorly capitalized firms.

With the derived distribution for debt portfolios, expected utility for direct lending and debt portfolios can be compared. I analyze under what conditions financial
intermediation yields higher expected utility for investors/consumers than no inter-
mediation. A main finding is that banks can only be started by rich investors, since
it requires substantial reserves; due to the correlated borrowers, a bank’s portfo-
lio will not generate a non-stochastic outcome. This raises the expected auditing
cost of banks and makes them inefficient, unless they carry reserves reducing the
expected auditing cost for depositors. That is, banks need to carry reserves that are
sufficiently large to almost eliminate the probability of a default. Moreover, banks
must have bargaining power both against borrowers and depositors. If consumers
have too much bargaining power, deposits will be too expensive for banking to be
lucrative.

The requirement that banking must be backed by reserves means that less
wealthy banks lead to a tightening of bank supplied loans. This is similar to the
result of Holmström and Tirole (1997), where they use moral hazard and perfect
correlation between entrepreneurs. An implication is that banks could enhance and
prolong business cycle downturns, since a macroeconomic shock that reduces banks’
reserve levels leads to less granted loans, thereby slowing down the recovery of the
economy.

I feel that my findings are well in line with stylized facts from credit markets, in
spite of the two-period model. Banks’ reserves are large to keep them afloat through
a rather deep recession, given that they have been involved in sound banking practice
which only involves lending to low risk projects. Reserve requirements only come
into play when banks have been involved in more or less reckless lending.

I can also explain why small provincial banks carry much larger reserves than
the large market leading banks. Since they cannot diversify geographical risk, their
portfolio will be more risky and it is optimal to fully insure depositors as the expected
auditing costs grow quickly.
References


Essay II
Optimal Rebalancing of Portfolio Weights under Time-varying Return Volatility

1 Introduction

This paper deals with horizon effects in optimal portfolio choice. In this context, horizon effects mean that investors portfolio weights depend on the remaining time before liquidation. The implication is that investors with different investment horizons, but identical risk aversion and wealth, choose different investment portfolios.

During the fall of 2000, Sweden launched a new pension plan consisting of two parts; an income based pay-as-you-go scheme and a fully funded part. The latter is individually managed through a government agency (PPM) where agents can select different privately managed mutual funds (currently close to 700). The money of agents who, for one reason or another, do not make any choice, will be invested in a portfolio managed by the government. In the new pension system, agents can rebalance the portfolio of the funded part at no transaction cost, but five different mutual funds at most can be held at any point in time.\(^1\)

When the new pension system was launched, several fund managers tailored mutual funds for different age groups and branded them “generation funds”. The idea is that an agent picks a mutual fund solely on basis of birth year, by selecting the appropriate “generation fund”.\(^2\) All “generation funds” have the same strategy,

---

\(^1\)The information is from the Premium Pension Authority PPM, www.ppm.nu

\(^2\)Fund managers differ slightly in how they define a generation, varying from between five to ten years.

---

*I would like to thank my advisor Hans Wijkander and opponents at Licenciate defence and final seminar, Paul Söderlind and Martin Holmén for comments and suggestions. The paper has also benefited from comments of participants at the 5\(^{TH}\) International Conference on Macroeconomic Analysis and International Finance in Crete, Enter Jamboree in Toulouse, “Measuring, Modelling and Forecasting Financial Volatility” workshop in Uppsala, and seminar participants at Stockholm University and Sveriges Riksbank. The author gratefully acknowledge computational power supplied by the Department of Mathematics and financial support from Sparbankernas forskningsstiftelse and SIAMON.*
namely to decrease the share of stocks in the portfolio over time. This is not surprising since the investment industry has for a long time been giving investment advice such as: hold a percentage equivalent to your age in Treasury bills and invest the remaining part of your portfolio in stocks or, hold 100% stocks if your investment horizon is longer than 5-10 years. The reason for this kind of advice is often that stocks are claimed to be less risky in the long run. When the new pension plan was launched, one of the major fund managers offered a decreasing scale from 90% stocks for agents born in the sixties to 50% stocks for agents born in the forties, thereby implying a 2% decrease in risky assets per year. This strategy is very similar to the first investment advice stated above, although with twice as fast a substitution between stocks and bills.

An important question is whether the behavior described above can be justified for utility maximizing agents? Until recently, academics have given a negative answer to that question on the basis of a result by Samuelson (1969) and Merton (1969), which is that optimal rebalancing agents with an isoelastic utility function should hold a fixed share of risky assets, regardless of the investment horizon. However, the result relies on an assumption of identically and independently distributed returns (i.i.d.). Merton (1971, -73) relaxed the assumption of i.i.d. returns and isoelastic utility when analyzing intertemporal portfolio choice. With these generalizations of the model, he shows that there can be a “hedging-demand” for risky assets, which leads to time-diversification. This horizon effect on portfolio weights is created by the desire to hedge against shifts to lower returns in the future, hence the name hedging-demand.

What type of return distribution is appropriate when I conduct an analysis of portfolio choice? Several recent empirical articles have rejected the assumption of i.i.d. returns, see e.g. Lo and MacKinlay (1988) or Bollerslev, Chou and Kroner (1992). Hence, more involved models should be used when modeling returns. During the nineties, several new attempts have been made to solve intertemporal investment models under time varying investment opportunities. Kim and Omberg (1996) use mean reverting risk premium in continuous time (an Ornstein-Uhlenbeck process), to show that the investment in risky assets is increasing in the investment horizon. Campbell and Viceira (1999) solve approximately for optimal portfolio weights for an infinitely lived investor, in discrete time, using a model with a time varying mean. Instead of analytically solving for optimal investment strategies, Barberis (2000) simulates optimal portfolio weights when returns are predictable and investors also face an estimation risk. He shows that in his setting, there are horizon effects on portfolio weights.

The above papers all have in common that they do not consider the evidence of volatility clustering, first noted by Mandelbrot (1969). He saw in data that large shocks to expected returns, negative or positive, are followed by new large shocks.
Later, Black (1976) noted that not only is there volatility clustering, there also seems to be a negative correlation between shocks to returns and volatility of returns; that is, large negative shocks to expected return tend to increase volatility more than do large positive shocks. These features are incorporated by Chacko and Viceira (2000) when they use stochastic volatility to model returns and derive approximate analytical solutions to the intertemporal investment problem in continuous time. They show that time varying volatility gives rise to changing hedging demand. However, like Campbell and Viceira (1999), Chacko and Viceira (2000) do not consider different investment horizons. Instead, they log-linearize the problem and obtain analytical expressions for portfolio weights for an investor with an infinite horizon and then, they study how these weights change for different investment horizons.

This paper investigates whether horizon effects are present under optimal portfolio rebalancing with time varying risk, in discrete time. Risk is measured as return volatility. I use models of the GARCH class to explore implications of volatility on portfolio weights. There are several reasons for using GARCH models for asset returns. First, they can capture some of the excess kurtosis that is present in empirical data and according to Dijk and Frances (2001), they are frequently used by practitioners. Second, Engle and Patton (2001) also point out that even if the true data generating process for assets is not GARCH, these models still serve as very good filter rules or approximations. Third, GARCH often results in models with few parameters which is tractable; see Hansen and Lunde (2001). I restrict the analysis to the growth of investors’ wealth, i.e. investors have no labor income. Campbell, Cocco, Gomes and Maenhout (2001) show that due to the low correlation between labor income and asset returns, labor income acts as a substitute for the risk-free asset. This will generate horizon effects on portfolio weights for pension savings, but is unrelated to the risky assets distribution over time.

In GARCH models, risk measured as volatility is perfectly forecastable and I investigate whether this is enough to induce horizon effects on portfolio choice. Two different types of volatility models are used to perform the analysis, the first with a constant mean and the second with a time-varying mean. The constant mean models used are Bollerslev's (1986) GARCH, Glosten, Jagannathan and Runkle's (1993) GJR-GARCH and Nelson’s (1991) EGARCH, the last two being asymmetric models. As a second step, I also analyze the effect of predictability of expected returns through the risk level, in accordance with Merton (1973). He shows that if returns are not \textit{i.i.d.}, the expected return on risky assets is related to the risk. I follow the work by Engle, Lilien and Robins (1987), using GARCH volatility in the mean equation, GARCH-M. In accordance with Merton (1973), they stipulated that when investors face higher volatility, they will demand a higher risk premium and thus, the authors included conditional volatility in the mean equation.

