Factor Selection for Beta Pricing Models

Olivier Ledoit*

November 1995

*The Anderson School at UCLA. This is a revised version of the last two chapters of my Ph.D. thesis at MIT. I thank Petr Adamek, Tony Bernardo, Tim Crack, Rich Lyons, Angel Serrat, Jiang Wang and especially John Heaton for comments. I also thank the chairman of my committee Andrew Lo.

Factor Selection for Beta Pricing Models

Abstract

The beta pricing literature usually makes enough unrealistic assumptions to obtain exact pricing results, and then engages in a debate over which factors have the best theoretical or empirical justification. It is more realistic to acknowledge that only approximate pricing can be obtained, so that the best factors are simply the ones that minimize the bound on mean squared deviations from the model. From this point of view, the optimal number of factors in the APT is one, the single-factor APT and the CAPM yield nearly the same bound on mean squared deviations, and this bound is too high for practical use.

1 Introduction

Beta pricing models have a rich history, going back to the Capital Asset Pricing Model (CAPM) and the Arbitrage Pricing Theory (APT). Atfer the CAPM and the APT were introduced, it was argued that they lacked testable implications, at least in their original formulations. Roll (1977) points out that the CAPM is not testable because the composition of the true market portfolio is unobservable. Shanken (1982, Appendix A) points out that the APT is not testable either, because the number of assets in our economy is finite.

In recent years, we have witnessed the emergence of a new paradigm that encompasses both the CAPM and the APT as special limiting cases, and is testable. Kandel and Stambaugh (1987) and Shanken (1987a) add to the CAPM the assumption that the correlation between the true and observed market portfolio returns is above some level specified *a priori*. Shanken (1992, Section III) adds to the APT the assumption that the maximum Sharpe ratio in the economy is below some level specified *a priori*. He also notes that this modified CAPM and this modified APT are intimately related. As a matter of fact, they only differ over what factors they use. Their testable implications are the same: an upper bound on the sum of squared deviations from an approximate beta pricing equation.

This common approach has a nice interpretation in terms of Hansen and Jagganathan's (1991) mean-standard deviation frontier for intertemporal marginal rates of substitution (IMRSs). MacKinlay (1995) exploits this framework to distinguish between theoretical explanations of stock pricing anomalies. Ait-Sahalia, Cochrane and Saá-Requejo (1995) use it to bound option prices in incomplete markets.

The present paper starts by giving a concise, self-contained presentation of the new theory,

emphasizing its versatility, and proposes to give it its own name: Risk Arbitrage Pricing Theory. This name comes from the theory's main economic assumption, which is to rule out risk arbitrage. A risk arbitrage is defined as an investment opportunity whose Sharpe ratio is too high. Determination of the exact meaning of "too high" is discussed in detail.

The Risk Arbitrage Pricing Theory provides a general framework encompassing all beta pricing models, including the CAPM and the APT, therefore it is particularly well-suited for analyzing the choice of factors. The main objective of this paper is to find satisfactory answers to key questions that were not well-formulated, but were nonetheless hotly debated, at the time when the original versions of the CAPM and the APT were introduced. These questions are: What objectives should determine factor selection? Should factors be exogenous aggregates or extracted from the covariance matrix of payoffs? What should the number of factors be? How accurate are the CAPM and the APT? Which one is better?

The next section reviews the Risk Arbitrage Pricing Theory. Section 3 uses it to analyze factor selection for beta pricing models. Section 4 updates this analysis to account for the effect of risk premium estimation error. Section 5 shows how to specify *a priori* an upper bound on the maximum Sharpe ratio. Section 6 acknowledges some of the limitations of the paper and gives directions for future research. The last section concludes.

2 Risk Arbitrage Pricing Theory

I define a very general economic model. The only restrictive assumption is that there is a riskless bond. It should be noted that, in practice, AAA-rated bonds generate almost riskfree payoffs at the usual horizons. Nonetheless, lifting this assumption will be an objective for future research.

Assumption 1 *The economy is populated by Von Neumann-Morgenstern utility maximizers. It is in equilibrium. In equilibrium, at least one agent is non-satiated.*

Assumption 2 *There is a finite universe of assets that are claims to random payoffs at the end of the period. These payoffs have finite second moments. If an asset trades, then it trades at a finite price at the beginning of the period. We consider a set of $n+k+1$ assets that trade without frictions, and whose payoffs are known.*

From now on, the term "asset" is reserved for the $n+k+1$ assets under consideration. The others are called "excluded assets". There may or may not be excluded assets. If there are, then they may or may not trade. If some of them trade, then they may do so with or without frictions, and their prices and/or payoffs may or may not be known. In short, there are no assumptions on excluded assets.

Assumption 3 *The $n+k+1$ assets are divided into 3 categories. There are n assets called "stocks" whose prices are not known. The goal is to characterize stock prices. There are k assets called "factors" whose prices are known. Finally, there is one asset called "bond" whose payoff is riskless and nonzero, and whose price is known.*

If stock prices are known, then they can be used to test implications of the theory. However, it should be obvious that these implications must be derived without looking at stock prices. This justifies the assumption that stock prices are not known.

Since at least one agent is non-satiated, the price of the bond must be nonzero and of the same sign as its payoff. We can assume without loss of generality that the payoff of the bond is one, and call its price $1/(1+r_F)$, where $r_F > -1$. The excess payoff of a portfolio of assets is

its payoff minus $(1 + r_F)$ times its price. The Sharpe ratio of a risky portfolio is the expectation divided by the standard deviation of its excess payoff.

If a portfolio of stocks and/or factors is riskless, then it has to earn the riskfree rate of return, or else a non-satiated agent would change her position to pocket the return differential, thereby disrupting equilibrium. Therefore we can assume without loss of generality that the only riskless portfolio of stocks and/or factors is the null portfolio. This implies that the $(n + k)$-dimensional covariance matrix of excess payoffs on risky assets is nonsingular. Therefore the orthogonal projection onto the space of asset payoffs is uniquely defined.

It may seem restrictive to require factors to be asset payoffs, since some authors allow APT factors to be exogenous aggregates instead. But such exogenous factors can always be projected onto the space of asset payoffs. If the goal is to explain stock prices in terms of covariances with factors, then factors can be replaced by their projections without loss of generality. This is the convention that we use here. An important point, however, is that the price of the projection must be known.

For the CAPM, the number of factors k can be set to one, and the single factor can be taken as the market portfolio. For the APT, the set of factors can be chosen to explain a large part of the variance of stock payoffs.

Let $\tilde{\mathrm{x}}$ denote the $(n + k)$-dimensional vector of excess payoffs on risky assets. Define its expectation $\boldsymbol{\mu} = \mathrm{E}[\tilde{\mathrm{x}}]$, and its nonsingular covariance matrix $\boldsymbol{\Sigma} = \mathrm{Cov}[\tilde{\mathrm{x}}, \tilde{\mathrm{x}}']$.

Let $\mathrm{w}_p = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ and $\tilde{x}_p = \mathrm{w}_p'\tilde{\mathrm{x}}$. The excess payoff on any portfolio of risky assets has its expectation equal to its covariance with $\tilde{x}_p$. In this sense, the portfolio w_p can be used to price

all the assets. I call it the *pricing* portfolio.[1] The pricing portfolio is mean-variance efficient and has maximum Sharpe ratio among combinations of assets.

Unfortunately, since the pricing portfolio depends on unknown stock prices, it is unobservable. Characterizing expected excess payoffs on stocks is possible if and only if we assume that we know something about the pricing portfolio. Roll's (1977) critique of the CAPM is that the pricing portfolio could be anything because of the influence of excluded assets. Shanken's (1992, Appendix A) critique of the APT is that the pricing portfolio could be anything when the economy is finite.

