AP 2017–01

Chair of Applied Stochastics and Risk Management

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg
Faculty of Economics and Social Sciences
Department of Mathematics/Statistics

Working Paper

Arbitrage Pricing Theory in Ergodic Markets

Gabriel Frahm

July 31, 2018

Working Paper

AP 2017–01
July 31, 2018

Arbitrage Pricing Theory in Ergodic Markets

Gabriel Frahm

Helmut Schmidt University
Faculty of Economics and Social Sciences
Department of Mathematics/Statistics
Chair of Applied Stochastics and Risk Management
Holstenhofweg 85, D-22043 Hamburg, Germany

URL: www.hsu-hh.de/stochastik
Phone: +49 (0)40 6541-2791
E-mail: frahm@hsu-hh.de

Working Paper
Please use only the latest version of the manuscript. Distribution is unlimited.

Supervised by: Prof. Dr. Gabriel Frahm
Chair of Applied Stochastics and
Risk Management

URL: www.hsu-hh.de/stochastik

Arbitrage Pricing Theory in Ergodic Markets*

Gabriel Frahm†
Helmut Schmidt University
Department of Mathematics and Statistics
Chair of Applied Stochastics and
Risk Management

July 31, 2018

Abstract

Traditional approaches to Arbitrage Pricing Theory (APT) propose a factor model, but empirical applications of APT are, nowadays, based on seemingly unrelated regression. I drop the factor model and assume only that the market is ergodic. This enables me to apply the theory of Hilbert spaces in a natural way. The expected return on any asset can always be approximated by an affine-linear function of its betas and we are able to estimate the relative number of assets that violate the APT equation by taking the expected returns and betas in the market into account. I present a simple sufficient condition for the APT equation in its inexact form. Further, I show that the APT equation holds true in its exact form if and only if an equilibrium market is exhaustive, which means that it must be possible to replicate the betas and idiosyncratic risk of each asset by some strategy that diversifies away all approximation errors in the market.

Keywords: Arbitrage Pricing Theory; beta; common risk; ergodicity; expected return; factor model; idiosyncratic risk; seemingly unrelated regression.

JEL Subject Classification: G12, G11.

---

*I am very grateful to Philipp Adämmer for his helpful comments on the manuscript and our valuable discussions.
†Phone: +49 40 6541-2791, e-mail: frahm@hsu-hh.de.

1. Motivation

The Arbitrage Pricing Theory (APT) and the Capital Asset Pricing Model (CAPM) are commonly seen as the most significant theories of capital market. Stephen Ross develops the APT in a working paper in 1971, which is published later on in Ross (1982). He discusses the APT in Ross (1976a) and in Section 9 of Ross (1976b). Ross vividly advocates the APT, typically in collaboration with Richard Roll, in a series of contributions in the 1980s. Roll and Ross (1984a) show how APT can be implemented in portfolio management. Roll and Ross (1980) investigate the APT empirically, which is followed by a reply (Roll and Ross, 1984b) to some critical remarks by Dhrymes et al. (1984). Further, Roll and Ross (1983) compare the APT with its main competitor, i.e., the CAPM, whereas Ross and Walsh (1983) describe how the APT can be used for international asset pricing.

The CAPM is based on the assumption that all market participants aim at maximizing a mean-variance objective function, whereas the APT makes (almost) no behavioral assumption. It presumes only that market participants try to exploit arbitrage opportunities and so it can be considered a robust alternative to the CAPM. The main idea of APT is simple: If there exists an arbitrage opportunity, the market cannot be in equilibrium and, conversely, if the market is in equilibrium, there cannot exist any arbitrage opportunity. Ross (1976a) argues that, in this case, the expected return on Asset i must essentially be

$$\mu_i = \lambda_0 + \sum_{j=1}^{m} \lambda_j \beta_{ij} \tag{1}$$

for any positive integer m. More precisely, the APT equation asserts that the cost of capital of Asset i, μ_i, is an affine-linear function of its betas, $\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}$, where λ_0 represents the time value of money and $\lambda_1, \lambda_2, \ldots, \lambda_m$ are referred to as the market prices of risk.

The behavioral assumption of APT, namely that the market is arbitrage-free, seems to be negligible from an economic point of view, but the structural assumptions that are usually made in the literature are quite strong. It is typically assumed that the return on Asset i follows a factor model

$$R_i = \mu_i + \sum_{j=1}^{m} \beta_{ij} F_j + \epsilon_i, \tag{2}$$

where F_j is some unobservable risk factor, β_{ij} is the associated factor loading, and ϵ_i denotes the idiosyncratic risk of Asset i. The reader can find a large number of structural conditions that guarantee that the idiosyncratic risks can be diversified away (see, e.g., Admati and Pfleiderer, 1985, Chamberlain and Rothschild, 1983, Connor and Korajczyk, 1995, Huberman, 1982, Ingersoll, 1984, Jarrow and Rudd, 1983, Reisman, 1988). The problem is that the factor model is very restrictive. In its most stringent form it presumes that the idiosyncratic risks are uncorrelated. Weaker forms allow $\epsilon_1, \epsilon_2, \ldots$ to be correlated, but the correlation must be sufficiently low in order to preserve their diversifiability.[1]

The CAPM is not based on a factor model but on the mean-variance portfolio-optimization

[1] For a nice overview of factor models that are used in APT see, e.g., Connor and Korajczyk (1995).