Why would we expect GARCH models to induce horizon effects? First, the
assumption of *i.i.d.* returns used in Samuelson (1969) and Merton (1969) does not hold and second, Merton (1973) showed that time-varying risk could potentially give rise to hedging demand. Third, Chacko and Viceira (2000) show that shocks to risk have different effects on different investment horizons. Remember that they analyze how the utility for an investor with an infinite investment horizon is affected on different horizons.

The analysis of optimal portfolio weights follows the method of Barberis (2000). Returns are assumed to follow estimated GARCH processes, and expected utility is then approximated by the average of a large random draw from this process. The simulation of portfolio weights is performed for integer percentages between 0-99%, where the optimal portfolio weight is the one rendering the largest “expected” utility.

For models with a constant mean, investors have constant portfolio weights independent of the investment horizon. The reason is that the ability to forecast risk does not affect the future expected return. Hence, there is no hedging demand for the risky asset with pure GARCH returns. When volatility is included in the mean equation, resulting in predictable returns, there are horizon effects. Optimal portfolio weights on shares decrease as the point of liquidation approaches, which seems to be the case in practice, although the effect is only present for the last 2-3 years. I show that size and duration of the horizon effect mainly depends on three things: First, it is the size of the coefficient on the mean effect that triggers the horizon effect and second, the persistence in conditional volatility is important, since it introduces persistence in returns. Third, the correlation between shocks to expected returns and volatility is important, which I model by the asymmetry parameter. In fact, the result in this paper supports the second investment advice stated above, although the effects are somewhat small in terms of magnitude.

The paper is organized as follows. In Section 2, the method of optimization of utility from Barberis (2000) is presented. The return models are discussed in Section 3, together with some statistical properties and implications for buy-and-hold strategies. Section 4 first presents the data used for estimation, CRSP 1954-94 for the US and the Frenberg and Hansson (1992) data for Sweden 1919-1999 (updated version). After presenting the data, I estimate the model parameters. Simulations of optimal portfolio weights are made in Section 5 together with a sensitivity analysis. My conclusions are stated in Section 6.

## 2 Optimizing Utility

Consider a model with a risk averse individual who at time $t$ has an endowment $W_t$ to invest for consumption at some future date $T$. This individual chooses an investment strategy balancing growth and risk so as to maximize expected utility. The problem can be related to the standard CAPM, where one seeks the mix of the
risk-free and risky assets that maximizes expected utility. Investors in my model face a multiperiod rather than a single period optimization problem. The solution then involves yet another dimension, namely dynamic rebalancing of their portfolios. The expected utility of final wealth will now depend on actions not yet taken by the investor.

What does this mean in practice? For an investor, with e.g. thirty periods to retirement, the problem is to select a portfolio weight today \( t=0 \) that maximizes the expected utility of wealth thirty periods ahead. For this purpose, he must form expectations over future developments in the asset market and how he will rebalance his portfolio according to these fluctuations. The result is a probability distribution for expected utility of final wealth that depends on the sequence of 30 chosen portfolio weights. In the next period \( t=1 \), the investor can revise his choice by selecting 29 new portfolio weights, leading to a new probability distribution for expected utility of final wealth. At \( t=29 \), the investor only selects a single portfolio weight that will decide the final wealth, depending on the outcome in the asset market. The only sequence of portfolio weights that will not result in a probability distribution over the expected utility of final wealth, is the choice of no risky assets at any point in time. This choice results in a final wealth of \( W_0 \exp \{ 30 r_f \} \) with certainty, since I assume a constant risk-free real return. The investor’s maximization problem is written as follows:

\[
J(W_t, T) = \max_{\{\omega_i\}} E_t U(W_{t+T}) \tag{1}
\]

where \( W_{t+i} = \left[ (1 - \omega_{t+i-1}) \exp (r_f) + \omega_{t+i-1} \exp (r_f + r_{t+i}) \right] W_{t+i-1} \tag{2} \)

\[
W_t = W_0 > 0, \quad i = 1, \ldots, T,
\]

In equation (1), \( \{\omega_i\} \) is the sequence of portfolio weights, which are to be selected from \( t \) through \( T \) to maximize the expected utility of final wealth. The risky returns \( r_{t+i} \) are excess log returns, i.e. continuously compounded returns and \( r_f \) is the real risk-free log return. Note that when weights are chosen for the last time, the problem will be equivalent to a buy-and-hold setting, \( \omega \) is selected in \( t + T - 1 \) and the portfolio is then held until \( T \).

I start the discussion when the agent only has one period left before retirement. The maximization problem is as follows when we consider a continuous probability distribution for returns:

\[
J(W_{t+T-1}, T | \hat{\theta}, h_{t+T}) = \max_{\omega_{t+T-1}} \int U(W_{t+T}) p \left( r_{t+T} | \hat{\theta}, h_{t+T}^2 \right) dr_{t+T}. \tag{3}
\]

In (3) \( p (\cdot) \) is the probability function for return given some parameter vector \( \theta \) and conditional volatility \( h_t^2 \). The parameter vector and conditional volatility will
be extensively presented in Section 3, where the different return models used in the paper are presented. The state in the return models considered is completely specified by the current shock to return, the history of shocks to return and the parameter vector. The maximization of the integral in (3) is approximated with the discrete maximization problem below

\[ J(W_{t+T-1}, T) \approx \max_{\omega_{t+T-1}} \frac{1}{I} \sum_{i=1}^{I} U \left( \left( 1 - \omega_{t+T-1} \right) \exp(r_f) + \omega_{t+T-1} \exp(r_f + r_{t+T}^i) \right) W_{t+T-1} \].

(4)

In the approximation (4) of the integral, superscript \( i \) on return indicates the \( i \):th draw from the return distribution, \( I \) is the size of the random draw to approximate the expected value. Barberis (2000) shows that in his setting with VAR returns, it is sufficient with a random draw, \( I \) of size 1,000,000.

When we step back one period or more and let the investor have two or more remaining periods for choosing portfolio weights, the problem becomes more interesting. In this case, where agents can rebalance, we are required to solve a dynamic optimization problem. From Kreps (1990), we know that finite horizon dynamic programming problems can always be solved recursively,\(^3\) due to Bellman’s principle of optimality. “If a strategy is optimal for each point in time at that point of time, given that an optimal strategy will be used thereafter, then the strategy is optimal.” Since in practice, it is possible for agents to rebalance in every period, the focus in this paper is on that particular case which we will call optimal rebalancing; although I will also briefly address multiperiod buy-and-hold strategies in section 3.

In the present problem, there is a continuum of states, which makes it impossible to exactly solve the dynamic problem. Instead, I first discretize the state space and then solve the problem recursively, by applying Bellman’s principle of optimality.

If writing the \( t \):th period value function (1) using the value function for the \( t + 1 \) period on the right-hand side, I obtain the following Bellman equation

\[ J(W_t, T) = \max_{\omega_t} E_t J(W_{t+1}, T). \]

(5)

To be able to simulate solutions to the investment problem presented above, I must specify a utility function. Arrow (1971) suggests that investors have a close to constant relative risk aversion. Hence, a plausible utility function should have this features. A common choice in the finance literature is to use the power utility presented in equation (6), see e.g. Barberis (2000), Mehra and Prescott (1985) and Samuelson (1969)

\[ U(W_{t+T}) = \frac{W_{t+T}^{1-A}}{1-A}. \]

(6)

\(^3\)The original reference is Bellman (1957).
One problem with power utility is that extreme values of the coefficient of relative risk aversion must often be used to fit models to empirical data; see, for example, the “equity premium puzzle” in Mehra and Prescott (1985). As will be seen later in section 5, this is not the case in this paper where coefficients of relative risk aversion of 6 and below are used. With power utility, it is easily seen that the coefficient of relative risk aversion equals $A$

$$R (W) = -W \frac{U'' (W)}{U' (W)} = A.$$  

(7)