Kandel and Stambaugh (1987) and Shanken's (1987a) modification of the CAPM and Shanken's (1992, Section III) modification of the APT are simply to impose a restriction on the pricing portfolio, and then derive the restriction that it implies on stock prices. In the first case, the assumption is that the multiple correlation coefficient between the pricing portfolio payoff and factor payoffs is above a prespecified level $\underline{\rho}$. In the second case, the assumption is that the Sharpe ratio of the pricing portfolio is below a prespecified level $\overline{S}$. $\underline{\rho}$ and $\overline{S}$ can be chosen so that these two conditions are equivalent.

Theorem 1 *Let S_p denote the Sharpe ratio of the pricing portfolio. Let S_F denote the maximum Sharpe ratio in the space of factors. Let $\rho_{(p,F)}$ denote the multiple correlation coefficient of the pricing portfolio payoff with factor payoffs. They are related by the equation:*

$$S_F = S_p \times \rho_{(p,F)}. \tag{1}$$

Proof of Theorem 1 See Shanken (1987, Corollary 1).

[1] It is related to Shanken's (1987a) maximally correlated portfolio, and the familiar tangency portfolio.

Yet another way to state the same assumption is to appeal to the optimal orthogonal portfolio $\tilde{x}_{h(F)}$ (Roll, 1980; MacKinlay, 1995). It is the residual of the projection of the pricing portfolio onto the factor space. Let $S_{h(F)}$ denote its Sharpe ratio. It is easy to verify that $S_p^2 = S_F^2 + S_{h(F)}^2$. Therefore any restriction on S_p translates into a restriction on $S_{h(F)}$.

As we said, the economic model defined so far has no testable implication. This can be changed by imposing the following restriction.

Assumption 4 *The economist knows enough about agent preferences to specify* a priori *a lower bound $\underline{\rho} > 0$ on the multiple correlation coefficient between the pricing portfolio payoff and factor payoffs. Or, equivalently, the economist specifies* a priori *an upper bound $\overline{S} = S_F/\underline{\rho}$ on the Sharpe ratio of the pricing portfolio. Or, equivalently, the economist specifies* a priori *an upper bound $\overline{S}_h = \sqrt{1/\underline{\rho}^2 - 1}\, S_F$ on the Sharpe ratio of the optimal orthogonal portfolio.*

From now on, the analysis is carried indifferently in terms of $\underline{\rho}$, $\overline{S}$ or $\overline{S}_h$. Translation is immediate thanks to Theorem 1.

I call a portfolio with Sharpe ratio higher than the prespecified level $\overline{S}$: a "risk arbitrage" opportunity. Assumption 4 in effect rules out risk arbitrage. As $\overline{S}$ goes to infinity, the notion of risk arbitrage converges to Ross's notion of arbitrage.[2] For finite $\overline{S}$, a risk arbitrage is not necessarily an arbitrage, since it can be risky. However, it involves such little risk compared to its expected excess payoff that it cannot exist in equilibrium (by definition).

Strictly speaking, the theory is not fully specified until a method for determining $\overline{S}$ is given. Several such methods are reviewed in Section 5. But there are two important points: 1) Nothing can be said about stock prices without imposing *a priori* some restriction similar to Assumption

[2]Section 6.2 discusses in more detail the competing notions of arbitrage.

4; 2) In terms of stock pricing implications, it does not matter at all *how* $\overline{S}$ is determined, as long as *some* $\overline{S}$ is determined. Thus, the Risk Arbitrage Pricing Theory is a versatile framework that encompasses a variety of more specialized models.

Let $\tilde{s}_1, \ldots, \tilde{s}_n$ denote excess payoffs on stocks. Let $\tilde{f}_1, \ldots, \tilde{f}_k$ denote excess payoffs on factors. Project stocks onto factors:

$$\tilde{s}_i = \alpha_i + \beta_{i1}\tilde{f}_1 + \ldots + \beta_{ik}\tilde{f}_k + \tilde{e}_i, \qquad \mathrm{E}\left[\tilde{e}_i \middle| \tilde{f}_1, \ldots, \tilde{f}_k\right] = 0, \qquad i = 1, \ldots, n, \tag{2}$$

where α_i is the expectation of the excess residual payoff (also called alpha), $\beta_{i1}, \ldots, \beta_{ik}$ are factor loadings (also called betas), and $\tilde{e}_i$ is the residual. Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)'$ denote the vector of alphas. Let $\tilde{\mathrm{e}} = (\tilde{e}_1, \ldots, \tilde{e}_n)'$ denote the vector of residuals. Let $\overline{\lambda}$ denote the largest eigenvalue of the covariance matrix of residuals $\boldsymbol{\Omega} = \mathrm{Cov}[\tilde{\mathrm{e}}, \tilde{\mathrm{e}}']$. Intuitively, if $\overline{\lambda}$ is not large, then residuals do not explain much of the risk: factors explain a lot of it.

Theorem 2 *Consider the approximate beta pricing formula:*

$$\mathrm{E}\left[\tilde{x}_i\right] \approx \beta_{i1}\mathrm{E}\left[\tilde{f}_1\right] + \ldots + \beta_{ik}\mathrm{E}\left[\tilde{f}_k\right], \qquad i = 1, \ldots, n. \tag{3}$$

Deviations from approximate beta pricing are measured by alphas:

$$\alpha_i = \mathrm{E}\left[\tilde{x}_i\right] - \left(\beta_{i1}\mathrm{E}\left[\tilde{f}_1\right] + \ldots + \beta_{ik}\mathrm{E}\left[\tilde{f}_k\right]\right), \qquad i = 1, \ldots, n. \tag{4}$$

Under maintained Assumptions 1-3, Assumption 4 is equivalent to the following bound on deviations from beta pricing:

$$\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha} \leq \overline{S}^2 - S_F^2, \tag{5}$$

where $\boldsymbol{\Omega}$ is the covariance matrix of residuals. Under maintained Assumptions 1-3, Assumption 4 implies the following bound on the mean squared deviation from beta pricing:

$$\frac{1}{n}\sum_{i=1}^{n}\alpha_i^2 \leq \frac{\overline{\lambda}\left(\overline{S}^2 - S_F^2\right)}{n}, \tag{6}$$

where $\overline{\lambda}$ is the largest eigenvalue of the covariance matrix of residuals.

Proof of Theorem 2 See Shanken (1987, Proposition 2).

Assumption 4 is the main economic assumption of the Risk Arbitrage Pricing Theory, and Equations (5) and (6) are its testable implications. The link between them is a mathematical tautology, therefore testing asset pricing implication (5), for example, is equivalent to testing that the economy admits no risk arbitrage opportunities as defined by Assumption 4. The proof of Theorem 2 is constructive: the theory is rejected by the data if and only if it is possible to construct a portfolio whose Sharpe ratio exceeds $\overline{S}$.

The Risk Arbitrage Pricing Theory is intuitively attractive: Economists first write down everything they know about what should drive stock returns by selecting factors, and then they put an upper bound on the influence of what they do not know. This bound is strictly positive, because there are some things that drive returns, but are unobservable to economists.[3]

[3]To use the (loaded) language of economic rationality, a little deviation from model predictions can be rational, but too much deviation has to be irrational, and therefore will not exist in equilibrium. The value of $\overline{S}$ defines the boundary between rational and irrational pricing behavior.