approach developed by Markowitz (1952). It does not make any distributional assumption—besides the basic assumption that the second moments of the asset returns are finite.[2] Hence, it seems to be highly desirable to drop the factor model in APT, which is actually done by McElroy et al. (1985). They observe that Eq. 2 is a linear regression equation and so the asset returns can be represented by a system of seemingly unrelated regression (SUR) equations. Nonetheless, their focus is empirical rather than theoretical. To understand their arguments, I should note that Roll and Ross (1980) propose a two-step estimation procedure for the market prices of risk. In the first step, the expected returns and betas are estimated by maximum likelihood, based on the given factor model of asset returns. In the second step, the market prices of risks are estimated by applying an ordinary least-squares (OLS) regression of the expected-return estimates on the beta estimates that have been obtained in the first step. Estimating the parameters of a factor model requires a distributional assumption. If this assumption is violated, the resulting estimators are inconsistent or, at least, inefficient. Moreover, the two-step estimation approach leads to a well-known errors-in-variables bias: Even if the proposed factor model is correct, the two-step estimator for the market prices of risk is inconsistent if the number of assets grows to infinity but the number of time-series observations stays fixed. To reduce the errors-in-variables bias, the assets are typically grouped into portfolios, which makes the entire approach somewhat arbitrary and, unfortunately, it turns out that the empirical results essentially depend on the chosen groups (Antoniou et al., 1998, Clare and Thomas, 1994). Another problem is that the factors are unobservable, which makes an economic interpretation difficult—not to say impossible (Chen and Jordan, 1993). These problems can be avoided by applying a nonlinear SUR estimator (McElroy et al., 1985),[3] which is based on observable variables and estimates the parameters of the SUR equations, i.e., the betas and the market prices of risk, simultaneously.[4]

To the best of my knowledge, there is no theoretical foundation of APT in the context of SUR. Is it still possible to justify APT if we drop the factor model? At first glance, this question seems to be of minor importance, at least from an econometric point of view, since the factor model can be considered a SUR in which the residuals are uncorrelated in the cross section.[5] Hence, the SUR estimators might be inefficient but, apparently, not inconsistent. I will show that this naive conclusion is utterly wrong. APT can only be justified if the SUR equations are properly specified, i.e., we must not omit any risk factor whose market price differs from zero. Nevertheless, even if the SUR equations are properly specified, we cannot guarantee that the APT equation holds true for each asset in the market. APT just states that Eq. 1 must *essentially* be satisfied in an equilibrium market. The problem is that the number of assets for which the APT equation is violated as well as the magnitude of each approximation error, i.e., $u_i := \mu_i - \lambda_0 - \sum_{j=1}^{m} \lambda_j \beta_{ij}$, can be arbitrarily large.

The main contributions of this work are as follows:

(i) APT is discussed in the context of ergodic markets. Ergodicity enables us to apply the

[2] It is not presumed that the asset returns have a joint normal distribution (Markowitz, 2012).

[3] The attribute "nonlinear" is somewhat misleading. In fact, the regression equations are linear, but the parameter restrictions mentioned by McElroy et al. (1985) are nonlinear.

[4] McElroy et al. (1985) presume that the time value of money, λ_0, equals the risk-free interest rate.

[5] I do not distinguish between observable and unobservable regressors.

theory of Hilbert spaces in a natural way and thus to weaken the basic assumptions of APT, substantially.

(ii) It is shown that the expected return on any asset can always be approximated by an affine-linear function of its betas. Further, a useful result concerning the approximation quality is provided.

(iii) It is proved that the approximation errors are essentially zero if and only if the market prices of all common risks that are omitted in the return equation equal zero, provided that there *exists* some properly specified return equation.

(iv) A simple sufficient condition for the APT equation in its *inexact* form is presented. It is shown that the condition is implied by the standard assumptions of APT.

(v) It is demonstrated that the APT equation holds true in its *exact* form if and only if it is possible to replicate all betas and idiosyncratic risks such that the approximation errors in the market vanish.

The rest of this work is organized as follows: In Section 2, I introduce the notation and discuss the basic assumptions. Section 3 represents the main part of this work. It is divided into two subsections. Section 3.1 presents the general results and Section 3.2 focuses on equilibrium markets. Section 3.1.1 contains the approximate APT equation, whereas in Section 3.1.2, I discuss the specification problem. Further, Section 3.2.1 contains the inexact APT equation and in Section 3.2.2, I derive the exact APT equation. Finally, Section 4 concludes this work. All proofs can be found in the appendix unless the given statement is trivial.

2. Notation and Basic Assumptions

Any assertion about a random quantity is meant to be valid *almost surely*. Let $\{X_n\}$ be any random sequence. Almost sure convergence is denoted by $X_n \to X$, convergence in probability is indicated by $X_n \xrightarrow{\text{p}} X$, and convergence in distribution is symbolized by $X_n \rightsquigarrow X$. Every tuple $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ is considered a *column* vector, and all random quantities that occur in this work shall have finite second moments. Each statement that involves an index i is meant to be true for *all* $i \in I$, where the index set I should always be clear from the context. The symbol $\mathbf{0}$ represents a vector of zeros, whereas $\mathbf{1}$ is a vector of ones. Further, the identity matrix is denoted by $\mathbf{I}$.