Since I solve the problem by numerical simulation over the Bellman equation in (5), the dependence on initial wealth must be removed to obtain a tractable problem. The following manipulations are made to remove the dependence of initial wealth in my Bellman equation. From the definition of the wealth dynamics in (2), final wealth can be written as:

$$W_T = \prod_{i=0}^{T} ((1 - \omega_{t+i}) \exp (r_f) + \omega_{t+i} \exp(r_f + r_{t+1+i})) W_t,$$

(8)

where $W_i$ is some initial wealth. Using this together with my functional form of utility (6) to rewrite the value function (5), initial wealth can be moved outside the maximization of the expectation, since it is known as of $t$ and can be regarded as a constant. This gives us equation (9)

$$J (W_t, T) = \frac{W_t^{1-A}}{1-A} \max_{(\omega_t)} \left[ \prod_{i=0}^{T} ((1 - \omega_{t+i}) \exp (r_f) + \omega_{t+i} \exp(r_f + r_{t+1+i})) \right]^{1-A}.$$  

(9)

I want to find the optimal path of portfolio weights independent of initial wealth to have a more tractable simulation problem. First I lead (9) from $t$ to $t + 1$

$$J (W_{t+1}, T) = \frac{W_t^{1-A}}{1-A} \max_{(\omega_{t+1})} \left[ \prod_{i=1}^{T} ((1 - \omega_{t+i}) \exp (r_f) + \omega_{t+i} \exp(r_f + r_{t+1+i})) \right]^{1-A}.$$  

(10)

By using the dynamics of wealth from equation (2), I can write $J (W_{t+1}, T)$ in terms of $W_t$, return in $t + 1$ and a function $Q (t + 1, T)$ consisting of the maximization part in (10).

$$J (W_{t+1}, T) = \frac{W_t^{1-A}}{1-A} ((1 - \omega_t) \exp (r_f) + \omega_t \exp(r_f + r_{t+1}))^{1-A} Q (t + 1, T)$$  

(11)

Naturally, the value function in $t$ can also be written in this form, dividing the function into a product of $W_t$ and $Q (t, T)$

$$J (W_t, T) = \frac{W_t^{1-A}}{1-A} Q (t, T).$$  

(12)
Both value functions now consist of two parts, wealth in $t$ and a maximand that is only defined over future returns. By just substituting these two equations into (5) $W_t$ cancels out and the final Bellman equation can be formulated without dependence on initial wealth, which makes the simulation problem more tractable

$$Q(t, T) = \max_{\omega_t} \left\{ [(1 - \omega_t) \exp(r_f) + \omega_t \exp(r_f + r_{t+1})]^{1 - \lambda} Q(t + 1, T) \right\}. \quad (13)$$

Using (13), I can recursively solve for optimal portfolio weights, using the fact that in the last period, the value of $Q(\cdot)$ is known. The problem thus reduces to solving a sequence of one-period optimizations. The dynamic problem can then be solved by the approximation used for the expected value of the one-period problem (4). However, I will only solve the value for $Q(t, T)$ for the prespecified grid of states, meaning that I will approximate the $i$:th drawn state with the closest grid state $j$; that is, the continuation value, $Q^i(t + 1, T)$ in (13), is not calculated for the drawn state, but for the grid state $j$ closest to the drawn value.

Deciding the size of the state grid is simply a question of a trade-off between speed and accuracy in the simulations. By decreasing the step size on the grid, the approximation between the drawn state and the grid state becomes better, but the time required for the simulation increases quickly. The grid in the Barberis (2000) paper is set to 25 to get smooth investment curves; in this paper this is not sufficiently large since the conditional variance is highly volatile. Even if I use a grid with 40 states, there is still some oscillation in the choice of portfolio weights, but it is judged to be sufficiently large to obtain reliable results.

The maximizing is done by varying the portfolio weight from 0-99% by integer steps for all different states and drawn returns in every period. Hence, I exclude short selling and impose a non-borrowing constraint. Mutual funds are not allowed to take short positions according to Swedish law, which is one justification for excluding short selling. There are several reasons for justifying the exclusion of borrowing. Theoretically, the integral in (3) is equal to $-\infty$, since expected utility is not bounded from below when agents are allowed to risk their entire wealth. For simulation purposes, it is intractable to have large grids for which to calculate optimal utility. Therefore, it is not advisable to simulate over too large a number of different portfolio weights. One choice is then to restrict the analysis to between 0 and 99%. Finally, the simulations in section 6 can be performed setting the coefficient of risk aversion such that agents choose to have between 0 and 99% stocks for all reasonable states. In fact, performing simulations with the borrowing constraint set to allow 30% borrowing together with lower risk aversion simply shifts the portfolio weight curve upward. The expected value of utility for each portfolio is calculated using the random draws and choosing the portfolio with the highest utility.

---

4 Mutual funds are allowed to use derivatives, if this will improve the efficiency in managing the assets.
3 Return Models

This section is an exposé over the GARCH models used to generate returns for the portfolio problem. First, I present the standard GARCH(p,q) model which captures the stylized fact that financial volatility is clustering. Due to Black’s (1976) observation that large negative shocks to expected return increase volatility more than large positive shocks, I then extend the standard GARCH(p,q) to asymmetric GARCH models where this effect is captured. To incorporate the intuitive possibility that investors require a higher expected return when risk is increased, I also consider GARCH-in-mean models. These include the conditional volatility in the expected return equation and are therefore consistent with the equilibrium model by Merton (1973). Finally, I briefly discuss the implications for buy-and-hold investors by studying aggregation properties of the standard GARCH, and the autocorrelation structure of the other models.

3.1 Symmetric Models

It is convenient to use i.i.d. stochastic processes when modeling economics. The most popular choice by far is to base the the process on the standard normal distribution. A model of expected returns would then be as follows:

\[ r_t = \mu + \varepsilon_t, \quad \varepsilon_t \sim N(0, h^2) , \]  

where \( h^2 \) is the variance of returns. However, such a model ignores the facts, presented in the introduction, that returns must be distributed according to a much more complicated process. Engle (1982) presented the path breaking model of autoregressive conditional heteroskedasticity, ARCH(q), by modeling volatility as a function of \( q \) lagged shocks to returns. In his setting, the variance is no longer constant as in (14), but a function varying over time, \( h_t^2 \). Later, his model was generalized and improved to GARCH(p,q) by not only including \( q \) lagged shocks, but also \( p \) lagged volatility terms, Bollerslev (1986)\(^5\)

\[ r_t = \mu + \varepsilon_t, \quad \varepsilon_t \sim N(0, h_t^2) \]  

\[ h_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}^2. \]

From (15), it is clear that the difference between i.i.d. returns and GARCH returns is that the variance conditioned on passed observations is time-varying, hence

\(^5\)The ARCH model often requires a large number of lagged shocks in the volatility equation to fit the data. A problem with many lags in the volatility equation is that an arbitrary declining structure must often be imposed in order not to violate the non-negativity constraint, Bollerslev (1986).
the $t$ subscript on $h_t^2$ in (15) and (16). In (14) and (15), I have assumed shocks to be normally distributed without loss of generality.\footnote{Other distributions can naturally also be used, such as the $t$-distribution which is in fact commonly used. But the choice of distribution for shocks mainly affects the techniques for estimating the models.} To ensure positive conditional variance and stationarity of the GARCH process, the following conditions must be fulfilled (note that the non-negativity constraints on all parameters are sufficient but not necessary):

$$\alpha_0 > 0, \quad \alpha_i \geq 0, \quad i = 1, \ldots, q \quad (17)$$

$$\beta_i \geq 0, \quad j = 1, \ldots, p \quad (18)$$

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_i < 1. \quad (19)$$

Although GARCH$(p,q)$ captures the volatility clustering of financial data, it cannot explain the evidence of shocks to expected return being negatively correlated with volatility, Nelson (1991).\footnote{Nelson also notes two features regarding stationarity and oscillation, which are not considered here.} That is, a positive/negative shock reduces/increases volatility. However, the GARCH$(p,q)$ only models the magnitude of shocks, since only the squares of past shocks are utilized together with past conditional volatility. This means that a negative shock to return will have exactly the same effect on conditional volatility as a positive shock. Next, I present two models extending the standard GARCH, so that it also captures the stylized fact of shocks to expected return being negatively correlated with volatility.