The Risk Arbitrage Pricing Theory is related to results by Hansen and Jagganathan (1991, henceforth HJ). Given the mean vector and the covariance matrix of returns on a set of securities that do not span the riskless bond, HJ compute the mean-standard deviation frontier $(\mu, \sigma(\mu))$ for the intertemporal marginal rate of substitution (IMRS). The mean of the IMRS μ can be interpreted as the price of the riskless bond with face value one. The standard deviation of the IMRS $\sigma(\mu)$ can be interpreted as the maximum Sharpe ratio in the economy made of the set of existing securities and a riskless bond with price μ. Therefore, if we assume that no Sharpe ratio can exceed $\overline{S}$, then the bond price must lie between the lower bound $\underline{\mu} = \min\{\mu : \ \sigma(\mu) \leq \overline{S}\}$ and the upper bound $\overline{\mu} = \max\{\mu : \ \sigma(\mu) \leq \overline{S}\}$. It is possible to generalize this approach to obtain bounds on the price of any new asset, and even any number of new assets. In the Risk Arbitrage Pricing Theory, HJ's set of existing securities becomes the set of factors, HJ's new asset becomes the set of stocks, and HJ's bound on the new asset's price becomes the bound on stock prices from Theorem 2.

There are three limit cases where the mean squared pricing deviation goes to zero, i.e. we obtain exact beta pricing. The first case is when $\overline{\lambda}$ goes to zero. It means that factors explain all the variance of stocks. This is the result of a noiseless APT with no idiosyncratic risk. The second case is when $\overline{S}^2 - S_F^2$ goes to zero. It means that factors span the pricing portfolio. This is the result of the CAPM if the market portfolio is observed, and of Connor's (1984) unified beta pricing theory. The third case is when n goes to infinity. This is the result of the APT.[4]

[4]Please note that I consider the APT as an exact beta pricing model, because the mean squared deviation vanishes. Some authors consider it an approximate beta pricing model, but I prefer to reserve this term for models with nonvanishing mean squared deviation.

The problem with these extreme cases is that they are not realistic. In practice, one would expect the mean squared deviation to be nonzero. The only question is: How small is it? The answer hinges critically on the choice of factors.

As mentioned in the introduction, the main objective of this paper is not to present the Risk Arbitrage Pricing Theory, since other authors have done it before. The main objective is to use the Risk Arbitrage Pricing Theory to analyze factor selection in beta pricing models. Indeed, the way the theory was derived, factors could be anything, so this framework is particularly well-suited for comparing the respective merits of different sets of factors. This problem has generated a lot of attention in the literature, but it could not be treated rigorously outside of the Risk Arbitrage Pricing Theory.

The specific questions to be investigated are: What objectives should determine factor selection? Should factors be exogenous aggregates or extracted from the covariance matrix of payoffs? What should the number of factors be? How accurate are the CAPM and the APT? Which one is better?

Section 3 examines these questions in the way the Risk Arbitrage Pricing Theory was presented above, i.e. ignoring estimation error. Please note that estimation error on expected excess payoffs on stocks is inconsequential, since stock prices are not known in the theory. It would only affect *tests* of the Risk Arbitrage Pricing Theory, and these tests are outside the scope of the paper. On the other hand, estimation error on expected excess payoffs on factors matters a lot. Its impact on the answers found in Section 3 is analyzed in Section 4. Thus, Sections 3 and 4 give increasingly complex and realistic answers to the same set of questions about factor selection in beta pricing models.

These answers do not depend at all on the method used for determining $\underline{\rho}$. This is why the discussion of such methods is relegated to Section 5. However, some answers do depend on the actual value of $\underline{\rho}$. For illustrative purposes, the value used here is $\underline{\rho} = 1/\sqrt{2}$, following Ross (1976, p.354) and MacKinlay (1995). This means that the maximum Sharpe ratio in the orthogonal of the factor space is equal to the maximum Sharpe ratio in the factor space. MacKinlay's empirical results show that this choice may overestimate the accuracy of beta pricing models. Where applicable, sensitivity of a result to the value of $\underline{\rho}$ is reported.

For the purpose of comparing the accuracy of beta pricing models with different sets of factors, it is Equation (6), rather than Equation (5), that is used. The mean squared deviation is a simpler and more intuitive measure of model accuracy than $\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha}$. This is in the tradition of the APT. The conclusion comes back to this issue.

3 Factor Selection Without Estimation Error

This section answers key questions about factor selection in the Risk Arbitrage Pricing Theory, ignoring estimation error on expected excess payoffs of factors.

3.1 Objective

The set of factors should be selected to minimize the upper bound on the mean squared deviation from beta pricing derived in Equation (6): $\overline{\lambda}(\overline{S}^2 - S_F^2)/n$. $\overline{S}$ and n are fixed with respect to the set of factors, therefore the objective is a combination of maximizing S_F with minimizing $\overline{\lambda}$. Maximizing S_F means taking factors that are as close as possible to spanning the pricing

portfolio. This is the core idea of the CAPM. Minimizing $\overline{\lambda}$ means taking factors that explain as much as possible of the risk of the stock market. This is the core idea of the APT. Therefore the Risk Arbitrage Pricing Theory acknowledges the respective contributions of the CAPM and the APT. Rather than opposing them, it combines them.

Both parts of the objective designate the market factor as a natural candidate: first, it is close to the pricing portfolio (if there is some truth to the CAPM); and second, it explains a lot of the risk. To be fair, there has been a lot of debate recently about the first point. Some empirical results appear to show that market betas are unrelated to expected returns. However, there is absolutely no doubt about the second point. Therefore the market factor deserves to be included in any beta pricing model, whether or not market betas are related to expected returns, just because it explains a lot of the variation in stock returns.

Other factors can be added if they explain a lot of the risk, as is customary in the APT. One idea that might have slipped in the cracks between the CAPM and the APT, is that it is also worthwhile adding factors that do not explain much of the risk, provided that they have high Sharpe ratios (and low correlation with the other factors). This is not a well-known justification for adding factors, yet it is intuitive and appears naturally in the Risk Arbitrage Pricing Theory. For example, Chen, Roll and Ross's (1986) factor selection scheme is not soundly justified by the standard APT, because it pays no attention to how much stock market risk is explained by factors. However, searching for factors with high Sharpe ratios, as these authors do, is justified by the Risk Arbitrage Pricing Theory.

In the remainder of the section, I pay scant attention to the maximization of S_F. I only consider sets of factors that include some version of the market factor, and set S_F equal to the

Sharpe ratio of the market factor S_M. When the market is a factor, S_M is a lower bound for S_F, therefore the inequality:

$$\frac{1}{n}\sum_{i=1}^{n}\alpha_i^2 \leq \frac{\overline{\lambda}\left(\overline{S}^2 - S_M^2\right)}{n} \tag{7}$$

is a consequence of Equation (6). I do not use the sharper equation because I believe that it is better to be conservative. Implicit in the use of some $S_F > S_M$ is the assumption that it is possible to publish a combination of factors that beats the market in terms of the Sharpe ratio, without having investor reaction adversely affect the Sharpe ratio of the combination. This implicit assumption strikes me as bold, and the disappearance of Banz's (1981) size effect in the eighties speaks for caution.[5] This argument can be interpreted as a criticism of the approach of Chen, Roll and Ross (1986). This is why the remainder of the section focuses solely on minimizing $\overline{\lambda}$.

Consider the following thought experiment: there are $n+k$ risky assets with known payoffs; but we can only obtain the prices of k linearly independent portfolios of these assets; these k portfolios are called "factors", and are used to obtain price bounds on n remaining "stocks"; what are the k best portfolios to designate as factors? This thought experiment exactly mimics what goes on when we select a set of factors for beta pricing. The remainder of the section shows how to select the set of factors that minimizes the measure of residual risk $\overline{\lambda}$.

[5]It is understood that the size effect may not be risk-based (see Daniel and Titman, 1995), and therefore may be unrelated to beta pricing, but the point of the anecdote is that investors react.

3.2 Exogenous vs. Endogenous Factors

Suppose that k exogenous aggregates explain expected excess payoffs on stocks. These factors can be basically any set of random variables outside the space of asset payoffs. The discussion below Assumption 3 states that it is at least indifferent to replace them with their projections onto the payoff space, i.e. with factor-mimicking portfolio payoffs. Theorem 3 shows that it is strictly better.

