LEARNING, DYNAMICS OF BELIEFS, AND ASSET PRICING

by

Tobias Adrian

Diplom-Volkswirt, University of Frankfurt (1995)
Maîtrise d'Économie Appliquée, University of Paris IX Dauphine (1995)
MSc Econometrics and Mathematical Economics, London School of Economics and Political Science (1998)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2003

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUN 0 9 2003
LIBRARIES

© Tobias Adrian

The author hereby grants to Massachusetts Institute of Technology permission to reproduce and
to distribute copies of this thesis document in whole or in part.

Signature of Author ....
Department of Economics
June 2003

Certified by ....
Olivier J. Blanchard
Chairman, Department of Economics
Thesis Supervisor

Certified by ....
Xavier Gabaix
Assistant Professor
Thesis Supervisor

Certified by ....
Stephen A. Ross
Franco Modigliani Professor of Financial Economics
Thesis Supervisor

Accepted by ....
Peter Temin
Departmental Committee on Graduate Studies

ARCHIVES

LEARNING, DYNAMICS OF BELIEFS, AND ASSET PRICING

by

Tobias Adrian

Submitted to the Department of Economics
on June 2003, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

In the first chapter, I study the impact of statistical arbitrage on equilibrium asset prices. Arbitrageurs have to learn about the long-run behavior of the stock price process. They condition their investment strategy on the observation of price and volume. The learning process of the statistical arbitrageurs leads to an optimal trading strategy that can be upward sloping in prices. The presence of privately informed investors makes the equilibrium price dependent on the history of trading volume. The response of prices to news is nonlinear, and little news can have large effects in some ranges of the prices.

In the second chapter, together with Francesco Franzoni, we develop an equilibrium model of learning about time-varying risk factor loadings. In the model, CAPM holds from investors' ex-ante perspective. However, positive mispricing can be observed when investors' expectations of beta are above ex-post realizations. This model is used to explain the 'value premium'. In a learning framework, the fact that value stocks used to be more risky in the past leads to investors' expectations of beta that exceed the estimates from more recent samples. We propose an empirical methodology that takes investors' expectations of the factor loadings explicitly into account when estimating betas. With the adjusted estimates of beta, we can explain the cross-section of average returns of the ten book-to-market portfolios, and account for the value premium in the relevant sample.

The third chapter investigates the role of contagion during the Great Depression. The Great Depression was a worldwide phenomenon, accompanied by financial crisis. I investigate whether financial contagion contributed to the spread of the Great Depression across countries. Contagion happens when idiosyncratic shocks are transmitted from one country to another. Asset price movements in the country affected by contagion are not justified by its own fundamentals. Contagion leads to an increase in the covariance of international financial markets during periods of financial crisis. Two particular events are tested: the stock market crash of 1929 and the Latin American debt crises of 1931. In both events the hypothesis that the crises spread contagiously is rejected with one exception: the French Stock Market.

Thesis Supervisor: Olivier J. Blanchard
Title: Chairman, Department of Economics

Thesis Supervisor: Xavier Gabaix
Title: Assistant Professor

Thesis Supervisor: Stephen A. Ross
Title: Franco Modigliani Professor of Financial Economics

Acknowledgements

First and foremost, I would like to thank my advisors Olivier Blanchard, Xavier Gabaix and Steve Ross. Their continuous encouragement and support made this thesis possible.

Olivier has always been there for me, giving cutting remarks, new perspectives, lots of moral support, smiles, and uncountable economic insights. Xavier helped polish my chapters, try out new ideas, think unconventionally, and kept me busy with hands-on research. Steve taught me most of what I know in asset pricing and inspired my enthusiasm for the field. He has always believed that I had 'something' for finance, has shared his wisdom about financial markets, and has given me brilliant ideas.

The faculty at MIT has been immensely stimulating during the lunches, seminars, lectures, and countless conversations in corridors and offices. I would especially like to thank Rudi Dornbusch for convincing me to come to MIT, reminding me continuously that research is about sitting down and writing, and being always there to talk. Sendhil Mullainathan, Dimitri Vayanos, and Jiang Wang have given me essential comments for my research at various stages. Daron Acemoglu, Marios Angeletos, John Campbell, Guido Kuersteiner, Nobu Kiyotaky, Jonathan Lewellen, Roberto Rigobon, Peter Temin, and Ivan Werning, discussed aspects of my research at different times, clarifying my thoughts, and making me think harder.

I am grateful to Daniel Gros for supervising my first research efforts as an undergraduate student, giving me my first job as an economist, and co-authoring my first papers. His thinking about economics is accompanying me in all of my research projects.

John Moore, François Ortalo-Magné, and Tony Venables from the LSE encouraged me to pursue economic research, and made it possible for me to attend MIT. Special thanks to Ronald Anderson for introducing me to financial economics.

The influence of my fellow Ph.D. students has been essential for my intellectual development. I would especially like to thank Manuel Amador, Francesco Franzoni, and Augustin Landier, for being there in many crucial moments, for working with me, stimulating my thoughts, and for their friendships. Daniel Bergstresser, Quy Toan Do, Francesco Franco, Jonathan Kearns, Malte Loos, Thomas Philippon have been with me as friends, companions, and fellow economists.

Many thanks to Stavros Pageanos, Jorge Rodriguez, and Mark Westerfield, for sharing their passion for asset pricing with me, and for always being there to talk about it.

Katherine Swan and Gary King have been extremely helpful from the first day I set foot into MIT. My students in the various classes that I have tought have given me much joy and fulfillement over the years.

I would like to thank my family and especially my parents, as well as the Hanks family, for supporting me throughout my studies.

My wife, Jennifer Vilas Hanks, has shared every day of the Ph.D. with me, with unlimited love and passion. This thesis is dedicated to her.

For Jennifer

Contents

Chapter 1

Inference and Arbitrage: The Impact of Statistical Arbitrage on Stock Prices

1.1 Introduction

Statistical arbitrageurs employ a variety of investment strategies to take advantage of mis-priced assets. The common feature of these strategies is that temporary deviations of prices from their long term level are exploited. The principal difficulty for arbitrageurs is to distinguish permanent from temporary movements in prices. In this chapter, the inference problem of statistical arbitrageurs that learn to distinguish between temporary and permanent deviations of prices is modeled explicitly.

Arbitrageurs face a trade-off between an **inference effect** and an **arbitrage effect**. When prices increase, the arbitrageurs have an incentive to sell assets, as they become more expensive given beliefs about the true value of the asset. However, a higher price also makes it more likely that the true expected payoff is high, which leads to an updating of beliefs. This is the inference effect, which makes the arbitrageurs trades upward sloping in price for a certain range of parameters and prices.

In the model, arbitrageurs trade against two classes of investors: noise traders and fundamental traders. The noise traders are causing deviations of prices from the fundamental valuation that the statistical arbitrageurs exploit. The fundamental traders have an informational advantage over the statistical arbitrageurs, but they have a short time horizon and are risk averse, so that their information is not fully revealed in the equilibrium price. Statistical arbitrageurs are assumed to be unconstrained, risk neutral and have a long-term investment horizon. They condition their trading strategies on the history of prices and trading volume. In order to assess the value stocks, arbitrageurs needs to estimate a model, i.e. need to learn from past observations of data on prices and volume. This model guides the arbitrageur in distinguishing price changes of an asset due to permanent from transitory disturbances. Arbitrageurs are assumed to have non-normal priors about the price process.

Intuitively, the trading strategy of the arbitrageurs can be upward sloping in prices for the following reason. For an intermediate range, as prices increase, statistical arbitrageurs infer that the long-run average price must be higher than previously thought and increase their asset holdings. When prices are very high or very low, a change in prices represents a buy or sell opportunity for statistical arbitrageurs. Movements in prices have different impacts on the arbitrageur's trading strategy depending on the level of prices. In the range of prices where the arbitrage effect dominates the inference effect, the statistical arbitrageurs learns a lot about the relative likelihood of the high or the low state, which makes the price move drastically. Small disturbances due to noise or fundamental information makes prices move very strongly in these ranges. When prices are very low or very high, not much is learned from new information, and the price reacts very little to either noise or news.

The absence of a simple relationship between the arrival of information and movements in prices has lead many to question the relevance of informational sources for movements in asset prices. Instead, it is often argued that noise traders or irrational speculators are causing movements without news and are mitigating the impact of new information. In the framework presented here, the stochastic structure leads to a nonlinear relationship between new information and prices that has such pricing behavior as a consequence. Little fundamental news moves prices dramatically at times, whereas big pieces of news have little impact on prices at other times.

The rest of the chapter is organized as follows. In section 1.2, the pricing in a benchmark economy in which noise trader risk is priced in equilibrium is derived. In section 1.3, the concept of statistical arbitrage is defined, and it is demonstrated that there exist statistical arbitrage opportunities for infinitely lived arbitrageurs in the benchmark economy. In section 1.4, the partial equilibrium is analyzed. In particular, the optimal investment strategy when the arbitrageur has no price impact and takes the noisy equilibrium price as given is analyzed. In section 1.5, the general equilibrium with risk-neutral, infinitively lived arbitrageurs is analyzed. The main finding is here that the equilibrium pricing function is nonlinear: at normal times, arbitrageur are stabilizing prices as their investment strategy is downward sloping in prices. However, when the inference effect dominates the arbitrage effect, the arbitrageur has an upward sloping equilibrium asset holding. This makes the equilibrium price more sensitive to noise trader shocks. Section 1.7 concludes.

1.2 The Noise Trader Economy

In order to analyze the impact of statistical arbitrageurs on the equilibrium price, we start by studying a benchmark economy without arbitrageurs. The key ingredient of this economy is that noise trader risk is priced. There are two assets in the economy, a risk free bond that pays a continuously compounded interest rate r, and that is in infinitively elastic supply. There is also a dividend paying stock with price P_t. Each stock yields dividends according to the following process:

$$dD_t = \mu dt + \sigma^D dZ_t^D$$

The drift of the dividend μ is assumed to be constant. The term Z_t^D denotes a Brownian motion. σ^D is the instantaneous standard deviation of the dividend.

In this section, there are only two types of agents in the economy: noise traders and fundamental investors. Noise traders demand the stock in an amount that evolves according to the following process:

$$du_t = \theta\left(\bar{u} - u_t\right) dt + \sigma^u dZ_t^u \tag{1.1}$$

where Z_t^u denotes a Brownian motion that is independent of Z_t^D. The noise trader demand is reverting back to the long-run mean $\bar{u}$, at rate θ.

The equilibrium price of the stock is assumed to be of the form:

$$dP_t = \eta_t dt + \sigma^P dZ_t^P$$

where $Z_t^P = \left[Z_t^u, Z_t^D\right]$ and $\sigma^P = \left[\sigma^{uP}, \sigma^{DP}\right]'$. The drift and variance of the stock price process is determined in equilibrium.

The other agents in the economy are fundamental investors. In each period, a new generation of investors is born, earns a non-random labor income, invests the labor income according to CARA utility with coefficient α, and consumes when old. Fundamental investors behave like instantaneous overlapping generations. The demand of these investors is:

$$y_t = \frac{\eta_t + D_t - rP_t}{r\alpha\sigma^{P2}} \tag{1.2}$$

The demand function is the standard demand of a CARA investor when the price process has constant drift, as demonstrated in Merton (1990). When the drift η_t is time varying and the investor is long-lived, the demand of equation (1.2) is not optimal, as a long-lived investor would hedge the time-variation in the drift rate. The assumption that fundamental investors are only instantaneously lived gives the demand of equation (1.2).

The equilibrium price in the economy is such that the market for the asset clears at all times. In the benchmark economy of this section, there are no arbitrageurs present. In particular, the equilibrium price is such that investors demand y_t plus the noise traders demand u_t equal total supply S:

$$y_t + u_t = S \tag{1.3}$$

From this market clearing condition, it follows that the pricing function is linear in D_t and u_t. The derivation of the pricing function is in the appendix. The equilibrium price when $\sigma^D > 0$ is:

Proposition 1.1

$$P_t = \frac{D_t}{r} + \frac{\mu}{r^2} - \alpha \left(\sigma^P\right)^2 (S - \bar{u}) + (\kappa/\theta)(u_t - \bar{u}) \tag{1.4}$$

where

$$\begin{aligned}
\sigma^{PD} &= \sigma^D/r \\
\sigma^{Pu} &= \frac{r+\theta}{r\alpha\sigma^u}\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 + \left(2\sigma^D/r\right)^2}\right) \\
\left(\sigma^P\right)^2 &= \left(\sigma^{PD}\right)^2 + \left(\sigma^{Pu}\right)^2 \\
\kappa &= r\alpha\left(\sigma^P\right)^2 (r+\theta)^{-1}\theta
\end{aligned}$$

In the limit when $\sigma^D = 0$, there are two equilibria. One is the limiting case of the noise trading equilibrium in the preposition as $\sigma^D \to 0$, the other equilibrium is the "fundamental" one with $P_t = r^{-1}D_t + r^{-2}\mu$. These two equilibria correspond to the two equilibria from DeLong, Shleifer, Summers, Waldmann (1990a). When the dividend is risky $\left(\sigma^D > 0\right)$, only the equilibrium with noise trader risk is possible.

Taking the derivative of the pricing function gives:

$$dP_t = \kappa\left(\bar{p} - (u_t - \bar{u})\right)dt + \sigma^P dZ^P$$

where $\bar{p} = \mu r^{-1}\kappa^{-1}$. The noise trader risk leads to time-varying drift of the equilibrium price process. When noise trader demand is high, it is pushing up prices contemporaneously. However, it is expected to lead to lower prices in the future, as noise trader demand is mean reverting. The long-run average level of the price is $\bar{p}$. The average price is governed by two parameters: μ and $\bar{u}$. The inference problem of the statistical arbitrageurs will be assumed to be about these two parameters.

In this section, only the fundamental investors and noise traders hold the asset. Taking the pricing function of this economy as given, the next section shows that there are statistical arbitrage opportunities in equilibrium. In the next section, all parameters are assumed to be observable to the statistical arbitrageurs. The inference problem is introduced in section 1.4.

1.3 Statistical Arbitrage

The equilibrium price of the benchmark economy in the previous section has a time varying drift rate. In this section, I define statistical arbitrage, and show that the benchmark economy has statistical arbitrage opportunities.

Statistical arbitrageurs are assumed to have an infinite time horizon, and to be risk neutral. The dynamic budget constraint of statistical arbitrageurs is:

$$dW_t = r\left(W_t - A_t P_t\right) dt + A_t dP_t + A_t D_t dt \tag{1.5}$$

where A_t denotes the trading strategy of the arbitrageur, and W_t the wealth at time t of the arbitrageurs. The definition of a statistical arbitrage is in terms of the discounted logarithm of wealth. with the following notation

$$\tilde{w}_t = \ln\left(e^{-rt} W_t\right) - \ln\left(W_0\right)$$

a statistical arbitrage is defined as:

Definition 1.1 *A statistical arbitrage is an investment strategy* $\{A_t\}_{t=0}^{\infty}$ *that satisfies the dynamic budget constraint 1.5 and the following conditions:*

1) Zero initial cost

$$\tilde{w}_t = 0$$

2) The arbitrage portfolio $\{A_t\}$ *generates a path of wealth* W_t *such that the expected cumulative payoff converges to a positive limit:*

$$\lim_{t \to \infty} E_0 \left[\tilde{w}_t \right] > 0$$

3) The expected cumulative payoff is non-negative a.s.:

$$\lim_{t \to \infty} \Pr \left[\tilde{w}_t < 0 \right] = 0$$

This definition of a statistical arbitrage is an extension to infinite horizon of the standard definition of an arbitrage. It is adopted from Hogan, Jarrow, and Warachka (2002). Note the similarity to the APT of Ross (1976). In the APT, asymptotic arbitrage is defined in the cross-section of stocks as the number of stocks asymptotes to infinity. In the definition of statistical arbitrage, arbitrage refers to the time-dimension of each stock as time asymptotes to infinity.

In the benchmark economy of the previous section, there exist statistical arbitrage opportunities. Consider the following trading strategy:

$$A_t = W_t A t^q \left(\bar{u} - u_t \right) \tag{1.6}$$

where q is a constant $q \geq 0$ and A is a positive constant. This trading strategy is contrarian: as noise trader demand pushes up prices (u_t is high), the drift rate of the pice is expected to decline. The trading strategy is proportional to wealth: as arbitrageurs accumulate more wealth, they trade more aggressively. I further allow that the size of the trading position grows with time at rate q. Using the trading strategy defined in equation 1.6, I can show that there exist statistical arbitrage opportunities in the benchmark economy of the previous section:

Proposition 1.2 *There exist statistical arbitrage opportunities in the benchmark economy with the pricing function from proposition 1.1 if and only if the following condition 1 holds:*

$$r\alpha \left(\sigma^P\right)^2 \sigma^{u2} > 2\theta \sigma^u \sigma^P \sigma^{Pu}.$$

In the proof of the proposition, I show that the cumulative value of the trading strategy evolves according to:

$$\tilde{w}_t = \int_0^t \lambda_{1s} \left(u_s - \bar{u}\right) ds + \int_0^t \lambda_{2s} \left(u_s - \bar{u}\right)^2 ds + \sigma^P \int_0^t \lambda_{3s} \left(u_s - \bar{u}\right) dZ_s^P$$

where λ_{is} are functions of t^q. These λ_{is} for $i = 1, 2, 3$ are specified in the proof of the proposition in the appendix. There are three different terms determining the value of the arbitrage strategy. The first term is proportional to $(u_t - \bar{u})$. This term originates from the noise trader risk that the arbitrageurs have to bear by holding the asset. The second term is proportional to $(u_t - \bar{u})^2$. This term originates from the contrarian nature of the trading strategy of the arbitrageurs: arbitrageurs buy when the price is high relative to the dividend process, and sell when the price is low relative to the dividend process. The further the noise traders drive the drive away from the no-noise fundamental price, the more do the arbitrageurs profit from this disturbance. The third term is the stochastic term. The coefficient λ_{3t} is actually negative: the trading strategy is negatively correlated with the price process, which again is from the contrarian nature of the trading strategy.

The expectation of the trading strategy is stochastically equal to:

$$E_0\left[\tilde{w}_t\right] = O\left(\frac{r\alpha\left(\sigma^P\right)^2\sigma^{u2} - 2\theta\sigma^u\sigma^P\sigma^{Pu}}{2\theta\left(q+1\right)}At^{q+1}\right)$$

The expected cumulative trading profit is of the order t^{q+1}, which shows that condition 2.) of the definition is satisfied if $r\alpha\left(\sigma^P\right)^2\sigma^{u2} > 2\theta\sigma^u\sigma^P\sigma^{Pu}$. This condition is satisfied when the risk aversion of the fundamental investors is large enough, so that noise trader risk is priced sufficiently high.

The variance of the trading strategy is of the order:

$$Var_0\left[\tilde{w}_t\right] = O\left(\frac{\sigma^{u2}\kappa^2\bar{u}^2}{2\theta\left(2q+1\right)}A^2t^{2q+1}\right)$$

Due to the normality of Brownian motions, what determines the limit of the probability of the trading strategy is:

$$\lim_{t\to\infty} E_0\left[\tilde{w}_t\right] / \sqrt{Var_0\left[\tilde{w}_t\right]}$$

The order of convergence of the probability is therefore $t^{.5}$. Note that the order of convergence is independent of q, in particular, it holds for $q = 0$ that the trading strategy 1.6 is asymptotically riskless. The details of the proof are in the appendix.

If there are statistical arbitrageurs in the economy, the pricing function from proposition

(1.1) cannot be an equilibrium. The arbitrageurs could scale the trading strategy from equation (1.6) arbitrarily, and make profits in the limit for sure. In the next section, an informational problem for the arbitrageurs is introduced, and then the equilibrium properties in the economy with arbitrageurs are derived in section 1.5.

1.4 Inference and Arbitrage in Partial Equilibrium

The previous section demonstrated that there are statistical arbitrage opportunities in the benchmark noise trader economy. This section is concerned with the optimal investment strategy of arbitrageurs, when the arbitrageur faces parameter uncertainty. The idea is that the arbitrageur is unsure whether buying or selling a stock is the optimal investment strategy as the long-run mean of the noise-trader risk is unknown. For certain levels of noise trader demand, the arbitrageur is unsure whether going long or short in the stock is optimal as it is unknown whether the deviation of price from the "fundamental" level is caused by temporary or permanent disturbance, i.e. whether it is caused by u_t or by $\bar{p} + \bar{u}$.

The trading strategy depends on the knowledge of $\bar{u}$:

$$A_t = W_t A \left(\bar{u} - u_t\right)$$

It is now assumed that the arbitrageurs do not know the long run average level of the price, that they are uncertain about $\bar{p}+\bar{u}$. Recall that $\bar{p} = r^{-1}\kappa^{-1}\mu$. I make the following assumptions about the priors of these values:

$$\Pr(\mu = \bar{D}) = \pi_0^D \qquad \Pr(\mu = 0) = 1 - \pi_0^D$$

$$\Pr(\bar{u} = \bar{U}) = \pi_0^u \qquad \Pr(\bar{u} = 0) = 1 - \pi_0^u$$

When the true level of $\bar{u}$ is unknown, the statistical arbitrageur cannot trade the strategy from the previous section. I assume that the long-run average of the noise can take 2 values, 0 or $\bar{U}$. Under this assumption, there is a range of noise in which the arbitrageur is unsure about the arbitrage:

$$0 < u_t < \bar{U}$$

When u_t is between these two critical values, the arbitrageur cannot not be sure whether the observed price level is too high or too low. When $u_t < 0$, the arbitrageur is sure that noise trader demand is unusually low, and the asset must be undervalued. When $u_t > \bar{U}$, the arbitrageur knows for sure that the stock is overpriced due to the noise-trader risk. In between these two boundaries, the arbitrageur must make a best guess as to what the appropriate investment strategy is.

Denoting the expectation of u_t conditional on the history of prices by m_t, and conditional on $\bar{u}$, the optimal filter for the noise is:

$$dm_t = \theta\left(\bar{u} - m_t\right)dt + \sigma_t^m dZ_t^m$$

The definition of $\sigma_t^m dZ_t^m$ is given in the appendix. The likelihood function of $\bar{u}$, conditional on the history $m^t = \{m_\tau : 0 \leq \tau \leq t\}$ can be computed from Küchler and Sørensen (1997):

$$\Pr\left(\bar{u} = \bar{U}|m^t\right) = \exp\left(\frac{\theta\left(m_0 - \bar{u}\right)^2}{2\sigma_0^{m2}} - \frac{\theta\left(m_t - \bar{u}\right)^2}{2\sigma_t^{m2}} - \frac{1}{2}\theta^2\int\frac{\left(u_s - \bar{u}\right)^2}{\sigma_s^{u2}}ds + \frac{1}{2}\theta t\right)$$

The Radon-Nikodym derivative of changing the measure of m_t from $\bar{u} = \bar{U}$ to $\bar{u} = \bar{U}$ is then:

$$\phi_t = \frac{\Pr\left(\bar{u} = \bar{U}|m^t\right)}{\Pr\left(\bar{u} = 0|m^t\right)} = \exp\left[\theta\bar{U}\left(\frac{m_t}{\sigma_t^{m2}} - \frac{m_0}{\sigma_0^{m2}} + \theta\int_0^t \frac{m_s - \bar{U}/2}{\sigma_s^{u2}}ds + \frac{\sigma_t^{m2} - \sigma_0^{m2}}{2\sigma_t^{m2}\sigma_0^{m2}}\bar{U}\right)\right] \quad (1.7)$$

There are two important elements in this expression.

The probability of the high state, conditional on the estimate of noise trader shocks is then according to Bayes rule:

$$\pi_t^u = \frac{\pi_0^u\phi_t}{\pi_0^u\phi_t + 1 - \pi_0^u}$$

The evolution of π can be found by taking the time derivative ofπ:

$$d\pi_t^u = \pi_t^u\left(1 - \pi_t^u\right)\frac{\theta\bar{U}}{\sigma_t^{m2}}\left(\theta\left(\bar{u} - \bar{U}\pi_t^u\right)dt + \sigma_t^m dZ_t^m\right)$$

The discounted wealth from buying one additional unit of the asset and holding it forever is:

$$E_t\left[e^{-r(T-t)}W_T\right] \to \frac{S - \bar{U}\pi_t^u}{r} + \frac{\bar{U}\pi_t^u - m_t}{r+\theta} \quad \text{as } T \to \infty$$

The important feature of figure 1-1 is that the sensitivity of expected discounted wealth with respect to the estimated level of noise trader demand is different according to the level of estimated noise trader demand m_t. For very high and very low levels of estimated noise trader demand, the expected discounted wealth level is not very sensitive to changes in the level of noise. However, in an intermediate range, the arbitrageur learns strongly about the profitability of an arbitrage opportunity, and the expected discounted wealth is very sensitive to changes in the estimated level of noise.

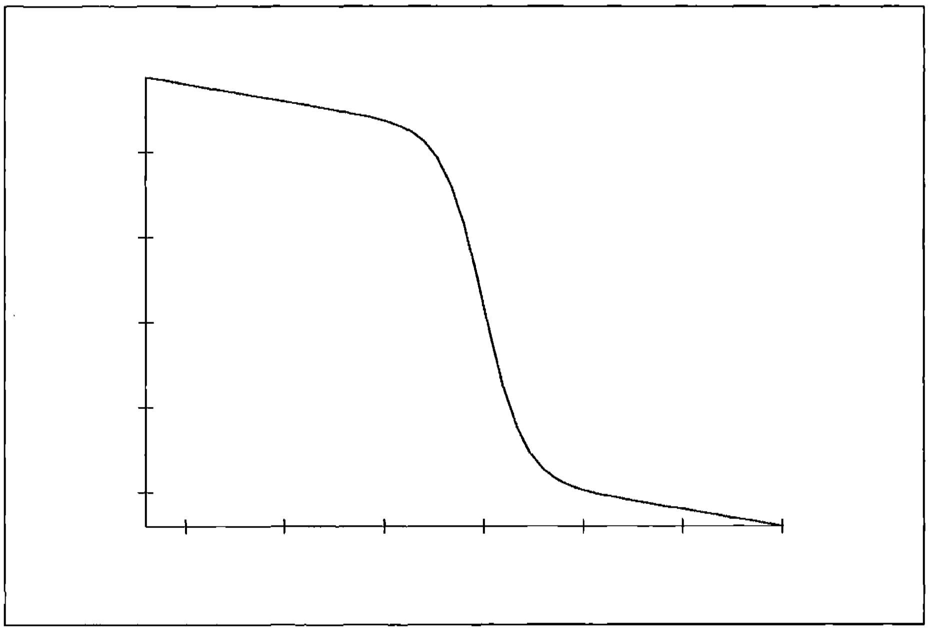


Figure 1-1: Expected Wealth from Arbitrage

The feature of the learning process that is not visible from figure 1-1 is that the location of the kink in the expected wealth depends on the history of the estimated noise shocks m. This feature can be seen from the likelihood function 1.7. A history of positive shocks to m makes it more likely that the high level of average noise trader demand is the true one, i.e. that the critical value of the price at which going long in the stock provides an arbitrage opportunity is actually lower.

The filtering problem as it is derived has the following property. Assuming that statistical arbitrageurs only conditions on the history of the price, it can be verified that learning about the long-run mean of noise $\bar{u}$ is the same learning problem as learning about the drift of the dividend μ, up to a linear transformation. In particular, application of theorem 12.7. of Liptser and Shiryaev (2000), assuming that μ instead of $\bar{u}$ is unknown leads to the following property:

$$E\left[\bar{u}|F_t^P\right] = \frac{r+\theta}{\theta\alpha r^2\sigma^{P2}} E\left[\mu|F_t^P\right]$$

where F_t^P denotes the filtration generated by the history of prices up to time t. As long as the statistical arbitrageur only learns from the price, it is thus indistinguishable whether the learning is about the drift of the dividend, or the long-run average of the noise trader demand. In this partial equilibrium analysis, both of these constants change the long-run level of the price, up to the multiplication by $\frac{r+\theta}{\theta\alpha r^2\sigma^{P2}}$.

If instead of learning from price, the arbitrageurs learn from volume, we get the following learning mechanism. As a proxy for volume, define the variable v_t as:

$$v_t = y_t + u_t$$

Using Itô's lemma gives the following process for v:

$$dv_t = \theta\left(\bar{u} - v_t + \frac{1}{2\theta}\sigma^{u2}\right)dt + \sigma^u dZ_t^u$$

Learning about the profitability of an arbitrage opportunity from volume is similar to learning from price. Applying again theorem 12.7 from Liptser and Shiryaev (2000), the expectation of $\bar{u} = \bar{U}$, conditional on the history of v_t is

$$d\pi_t^v = \pi_t^v\left(1 - \pi_t^v\right)\frac{\theta\bar{U}}{\sigma^{u2}}\left[\left(\bar{u} - \pi_t^v\bar{U}\right)dt + \sigma^u dZ_t^u\right]$$

Learning from volume inherits many of the properties of learning from price. In particular, the likelihood function displays a similar nonlinearity as the one in the case of learning from the price, and the critical condition for expected wealth has a similar shape as the one displayed in figure 1-1.

Having introduced the inference problem into the noise trader economy in partial equilibrium, I will study the impact of statistical arbitrageurs on prices in general equilibrium.

1.5 The Impact of Arbitrageurs in General Equilibrium

Having studied the investment strategy of the arbitrageurs in partial equilibrium when the noise trader risk is priced, I will now study the general equilibrium. The key trade-off between inference and arbitrage that causes the expected discounted wealth from holding a stock to have a nonlinear form will translate into a nonlinear price. It is shown that the economics behind this nonlinearity are the stock holdings of the arbitrageur. The equilibrium asset holdings of arbitrageurs are determined by 2 opposing forces, the inference and the arbitrage effect. When the inference effect dominates the arbitrage effect, the equilibrium asset holdings of the arbitrageur are upward sloping in prices.