### 3.2 Asymmetric models

I start by presenting Nelson’s (1991) EGARCH, which models the negative correlation between shocks to expected return and volatility. By modeling the log of the conditional variance, he avoids the problem of having to constrain the parameters in the model to be positive. Thus, conditional variance can be modeled by the untransformed shocks, instead of the square of shocks, which makes it possible to introduce a negative correlation between news and conditional variance

$$\ln \left( h_t^2 \right) = \alpha_0 + \beta \ln \left( h_{t-1}^2 \right) + \alpha_1 \frac{|\varepsilon_{t-1}|}{h_{t-1}} + \gamma \frac{\varepsilon_{t-1}}{h_{t-1}}. \quad (20)$$

Conditional variance is modeled by equation (20), the mean equation will be of the same form as for the above described standard GARCH in (15). The $\alpha_1$ captures the magnitude effect of a shock by taking the absolute value of the shock, instead of the square, as in regular GARCH models. $\gamma$ captures the correlation between shocks to...
expected return and conditional variance by letting the shock appear untransformed in the conditional variance. According to the findings of earlier research, e.g., Black (1976), the sign on $\gamma$ should be negative, indicating a negative correlation between shocks to return and conditional variance. The shock in the mean equation is still assumed to be normal, as in the symmetric GARCH case.\footnote{Nelson (1991) uses the Generalized Error Distribution GED, which includes the normal distribution as a special case.} The condition for weak stationarity of this model is simply that $|\beta| < 1$.

Zakonian and Rabemananjara (1993) and Glosten, Jagannathan and Runkle (1993) independently developed an alternative to Nelson’s EGARCH to model asymmetric effects between shocks to return and volatility, threshold GARCH. In this paper, the model by Glosten et. al. is used, which is commonly referred to as GJR-GARCH, due to its authors. The model base is a standard GARCH but also includes a parameter capturing asymmetry. By including the square of shocks to expected return multiplied by an indicator function, which takes the value of 1 when $\varepsilon_t < 0$ and 0 when $\varepsilon_t \geq 0$, I can model the asymmetry

\begin{align}
 h_t^2 &= \alpha_0 + \beta h_{t-1}^2 + \alpha_1 \varepsilon_{t-1}^2 + \gamma \varepsilon_{t-1}^2 I_{t-1} \\
 I_t &= 1 \text{ if } \varepsilon_t < 0 \text{ otherwise } 0.
\end{align}

This model, GJR-GARCH, nests the standard GARCH and is as restrictive as the ordinary GARCH model in terms of parameter restrictions; that is, $\alpha_0$, $\alpha_1$ and $\beta$ must be positive and $\gamma + \alpha_1 \geq 0$ for the model to guarantee positive conditional variance. The stationarity condition of the process for conditional variance is, $\gamma/2 + \alpha_1 + \beta < 1$, due to the symmetric standardized errors, which is intuitive since the $\gamma$ is effective for half of the errors only. The advantage of this model over EGARCH is that I can test the standard GARCH model against asymmetry, with the hypothesis $H_0 : \gamma = 0$ and $H_1 : \gamma \neq 0$.

### 3.3 Time-Varying Risk Premium

Merton (1973) shows that if returns are not i.i.d., the expected return will be linked to risk through the coefficient of risk aversion. The intuition behind Merton’s result is that a more volatile market should make investors require a higher risk premium, and also that a more risk averse market would require a larger risk premium; that is, expected return and volatility should be positively correlated.

The models presented so far only focus on the volatility of shocks and therefore ignore the possibility that the expected returns are not constant. Engle, Lillian and Roberts (1987) model that by introducing conditional volatility in the mean equation, ARCH-M. GARCH-M models include predictable returns just as in Barberis (2000), but they are still pure volatility models since no exogenous variables are
used to predict returns. I use both conditional variance and conditional standard deviation as the measure for risk. The argument for the latter is that standard deviation has the same scale as returns; they are both measured as percentages in the model

\[ r_t = \mu + \delta h_{t-1}^i + \varepsilon_t, \quad \varepsilon_t \sim N \left(0, h_t^2\right), \quad i = 1, 2. \] (23)

The mean equation now takes the form of equation (23). Conditional volatility is modeled as for the constant mean models in (16), (20) or (21). Nelson (1991) points out that it can be shown that volatility in the mean might introduce some negative/positive effect between return and volatility on average. This is due to the fact that although \( h_t^2 - h_{t-1}^2 \) and \( r_t \) are conditionally uncorrelated, \( E \left| \varepsilon_t \right| r_t < 0 \right| > E \left| \varepsilon_t \right| | r_t \geq 0 \) if \( \mu \) and \( \delta \) are positive.

Before I estimate parameters and simulate expected utilities to find the optimal portfolio weights for dynamically rebalancing agents, I briefly discuss implications for buy-and-hold investors of the presented return models in the next subsection.

### 3.4 Buy-and-hold

As buy-and-hold investors cannot rebalance, the return distribution of interest for them is the one for returns aggregated over time. To see how buy-and-hold investors will be affected if returns are GARCH and not i.i.d., the results by Drost and Nijman (1993) can be used. They derive properties for finite time aggregation of some GARCH-models showing that a strong GARCH model, i.e. the shocks \( (\varepsilon_t) \) follow a specified distribution as in (15) and (23), aggregates to a weak GARCH model. In a weak GARCH model, the distribution of the shocks \( (\varepsilon_t) \) is not known but shocks fulfill the following conditions, \( P[\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, ...,] = 0 \), \( P[\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, ...,] = h_t^2 \), where \( P[\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, ...,] \) denotes the best linear prediction of \( \varepsilon_t \) given \( 1, \varepsilon_{t-1}, \varepsilon_{t-2}, \) and so forth.

If I let \( r_{m,t} \) denote returns aggregated \( m \) periods, it will follow the weak GARCH specified in (24) through (26)

\[ r_{m,t} = m\mu + \varepsilon_{m,t} \] (24)

\[ h_{m,t}^2 = \alpha_0 + \alpha_1 \varepsilon_{m,t-m}^2 + \beta_1 h_{m,t-m}^2 \] , where (25)

\[ P[\varepsilon_{m,t} | \varepsilon_{m,t-m}, \varepsilon_{m,t-2m}, \ldots] = 0 \]

\[ P[\varepsilon_{m,t}^2 | \varepsilon_{m,t-m}, \varepsilon_{m,t-2m}, \ldots] = h_{m,t}^2 \] (26)

(24) is the low frequency counterpart of (15), the main difference is that shocks are no longer normally distributed. \( m\mu \) is due to the fact that the mean grows linearly with the number of periods. Aggregation of \( r_{m,t} \) and \( \varepsilon_{m,t} \) is just the sum of their \( m \) monthly observations. Parameters in the volatility equation (25) are functions of the high frequency parameters and original process kurtosis. The constant in the volatility equation is a function of the original parameters \( \alpha_0, \alpha_1, \beta \) and the
Figure 1: Difference between GARCH(1,1) conditional variance scaled by T and conditional variance from the true T period aggregated GARCH(1,1) conditional variance.

\[ m \text{ periods of aggregation} \]

\[ \alpha_{0,m} = m \alpha_0 \frac{1 - (\alpha_1 + \beta)^m}{1 - (\alpha_1 + \beta)}. \quad (27) \]

It is obvious that \( \alpha_{0,m} \) is increasing in \( m \), since \( \alpha_1 + \beta \) is defined to be within the unit circle (19). This also means that when a large number of returns are aggregated, \( \alpha_{0,m} \) increases in proportion to the number of periods as the fraction in (27) converges to \( 1/(1 - \alpha_1 - \beta) \). The relationship between the aggregated parameters on squared shocks, lagged volatility and the original parameters is as follows

\[ \alpha_{1,m} = (\alpha_1 + \beta)^m - \beta_m. \quad (28) \]

The first term on the right hand side goes to zero as the number of aggregating periods increases and the second term is the aggregated parameter on the lagged volatility. Even if we do not know the value or function of \( \beta_m \), it is evident from (28) that conditional heteroskedasticity goes to zero. By just moving \( \beta_m \) to the left-hand side in (28) \( \alpha_{1,m} + \beta_m = (\alpha_1 + \beta)^m \), we see that the sum of the aggregated parameters approaches zero as \( m \) increases. This implies that the parameters themselves go to zero, due to the non-negativity constraints (17) and (18). Although it is clear that GARCH returns are not i.i.d., they asymptotically approach an i.i.d. process. This result implies that, at least for very long horizons, there will be no effect on portfolio weights for different length investments for buy-and-hold investors.