Theorem 3 *Replacing factors by their projections onto asset returns reduces the maximum eigenvalue of the covariance matrix of residuals.*

Proof of Theorem 3 See Appendix A.

This is an easy way to give more bite to the Risk Arbitrage Pricing Theory. For a given value of k, Theorem 3 implies that optimal factors are payoffs on carefully chosen portfolios. The exact identity of the optimal set of factors is pinned down by Theorem 4.

Theorem 4 *Let the jth factor be the return on the portfolio whose weight vector is the eigenvector of the covariance matrix of asset returns corresponding to the jth largest eigenvalue. This choice of factors minimizes the largest residual eigenvalue $\overline{\lambda}$.*

Proof of Theorem 4 This is a well-known result from matrix algebra.

The way to give the most bite to the Risk Arbitrage Pricing Theory is to choose the factors associated with the first k eigenvectors of the covariance matrix of asset returns. Relying on exogenous factors makes residuals riskier than they need be, hence reduces the accuracy of beta pricing.

It is worth checking the practical importance of Theorem 4. Stock returns are extracted from the Center for Research on Security Prices (CRSP) database. In order to maximize the number of stocks n, the universe consists of all the stocks traded on the New York Stock Exchange (NYSE) or the American Stock Exchange (AMEX) with less than 10% missing observations over the period considered. In order to minimize estimation error on the covariance matrix of payoffs (see Section 6.1), a high sampling frequency is chosen: daily. Nevertheless, the first and the second moments of returns are always quoted on an annual basis. At the daily frequency, non-synchronous trading is an important issue: I apply Korkie's (1989) refinement of Shanken's (1987b) technique to adjust covariance and beta estimates for non-synchronous trading up to three lags. To minimize estimation error on expected excess payoffs (see Section 4), a long period is chosen: 20 years. The period covers the first 20 years of the CRSP daily database (July 1962 to June 1982). The last twenty years might have been more relevant, but they contain the Crash of 1987, an outlier which severely affects second moments. The rule here is to put beta pricing models in their best light. The data contain 5017 daily returns on 1019 stocks. In addition, the CRSP value-weighted NYSE and AMEX index return including dividends is used as exogenous market factor.

It is well-known that the value-weighted index and the portfolio corresponding to the first eigenvector of the covariance matrix have highly correlated payoffs. This suggests that, when $k = 1$, the single factor can be chosen either way indifferently. With the data described above, the maximum residual eigenvalue $\overline{\lambda}$ is equal to 3.9 if the factor is the value-weighted index, vs. 1.9 for the first eigenvector. Therefore the mean squared deviation can be cut in half by moving from the value-weighted index to the first eigenvector. In general, results are even more impressive

for higher numbers of factors, but they vary a lot depending on the set of exogenous factors.

3.3 Number of Factors

Since the original formulation of the APT assumes $n \to \infty$, some authors have tried to find the number of factors by considering a sequence of increasingly large universes of stocks. The sequence is obviously bounded by the finite universe of stocks that are actually traded, but supposedly some information can be gleaned about behavior at infinity. I believe that this is a case of reverse logic. The vision underlying such efforts is that our economy is an approximation, and that the truth is the infinite economy; whereas the correct statement is that our economy is the true one, and that the infinite economy is but a tractable approximation. The only question is whether or not it is a good approximation. One advantage of the Risk Arbitrage Pricing Theory is that it shows explicitly how closely our economy with finite n is approximated by the infinite APT.

In the APT, the distinction between factors and residuals is that factor eigenvalues are infinitely larger than residual eigenvalues (see Chamberlain and Rothschild, 1983). For finite n this criterion cannot be taken litterally but, since the difference between factors and residuals is supposed to be nearly infinite, it should at least be apparent to the naked eye. Figure 1 plots the 100 largest eigenvalues of the covariance matrix of stock returns formed with the data described above.[6] The naked eye can only see one, or maybe two or three, factors. For higher values of k, it does not make sense to pretend that factor eigenvalues are infinitely larger than residual eigenvalues.

[6]The Risk Arbitrage Pricing Theory is presented in Section 2 in terms of payoffs rather than returns. Because of limited liability, stock prices are positive, so it is possible to normalize payoffs so that every stock price is equal to one. With this convention, payoffs are equal to returns.

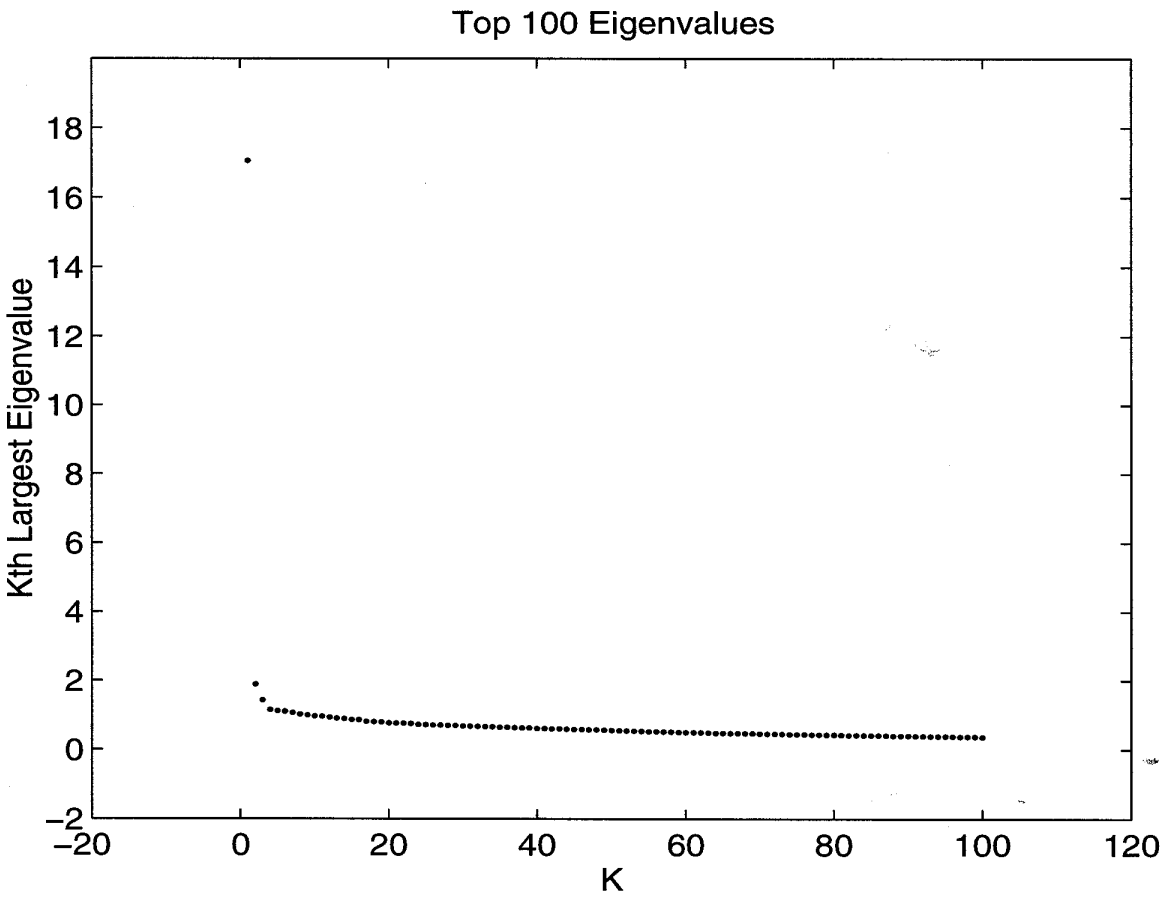

Figure 1: Top 100 Eigenvalues of the Covariance Matrix of Stock Returns. The covariance matrix of NYSE and AMEX stock returns is estimated from daily CRSP data over 7/62-6/82. There are 5017 observations on 1019 stocks.