With free entry of arbitrageurs, the equilibrium price must be such that the expected payoff from a zero wealth investment yields an expected discounted payoff of 0 asymptotically:

$$\lim_{T\to\infty} E_t\left[e^{-rT}W_T\right] = 0$$

Following Santos and Woodford (1997), I also impose the transversality condition in order to rule out rational bubbles as in Blanchard and Watson (1982):

$$\lim_{T\to\infty} E_t\left[e^{-rT}P_T\right] = 0$$

The equilibrium price is then the discounted future payoff of dividends:

$$P_t = E_t\left[\int_t^\infty e^{-rs}D_s ds\right]$$

I assume that the arbitrageurs face uncertainty about the drift of the dividend payoff. Recall that the drift can take two values, $\mu = \left\{0, \bar{D}\right\}$. The equilibrium price is then:

$$P_t = \frac{D_t}{r} + \frac{E_t\left[\mu\right]}{r^2}$$

In general equilibrium, the statistical arbitrageurs set the price, so that there is no information in the price that is of value to the arbitrageurs. Learning about the drift of the dividend is like learning about the right level of the price. With a drift of $\bar{u} = 0$, the correct level of the price is D_t/r. With a drift of $\bar{D}$, the correct level of the price is $D_t/r + \bar{D}/r^2$.

I assume that the short-horizon investors know the drift μ of the dividend. In order to make the model tractable, I assume that the noise trader process is still given by 1.1, but that the u now refers to the belief of the noise traders about the true level of the drift of the dividend.[1] In particular, the drift of the price process is $r^{-1}\mu + r^{-2}E_t\left(d\pi_t^D\right)$. Whereas informed investors know μ, the noise traders have stochastic belief about μ.

[1]The pricing model of the previous sections could be rewritten such that noise traders have the static CARA demand functions of equation (1.2). All of the previous results would remain unchanged.

I will start by analyzing the economy when the arbitrageurs can observe the demand schedule submitted by the noise traders plus the informed traders. This is a similar set-up to Kyle (1985), with a different stochastic structure of the shocks: whereas Kyle assumes that μ is normally distributed, it is assumed here that μ has a bimodal distribution.

I will show that the statistic that the arbitrageurs can infer from observing the demand of the rest of the market is:

$$x_t = \mu + u_t$$

From the specification of the noise trader process 1.1, the process for x is therefore:

$$\begin{aligned} dx_t &= \theta\left(\bar{x} - x_t\right) dt + \sigma^u dZ^u \\ \bar{x} &= \bar{u} + \mu \end{aligned}$$

Furthermore, corresponding to the different drifts μ, the variables $\bar{X} = \bar{u} + \bar{D}$, and $X = \bar{u}$ are defined. From Küchler and Sørensen (1997), the likelihood function of x_t conditional on $\bar{x}$ and a path of $x^t = \{x_\tau : 0 \leqslant \tau \leqslant t\}$ is:

$$f\left(x_t | x^t, \bar{x}\right) = \exp \frac{\theta}{2\sigma_u^2}\left(-\left(x_t - \bar{x}\right)^2 + \left(x_0 - \bar{x}\right)^2 - \theta \int_0^t \left(x_s - \bar{x}\right)^2 ds + t/\sigma_u^2\right)$$

from this likelihood function, the Radon-Nikodym derivatives of changing the measures from $\bar{x} = \bar{X}$ to $\bar{x} = X$ can be computed. I normalize $X = 0$:

$$\phi_t = \exp \frac{\theta \bar{X}}{\sigma^{u2}}\left[x_t - x_0 + \theta \int_0^t \left(x_s - \bar{x}/2\right) ds\right]$$

The relative likelihood of $\bar{x} = \bar{X}$ versus $\bar{x} = 0$ therefore depends on two main statistics. First, the current level of x_t minus the initial level x_0. This is a level effect: x_t is likely to be higher when the mean reversion level of x is higher. The second term is depending on the history of realizations of x_t. In particular, the average time spend above the long-run mean $\bar{x}$.

The Bayesian estimate of $\bar{x}$ is then:

$$\Pr\left[\bar{x} = \bar{X}|x^t\right] = \pi_t = \frac{\phi_t}{1 - 1/\pi_0 + \phi_t}$$

where $x^t = \{x_\tau : 0 \leqslant \tau \leqslant t\}$. I normalize the $\bar{u} + \mu$ such that $X = 0$. Then, the law of motion of π is:

$$d\pi_t = \pi_t (1 - \pi_t)\, \theta \bar{X} \left(\sigma^u\right)^{-2} \left[\left(\bar{x} - \bar{X}\pi\right) dt + \sigma^u dZ^u\right]$$

In terms of the observation of the x_t, this can be rewritten as follows:

$$d\pi_t = \pi_t (1 - \pi_t)\, \theta \bar{D} \left(\sigma^u\right)^{-2} \left[dx_t + \theta\left(\bar{X}\pi - x_t\right) dt\right]$$

Under the arbitrageurs information set, the process of π is a martingale. However, under the investors information set, as investors know $\bar{X}$, the process of π has a positive drift if $\bar{x} = \bar{X}$ and has a negative drift if $\bar{x} = 0$.

Using the notation

$$dP_t = \eta_t^P dt + \sigma_t^P dZ$$

and taking the derivative of the pricing function we find:

$$\begin{aligned}\eta_t^P &= r^{-1}\mu + r^{-2}\pi_t(1-\pi_t)\theta\bar{X}(\sigma^u)^{-2}(\bar{x}-\bar{X}\pi) \\ \left(\sigma_t^P\right)^2 &= r^{-2}\left[\sigma_D^2 + r^{-2}\pi_t^2(1-\pi_t)^2\theta^2\bar{X}^2(\sigma^u)^{-2}\right]\end{aligned}$$

The demand from the investors is then:

$$\frac{\eta_t + D_t - rP_t}{r\alpha\left(\sigma_t^P\right)^2} = \frac{r^{-1}\left(\mu - \bar{D}\right)\pi + r^{-2}\pi_t(1-\pi_t)\frac{\theta\bar{D}}{\sigma^{u2}}\left(\mu - \bar{D}\pi\right)}{r\alpha\left(\sigma_t^P\right)^2} = \frac{\mu - \bar{D}\pi}{r\alpha\tilde{\sigma}_t^2}$$

where

$$\tilde{\sigma}_t^2 = \frac{\left[r^{-1}\sigma_D + r^{-2}\pi_t(1-\pi_t)\theta\bar{X}(\sigma_u)^{-1}\right]^2}{\left[r^{-1} + r^{-2}\pi_t(1-\pi_t)\theta\bar{X}(\sigma^u)^{-2}\right]}$$

The demand of the noise traders together with the informed investors is then:

$$\frac{x_t - \bar{X}\pi_t}{r\alpha\tilde{\sigma}_t^2}$$

As the arbitrageurs set the price, their holding is the supply S minus the demand of the noise traders plus informed investors. The asset holdings are represented in figure 1-3 as a function of price. The equilibriumprice as a function of x is depicted in figure 1-2. The key feature is that the equilibrium asset holdings of the arbitrageur are downward sloping for high and low levels of the price, and upward sloping for intermediate levels. The upward sloping portion of the investment strategy corresponds to the relatively steep section of the pricing function (see figure 1-3). The intuition is that there is an inference and an arbitrage effect. When prices are very high or very low, the arbitrageur is a contrarian investor. However, for an intermediate range, the arbitrageur learns from the information that is contained in prices: as prices increase, the arbitrageur increases equilibrium asset holdings, as the arbitrageur updates his beliefs about the long-run level of prices.

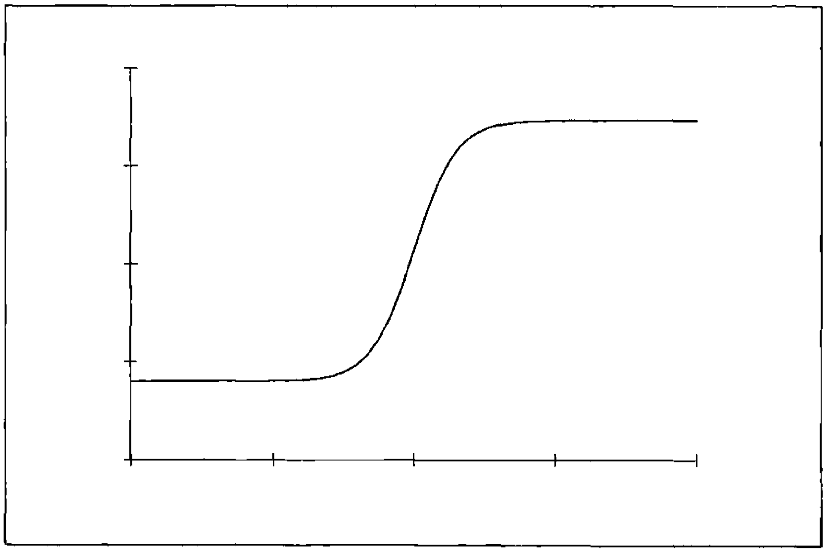


Figure 1-2: Equilibrium Price P(x)

1.6 Related Literature

Recently, a number of authors have studied the impact of arbitrageurs, hedge funds, or convergence traders on equilibrium prices. Xiong (2001) studies convergence traders that are wealth constrained. When prices drop sharply, the arbitrageurs' wealth drops, which can amplify drops in prices. In this chapter, arbitrageurs are unconstrained and risk-neutral, and the amplification effect is due to learning. Gromb and Vayanos (2002) investigate the welfare implications of margin requirements in segmented markets.[2] They show that financial constraints can lead

[2] Liu and Longstaff (2000) also study the effect of margin constraints on arbitrage.

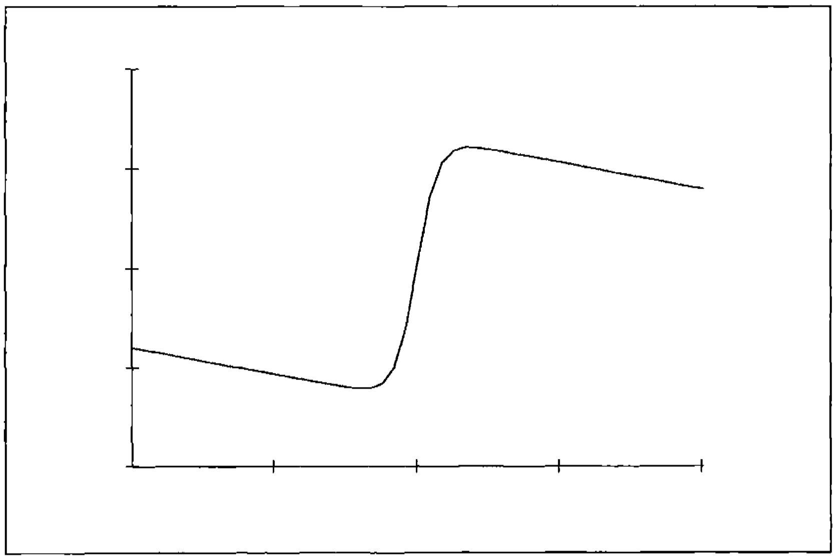


Figure 1-3: The Arbitrageur's Stock Holdings

to too much or too little risk taking behavior. In the (unconstrained) first-best of the model considered here, the absence of private information would always lead to less volatile prices, and the equilibrium is constrained efficient. Abreu and Brunnermeier (2002) study the coordination problem of arbitrageurs in reaction to a bubble. Even though rational arbitrageurs know that there is a bubble in asset prices, they do not know when the other arbitrageurs will start to trade against the market. Once coordination happens, the bubble crashes.[3] [4]

My model is driven by asymmetric information between the risk-neutral arbitrageurs and informed, risk-averse investors. The focus of this chapter is on the relationship between fundamental news and prices, and volume and prices. Romer (1993) also studies price movements in an asymmetric information setting. He shows that information aggregation can lead to a

[3]Goetzmann, Ingersoll, Ross (1998) study the incentives provided by high water mark provisions in the hedge fund industry.

[4]For a description of the technology bubble see Ofek and Richardson (2002). For a systematic account of extreme movements in stock prices see Gabaix et al. (2002 a,b).

situation where information is suddenly revealed, even though no new fundamental information was revealed. Gennotte and Leland (1990) show in an asymmetric information model that the demand function for assets can become backward sloping, so that crashes can occur. In the model presented here, the demand function is never backward sloping, but it is nonlinear. The model in this chapter is not a model of crashes, but one of amplification. The results on the relationship between volume and prices are an extension of results obtained by Brown and Jennings (1989), Grundy and McNichols (1989). Both of these papers study technical analysis in an asymmetric information framework, extending the intuitions outlined in Treynor and Ferguson (1985).

The nonlinearity that arises from the bimodal distribution is similar to the one studied by Veronesi (1999), where the (unobserved) drift of the dividend jumps between a high and a low state. This leads to an amplification mechanisms similar to the one studied here. The main difference is that hedge funds in the model studied here face asymmetric information, and that gives rise to interesting implications about volume.[5]

Limits to arbitrage due to irrationality are studied by Shleifer and Vishny (1997). They point out that arbitrageurs might have the largest arbitrage opportunities at times when they are the most constrained, as investors withdraw their capital when arbitrageurs lose money from a position.

Recent empirical literature on hedge funds include Brunnermeier and Nagel (2002), Baker and Savasoglu (2002), Brown and Goetzmann (2001), Brown, Goetzmann and Park (1998), Wurgler and Zhuravskaya (2002), Mitchell and Pulvino (2002), Fung and Hsieh (2000, 2001).

[5] Nonlinearities are very important for empirical work. Boudoukh et al. (2001) find that a nonlinear regression can explain a substantial amount of the variation of FCOJ prices uncovered by Roll (1984).

1.7 Conclusion

In this chapter, I analyze a model with statistical arbitrageurs that are risk-neutral, unconstrained, and have a long term investment horizon. The arbitrageur's inference problem is that it is not known whether the unconditional payoff of the trading strategy is zero or positive. The key assumption is that there are only two states: in the partial equilibrium, it is assumed that arbitrageurs do not know the long-run average level of the price, wich can be either 0 or positive. In general equilibrium, it is demonstrated that only the drift of the dividend process matters for pricing, this drift could be 0 or positive. This twin-peaked distribution leads to interesting implications for the pricing function, and illustrates the arbitrageurs fundamental trade-off. Throughout the chapter, it is assumed that the arbitragers only learn from the observation of prices and volume, not from the observation of dividends.

The arbitrageurs face a trade-off between inference and arbitrage: low demand for the stock means that prices are driven lower, relative to fundamental valuation. However, the arbitrageur can learn from demand shocks as well. In a certain range of prices, this learning is dominating the arbitrage effect, and the arbitrageur has upward sloping stock holdings. Depending on the relative strength of the inference and the arbitrage effect, arbitrageurs thus have either downward sloping or upward sloping asset holdings as a function of price.

The history of demand matters for prices. Prices can drop (increase) sharply when enough selling (buying) pressure has accumulated in the market. A situation is possible where the arbitrageur absorbs all the selling from the market, and prices do not change. Suddenly, the arbitrageur realizes that the cumulative sales from the rest of the market must be very informative, the arbitrageur starts to sell heavily, and prices drop sharply.

1.8 Appendix

1.8.1 Equilibrium Price in the Noise trader Economy

Proof. Proposition (1.1) : Let's start with the guess that the drift of the price is a linear function of the noise trader demand::

$$\eta_t = C_0 + C_1 u_t$$

From the market clearing condition 1.3, we find:

$$P_t = r^{-1}\left(C_0 - r\alpha\sigma^{P2}S\right) + r^{-1}D_t + r^{-1}\left(C_1 + r\alpha\sigma^{P2}\right)u_t \tag{1.8}$$

Differentiating and replacing for the dividend and the noise trader process gives:

$$dP_t = r^{-1}\left(C_1 + r\alpha\sigma^{P2}\right)\theta\left(\bar{u} - u_t\right)dt + r^{-1}\mu dt + r^{-1}\sigma^D dZ^D + r^{-1}\left(C_1 + r\alpha\sigma^{P2}\right)\sigma^u dZ^u$$

Matching coefficients for the drift term:

$$C_0 = r^{-1}\mu + \tfrac{r\alpha\theta}{r+\theta}\left(\sigma^P\right)^2\bar{u} \quad C_1 = -\tfrac{r\alpha\theta}{r+\theta}\left(\sigma^P\right)^2$$

Replacing back into 1.8 gives the pricing function:

$$P_t = \frac{D_t}{r} + \frac{\mu}{r^2} + \alpha\sigma^{P2}\left[\frac{\theta}{r+\theta}\left(\bar{u} - u_t\right) + \left(u_t - S\right)\right] \tag{1.9}$$

Differentiating the pricing function 1.9, and matching the volatility gives the following equation:

$$0 = \left(\sigma^P\right)^4 \left(\frac{r\alpha}{r+\theta}\right)^2 (\sigma^u)^2 - \left(\sigma^P\right)^2 + \left(\frac{1}{r}\right)^2 \left(\sigma^D\right)^2$$

Using the independence between innovations to D and innovations to u that implies that $\left(\sigma^{Pu}\right)^2 + \left(\sigma^{PD}\right)^2 = \left(\sigma^P\right)^2$ we get:

$$\begin{aligned} \sigma^{PD} &= \sigma^D / r \\ \sigma^{Pu} &= \frac{r+\theta}{r\alpha\sigma^u}\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 + \left(2\sigma^D/r\right)^2}\right) \end{aligned}$$

Which concludes the proof of the proposition. ■

1.8.2 Proof of Statistical Arbitrage

Proof. Proposition (1.2) : Recall that the equilibrium pricing function evolves according to the following process:

$$dP_t = \kappa\left(\bar{p} - (u_t - \bar{u})\right) dt + \sigma^P dZ^P$$

where $\kappa = r\alpha\left(\sigma^P\right)^2 (r+\theta)^{-1}\theta$ and $\bar{p} = \mu r^{-1}\kappa^{-1} - \bar{u}$. The dynamic budget constraint of the arbitrageurs is:

$$dW_t = r\left(W_t - A_t P_t\right) dt + A_t dP_t + A_t D_t dt$$

Let's consider the following trading strategy:

$$A_t = W_t t^q \left(\bar{u} - u_t\right)$$

where $q > 0$. Replacing for the price and the price process in the budget constraint gives:

$$dW_t/W_t = rdt + \gamma_{1_t}(u_t - \bar{u})\,dt + \gamma_{2_t}(u_t - \bar{u})^2\,dt + \gamma_{3_t}(u_t - \bar{u})\,\sigma^P dZ^P$$

where the following constants are defined:

$$\begin{aligned}
\gamma_{1_t} &= At^q\left(\kappa\bar{u} - r\alpha\left(\sigma^P\right)^2(S - \bar{u})\right) \\
\gamma_{2_t} &= At^q r\alpha\left(\sigma^P\right)^2 \\
\gamma_{3_t} &= -At^q
\end{aligned}$$

Define the logarithm of the discounted wealth process as:

$$\tilde{w}_t - \tilde{w}_0 = \ln\left(e^{-rt}W_t\right) - \ln(W_0)$$

By Ito's lemma we obtain:

$$\tilde{w}_t = \gamma_{1t}(u_t - \bar{u})\,dt + \left(\gamma_{2t} - \frac{1}{2}t^{2q}\sigma^{P2}\right)(u_t - \bar{u})^2\,dt + \gamma_{3_t}(u_t - \bar{u})\,\sigma^P dZ^P$$

To simplify notation, define:

$$\begin{aligned}
\lambda_{1t} &= At^q\left(\kappa\bar{u} - r\alpha\left(\sigma^P\right)^2(S - \bar{u})\right) \\
\lambda_{2t} &= At^q r\alpha\left(\sigma^P\right)^2 - \frac{1}{2}At^{2q}\sigma^{P2} \\
\lambda_{3t} &= -At^q
\end{aligned}$$

Integrating the dynamic budget constraint gives:

$$\tilde{w}_t = \int_0^t \lambda_{1s}\left(u_s - \bar{u}\right) ds + \int_0^t \lambda_{2s}\left(u_s - \bar{u}\right)^2 ds + \sigma^P \int_0^t \lambda_{3s}\left(u_s - \bar{u}\right) dZ_s^P$$

Integrating the noise trader prices:

$$\begin{aligned} u_s - \bar{u} &= e^{-\theta s}\left(u_0 - \bar{u}\right) + \sigma^u e^{-\theta s}\int_0^s e^{\theta v} dZ_v^u \\ \left(u_s - \bar{u}\right)^2 &= e^{-2\theta s}\left(u_0 - \bar{u}\right)^2 + \sigma^{u2}\frac{1 - e^{-2\theta s}}{2\theta} + 2\left(u_0 - \bar{u}\right)\sigma^u e^{-2\theta s}\int_0^s e^{\theta v} dZ_v^u \end{aligned}$$

Replacing the noise trader process into the budget constraint gives:

$$\begin{aligned} \tilde{w}_t &= \left(u_0 - \bar{u}\right)\int_0^t \lambda_{1s} e^{-\theta s} ds + \left(u_0 - \bar{u}\right)^2 \int_0^t \lambda_{2s} e^{-2\theta s} ds + \frac{\sigma^{u2}}{2\theta}\int_0^t \lambda_{2s}\left(1 - e^{-2\theta s}\right) ds \\ &\quad + \sigma^u \sigma^P \int_0^t \lambda_{4s}\left(e^{-\theta s}\int_0^s e^{\theta v} dZ_v^u\right) dZ_s^P \\ &\quad + 2\left(u_0 - \bar{u}\right)\sigma^u \int_0^t \lambda_{2s} e^{-2\theta s}\left(\int_0^s e^{\theta v} dZ_v^u\right) ds + \sigma^P\left(u_0 - \bar{u}\right)\int_0^t \lambda_{3s} e^{-\theta s} dZ_s^P \end{aligned}$$

Without loss of generality, assume that

$$\begin{aligned} S &= \bar{u} \\ u_0 &= \bar{u} \end{aligned}$$

Then the budget constraint simplifies to:

$$\begin{aligned} \tilde{w}_t &= \frac{\sigma^{u2}}{2\theta}\int_0^t \lambda_{2s}\left(1 - e^{-2\theta s}\right) ds + \sigma^u \sigma^P \sigma^{Pu}\int_0^t \lambda_{3s} ds \\ &\quad + \sigma^u \int_0^t \lambda_{1s} e^{-\theta s}\left(\int_0^s e^{\theta v} dZ_v^u\right) ds \end{aligned}$$

Taking expectations of the wealth process gives:

$$\begin{aligned}
E_0\left[\tilde{w}_t\right] &= \frac{\sigma^{u2}}{2\theta}\int_0^t \lambda_{2s}\left(1-e^{-2\theta s}\right)ds+\sigma^u\sigma^P\sigma^{Pu}\int_0^t \lambda_{3s}ds \\
&= \tilde{w}_0+r\alpha\left(\sigma^P\right)^2\frac{\sigma^{u2}}{2\theta}\int_0^t As^q\left(1-e^{-2\theta s}\right)ds-\sigma^u\sigma^P\sigma^{Pu}\int_0^t As^q ds \\
&= \tilde{w}_0+r\alpha\left(\sigma^P\right)^2\frac{\sigma^{u2}}{2\theta}A\left(\frac{t^{q+1}-1}{q+1}-\int_0^t s^q e^{-2\theta s}ds\right)-\sigma^u\sigma^P\sigma^{Pu}A\left(\frac{t^{q+1}-1}{q+1}\right) \\
&= O\left(\frac{r\alpha\left(\sigma^P\right)^2\sigma^{u2}-2\theta\sigma^u\sigma^P\sigma^{Pu}}{2\theta\left(q+1\right)}At^{q+1}\right)
\end{aligned}$$

Clearly, condition 3 for statistical arbitrage is satisfied if and only if $r\alpha\left(\sigma^P\right)^2\sigma^{u2}>2\theta\sigma^u\sigma^P\sigma^{Pu}$.

Similarly, taking the variance of wealth from the budget constraint:

$$\begin{aligned}
Var_0\left[\tilde{w}_t\right] &= \sigma^{u2}\int_0^t \lambda_{1s}^2 e^{-2\theta s}\left(\int_0^s e^{2\theta v}ds\right)ds \\
&= \sigma^{u2}\kappa^2\bar{u}^2\int_0^t A^2 s^{2q}\frac{1-e^{-2\theta s}}{2\theta}ds \\
&= O\left(\frac{\sigma^{u2}\kappa^2\bar{u}^2}{2\theta\left(2q+1\right)}A^2t^{2q+1}\right)
\end{aligned}$$

Due to the normality of the Brownian motion, the probability distribution of wealth is normal:

$$\begin{aligned}
\Pr\left(\tilde{w}_t<0\right) &= \Pr\left(\vartheta<\frac{-E_0\left[\tilde{w}_t\right]}{\sqrt{Var_0\left[\tilde{w}_t\right]}}\right) \\
&= N\left(\frac{-E_0\left[\tilde{w}_t\right]}{\sqrt{Var_0\left[\tilde{w}_t\right]}}\right)
\end{aligned}$$

where ϑ is a standard normal. From the previous results we obtain:

$$\begin{aligned}
\frac{E_0\left[\tilde{w}_t\right]}{\sqrt{Var_0\left[\tilde{w}_t\right]}} &= O\left(\frac{r\alpha\left(\sigma^P\right)^2\sigma^{u2}-2\theta\sigma^u\sigma^P\sigma^{Pu}}{\left(q+1\right)\sqrt{2\theta\sigma^{u2}\kappa^2\bar{u}^2}/\sqrt{2q+1}}\frac{t^{q+1}}{\sqrt{t^{2q+1}}}\right) \\
&= O\left(\xi t^{1/2}\right)
\end{aligned}$$

where $\xi = \frac{r\alpha(\sigma^P)^2\sigma^{u2} - 2\theta\sigma^u\sigma^P\sigma^{Pu}}{(q+1)\sqrt{2\theta\sigma^{u2}\kappa^2\bar{u}^2}/\sqrt{2q+1}}$. In order to evaluate the order of convergence of the probability, consider the following Taylor approximation to the normal distribution:

$$N(x) = N'(x)\left(a_1\frac{1}{1+\gamma x} + a_2\frac{1}{(1+\gamma x)^2} + a_3\frac{1}{(1+\gamma x)^3} + h.o.t.\right)$$

where a_i and γ are constants. The rate of convergence is then:

$$\Pr(\tilde{w}_t < 0) = O\left(\frac{a_1}{\sqrt{2\pi}}\frac{\exp(-\xi^2 t)}{1+\gamma\xi t^{1/2}}\right)$$

So that it can be concluded that the $\Pr(\tilde{w}_t < 0)$ converges to 0 at a rate faster than $exp(-\xi^2 t)$.

■

1.8.3 Optimal Filtering of Noise Trader Demand

Proof. The filtering problem is the following. The arbitrageur observes the price that evolves according to the process:

$$dP_t = \left(\frac{\mu}{r} + \frac{r\alpha\sigma^{P2}}{r+\theta}\theta(\bar{u} - u_t)\right)dt + \sigma^{uP}dZ^u + \sigma^{PD}dZ^D$$

The arbitrageur wants to learn about the current level of noise trader demand from observing the price process. The unobserved noise trader demand evolves according to:

$$du_t = \theta(\bar{u} - u_t)\,dt + \sigma^u dZ^u$$

All the conditions for Theorem 12.7 from Liptser and Shiryaev (2000) are satisfied, so that the filtered process can be directly computed. Denote the filtration generated by P buy $F_t^P = \{P_\tau : \tau < t\}$, the conditional expectation $m_t = E\left[u_t | F_t^P\right]$ and the forecast error $\gamma_t = E\left[\left(m_t - E\left[u_t | F_t^P\right]\right)^2 | F_t^P\right]$, then we find:

$$\begin{aligned}
dm_t &= \theta\left(\bar{u} - m_t\right) dt + \sigma_t^m dZ^m \\
\sigma_t^m dZ^m &= \Gamma_t dP_t - \Gamma_t \left(\frac{\mu}{r} + \frac{r\alpha\left(\sigma^P\right)^2}{r+\theta}\theta\left(\bar{u} - m_t\right)\right) dt \\
\Gamma_t &= \left(\sigma^{uP}\right)^{-2}\left(\sigma^u\sigma^{uP} + \gamma_t r\alpha\theta\left(r+\theta\right)^{-1}\left(\sigma^P\right)^2\right) \\
d\gamma_t &= -2\theta\gamma_t + \left(\sigma^u\right)^2 - \left(\sigma^{uP}\right)^{-2}\left(\sigma^u\sigma^{uP} + \gamma A_1\right)^2
\end{aligned}$$

■

Chapter 2

Learning about Beta: An Explanation of the Value Premium

with Francesco Franzoni, University Pompeu Fabra

2.1 Introduction

Since its invention by Sharpe (1964) and Lintner (1965), the Capital Asset Pricing Model (CAPM) has experienced varying fortune. Although the early tests by Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973) established its empirical success, the CAPM anomalies were discovered soon after. Two of the main empirical failures - the value and small stock premia - were pointed out by Basu (1977) and Banz (1981) respectively. Fama and French

(1992, 1993) present these failures of CAPM in the most dramatic way. They show that market risk, as measured by beta, does not explain the cross-section of average returns for portfolios sorted on past betas, and on size and book-to-market (B/M).

We provide a novel explanation for the value premium, and propose a new empirical methodology for testing asset pricing models. The motivation for our work comes from the observation in Franzoni (2002) that the beta of value and small stocks has varied significantly over the past sixty years. In particular, the value stocks beta has dropped by about 77%, from 2.2 in the early forties to below 0.50 in the late nineties.