To show how quickly the heteroskedasticity disappears, I aggregate an estimated model. Figure 1 compares the ten-year aggregated GARCH(1,1) volatility, \( h_{m,t}^2 \), with
the scaled one-month conditional variance, $T \cdot h_T^2$. It is obvious that the $T$-period conditional variance from the GARCH model is almost constant, whereas the scaled conditional volatility is extremely volatile. This means that investors with a long horizon face a constant risk, but investors with a short horizon face a time-varying risk.

The results of Drost and Nijman (1993) indicate that for long horizons, the aggregated return process is close to i.i.d.. From Barberis (2000), we know that if returns are i.i.d. there are no horizon effects for the buy-and-and hold investor. Campbell and Viceira (2002) argue that there should be a small effect due to the non-linear characteristics of the problem, but they also argue that it is too small to show in the simulations performed by Barberis.

It is harder to say anything about the implications for buy-and-hold investors, when returns follow asymmetric GARCH processes. The aggregation properties for these models are not yet known. However, due to the results by He and Teräsvirta (1999), we know that the constant mean models described here all have zero autocorrelation. This means that the risk to return relationship remains constant under time-aggregation and only very small amounts of time-diversification could be present under buy-and-hold strategies, as this would have to come from higher-order dependencies.

When the returns are predictable, Barberis (2000) shows there to be horizon effects for buy-and-hold investors. This is very intuitive, since the relationship between the expected return and risk no longer needs to be constant; that is, the reward for bearing risk relative to risk itself can vary for different investment horizons. To find out whether the expected return grows faster or slower than risk when returns are aggregated, I use the autocorrelation structure.

For the case where $i = 2$ in (23) and conditional variance is as in (21), Hong (1991) showed the autocorrelation to be positive, equations (29)-(30)

$$\rho_1 = \frac{2\delta^2 \alpha_0 \alpha_1^2 (\alpha_1 + \beta)}{(1 - \beta^2 - 2\beta \alpha_1 - 3\alpha_1^2) (1 - \alpha_1 - \beta) + 2\delta^2 \alpha_0 \alpha_1^2}$$

$$\rho_k = (\alpha_1 + \beta)^{k-1} \rho_{k-1}. \quad (30)$$

This implies that risk accumulates more quickly than expected return, which is illustrated for two periods in example (31)

$$V(r_1 + r_2) = 2\sigma^2 (1 + \rho_1). \quad (31)$$

Buy-and-hold investors should then carry less risky assets, the longer is the investment horizon.

The autocorrelation structure for GJR-GARCH-M model is more complicated. But if I simulate the autocorrelation for the estimated models in Section 4, it is
negative, implying that buy-and-hold investors should carry more risky assets, the longer is the investment horizon because risk accumulates more slowly than expected return. Hence, it would be crucial for buy-and-hold investors to know whether conditional volatility is asymmetric or not to make the right decision.

4 Estimating return model parameters

To make my findings as realistic as possible, I estimate parameter values for the different return models I use. I use two excess return series when estimating plausible parameter values. First, I use the CRSP equally weighted index 1952-94 (post treasury accord) to capture the small firm effect that is well documented in the financial literature, e.g. Fama and French (1992). Due to the pronounced use of time diversification in the Swedish market, discussed in the introduction, I also include the Swedish value weighted index on excess returns in 1919-99, constructed by Frennberg and Hansson (1992). Summary statistics of the data series are presented in table 1.

To calculate the excess return for the US, three-month Treasury bills are used as a proxy for the “risk-free rate”. Returns are backed out from the yields, assuming a linear increase in the bill price from $t$ to $T$ to obtain the price of the three-month T-bill with two months to maturity. This approximation is judged to be adequate for the present work, mainly because the parameter values are virtually the same whether excess or just plain returns are used. All Swedish data are from Frennberg et al. (1992). Their approximation for the risk-free return consists of the one-month discount rate (1919-80) and an average of one-month Commercial Bank Certificates and one-month Treasury bills (1980-99). The reason for using the discount rate and Commercial Bank Certificates is the lack of an unbroken series of Treasury bills in Sweden.

To estimate the parameters of the respective volatility models below, I use maximum-likelihood. The iteration method utilized is the Marquardt algorithm, a modification of the Gauss-Newton method.\footnote{See Judge et. al. (1985) for an extensive presentation.} In most cases, the standardized errors are not normally distributed and hence, all inference is performed using the Quasi-

Table 1

Descriptive Statistics for US 1952-94 and Swedish 1919-99
Excess-returns and their Respective Risk-free Proxy

The table contains two US data series, excess returns on CRSP equally-weighted index together with a proxy for returns on three-month T-bills, my proxy for the American risk-free rate of return. It also includes two Swedish data series, excess returns on the Hansson-Frennberg value-weighted index together with their proxy for returns on the risk-free interest rate. The Swedish proxy for the risk-free rate of return is an average of data on one-month Swedish T-bills and Commercial Bank Certificates from 1980 to -99, and the one-month discount rate in the rest of the years. The series are called \( r_t \), where superscript denotes country, i.e. US and Swe, and subscript the type of series, ew for equally-weighted and vw for value-weighted, TB for T-bills and D for discount rate (the Swedish risk-free proxy). The estimated statistics presented are: mean \( \hat{\mu} \), standard deviation \( \hat{\sigma} \) and first autocorrelation \( \hat{\rho}_1 \), together with t-statistic.

\[
\hat{\rho}_1 = \frac{\sum_{t=2}^{T} (r_t - \bar{r})(r_{t-1} - \bar{r})}{\sum_{t=1}^{T} (r_t - \bar{r})^2}
\]

<table>
<thead>
<tr>
<th>statistic</th>
<th>( r_{US}^{ew} )</th>
<th>( r_{US}^{TB} )</th>
<th>( r_{Swe}^{vw} )</th>
<th>( r_{Swe}^{TB} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\mu} )</td>
<td>0.00813</td>
<td>0.00117</td>
<td>0.00316</td>
<td>0.00477</td>
</tr>
<tr>
<td>( \hat{\sigma} )</td>
<td>0.0491</td>
<td>0.0010</td>
<td>0.0460</td>
<td>0.0021</td>
</tr>
<tr>
<td>( \hat{\rho}_1 )</td>
<td>0.188</td>
<td>0.749</td>
<td>0.173</td>
<td>0.989*</td>
</tr>
</tbody>
</table>

* Unit root is rejected with a Philips-Peron test at the 1% level.

4.1 Constant mean models

The estimated parameters for the GARCH(1,1) render the standard results that \( \alpha_1 + \beta \) is close to 1 for both data sets as expected. It can be seen in Table 2 that the sums of \( \alpha_1 \) and \( \beta \) are .95 and .99 for US and Sweden, respectively, with \( \beta \) larger than .8 for both models. This means that there is high persistence in conditional volatility in both data sets. The main difference between the estimated models for US and Sweden is that lagged news has a larger impact on conditional volatility in Sweden, whereas persistence is lower.

According to the aggregation results of Drost and Nijman (1991) described in section 3, both pure GARCH models can be aggregated. Both models have autocor-
relation problems in the standardized residuals. But, when lagged return is included in the mean equation, the autocorrelation disappears and the parameter values of all GARCH parameters are basically unchanged.