In spite of the argument that a theoretically infinite gap should be apparent to the naked eye, all the literature on the number of APT factors asks whether there is enough information in the data to isolate k factors at the usual confidence level. Given enough observations, the answer is yes for any value of $k < n$. Even if the nth factor explains a microscopic fraction of payoff variability, we can reject an $(n-1)$-factor structure, as long as we accumulate enough observations. Any other conclusion would require some nongeneric restriction, i.e. one that is only verified on a measure zero subset of the parameter space.[7]

The approach in the literature gives more importance to the number of observations than is intuitively satisfying. When we ask whether the economy can be approximated by a k-factor structure, I believe that the correct meaning of "approximated" is not a statistical one, but an economic one. The question is whether deviations from a k-factor beta pricing model are economically small. The typical deviation can be measured by the square root of the mean squared deviation. Figure 2 plots it, as given by Equation (7). If I say arbitrarily that 1% deviation on annual expected returns is economically small, then this implies at least $k = 68$ factors.

Notice that this result conflicts with the one based on an infinite gap between eigenvalues. The conflict suggests that the standard APT may not be a realistic model of the economy. Fortunately, its core contribution is captured in the more realistic framework of the Risk Arbitrage Pricing Theory.

[7]This is known as Lindley's paradox in statistics.

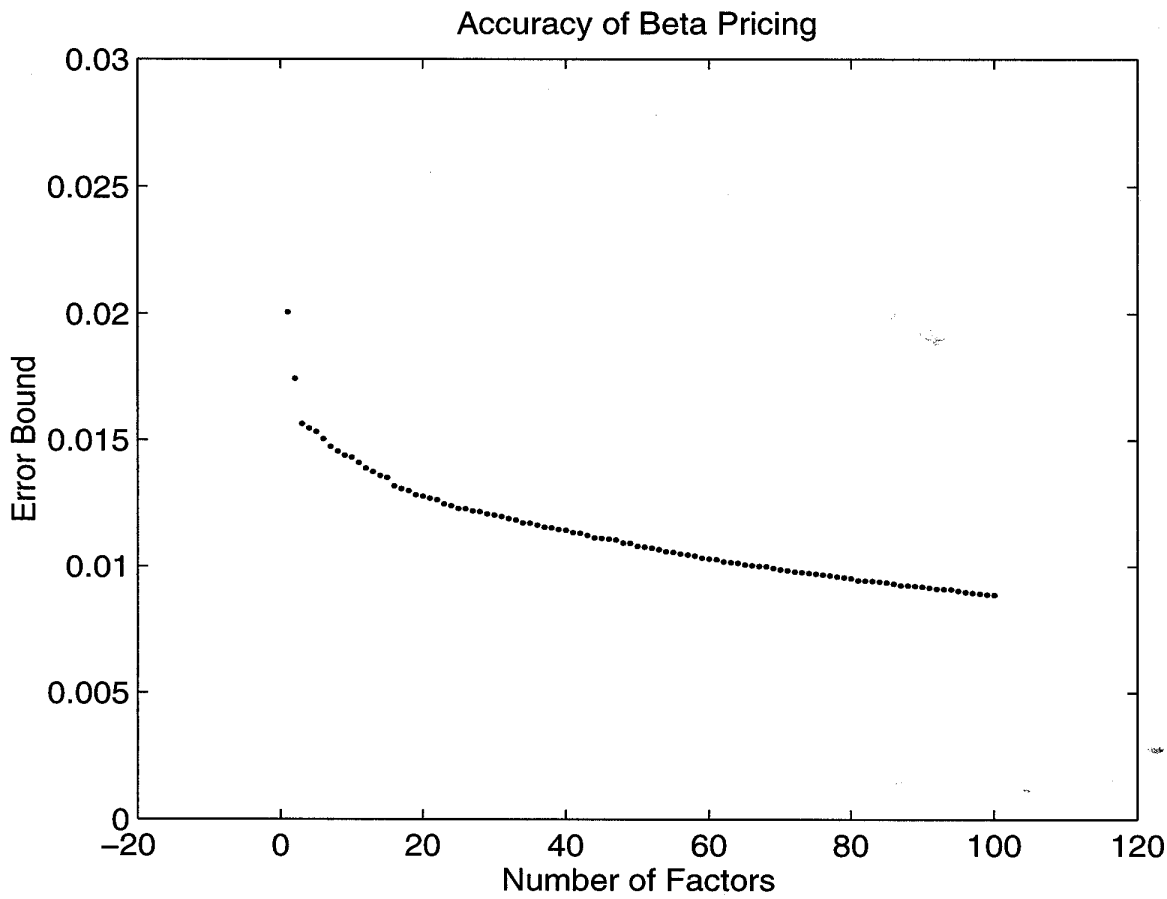

Figure 2: Beta Pricing Error Bound. This graph plots the upper bound on mean squared deviations from beta pricing in Equation (7).

3.4 Accuracy

The accuracy of our version of the CAPM can be gauged from Equation (8), with the following inputs: $\overline{\lambda} = 3.9$, $\underline{\rho} = 1/\sqrt{2}$, $S_M = 0.47$. The market Sharpe ratio S_M comes from the value-weighted NYSE and AMEX index including dividends, using the whole CRSP daily database (July 1962 to December 1994).[8] Typical deviation of annual expected returns from the model is 2.9%. This is not very accurate. It is almost as high as the cross-sectional standard deviation of the expected returns predicted by the model: 3.1%. Even if expected returns were observable without error, an almost flat relationship between returns and market betas would be compatible with the model.

The APT with factors extracted from the covariance matrix of stock returns fares much better. With $k = 68$ factors, the typical deviation of annualized expected stock returns from the beta pricing equation can be pushed down to 1%.

4 Effect of Risk Premium Estimation Error

The conclusions reached above ignore estimation error. This section updates them by accounting for estimation error of expected excess payoffs on factors, i.e. of factor risk premia.

4.1 Objective

Factor risk premium estimation error impacts the value of S_F on the right hand side of Equation (6). The sample estimator of S_F has upward bias because it comes from a maximization program

[8] It is higher than the one obtained from monthly sampling because market indices are heavily autocorrelated (Lo and MacKinlay, 1988).

with erroneous inputs: the program accumulates input errors in its favor. This is a familiar data mining problem. Searching for the portfolio of factors with the maximum Sharpe ratio amounts to mining the data for the best ex-post investment opportunity, with no guarantees as to its ex-ante characteristics. The magnitude of this problem increases dramatically in the number of factors. The solution is simply to reduce the number of factors over which the maximization is conducted to one, for example by setting S_F equal to S_M when the market factor is present. This reinforces on statistical grounds the recommendation made in Section 3.1 on economic grounds.

Even though the sample estimator of S_M is unbiased, it contains error. This affects the right hand side of Equation (7). At this stage, it matters whether Assumption 4 is stated in terms of $\underline{\rho}$, $\overline{S}$ or $\overline{S}_h$. The most elegant solution is to state the assumption directly in terms of $\overline{S}_h$, i.e. to rewrite Equation (7) as:

$$\frac{1}{n}\sum_{i=1}^{n}\alpha_i^2 \leq \frac{\overline{\lambda}\,\overline{S}_h^2}{n}. \tag{8}$$

Remember that $\overline{S}_h$ is the upper bound specified *a priori* by the economist on the Sharpe ratio of the optimal orthogonal portfolio.