I will frequently use the Linear Regression Theorem, which is a fundamental result of econometrics:

Theorem 1 (Linear Regression Theorem). *Let Y be a random variable, $X_0 \equiv 1$, and $X = (X_1, X_2, \ldots, X_m)$ an m-dimensional random vector with $m \in \mathbb{N}$.*[6] *There exists a vector $(\beta_0, \beta_1, \ldots, \beta_m)$ that satisfies the following equivalent conditions:*

(i) $(\beta_0, \beta_1, \ldots, \beta_m) = \arg\min_{(b_0,b_1,\ldots,b_m)} \mathrm{E}\left(\left(Y - \sum_{j=0}^{m} b_j X_j\right)^2\right)$,

[6] In the case of $m = 0$, the random vector X is void.

(ii) $\mathrm{Var}(X)(\beta_1, \beta_2, \dots, \beta_m) = \mathrm{Cov}(X, Y)$ *and* $\beta_0 = \mathrm{E}(Y) - \sum_{j=1}^m \beta_j \mathrm{E}(X_j)$,

(iii) $Y = \sum_{j=0}^m \beta_j X_j + \epsilon$ *with* $\mathrm{E}(\epsilon) = 0$ *and* $\mathrm{Cov}(X_j, \epsilon) = 0$.

If the covariance matrix of X is positive definite, the vector $(\beta_0, \beta_1, \dots, \beta_m)$ *is unique. It holds that*

$$(\beta_1, \beta_2, \dots, \beta_m) = \mathrm{Var}(X)^{-1}\mathrm{Cov}(X, Y) \tag{3}$$

and

$$\beta_0 = \mathrm{E}(Y) - \sum_{j=1}^m \beta_j \mathrm{E}(X_j). \tag{4}$$

In any case, the orthogonal projection $\sum_{j=0}^m \beta_j X_j$ *and the residual* ϵ *are unique.*

Throughout this work, I say that

$$Y = \beta_0 + \sum_{j=1}^m \beta_j X_j + \epsilon \tag{5}$$

is a linear equation, and it is a linear *regression* equation if and only if the three equivalent conditions mentioned by the Linear Regression Theorem are satisfied. The Linear Regression Theorem does not require any other assumption. It can be applied both to empirical and to theoretical distributions. If the distribution is empirical, we obtain the well-known results from OLS regression. Otherwise, the Linear Regression Theorem provides some useful insights regarding the linear relationship between the random variables $X_1, X_2, \dots, X_m$ and Y.

I assume that the market contains an infinite number of assets, which can either be risky or riskless. Let R_i be the (rate of) return on Asset i after some investment period. Dividends that occur during the investment period are considered part of the corresponding asset return. Further, let $X = (X_1, X_2, \dots, X_m)$ be any random vector. Suppose, without loss of generality, that the components of X are centralized, i.e., $\mathrm{E}(X_j) = 0$, and that the covariance matrix of X is positive definite, i.e., $\mathrm{Var}(X) > 0$.[7] We can always express the return on Asset i by the linear equation

$$R_i = \mu_i + \sum_{j=1}^m \beta_{ij} X_j + \varepsilon_i. \tag{6}$$

The intercept μ_i denotes the expected return on Asset i and the slope coefficients $\beta_{i1}, \beta_{i2}, \dots, \beta_{im}$ are referred to as its betas. Further, the random variables $X_1, X_2, \dots, X_m$ represent some common risks, whereas ε_i is the idiosyncratic risk of Asset i. Hence, $\sum_{j=1}^m \beta_{ij} X_j$ can be considered the systematic part and ε_i the unsystematic part of the unexpected return $R_i - \mu_i$. From $\mathrm{E}(R_i) = \mu_i$ and $\mathrm{E}(X_j) = 0$ it follows that $\mathrm{E}(\varepsilon_i) = 0$.

Eq. 6 can be represented in matrix form as $R = \mu + \mathrm{B}X + \varepsilon$ with $R = (R_1, R_2, \dots, R_n)$,

[7] This assumption is justified in Section A.1.1.

$\mu = (\mu_1, \mu_2, \ldots, \mu_n)$,

$$B = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m} \\ \vdots & \vdots & & \vdots \\ \beta_{n1} & \beta_{n2} & \cdots & \beta_{nm} \end{bmatrix}, \tag{7}$$

and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$. I omit the number n for notational convenience, but the reader should keep in mind that all vectors and matrices refer to the first n assets in the market.

I do not require a strict factor model, i.e., the idiosyncratic risks $\varepsilon_1, \varepsilon_2, \ldots$ need not be uncorrelated. Further, the given return equation (6) need not represent an approximate factor model (Chamberlain and Rothschild, 1983), where the sequence of the largest eigenvalues associated with $\{\mathrm{Var}(\varepsilon)\}$ is bounded above. Further, I do not presume that the common risks are strictly exogenous, i.e., $\mathrm{E}(\varepsilon_i \mid X) = 0$, which is a typical assumption in APT.[8] In factor analysis it is assumed that the common risks are unobservable, whereas they must be observable for SUR estimation. Burmeister and McElroy (1988) propose an estimation procedure that works with both observable and unobservable variables.[9] From a theoretical point of view, it is irrelevant whether the common risks are observable or unobservable. In any case, the common risks are usually correlated with each other and so we may allow the idiosyncratic risk of Asset i, ε_i, to be correlated with $X_1, X_2, \ldots, X_m$, too. Hence, a key observation of this work is that Eq. 6 need not be a linear *regression* equation. This important remark can be found also in Renault et al. (2017), but I did not find any other author drawing attention to this point. The given approach makes (almost) no distributional assumptions regarding the asset returns and common risks. We will see that ergodic markets provide enough structure in order to obtain the typical results of APT, and we are able to produce even more useful insights.