In an environment where the risk-factor loadings change dramatically, investors may not know the exact riskiness of the portfolios that they are going to hold. Consequently, they need to form beliefs about the betas, and these beliefs are affected by past levels of the loading. Moreover, as the riskiness of value stocks has been decreasing over time, it is likely that investors' expected beta is significantly higher than the actual beta. The implication is that while investors require an expected excess return that is proportional to the riskiness they perceive, the econometrician observes a premium in excess of the realized riskiness of these stocks, and the CAPM is rejected.

We develop this argument through an equilibrium model of learning, with unobservable and time-varying factor loadings, in the model CAPM holds from investors' ex-ante perspective. However, an econometrician examining realized returns can observe positive mispricing relative to CAPM when expected beta differs from realized beta, and the time-variation of betas is not properly taken into account. We propose that asset pricing tests should take the time-variation explicitly into account.

In practice, we derive an empirical methodology in which we obtain a proxy for investors' expectations of beta by applying the Kalman filter to realized portfolio returns. This measure of beta is then used to explain the cross-section of average returns on a set of ten B/M sorted portfolios. While the standard tests reject CAPM on these data, our version of CAPM augmented with learning is not rejected. Therefore, we provide an account for the valuc premium. These findings are robust to modifications in the empirical strategy. In particular, they extend to portfolios sorted on size, as well as on B/M.

The driving force of our results is the assumption that both the current level as well as the long-run average of beta are unobserved by investors. Learning about these two elements causes investors' expectations to adjust slowly to the new realizations of the loading, and to stick more closely to historical values of beta.

This chapter is organized as follows. In Section 2.2, we develop a model in which CAPM is augmented with learning about factor loadings. The capacity of the model to explain the value premium is assessed through simulations. Section 2.3 proposes an empirical strategy for testing asset pricing models that is consistent with investors learning about the relevant parameters. Moreover, we present empirical evidence that our model explains the cross-section of average returns for portfolios sorted on B/M. Section 2.4 presents robustness checks and extensions of our findings. Our results are compared to the literature in section 2.5. Section 2.6 draws conclusions.

2.2 The Learning-CAPM

Franzoni (2002) demonstrates that the systematic risk of stocks varies substantially over time. This finding suggests that investors may not know the true riskiness of assets precisely. In a world with uncertainty about relevant parameters, investors have to infer the factor loadings from the observable information.

Depending on how fast learning occurs, investors' beliefs can diverge more or less significantly from true factor loadings. Since investors' expectations of factor loadings determine the risk premium required to hold assets, it can be the case that the current value of the riskiness of a portfolio, does not entirely explain expected returns.

In this section, a CAPM model is set-up and analyzed in which investors must form expectations on the true level of systematic risk. The key result is that CAPM holds under investors' information set, but the econometrician who looks at realized returns can observe mispricing.

2.2.1 The set-up

There are N risky assets in the economy indexed by i, where $i = 1, .., N$. Each stock is paying dividends D_t^i that are assumed to be generated by the following factor structure:

$$D_{t+1}^i = \bar{D}^i + b_{t+1}^i x_{t+1} + \varepsilon_{t+1}^i \qquad (2.1)$$

Dividends are determined by a common risk factor x_{t+1}, idiosyncratic risk ε_{t+1}^i, and the factor loading b_{t+1}^i. The factor loading is assumed to be time-varying, according to the following process:

$$b_{t+1}^i = B^i + F^i b_t^i + u_{t+1}^i \qquad (2.2)$$

B^i and F^i are constants drawn at time 0. The shocks to factor loadings u_{t+1}^i are independent across stocks, *i.i.d.* through time, and normally distributed $u_{t+1}^i | x_t \sim N\left(0, \sigma_u^{i2}\right) \quad \forall i, \forall t$.

Furthermore, ε_t^i and u_t^j are independent for all i, j. Idiosyncratic risk is also normally distributed with mean zero conditional on x_t, so that $\varepsilon_t^i | x_t \sim N\left(0, \sigma_\varepsilon^{i2}\right) \quad \forall i, \forall t$. Idiosyncratic risk is independent across stocks, and *i.i.d.* through time. The common factor x_t is assumed to be normally distributed $N\left(0, \sigma_x^2\right)$. Therefore, $\bar{D}^i$ is the unconditional mean of dividends. The cross-sectional distribution of $\bar{D}^i$ is assumed to have bounded expectation, $E\left[\bar{D}^m\right] < \infty$.

There is an infinite number of overlapping generations of representative investors in the economy. The investor of each generation t is working when young, receiving labor income y_t, and consuming c_{t+1} when old. Labor income at time t is denoted by y_t, and assumed to be growing deterministically at rate g. The only decision that the investor has to make is the portfolio choice, i.e. deciding how to save between young and old age. Investors solve:

$$\begin{aligned} & \underset{\left\{\mathbf{a}_t^i\right\}_{i=1}^N}{Max} \quad E\left[-e^{-Ac_{t+1}} | \Im_t\right] \\ st. \ c_{t+1} &= (1+r)\, y_t + \sum_{i=1}^{N} \mathbf{a}_t^i \left(D_{t+1}^i + P_{t+1}^i - (1+r)\, P_t^i\right) \end{aligned} \tag{P}$$

We thus assume that investors have constant absolute risk aversion A. Labor income is deterministic, and grows at rate $(1+g)$. The risk-free asset is assumed to be in infinite supply at rate r. The number of shares of asset i owned by the investor at time t is denoted by $\mathbf{a}_t^i$.

The assumption that there are overlapping generations of investors in the economy allows one to focus on learning as the main intertemporal linkage. In particular, the OLG assumption implies that each investor has a one-period horizon. Changes in the investment opportunity set due to the time-variation of b_t^i only affect demand via a discount rate effect. We abstract from more general effects that would arise if there was correlation between consumption and changes in systematic risk, and that would lead to a multifactor pricing model in the spirit of

Merton (1971). We deliberately focus on a set-up where CAPM holds, and study the impact of learning in this world. It is straightforward to extend the framework to include state variable.

The agents information set evolves according to the following filtration $\Im_t$:

$$\Im_t = \left\{\Im_0, D_s^i, x_s, y_s \text{ for } s \leqslant t, \forall i\right\}$$

where $\Im_0 = \left\{x_0, y_0, \bar{D}^i, F^i, \bar{\mathbf{a}}^i \ \ \forall i\right\}$. The supply of shares outstanding denoted $\bar{\mathbf{a}}^i$, is thus assumed to be drawn at time 0 and known to the investor. The number of stocks is assumed to be large, so that the law of large number can be applied cross-sectionally. This assumption simplifies the learning process that is introduced in the next section.

2.2.2 Learning

In the specification of the investor's filtration, it has been assumed that neither the systematic risk factor loadings b_t^i, nor B^i are observable to the investor. Investors must therefore form expectations about true factor loadings b_t^i, as well as the long-run behavior of factor loadings that is governed by B^i. Investors are assumed to behave rationally, and forecast changes in systematic risk according to Bayes rule. As the systematic risk of factor loadings changes constantly, learning occurs according to the Kalman filter:

$$b_{t+1|t}^{ie} = B_{t-1}^{ie} + F^i b_{t|t-1}^{ie} + k_t^i \left(D_t^i - \bar{D}^i - b_{t|t-1}^{ie} x_t\right) \tag{2.3}$$

where $B_{t-1}^{ie} = E\left[B^i|\Im_{t-1}\right]$, and $b_{t+1|t}^{ie} = E\left[b_{t+1}^i|\Im_t\right]$. The details of the Kalman filter are given in the appendix. The optimal rule is to use the unexpected part of the current dividend realization to update the previous period's estimate of systematic risk. k_t^i is the "gain" and can be interpreted as a regression coefficient. Exact expressions are given in the appendix. It can be seen from the filter that $E\left[b^i_{t+1|t}|\Im_{t-1}\right] = B^{ie}_{t-1} + F^i b^{ie}_{t|t-1}$, i.e. the one-period ahead forecast of systematic risk is a combination of the long-run behavior of b^i_t, as captured by B^i, and the current estimate of the level of risk. The updating equation for expectations about B^i is:

$$B^{ie}_t = B^{ie}_{t-1} + K^i_t \left(D^i_t - \bar{D}^i - b^{ie}_{t|t-1} x_t\right) \tag{2.4}$$

Note that it follows from this equation that B^{ie}_t is a martingale under the investors information set: $B^{ie}_t = E\left[B^{ie}_{t+1}|\Im_t\right]$. The gain matrix for the filter of B^i, which is interpreted as a time-varying regression coefficient, is defined in the appendix.

A key assumption of the model is that dividend news occurs in discrete time. In fact, if news arrived in continuous time, investors could learn beta without error. We think that discrete time is a realistic assumption with regards to information that is relevant to form expectations on beta. For example, one can think that estimates of beta are updated on a quarterly basis, when earnings announcements are released. More generally, our implicit assumption is that very high frequency information is likely to be more qualitative, and difficult to translate into covariances.

2.2.3 The equilibrium pricing function

The equilibrium in the economy has to fulfill three standard criteria:

Definition 2.1 *An equilibrium is defined such that:*

1. *Investors of each generation* $t = 0, .., \infty$ *solve the maximization program* (P)

2. The good market clears:

$$c_t = y_t + \sum_{i=1}^{N} \bar{\mathbf{a}}^i D_t^i \quad \forall t$$

3. Asset markets clear:

$$\mathbf{a}_t^i = \bar{\mathbf{a}}^i \quad \forall i, \forall t$$

This definition of equilibrium is standard. As investors are assumed to form expectations rationally, equilibrium implies that they employ the Kalman filter developed in Section 2.2.2.

We can show that there exists a linear pricing function:

Proposition 2.1 *There exists a linear equilibrium pricing function:*

$$P_t^i = \frac{\bar{D}^i}{r} - \theta^i \left(b_{t+1|t}^{ei} + \frac{B_t^{ei}}{r} \right) \tag{2.5}$$

$\theta^i = \sigma_x^2 A / \left(1 + r - F^i\right)$

The proof is in the appendix. The price is the average discounted expected dividend, less a risk premium. The risk premium consists of the expectation of tomorrow's factor loading - $b_{t+1|t}^{ei}$- plus the discounted expected long-term average of the factor loading - $\frac{1}{r} B_t^{ei}$. The result that prices only vary due to changes in factor loadings arises from the stylized set-up of the model. It is straightforward to extend the model to cases where other shocks cause movements in prices.[1]

[1]If we change the dividend process from equation (2.1) to $D_{t+1}^i = D_t^i + b_{t+1}^i x_{t+1} + \varepsilon_{t+1}^i$, we can show that the equilibrium pricing function is $P_t^i = \frac{D_t^i}{r} - \frac{(1+r)^2 \sigma_x^2 A}{r^2 \left(1+r-F^i\right)} \left(b_{t+1|t}^{ei} + \frac{B_t^{ei}}{r} \right)$.

2.2.4 Returns, Expected Returns, and CAPM

Returns are defined as absolute excess returns:

$$R^i_{t+1} = D^i_{t+1} + P^i_{t+1} - (1+r)\, P^i_t$$

The reason to study absolute excess returns is analytical tractability, and is standard in a CARA-normal set-up. At the stage of calibration, we will transform the relevant quantities into relative returns. Using the pricing function, we can solve for returns:

$$R^i_{t+1} = b^i_{t+1} x_{t+1} + \varepsilon^i_{t+1} + \frac{\sigma_x^2 A}{1+r-F^i} \left(b^{ei}_{t+1|t} - b^{ei}_{t+2|t+1} + \left(B^{ei}_t - B^{ei}_{t+1} \right) / r + r \left(b^{ei}_{t+1|t} + B^{ei}_t / r \right) \right) \tag{2.6}$$

There are three terms to returns. First, $b^i_{t+1} x_{t+1} + \varepsilon^i_{t+1}$ constitutes the unexpected realization of dividends, consisting of systematic and idiosyncratic risk. Second, returns are an increasing function in downward-revisions of factor loadings, $b^{ei}_{t+1|t} - b^{ei}_{t+2|t+1}$, and of the long-run factor loading $\left(B^{ei}_t - B^{ei}_{t+1} \right)$. The intuition is straightforward: when expected factor loadings decline unexpectedly from one period to the next, prices increase as investors require a lower risk premium to hold the stock. The third term is driven by the short-run and long-run components of factor loadings, $b^{ei}_{t+1|t}$ and B^{ei}_t. This is the usual CAPM pricing effect: stocks with higher systematic risk have higher returns, as investors need to be compensated to hold the asset.

The effect of learning on returns can be seen by replacing for the filtering equations (2.3) and (2.4):

$$R_{t+1}^{i} = b_{t+1}^{i}x_{t+1} + \varepsilon_{t+1}^{i} + b_{t+1|t}^{ei}\sigma_x^2 A + \frac{\sigma_x^2 A\left(k_{t+1}^{i} + K_{t+1}^{i}/r\right)}{1+r-F^{i}}\left(x_{t+1}\left(b_{t+1|t}^{ei} - b_{t+1}^{i}\right) + \varepsilon_{t+1}^{i}\right) \quad (2.7)$$

Returns are determined by three parts: dividends news $b_{t+1}^{i}x_{t+1} + \varepsilon_{t+1}^{i}$, a risk premium that is determined by the one-period ahead expectation of systematic risk, $b_{t+1|t}^{ei}\sigma_x^2 A$, and the returns that are driven by surprises in the evolution of systematic risk, $\left(b_{t+1|t}^{ei} - b_{t+1}^{i}\right)$. Note that the one-period ahead expectation of last term in the excess returns is zero from the investors point of view, and uncorrelated with x_t. It is uncorrelated with x_t because we assumed that expectations are formed rationally, and the evolution of systematic risk is uncorrelated with x_t per assumption.

Using equation (2.7), the market return can be derived:[2]

Proposition 2.2 *The absolute market return in excess of the risk-free rate is:*

$$R_{t+1}^{m} = x_{t+1} + \sigma_x^2 A$$

The market return is solely driven by the common risk factor x_t. This comes from the assumption that innovations to risk factors b_t^i are idiosyncratic, and average out for the market as a whole. This means that both idiosyncratic and expected factor loadings average to 1 across

[2] In this proof, we use a number of technical assumptions that ensure that the variation in *beta* averages out in the cross-section. The cross-sectional distribution of $\bar{\mathbf{a}}^i$, is independent of the shocks ε_t^i, u_t^i, and x_t, and independent of the cross-sectional distribution of F^i. It is assumed that $\bar{\mathbf{a}}^i = O\left(N^{-1}\right)$, so that every stock is small relative to the market as the number of stocks gets large. We distinguish two cases, $F^i = 1$ and $F^i < 1$. In the case when $F^i < 1$, it is assumed that there exists a transformation of B^i, such that $\tilde{B}^i\left(1 - F^i\right) = B^i$, and $\sum_{i=1}^{N}\bar{\mathbf{a}}^i\tilde{B}^i = 1$. Furthermore, the economy is in the steady state at time 0, so that $b_0^i = \tilde{B}^i \quad \forall i$. In the case $F^i = 1 \ \forall i$, it is assumed that $\sum_{i=1}^{N}\bar{\mathbf{a}}^i B^i = 0$ and initially, $\sum_{i=1}^{N}\bar{\mathbf{a}}^i b_0^i = 1$. These assumptions are used to prove proposition 2.2.

stocks, and the updating part of returns averages to zero across stocks. From proposition 2.2, the expected market excess return - or equity premium- is $E\left[R_{t+1}^m|\Im_t\right]=\sigma_x^2 A$. The model thus abstracts from time-variation in expected market returns. This is a deliberate choice, as the focus of this chapter is the impact of time-varying betas on cross-sectional pricing anomalies. For individual stocks, we can show that a form of CAPM holds:

Proposition 2.3 *Denote* $\beta_t = Cov\left(R_{t+1}^i, R_{t+1}^m|\Im_t\right)/Var_t\left(R_{t+1}^m|\Im_t\right)$. *CAPM holds:*

$$E\left[R_{t+1}^i|\Im_t\right]=\beta_t E\left[R_{t+1}^m|\Im_t\right]$$

Furthermore:

$$\beta_t = b_{t+1|t}^{ei}$$

Under the investor's information set, the Capital Asset Pricing Model proposed by Sharpe (1964) and Lintner (1965) holds period by period: expected returns in period t are the asset's beta times the expected returns on the market portfolio. The key difference to the static set-up is the time-variation in the asset's betas. In our model, the beta for each asset equals the expected factor loading of the dividends of that asset in period $t+1$, conditional on information in period t.

2.2.5 Unobserved b, and ex-post measures of α

The model so far implied that CAPM holds under the investors information set, i.e. the beta that is determining returns corresponds to investors' estimate of the factor loading. In this section, we will contrast investors' expectations with the ex-post measures of risk by the econometrician. The key is now what we assume about the econometrician's information set. This subsection starts with a general result, that is then used to derive empirical implementations of the model. To start off, assume that the econometrician observes the history of returns:

$$\Im_t^E = \left\{R_s^i, R_s^m \text{ for } t_0 \leqslant s \leqslant t, \forall i\right\}$$

The majority of tests of CAPM are performed using only return data. When betas are not time-varying, returns are a sufficient statistic for the riskiness that is expected by market participants.

Proposition 2.4 *Under the econometrician's information set, we find that estimated beta is:*

$$\hat{\beta}_t^{iE} = \frac{Cov\left(R_{t+1}^i, R_{t+1}^m | \Im_t^E\right)}{Var\left(R_{t+1}^m | \Im_t^E\right)} = E\left[b_{t+1}^i | \Im_t^E\right]$$

The proposition says that the econometrician's estimate of systematic risk, $\hat{\beta}_t^{iE}$, is the econometrician's conditional expectation of the true factor loading, $E\left[b_{t+1}^i | \Im_t^E\right]$. This result stems from the assumption that market returns are uncorrelated with changes in systematic risk, and therefore the Kalman filter updates are uncorrelated with changes in the market return.

Proposition 2.5

$$E\left[R_{t+1}^i | \Im_t^E\right] = \hat{\alpha}_t^{iE} + \hat{\beta}_t^{iE} E\left[R_{t+1}^m | \Im_t^E\right]$$

where

$$\hat{\alpha}_t^E = E\left[b_{t+1|t}^{ei} - b_{t+1}^i | \Im_t^E\right] (1 + G_t)\, \sigma_x^2 A \tag{2.8}$$

and

$$G_t = \frac{E\left[x_{t+1}\left(k^i_{t+1} + K^i_{t+1}/r\right)|\Im^E_t\right]}{1+r-F^i} > 0$$

Proposition 2.5 is the key result of the model. If ex-post estimation of factor loadings differ from estimation of ex-ante expectations of factor loadings, the model predicts that the econometrician will observe positive mispricing. This corresponds to the term $E\left[b^{ei}_{t+1|t} - b^i_{t+1}|\Im^E_t\right]$.

The proposition also shows the role of learning. If there is a wedge in the measurement of ex-ante expectations and ex-post realizations of factors, such that $E\left[b^{ei}_{t+1|t} - b^i_{t+1}|\Im^E_t\right] > 0$, learning amplifies this wedge, through the term G_t, which is always positive. The learning process of investors is underlying G_t: the first term in the numerator, $E\left[x_{t+1}k^i_{t+1}|\Im^E_t\right]$, is the covariation of learning on b^i_{t+1} with the market factor, x_{t+1}. The second term in the numerator, $E\left[x_{t+1}K^i_{t+1}/r|\Im^E_t\right]$, is the (discounted) covariation of learning on B^i with the market factor x_{t+1}. These positive covariations of the learning process with the market factor are a direct consequence of the dividend process in equation (2.1): higher dividends can be either generated by higher x_{t+1} or a higher factor loading b^i_{t+1}. The role of learning on the long-run mean is particularly important: when B^i is known, $K^i_{t+1} = 0$.

The effect of learning and econometric misspecification is multiplicative: we will show in paragraphs below that misspecification of the econometric model gives rise to

$$E\left[b^{ei}_{t+1|t} - b^i_{t+1}|\Im^E_t\right] \neq 0$$

This mismeasurement is multiplied by $(1+G_t)$, and particularly with the effect of learning about the long-run mean.

In simulations in subsection 2.2.6, we examine the extent to which learning matters for measured mispricing with respect to CAPM. Equation (2.8) is directly tested in simulated data. The effects of learning about b_t^i and B^i are disentangled from the effects of surprise moves in factor loadings. A measure of the difference between the ex-post estimation and the ex-ante expectation of factor loadings can be seen in Figure 2-1, that is discussed in more detail in Section 2.3.

One of the implications of equation (2.8) is that there should be no mispricing if the econometric model is properly specified, and the time-variation of b_{t+1}^i is properly taken into account. This observation will be tested in section 2.3. In particular, we will show that the model implies that returns should be estimated with a Kalman filter, and that no mispricing with respect to CAPM should be observed once this is done. This prediction is tested cross-sectionally using estimated betas that result from Kalman filtering returns.

Our theory links the time-variation in beta to the ex-post measurement of alpha. Performing OLS regressions implicitly assume that betas are constant, at least over some time-interval. The OLS slope coefficient, from the regression that uses the whole series of data, is the correct estimator when the data generating process is:

$$H^{ols} : R_t^i = \beta^{i,ols} R_t^m + \varepsilon_t^i$$

In this section, we will examine more closely what the OLS assumption does to ex-post CAPM regressions. Let us assume now that the econometrician's information set consists of the history of returns, and in addition the null hypothesis H^{ols} that we labeled OLS:

$$\Im^{ols} = \left\{ R_s^i, R_s^m, \forall s \leq t; \forall i; H^{ols} \right\}$$

Let $\hat{\beta}_{\tau}^{i,ols} = E\left[b_{\tau}^{i}|\Im_t^{ols}\right]$ and $\hat{b}_t^{ei,ols} = E\left[b_{\tau}^{ei}|\Im_t^{ols}\right]$. Due to the misspecification, we find from proposition 2.5 that:

$$E\left[R_{t+1}^{i}|\Im^{ols}\right] = \hat{\alpha}_t^{i,ols} + \hat{\beta}_t^{i,ols} E\left[R_{t+1}^{m}|\Im_t^{E}\right]$$

where

$$\hat{\alpha}_t^{i,ols} = \left(\hat{b}_t^{ei,ols} - \hat{\beta}_t^{i,ols}\right)(1+G_t)\,\sigma_x^2 A \tag{2.9}$$

The estimate of the intercept in a CAPM regression is thus proportional to the bias that arises from differences in the estimates between realized and expected returns.

For value stocks, factor loadings have declined dramatically over the past decades. Performing OLS on the whole realization of the data implicitly assumes that investors expectations are well proxied by realized factor loadings. However, OLS on ex-post data tends to underestimate ex-ante risk. The intuition for this result is that the decrease in beta over the sample period causes investors' expectations to be above realized factor loadings, and the α for value stocks is picking up this wedge between expectations and factor loadings.

2.2.6 The Role of Learning: Simulations

This section presents the results of simulations of the model for value stocks. The simulations are done in order to disentangle the effects that arise from surprise moves in factor loadings from the effects of learning about time-varying beta. The main conclusion of these simulations is that the model can explain significant amounts of the value premium when learning occurs slowly enough. This is the case if investors have to learn on both the current level of the loading, b_t, and the long-run mean, B. Moreover, the autoregression coefficient F has to be sufficiently low, namely below one. The reader for whom these results are intuitive already at this stage, can skip directly to the empirical analysis in Section 2.3.

First, we consider the case when both the short-run and the long-run behavior of factor loadings - b and B - are known. In this case, the only uncertainty arises from the period-to-period change in b. Depending on the beliefs that investors have about the data generating process of factor loadings, varying degrees of the value premium can be explained.

Second, the effect of learning about the long-run behavior governed by B is studied, when the variation in factor loadings b is observable. Learning about B gives more weight to past observations, and increases the fraction of the value premium that can be explained. Lastly, it is shown that unobservability about the short-run and long-run systematic risk - b and B - can account for a large fraction of the value premium. In particular, we argue that the decline in the riskiness of value stocks leads to an ex-post bias to underestimate risk when OLS regressions of portfolio returns on the market return are performed.

In the simulations, the path for $\{b_t^i\}$ and $\{x_t\}$ are fixed. The only variables that are randomly generated are the paths of idiosyncratic risk, $\{\varepsilon_t^i\}$. The calibration and data description is in the appendix, in Section 2.7.7. In the case when neither b_t nor B are observable to the investor, Kalman filtering equations discussed in Section 2.2.2 are used to compute investors' expectations. For each time t, the Kalman filter updates in expectations can be computed. In particular, equations (2.3) and (2.4) can be rewritten as:

$$\begin{aligned} b_{t+1|t}^{ie} &= B_{t-1}^{ie} + F^i b_{t|t-1}^{ie} + k_t^i \left(\varepsilon_t^i + \left(b_{t+1} - b_{t|t-1}^{ie}\right) x_t\right) \\ B_t^{ie} &= B_{t-1}^{ie} + K_t^i \left(\varepsilon_t^i + \left(b_{t+1} - b_{t|t-1}^{ie}\right) x_t\right) \end{aligned}$$

The updating of investor's beliefs from period $t-1$ to t is solely a function of the realizations

Table 2.1: Simulation Results. Each cell in the table corresponds to averages accross 500 simulations. In each simulation, series of monthly returns for 1931:7 - 2001:12 are generated, investors' expectations are computed, and ex-post CAPM regressions are performed on 10-year rolling windows. The first two rows correspond to different assumptions about the observability of b and B. Different values for F correspond to different beliefs of investors. When B is unobservable, the values reported in the first column correspond to investors' priors. Columns denoted (1) report the percentage of simulated alphas that are significant in estimation windows when the alphas on real data are also sgnificant. The significance of the simulated alpha is computed using the standard deviation across repetitions. Columns denoted (2) report the average ratio of the simulated relative to the real alpha in windows when both are significant. Columns denoted (3) report the ratio of the simulated alpha to the real alpha obtained from a single regression in the 1963:7 - 2001:12 subsample. If the ratio is negative, zero is reported. In that sample, the real alpha is 0.45% monthly.

		Observable b Observable B			Observable b Unobservable B			Unobservable b Observable B			Unobservable b Unobservable B		
F	B	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
1	-.01	10	.18	.00	26	.14	.02	0	.00	.00	31	.14	.03
1	.00	21	.15	.01	10	.30	.02	29	.18	.09	30	.13	.07
1	.01	39	.14	.09	21	.15	.05	100	.51	.81	22	.17	.04
F	$\frac{B}{1-F}$	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
.99	1	14	.21	.01	13	.25	.02	12	.29	.00	94	.18	.29
.99	1.2	29	.15	.03	13	.20	.02	83	.14	.15	100	.21	.29
.99	1.5	47	.14	.09	11	.27	.05	100	.20	.34	92	.16	.29
.97	1	10	.28	.00	13	.23	.04	13	.29	.00	100	.37	.58
.97	1.2	44	.12	.08	49	.14	.09	87	.16	.24	100	.35	.55
.97	1.5	56	.15	.12	29	.15	.07	100	.41	.62	100	.37	.57
.95	1	10	.30	.00	65	.14	.17	12	.36	.02	100	.44	.69
.95	1.2	26	.16	.06	57	.14	.15	100	.20	.30	100	.47	.73
.95	1.5	34	.17	.15	52	.13	.12	100	.44	.70	100	.42	.69

of $(\varepsilon_t^i, b_{t+1}^i, x_t)$, and the parameters. In part 2.7.1 of the appendix , the expression for k_t^i and K_t^i are given, and they depend on the same variables and parameters. Given a path of $\{b_t^i\}$, beliefs are fully determined. Once beliefs are determined, portfolio returns are computed from equation (2.6). These simulated returns are then regressed on the market return, and summary statistics are reported in Table 2.1.

Results from two different types of regressions are reported in this table. The first type of regressions are for 10-year rolling windows, and results are reported in the columns denoted (1) and (2). The second type of regressions are over the period 1963:7-2001:12. Each table reports summary statistics from 500 repetitions. Average alphas are estimated from the simulated data under different assumptions about the observability of b and B. The values for F and B are the assumed values for the time-series process of b, in equation (2.2).

There are a number of comparative static results in the case that B is observable. First, lower values of F lead to a higher fraction of the value premium that can be accounted for in the 1963-2001 sample, holding B fix. This can be seen in the columns denoted (3). The intuition is that lower F implies a less persistent process for b, leading systematic risk to have a stronger tendency to revert back to the long-run mean. Thus, the lower F, and the higher B, the higher the fraction of the value premium that is explained.

When B is unobservable, the values for B and $B/(1-F)$ correspond to initial values of the Kalman filter. It is clear from looking at the columns denoted (3) that these initial conditions do not matter much for the 1963-2001 sample as long as $F < 1$. Priors about B do not have a very persistent effect, unless the process has a unit root, i.e. $F = 1$. This is true regardless of whether b is observable or not.

A key insight of the simulations comes from looking at the effect of learning about the long-run mean. The unobservability of B leads to measures of α that are similar to assuming that B is observable and large. Learning about B makes the learning process slower as long as $F < 1$.

The effect of learning about the long-run mean is even more pronounced in the case of unobservable b. Take the case when $F = .97$. In this case, column (3) gives a fraction of the value premium explained in the 1963-2001 sample of .24 and .62 when B is assumed to be 1.2 and 1.5 respectively. In the case that B and b are unobservable and $F = .97$, the corresponding values from column (3) are .55 and .57. This shows that the prior about B does not matter much, and it shows that the unobservability of B pushes the fraction of the value premium that can be explained close to the level where $B = 1.5$ is assumed. In the case that $F = .97$, learning does drive the results. In particular, comparing the values of column (3) horizontally for the case $F = .97$ and $B/(1-F) = 1.2$ shows that the fraction of the value premium that the model can explain goes from 8% to 55%. Most of the increase in explanatory power comes from the interaction in learning about b and B.