Table 2
Volatility Models with Constant Mean
Estimated for GARCH, GJR-GARCH and EGARCH

The estimation is performed by the Marquardt algorithm. Since standardized errors do not appear to be normal, the method described in Bollerslev and Wooldridge (1992), Quasi-Maximum likelihood, is used for inference. Parameters are as defined below in the mean and volatility equations, \( r_t \) is the excess-return on the respective index and \( h_t \) is the conditional variance. Each parameter is presented with its t-statistic below.

\[
\begin{align*}
\text{mean:} & \quad r_t = \mu + \varepsilon_t, \quad \varepsilon_t \sim N(0, h_t) \\
\text{GARCH:} & \quad h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma I_{t-1} \varepsilon_{t-1}^2 + \beta h_{t-1}, \quad I_t = 1 \text{ for } \varepsilon_t < 0, \text{ else } I_t = 0 \\
\text{EGARCH:} & \quad \ln h_t = \alpha_0 + \alpha_1 \left| \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \right| + \gamma \left( \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \right) + \beta_1 \ln h_{t-1}
\end{align*}
\]

| \multicolumn{2}{c}{GARCH} | \multicolumn{2}{c}{GJR-GARCH} | \multicolumn{2}{c}{EGARCH} |
|------------------|------------------|------------------|------------------|------------------|
|                      | \( U_{S_{ew}} \) | \( S_{we_{vw}} \) | \( U_{S_{ew}} \) | \( S_{we_{vw}} \) | \( U_{S_{ew}} \) | \( S_{we_{vw}} \) |
| \( \mu \)           | .0068            | .0063            | .0069            | .0059            | .0063            | .0059            |
|                    | .0001            | .0000            | .0002            | .0000            | .0002            | .0000            |
| \( \beta \)         | .020             | .018             | .011             | .014             | .020             | .018             |
| \( \gamma \)        | .020             | .018             | .011             | .014             | .020             | .018             |
| \( \beta_1 \)       | .000             | .000             | .000             | .000             | .000             | .000             |
| \( \alpha_0 \)      | .001             | .000             | .000             | .000             | .000             | .000             |
| \( \alpha_1 \)      | .000             | .000             | .000             | .000             | .000             | .000             |

* Denote that estimation performed imposing the non-negativity constraint on \( \alpha_1 \), which is binding.

As mentioned in the introduction, it has long been noted that there seems to be a negative correlation between shocks to expected return and volatility. Therefore, it is natural to proceed and also estimate the two asymmetric volatility models presented in section 3. First, I estimate GJR-GARCH, mainly because the regular GARCH is nested under GJR-GARCH and therefore, it can directly be seen if the asymmetric parameter should be included in the model.

Table 2 shows there to be a significant asymmetric effect in the US data but not in the Swedish ones. Further, the parameter on lagged shocks in the US data, \( \alpha_1 \), is negative which, in theory, makes it possible for conditional variance to become
negative. Therefore, the non-negativity constraint is imposed in the presented estimation.\footnote{The restricted estimation is made using Ox version 2.20 (see Dornik 1999) and the G@RCH2.0 package (Laurant and Peters 2001).} The estimated model on US data is also stationary, with $\gamma/2 + \alpha_1 + \beta$ summing to .91.

With EGARCH, I do not have to restrict any parameter values to be positive to ensure non-negative conditional variance. I assume a linear relationship in the logs of conditional variance and the non-negativity condition is no longer binding. The estimated model is presented in Table 2. Once more, I cannot detect any asymmetric effect in the Swedish data, which confirms the results with GJR-GARCH. In the US data, the asymmetry parameter $\gamma$ is significant and negative as expected. Hence, a negative shock raises the logarithm of the conditional variance more than a positive shock. The condition for weak stationarity $|\beta| < 1$ is fulfilled for both estimations.

### 4.2 Volatility in Mean

When I add predictability through volatility, there are two immediate effects on estimation. First, it results in some explanatory power. For the value weighted US data, the $R^2$ is in line with the dividend yield model in Barberis (2000), 0.003. Such a low $R^2$ is common in financial research. Second, volatility in mean also reduces the problem with autocorrelation in the estimation error; in some cases it is even removed.

Estimated parameters for the mean effect with the US data and the GARCH-M model have the expected sign and size. Remember that, according to Merton (1973), the mean parameter $\delta$ is the coefficient of relative risk aversion when we use the conditional variance as the risk variable. Therefore, the estimated models in Table 3 indicate a relative risk aversion of 4-6 in the US. Although 4-6 is fairly large, it is far lower than what is implied from the equity premium puzzle, which is often as large as 40. Gyhsels, Santa-Clara and Valkanov (2005) estimate a coefficient on volatility in the mean between 1.5 and 3 using a more elaborate empirical model, although their model is not suitable for my purpose due to dimensionality\footnote{They use both high and low frequency data and allow symmetric and asymmetric effects in volatility to have different persistence.}. I do not estimate the EGARCH-M and GJR-GARCH-M models for the Swedish data series, since I did not find any evidence of asymmetry in the constant mean estimations. The parameter values in the volatility equations are about the same for GARCH-M as for GARCH, the main difference being in the model’s accuracy. As mentioned above, when I include volatility in the mean, this holds for all models, the autocorrelation problem is dramatically reduced.
Table 3

Return and Volatility Models Estimated using GARCH-M, GJR-GARCH-M, and EGARCH-M

Estimation is performed by the Marquardt algorithm. Since standardized errors do not appear to be normal, the method described in Bollerslev and Wooldridge (1992), Quasi-Maximum likelihood, is used for inference. Parameters are as defined below in the mean and volatility equations, \( r_t \) is the excess-return on the respective index and \( h_t^i \) is the conditional standard deviation when \( i = 1 \), and conditional variance when \( i = 2 \). Each parameter is presented with its t-statistic below. \( \text{Ew} \) denotes equally weighted and \( \text{vw} \) denotes value weighted index. Parameters not included in the estimations are left as blank.

\[
\begin{align*}
\text{mean:} & \quad r_t = \mu + \delta h_t^i + \varepsilon_t, \quad i = 1, 2 \\
\text{GARCH:} & \quad h_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma I_{t-1} \varepsilon_{t-1}^2 + \beta_1 h_{t-1}^2, \quad I_t = 1 \text{ for } \varepsilon_t < 0, \text{ else } I_t = 0 \\
\text{EGARCH:} & \quad \ln h_t^2 = \alpha_0 + \alpha_1 \frac{\varepsilon_{t-1}}{h_{t-1}} + \beta_1 \ln h_{t-1}^2 + \beta_2 \ln h_{t-2}^2
\end{align*}
\]

<table>
<thead>
<tr>
<th>( i )</th>
<th>( US_{ew} )</th>
<th>( S\text{we}_{ew} )</th>
<th>( US_{ew} )</th>
<th>( S\text{we}_{ew} )</th>
<th>( US_{ew} )</th>
<th>( US_{ew} )</th>
<th>( US_{ew} )</th>
<th>( US_{ew} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51</td>
<td>1999</td>
<td>95</td>
<td>1999</td>
<td>95</td>
<td>1999</td>
<td>95</td>
<td>1999</td>
</tr>
<tr>
<td>( \mu )</td>
<td>-0.0240</td>
<td>0.049</td>
<td>-0.0059</td>
<td>0.000</td>
<td>-0.0150</td>
<td>0.000</td>
<td>-0.0013</td>
<td>0.000</td>
</tr>
<tr>
<td>( \delta )</td>
<td>0.0966</td>
<td>0.012</td>
<td>-0.2010</td>
<td>0.000</td>
<td>6.262</td>
<td>0.000</td>
<td>-1.225</td>
<td>0.000</td>
</tr>
<tr>
<td>( \alpha_0 )</td>
<td>0.0002</td>
<td>0.000</td>
<td>0.0001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.0002</td>
<td>0.000</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>0.1088</td>
<td>0.000</td>
<td>0.1834</td>
<td>0.000</td>
<td>0.1057</td>
<td>0.000</td>
<td>0.1788</td>
<td>0.000</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>0.0001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0.8343</td>
<td>0.000</td>
<td>0.8091</td>
<td>0.000</td>
<td>0.8382</td>
<td>0.000</td>
<td>0.8117</td>
<td>0.000</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The estimation results for GARCH-M and GJR-GARCH-M on US data show that intercepts in the mean equation are in general negative and the coefficients on volatility in mean are positive. This is in accordance with both the discussion in Engle, Lilian and Roberts (1987) and their estimated models for excess return on bonds. The opposite is true for Swedish data, which is in accordance with findings by Glosten, Jagannathan and Runkle (1993) on US data. For both data sets, the
volatility in the mean parameter is “less” significant when using conditional variance than standard deviation.