Let $\theta_i := (\mu_i, \beta_{i1}, \beta_{i2}, \ldots, \beta_{im})$ be a vector that contains the expected return on Asset i and its betas. Consider the empirical distribution function of the parameter vectors $\theta_1, \theta_2, \ldots, \theta_n$, i.e.,

$$x \longmapsto F_n(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\theta_i \leq x}, \quad \forall\, x = (x_0, x_1, \ldots, x_m) \in \mathbb{R}^{m+1}. \tag{8}$$

Here $A \mapsto \mathbf{1}_A$ denotes the indicator function, i.e., $\mathbf{1}_A = 1$ if A is true and otherwise $\mathbf{1}_A = 0$. Further, let F be any cumulative distribution function on $\mathbb{R}^{m+1}$. I assume that $F_n(x) \to F(x)$ for all $x \in \mathbb{R}^{m+1}$. Hence, F represents the empirical distribution function of *all* expected returns and betas in the market.

The expected returns and betas are deterministic.[10] Nonetheless, we can *imagine* an $\mathbb{R}^{m+1}$-valued random vector $\theta = (\vartheta_0, \vartheta_1, \ldots, \vartheta_m)$ with cumulative distribution function F. More precisely, let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, where $\mathcal{A}$ is the Borel σ-algebra on $\Omega = \mathbb{R}^{m+1}$ and $\mathbb{P}$ is the probability measure on $\mathcal{A}$ induced by F. Put another way, $\mathbb{P}$ is the *empirical* measure

[8] See, for example, Connor and Korajczyk (1995, p. 97) as well as Roll and Ross (1980, p. 1076). However, strict exogeneity can be important in empirical applications, in particular if somebody wants to estimate the expected returns and betas.

[9] See also Chen and Jordan (1993) for an empirical investigation of factor and time-series analysis in the context of APT.

[10] Gagliardini et al. (2016) propose a similar approach, but they consider $\theta_1, \theta_2, \ldots, \theta_n$ stochastic.

of the expected returns and betas that can be "observed" in the market.[11] Let f be any (Borel) measurable function from $\mathbb{R}^{m+1}$ to $\mathbb{R}$, so that $f(\theta)$ can be considered a random variable. If the integral of $f(\theta)$ with respect to $\mathbb{P}$ exists, it is denoted by $\mathbb{E}\big(f(\theta)\big)$. Any second moments that are based on the empirical measure $\mathbb{P}$ are indicated, in the same way, by the symbols "$\mathbb{V}$ar" and "$\mathbb{C}$ov." These moments should be distinguished from E, Var, and Cov, which refer to the asset returns, the common risks, and the idiosyncratic risks.

I assume that the parameter sequence $\{\theta_n\}$ is ergodic. This means that

$$\frac{1}{n}\sum_{i=1}^{n} f(\theta_i) \longrightarrow \mathbb{E}\big(f(\theta)\big) \tag{9}$$

for each integrable function f of θ, i.e., $\mathbb{E}\big(|f(\theta)|\big) < \infty$. Hence, we can treat the expected returns and betas in the market *as if* they were random variables. However, these parameters are in fact deterministic and thus $\mathbb{E}\big(f(\theta)\big)$ should be understood as an empirical mean rather than a probabilistic expectation. Any market that satisfies the aforementioned requirements is said to be ergodic.

Ergodicity is a typical assumption in econometrics. It guarantees that the sample distribution of time-series or cross-sectional observations converges to their stationary or population distribution, respectively. Thus, although the observations may depend on each other, the empirical distribution should be close to the theoretical one if the number of observations is sufficiently large. However, in our context the "sample" $\Theta = \big[\theta_1\ \theta_2\ \cdots\ \theta_n\big]' = \big[\mu\ \mathrm{B}\big] \in \mathbb{R}^{n\times(m+1)}$ contains the expected returns and betas in the market. Further, n is not a number of observations—it is the number of assets that are taken into consideration. This must be emphasized because ergodic theory usually is applied to time-series or cross-sectional data. By contrast, here I refer to the parameters of the linear equation $R = \mu + \mathrm{B}X + \varepsilon$, which are deterministic but not stochastic. This seems to be a novel approach in APT.[12]

The ergodicity assumption enables us to consider ϑ_j an element of a Hilbert space $\mathcal{H}$, where ϑ_0 represents the expected return and $\vartheta_1, \vartheta_2, \ldots, \vartheta_m$ are the betas of an asset in an *abstract* sense. The inner product of $\mathcal{H}$ is $\mathbb{E}(\vartheta_j\vartheta_k) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\theta_{ij}\theta_{ik}$, i.e., the vectors $\vartheta_j, \vartheta_k \in \mathcal{H}$ are orthogonal if and only if $\frac{1}{n}\sum_{i=1}^{n}\theta_{ij}\theta_{ik} \to 0$. The well-known Projection Theorem guarantees that there exists a *unique* orthogonal projection of ϑ_0 onto the subspace spanned by $1, \vartheta_1, \vartheta_2, \ldots, \vartheta_m$. We can use the Linear Regression Theorem in order to understand the projection in more detail. In the case of $\mathbb{V}\mathrm{ar}(\vartheta) > 0$ with $\vartheta := (\vartheta_1, \vartheta_2, \ldots, \vartheta_m)$, it yields the parameters

$$\lambda := (\lambda_1, \lambda_2, \ldots, \lambda_m) = \mathbb{V}\mathrm{ar}(\vartheta)^{-1}\mathbb{C}\mathrm{ov}(\vartheta, \vartheta_0) \tag{10}$$

and

$$\lambda_0 := \mathbb{E}(\vartheta_0) - \sum_{j=1}^{m}\lambda_j\mathbb{E}(\vartheta_j) \tag{11}$$

[11] Hence, the random vector θ is just the identity $\theta\colon \omega \mapsto \omega$.