The most important effect of risk premium estimation error is that the expected excess payoff of stocks predicted by the beta pricing model are erroneous. Let $\tau_j = \mathrm{E}[\widetilde{f}_j]$ denote factor risk premia, and let $\widehat{\tau}_j$ denote unbiased estimators ($j = 1, \ldots, k$). We must rewrite Equation (8) as:

$$\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{E}\left[\widetilde{x}_i\right] - \sum_{j=1}^{k}\beta_{ij}\widehat{\tau}_j\right)^2\right] \leq \frac{\overline{\lambda}\,\overline{S}_h^2}{n} + \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}\left[\sum_{j=1}^{k}\beta_{ij}\widehat{\tau}_j\right]. \tag{9}$$

Therefore the objective is not only to minimize $\overline{\lambda}$, but also to minimize the effect of risk premium estimation error.

4.2 Exogenous vs. Endogenous Factors

The variance term on the right hand side of Equation (9) is decreasing in the length of the estimation period. Inevitably we must get into a discussion of nonstationarity. If we extract factors from the covariance matrix of stock returns formed with the past 20 years of data, we implicitly assume that the covariance structure of the stock market has not changed much over the past 20 years. Most people would feel that this is a bit of a stretch, and recommend 10 or even 5 years instead. Even the list of traded stocks changes dramatically over 20 years. Therefore risk premia on endogenous factors are estimated over relatively short periods, which means that they have relatively high variances.

By contrast, exogenous factors with an economic interpretation can have their risk premia estimated over much longer periods. If exogenous factor risk premia are characteristic of the whole economy, then they change more slowly than endogenous factor risk premia tied to individual stocks. Companies are created and liquidated every day, but the economy stays more or less the same. Even if the exogenous factors are projected onto asset payoffs, the projection can be run anew every 5 years without problem. For example, the market risk premium is often estimated over the whole CRSP monthly database: 1926-1994, i.e. 68 years. This results in relatively low variance. This effect compensates to some extent for the disadvantages of exogenous factors described in Section 3.2.

Accounting for risk premium estimation error, the typical deviation from a single factor beta pricing model is as follows. For the market factor: 3.7%; for the first covariance matrix eigenvector: 3.5%. The risk premium for the market factor was estimated over 68 years, vs. 20 years for the first eigenvector. As anticipated, the gap has narrowed down: the disadvantage to

using exogenous factors is not as decisive as in the previous section.

4.3 Number of Factors

In the case where factors are covariance matrix eigenvectors, the variance term in Equation (9) can be characterized further. Let t denote the number of observations available. I assume that they are independent and identically distributed (iid). The jth factor risk premium $\tau_j = \mathrm{E}[\widetilde{f}_j]$ is estimated by the sample mean $\widehat{\tau}_j$ of the return on factor $\widetilde{f}_j$ over the t observations. Note that the variance of the return on $\widetilde{f}_j$ is the jth largest eigenvalue λ_j of the covariance matrix of asset returns Σ. Therefore the estimation error on the jth factor risk premium is: $\mathrm{E}[(\widehat{\tau}_j - \tau_j)^2] = \lambda_j/T$. Also note that the largest residual eigenvalue $\overline{\lambda}$ is equal to λ_{k+1}, if there are k factors. Theorem 2 can be modified to account for risk premium estimation.

Theorem 5 *With the above notation, Assumptions 1-4 imply the following bound on deviations from beta pricing with estimated factor risk premia:*

$$\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{E}\left[\widetilde{x}_i\right] - \sum_{j=1}^{k}\beta_{ij}\widehat{\tau}_j\right)^2\right] \leq \frac{\lambda_{K+1}\,\overline{S}_h^2}{n} + \frac{1}{nt}\sum_{j=1}^{k}\lambda_j. \tag{10}$$

Proof of Theorem 5 See Appendix B.

The first term on the right hand side of Equation (10) accounts for residual risk, and the second term for risk premium estimation error. The first term decreases in the number of factors k, while the second term increases in k. Theorem 5 shows how to determine the optimal number of factors as a trade-off between accuracy ($\lambda_{k+1}\overline{S}_h^2/n$) and parsimony ($\sum_{j=1}^{k}\lambda_j/nt$) in the beta pricing equation. Of course, this trade-off depends critically on the length of the estimation period

t.

The desire for parsimony instilled by risk premium estimation error is welcome, since Section 3.3 pointed to numbers of factors that were intuitively too large ($k = 68$ and above). Using Theorem 5, Figure 3 plots the square root of the mean squared deviation from beta pricing due to residual risk alone, risk premium estimation error alone, and total. Large values of k are heavily penalized by risk premium estimation error. It makes typical deviations jump to 3.52%, 3.51% and 3.52% for $k = 1$, 2 and 3 respectively. For values of k above 3, it grows regularly. Mean squared deviation is minimized at $k = 2$.[9]

4.4 Accuracy

The beta pricing model based on the market factor is not very accurate: it is consistent with typical deviations of 3.7% annually. This is higher than the cross-sectional dispersion of expected returns predicted by the model (3.1%). Even if expected stock returns were observed without error, a negative relationship between returns and market betas would be compatible with the CAPM, provided that the correlation between true and observed market portfolio returns is $1/\sqrt{2}$. This model allows too much deviation to be of practical use.

The APT with factors extracted from the covariance matrix of stock returns fares only slightly better. Using the optimal number of factors $k = 2$, the typical deviation is 3.5%. While it is better than using the market factor, it is not good enough for the model to be of practical use either.

[9]The optimal number of factors is sensitive to the choice of $\overline{S}_h$. Higher values of $\overline{S}_h$ give more importance to the residual risk term, which tends to increase the optimal number of factors.