For the case that $F = .99$, the increase in explanatory power comes from the assumption that b is unobservable. This is because mean reversion of b is the less important, the closer F is to 1. Finally, it might be worthwhile pointing out that the case $\{F = 1, B = .01, b$ unobserved and B observed$\}$ has high explanatory power (the fraction of the premium explained in the 1963:7-2001:12 sample is .81), but this is due to an unrealistic assumption: investors belief in this case that beta is increasing on average, even though it is decreasing in reality.

Our preferred estimates are the ones when $F = .97$ or $F = .99$ and both b and B are unobserved. In these cases, initial conditions do not matter much, and 29% or 55% – 58% of the value premium can respectively be explained. Furthermore,in the columns denoted (1) and (2) show that the α in the simulated data is significantly different from 0 in 92% – 100% of the cases when the α in the real data is significant.[3]

As a conclusion to this section, the simulations demonstrate that the model can account for a large fraction of the value premium when both b and B are unobservable. Learning about

[3] As a comparison, we have also computed the corresponding values of columns (1) – (3) in the case that investors learn using 5-year rolling window regressions. The simulated values we obtain are 18%, .20 and .06 for the three columns. This low explanatory power of the 5-year rolling window regression occurs because not enough information from the past is taken into account. On the contrary, when investors learn about the long-run mean, their expectations are affected by realizations of b that lie far back in the past.

the long-run mean gives weight to realizations of b that are far back in the past, and thus lead to expectations about the level of systematic risk that are relatively high.

We conjecture that our results are robust to set-ups that include more general stochastic processes for b, as long as investors have to learn about the long-run properties of the process of b. This is really the key lesson of the simulation exercise: a large fraction of the value premium can be explained as long as observations about b from the past matter a lot. In a set-up where the long-run mean can switch between states, our result would go through as long as investors have to estimate what those states are. The implication that b is mean-reverting in cases when $F < 1$ is sensible, as it is hard to believe that factor loadings will diverge to infinity with certainty, as would be implied by a nonstationary process.

2.3 Testing the Learning CAPM

2.3.1 The empirical predictions of the Learning CAPM

The model presented in Section 2.2 suggests that CAPM holds under investors' probability distribution. Hence, in order to test CAPM in a consistent way, the econometrician needs to fulfill two requirements. First, the information set that is used to compute betas has to be a subset of investors' information set. Second, all the available information has to be used in an 'optimal' way, meaning that the econometrician has to replicate the filtering process that investors undertake. We label the empirical model that follows from these prescriptions Learning CAPM (LCAPM).

Our argument can be made formally by referring to proposition 2.5. If the econometrician only uses information up to time t, her information set is a subset of the investors' one: $\Im_t^E \subseteq$

$\Im_t$[4]. Moreover, if the econometrician 'optimally' uses this information to compute betas, the conditional expectation is the proper operator to express her forecast of factor loadings. Under these two conditions, the law of iterated expectations can be applied to proposition 2.5, and prove that mispricing relative to CAPM is zero also from the econometrician's point of view[5]:

$$\begin{aligned}
\hat{\alpha}_t^E &= E\left[b_{t+1|t}^{ei} - b_{t+1}^{i} | \Im_t^E\right](1+G_t)\,\sigma_x^2 A \\
&= E\left[b_{t+1|t}^{ei} - E\left[b_{t+1}^{i} | \Im_t\right] | \Im_t^E\right](1+G_t)\,\sigma_x^2 A \\
&= 0
\end{aligned}$$

Among the implications of this result is the fact that CAPM is not expected to hold when betas are computed at time t with information not yet available to investors ($\Im_t^E \not\subseteq \Im_t$). This is particularly true if betas are time-varying. Let us abstract for a moment from the fact that investors have to learn about factor loadings , and focus only on the variability of beta over time. Suppose, for example, that the true underlying beta is decreasing from t to $t+T$, as in the case of value stocks. If the econometrician assumes it is constant, and beta is estimated using the whole interval of data, a biased estimate of the underlying β_τ for $\tau \in [t; t+T]$ is obtained. In particular, the estimate is biased downward for values of τ close to t, and biased upward for τ close to $t+T$. The argument extends for any variability of the underlying factor loading. As a consequence of estimating β_τ with a bias, the econometrician introduces measurement error in the right-hand side variable, and the tests of the CAPM lose power.

[4]It is a subset because the econometrician only observes returns, while investors' information set also includes dividends. We could expand the econometrician's information set to dividends, but while not altering the conclusions of our argument, it would unnecessarily force us to move from the space of returns to the space of dividends, which is less preferred in empirical tests.

[5]A further assumption subsumed by this result is that the econometrician has in mind the same process for the factor loadings as the one used by investors in their filtering mechanism.

This problem is the more serious, the longer the interval of time over which a constant beta is estimated, and the larger the variability of the underlying factor loading. Therefore, standard cross-sectional tests of CAPM, that assume that beta is constant over almost forty years, and that use portfolios sorted on B/M and size, are biased against CAPM. The evidence in Franzoni (2002) that betas for B/M and size portfolios have been changing dramatically over the past six decades exacerbates the bias. The time-series tests are equally subject to error, as they also assume that beta is constant.

The implications of time-variation in betas for CAPM tests, which are evident from our model, are not new to the literature. Time variation in the underlying factor loadings has been taken into account in a number of ways. One alternative is the Fama-MacBeth methodology (1973), which provides more flexibility to the cross-sectional tests by shortening the window over which beta is estimated. In this approach beta is usually computed over the prior five years of data (Fama and French, 1992). A more elaborate alternative is to model the variation of conditional moments as an autoregressive-conditional heteroskedasticity in the mean (ARCH-M) model (Engle, Lilien, and Robbins, 1987). Another alternative is the conditional CAPM methodology, which uses state variables that are available to investors to predict beta in each period. One of the first studies in this strand of literature was Harvey (1989), and one of the latest examples is Lettau and Ludvigson (2001). More recently, Ang and Chen (2002) have proposed a bootstrapping methodology to adjust confidence intervals in order to account for time-variation in beta.

The contribution of our model, however, is to point out that not only betas are time-varying and unobserved. This fact implies that investors have to learn about betas using the available information in an 'optimal' way. The simulation results in Section 2.2.6 show that the assumption of learning, combined with a decreasing factor loading, is able to account for a large part of the estimated value premium. As investors tie their expectations to the high past realizations of the loading, they require an expected return that is not justified by the betas that are measured ex-post, after the decrease has occurred.

We propose a new empirical methodology for testing CAPM that fully takes into account the predictions of our model. Since CAPM holds from investors' point of view, the econometrician needs to estimate betas by replicating the filtering process, which investors undertake. In our approach, the betas that explain the cross-section of returns at time $t+1$ are the result of Kalman filtering returns up to time t. Therefore, the first step in implementing a test of a Learning CAPM consists of obtaining estimates of the factor loading from the Kalman filter of returns.

2.3.2 Kalman filtering returns

When the econometrician's information set is a subset of investors' information, using equation (2.7) we can rewrite returns:

$$R^i_{t+1} = b^i_{t+1} R^m_{t+1} + \eta^i_{t+1}$$

where

$$\eta^i_{t+1} = \left(1 + g^i_{t+1}\right)\left(x_{t+1}\sigma^2_x A\left(b^{ei}_{t+1|t} - b^i_{t+1}\right) + \varepsilon^i_{t+1}\right)$$

and

$$g_{t+1} = \left(1 + r - F^i\right)^{-1}\sigma^2_x A\left(K^{ib}_{t+1} + K^{iB}_{t+1}/r\right)$$

The result that $E\left[\eta^i_{t+1}|\Im_t\right] = 0$ follows from the properties of the Kalman filter. From here, by the law of iterated expectations, we prove that $E\left[\eta_{t+1}^{i}|\Im_{t}^{E}\right] = 0$. Furthermore, $E\left[\eta_{t+1}^{i}|\Im_{t}, x_{t+1}\right] = E\left[\eta_{t+1}^{i}|\Im_{t}^{E}, x_{t+1}\right] = 0$.

So, the correct specification of the state space system from the econometrician's point of view is very simple:

$$\begin{aligned} Observable &: R_{t+1}^{i} = \beta_{t+1} R_{t+1}^{m} + \eta_{t+1}^{i} \\ State\ Eq\ 1 &: b_{t+1}^{ie} = B^{i} + F^{i} b_{t}^{i} + u_{t+1}^{i} \\ State\ Eq\ 2 &: B^{i}\ \ constant \end{aligned} \tag{2.10}$$

These equations imply that the econometrician applies the Kalman filter directly to realized returns.

In practice, we use the Kalman filtering procedure for a model with time-varying coefficients as in Hamilton (1994). For more details on the filtering equations, we refere to Section 2.7.1 of the appendix[6]. This methodology allows us to derive a filtered beta series for each of the ten B/M decile portfolios in the 1931:7-2001:12 sample. The data points in the 1926:7-1931:6 interval are used to compute OLS estimates that provide initial conditions for β and the forecast error on β in the Kalman filter. This choice of initial conditions is not crucial for our results.

Note that one limitation of this approach is that it is not possible to have investors, or the econometrician, learn about the autocorrelation F^{i} of factor loadings using the Kalman filter. The reason is that the Kalman filter is linear, except for the conditioning variable x_t. Learning about both the current state of the factor loading, and the autocorrelation F^{i} is inherently

[6]The equations in the appendix refer to the filter for investors' problem. To obtain the filtering system that is used in the empirical analysis, one just needs to replace D_{t+1}^{i} with R_{t+1}^{i}, and x_{t+1} with R_{t+1}^{m} in the Observable equation.

nonlinear.

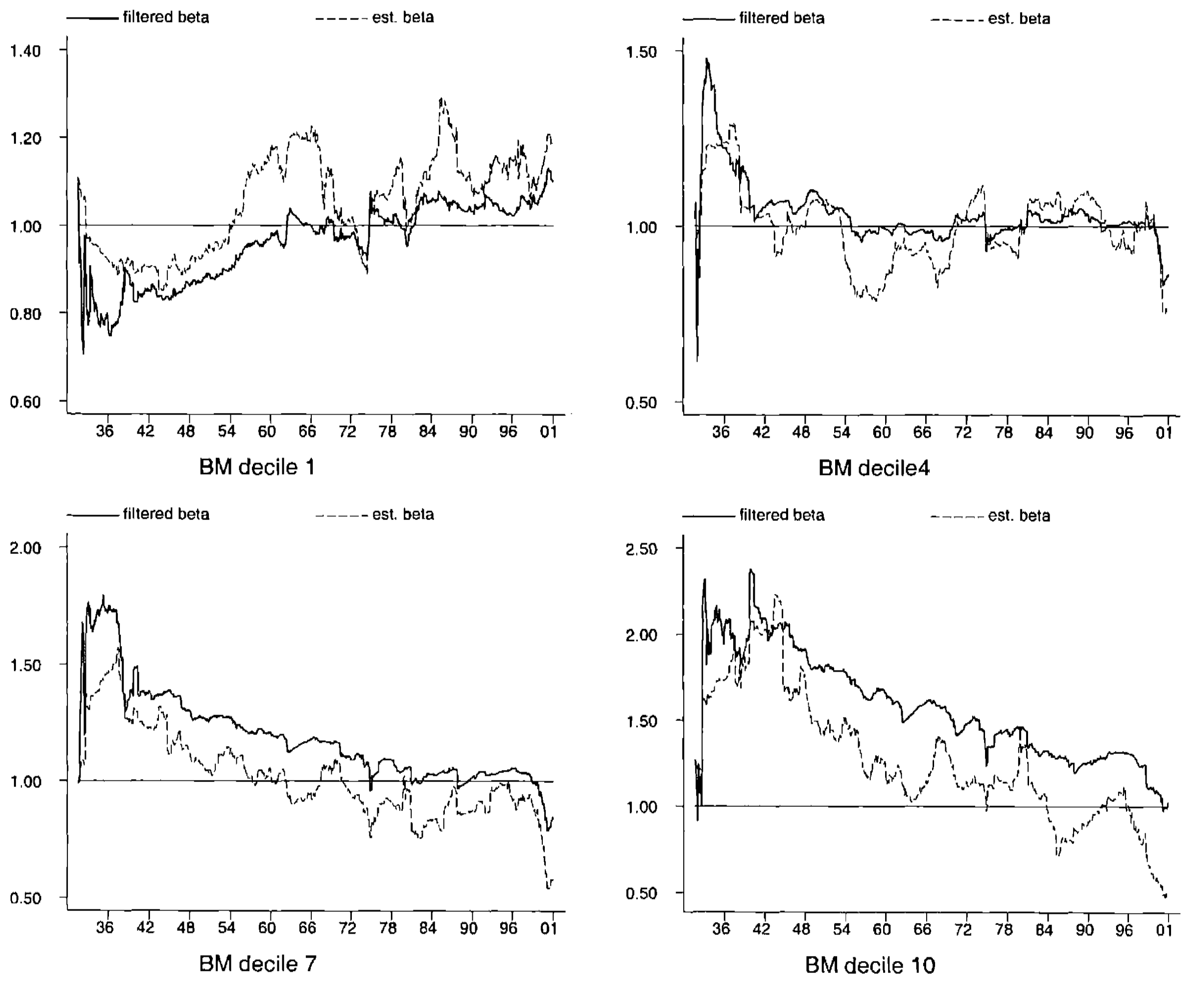
However, we can directly estimate the parameters of the state equations in (2.10). In particular, the autocorrelation F^i, and the variance-covariance matrix of error terms is estimated using maximum likelihood on the whole history of portfolio and market returns. One of the identifying assumptions of our model is that innovations to u_t^i are uncorrelated with innovations to η_t^i, for each stock, and error terms are uncorrelated across stocks. We impose that the autocorrelation of factor loadings $F^i = F \; \forall i$. The ML-estimate of factor loadings from the Kalman filter on the ten B/M portfolios is around .97 monthly.

The simulations in Section 2.2.6 show that a value for F of .97 delivers expectations of beta that do not adjust too quickly to new realizations. So, expected returns today are still affected by the history of beta. In particular, Table 2.1 tells us that in the simulated data when investors assume F to be equal to .97, we can account for almost 60% of the value premium.

The graphs in Figure 2-1 provide a visual impression of the speed of adjustment of the filtered beta to new information. Each graph reports the series of beta coming out of the Kalman filter (with $F = .97$) for B/M deciles 1, 4, 7, and 10 respectively, along with the series of beta resulting from five-year rolling window regressions. It is evident how the filtered series gives more weight to past information. In particular, when the rolling window beta is increasing, as in the case of lower B/M deciles, the filtered series is below the estimated series. On the other hand, for higher B/M deciles the filtered beta lies above the estimated one, because there is learning about a decreasing series. For example, in the case of the 10th B/M decile the beta estimated on the last five years of data (1:1997-12:2001) is .51, while the filtered beta in December 2001 is 1.03, more than twice as much[7]. More information on the behavior of the

[7]The relative behavior of estimated and filtered betas for the remaining B/M deciles resembles that for the

Figure 2-1: Estimated beta Vs. Filtered beta, by B/M deciles. Each of the graphs plots a series of estimated betas from five-year rolling window regression and a series of beta derived from Kalman filtering returns, with an autoregression coefficient F=.97. The return series are for B/M decile portfolios. Deciles 1, 4, 7, and 10 (clockwise from the top-left graph) are reported.



filtered beta series is provided in Tables 2.2 and 2.3. Again, notice that for the higher B/M deciles, the filtered betas tend to be above the OLS estimates, as the filter gives more weight than OLS to the past, and the beta of value stocks has been decreasing. The opposite occurs for the lower deciles.

The wedge between the Kalman filtered betas and the betas from simple rolling window regressions captures the discrepancy that exists in our model between investors' expectations of the loading and the current value of the loading. The key to interpret our empirical strategy is to realize that the econometrician has to take this wedge into account, because investors' expectations of the loadings are what matters for pricing.

2.3.3 Empirical Implementation

Our focus is to explain the 'value premium'. Hence, the empirical analysis of this section uses the ten B/M decile portfolios as test assets. Davis, Fama, and French (2002) provide details on the construction of B/M and size sorted portfolios that span the same sample period as the one we use. Our test assets are formed in a similar fashion, the only difference is that the sorting is performed only along the B/M dimension. Tables 2.2 and 2.3 report summary statistics for these portfolios. Notice in particular that while value stocks have higher excess returns than growth stocks in each sample, their beta is lower in the second sub-sample. This phenomenon is the basis of the 'value premium', as we will argue later.

Our data starts in July 1926. As we need the first five years for the first estimation window, results for the period 1931:7-2001:12 are reported. We also report results for the subperiods

deciles presented in Figure 2-1. In particular, for the lower deciles the estimated series is above the filtered series, as they are both increasing; for the higher deciles the estimated beta is below the filtered beta, as the loading is decreasing.

Table 2.2: Summary statistics (1931:7-2001:12). The table reports summary statistics for the 10 B/M decile portfolios (constructed as in Davis, Fama, and French, 2000), for the market portfolio (CRSP value-weighted), and for the risk free rate (return on 3-month T-Bills). The relevant sample for this table is 1931:7-2001:12. For each B/M portfolio, and for the market the reported statistics are the average annualized return in excess of the risk free rate, with its standard error in parenthesis. For the risk free rate the simple annualize return is given. Also, for each decile portfolio the beta from an OLS time-series regression in the given sample is reported. For each B/M portfolio we report the average, the minimum, and the maximum of the Kalman filtered beta series in the given sample. The description of the estimation of Kalman filtered betas is given in Section 2.3.2.

	1931:7-2001:12				
	$\bar{R}^{e\,i}$	$\hat{\beta}^{ols}$	$\hat{\beta}^{kf}$		
			avg.	min	max
dec. 1	7.17 [2.36]	1.00	.95	.70	1.13
dec. 2	8.31 [2.25]	.97	.96	.72	1.04
dec. 3	7.92 [2.20]	.94	.93	.70	1.03
dec. 4	9.05 [2.51]	1.06	1.03	.61	1.47
dec. 5	9.43 [2.32]	.97	.96	.79	1.29
dec. 6	9.75 [2.59]	1.08	1.10	.86	1.72
dec. 7	10.74 [2.80]	1.14	1.19	.78	1.79
dec. 8	12.37 [2.85]	1.15	1.24	.92	1.93
dec. 9	13.74 [3.36]	1.32	1.41	.85	2.23
dec. 10	14.15 [3.86]	1.43	1.57	.91	2.37
Mkt	8.20 [2.23]				
Rf	3.79 [0.10]				

Table 2.3: Summary statistics (1931:7-1963:6, and 1963:7-2001:12). The table reports summary statistics for the 10 B/M decile portfolios (constructed as in Davis, Fama, and French, 2000), for the market portfolio (CRSP value-weighted), and for the risk free rate (return on 3-month T-Bills). The relevant samples for this table are 1931:7-1963:6, and 1963:7-2001:12. For each B/M portfolio, and for the market the reported statistics are the average annualized excess return in excess of the risk free rate, with its standard error in parenthesis. For the risk free rate the simple annualize return is given. Also, for each decile portfolio the beta from an OLS time-series regression in the given sample is reported. For each B/M portfolio we report the average, the minimum, and the maximum of the Kalman filtered beta series in the given sample. The description of the estimation of Kalman filtered betas is given in Section 2.3.2.

	Panel A: 1931:7-1963:6					Panel B: 1963:7-2001:12				
	$\bar{R}^{e\,i}$	$\hat{\beta}^{ols}$	$\hat{\beta}^{kf}$			$\bar{R}^{e\,i}$	$\hat{\beta}^{ols}$	$\hat{\beta}^{kf}$		
			avg.	min	max			avg.	min	max
dec. 1	10.31 [3.78]	.94	.88	.70	1.10	4.56 [2.97]	1.10	1.02	.92	1.13
dec. 2	11.03 [3.78]	.94	.92	.72	1.04	6.05 [2.66]	1.02	1.00	.90	1.04
dec. 3	10.24 [3.72]	.89	.88	.70	1.00	5.99 [2.67]	1.02	.97	.89	1.03
dec. 4	10.78 [4.65]	1.12	1.07	.61	1.47	5.78 [2.62]	.96	1.00	.83	1.05
dec. 5	13.49 [4.19]	1.02	1.01	.82	1.29	6.06 [2.45]	.89	.93	.79	0.98
dec. 6	12.46 [4.89]	1.19	1.21	.93	1.72	7.49 [2.45]	.89	1.01	.86	1.08
dec. 7	13.60 [5.43]	1.32	1.35	.98	1.79	8.37 [2.42]	.83	1.05	.78	1.19
dec. 8	16.45 [5.56]	1.34	1.41	.95	1.93	8.98 [2.41]	.83	1.10	.82	1.25
dec. 9	18.50 [6.72]	1.58	1.65	1.11	2.23	9.79 [2.57]	.87	1.20	.85	1.39
dec. 10	18.11 [7.72]	1.72	1.84	.91	2.37	10.86 [2.98]	.94	1.35	.97	1.62
Mkt	11.27 [3.90]					5.64 [2.49]				
Rf	1.06 [0.05]					6.07 [0.11]				

1931:7-1963:6 and 1963:7-12:2001. There are two reasons for this split. Firstly, Fama and French (1992, 1993) begin their estimation in 1963:6. Secondly, the value premium only appears in the second sub-sample.

The strategy we choose for estimating the learning CAPM resembles the cross-sectional tests of CAPM. These tests involve two stages. First, beta for each portfolio is estimated on the whole history of data with time-series regressions. Then, portfolio average returns are regressed on the estimated betas. The main implication of CAPM in its most general version is that beta is a significant predictor of the cross-section of average returns. In particular, the coefficient on beta should be equal to the market risk premium. In the Sharpe (1964) and Lintner (1965) version with risk free rate the constant in the regression should be equal to the risk free rate, or to zero if excess returns are used.

The two-stage methodology that we propose entails some important modifications of the one just described. In the first stage we obtain a series of Kalman filtered betas for each portfolio, according to the procedure that was exposed in the previous subsection, and using a value for F of 0.97.

In the second stage we relate the cross-section of average returns to the filtered betas. Unlike the standard tests, we have a different beta for each time period. So, in order to derive a cross-sectional test, we take unconditional expectations of the conditional version of CAPM. If CAPM holds conditionally, as is assumed by our model, it is the case that:

$$E_t\left[R^i_{t+1}\right] = \beta^i_t E_t\left[R^m_{t+1}\right] \tag{2.11}$$

where R_{t+1} denotes excess returns.

By the law of iterated expectations, taking the unconditional expectation of each side yields[8]:

$$E\left[R_{t+1}^{i}\right] = E\left[\beta_{t}^{i} R_{t+1}^{m}\right] \tag{2.12}$$

Equation (2.12) is the basis for the cross-sectional test. For each portfolio, we multiply the Kalman filtered beta series by the market return one period ahead. Then, we regress average portfolio returns on the time-series mean of these new series. In addition to the standard CAPM predictions, this methodology produces the theoretical restriction that the coefficient on the interaction between filtered betas and the market return has to equal one.

Hence, defining M_t^i as $(\hat{\beta}_t^{i,kf} \cdot R_{t+1}^m)$, where $\hat{\beta}_t^{i,kf}$ is the filtered beta series, the regression that we run in the case of LCAPM is

$$\bar{R}^i = g_0 + g_1 \bar{M}^i + e^i \tag{2.13}$$

where the bar above a variable denotes the time-series mean.

We compare the results of our tests with the standard cross-sectional tests of CAPM, for which the estimated regression is:

$$\bar{R}^i = g_0 + g_1 \hat{\beta}^{i,ols} + e^i \tag{2.14}$$

where $\hat{\beta}^{i\,ols}$ is the OLS estimate of portfolios beta on the whole sample.

At this point it is important to remark that the g_1 coefficients in equations (2.13) and (2.14)

[8]See Cochrane (2001) for a detailed discussion of the difference between conditional and unconditional asset pricing equations.

Table 2.4: Cross-sectional Tests(10 B/M portfolios, whole sample). The table reports estimates g_1 from cross-sectional regressions of average excess returns on betas (for the CAPM), and on the time-series average of the product between the kalman-filtered beta and the excess market return (LCAPM). The regressions also include a constant (g_0). P-values and 5% confidence intervals are provided. The test assets are the ten B/M decile portfolios. The estimation samples are given in the table.

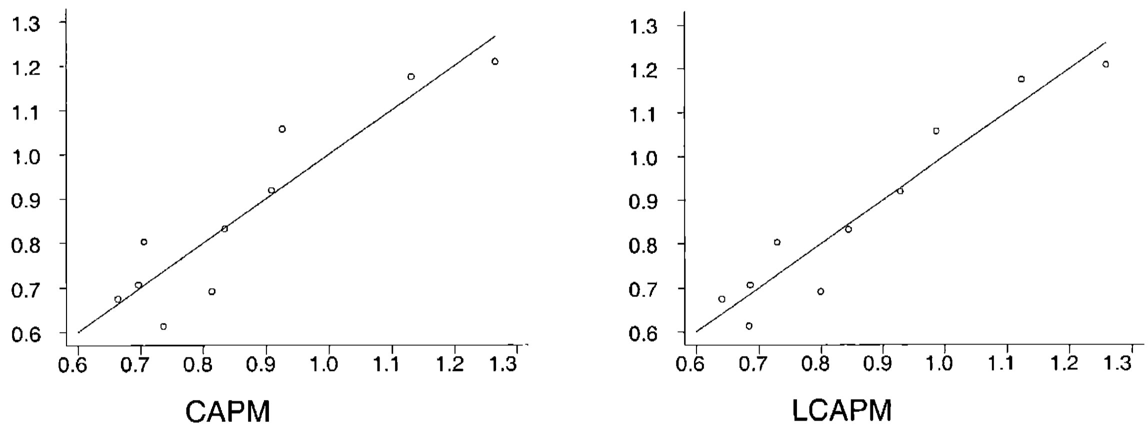
	1931:7-2001:12	
	CAPM	LCAPM
g_0	-.48	-.04
p-value	[.04]	[.64]
conf. int.	$(-.95; -.01)$	$(-.27; .18)$
g_1	1.21	1.12
p-value	[.00]	[.00]
conf. int.	$(.79; 1.63)$	$(.84; 1.39)$
adj. R^2	83.01%	90.70%

have different interpretations. The estimate of g_1 in the CAPM regression (2.14) is expected to equal the market premium. Instead, the g_1 in the LCAPM equation (2.13) is constrained by the theory to be equal to one, as the variable $\bar{M}^i$ already includes the market return.

From Table 2.4 we can assess the performance of our LCAPM against the standard CAPM tests in the 1931:7-2001:12 sample. In these data, the traditional version of CAPM does fairly well in terms of explaining the cross-section of average returns, as it captures about 83% of the variation. However, the constant in the regression is significantly negative, and the estimated market premium is a preposterous 1.21% per month (14.5% annually). Comparatively, our LCAPM performs well on all fronts. The adjusted R^2 is about 90%, and the constant is not significantly different from zero. Moreover, the theoretical restriction that g_1 has to equal one is not rejected by the data.[9] Figure 2-2 provides a graphical impression of these results. The

[9] We have preliminary evidence that our results are confirmed when we use a wild bootstrapping methodology to compute confidence intervals.

Figure 2-2: Performance of the CAPM and LCAPM, 1931:7-2001:12. The diagram on the left corresponds to the CAPM, the one on the right to the LCAPM. The horizontal axes correpspond to the predicted average excess returns, and the vertical axes to the sample average excess returns, for the 10 B/M decile portfolios. The predicted average excess returns are based on the regressions in Table 2.4. The sample period is 1931:7-2001:12. The straight line is the 45-degree line.



graphs plot portfolio average returns against the predictions of each of the two models. If a model predicted 100% of the variation of average returns, all the points would fall on the 45-degree line.

It is well known that the performance of the CAPM varies across sub-samples. Therefore, table 2.5 and figures 2-3 and 2-4 report results by sub-samples. In the early data (Table 2.5) the standard CAPM does a fair job in explaining the cross-section of returns. The restriction on the constant is satisfied as well. Our LCAPM slightly outperforms it by achieving an R^2 of almost 87%.

In the second sub-sample (Table 2.6), the rejection of CAPM by the standard tests is well known, and it constitutes the so-called 'value puzzle'. This fact can be explained as follows. Although the average returns on value stocks are significantly higher than those on growth

Table 2.5: Cross-sectional Tests (10 B/M porfolios, by subsamples). The table reports estimates g_1 from cross-sectional regressions of average excess returns on betas (for the CAPM), and on the time-series average of the product between the kalman-filtered beta and the excess market return (LCAPM). The regressions also include a constant (g_0). P-values and 5% confidence intervals are provided. The test assets are the ten B/M decile portfolios. The estimation samples are given in the table.

	1931:7-1963:6	
	CAPM	LCAPM
g_0	.07	.28
p-value	[.65]	[.04]
conf. int.	$(-.28; .43)$	$(.01; .55)$
g_1	.90	.73
p-value	[.00]	[.00]
conf. int.	$(.61; 1.19)$	$(.51; .95)$
adj. R^2	84.74%	86.90%

Figure 2-3: Performance of the CAPM and LCAPM, 1931:7-1963:6. The diagram on the left corresponds to the CAPM, the one on the right to the LCAPM. The horizontal axes correpspond to the predicted average excess returns, and the vertical axes to the sample average excess returns, for the 10 B/M decile portfolios. The predicted average excess returns are based on the regressions in Table 2.5. The sample period is 1931:7-2001:12. The straight line is the 45-degree line.