[12] A similar approach can be found in Ingersoll (1984, 1987, Chapter 7). However, the ergodic-market approach chosen in this work can be considered more general. This will be explained in the subsequent analysis.

of the linear regression equation

$$\vartheta_0 = \lambda_0 + \sum_{j=1}^{m} \lambda_j \vartheta_j + v\,. \tag{12}$$

I assume that the investors know the expected returns and betas in the market, which means that they are able to calculate the parameters $\lambda_0, \lambda_1, \dots, \lambda_m$.

3. Ergodic Markets

3.1. The General Case

3.1.1. Approximate APT Equation

In an ergodic market, we can always approximate the cost of capital of Asset i, μ_i, by the affine-linear function $\lambda_0 + \sum_{j=1}^{m} \lambda_j \beta_{ij}$. Here, the parameters $\lambda_0, \lambda_1, \dots, \lambda_m$ are given by Eq. 10 and Eq. 11, i.e., they minimize the variance of the residual v in Eq. 12. Then we obtain the expectation equation

$$\mu_i = \lambda_0 + \sum_{j=1}^{m} \lambda_j \beta_{ij} + u_i, \tag{13}$$

where u_i represents an approximation error. This holds true for any number and type of common risks, $X_1, X_2, \dots, X_m$, irrespective of whether the market is in equilibrium or not.

The parameters $\lambda_0, \lambda_1, \dots, \lambda_m$ of Eq. 13 are determined by the empirical distribution of *all* expected returns and betas in the market. To the best of my knowledge, most approaches that can be found in the literature are based on an orthogonal or oblique projection of μ onto the column space of $\begin{bmatrix}\mathbf{1} & \mathrm{B}\end{bmatrix}$ (see, e.g., Gagliardini et al., 2016, Ingersoll, 1984, 1987, Chapter 7), and so the parameters $\lambda_0, \lambda_1, \dots, \lambda_m$ depend on the number of assets, i.e., n. By contrast, here the parameters of the expectation equation are independent of n. In fact, they depend only on the second moments of the random vector $\theta = (\vartheta_0, \vartheta_1, \dots, \vartheta_m)$, whose components take place in the Hilbert space $\mathcal{H}$. They are *uniquely* determined by F if the covariance matrix of ϑ is positive definite. This is equivalent to the statement that $\frac{1}{n}\begin{bmatrix}\mathbf{1} & \mathrm{B}\end{bmatrix}'\begin{bmatrix}\mathbf{1} & \mathrm{B}\end{bmatrix}$ converges to a matrix with full rank.[13] Otherwise, there exist infinitely many λ's such that $\mathbb{V}\mathrm{ar}(\vartheta)\lambda = \mathbb{C}\mathrm{ov}(\vartheta, \vartheta_0)$. Then we can set $m - \operatorname{rank} \mathbb{V}\mathrm{ar}(\vartheta)$ components of λ to zero and so we may ignore the associated components of ϑ. Hence, I assume that $\mathbb{V}\mathrm{ar}(\vartheta) > 0$ without loss of generality.

Although it is always possible to approximate μ_i by $\lambda_0 + \sum_{j=1}^{m} \lambda_j \beta_{ij}$, the quality of approximation essentially depends on $\sigma_v^2 := \mathbb{V}\mathrm{ar}(v)$. This is expressed by the next two theorems.

Theorem 2 (Approximate APT equation)**.** *The expected return on Asset i amounts to*

$$\mu_i = \lambda_0 + \sum_{j=1}^{m} \lambda_j \beta_{ij} + u_i, \tag{14}$$

[13] This assertion is proved in Section A.1.2.

where the parameters $\lambda_0, \lambda_1, \ldots, \lambda_m$ *are given by* $\lambda = \mathbb{V}\mathrm{ar}(\vartheta)^{-1}\mathbb{C}\mathrm{ov}(\vartheta, \vartheta_0)$ *and* $\lambda_0 = \mathbb{E}(\vartheta_0) - \sum_{j=1}^{m} \lambda_j \mathbb{E}(\vartheta_j)$*. Further, we have that*

$$\frac{1}{n}\sum_{i=1}^{n}\left(\mu_i - \lambda_0 - \sum_{j=1}^{m}\lambda_j\beta_{ij}\right)^2 \longrightarrow \sigma_v^2 < \infty \tag{15}$$

with $\sigma_v^2 = \mathbb{V}\mathrm{ar}(\vartheta_0) - \mathbb{C}\mathrm{ov}(\vartheta, \vartheta_0)'\mathbb{V}\mathrm{ar}(\vartheta)^{-1}\mathbb{C}\mathrm{ov}(\vartheta, \vartheta_0) \geq 0$*.*