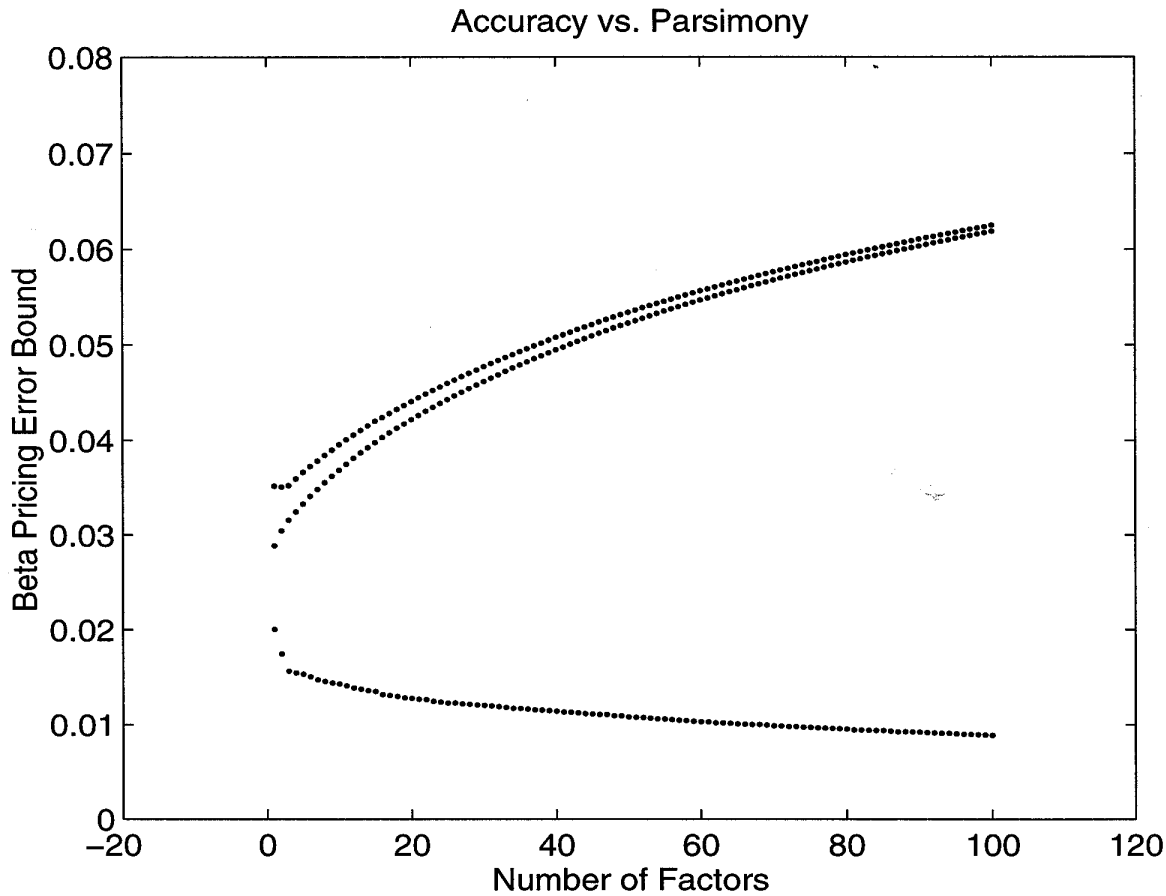

Figure 3: Accuracy vs. Parsimony. The lower dots plot the square root of the first term on the right hand side of Equation (10). It represents the deviation from beta pricing due to residual risk. It decreases in the number of factors k. The middle dots plot the square root of the second term on the right hand side of Equation (10). It represents the deviation from beta pricing due to risk premium estimation error. It increases in the number of factors k. The upper dots plot the square root of the sum of the two terms on the right hand side of Equation (10). It represents the total deviation from beta pricing. It is minimized for $k = 2$ factors. Choosing k to minimize the total deviation from beta pricing involves a trade-off between accuracy (with residual risk) and parsimony (with risk premium estimation error). The solution of this trade-off is the optimal number of factors. The optimal number of factors is quite small. Even at the optimum, deviations from beta pricing are rather large.

5 Defining Risk Arbitrage

The definition of risk arbitrage requires specifying *a priori* an upper bound $\overline{S}$ on Sharpe ratios. There are several ways to do this.

First, the best way is probably to let readers decide for themselves. It is always possible to consider a wide range of values for $\overline{S}$, and report the mapping of $\overline{S}$ to the results. On a smaller scale, it is always desirable to give an idea of the sensitivity of results to the choice of $\overline{S}$.

The second best way is to estimate $\overline{S}$ over the past, and then make the economic assumption that it will stay the same in the future. The sample maximum Sharpe ratio has severe upwards bias, but Jobson and Korkie (1980) give a modified estimator that is unbiased under normality. Using this estimator, MacKinlay (1995) finds $[0.49, 1.40]$ as a 90% confidence interval for the maximum annualized Sharpe ratio. By comparison, he also reports the sample Sharpe ratio of a value-weighted index from the CRSP monthly database: 0.33.

The other ways to determine $\overline{S}$ require stronger assumptions on investor preferences. For example, assuming mean-variance preferences, every investor holds a portfolio that is mean-variance efficient in the universe of all traded assets, possibly after hedging her endowment in non-traded assets. The Sharpe ratio of this portfolio is a function of the risk aversion of the population. Laboratory experiments can give the risk aversion of a subset of the population. Then the results can be extrapolated to the population as a whole. This pins down the maximum Sharpe ratio in the universe of all traded assets. It can be used as an upper bound for the maximum Sharpe ratio among portfolios of the $n + k + 1$ assets under consideration.

In general, the more structure one is willing to impose, the easier it is to come up with a value for $\overline{S}$. The following theorem illustrates this in a world with non-traded assets.

Theorem 6 *Assume that agents exhibit constant absolute risk aversion and that future asset values are normally distributed. Assume that all excluded assets are non-traded. Let v_T (respectively v_{NT}) denote the variance of the future dollar value of all traded (resp. non-traded) assets in the economy. If the correlation between traded and non-traded assets is non-negative then, in equilibrium, the maximum Sharpe ratio among portfolios of traded assets cannot exceed*

$$\sqrt{1+\frac{v_{NT}}{v_T}}\,S_M \tag{11}$$

where S_M is the Sharpe ratio of the value-weighted portfolio of traded assets.

Proof of Theorem 6 See Appendix C

In this setup, $\overline{S}$ can be set equal to the quantity in Equation (11), if we can measure the variance of the future dollar value of non-traded assets. For example, if the variances of the values of traded and non-traded assets are the same, then we can take $\overline{S} = \sqrt{2}S_M$.

Still assuming that all agents have mean-variance preferences and that all excluded assets are non-traded, the only cause for heterogeneity across investor holdings is heterogeneity in non-traded asset endowments. The only cause of inefficiency for the value-weighted portfolio of traded assets is aggregate endowment in non-traded assets. Since heterogeneity across investor holdings is observable, we know a lot about heterogeneity in non-traded asset endowments. It would be tempting to try to turn it into information about the variance of the aggregate endowment in non-traded assets. It is basically like trying to infer the variance of the market factor from observing each stock's idiodyncratic variance. It is possible if we know the typical correlation between two stocks. Here, it means knowing the typical correlation between two agents' non-

tradeable endowments. If we do, then we can get a handle on the inefficiency of the market portfolio, i.e. on $\overline{S}$, by measuring investor holding heterogeneity.

Ross (1976, p.354) and MacKinlay (1995) specify *a priori* the value $\overline{S} = \sqrt{2}S_M$, where S_M is the Sharpe ratio of a broad-based value-weighted stock market index. There is an elegant interpretation of this number: it means that the maximum Sharpe ratio among portfolios uncorrelated with the market portfolio is equal to the Sharpe ratio of the market portfolio. This choice implies that no portfolio that consistently earns the same average return as the market portfolio with less than half the variance can survive in equilibrium. Nothing could stop investors from pouring enough money into such a portfolio to bring down its Sharpe ratio. Breen, Glosten and Jagganathan (1989) show that information in Treasury Bill yields can be used to construct a dynamic portfolio of stocks whose Sharpe ratio is just about equal to $\sqrt{2}S_M$.

In summary, there is hope that comparing the results from fundamentally different approaches can give us firm guidance about specifying $\overline{S}$. $\overline{S}$ depends a lot on the set of assets under consideration: it will not be the same if they are stocks only, long-term or corporate bonds, or even options. For a universe of stocks, it seems that most reasonable values of $\overline{S}$ are near or above $\sqrt{2}S_M$, where S_M is the Sharpe ratio of a market index.

6 Limitations and Extensions

This section acknowledges some of the limitations of the paper and outlines extensions that are open for future research.

6.1 Effect of Covariance Matrix Estimation Error

In the numerical results presented above, the most striking fact is that the maximum residual eigenvalue $\overline{\lambda}$ is very large. This could be partially due to covariance matrix estimation error. It is well known that the maximum eigenvalue of a matrix estimated with error is severely biased upwards. Therefore the conclusions reached about the accuracy of beta pricing models should be taken with caution. However, this type of estimation error cuts both ways. It also creates effects that may compensate for the one just mentioned.

One such effect is to blur the distinction between factors and residuals. Suppose that we only have an estimate the true covariance matrix Σ. Beta pricing theories call for decomposing the portfolio space into an estimated factor space and an estimated residual space. By construction, the estimated correlation of a factor with a residual is zero; but the true correlation could be nonzero. Preliminary evidence indicates that the maximum true correlation between estimated factors and estimated residuals can be extremely high as soon as there are more than a few factors. This is a disturbing twist on the properties of factors and residuals.