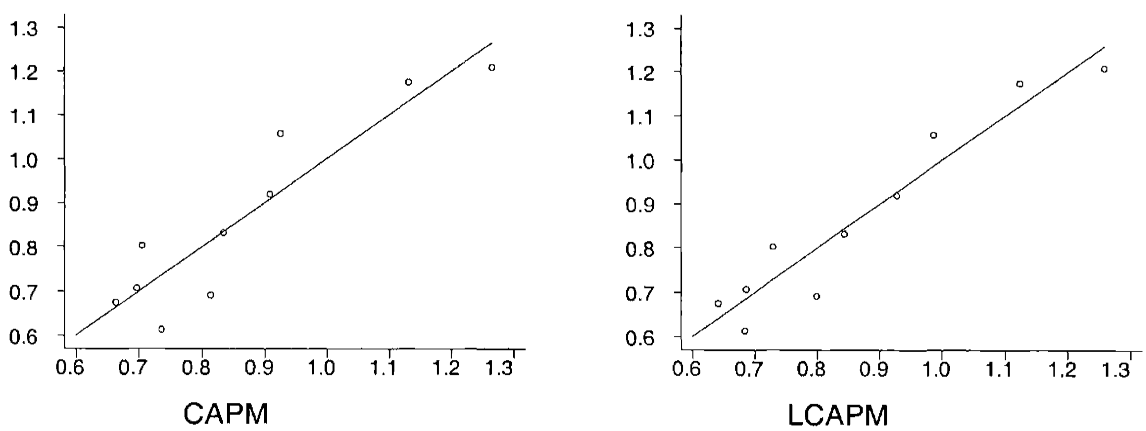
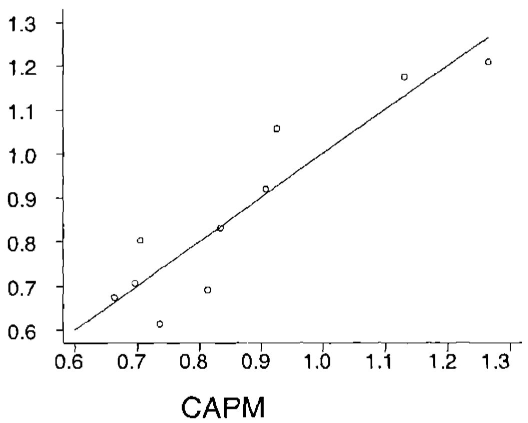
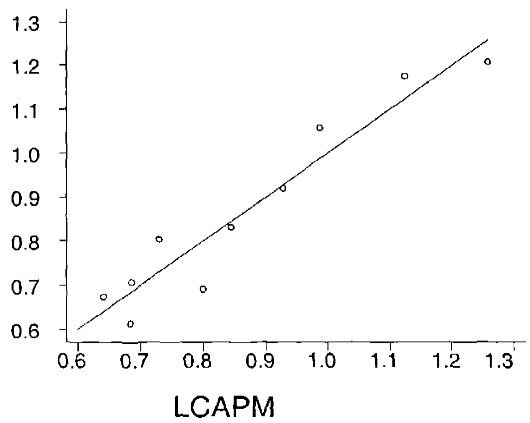


Table 2.6: Cross-sectional Tests (10 B/M porfolios, by subsamples). The table reports estimates g_1 from cross-sectional regressions of average excess returns on betas (for the CAPM), and on the time-series average of the product between the kalman-filtered beta and the excess market return (LCAPM). The regressions also include a constant (g_0). P-values and 5% confidence intervals are provided. The test assets are the ten B/M decile portfolios. The estimation samples is given in the table.

	1963:7-2001:12	
	CAPM	LCAPM
g_0	1.78	-.75
p-value	[.00]	[.07]
conf. int.	(.67; 2.89)	(−1.58; .08)
g_1	-1.24	2.74
p-value	[.04]	[.00]
conf. int.	(−2.42; −.06)	(1.08; 4.40)
adj. R^2	35.47%	59.96%

Figure 2-4: Performance of the CAPM and LCAPM, 1963:7-2001:12. The diagram on the left corresponds to the CAPM, the one on the right to the LCAPM. The horizontal axes correpspond to the predicted average excess returns, and the vertical axes to the sample average excess returns, for the 10 B/M decile portfolios. The predicted average excess returns are based on the regressions in Table 2.6. The sample period is 1931:7-2001:12. The straight line is the 45-degree line.





stocks, the ranking on betas is inverted: high beta for growth and low beta for value. For example, from Panel B of Table 2.3 we learn that the average return on the 10th B/M decile portfolio exceeds the one on the 1st B/M decile portfolio by 6.2% per annum in the 1963:7-2001:12 interval. On the other hand, the estimated beta of the 10th B/M decile portfolio in the same sample is .94, while the beta for the 1st decile portfolio is 1.10. This fact explains why in Panel B of Table 2.6 the coefficient on beta for the CAPM column is negative, in strong violation of the theoretical prediction[10].

In comparison to the rejection of CAPM by the standard tests, our LCAPM achieves its most striking success in the second sub-sample. The LCAPM can explain almost 60% of the cross-sectional variation of returns. Moreover, the constant is insignificantly different from zero at the 5% confidence level. Finally, the theoretical restriction that g_1 is equal to one is not rejected at the 99% confidence level. Consistent with these results, Figure 2-4 shows that for the LCAPM the dots are aligned around the 45-degree line, while for the CAPM they are very dispersed.

The economic intuition for the empirical success of the LCAPM is the following. Taking into account that factor loadings are time-varying and unobserved by investors, we can estimate investors expectations of factor loadings explicitly. Because the beta of value stocks has decreased over time, optimal learning by investors ties current expectations to past levels of factor loadings. Given that a symmetric development that characterizes growth stocks and value stocks, it follows that the ranking in expected betas reflects the average ranking in returns. The evidence that the Kalman filtered beta for the 10th B/M portfolio is above the corresponding series for

[10]The R^2 of 35% is meaningless as a summary measure of the performance of the model, as it is achieved with a negative slope coefficient.

the 1st decile portfolio over most the 1963:7-2001:12 period is supportive of this argument (see Figure 2-1).

The trending behavior of the betas of B/M sorted portfolios make our methodology particularly suitable to test CAPM on this set of assets. However, our approach extends to other assets as well, and the next section presents some evidence in this direction.

We think that our results show the importance of modeling investors' expectations of unobservable factor loadings, and the necessity of incorporating them into asset pricing theory and tests, especially in the presence of high parameter variability. This contribution is relevant for the tests of other asset pricing models as well.

2.4 Robustness and Extensions

2.4.1 Additional tests

The results presented in the previous section 2.3 will be extended in several dimensions.

First, one might think that using the procedure proposed by Fama-MacBeth (1973) to test CAPM accounts for the variability of beta. In Section 2.3, we argued that this procedure is not enough to capture investors' expectations of the factor loading. While the Fama-MacBeth procedure accounts for the variability of beta over time, using rolling windows, it does not incorporate the history of the loading, which matters for investors' expectations when investors are Bayesian and apply the Kalman filter. Using rolling window regressions also does not take into account the fact that betas are latent variables. Here we present the results of applying the Fama-MacBeth procedure to the ten B/M decile portfolios.

For the second sub-sample (1963:7-2001:12), we find that the time-series average of the slope coefficient from cross-sectional CAPM regressions, in which the betas estimated on the prior five years of data are the explanatory variable, is significantly negative (-0.02 percent monthly). Hence, even allowing the estimates of beta to change every month does not prevent the rejection of CAPM.

However, when we use the Kalman-filtered beta series as an explanatory variable, the time-series average of the slope coefficients is 0.70 percent monthly and significant. This estimate implies an 8% risk premium on an annual basis, which is not too far from the observed equity premium. [11]

The conclusion that we draw from this analysis is that allowing for time variation in beta is not enough to account for the better performance of Kalman-filtering methodology. In fact, the LCAPM is also successful because it provides a proxy for investors' expectations of the factor loadings.

It is interesting to assess the LCAPM vis-a-vis the Fama-French three-factor model. On the ten B/M portfolios, the Fama-French model naturally performs well, as the factors replicate the variation in the test portfolios. Consequently, the R^2 for this model is above 90% in all samples. However, when our $\bar{M}^i$ variable is included in the cross-sectional regressions along with the loadings on HML and SMB, the latter two are no longer significant for the whole sample, as well as in the early sub-sample. We take this result as evidence of collinearity between the Fama-French factors and our learning based explanation. However, in the second sub-sample, HML remains significant even when we include our $\bar{M}^i$ variable.

[11] As reported in tables 2.2 and 2.3, the average excess market return is 8.2% annually in the 1931:7-2001:12 sample. It is 11.2% in the 1931:7-1963:6 subsample, and 5.6% in the 1963:7-2001:12 subsample.

2.4.2 Sensitivity of F

A relevant element in the empirical implementation of the LCAMP is the choice of the value for the autoregression coefficient F, which is relevant in the filtering process. We based our choice of F (0.97) on maximum likelihood estimation performed on the whole sample. However, it is not necessarily the case that investors know this value of F ex-ante. Unlike the econometrician, the whole sample of data is not available to the investors.

We test the robustness of our results to different specifications for F. As one might expect in the light of the simulation results, higher values of F reduce the wedge between filtered betas and estimated betas, by increasing the responsiveness to new information. On the other hand, a lower F makes updating less quick, and history matters more. For F equal to 0.98 the R^2 in the whole sample is still a remarkable 90%. In the second sub-sample it falls to 33%, but the coefficient on g_1 is still significantly positive, and the constant is insignificant. Therefore, for F equal to 0.98, LCAPM performs well in explaining the cross-section of B/M stocks. When F takes on values larger than 0.98, the LCAPM still performs well over the whole sample with an R^2 of about 90%, but it has no explanatory power in the second sub-sample. On the other hand, F equal to 0.95 greatly the ability of the LCAPM to explain the cross-section of average returns. In the longer sample, the R^2 is about 91%, while in the more recent sample, it jumps up to 77%. These results highlight the role that the speed of learning plays in our explanation, and they are in line with the simulation results presented in section 2.2.6.

2.4.3 Book-to-Market and Size

Finally, an important extension of the results from the previous section consists of testing the LCAPM on a different set of assets. Besides the value premium, another famous anomaly is the 'size effect' (see Fama and French, 1992), namely the fact that small stocks pay a significantly positive premium relative the expected return predicted by CAPM. Although this anomaly disappeared after its first publication in the eighties (Banz, 1981), it represents a challenge to CAPM. Following Campbell and Vuolteenaho (2002), we focus on only 24 of the 25 portfolios constructed by Davis, Fama, and French (2000), leaving out the small-growth portfolio. The small growth portfolio is an outlier in most models, including the learning CAPM. In fact, recent evidence by Lamont and Thaler (2001), Mitchell, Pulvino, and Stafford (2002), and D'Avolio (2002) suggests that the equilibrium expected returns on this portfolio are heavily affected by short-sale constraints and other limits to arbitrage. These elements are not accounted for by our theory, which is centered around the role of learning.

The results of the cross-sectional tests for the whole sample are presented in Table 2.7, while Tables 2.8 and 2.9 cover the 1931:7-1963:6 and 1963:7-2001:12 intervals. As in the case of the ten B/M portfolios, the traditional CAPM performs well in the whole sample. Still, our approach improves the R^2 by about 10%, while respecting the restrictions on the slope and intercept imposed by the theory.

Consistent with the results in Section 2.3, the most relevant contribution of our methodology occurs in the modern sub-sample (1963:7-2001:12). While the traditional version of CAPM has no explanatory power in that period, our LCAPM can account for about 34% of the cross-sectional variation in average returns. Again, the predictions that the theory imposes on this model are not rejected by the data. The constant is insignificantly different from zero, and we do not reject the hypothesis that g_1 equals one. While the results for the 24 portfolios extend the scope of the LCAPM explanatory power, they confirm the important contribution made by this testing approach.

Table 2.7: Cross-sectional Tests (24 B/M and Size sorted portfolios, whole sample). The table reports estimates g_1 from cross-sectional regressions of average excess returns on betas (for the CAPM), and on the time-series average of the product between the kalman-filtered beta and the excess market return (LCAPM). The regressions also include a constant (g_0). P-values and 5% confidence intervals are provided. The test assets are 24 B/M and size sorted portfolios, i.e. the 25 portfolios of Davis, Fama, and French (2002), minus the small-growh portfolio. The estimation samples are given in the table.

	1931:7-2001:12	
	CAPM	LCAPM
g_0	-.24	-.39
p-value	[.40]	[.14]
conf. int.	$(-.82; .34)$	$(-.92; .14)$
g_1	1.02	1.47
p-value	[.00]	[.00]
conf. int.	$(.54; 1.50)$	$(.92; 2.03)$
adj. R^2	45.15%	55.91%

The results from the whole set of 25 B/M and size sorted portfolios are less strong, but they confirm the success of the LCAPM, as it appears from Tables 2.10, 2.11 and 2.12. In the long sample the R^2 for LCAPM is 33% (14% for CAPM), the constant is insignificantly different from zero at the 5% level, and g_1 is not different from one. The same occurs in the early subsample, where the R^2 is 43% (39% for CAPM). In the modern subsample, the R^2 is 7.3%, which is definitely an improvement over CAPM, for which a significantly negative slope coefficient explains the cross-section of average returns. Furthermore, the constant in the LCAPM regression is not significantly different from zero, and the constraint on g_1 is not rejected at the 95% confidence level. In this sample, g_1 is only significant with 90% confidence.

Table 2.8: Cross-sectional Tests (24 B/M and Size sorted portfolios, by subsamples). The table reports estimates g_1 from cross-sectional regressions of average excess returns on betas (for the CAPM), and on the time-series average of the product between the kalman-filtered beta and the excess market return (LCAPM). The regressions also include a constant (g_0). P-values and 5% confidence intervals are provided. The test assets are 24 B/M and size sorted portfolios, i.e. the 25 portfolios of Davis, Fama, and French (2002), minus the small-growh portfolio. The estimation samples are given in the table.

	1931:7-1963:6	
	CAPM	LCAPM
g_0	.13	-.11
p-value	[.52]	[.69]
conf. int.	(−.30; .58)	(−.68; .45)
g_1	.92	1.07
p-value	[.00]	[.00]
conf. int.	(.59; 1.24)	(.66; 1.48)
adj. R^2	58.88%	55.39%

Table 2.9: Cross-sectional Tests (24 B/M and Size sorted portfolios, by subsamples). The table reports estimates g_1 from cross-sectional regressions of average excess returns on betas (for the CAPM), and on the time-series average of the product between the kalman-filtered beta and the excess market return (LCAPM). The regressions also include a constant (g_0). P-values and 5% confidence intervals are provided. The test assets are 24 B/M and size sorted portfolios, i.e. the 25 portfolios of Davis, Fama, and French (2002), minus the small-growh portfolio. The estimation samples are given in the table.

	1963:7-2001:12	
	CAPM	LCAPM
g_0	1.12	-.59
p-value	[.00]	[.12]
conf. int.	(.55; 1.69)	(−1.36; .17)
g_1	-.38	2.19
p-value	[.15]	[.00]
conf. int.	(−.93; .15)	(.91; 3.46)
adj. R^2	5.00%	33.83%

Table 2.10: Cross-sectional Tests (25 B/M and Size sorted portfolios, whole sample). The table reports estimates g_1 from cross-sectional regressions of average excess returns on betas (for the CAPM), and on the time-series average of the product between the kalman-filtered beta and the excess market return (LCAPM). The regressions also include a constant (g_0). P-values and 5% confidence intervals are provided. The test assets are the 25 portfolios of Davis, Fama, and French (2002). The estimation samples is 1931:7-2001:12.

	1931:7-2001:12	
	CAPM	LCAPM
g_0	.28	−.10
p-value	[.37]	[.73]
conf. int.	(−.37; .95)	(−.75; .53)
g_1	.57	1.15
p-value	[.03]	[.00]
conf. int.	(.04; 1.09)	(.48; 1.82)
adj. R^2	14.36	32.91

Table 2.11: Cross-sectional Tests (25 B/M and Size sorted portfolios, by subsamples). The table reports estimates g_1 from cross-sectional regressions of average excess returns on betas (for the CAPM), and on the time-series average of the product between the kalman-filtered beta and the excess market return (LCAPM). The regressions also include a constant (g_0). P-values and 5% confidence intervals are provided. The test assets are the 25 portfolios of Davis, Fama, and French (2002). The estimation samples are given in the table.

	1931:7-1963:6	
	CAPM	LCAPM
g_0	.37	.05
p-value	[.13]	[.85]
conf. int.	(−.13; .87)	(−.56; .67)
g_1	.72	.94
p-value	[.00]	[.00]
conf. int.	(.36; 1.09)	(.49; 1.38)
adj. R^2	39.79	43.16

Table 2.12: Cross-sectional Tests (25 B/M and Size sorted portfolios, by subsamples). The table reports estimates g_1 from cross-sectional regressions of average excess returns on betas (for the CAPM), and on the time-series average of the product between the kalman-filtered beta and the excess market return (LCAPM). The regressions also include a constant (g_0). P-values and 5% confidence intervals are provided. The test assets are the 25 portfolios of Davis, Fama, and French (2002). The estimation samples are given in the table.

	1963:7-2001:12	
	CAPM	LCAPM
g_0	1.26	−.02
p-value	[.00]	[.95]
conf. int.	(.74; 1.79)	(−.91; .86)
g_1	−.53	1.20
p-value	[.03]	[.10]
conf. int.	(−1.03; −.04)	(−.25; 2.66)
adj. R^2	14.64	7.32

2.5 Related literature

Our model incorporates insights from different strands of the empirical and theoretical asset pricing literature. Fama and French (1993, 1995, 1996) propose a risk-based explanation to resolve the value and size puzzles. By including size and B/M factors as proxies for some underlying distress risk, they show that they can account for a large fraction of the cross-sectional variation of stocks. Fama and French motivate their findings with multifactor asset pricing models, such as Merton's (1971) ICAPM, and Ross's (1976) APT. Vassalou and Xing (2002) examine the link between the SMB and HML factors and distress risk in detail. They conclude that "although SMB and HML contain default-related information, this is not the reason that the FF model can explain the cross-section. SMB and HML appear to contain important price information, unrelated to default risk." (Vassalou and Xing 2002). We provide an explanation for this finding. In our framework, deviations from CAPM such as the significance of the SMB and HML factors detected in ex-post tests stem from the fact that the systematic riskiness of value and small stocks is unobserved and variable. We postulate that market risk is priced ex-ante only.

The behavioral finance literature suggests that characteristics, rather than risks, are priced in equilibrium. Daniel and Titman (1997), for example, argue that investors have an exaggerated perception of the riskiness of small and value stocks, and require a premium to hold them. Lakonishok, Shleifer and Vishny (1994) and LaPorta, Lakonishok Shleifer, and Vishny (1997) provide evidence that investors have biased expectations about the earnings of value and growth stocks. Our framework explains why the perception of risk for value stocks is understated by usual CAPM regressions.

More recently, Lettau and Ludvigson (2001) have produced favorable evidence for the Consumption CAPM of Breeden (1979) and Lucas (1978). In particular, they argue that the CCAPM holds under conditional probability distributions. Hence, the correct empirical implementation of the model requires the use of scaled factors, i.e. risk factors multiplied by appropriate state variables. Our theory has elements of the Lettau and Ludvigson's (2001) explanation, and more generally of the conditional CAPM literature, as it stresses the importance of studying conditional moments of the return distribution. Lettau and Ludvigson build on the important contribution of Jaganathan and Wang (1996), who estimate a conditional CAPM with labor income as conditioning variable.[12] Campbell and Cochrane (2000) explain the cross-section of returns with conditioning variables that result from a habit-formation model.[13] A recent paper that applies a conditional CAPM approach to focus exclusively on the value premium is Petkova and Zhang (2002).

[12] See Santos and Veronesi (2001) for a more recent version of using labor income as conditioning varible.

[13] Habit formation models were treated by Constantinides (1990) and Sundaresan (1989), and more recently by Menzly, Santos and Veronesi (2002).

A number of recent papers make the distinction between ex-ante expectations and ex-post measurements of returns. Fama and French (2002) extend work by Blanchard (1993) to estimate ex-ante expectations of the equity premium, pointing out that average equity returns are likely to be above ex-ante expectations. Elton (1999) warns of the dangers of equating average returns with expected returns, and gives a number of examples where average realized returns are clearly different from expected returns. Brav, Lehavy, and Michaely (2002) examine the expectations of stock analysts by analyzing earnings forecasts. The authors find that CAPM cannot be rejected for these expectations. Furthermore, size and B/M factors are insignificant for stock analyst expectations. This evidence is in the spirit of our model, as we argue that expected returns are driven by CAPM, and the size and B/M factors proxy for the difference between ex-post measures and ex-ante expectations. Other recent evidence by Doukas, Kim and Pantzalis (2002) shows that expectations of future earnings of stock analysts are unbiased for value stocks and small stocks, implying that the abnormal returns of value stocks on earnings announcement days found by LaPorta, Lakonishok Shleifer, and Vishny (1997) are caused by a different mechanism. We are suggesting a possible mechanism: it might not be a surprise in the level in earnings, but the news about beta that has caused the excess returns of value stocks.

The reduced form of the general equilibrium pricing model that we derive is a Kalman filter. Some of the early proponents of applying Kalman filtering to economics are Engle and Watson (1987). The reduced form of our pricing model is a GARCH model, resembling the one estimated by Bollerslev, Engle, and Wooldridge (1988). The resemblance to GARCH models is directly resulting from our focus on modelling time-varying betas explicitly. Moskowitz (2002) estimates a GARCH model, including estimates of time-varying covariances of stocks with the value and size factors. Campbell and Hentschel (1992) is another example of a model that gives a GARCH specification in the reduced form.

Learning in a symmetric information setting has recently attracted a considerable amount of attention in the academic research literature[14]. Brennan and Xia (2001) address the equity premium puzzle in a learning model. In their model, the drift of the aggregate dividend process is unobservable. Investors update their belief about the true drift of the dividend process, which leads to a learning premium in aggregate asset prices. In our model, we focus on cross-sectional asset pricing anomalies, and calibrate parameters, in particular risk aversion, to reproduce the observed equity premium. The aggregate equity premium is not the focus of this chapter. Veronesi (1999) also examines the impact of latent variables on the aggregate stock market: unobserved switches from booms to recessions exacerbate the volatility of the stock market.

A paper that is closely related to ours is the one by Lewellen and Shanken (2002). The authors assume that the mean of the dividend process is unobserved. Their paper accounts for both predictability and excess variance of returns (Shiller, 1981). Ex-ante investors use all the available information, so there are no arbitrage opportunities. Predictability is observable ex-post, and is driven by the Bayesian updating of investors' beliefs about the unobservable mean of the dividend. The authors demonstrate that an econometrician will detect excess volatility as a consequence of learning, as new realizations of dividends cause investors to constantly revise their beliefs. Lewellen and Shanken explain cross-sectional anomalies as the result of learning on fundamentals. Our approach, instead, focuses on learning about riskiness, in an environment with time-varying factor loadings.

Other papers have focused on the portfolio allocation problem for investors with incomplete

[14] Earlier literature in learning include Genotte (1986) treats portfolio choice with parameter uncertainty, and Bawa, Brown and Klein (1979) examine the influence of estimation risk on portfolio choice.

information. In Barberis (2000), investors are unsure whether returns are predictable or not. This uncertainty leads to an excessive allocation of wealth to stocks compared to the market, and this allocation is larger, the longer the investment horizon. Barberis (2000) does not address cross-sectional implications of incomplete information, which is the focus of this chapter. Finally, Pastor (2000) examines the Bayesian decision problem of investors who are unsure whether CAPM holds. One of the alternatives to the CAPM that Pastor (2000) considers is the multifactor model proposed by Fama and French (1993). Pastor (2000) finds that even if an investor strongly believes that the market portfolio is mean variance efficient, he should invest a substantial amount of her wealth in value stocks. Unlike this chapter, Pastor (2000) does not examine the general equilibrium implications of Bayesian investors, which is the focus of this chapter.

There are two recent papers that exploit the time-variation in beta explicitly. Campbell and Vuolteenaho (2002) decompose the time-variation in betas into variation driven by dividend news of individual stocks, and variation driven by expected return news by the market portfolio. This approach is complementary to ours, as we abstract completely from time-variation in expected market returns. In our approach, it is the time variation in expected returns of individual stocks that is driving the results. The work of Ang and Chen (2002) is closely related to ours, as they take the time-variation in betas explicitly into account. Using Bayesian statistics, Ang and Chen estimate the parameters of the time-series process for betas. Then, through a bootstrapping procedure they show that the CAPM is not rejected on B/M sorted portfolios, once time variation in betas is taken into account. They make a statistical argument that is complementary to ours. Whereas Ang and Chen focus on the correction to standard that needs to be done to account for the time-variation in betas, we focus on the implication for the level of betas. In particular, we propose to estimate betas using the Kalman filter, as implied by the general equilibrium asset pricing model that we study. Our conclusions are consistent with the ones by Ang and Chen (2002), as we find, as them, that CAPM holds once the variability and unobservability of betas is taken explicitly into account.

Our approach gives a new perspective on the argument made by Roll and Ross (1994) concerning the mean-variance efficiency of the market portfolio. Roll and Ross (1994) argue that because the true market portfolio is not observable, what is used as a proxy to the market portfolio in empirical studies must not lie on the mean-variance efficient frontier.[15] Put differently, even if the true market portfolio lies on the mean variance efficient frontier, that does not imply that the proxy of stocks used in empirical tests lies on the frontier. The measurement error introduced by not observing the true portfolio makes it impossible to reject mean-variance efficiency. We point out that an additional reason why measurement error in mean-variance space is introduced is the time-variation of betas. Once this time-variation of betas is taken into account, we show that we cannot reject that the market portfolio is mean-variance efficient, i.e. betas proxy for expected returns, and the alpha in cross-sectional regressions is statistically zero.

2.6 Conclusions

The assumption that investors know the variance-covariance structure of asset returns is implausible if the parameters of the model vary over time. Dramatic changes in factor loadings, like the ones documented by Franzoni (2002) for value and small stocks, cause investors to

[15] Our results also apply to MacKinley (1994).

continuously revise their expectations of the riskiness of assets. Depending on the speed of learning, and on the amount of noise in returns, these expectations can diverge significantly from the true level.

We have developed an equilibrium model of learning about risk factor loadings. In the model, CAPM holds from investors' point of view, but the econometrician observes mispricing when the expected factor loadings diverge from the actual level of systematic risk.

The simulations of the model show that in order to produce levels of mispricing that are close to the value premium, the (optimal) learning process of investors needs to be such that they put sufficient weight on past observations. This is the case when factor loadings follow a mean reverting process, and when investors do not know the long run mean of the loading.

The first condition assures that expectations of the loading are fairly stable in the face of new information, as they tend to adhere to the long run mean. The second condition causes the belief of the loading to decrease slowly, while investors learn about the long run mean as well.

We propose an approach for testing the CAPM that is consistent with the implications of the model, and we label the resulting empirical model Learning CAPM (LCAPM).

The idea behind this testing strategy is that investors' expectations of the factor loading determine equilibrium expected returns. Therefore, the econometrician who is testing a pricing model, needs to replicate the filtering process that investors undertake. In particular, the econometrician should not use more information than the one available to investors. Also, this information needs to be processed in an 'optimal' way.

We translate these prescriptions into practice by obtaining series of portfolio betas from Kalman filtering realized returns. Then, we use these filtered series to test the CAPM prediction that beta explains the cross-section of expected returns.

The fundamental result of this chapter is that when we apply this methodology to the ten B/M decile portfolios, the learning augmented version of CAPM is not rejected by the data. This conclusion also holds in the last forty years of data, in which the standard tests systematically reject CAPM. Therefore, we conclude that learning provides an account for the value premium. We have presented a extensions of these results. In particular, the LCAPM also performs well in explaining the cross-section of average returns on portfolios sorted on size, as well as on B/M.