The residual u_i represents the error of the affine-linear approximation $\mu_i \approx \lambda_0 + \sum_{j=1}^{m} \lambda_j \beta_{ij}$ and thus $\frac{1}{n}\sum_{i=1}^{n} u_i^2$ is the mean square approximation error. In general, we cannot expect that $\frac{1}{n}\sum_{i=1}^{n} u_i^2 \to 0$. However, Theorem 2 guarantees that the limit of the mean square approximation error is finite and it enables us to assess the number of assets in the market for which the approximation error is sufficiently low. Before I come back to this point, note that

$$\mathbb{V}\mathrm{ar}(\vartheta_0) = \lim_{n\to\infty}\frac{\mu'\mu}{n} - \left(\lim_{n\to\infty}\frac{\mathbf{1}'\mu}{n}\right)^2, \tag{16}$$

$$\mathbb{C}\mathrm{ov}(\vartheta, \vartheta_0) = \lim_{n\to\infty}\frac{\mathrm{B}'\mu}{n} - \left(\lim_{n\to\infty}\frac{\mathrm{B}'\mathbf{1}}{n}\right)\left(\lim_{n\to\infty}\frac{\mathbf{1}'\mu}{n}\right), \tag{17}$$

and

$$\mathbb{V}\mathrm{ar}(\vartheta) = \lim_{n\to\infty}\frac{\mathrm{B}'\mathrm{B}}{n} - \left(\lim_{n\to\infty}\frac{\mathrm{B}'\mathbf{1}}{n}\right)\left(\lim_{n\to\infty}\frac{\mathbf{1}'\mathrm{B}}{n}\right). \tag{18}$$

Thus, if the number of assets that we take into consideration is large, it is possible to approximate the second moments of θ by

$$\mathbb{V}\mathrm{ar}(\vartheta_0) \approx \frac{1}{n}\sum_{i=1}^{n}(\mu_i - \bar{\mu})^2, \tag{19}$$

$$\mathbb{C}\mathrm{ov}(\vartheta, \vartheta_0) \approx \frac{1}{n}\sum_{i=1}^{n}(\beta_i - \bar{\beta})(\mu_i - \bar{\mu}), \tag{20}$$

and

$$\mathbb{V}\mathrm{ar}(\vartheta) \approx \frac{1}{n}\sum_{i=1}^{n}(\beta_i - \bar{\beta})(\beta_i - \bar{\beta})', \tag{21}$$

where $\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mu_i$, $\beta_i = (\beta_{i1}, \beta_{i2}, \ldots, \beta_{im})$, and $\bar{\beta} = \frac{1}{n}\sum_{i=1}^{n}\beta_i$. Similarly, we are able to approximate σ_v^2 by the mean square approximation error $\frac{1}{n}\sum_{i=1}^{n} u_i^2$. Geweke and Zhou (1996) consider $\frac{1}{n}\sum_{i=1}^{n} u_i^2$ a measure of pricing errors, which suggests that some assets are mispriced. However, up to now we cannot say why any μ_i should be "incorrect." For this reason, I prefer the term approximation error.

The abstract residual $v = \vartheta_0 - \lambda_0 - \sum_{j=1}^{m}\lambda_j\vartheta_j$ of Eq. 12, which takes place in the Hilbert space $\mathcal{H}$, is a measurable function of θ and the same holds true for $\mathbf{1}_{|v|\geq\tau}$, where τ is any positive number. This leads to the following corollary, which is an immediate consequence of Chebyshev's inequality and thus I skip its proof.

Corollary 1 (Approximation error)**.** *For every $\tau > 0$ we have that*

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{|u_i|\geq\tau} \longrightarrow \mathbb{P}\big(|v| \geq \tau\big) \leq \frac{\sigma_v^2}{\tau^2}. \tag{22}$$

Hence, the proportion of assets whose approximation error exceeds the interval $(-\tau,\tau)$ cannot be greater than $\min\{(\sigma_v/\tau)^2, 1\}$. For example, suppose that the market contains 10000 assets and let the standard deviation of v be 1%. Corollary 1 tells us that less than about $10000 \cdot (0.01/0.02)^2 = 2500$ assets in the market have an approximation error that exceeds the critical amount of $\pm 2\%$. This estimation is based on Chebyshev's inequality and so it is very conservative. If we know the probability distribution of v, we can essentially refine our statement about the approximation quality. If we knew, e.g., that v is normally distributed, we could say that only about $10000 \cdot 2\,\Phi(-0.02/0.01) = 455$ assets in the market are affected by an approximation error whose absolute value is greater or equal than 2%.

3.1.2. Proper Specification

In APT we are particularly interested in the case of $\sigma_v^2 = 0$, i.e., $v = 0$, which means that $\vartheta_0 = \lambda_0 + \sum_{j=1}^{m}\lambda_j\vartheta_j$. Put another way, in this case the APT equation $\mu_i = \lambda_0 + \sum_{j=1}^{m}\lambda_j\beta_{ij}$ is *essentially* satisfied. An economic argument for v to be zero is provided in Section 3.2.1. The main focus of this section is on the question of whether or not Eq. 6 is properly specified in the following sense.[14]

Definition 1 (Proper specification)**.** *The return equation $R_i = \mu_i + \sum_{j=1}^{m}\beta_{ij}X_j + \varepsilon_i$ is said to be properly specified if and only if $\sigma_v^2 = 0$.*

Suppose that there exists an arbitrarily large number of common risks $X_1, X_2, \ldots, X_M$ such that the return equation

$$R_i = \mu_i + \sum_{j=1}^{M}\beta_{ij}X_j + \epsilon_i \tag{23}$$

is properly specified. More precisely, the residual ν of the linear regression equation

$$\vartheta_0 = \kappa_0 + \sum_{j=1}^{M}\kappa_j\vartheta_j + \nu \tag{24}$$

equals zero. The problem is that the number of common risks, M, can be very large. Is it possible to omit some common risks in Eq. 23? More precisely, is the return equation given by Eq. 6 properly specified for any (small) number $m < M$ of common risks? The next theorem states that this is true if and *only* if the market prices of the omitted risks, i.e., $\kappa_{m+1}, \kappa_{m+2}, \ldots, \kappa_M$, equal zero.