Consider the problem facing the risk arbitrageur. This is the agent whose existence is assumed by the Risk Arbitrage Pricing Theory. He enforces the theory's predictions by taking advantage of risk arbitrage opportunities tirelessly until they disappear. He locates the residual space, determines whether some residuals command higher returns than allowed by the theory and, if so, invests in them until their prices rise. His problem is that he can only use an erroneous covariance matrix estimator to try and locate the residual space. Some portfolios in the estimated residual space have high true correlations with factors. The risk arbitrageur may end up bearing unwanted factor risk. He may feel that his expected return compensates him enough for exposure

to residual risk, but not to factor risk. For this reason, he may drop out of the business of risk arbitrage altogether, and deviations from beta pricing may turn out to be even higher than the Risk Arbitrage Pricing Theory predicts.

Another way in which covariance matrix estimation error blurs the distinction between factors and residuals is as follows. In the standard APT, factor eigenvalues are infinitely larger than residual eigenvalues. In a finite economy, this cannot be the case, but at least we would expect factor eigenvalues to be larger than residual eigenvalues. Preliminary results indicate the contrary: as soon as there are more than a few factors, some true eigenvalues in the estimated factor space are smaller than some true eigenvalues in the estimated residual space. Confirmation of this result would be the final nail in the coffin for the idea that Nature separates factors from residuals.

Finally, consider the impact of covariance matrix estimation error on eigenvalues and eigenvectors. If we knew the true covariance matrix, then we could choose the k factors optimally as in Theorem 4, and the maximum residual eigenvalue would be equal to λ_{k+1}, the $(k+1)^{\text{th}}$ largest eigenvalue. However, we only know an estimate of the covariance matrix, so we choose the k factors suboptimally. This implies that the largest eigenvalue in the estimated residual space is larger than λ_{k+1}.

Overall, the effect of covariance matrix estimation error on the conclusions reported above is ambiguous, and further research is warranted.

6.2 Notions of Arbitrage

The paper can be criticized for relying too much on the Sharpe ratio. The Sharpe ratio need not summarize the attractiveness of a portfolio, unless all investors have mean-variance preferences,

and that is a very restrictive assumption. A reply to this criticism is that the Risk Arbitrage Pricing Theory does not assume that investors hold mean-variance efficient portfolios. Some amount of deviation from mean-variance preferences is built into the theory through Assumption 4. There is also another reply that opens up an interesting direction for future research. It requires first a review of the competing notions of arbitrage.

In a mean-variance framework, at least three notions of arbitrage can be defined. As we do throughout the paper, we assume the existence of a riskless asset and use it to construct excess payoffs.

Definition 1 *A* pure arbitrage *opportunity is a portfolio whose excess payoff has strictly positive expectation and zero standard deviation, i.e. its Sharpe ratio is equal to* $+\infty$.

Definition 2 *A* limiting arbitrage *opportunity is a sequence of portfolios whose excess payoffs have expectations bounded below by a strictly positive number, and standard deviations converging to zero, i.e. have Sharpe ratios going to infinity.*

Definition 3 *A* risk arbitrage *opportunity is a portfolio whose Sharpe ratio exceeds the prespecified level* $\overline{S}$.

These three notions are clearly related. Very loosely speaking, risk arbitrage is a watered-down version of limiting arbitrage, which is itself a watered-down version of pure arbitrage. Definition 2 is close to what is used in the APT, and Definition 3 is exactly what is used in the Risk Arbitrage Pricing Theory.

The only problem with these definitions is that they have a strong mean-variance flavor. There exists a more general notion of pure arbitrage, which is perhaps more widespread and more satisfying than Definition 1, because it is not tied to mean and variance.

Definition 4 *A pure arbitrage opportunity is a portfolio whose excess payoff is nonnegative with probability one, and is strictly positive with strictly positive probability.*

This is a direct competitor to Definition 1. A pure arbitrage according to Definition 4 is a pure arbitrage according to Definition 1, but the converse is not necessarily true. Consider a lottery whose excess payoff is \$1 or \$2, with probability $1/2$ each. Most people would describe this lottery as a pure arbitrage opportunity, and indeed it is one according to Definition 4, but not according to Definition 1. Therefore Definition 1 is not fully satisfying. Obviously, this spills over to Definitions 2 and 3. For example, the standard APT does not rule out the existence of this lottery. This is a weakness of the whole mean-variance approach, and the Risk Arbitrage Pricing Theory suffers from it too.

It would be interesting to construct definitions of limiting arbitrage and risk arbitrage that are watered-down versions of Definition 4. This requires creating a ratio that does for Definition 4 what the Sharpe ratio does for Definition 1. I introduce such a ratio below. Decompose the excess payoff x into $x = x^+ - x^-$, where $x^+ = \max(x, 0)$ is the gain, and $x^- = \max(-x, 0)$ is the loss. A pure arbitrage opportunity according to Definition 4 is a portfolio whose gain x^+ has strictly positive expectation, and whose loss x^- has zero expectation. Therefore it seems natural to introduce the following ratio:

$$\frac{\mathrm{E}[x^+]}{\mathrm{E}[x^-]}. \tag{12}$$

This ratio is well-defined if payoffs have finite first moments, except that it is indeterminate for the riskless asset. It is invariant with respect to multiplication by a scalar and addition of the riskless asset. The ratio is always nonnegative and can be equal to $+\infty$. It is infinite if and only if the security is a pure arbitrage opportunity according to Definition 4. While the Sharpe ratio

weighs the mean against the standard deviation, this ratio weighs the gain against the loss. It is now possible to redefine limiting arbitrage and risk arbitrage in the spirit of Definition 4.

Definition 5 *A* limiting arbitrage *opportunity is a sequence of portfolios whose ratios defined by Equation (12) go to infinity.*

Definition 6 *A risk arbitrage opportunity is a portfolio whose ratio defined by Equation (12) exceeds some prespecified level.*

The paper does rely on a somewhat restrictive notion of arbitrage through the use of the Sharpe ratio, but it would be possible to switch to the more general notion of arbitrage by using the ratio defined in Equation (12) instead.[10]

7 Conclusion

Beta pricing models are presented in exact form most of the time. Since a model is by definition a simplified representation of reality, it is well understood that it cannot hold exactly, and usually the discussion stops here. Some authors have gone far enough to give a back-of-the-envelope evaluation of the accuracy of versions of the APT (Ross, 1976; Dybvig, 1983; Grinblatt and Titman, 1983; Shanken, 1992) because of concern over an approximation in the proofs.

The view that I advocate here is quite different: the accuracy of beta pricing models has a fundamental importance. It deserves careful study. It is the key to choosing between models with alternatives sets of factors, and even to figuring out whether beta pricing has any practical use at all.

[10]The author is currently participating in a research project in this area.

It is interesting to see how fast risk premium estimation error penalizes models with more than one factor. Judging from the results reported here, a beta pricing model with a single factor representing the market is still the best we have.

Another interesting point is how much is lost in moving from a tight bound on $\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha}$ to a less tight bound on the root mean squared alpha $\sqrt{\boldsymbol{\alpha}'\boldsymbol{\alpha}/n}$. It amounts to replacing every eigenvalue in the covariance matrix of residuals $\boldsymbol{\Omega}$ by the maximum eigenvalue $\overline{\lambda}$, and $\overline{\lambda}$ is much larger than the typical eigenvalue in $\boldsymbol{\Omega}$. This is why beta pricing models perform so poorly in the root mean squared alpha metric. It is unfortunate, since is metric this the one most easily understood by non-specialists. But we now we have to determine the practical use of beta pricing models that are only accurate in the metric $\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha}$.

Multivariate tests of asset pricing models use the right metric, but cross-sectional tests use the wrong one. Most cross-sectional regressions of returns on betas use Ordinary Least Squares, which is equivalent to minimizing $\sqrt{\boldsymbol{\alpha}'\boldsymbol{\alpha}/n}$. Therefore they are much less meaningful than the seldom used Generalized Least Squares regressions that would minimize $\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha}$.

The Risk Arbitrage Pricing Theory provides a framework well suited to the analysis of these important questions.