2.7 Appendix

2.7.1 Derivation of the Kalman-Filter

Let us introduce the following notation:

$$\xi_t^i = \begin{pmatrix} B^i \\ b_t^i \end{pmatrix} \quad \tilde{F}^i = \begin{pmatrix} 1 & 0 \\ 1 & F^i \end{pmatrix} \quad H_t = \begin{pmatrix} 0 \\ x_t \end{pmatrix} \quad U_{t+1}^i = \begin{pmatrix} 0 \\ u_{t+1}^i \end{pmatrix}$$

Then equations (2.1) and (2.2) can be written as:

$$\begin{aligned} \xi_{t+1}^i &= \tilde{F}\xi_t^i + U_{t+1}^i \ \ \forall i \\ D_t^i - \bar{D}^i &= H_t'\xi_t^i + \varepsilon_t^i \ \ \forall i \end{aligned}$$

Furthermore denote the variance-covariance matrix of the forecast error as follows:

$$\Gamma_{t+1|t}^i = E\left[\left(\xi_{t+1}^i - E\left[\xi_{t+1}^i|\Im_t\right]\right)\left(\xi_{t+1}^i - E\left[\xi_{t+1}^i|\Im_t\right]\right)'|\Im_t\right]$$

With this notation, the Kalman-Filter from Hamilton (1994, chap. 13) or Liptser and Shiryaev (2000, chap. 14) can be directly applied:

$$E\left[\xi_{t+1}^{i}|\Im_{t}\right] = \tilde{F}^{i}E\left[\xi_{t}^{i}|\Im_{t-1}\right]+\kappa_{t}^{i}\left(D_{t}^{i}-\bar{D}^{i}-H_{t}'E\left[\xi_{t}^{i}|\Im_{t-1}\right]\right) \tag{2.15}$$

$$\kappa_{t}^{i} = \tilde{F}^{i}\Gamma_{t|t-1}^{i}H_{t}\left(H_{t}'\Gamma_{t|t-1}^{i}H_{t}+\sigma_{\varepsilon}^{i2}\right)^{-1}$$

Now, using the notation introduced in Section 2.2.2, $b_{t+1|t}^{ie} = E\left[b_{t+1}^{i}|\Im_{t}\right]$ and $B_{t}^{ie} = E\left[B^{i}|\Im_{t}\right]$. Using this in 2.15, we obtain:

$$\begin{bmatrix} B_{t}^{ie} \\ b_{t+1|t}^{ie} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & F^{i} \end{bmatrix}\begin{bmatrix} B_{t-1}^{ie} \\ b_{t|t-1}^{ie} \end{bmatrix} + \kappa_{t}^{i}\left(D_{t}^{i}-\bar{D}^{i}-x_{t}b_{t|t-1}^{ie}\right)$$

where

$$\kappa_{t}^{i} = \begin{bmatrix} 1 & 0 \\ 1 & F^{i} \end{bmatrix}\Gamma_{t|t-1}^{i}\begin{bmatrix} 0 \\ x_{t} \end{bmatrix}\left(\begin{bmatrix} 0 \\ x_{t} \end{bmatrix}'\Gamma_{t|t-1}^{i}\begin{bmatrix} 0 \\ x_{t} \end{bmatrix}+\sigma_{\varepsilon}^{i2}\right)^{-1}$$

Or, writing each updating equation separately:

$$B_{t}^{ie} = B_{t-1}^{ie}+K_{t}^{i}\left(D_{t}^{i}-\bar{D}^{i}-x_{t}b_{t|t-1}^{ie}\right) \tag{2.16}$$

$$b_{t+1|t}^{e} = B_{t-1}^{ie}+F^{i}b_{t|t-1}^{ie}+k_{t}^{i}\left(D_{t}^{i}-\bar{D}^{i}-x_{t}b_{t|t-1}^{ie}\right) \tag{2.17}$$

where

$$K_{t}^{i} = \frac{\gamma_{t|t-1}^{i(1,2)}}{\gamma_{t|t-1}^{i(2,2)}x_{t}^{2}+\sigma_{\varepsilon}^{i2}}x_{t} \qquad k_{t+1}^{i} = \frac{\gamma_{t|t-1}^{i(1,2)}+\gamma_{t|t-1}^{i(2,2)}F^{i}}{\gamma_{t|t-1}^{i(2,2)}x_{t}^{2}+\sigma_{\varepsilon}^{i2}}x_{t}$$

and $\gamma_{t|t-1}^{i(q,r)}$ is are the (q-th, r-th) element of the matrix $\Gamma_{t|t-1}^{i}$ and: $\kappa_t^i = \begin{pmatrix} K_t^i & k_t^i \end{pmatrix}'$. From Hamilton (1994), we find the evolution of the forecast error:

$$\Gamma_{t+1|t}^{i} = \left(\tilde{F}^i - \kappa_t^i H_t^{i\prime}\right) \Gamma_{t|t-1}^{i} \left(\tilde{F}^{i\prime} - H_t^i \kappa_t^{i\prime}\right) + \sigma_\varepsilon^2 \kappa_t^i \kappa_t^{i\prime} + \begin{pmatrix} 0 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}$$

2.7.2 Proof of Proposition 2.1

Before the proof of Proposition 2.1, we need the following Lemma:

Lemma 2.1 *Under the distributional assumptions of sections 2.1.-2.3.,*

$$c_{t+1} \to \bar{D}^m + x_{t+1} + y_{t+1} \quad \text{as } N \uparrow \infty$$

where $\bar{D}^m = \sum_{i=1}^{N} \bar{\mathbf{a}}^i \bar{D}^i$

Proof. Lemma 2.1

Solving the process for b recursively gives:

$$b_t^i = \left(F^i\right)^t b_0^i + \sum_{s=1}^{t} \left[\left(F^i\right)^{s-1} \left(B^i + u_{t-s+1}^i\right)\right]$$

Let us first consider the case $F^i = 1 \;\; \forall i$. Taking the market-weighted sum gives:

$$\sum_{i=1}^{N} \bar{\mathbf{a}}^i b_t^i = \sum_{i=1}^{N} \bar{\mathbf{a}}^i b_0^i + \sum_{s=1}^{t} \left[\left(\sum_{i=1}^{N} \bar{\mathbf{a}}^i B^i + \sum_{i=1}^{N} \bar{\mathbf{a}}^i u_{t-s+1}^i\right)\right]$$

It is assumed that $\bar{\mathbf{a}}^i = O\left(N^{-1}\right)$, which implies that we can define $\tilde{\mathbf{a}}^i = \bar{\mathbf{a}}^i N$ with the property that

$$\sum_{i=1}^{N} \tilde{\mathbf{a}}^i < C < \infty \quad \text{as } N \uparrow \infty$$

for some constant C. Furthermore, it has been assumed that the distribution of shares outstanding, $\bar{\mathbf{a}}^i$, is independent of the shocks ε_t^i, u_t^i, and x_t for all t and all i. The cross-sectional expectation of $\tilde{\mathbf{a}}^i$ is finite: $E\left[\tilde{\mathbf{a}}^i\right] < \infty$. Initially, $\sum_{i=1}^{N} \bar{\mathbf{a}}^i b_0^i = 1$ and $\sum_{i=1}^{N} \tilde{\mathbf{a}}^i B^i = 0$ when $F^i = 1$. Using these assumptions, we get:

$$\begin{aligned}
\sum_{i=1}^{N} \bar{\mathbf{a}}^i b_t^i &= 1 + \sum_{s=1}^{t} \left[\left(\frac{1}{N} \sum_{i=1}^{N} \tilde{\mathbf{a}}^i B^i + \frac{1}{N} \sum_{i=1}^{N} \tilde{\mathbf{a}}^i u_{t-s+1}^i \right) \right] \\
&\overset{N\uparrow\infty}{\longrightarrow} 1 + \sum_{s=1}^{t} E\left[\tilde{\mathbf{a}}^i u_{t-s+1}^i\right] = 1 + \sum_{s=1}^{t} E\left[\tilde{\mathbf{a}}^i\right] E\left[u_{t-s+1}^i\right] = 1
\end{aligned}$$

The case when $F < 1$ is similar. Recall from Section 2.1. that it is assumed that there exists a transformation of B^i, such that $\tilde{B}^i\left(1 - F^i\right) = B^i$, and $\sum_{i=1}^{N} \bar{\mathbf{a}}^i \tilde{B}^i = 1$. We can therefore write:

$$\begin{aligned}
b_t^i &= \left(F^i\right)^t b_0^i + \sum_{s=1}^{t} \left[\left(F^i\right)^{s-1} \left(\tilde{B}^i \left(1 - F^i\right) + u_{t-s+1}^i \right) \right] \\
&= \tilde{B}^i + \left(F^i\right)^t \left(b_0^i - \tilde{B}^i \right) + \sum_{s=1}^{t} \left(F^i\right)^{s-1} u_{t-s+1}^i
\end{aligned}$$

Recall that it is assumed that the economy is in the steady state at time 0, so that $b_0^i = \tilde{B}^i$. Taking the market weighted sum gives:

$$\sum_{i=1}^{N} \bar{\mathbf{a}}^i b_t^i = \sum_{i=1}^{N} \bar{\mathbf{a}}^i \tilde{B}^i + \sum_{s=1}^{t} \left(\sum_{i=1}^{N} \bar{\mathbf{a}}^i \left(F^i\right)^{s-1} u_{t-s+1}^i \right) \overset{N\uparrow\infty}{\longrightarrow} 1$$

Therefore we conclude that: $\sum_{i=1}^{N} \bar{\mathbf{a}}^i b_t^i \to 1 \quad \text{as } N \uparrow \infty \ \ \forall t.$

From this, it follows immediately that the market dividend, D_{t+1}^m is:

$$D_{t+1}^m = \sum_{i=0}^{N} \bar{\mathbf{a}}^i D_{t+1}^i = \sum_{i=0}^{N} \bar{\mathbf{a}}^i \left(\bar{D}^i + b_t^i x_t + \varepsilon_t^i \right) \overset{N \uparrow \infty}{\longrightarrow} \bar{D}^m + x_{t+1}$$

as $\sum_{i=0}^{N} \bar{\mathbf{a}}^i \varepsilon_t^i = \frac{1}{N} \sum_{i=0}^{N} \tilde{\mathbf{a}}^i \varepsilon_t^i \to E\left[\tilde{\mathbf{a}}^i \varepsilon_t^i\right] = E\left[\tilde{\mathbf{a}}^i\right] E\left[\varepsilon_t^i\right] = 0.$

Taking all these results together shows that total consumption is:

$$\begin{aligned} c_{t+1} &= y_{t+1} + \sum_{i=0}^{N} \bar{\mathbf{a}}^i D_{t+1}^i \\ &= y_{t+1} + \sum_{i=0}^{N} \bar{\mathbf{a}}^i \bar{D}^i + \sum_{i=0}^{N} \bar{\mathbf{a}}^i b_t^i x_t + \sum_{i=0}^{N} \bar{\mathbf{a}}^i \varepsilon_t^i \\ &\to y_{t+1} + \bar{D}^m + x_{t+1} \end{aligned}$$

■

With this lemma, the pricing function from proposition 2.1 can be derived:

Proof. Proposition 2.1:

The FOC for investor's optimization problem is:

$$E\left[e^{-Ac_{t+1}} \left(D_{t+1} + P_{t+1} - (1+r) P_t\right) | \Im_t\right] = 0$$

Replacing for total consumption by application of Lemma 2.1 gives:

$$(1+r) P_t^i E\left[e^{-Ax_{t+1} - A\left(\bar{D}^m + y_{t+1}\right)} | \Im_t\right] = E\left[e^{-Ax_{t+1} - A\left(\bar{D}^m + y_{t+1}\right)} \left(D_{t+1}^i + P_{t+1}^i\right) | \Im_t\right]$$

As labor income y and $\bar{D}^m$ are independent of x, and both the dividend process and the variables in the pricing function are independent of labor income, this reduces to:

$$(1+r) P_t^i E\left[e^{-Ax_{t+1}}|\Im_t\right] = E\left[e^{-Ax_{t+1}}\left(D_{t+1}^i + P_{t+1}^i\right)|\Im_t\right]$$

A linear pricing function is:

$$P_t^i = \nu \bar{D}^i + \omega^i b_{t+1|t}^{ie} + \upsilon^i B_t^{ie}$$

Replacing for the guess of the pricing function gives:

$$E\left[e^{-Ax_{t+1}}\left(x_{t+1}b_{t+1}^i + \varepsilon_{t+1}^i + \bar{D}^i\left(1+\nu^i\right) + \omega^i b_{t+2|t+1}^{ei} + \upsilon^i B_{t+1}^{ei}\right)|\Im_t\right] = (1+r) P_t^i E\left[e^{-Ax_{t+1}}|\Im_t\right]$$

By the Law of Iterated Expectations and the fact that b_{t+1}^i and x_{t+1} are independent, it follows that:

$$E\left[e^{-Ax_{t+1}}x_{t+1}b_{t+1}^i|\Im_t\right] = E\left[e^{-Ax_{t+1}}x_{t+1}\right] E\left[b_{t+1}^i|\Im_t\right] = E\left[e^{-Ax_{t+1}}x_{t+1}|\Im_t\right] b_{t+1|t}^{ie}$$

Using this together with $E\left[\varepsilon_{t+1}^i|\Im_t, x_{t+1}\right] = 0$, the pricing equation reduces to:

$$\begin{aligned} & (1+r) P_t^i E\left[e^{-Ax_{t+1}}|\Im_t\right] \\ = \;& E\left[e^{-Ax_{t+1}}\left(x_{t+1}b_{t+1|t}^{ie} + \bar{D}^i\left(1+\nu^i\right) + \omega^i b_{t+2|t+1}^{ie} + \upsilon B_{t+1}^{ie}\right)\Im_t\right] \end{aligned}$$

Using the updating equations (2.16) and (2.17), we can replace for $E\left[B_{t+1}^{ie}|\Im_t\right] = B_t^{ie}$ and $E\left[b_{t+2|t+1}^{ie}|\Im_t\right] = B_t^{ie} + F^i b_{t+1|t}^{ie}$:

$$
\begin{aligned}
& E\left[e^{-Ax_{t+1}}\left(x_{t+1}b_{t+1|t}^{ie} + \bar{D}^i\left(1+\nu^i\right) + \omega^i\left(B_t^{ie} + F^i b_{t+1|t}^{ie}\right) + \upsilon^i B_t^{ie}\right)|\Im_t\right] \\
= \; & \left(1+r\right)P_t^i E\left[e^{-Ax_{t+1}}|\Im_t\right]
\end{aligned}
$$

Now we use again the iterated expectations, and the independence between b_{t+1} and x_{t+1}, this equation simplifies:

$$
\begin{aligned}
& \left\{\bar{D}^i\left(1+\nu^i\right) + \omega^i\left(B_t^{ie} + F^i b_{t+1|t}^{ie}\right) + \upsilon^i B_t^{ie}\right\} E\left[e^{-Ax_{t+1}}|\Im_t\right] + b_{t+1|t}^{ie} E\left[e^{-Ax_{t+1}}x_{t+1}|\Im_t\right] \\
= \; & \left(1+r\right)P_t^i E\left[e^{-Ax_{t+1}}|\Im_t\right]
\end{aligned}
$$

Computing the expectations shows:

$$
E\left[e^{-Ax_{t+1}}|\Im_t\right] = e^{\frac{1}{2}A^2\sigma_x^2}
$$

and

$$
\begin{aligned}
E\left[e^{-Ax_{t+1}}x_{t+1}|\Im_t\right] &= \frac{1}{\sqrt{2\pi\sigma_x^2}}\int_{-\infty}^{\infty} x_{t+1}\exp -Ax_{t+1}\exp\left(-\frac{x_{t+1}^2}{2\sigma_x^2}\right)dx_{t+1} \\
&= e^{\frac{1}{2}A^2\sigma_x^2}\frac{1}{\sqrt{2\pi\sigma_x^2}}\int_{-\infty}^{\infty} x_{t+1}\exp\left(-\frac{1}{2\sigma_x^2}\left(x_{t+1}^2 + 2\sigma_x^2 Ax_{t+1} + \sigma_x^2 A^2\right)\right)dx_{t+1} \\
&= -e^{\frac{1}{2}A^2\sigma_x^2}\sigma_x^2 A
\end{aligned}
$$

and replacing into the pricing equation gives:

$$P_t^i = \bar{D}^i \frac{1+\nu^i}{1+r} + \frac{\left(\omega^i F^i - \sigma_x^2 A\right)}{1+r} b_{t+1|t}^{ie} + \frac{\omega^i + v^i}{1+r} B_t^{ie}$$

By the method of undetermined coefficients, we obtain three equations in three unknowns:

$$\begin{aligned} 1+\nu^i &= (1+r)\,\nu^i \\ \omega^i F^i - \sigma_x^2 A &= (1+r)\,\omega^i \\ \omega + v^i &= (1+r)\,v^i \end{aligned}$$

Therefore the pricing function becomes:

$$P_t^i = \frac{\bar{D}^i}{r} - \frac{A\sigma^2}{1+r-F^i}\left(b_{t+1|t}^{ie} + \frac{B_t^{ie}}{r}\right)$$

■

2.7.3 Proof of Proposition 2.2

Rewriting equation (2.7) gives:

$$R_{t+1}^i = b_{t+1}^i \left(x_{t+1} + \sigma_x^2 A\right) + \left(1 + \frac{\left(k_{t+1}^i + K_{t+1}^i / r\right)}{1+r-F^i}\right)\left(\sigma_x^2 A x_{t+1}\left(b_{t+1|t}^{ei} - b_{t+1}^i\right) + \varepsilon_{t+1}^i\right)$$

Summing the last term in this equation gives:

$$\frac{1}{N}\sum_{i=0}^{N} \tilde{\mathrm{a}}^i \left(b_{t+1|t}^{ei} - b_{t+1}^i\right) \longrightarrow 0 \text{ as } N \uparrow \infty$$

This follows directly from the assumption that innovations in b^i are independent across assets. Furthermore, the assumptions about the distribution of F^i together with the properties of the Kalman filter give:

$$\begin{aligned}\sum_{i=0}^{N} \tilde{\mathbf{a}}^i R_{t+1}^i &= \frac{1}{N}\sum_{i=0}^{N} \tilde{\mathbf{a}}^i b_{t+1}^i x_{t+1} + \frac{1}{N}\sum_{i=0}^{N} \tilde{\mathbf{a}}^i \varepsilon_{t+1}^i \\ &+ \frac{1}{N}\sum_{i=0}^{N} \tilde{\mathbf{a}}^i \left(1 + \frac{\left(k_{t+1}^i + K_{t+1}^i/r\right)}{1+r-F^i}\right)\left(\sigma_x^2 A x_{t+1}\left(b_{t+1|t}^{ei} - b_{t+1}^i\right) + \varepsilon_{t+1}^i\right) \\ &\overset{N\uparrow\infty}{\longrightarrow} x_{t+1} + \sigma_x^2 A\end{aligned}$$

where results from Lemma 2.1 were used.

■

2.7.4 Proof of Proposition 2.3

Taking conditional expectations of the return equation (2.6) gives:

$$\begin{aligned}E\left[R_{t+1}^i|\Im_t\right] &= E\left[b_{t+1}^i x_{t+1}|\Im_t\right] + E\left[\varepsilon_{t+1}^i|\Im_t\right] \\ &+ \frac{\sigma_x^2 A}{1+r-F^i}\left(b_{t+1|t}^{ei} - E\left[b_{t+2|t+1}^{ei}|\Im_t\right] + \left(B_t^{ei} - E\left[B_{t+1}^{ei}|\Im_t\right]\right)/r + r b_{t+1|t}^{ei} + B_t^{ei}\right)\end{aligned}$$

By the law of iterated expectation,

$$E\left[b_{t+1}^i x_{t+1}|\Im_t\right] = E\left[E\left(b_{t+1}^i|\Im_t, x_{t+1}\right)x_{t+1}|\Im_t\right] = b_{t+1|t}^{ie} E\left[x_{t+1}|\Im_t\right] = 0$$

Furthermore, we have that $E\left[\varepsilon_{t+1}^i|\Im_t\right] = 0$ per assumption, $E\left[B_{t+1}^{ei}|\Im_t\right] = B_t^{ei}$ from equation 2.16 and $E\left[b^{ei}_{t+2|t+1}|\Im_t\right] = B^{ei}_t + F^i b^{ei}_{t+1|t}$ from equation (2.17). Replacing back into the expected returns yields:

$$E\left[R^i_{t+1}|\Im_t\right] = \sigma^2_x A b^{ei}_{t+1|t}$$

Using proposition 2.2, we find

$$E\left[R^i_{t+1}|\Im_t\right] = b^{ei}_{t+1|t} E\left[R^m_{t+1}|\Im_t\right]$$

Now Let us compute the investor's beta:

$$\beta_t = \frac{Cov\left(R^i_{t+1}, R^m_{t+1}|\Im_t\right)}{Var\left(R^m_{t+1}|\Im_t\right)}$$

Taking into account the expression for market returns from proposition 2.2 this yields:

$$\beta_t = \frac{1}{\sigma^2_x} Cov\left(R^i_{t+1}, x_{t+1}|\Im_t\right)$$

Replacing for individual returns, using the fact that variables conditional on time t information drop out of the covariance, as they are not random conditional on information at time t, and using $E\left[\varepsilon^i_{t+1}|x_{t+1}\right] = 0$, gives:

$$\beta_t = \frac{1}{\sigma^2_x} Cov\left(b^i_{t+1} x_{t+1}, x_{t+1}|\Im_t\right) - \frac{1}{\sigma^2_x} \frac{\sigma^2_x A}{1 + r - F^i} Cov\left(b^{ei}_{t+2|t+1} + B^{ei}_{t+1}/r, x_{t+1}|\Im_t\right) \quad (2.18)$$

Taking into account the Kalman-filtering equations (2.16) and (2.17), we find that the second term in equation (2.18) reduces to 0:

$$Cov\left(b^{ei}_{t+2|t+1} + B^{ei}_{t+1}/r, x_{t+1}|\Im_t\right) = 0$$

This is a consequence of the fact that innovations to factor loadings are assumed to be independent of x_t. As investor's are acting like Bayesian's, this implies that the filter for b and B is uncorrelated with x_t.

Taking this result into account, beta reduces to:

$$\begin{aligned}\beta_t &= \frac{1}{\sigma_x^2}Cov\left(b^i_{t+1}x_{t+1}, x_{t+1}|\Im_t\right) = \frac{1}{\sigma_x^2}E\left[b^i_{t+1}x^2_{t+1}|\Im_t\right] \\ &= \frac{1}{\sigma_x^2}E\left[E\left[b^i_{t+1}|\Im_t, x_{t+1}\right]x^2_{t+1}|\Im_t\right] = b^{ie}_{t+1|t}\frac{1}{\sigma_x^2}E\left[x^2_{t+1}|\Im_t\right] = b^{ie}_{t+1|t}\end{aligned}$$

which proofs the proposition.

■

2.7.5 Proof of Proposition 2.4

Recall the notation $b^{iE}_{t|t-1} = E\left[b^i_t|\Im^E_{t-1}\right]$

$$\begin{aligned}\hat{\beta}^E_{t-1} &= \frac{Cov\left[R^i_t, R^m_t|\Im^E_{t-1}\right]}{Var\left[R^m_t|\Im^E_{t-1}\right]} \\ &= E\left[b^i_t|\Im^E_{t-1}\right] - \frac{\theta^i}{\sigma_x^2}Cov\left[\left(b^i_t - b^{ie}_{t|t-1}\right)\left(k^i_t x_t + K^i_t x_t/r\right), x_t|\Im^E_{t-1}\right] \\ &= E\left[b^i_t|\Im^E_{t-1}\right] - \frac{\theta^i}{\sigma_x^2}E\left[b^i_t - b^{ie}_{t|t-1}|\Im^E_{t-1}\right]Cov\left[\left(k^i_t x_t + K^i_t x_t/r\right), x_t|\Im^E_{t-1}\right] \\ &= b^{iE}_{t|t-1} - \frac{\theta^i}{\sigma_x^2}E\left[b^i_t - b^{ie}_{t|t-1}|\Im^E_{t-1}\right]\left(\frac{1+r}{r}\gamma^{i(1,2)}_{t|t-1} + F^i\gamma^{i(2,2)}_{t|t-1}\right)Cov\left[\frac{x^2_t}{\gamma^{i(2,2)}_{t|t-1}x^2_t + \sigma^{i2}_\varepsilon}, x_t|\Im^E_{t-1}\right]\end{aligned}$$

Now note that the term $x_t^2\left(\gamma_{t|t-1}^{i(2,2)}x_t^2+\sigma_\varepsilon^{i2}\right)^{-1}$ is a symmetric function around 0. Furthermore, the distribution of x_t is symmetric around 0, as it is a normal with 0 mean. Therefore,

$$Cov\left(\frac{x_t^2}{\gamma_{t|t-1}^{i(2,2)}x_t^2+\sigma_\varepsilon^{i2}}, x_t|\Im_{t-1}^E\right)=0$$

and we find that the econometrician's beta is $\hat{\beta}_{t-1}^E = b_{t|t-1}^{iE}$.

■

2.7.6 Proof of Proposition 2.5

Taking expectations of individual returns from equation (2.7) gives under the econometricians information set:

$$E\left[R_{t+1}^i|\Im_t^E\right]=(1+G_t)\,E\left[b_{t+1|t}^{ei}|\Im_t^E\right]\sigma_x^2A-G_tE\left[b_{t+1}^i|\Im_t^E\right]\sigma_x^2A$$

where the following substitution was made:

$$G_t=\frac{E\left[x_{t+1}\left(K_{t+1}^{bi}+K_{t+1}^{Bi}/r\right)|\Im_t^E\right]}{1+r-F^i}$$

Now recall from proposition 2.4 that the beta of the econometrician, $\hat{\beta}_t^E=E\left[b_{t+1}^i|\Im_t^E\right]$. We can therefore write:

$$E\left[R_{t+1}^i|\Im_t^E\right]=E\left[b_{t+1|t}^{ei}-b_{t+1}^i|\Im_t^E\right](1+G_t)\,\sigma_x^2A+\hat{\beta}_t^E E\left[R_{t+1}^m|\Im_t^E\right]$$

The proof of the proposition is concluded by recognizing that

$$\hat{\alpha}_t^E = E\left[b_{t+1|t}^{ei} - b_{t+1}^{i}|\Im_t^E\right](1+G_t)\,\sigma_x^2 A$$

■

2.7.7 Calibration and Data Description

The calibration of the model is relatively straightforward, as there are few variables and parameters. We use monthly return data from July 1926 to December 2001.[16] Equation (2.2) shows that x_t is the market return, less the equity premium. The market return is computed as the value-weighted portfolio of the universe of stocks in CRSP. The data set also contains returns on B/M decile portfolios, formed as in Fama and French (1993). The simulations are done only for the 10th B/M decile, i.e. the value stocks that were also investigated by Franzoni (2002).

The model allows the price of the market portfolio to be normalized to one. This normalization implies that the relative market return is equal to the absolute market return, while relative portfolio returns are equal to the absolute returns times the weight of the portfolio in the market. σ_x^2 is set to equal the variance of the excess market return, which is 5.5% monthly in the data set. The constant of absolute risk aversion A is chosen to match the theoretical equity premium, $A\sigma_x^2$, with the average realized equity premium, which is 0.68% monthly. The risk free rate r equals the average realized value of 0.31%.

The variance of idiosyncratic risk σ_ε^2 represents the noise in the observation equation of the

[16]The data has been obtained from Ken French's website.

Kalman system. It is matched to the idiosyncratic risk of the 10th B/M portfolio. σ_ε^2 is set to a level such that the R^2 in a regression of the portfolio return is 70%. In the simulation, factor loadings b_t are proxied by estimates of five-year rolling window regressions with one-month increments of portfolio excess returns on the market excess return. σ_u^2 is the sample variance of fitting an $AR(1)$ on the estimated time-series of $\{b_t\}$.

A separate issue is the estimate of F, which is a crucial parameter in the learning process. Dickey-Fuller unit root tests do not reject a unit root for the time-series of betas from the rolling-window estimation. However, the rolling window procedure mechanically induces a unit root in the data, as 59 out 60 observations that are used for estimating beta are overlapping in two adjacent windows.

A more correct procedure to obtain an estimate of F is to use the Kalman filter on returns along with maximum likelihood estimation, as it is explained in Section 2.3. This strategy yields estimates of F that are between .94 and .97, and these estimates are statistically different from 1.

In order to understand the impact of different values of F on the learning process, we perform simulations for values of F ranging .95 to 1.

In the cases that B is observable, the drift (in cases when $F = 1$) or the level of the long-run mean of b (in cases when $F < 1$) must be calibrated. The ex-post, unconditional mean of b over the sample period is 1.2, and the average decline is -0.1 monthly. In order t

o study the impact of different beliefs about these parameters, simulation results for $B = \{-0.01, 0, 0.01\}$ when $F = 1$ and $B/(1-F) = \{1.5, 1.2, 1.0\}$ when $F < 1$ are reported.

The initial conditions for the variance-covariance matrix of the forecast error is estimated using the first 5 years of the estimates of betas from the first window of the sample. The initial condition for b_t^e is equal to the estimate of beta in that first window. Note that choice of the initial conditions turns out not to be crucial empirically. This fact is discussed in the text in some detail. The exact formulas for the evolution of the forecast error is given in Section 2.7.1 of the appendix.

Chapter 3

Contagion and the Great Depression

3.1 Introduction

The Great Depression was an extraordinary decline in output worldwide, accompanied by widespread financial crisis. The depression was quickly propagated across countries. Certain shocks, such as the restrictive monetary policy induced by the Gold Standard, were common, whereas other shocks, such as banking crisis, were more idiosyncratic. The transmission channels for shocks originating in a particular country are international trade flows and the international flow of capital.

The question addressed in this chapter is whether the propagation mechanism of financial shocks changed during the Great Depression. The Great Depression was associated with a multitude of financial crises such as stock market crashes, bank runs, speculative attacks and sovereign defaults. Furthermore, these financial crisis spread quickly across countries.

Contagion is the transmission of idiosyncratic shocks to asset prices from one country to another. The transmission of idiosyncratic shocks occurs due to portfolio rebalancing: a movement in asset prices in one country leads to an adjustment in the portfolio position of investors which induces a change in the prices of another country. This change of prices in the country affected by contagion is not driven by a change in its fundamentals. Contagion causes an increase in the correlation structure of asset returns during periods of financial turmoil. If the correlation of asset returns increases precisely at the time when asset prices fall, contagion constitutes a powerful transmission mechanism, over and above the transmission mechanism via real economic activity.

I study a factor model of asset returns, that decomposes price movements into systematic and idiosyncratic risk. Such a factor model explains average returns well during normal times. I then go on to test whether the covariance structure of shocks changes during periods of financial crisis, i.e. whether idiosyncratic shocks are transmitted across countries during periods of financial crisis.

Two sets of data are used in order to test the hypothesis whether financial crises spread contagiously. The first set consists of stock market indices of the US, France, Germany and Great Britain. The US stock market crash is important as it marked the beginning of the world-wide depression. The second test concerns the debt crises of Latin America in 1931. Five of the 6 Latin American countries in my sample defaulted on their sovereign debt in 1931. The coincidence of defaults suggests that contagion the debt crisis might have spread contagiously across Latin America.

I find that the financial crisis did not spread due to contagion, except in the case of France. The comovement of financial markets during the Great Depression was thus caused by the deterioration of a common risk factors, and not the transmission of idiosyncratic shocks.