For estimations of the EGARCH-M model, conditional volatility is not significant. I must also include volatility lagged twice in the volatility equation to model the persistence in volatility. Including an AR(1) term in the EGARCH-M equation renders a significant volatility in the mean parameter. However, since EGARCH-M requires two lagged volatility terms, I cannot use this model for simulations. Using a grid with 40 states would result in a simulation over 1,600 combinations of states, which is not possible. Thus, I focus on GARCH-M and GJR-GARCH-M when expected utility is calculated in section 5.

5 Optimal Portfolio Weights

In this section, I will investigate what sequence should optimally be chosen by an investor with the preferences specified in section 2, when he faces the return models specified in section 3. The question of interest is whether the sequence chosen by the investor contains horizon effects.

In my model, conditional volatility is the state variable determining the risk in the next period. Before optimal portfolio rules can be found by simulation, a suitable size of the state grid must be set. The number of grid points need to be large to result in simulations which are good approximations to expected utility. However, the needed computer time increases rapidly with the number of grid points. By testing some different sizes, I judge 40 grid points to be sufficient. Some oscillation in the choice of portfolio weights is still present with this choice of grid size, but it is so insignificant that it does not cloud my conclusions.

Due to the fact that the empirical conditional variances are heavily skewed, with occasional large “outliers”, I decide not to cover all realized conditional variances. The grid is set to cover 90-98% of the estimated series of conditional variances. For most models, this is within 0.0005-0.006. Thus, I obtain a state grid with a very small distance between the grid points. Unevenly spaced grids have also been tried, but with poor results; hence evenly spaced grids are used.

The simulation method for all models below are identical, the only difference is the return generating model, which will vary. Beginning with the last period, random errors are drawn from a standard normal distribution. For each state $i=1,...,40$, the error series is scaled with the conditional standard deviation (state $i$) to generate the simulated shocks to conditional mean returns. Varying the portfolio weight on the risky asset from zero to 99 % in equation (13), once more stated below as a reminder, the optimal portfolio is found by choosing $\omega$, which results in the largest
utility $Q(t, T)$

$$Q(t, T) = \max_{\omega_t} \left\{ [(1 - \omega_t) \exp(r_f) + \omega_t \exp(r_f + r_{t+1})]^{1-A} Q(t + 1, T) \right\}. $$

In most cases, I follow Barberis (2000) and draw 1,000,000 errors; he shows this to be a sufficiently large draw. With this size of the random draw, the optimal path of portfolio weights is identical when repeated. In some of the sensitivity analyses, I use a more efficient simulation method with antithetic errors, which allows me to reduce the size of the simulation to 250,000. This means that I impose the symmetry of the normal distribution. Only half the needed errors are drawn with this method, the other half just constitute the negative of the drawn errors.\(^{12}\)

![Graph showing optimal allocation to stocks when initial shock is varied, with EGARCH(1,1), A=5.](image)

Figure 2: The graph shows how demand for risky assets changes when we change the initial state $h_{t-1}$. Parameter values are for the equally weighted post Treasury accord US-data.

### 5.1 Constant Mean Models

When the impact of forecastable volatility on optimal portfolio weights is investigated, with the numerical maximization of the utility function, it turns out that portfolio weights are myopic for both symmetric and asymmetric GARCH-models.

\(^{12}\)A more elaborate description can be found in Hendry (1997), for example.
In Figure 2, the optimal portfolio weights for EGARCH(1,1) are displayed. Parameters are as from the estimation on US data. It shows there to be no evidence of time diversification. Some small initial drops are present in many of the states, but they are merely an artifact from having a discretized state variable. The Sammelson (1969) and Merton (1969) result of no horizon effects on portfolio weights when return is \textit{i.i.d.} holds true, also when the mean is constant but shocks are heteroskedastic. Figure 2 also shows the marginal change of optimal portfolio weights on the risky asset to increase as risk decreases. Thus, reducing the initial conditional variance from .0015 to .0014 yields a 4% increase in optimal portfolio weights, while reducing the initial conditional variance from .0011 to .001 yields an increase of 7%.

Why is there no hedging demand of the Merton (1973) type when risk is time-varying and forecastable? First of all, as mentioned before, I study a portfolio problem with a fixed horizon where the agent can rebalance optimally. Chacko and Viceira (2000) study the effect of shocks to expected return on the expected utility for an investor with an infinite horizon, where they do find small effects. In my setting, expected return is the same in every period, regardless of conditional volatility. Knowing how the risk will evolve over time does not affect the expected return; hence it is of no importance whether you have one or two periods left to capitalization. More risk will naturally reduce the expected utility, but it will affect the expected utility for long- and short-term investors in the same way.

Hence, both buy-and-hold and dynamically rebalancing agents will choose not to diversify over time. This is in contrast with the conjecture that agents possibly choose different sequences when return volatility is heteroskedastic. Being able to forecast risk is evidently not sufficient to increase/decrease the demand for risky assets with a longer horizon.

5.2 Volatility in Mean

When I turn to the models with both time varying mean and variance, the case should be that I get some effect from the ability to forecast returns and the size of this effect should also be dependent on the persistence of volatility. If there is very low persistence in volatility, I can only forecast returns one period ahead, which can hardly induce any horizon effects beyond that point.

First, simulations are made using no asymmetry, which is the only volatility-in-mean model I was able to fit to Swedish data. The simulation is performed for the GARCH-M estimated on the US portfolio, since it has the best fit of the models with a positive mean effect. The difference to previous simulations is that the current state, from the state grid, both scales the errors and enters the mean equation. The results are presented in Figure 3 together with the results for the GJR-GARCH-M model. It is clear that the increase in portfolio weights is marginal, in total only
5%. Most of the effect is realized in the last year, so that a GARCH-M together with power utility cannot explain the advice to investors to decrease their weight on stocks as they grow older. However, this means that with the estimated parameters in this paper, investors do choose different sequences as time goes by. If this is compared to the example in Section 2, investors choose the same sequence for the first 29 years, but then begin to reduce the investment in the risky asset in the last year before retirement.

![Graph of portfolio weights for GARCH-M and GJR-GARCH-M simulations.](image)

Figure 3: Graph of portfolio weights for GARCH-M and GJR-GARCH-M simulations. Models are estimated on the US equally weighted portfolio 1952-94. The initial state is set to .0018, the median conditional variance for the GJR-GARCH-M model. The median for GARCH-M is slightly higher, .002.

To investigate the effect of asymmetry together with conditional volatility in mean, I use the estimated GJR-GARCH-M model, with conditional standard deviation as the mean effect. From Figure 3, it is obvious that the effect is larger (12%) and persists for a longer period (2-3 years) for the GJR-GARCH-M than for GARCH-M. The sequence chosen by the investors would be the same if they had longer investment horizons than three years. With less than three years left before capitalization, they start to substitute away from risk at a moderate rate, which is increasing over time.

To say that this investigation supports investments such as holding a percentage equalling your age in bonds and the rest of your portfolio in stocks would be to stick out your neck. But it does support advice such as: if your investment horizon is longer than five years, stick to stocks. It is clear that, for example, “generation
funds” sold in Sweden cannot be justified from a pure wealth perspective. However, I have not studied the effect of labor income. It could be viewed as a substitute for the risk-free asset and could therefore result in larger time-diversification, see e.g. Campbell, Cocco and Maenhout (2001).

5.3 Sensitivity analysis

To analyze the results, I perform a sensitivity analysis by altering the parameter values and re-simulating the GJR-GARCH-M, with conditional standard deviation as the mean effect. I vary the coefficients for asymmetry and persistence in a GJR-GARCH-M model with the following parameter values: \( \mu = -.015, \delta = .5, \alpha_0 = .0002 \) and \( \alpha_1 = .0171 \). Then, I let the \( \beta \) (persistence parameter) and the \( \gamma \) (asymmetry parameter) vary, but the sum of the two parameters always equals 0.95. The result is shown in Figure 4. An interesting point revealed in the diagram, we can see that unless there is large persistence, the horizon effect dies out very quickly. When \( \beta \) is as low as 0.4, the horizon effect only lasts for about five months. As the persistence in volatility \( \beta \) increases and approaches 0.8, the horizon effect lasts for close to thirty months. This is a highly expected result, since the increased sluggishness in the forecasting variable makes it possible to forecast further into the future with some accuracy.