Theorem 3 (Proper specification)**.** *Suppose that the return equation $R_i = \mu_i + \sum_{j=1}^{M}\beta_{ij}X_j + \epsilon_i$ is properly specified. The return equation $R_i = \mu_i + \sum_{j=1}^{m}\beta_{ij}X_j + \varepsilon_i$ with $m < M$ is properly specified if*

[14] I still do not require an equilibrium market.

and only if $\kappa_j = 0$ for all $j > m$, in which case it holds that $\lambda_j = \kappa_j$ in $\vartheta_0 = \lambda_0 + \sum_{j=1}^{m} \lambda_j \vartheta_j + v$ for all $0 \leq j \leq m$.

Hence, the investors may ignore the omitted risks $X_{m+1}, X_{m+2}, \ldots, X_M$ if (and only if) $\kappa_j = 0$ for all $j > m$. Of course, this is possible only if they *know* all expected returns and betas in the market, which is a standard assumption in APT. In real life, however, the market participants have to apply an econometric test for the null hypothesis $H_0 : \kappa_{m+1}, \kappa_{m+2}, \ldots, \kappa_M = 0$ that is based on some estimation procedure for the expected returns and betas. This can be done by using a (nonlinear) SUR estimator. As already mentioned, SUR estimation avoids the well-known errors-in-variables bias, which is often prevalent when applying APT in practice.

If some investor rejects H_0, he should expect that the return equation (6) is *not* properly specified, in which case he must not ignore the omitted risks. Otherwise, the mean square approximation error does not tend to zero and the return equation might suffer from an omitted-variables bias. This means that the chosen estimators are consistent for $\lambda_0, \lambda_1, \ldots, \lambda_m$, but not for the regression parameters $\kappa_0, \kappa_1, \ldots, \kappa_m$, which belong to Eq. 24. By contrast, he could assume that Eq. 6 is properly specified if the null hypothesis cannot be rejected. Put another way, in this case, the (small) set of m common risks seems to be sufficient.

3.2. Market Equilibrium

In this section, I investigate the question of whether it is possible to strengthen Theorem 2 by assuming that the market is in equilibrium. We obtain the typical result of APT, namely the inexact APT equation. Moreover, we are able to derive the exact APT equation by assuming that the market is exhaustive. The precise meaning of "exhaustive" will be clarified below.

3.2.1. Inexact APT Equation

The quality of approximation is not directly related to the joint distribution of the idiosyncratic risks. It rather depends on the orthogonal projection of ϑ_0 onto the subspace of $\mathcal{H}$ that is spanned by $(1, \vartheta)$. Nonetheless, we can expect that there exists an indirect relationship between the approximation errors and the idiosyncratic risks. I will come back to this point later on.

Let $\{w_n/n\}$ with $w_n = (w_{n1}, w_{n2}, \ldots, w_{nn}) \in \mathbb{R}^n$ be any investment strategy, i.e., any sequence of portfolios. More precisely, the vector w_n contains the components of the last row of a triangular array

$$\begin{bmatrix} w_{11} & & & \\ w_{21} & w_{22} & & \\ \vdots & \vdots & \ddots & \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} \tag{25}$$

and thus it may depend on n, i.e., the number of assets that we take into consideration. Let ξ be any random variable. I assume that $(X, \xi_n) \rightsquigarrow (X, \xi)$ with $\xi_n := \frac{1}{n} \sum_{i=1}^{n} w_{ni} \varepsilon_i$.

Now, consider the portfolio $u/n \in \mathbb{R}^n$ with $u = (u_1, u_2, \ldots, u_n)$, which will play a major role in the subsequent analysis. Since the market is ergodic, we have that $\frac{1}{n} \sum_{i=1}^{n} u_i \to 0$. This means that the strategy $\{u/n\}$ is *asymptotically* self-financing.[15] It cannot have any rate of return because (in the limit) the invested capital is zero. However, its profit amounts to

$$P_n := \frac{1}{n}\sum_{i=1}^{n} u_i R_i = \frac{1}{n}\sum_{i=1}^{n} u_i \left(\mu_i + \sum_{j=1}^{m} \beta_{ij} X_j + \varepsilon_i\right) \tag{26}$$

$$= \frac{1}{n}\sum_{i=1}^{n} u_i \mu_i + \sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n} \beta_{ij} u_i\right) X_j + \frac{1}{n}\sum_{i=1}^{n} u_i \varepsilon_i\,. \tag{27}$$