The remainder of this chapter is organized as follows. In section 3.2, a factor model is described, and financial contagion is defined within this model. Recent theoretical models of contagion are discussed within this factor model, and the empirical implementation is specified. In section 3.3, it the results of the contagion test for the stock markets of industrialized countries are reported. In section 3.4, the Latin American debt crisis is analyzed. Section 3.5 links the results of this chapter to the literature on the economics of the Great Depression. Section 3.6 concludes.

3.2 Financial Contagion

3.2.1 Factor Model of Asset Returns

Financial contagion is a change in the covariance structure of returns during periods of financial crisis. Consider the following factor model of returns:

$$R_t^i = \alpha^i + \sum_v \beta^{iv} f_t^v + \varepsilon_t^i \tag{3.1}$$

In this set-up, the superscript i refers to country i, and returns of a countries stock market index at time t are denoted by R_t^i. Returns are assumed to be generated by factors f_t^v, which is the realization of the v's factor at time t. The coefficients β^{iv} are the factor loadings. The key assumption in this set-up is that all the systematic risk is captured by the common factors, and that the error terms ε_t^i are independently distributed across countries, i.e. that $E\left(\varepsilon_t^i \varepsilon_t^j\right) = 0$ for all i, j.

The factor model specified in equation 3.1 encompasses a wide variety of asset pricing models. In particular, in the case of the CAPM, the constant α^i is the risk-free rate, and there is only one common factor, which is the value weighted world stock index. In the case of the CCAPM of Breeden (1978) or Lucas (1978), the single factor is consumption growth. The factor structure follows from a linearization of the Euler Equation. At a more general level, the ICAPM by Merton (1971) generates a factor structure such as 3.1, where the factors are state variables that shift the investment opportunity set. The model is closest to Ross's (1976) APT, where equation 3.1 holds approximately.

It is important to point out that no assumptions were made about the variance structure of shocks. In particular, the model allows for heteroscedastic factors and heteroscedastic idiosyncratic risk. The variance of the shocks is finite, but it can vary over time. In order to fix notation, denote the variance of the idiosyncratic shock at time t of country i by σ_t^i, and the variance of factor v at time t by σ_t^v.

Definition 3.1 *Contagion in a given time period* $\tau \in [t', t'']$ *between two countries* i, j *is an increase in the covariance of returns between those countries due to correlated idiosyncratic shocks* $E\left(\varepsilon_\tau^i \varepsilon_\tau^j\right) > 0$.

Whereas is "normal" times, idiosyncratic risk is uncorrelated across countries, it becomes correlated during crisis. A correct specification of the factor model is thus to introduce a crisis factor. A financial crisis can be defined as a decline in asset prices in excess of a particular threshold. In the case of the US stock market crash of 1929, prices declined by 23% within 2 days. If the decline in asset prices is caused by a decline in a common factor f^v, then there is no contagion. However, if the financial market crisis is idiosyncratic to one country, and yet asset prices in another country decline without a decline in the systematic factors, then this is contagion. In terms of the factor model specified above, contagion happens when the idiosyncratic shocks are correlated across two countries during the time of financial crisis. An alternative way to write down such a model is that there is a financial crisis factor, i.e. the first $V-1$ factors are factors during normal times, whereas the V's factor only appears during a financial crisis. The empirical implications of this definition will be discussed below.

3.2.2 Theories of Contagion

The factor model developed in the previous section can be interpreted as a reduced form of an equilibrium model of asset prices. Theories of contagion share the following key ingredient: idiosyncratic shocks to the fundamentals of one country are transmitted to other countries (or assets) via portfolio effects. There are currently three mechanisms that lead to contagion: the wealth effect, inference in an asymmetric information setting, and learning about an underlying state variable under symmetric information.

Kyle and Xiong (2001) develop an equilibrium asset pricing model where contagion is caused by the wealth effect. Rational convergence traders trade against exogenous noise traders and liquidity providers, investing in two risky and one riskless asset. When the wealth of convergence traders due to an idiosyncratic shock to the demand of one of the risky assets falls, and this decline in wealth is sufficiently large, they liquidate the position in the other market as well, even though fundamentals in the markets are unchanged. The volatility of both asset increases, and correlation of the returns of the two risky assets increase as well.

Kodres and Pritsker (2001) study an asymmetric information general equilibrium model with many assets. They show that cross-correlation between markets can occur, even if those markets have no common systematic risk factors. In their setting, contagion can not only occur due to portfolio rebalancing, it can also be caused by information asymmetries between markets. The authors also study the importance of hedging demand (such as portfolio insurers) for contagion. As an example, consider a world with 3 assets and two systematic risk factor. Countries 1 and 3 do not share a systematic risk factor, increase their position in country 2 in order to keep the exposure to the first systematic risk factor, and reduce the exposure in country 2. Thus an idiosyncratic shock to country one is transmitted to country 3 via portfolio rebalancing. In general, contagion tends to be magnified by the presence of agents that conduct dynamic hedging.

In the theory of Ribeiro and Veronesi (2002), the variance-covariance matrix of fundamentals is constant over time. However, the growth rate varies according to a Markov process and is unobservable. Investors thus have to infer the most likely state of the current growth rate from the observation of fundamentals. Switches in the growth rate are common across countries, and are interpreted as a global business cycle. When investors are uncertain about the true state of the cycle, excess variance and covariance in asset returns relative to fundamentals is created. This theory is again compatible with the factor specification.

3.2.3 Data

The test for contagion in the Great Depression is done on two sets of data. The data set for the stock market crash of 1929 is taken from the NBER Historical Database.[1] For Great Britain, the security price index (series 11011) computed by Bankers Magazine is chosen. For Germany, the 'index of stock prices' (series 11023) originally published in the 'Viertelsjahrhefte zur Konjunkturforschung'. For US stocks, the index of all common stocks published by the Cowles Commission and Standard and Poors (series 11025) is taken. The Open market rate

[1] http:\\www.nber.org\databases\macrohistory\

for Great Britain (series 13016) is from 'The Economist', for Germany from (series 13018) from 'Statistisches Jahrbuch fuer das Deutsche Reich' and for the US (series 13002) from the Commercial and Financial Chronicle. The data for Latin America are from Dornbusch (1990). They consist of monthly sovereign bond prices in constant US$.

3.2.4 Empirical Implementation

In order to test for contagion, the factor model from equation (3.1) is estimated first. Then I test whether the residuals from the factor model are correlated during different time spans of the Great Derpession such as the time period of the stock market crash. The empirical implementation of the test is difficult for two reasons. First, the common factors of the factor model are unobserved. In particular, it is likely that some of the systematic risk factors are not accounted for. This is a familiar problem in the closed economy asset pricing literature. The common approach is to regress country returns on the contemporaneous returns of the other countries in the sample, but it is more than likely that such a regression does not capture all of the systematic risk. In such a regression, there is therefore residual covariance of the error terms. The second difficulty is that the error terms are heteroscedastic, which makes the statistical inference difficult. The time periods of interest when studying contagion are precisely the times when the variance of idiosyncratic shocks is large. The approach that I am taking is to estimate a factor model that is closely related to general equilibrium asset pricing theories such as the CAPM or the CCAPM. Returns of country i are regressed on the contemporaneous returns of the other countries in the sample, as well as interest rates of all countries in the sample. These first stage regressions are reported in the appendix. In order to control for autocorrelation, I also include lagged returns and lagged interest rates in the first stage regressions. From these first stage regressions, the residuals are recovered. Then the variance-covariance matrix of residuals is estimated for the crisis period and the normal period. In order to test for a change in the variance-covariance matrix, the determinant change in covariance test (or simply dcc-test) is performed. It tests the statistical significance in the change of the variance-covariance matrix of residuals. The distribution of the variance covariance matrix is obtained by bootstrapping. This test is developed by Rigobon (2002 a, b), and has the advantage that it is robust to heteroscedasticity of the error structure. In particular, no distributional assumptions about the residuals are needed. The variance-covariance matrix can have any form. The test reports whether the covariances of the residuals change significantly between the crisis and the normal period.

3.3 The Stock Market Crash of 1929

3.3.1 Economic Environment

The US stock market crash of October 1929 marked the beginning of the Great Depression. US industrial production peaked in August 1929 and declined 1.8% between August and October, by 9.8% during the last 2 months of 1929, and by 23.9% in 1930. The decline of industrial production only reversed in 1932. German and British industrial production show similar patterns; peaking in 1929 and declining steadily until 1932. The magnitude of the depression, however, was much less pronounced in Britain than in the US or Germany.

Contractionary open market sales by the Federal Reserve started at the beginning of 1928 according to Hamilton (1987). This contractionary monetary policy, together with an increased demand for loans caused an increase in interest rates in 1929. Figure 3-2 shows that the interest

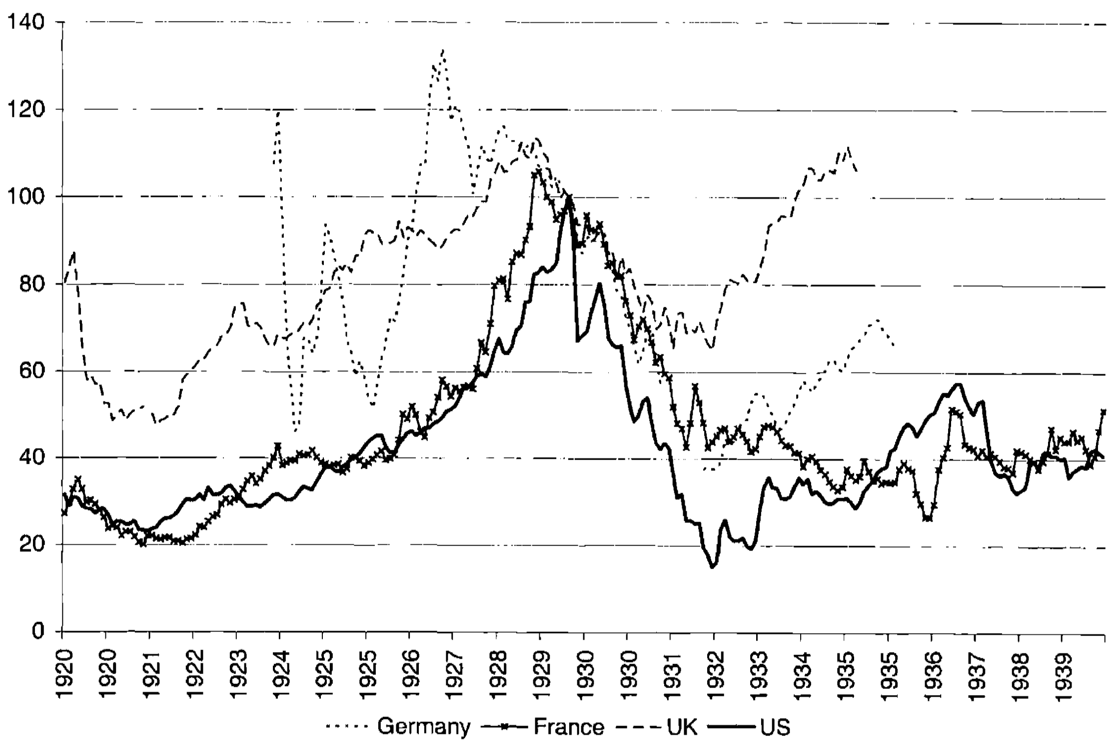


Figure 3-1: Stock Indices in France, Germany, Great Britain, and the US. Indices normalized at 100 in 10:1929.

rate in the US, Great Britain and Germany increased significantly during the twelve months prior to the Stock market crash (from 4.34% to 6.15% in Great Britain, from 6.58% to 7.28% in Germany and from 5.51% to 6.12% in the US).

The literature disagrees about the link between the Great Depression and the US stock market crash of 1929. Temin (1989) argues that the stock market crash in the US was not the cause for the Great Depression. A series of negative shocks, the combination of bad policy, the aftermath of WWI, and the Gold Standard as international policy regime all contributed to the Great Depression. Romer (1990), however, does show that the crash contributed significantly to the slump in US consumption of 1930. Romer demonstrates that the increase in uncertainty

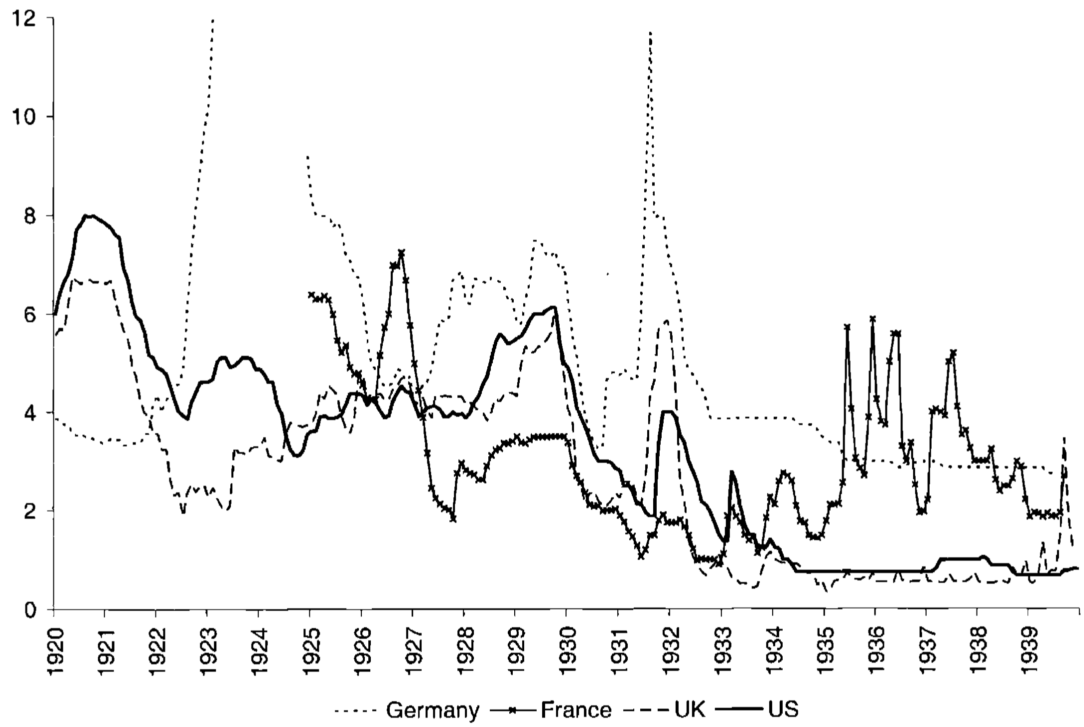


Figure 3-2: Interest Rates in Germany, France, Great Britain, and the US.

caused by increased stock market volatility dramatically led consumers and producers to reduce investments in durable goods. White (1990) argues that the US crash constituted the burst of a speculative bubble. White deomonstrates that dividends moved closely with the stock index prior to 1926, but were overshot significantly by the latter during the months prior to the crash.

The US stock market crash did not coincide with the decline in Germany, Great Britain, or France. The French stock market began to decline in February 1929, the German one in June 1928, and the British one in January 1929. The decline of industrial production of Germany, France and Britain, however, coincide in their timing with the decline in the US. If it is true that the increase in volatility of the stock market is a major explanation for the decline of output in the US, then one would also expect this to be a possible mechanism for the two other countries.

This is especially true for Germany. The decline in German output is of very similar magnitude to the US and the beginning of the German depression coincides with the US. Figure 3-1 shows that the US and German stock market indices move closely together after October 1929, but not before that date.

In tests for contagion, three periods are distinguished. The crash period is dated October 1929 to December 1931, which is the first part of the great depression. The great depression period is dated October 1929 to December 1934, marks the beginning of the recovery from the Great Depression. As some believe that the stock market crash of 1929 was preceded by a speculative bubble, I also define a bubble period from January 1926 to December 1931.

3.3.2 Testing for Contagion in France

The French stock market displays the greatest similarity with the US stock market, gaining roughly 80% during the 1920's, and then dropping sharply in the 1930's, reverting back to its level of the early 1920's by the mid-1930's. This similarity in the stock market is accompanied by differences in the evolution of the interest rate. In particular, the French interest rates peaked in 1927, as France needs to defend its adherence to the Gold standard.

The correlation of the French stock market with the US stock market are reported in the first column of table 3.1. Whereas the correlation with the US market is 37% over the sample period, it increases to 72% during the crash period. However, this increase in correlation is not accompanied by an increase in the volatility of the French stock market, the standard deviation is actually 4.5% during the crash period, compared with 5.8% during the sample period.

The dcc-test described in section 2.3 shows that there was a significant change in the US-French variance-covariance matrix during the crash period and also during the longer Great

Depression period. The p-value for both of these tests is above 99%, and thus the stability of the variance-covariance matrix is rejected.

Table 3.1: This table reports the test results for the change in the variance-covriance matrix of residuals of France and the US between the normal and the abnormal periods. The first-stage regressions are reported in Table 6 of the Appendix. $\Delta\lambda$ refers to the change of the determinant of the covariance matrix. The distribution of the covariance matrix is obtained by bootstrapping the residuals 10,000 times. $\sigma(\lambda)$ denotes the standard deviation of the change in the determinant of the covariance matrix, computed from the bootstrapped distribution. The z-stat is the ratio of $\Delta\lambda$ over $\sigma(\lambda)$. The p-value the probability of observing the estimated $\Delta\lambda$. Contagion is accepted for p-values below 2.5% or above 97.5%.

Testing Contagion: France and US							
Regressions	$\rho(us, fr)$	$\sigma(fr)$	$\Delta\lambda$	$\sigma(\Delta\lambda)$	z-stat	p-value	Contagion
Sample 20:1 - 39:12	.37	.058					
Crash 29:10 - 31:12	.72	.045	-9.37e-6	5.23e-6	-1.79	.992	yes
Gr Depr 29:10 - 34:12	.56	.053	-1.10e-5	3.85e-6	-2.87	.999	yes
Bubble 26:1 - 31:12	.57	.055	8.20e-7	1.65e-6	0.49	.309	no

3.3.3 Testing for Contagion in Germany

Due to the hyperinflation in 1921, the German economy was facing very different shocks than the other countries in the sample. This is most obvious from the interest rate path during the 1920's. In 1922, the German interest rate increased sharply from roughly 4% to over 12%, as the monetary authorities in Germany contracted money supply. This period of high interest rates lasted until 1926, when German interest rates fell back to 4%, comparable with the US interest rate. Thereafter, the German interest rate is more volatile than the other interest rates in the sample, but the correlation is clearly positive.

These shocks to the German economy caused stocks to be substantially more volatile during the 1920's than the stocks of the other countries. The correlation of German stock returns with US stock returns is only 22% over the sample period, as reported in table 2. During the crash, this correlation increases to 49%, and to 37% during the Great Depression period. At the same time, the volatility of German stock returns actually lower during the crash period than during the whole sample period, which again is attributable to the aftermath of the German hyperinflation. Over the first half of the sample period, much of the volatility in the German stock market is due to idiosyncratic shocks. This is also the picture that emerges from the analysis of the residual variance covariance matrix, reported in Table 2.

The change in the determinant of the variance-covariance matrix turns out to be insignificant for all of three subsamples. The change of the determinant is negative, which as the variance of the German returns is lower in the test periods, and the covariance is increasing during those periods. This is in contrast to the British case discussed below, where the financial crisis shows the usual pattern of an increase in the variance that dominates the increase in the covariance (leading to a positive change in the determinant of the covariance matrix).

Table 3.2: This table reports the test results for the change in the variance-covriance matrix of residuals of Germany and the US between the normal and the abnormal periods. The first-stage regressions are reported in Table 6 of the Appendix. $\Delta\lambda$ refers to the change of the determinant of the covariance matrix. The distribution of the covariance matrix is obtained by bootstrapping the residuals 10,000 times. $\sigma(\lambda)$ denotes the standard deviation of the change in the determinant of the covariance matrix, computed from the bootstrapped distribution. The z-stat is the ratio of $\Delta\lambda$ over $\sigma(\lambda)$. The p-value the probability of observing the estimated $\Delta\lambda$. Contagion is accepted for p-values below 2.5% or above 97.5%.

Testing Contagion: Germany and US							
Regression	$\rho\,(us, ge)$	$\sigma\,(ge)$	$\Delta\lambda$	$\sigma\,(\Delta\boldsymbol{\lambda})$	z-stat	p-value	contagion
Sample 24:02 - 35:12	.22	.075					
Crash 29:10 - 31:12	.49	.046	-.0000029	.00000422	-0.71	.875	no
Gr Depr 29:10 - 34:12	.37	.046	-.00000120	.00000393	-0.30	.688	no
Bubble 26:01 - 31:12	.32	.053	-.00000206	.00000180	-1.13	.907	no

3.3.4 Testing for Contagion in Great Britain

Figure 3-2 shows that British interest rates were strongly correlated with the US interest rates over the whole sample period. [add date of Britain abandoning the gold standard]. It is also visible from figure 3-1 that the percentage increase in the stock market during the 20's was somewhat less pronounced in Britain compared to the US (60% from 1921 to 1929 in Britain versus 80% in the US). After 1932, the British stock market recovered much quicker than the US Stock market, returning to its level of 1929 already in 1935. This recovery of the stock index does not take place in the US for a much longer time, in fact, it is only in the late 40's that the US stock market recovers.

Table 3 shows that the correlation of British stock returns with US returns is slightly lower during the crash period than during the whole sample (35% versus 37%). At the same time, the variance of the British stock return increases during the crisis period, and also during the more extended Great Depression period. This increase in volatility, however, is not due to contagion. The covariance structure of residuals does not change, and the contagion test is clearly rejected. Note that the contagion test is a two-sided test, so contagion is accepted if the p-value is either below 2.5%, or above 97.5%. The increase in volatility of the British stock market after the US stock market crash is thus due to a common factor, and not due to contagion.

3.4 The Latin American Debt Crisis

3.4.1 The Economic and Political Environment

Latin America experienced a period of extraordinary growth during the 1920s. In the period of 1925-1929, real GDP rose on average by 5.7% in Argentina, 7.2% in Brazil, 10.8% in Chile

Table 3.3: This table reports the test results for the change in the variance-covriance matrix of residuals of Great Britain and the US between the normal and the abnormal periods. The first-stage regressions are reported in Table 6 of the Appendix. $\Delta\lambda$ refers to the change of the determinant of the covariance matrix. The distribution of the covariance matrix is obtained by bootstrapping the residuals 10,000 times. $\sigma(\lambda)$ denotes the standard deviation of the change in the determinant of the covariance matrix, computed from the bootstrapped distribution. The z-stat is the ratio of $\Delta\lambda$ over $\sigma(\lambda)$. The p-value the probability of observing the estimated $\Delta\lambda$. Contagion is accepted for p-values below 2.5% or above 97.5%.

Testing Contagion: Great Britain and US							
	$\rho\,(us, uk)$	$\sigma\,(uk)$	$\Delta\lambda$	$\sigma\,(\lambda)$	z-stat	p-value	Contagion
Sample period 20:1 - 35:03	.37	.038					
Crash 29:10 - 31:12	.35	.048	4.49e-6	3.51e-6	1.28	.060	no
Gr Depr 29:10 - 34:12	.41	.041	1.05e-6	1.93e-6	0.54	.335	no
Bubble 29:10 - 39:12	.39	.035	1.57e-7	2.52e-7	0.63	.347	no

and 7.5% in Columbia (Eichengreen and Portes (1987)). The Latin American debt crises of the 1930s is preceded by political turmoil in the region. Following widespread labor protests, President Hernando Siles of Bolivia was overthrown in May 1930. In August 1930, political unrest forced the dictator of Peru, Augusto Leguia to resign. In the same month, a coup d'etat forced the Argentinian president, Hipolito Irigoyen out of office. Starting in October 1930, revolts in the southern provinces of Brazil took place which lead to a new head of state, Getulio Vargas. Social conflict together with a stagnating economy force the Chilean dictator Carlos Ibanez to resign in July 1931.

The major cause of the economic crises is the fall in Latin American export prices. Maddison (1985) and Marichal (1989) point out that the decline in Latin American exports was the single most important cause for the subsequent debt crises. Total Latin American exports fell from roughly US$ 3 Billion in 1929 to US$1.2 Billion in 1932 (Marichal 1989).

Despite the critical economic and political situation in Latin America, all countries serviced their external debt in 1929 and 1930. This led to a fall in foreign reserves of more than US$1

Table 3.4: This table reports the time period of adherence to the gold standard, and the month of default for the countries in the sample. Sources: Eichengreen (1992) and Marichal (1989)

	On Gold	Default
Argentina	27 - 29	No Default
Bolivia	28 - 31	Jan. 31
Brazil	27 - 30	Oct. 31
Chile	26 - 32	Jul. 31
Columbia	23 - 31	Feb. 32
Peru	31 - 32	May 31

Bio. The fall in foreign reserves and gold reserves as well as the dramatic decrease in export revenues made it increasingly difficult to service foreign debt. In January 1931, Bolivia was the first country to declare a unilateral moratorium on its foreign debt. Defaults of Peru, Chile and Brazil followed in the same year. Columbia partially suspended payments in 1932. Argentina was the only large Latin American economy which did not default in the 1930s. (See Table 1).

The economic and political turmoil in Latin America caused a dramatic drop in bond prices, as depicted in figure 3-3. This dramatic drop in bond prices reflects the increased default risk due to the economic recession that spilled over from the US, and the political instability caused by the recession. Table 3.4 shows that Argentina was the only country that did not default on its debt, and its bond price is the only one that reverts back to its pre-crisis level towards the end of the 30's. All the other countries in the sample experience a sharp drop in bond prices during 1931, from 80-90 before 1931 to 10-40 after 1931, and stay at this low level. The Latin American debt crisis also forced all of the countries out of the Gold Standard. Argentinia, which was the only country to avoid default is also the country that left the Gold Standard the earliest, in 1929. This confirms the finding of Eichengreen (1992) that countries that left the Gold Standard earlier recovered more quickly.

Eichengreen and Portes (1989) find for Latin America that 'interwar investors exhibited

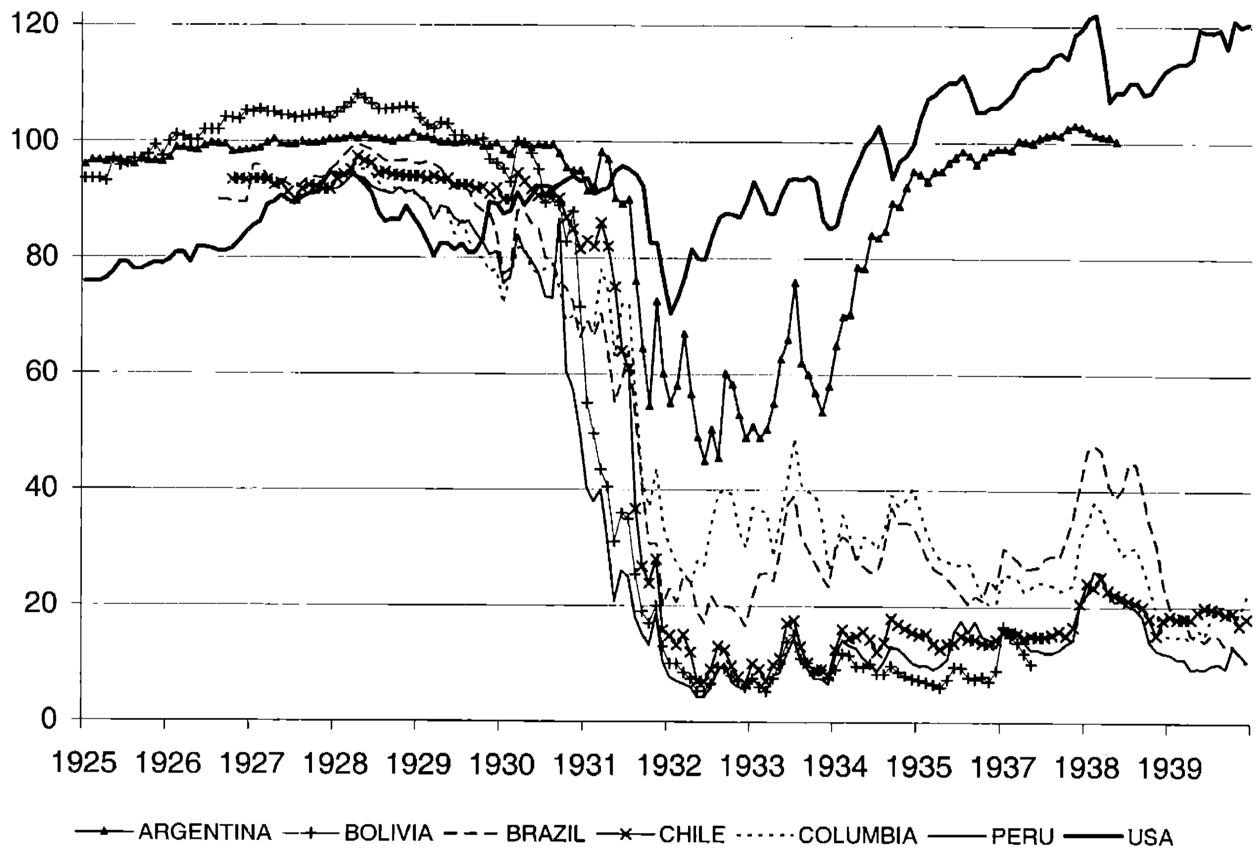


Figure 3-3: Bond Prices in Latin America.

sophistication and foresight at the lending stage. ... There is little evidence that capital markets have grown more sophisticated over time, or that banks have a comparative advantage in processing the relevant information'. This sophistication is visible in the evolution of bond prices, as the drop in prices precedes the default in most of the countries, i.e. it was the anticipation of a possible default that drove bond prices down. Argentina is again the exception, its bond price falls less than the price of the other countries, and recovers more quickly.