The reason why asymmetry is important is less clear, although the asymmetry parameter naturally adds to the sluggishness in the sense of its making the impact of a shock on future volatility larger. However, I think that the main reason for the importance of the asymmetry can be attributed to the “sharpening” of the impact of my state parameter, i.e. conditional volatility. When hit with a low return today, investors are compensated by a higher expected return tomorrow, even if asymmetric effects on volatility are absent. If volatility contains an asymmetric effect which is negatively correlated with shocks to expected returns, this will increase the expected return tomorrow by much more than a symmetric effect, after a negative shock. It can be said that with a negative correlation between shocks to expected return and volatility, investors are compensated more when this is most needed.

To investigate the nature of the hedging effect, I make an experiment where I alter the signs of the coefficients on lagged shocks to expected return \( (\varepsilon_{t-1}) \) and re-simulate using antithetic variates. To fulfill non-negativity and stationarity conditions, I must also manipulate the size of the estimated GARCH coefficients. The baseline model is GJR-GARCH(1,1), with a conditional variance in the mean. First, the coefficient on the mean effect is now set to -3.9185, i.e. -1 times the estimated coefficient. Second, I want to see how a change in asymmetry affects the optimal portfolio choice over time, thus the asymmetry parameter, \( \gamma \), is also altered. To fulfill the non-negativity constraint, the symmetric effect \( \alpha_1 \) is increased to .11636 and \( \gamma \) is reduced to .11636.
II: Optimal Rebalancing of Portfolio Weights under Time-varying Return Volatility

The effect of varying persistence and asymmetry in GJR-GARCH-M

Figure 4: In the 3D plot, I show how the allocation to stocks varies when I let the parameters vary for the GJR-GARCH-M model estimated on the US return series, $\delta = .5$ rather than the estimated .49. I have chosen to let the sum of the volatility parameters remain unchanged and equal .95, in order to obtain the “same” persistence in the simulation. Hence, if the GARCH effect, $\beta$, is reduced by .1 I increase the asymmetry parameter $\gamma$ by 0.1.

The resulting optimal paths are also smoothed by a polynomial. In Figure 5, it can be seen that with a negative mean effect, as estimated for Sweden, optimal portfolio weights are decreasing in the investment horizon. It is interesting that if I also alter the sign of the asymmetric effect, the optimal portfolio weights are increasing with the investment horizon. This indicates that when there is a risk-return relationship, i.e. the expected return depends on the level of risk, the type of time diversification is determined by the correlation between risk and return.

Throughout the simulations, it is clear that we do not have to set the coefficient of relative risk aversion to values much higher than what has been “justified” in earlier work, as mentioned in section 2. With the return models and the specific parameterization, there is no need to use coefficients of relative risk aversions larger than 6. To see how my results depend on the level of relative risk aversion, I simulate expected utility for different levels of risk aversion. Campbell and Viceira (1999) show that if investors have log-utility, that is a coefficient of relative risk equal to
one, no horizon effects on portfolio weights should be present. In Figure 6, the effect of different values on relative risk aversion is considered, where I let the coefficient vary between 2 and 9. To make the 3D graph illustrative, I adjust the intercept of the return equation, by altering the real risk-free rate, so that the allocation to the stock index with a one-month horizon is equal to 63%. This is done since this was the estimated GJR-GARCH(1,1)-M model simulated above, using a coefficient of relative risk aversion of 6, i.e. the result shown in Figure 3 earlier.

Figure 6 shows my simulations to be coherent with earlier theoretical findings that with log-utility there is no time-diversification, e.g. Campbell and Viceira (1999). As I reduce the coefficient of relative risk aversion from 9 to 2, it is clear that the horizon effects are diminishing. For $A = 2$, a small horizon effect for the first few months is present. I also find that as the coefficient of relative risk aversion is set to 9, the time-diversification seems to decrease slightly as compared to a risk aversion of 6. I believe this to be due to the fact that when the investor becomes more risk averse, he becomes very reluctant to take on any risk at all.

My sensitivity analysis shows that it is not only the correlation between expected return and risk that is crucial if investors, that are acting optimally, should choose to invest more in stock when they have a longer investment horizon. Even if more risk lowers expected returns there can be a positive hedging demand, given that
Figure 6: The 3D graph shows how the optimal portfolio weights change with the coefficient of relative risk aversion $A$, when all other parameters are as in the GJR-GARCH-M model for the US data. First, note that the risk-free rate is changed in the different simulations to make the one-month allocation equal for all values of RRA. Second, that the curves are smoothed by a three-degree polynomial to better display the change in time-diversification.

Shocks to expected return are negatively correlated to conditional risk. In this case expected returns are smoothed by the volatility process. Further, asymmetry and persistence in volatility are both crucial for the horizon effect to be of any magnitude and duration. Finally I also showed that horizon effects depend on the level of risk aversion.

6 Conclusion

Horizon effects on portfolio weights cannot be generated with the ability to forecast volatility alone, i.e. with constant mean models. The investors choose the same sequence of weights at every rebalancing point. Hence, when returns are assumed to be generated by symmetric or asymmetric GARCH models, there is no support for the strategy to decrease the share of risky assets as in the Swedish “generation funds”. This confirms the results from the $i.i.d.$ analysis made by Samuelson (1969).
and Merton (1969). However, there are large effects of shocks to volatility which have the same magnitude regardless of the investment horizon, which is the reason why there are no horizon effects.

When returns are assumed to have a positive relationship with the conditional standard deviation, as in GARCH-M models, investors do choose different sequences of portfolio weights when they get very close to retirement. This is in accordance with the second investment advice stated in the introduction, i.e. that investors should be fully invested in the risky asset if their horizon is longer than five years. The estimated persistence in the horizon effect with the GARCH in the mean model is slightly shorter, about three years. During these last three years before retirement (or the liquidation of investment for other reasons), investors substitute away from the risky asset. At first, this substitution is slow, just over 1% in the third year before retirement, but it accelerates and in the last year before retirement, the effect is 1/2% per month. This substitution away from the risky asset during the last three years before retirement adds up to about 12% in total.

My analysis shows that it is not the volatility in mean that solely explains the horizon effect, although it is the factor that triggers the horizon effect. It is the combination of asymmetry and persistence in the volatility equation that results in large effects. From comparative simulations, I show that a large coefficient on the volatility in mean only generates small effects on portfolio weights, unless it is accompanied by large coefficients on persistence and asymmetry in conditional volatility. The analysis shows that when the persistence in conditional volatility decreases, so does the horizon effect. However, if the asymmetric effect is removed, the horizon effect becomes very small, despite a large persistence in volatility.

The hedging demand for risky assets with longer investment horizons is due to the shift in expected return. When volatility is high, so is the expected return, if the coefficient on the mean effect is positive. Moreover, if we have asymmetric effects, hedging demand can be positive regardless of the sign on the coefficient of the mean effect. That is, the combination of signs of the mean effect and the asymmetry parameter in the volatility equation determines if hedging demand is positive or negative. For positive hedging demand, it is essential that there is a negative relationship between realized returns and expected returns. That is, low return today yields high expected return tomorrow.

A comparative analysis of the curvature of the utility function shows the results to be sensitive to the choice of relative risk aversion. When the coefficient of relative risk aversion increases, the horizon effect persists for a longer period. Though this effect is present only when risk aversion is reasonably low, coefficient of risk aversion range from 2-6. When the coefficient of risk aversion is raised to 9, it seems like the horizon effect declines slightly.

An interesting question not answered in this paper is how my portfolio impli-
ations would effect an equilibrium model with endogenous returns on risky assets. Changes in demography could then potentially change the expected return process so that other optimal rules would apply. This would be due to the demand shocks for risky assets when a larger share of the population desires less risky assets due to old age.
References


Engle, Robert, and Andrew Patton, (2001), “What Good is a Volatility Model?” working paper Stern School of Business NYU
Essay III
Can Parameter Uncertainty Help Solve the Home Bias Puzzle?
Published in Ekonomia volume 6 issue 2, 2003
http://ideas.repec.org/a/ekn/ekonom/v6y2003i2p187-209.html