3.4.2 Testing for Contagion in Latin America

The correlations of bond returns increased during the debt crisis of 1931, as is shown in 3-4. In the figure, the correlations the correlations between bond returns in 1931 are depicted on the x-axes, plotted against the correlations during normal times on the y-axis. Points below the 45°-line are pairwise correlations that increase during the crisis period. Except for the Peru-Chile and Bolivia-Chile correlations, all of the correlations increase during 1931. In particular, correlations with the US bond price increase during 1931. As can be seen from 3-3, the US bond price started to decline towards the end of 1931. This coincides with the "second part" of the Great Depression (Friedman and Schwartz 1963, Temin 1989). Even though the correlations among Latin American bond prices increase, this is not a prove of financial contagion, a common factor could have become more volatile.

The results from the testing for contagion are reported in table 3.5. This table reports the dcc-test described in section 2.3 for two different periods, the year 1931, and a longer crisis period from June 1930 to June 1932. Both of these tests reject contagion.

Eichengreen and Portes (1987) compare the Latin American debt crisis of the 1930s with that of the 1980s. They point out that there is a triangular relationship between debt defaults, bank failures and exchange market disturbances. They find empirically that asset-market linkages running from debt defaults and exchange market disturbances to banking crises are the most important in understanding the international spread of financial crises in the 1930s.

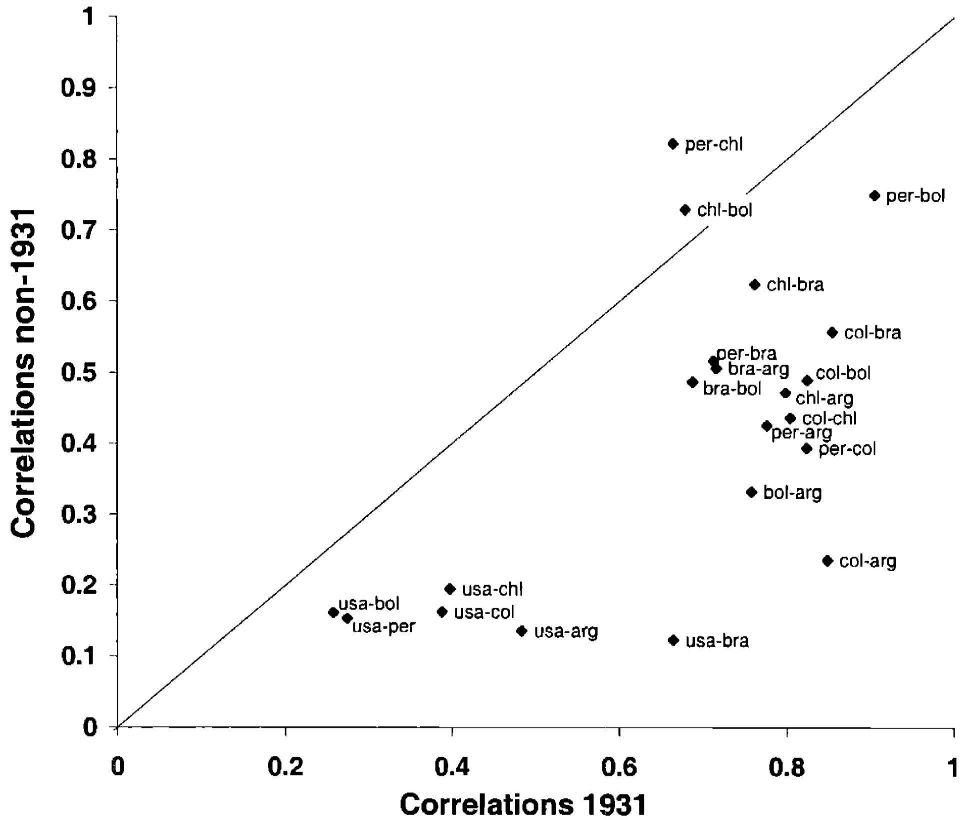


Figure 3-4: Correlations in Latin America during the debt crisis and during normal times.

3.5 International Transmission of the Great Depression

Recent attempts in explaining the causes of the Great Depression have emphasized international linkages. Kindleberger (1973) was an early proponent of the view that the Great Depression should be analyzed from a global point of view, even if it was caused in the US (Romer 1993). The primary transmission mechanism emphasized in the literature is the Gold Standard: monetary authorities in the US conducted restrictive monetary policies starting in 1928, and these policies were transmitted internationally via the Gold Standard.

The results in this chapter demonstrate that financial contagion was not one of the factors

Table 3.5: This table reports the test results for the change in the variance-covriance matrix of residuals of Latin American countries between the normal and the Default / Crisis Periods. The first-stage regressions are reported in Table 7 of the Appendix. $\Delta\lambda$ refers to the change of the determinant of the covariance matrix. The distribution of the covariance matrix is obtained by bootstrapping the residuals 10,000 times. $\sigma(\lambda)$ denotes the standard deviation of the change in the determinant of the covariance matrix, computed from the bootstrapped distribution. The z-stat is the ratio of $\Delta\lambda$ over $\sigma(\lambda)$. The p-value the probability of observing the estimated $\Delta\lambda$. The test is two sided: contagion is accepted for p-values below 2.5

Testing contagion in Latin America Argentina, Bolivia, Brazil, Chile, Columbia, Peru, US					
Regression Analysis	$\boldsymbol{\Delta\lambda}$	$\boldsymbol{\sigma(\lambda)}$	**z-stat**	**p-value**	**Contagion**
Default 31:01 - 31:12	43,220,709	4,466,817,800	.0096	.695	no
Crisis 30:06 - 32:06	-49,154,963	756,184,220	-.060	.610	no

that contributed to the spread of the great depression. The only exception is the case of France, where stock market contagion was detected. The spread of the Great Depression globally is thus caused by deterioration in the common fundamentals, such as monetary policy and trade linkages.

Temin (1989) and Eichengreen (1992) point out that the Gold Standard was associated with certain beliefs about the conduct of economic policy. The believe shared by policy makers at the time was that the adherence to the Gold standard was essential. The costs of restrictive monetary and fiscal policy in terms of output and unemployment were underestimated. Eichengreen and Sachs (1985) demonstrate that it is likely that a coordinated devaluation of currencies vis-a-vis the price of Gold would have boosted output of industrialized countries substantially.

Besides the emphasize on the role of the Gold Standard, much of the recent research has emphasized the role of credit market imperfections. The length of the Great Depression raises the question how financial and real shocks were propagated over time. An idea, going back to Fisher and Keynes, is the role of debt deflation. Falling asset prices create pressure on debtors, who are forced to sell, causing further declines in asset prices. As nominal claims of banks are replaced by real claims due to bankruptcy, deflation threatens to cause the collapse of the banking system. Bernanke (1983) demonstrates that such a mechanism can have real effects. As the net-worth of debtors falls, firms become credit constraint, and productive investment opportunities are forgone. Bernanke (1995) reports that banking crises have a statistically large and significant impact on employment and production. Furthermore, he shows that countries on the gold standard were more likely to be affected by banking crises. In the light of the current chapter, it is unlikely that banking crisis and the debt-default mechanism spread contagiously, it seems likely that it is spread due to common risk factors that deteriorated. The international spread of banking crisis is, however, left for future research.

3.6 Conclusion

Only the French Stock market was significantly affected by contagion from the US stock market crash of 1929. The crash did not spread contagiously to Germany or Great Britain. In the case of the Latin American debt crisis, contagion did not contribute to the widespread default of sovereign bonds. It is interesting to contrast this finding with more recent crisis. Many studies have documented that contagion has played a major role in the determination of asset prices during crisis periods such as the LTCM default in 1998, the Russian default in 1998, and the Asian crisis in 1997. This contrasting finding can be interpreted in light of the theories of contagion. The common element in theories about contagion is that portfolio effects led to increased volatility and correlations. It is very plausible that these portfolio effects were less important during the Great Depression, as the global economy was less integrated at the time.

The conclusion of this chapter is that the traditional channels for the international transmission of the Great Depression must play the predominant role in the spread of financial crisis. Adherence to the Gold Standard transmitted the excessively restrictive monetary policy implemented in the US to other countries. The deterioration of the real economy, and in particular the fall in trade flows constitute the common factor that lead to the extraordinary meltdown of global financial markets starting in 1929.

3.7 Appendix

France: First Stage Regressions							
	R_{US}	std. err.	i_{FR}	std. err.	i_{US}	std. err.	R^2
Sample 24:02 - 35:12	.272	(.054)	-.0002	(.002)	.001	(.0025)	.14
Crash 29:10 - 31:12	.312	(.065)	-.011	(.0275)	.012	(.0163)	.54
Gr Depr 29:10 - 34:12	.251	(.051)	-.002	(.010)	.002	(.0051)	.31
Bubble 26:01 - 31:12	.4245	(.0779)	.0035	(.0018)	.0018	(.0052)	.38

Germany: First Stage Regressions							
	R_{US}	std. err.	i_{GE}	std. err.	i_{US}	std. err.	R^2
Full Sample 24:02 - 35:12	.133	(.057)	-.012	(.0045)	.0051	(.004)	.18
Crash 29:10 - 31:12	.250	(.111)	.009	(.010)	-.004	(.011)	.28
Gr Depr 29:10 - 34:12	.013	(.009)	.013	(.009)	-.015	(.006)	.24
Bubble 26:01 - 31:12	.26	(.101)	-.057	(.007)	.004	(.008)	.10

Great Britain: First stage regressions							
	R_{US}	std. err.	i_{UK}	std. err.	i_{US}	std. err.	R^2
Sample 24:02 - 35:12	.194	(.038)	-.002	(.003)	-.003	(.002)	.18
Crash 29:10 - 31:12	.204	(.120)	.008	(.009)	-.010	(.0101)	.34
Gr Depr 29:10 - 34:12	.161	(.048)	.0008	(.005)	-.005	(.005)	.23
Bubble 26:01 - 31:12	.208	(.062)	.006	(.005)	-.010	(.005)	.16

Table 3.6: First Stage Regressions Europe

	Latin America: First stage regressions					
	Argentina	**Bolivia**	**Brazil**	**Chile**	**Columbia**	**Peru**
R_{ARG}		-.179	.370*	.428*	-.085	.163
$R_{ARG}(-1)$	-.246	.028	.148	.348*	-.233	-.366
R_{BOL}	-.049		-.040	.195*	.179*	.455*
$R_{BOL}(-1)$	.052	.061	.056	-.107	-.076	.087
R_{BRA}	.166*	-.065		.261*	.431*	.013
$R_{BRA}(-1)$	-.196*	.041	.073	.051	-.089	.080
R_{CHL}	.037	.268*	.223*		.064	.518
$R_{CHL}(-1)$	.237*	.037	-.136	-.142	-.242*	-.009
R_{COL}	.163*	.285*	.425*	.074		-.007
$R_{COL}(-1)$	-.019	-.066	-.097	.106	.022	-.040
R_{PER}	.046	.466*	.008	.386*	-.004	
$R_{PER}(-1)$	-.017	.002	-.003	.037	.000	-.033
R_{USA}	.366*	.179	.025	.193	-.029	-.056
$R_{USA}(-1)$	-.417	-.365	-.146	-.065	.590	.173
Constant	1.4	-.51	-.41	-1.8	-.23	1.4
R^2	.53	.62	.52	.73	.47	.71

Table 3.7: First Stage Regressions Latin America

Bibliography

[1] Abreu, D., and M. Brunnermeier, 2002, "Bubbles and Crashes", *Econometrica*, forthcoming.

[2] Ang, A., and J. Chen, 2002, "CAPM Over the Long-Run: 1926-2001", mimeo, University of Southern California and Columbia University.

[3] Baker, M., and S. Savasoglu, 2002, "Limited Arbitrage in Mergers and Acquisitions", *Journal of Financial Economics 64*, 91-116.

[4] Banz, R., 1981, "The Relationship between Return and Market Value of Common Stocks", *Journal of Financial Economics 9*, pp. 3-18.

[5] Barberis, N., 2000, "Investing for the Long Run when Returns are Predictable", *Journal of Finance 55*, pp. 225-264.

[6] Basu, S., 1977, "The Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hyothesis", *Journal of Finance, 32*, pp. 663-682.

[7] Bawa, V. S., S. Brown, and R. Klein, 1979, *Estimation Risk and Optimal Portfolio Choice*, North Holland.

[8] Bekaert, G., and Liu, J. 2001, "Conditioning Information and Variance Bounds on Pricing Kernels", NBER Working Paper 6880.

[9] Bernanke, B. 1983 "Nonmonetary effects of the Financial Crisis in the Propagation of the Great Depression ", *American Economic Review 73*: pp. 257-76.

[10] Bernanke, B. 1995 "The Macroeconomics of the Great Depression: A Comparative Approach", *Journal of Money, Credit, and Banking 27*: pp. 1-28.

[11] Black, F., M. Jensen, and M. Scholes, 1972, "The Capital Asset Pricing Model: Some Empirical Tests", in M. Jensen ed., *Studies in the Theory of Capital Markets*, (Praeger, New York, NY).

[12] Blanchard, O. J., 1993, "Movements in the Equity Premium", *Brookings Papers on Economic Activity, 1993-2*, pp. 75-138.

[13] Blanchard, O. J., and M. Watson, 1982, "Bubbles, Rational Expectations and Financial Markets,", in Paul Wachtel (ed.), *Crises in the Economic and Financial Structure,* Lexington Books, pp. 295-316.

[14] Bollerslev, T., R. Engle, and J. Wooldridge, 1988, "A Capital Asset Pricing Model with Time Varying Covariances", *Journal of Political Economy 96,* pp. 116-131.

[15] Boudoukh, J., M. Richardson, Y. Shen, and R. Whitelaw, 2002, "Do Asset Prices Reflect Fundamentals? Freshly Squeezed Evidence from the FCOJ Market", mimeo, NYU.

[16] Brav, A., R. Lehavy, and R. Michaely, 2002, "Expected Returns and Asset Pricing", mimeo, Duke University, University of Michigan, and Cornell University.

[17] Breeden, D., 1979, "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities", *Journal of Financial Economics, 7*, pp. 265-296.

[18] Brennan, M. J. and Y. Xia, 2001, "Stock Return Volatility and Equity Premium", *Journal of Monetary Economics 47*, pp. 249-283.

[19] Brown, D., and R. Jennings, 1989, "On Technical Analysis", *Review of Financial Studies 2*, pp. 527-551.

[20] Brown, S., and W. Goetzmann, 2001, "Hedge Funds with Style", NBER Working paper 8173.

[21] Brown, S., W. Goetzmann, and J. Park, 1998, "Hedge Funds and the Asian Currency Crisis", mimeo, NYU and Yale School of Management.

[22] Brunnermeier, M., and S. Nagel, 2002, "Arbitrage at its Limits: Hedge Funds and the Technology Bubble", mimeo, Princeton University and LBS.

[23] Campbell, J. Y., and L. Hentschel, 1992, "No News is Good News: an Asymmetric Information Model of Changing Volatility in Stock Returns", *Journal of Financial Economics 31*, 281-318.

[24] Campbell, J. Y., and R. Shiller, 1988, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors", *Review of Financial Studies 1*, pp. 195-227.

[25] Campbell, J. Y., and T. Vuolteenaho, 2002, "Bad Beta, Good Beta", mimeo, Harvard University.

[26] Campbell, J.Y., and J. Cochrane, 2000, "Explaining the Poor Performance of Consumption-based Asset Pricing Models", *Journal of Finance 55*, pp. 2863-2878.

[27] Campbell, J.Y., M. Lettau, B. Malkiel, and Y. Xu, 2001, "Have Individual Stocks become more Volatile? An Empirical Exploration of Idiosyncratic Risk", *Journal of Finance 56*, pp. 1-43.

[28] Chen, N.-F., and F. Zhang, 1998, "Risk and Return of Value Stocks", *Journal of Business 71*, 501-536.

[29] Cochrane, J. H., 2001, *Asset Pricing*, Princeton University Press, Princeton, NJ.

[30] Cohen, R. B., C. Polk, and T. Vuolteenaho, 2002, "The Value Spread", *Journal of Finance*, forthcoming.

[31] Constantinides, G., 1990, "Habit Formation: A Resolution of the Equity Premium Puzzle", *Journal of Political Economy 98*, pp. 519-543.

[32] Daniel, K., and S. Titman, 1997, "Evidence on the characteristics of cross-sectional variation in stock returns", *Journal of Finance 52*, pp. 1-33.

[33] Davis, J., E. Fama, and K. French, 2000, "Characteristics, Covariances, and Average Returns: 1929 to 1997", *Journal of Finance 55*, pp. 389-406.

[34] D'Avolio, G., 2002, "The Market for Borrowing Stock", *Journal of Financial Economics*, forthcoming.

[35] DeLong, J. B., A. Shleifer, L. Summers, and R. Waldmann, 1990a, "Noise Trader Risk in Financial Markets", *Journal of Political Economy 98*, pp. 703-738.

[36] DeLong, J. B., A. Shleifer, L. Summers, and R. Waldmann, 1990b, "Positive Feedback Strategies and Destabilizing Rational Speculation", *Journal of Finance 45*, pp. 375-395.

[37] Dornbusch, R. 1990 " Brazilian Debt: A Requiem for Muddling Through" , in: Sebastian Edward and Felipe B. Larrain "*Debt, Adjustment, and Recovery* " , Basil Blackwell.

[38] Doukas, J., C. F. Kim, and C. Pantzalis, 2002, "A test of the Errors-in-Expectations Explanation of the Value/Glamour Stock Returns Performance: Evidence from Analysts' Forecasts", *Journal of Finance 57*, pp.2143-2165.

[39] Eichengreen, B. 1992 "*Golden Fetters: The Gold Standard and the Great Depression, 1919-1939*", New York: Oxford University Press.

[40] Eichengreen, B., and J. Sachs 1985 "Exchange Rates and the Economic Recovery in the 1930s", *Journal of Economic History 45*:pp. 925 - 946.

[41] Eichengreen, B., and R. Portes 1987 "The Anatomy of Financial Crises", in: Portes and Swoboda (1987): 10- 66.

[42] Elton, E., 1999, "Expected Returns, Realized Return, and Asset Pricing Tests", *Journal of Finance 54*, pp.1199-1220.

[43] Engle, R. D. Lilien, and R. Robins, 1987, "Estimating Time-varying Risk Premia in the Term Structure: the ARCH-M Model", *Econometrica 55*, pp. 391-407.

[44] Engle, R., and M. W. Watson, 1987, "The Kalman Filter: Applications to Forecasting and Rational Expectations Models", in: *Advances in Econometrics Fifth World Congress*, Volume I, ed. T. Bewley, Cambridge University Press, pp. 245-283.

[45] Fama, E. and J. MacBeth, 1973, "Risk, Return, and Equilibrium: Empirical Tests", *Journal of Political Economy, 81*, pp. 607-636.

[46] Fama, E., and K. French, 1992, "The Cross-section of Expected Stock Returns", *Journal of Finance 47*, pp. 427-465.

[47] Fama, E., and K. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics 33*, pp. 3-56.

[48] Fama, E., and K. French, 1995, "Size and Book-to-Market Factors in Earnings and Returns", *Journal of Finance 50*, pp. 131-155.

[49] Fama, E., and K. French, 1996, "Multifactor Explanations of Asset Pricing Anomalies", *Journal of Finance 51*, pp. 55-84.

[50] Fama, E., and K. French, 2002, "The Equity Premium", *Journal of Finance 57*, pp. 637-659.

[51] Franzoni, F., 2002, "Where is beta going? The Riskiness of Value and Small Stocks", Ph.D. Dissertation, MIT.

[52] Friedman, M., and A. Schwartz 1963 "*A Monetary History of the United States, 1867-1960*", Princeton: Princeton University Press.

[53] Fung, W., and D. Hsieh, 2000, "Measuring the Market Impact of Hedge Funds", *Journal of Empirical Finance 7*, pp.1-36.

[54] Fung, W., and D. Hsieh, 2001, "The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers", *Review of Financial Studies 14*, pp. 313-341.

[55] Gabaix, X., P. Gopikrishnan, V. Plerou, and H. E. Stanley, 2002a, "Understanding large movements in stock market activity", mimeo, MIT.

[56] Gabaix, X., P. Gopikrishnan, V. Plerou, and H. E. Stanley, 2002b, "A simple theory 'cubic' laws of financial fluctuations", mimeo, MIT.

[57] Gennotte, G. and H. Leland, 1990, "Market Liquidity, Hedging and Crashes", *American Economic Review 80*, pp. 999-1021.

[58] Gennotte, G., 1986, "Optimal Portfolio Choice under Incomplete Information", *Journal of Finance 61*, pp. 733-749.

[59] Goetzmann, W., J. Ingersoll, and S. Ross, 1998, "High Water Marks", NBER Working Paper 6413.

[60] Gromb, D., and D. Vayanos, 2002, "Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs", *Journal of Financial Economics 66*, pp. 361-407.

[61] Grossman, S., and J. Stiglitz, 1980, "On the Possibility of Informationally Efficient Markets", *American Economic Review 70*, pp. 393-408.

[62] Grundy, B., and M. McNichols, 1989, "Trade and Information of Prices through Trade and Direct Disclosure", *Review of Financial Studies 2*, pp. 495-527.

[63] Hamilton, J. (1987) “Monetary Factors in the Great Depression“, *Journal of Monetary Economics 19*, pp. 145-69.

[64] Hamilton, J. D., 1994, *Time Series Analysis.* Princeton, NJ: Princeton University Press.

[65] Hansen, L. and Jagannathan, R., 1997, "Assessing Specification Errors in Stochastic Discount Factor Models" *Journal of Finance 52*, pp. 557–590.

[66] Harvey, C., 1989, "Time-Varying Conditional Covariances in Tests of Asset Pricing Models", *Journal of Financial Economics 24*, pp. 289-317.

[67] Hodrick, R. and Zhang, X. 2001, "Evaluating the Specification Errors of Asset Pricing Models", *Journal of Financial Economics 62*, pp. 327-376.

[68] Hogan, S., R. Jarrow, and M. Warachka 2002 "Statistical Arbitrage and Tests of Market Efficiency", mimeo, Singapore Research University.

[69] Jagannathan, R., and Z. Wang, 1996, "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance 51*, pp. 3-53.

[70] Küchler, U. and Sørensen, M. 1997, *Exponential Families of Stochastic Processes*, Berlin: Springer Verlag.

[71] Kindleberger, C. 1973 “*The World in Depression* “, Berkeley: University of California Press.

[72] Kodres, L., and M. Pritsker 2001, "A Rational Expectations Model of Financial Contagion", *Journal of Finance 57*, pp. 769-99.

[73] Kyle, A., 1985, "Continuous Auctions and Insider Trading", *Econometrica 53*, pp. 1315-1335.

[74] Kyle, A., and W. Xiong 2001, "Contagion as a Wealth Effect", *Journal of Finance 56*, pp. 1401-40.

[75] La Porta, R., J. Lakonishok, A. Shleifer, and R. Vishny, 1997, "Good News for Value Stocks: Further Evidence on Market Efficiency", *Journal of Finance 52*, pp. 859-874.

[76] Lakonishok, J., A. Shleifer, and R. Vishny, 1994, "Contrarian Investment, Extrapolation, and Risk", *Journal of Finance 50*, pp. 1541-1578.

[77] Lamont, O., and R. Thaler, 2001, "Can the Market add and substract? Mispricing in Tech-stock carve-outs", mimeo, GSB Chicago.

[78] Lettau, M., and S. Ludvigson, 2001, "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying", *Journal of Political Economy*, 109, pp. 1238-1287.

[79] Lettau, M., and S. Ludvigson, 2002, "Expected Returns and Expected Dividend growth", mimeo, NYU.

[80] Lewellen, J. and J. Shanken, 2002, "Learning, Asset-pricing Tests, and Market Efficiency", *Journal of Finance 57* , pp. 1113-1145.

[81] Lintner, J., 1965, "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets", *Review of Economics and Statistics*, 47, pp. 13.-37.

[82] Liptser, R., and A. Shiryaev, 2000, *Statistics of Random Processes*, Vol. I & II, 2nd Edition, Berlin: Springer Verlag.

[83] Liu, J., and F. Longstaff, 2000, "Losing Money on Arbitrages: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities", mimeo, UCLA.

[84] Lucas, R., 1978, "Asset Prices in an Exchange Economy", *Econometrica 46*, pp.1429-45.

[85] MacKinley, A. C., 1995, "Multifactor Models do not Explain Deviations from CAPM", *Journal of Financial Economics 38*, pp. 3-28.

[86] Maddison, A. 1985 "*Two Crises: Latin America and Asia 1929-38 and 1973-83*", Paris: OECD.

[87] Marichal, C. 1989 "*A Century of Debt Crises in Latin America*", Princeton: Princeton University Press.

[88] Mehra, R. and E. Prescott, 1985, "The Equity Premium: A Puzzle", *Journal of Monetary Economics 15*, pp. 145-161.

[89] Merton R. C., 1973 "An Intertemporal Asset Pricing Model", *Econometrica 41*, pp. 867-887.

[90] Merton, R. C., 1990, *Continuous Time Finance*, Oxford: Blackwell Publishers.

[91] Mitchell, M., and T. Pulvino, 2002, "Characteristics of Risk and Return in Risk Arbitrage", *Journal of Finance*, forthcoming.

[92] Mitchell, M., T. Pulvino, and Erik Stafford, 2001, "Limited Arbitrage in Equity Markets", *Journal of Finance 57*. pp. 551-584.

[93] Moskowitz, T., 2002, "An Analysis of Covariance Risk and Pricing Anomalies", mimeo, University of Chicago.

[94] Ofek, E., and M. Richardson, 2002, "Dotcom Mania: The Rise and Fall of Internet Stock Prices", *Journal of Finance*, forthcoming.

[95] Pastor, L., 2000, "Portfolio Selection and Asset Pricing Models", *Journal of Finance 55*, pp. 179–223.

[96] Portes, R., and A. K. Swoboda 1987 "*Threats to International Financial Stability*", New York: Cambridge University Press.

[97] Ribeiro, R., and P. Veronesi, 2002, "The Excess Comovement of International Stock Markets in Bad Times: A Rational Expectations Equilibrium Model", mimeo GSB Chicago.

[98] Rigobon, R. 2002a, "Identification through Heteroscedasticity", *Review of Economics and Statistics*, forthecoming.

[99] Rigobon, R. 2002b, "On the Measurement of the International Propagation of Shocks: Are they Stable?", *Journal of International Economics*, forthcoming.

[100] Roll, R. and S. Ross, 1994, "On the Cross-sectional Relation between Expected Return and Betas", *Journal of Finance 50*, pp. 101-121.

[101] Roll, R., 1977, "A critique of the Asset Pricing Theory's Tests: Part I", *Journal of Financial Economics 4*, pp.129-176.

[102] Roll, R., 1984, "Orange Juice and Weather", *American Economic Review 74*, pp. 861-880.

[103] Romer, C. D. 1990 "The Great Crash and the Onset of the Great Depression", *Quarterly Journal of Economics 105* , pp. 597-624.

[104] Romer, C. D. 1993 "The Nation in Depression", *Journal of Economic Perspectives 7*,: pp. 19-39.

[105] Romer, D., 1993, "Rational Asset Price Movements without News", *American Economic Review 83*, pp. 1112-1130.

[106] Ross, S., 1976, "The Arbitrage Theory of Capital Asset Pricing", *Journal of Economic Theory 13*, pp. 341-360.

[107] Santos, M. S., and M. Woodford, 1997, "Rational Asset Pricing Bubbles", *Econometrica 65*, pp. 19-57.

[108] Santos, T., and P. Veronesi, 2001, "Labor Income and Predictable Stock Returns, mimeo, GSB Chicago.

[109] Shanken, J., 1990, "Intertemporal Asset Pricing: An Empirical Investigation", *Journal of Econometrics*, 45, pp. 99-120.

[110] Sharpe, W., 1964, "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk", *Journal of Finance 19*, pp. 425-442.

[111] Shiller, R. J., 1981 "Do Stock Prices Move too much to be Explained by Subsequent Changes in Dividends?", *American Economic Review 71*, pp. 421-436.

[112] Shleifer, A., and R. Vishny, 1997, "The Limits of Arbitrage", *Journal of Finance 52*, pp.35-55.

[113] Sundaresan, S., 1989, "Intertemporally Dependent Preferences and the Volatility of Consumption and Wealth," *Review of Financial Studies 3*, pp. 73-89.

[114] Temin, P. 1989 "*Lessons from the Great Depression* ", Cambridge: The MIT Press.

[115] Temin, P. 1993 "Transmission of the Great Depression", *Journal of Economic Perspectives 7*, pp. 87-102.

[116] Treynor, J., and R. Ferguson, 1985, "In Defense of Technical Analysis", *Journal of Finance 40*, pp.757-551.

[117] Vassalou, M., and Xing, Y., 2002, "Default Risk and Equity Returns", *Journal of Finance*, forthecoming.

[118] Veronesi, P., 1999, "Stock Market Overreaction to Bad Newsin Good Times: A Rational Expectations Equilibrium Model", *Review of Financial Studies 12*, 975-1007.

[119] Wang, J., 1994, "A Model of Intertemporal Asset Prices Under Asymmetric Information", *Review of Economic Studies 60*, pp. 249-282.

[120] White, E. N. 1990 "The Stock Market Boom and Crash of 1929 Revisited", *Journal of Economic Perspectives 4*, pp. 67-83.

[121] Wurgler, J. and E. Zhuravskaya, 2002, "Does Arbitrage flatten Demand Curves?", *Journal of Business*, forthcoming.

[122] Xiong, W., 2001, "Convergence Trading with Wealth Effects: an Amplification Mechanism in Financial Markets", *Journal of Financial Economics 62*, pp. 247-292.