Since the market is ergodic, it holds that

$$\frac{1}{n}\sum_{i=1}^{n} u_i \mu_i = \frac{1}{n}\sum_{i=1}^{n} u_i \left(\lambda_0 + \sum_{j=1}^{m} \lambda_j \beta_{ij} + u_i\right) \tag{28}$$

$$= \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} u_i\right)}_{\to 0}\lambda_0 + \sum_{j=1}^{m}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} \beta_{ij} u_i\right)}_{\to 0}\lambda_j + \underbrace{\frac{1}{n}\sum_{i=1}^{n} u_i^2}_{\to \sigma_v^2} \longrightarrow \sigma_v^2 \tag{29}$$

and

$$\sum_{j=1}^{m}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} \beta_{ij} u_i\right)}_{\to 0} X_j \longrightarrow 0\,. \tag{30}$$

Hence, it follows that $P_n \rightsquigarrow P = \sigma_v^2 + \xi$ with $\sigma_v^2 \geq 0$ and $\mathrm{Var}(\xi) \geq 0$. If the market is efficient in the sense of Fama (1970), $\mathrm{E}(P) = \sigma_v^2$ typically increases with $\mathrm{Var}(P) = \mathrm{Var}(\xi)$.[16] Hence, the approximation errors and the idiosyncratic risks are interconnected.

The following definition is inspired by Ross (1976a).

Definition 2 (Well-diversified strategy)**.** *The strategy $\{w_n/n\}$ is said to be well-diversified if and only if*

$$\frac{1}{n}\sum_{i=1}^{n} w_{ni}\varepsilon_i \xrightarrow{\text{P}} 0\,. \tag{31}$$

Some authors (see, e.g., Chamberlain and Rothschild, 1983) require that some traders are risk averse and assume that

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} w_{ni}\varepsilon_i\right) \longrightarrow 0. \tag{32}$$

Put another way, the sequence $\left\{\frac{1}{n}\sum_{i=1}^{n} w_{ni}\varepsilon_i\right\}$ must converge in mean square to zero, which implies that $\{w_n/n\}$ is well-diversified. However, focusing on the variance is not necessary and, actually, APT does not require any risk-averse investor.

I make only the following assumption:

A. There exists a rational investor possessing a continuous and strictly increasing utility

[15] Any portfolio $w = (w_1, w_2, \ldots, w_n) \in \mathbb{R}^n$ is said to be self-financing if and only if $\sum_{i=1}^{n} w_i = 0$.
[16] Here, it is implicitly assumed that $\mathrm{E}(\xi) = 0$, which does not follow from $\mathrm{E}(\xi_n) = 0$.

function $u\colon \mathbb{R} \to \mathbb{R}$. Let c be the initial capital of the investor. If $P_n \xrightarrow{\mathrm{P}} p \in \mathbb{R}$ then

$$\mathrm{E}\big(u(c + aP_n)\big) \longrightarrow u\big(c + ap\big), \quad \forall\, a > 0. \tag{33}$$

Assumption **A** is equivalent to the statement that the sequence $\{u(c + aP_n)\}$ is asymptotically uniformly integrable, provided that P_n converges in probability to p (van der Vaart, 1998, Chapter 2.5). Note that if $\{u/n\}$ is well-diversified, we have that $P_n \xrightarrow{\mathrm{P}} \sigma_v^2$.

The next theorem can be considered the main result of APT.

Theorem 4 (Inexact APT equation)**.** *Let Assumption* **A** *be satisfied and suppose that the market is in equilibrium. If* $\frac{1}{n}\sum_{i=1}^{n} u_i\varepsilon_i \xrightarrow{\mathrm{P}} 0$*, then*

$$\frac{1}{n}\sum_{i=1}^{n}\left(\mu_i - \lambda_0 - \sum_{j=1}^{m}\lambda_j\beta_{ij}\right)^2 \longrightarrow 0\,. \tag{34}$$

Hence, if the market is in equilibrium, the variance of v must be zero—provided that $\{u/n\}$ is a well-diversified strategy, i.e., $\frac{1}{n}\sum_{i=1}^{n} u_i\varepsilon_i \xrightarrow{\mathrm{P}} 0$. The latter condition is essential and ensures that the inexact APT equation is not trivially satisfied for every arbitrary set of common risks. I will come back to this point below.

Theorem 4 does *not* claim that the APT equation $\mu_i = \lambda_0 + \sum_{j=1}^{m}\lambda_j\beta_{ij}$ is satisfied for each asset in the market. It only asserts that $\frac{1}{n}\sum_{i=1}^{n} u_i^2 \to 0$. This means that the APT equation must hold true only in the abstract sense, i.e., $\vartheta_0 = \lambda_0 + \sum_{j=1}^{m}\lambda_j\vartheta_j$. Put another way, in an equilibrium market, we must have that

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{u_i\neq 0} \longrightarrow \mathbb{P}\big(v \neq 0\big) = 0\,. \tag{35}$$

Hence, Eq. 1 indeed can be violated, but only for an *evanescent* proportion of assets. This means that the relative number of approximation errors must vanish if the number of assets that we take into consideration, n, grows to infinity. To the best of my knowledge, Al-Najjar (1998) is the first author who comes to the same conclusion. Similar arguments can be found also in Gagliardini et al. (2016) as well as Renault et al. (2017). However, the absolute number and magnitude of the approximation errors can be arbitrarily large and so, in general, u_i cannot be considered negligible.

There is no universal answer to the question of how to pick the common risks. Of course, this depends on the particular case, but at least we can check whether Eq. 6 is properly specified by applying a nonlinear SUR estimator. This leads to the estimated residuals $\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n$. According to Theorem 4, we may expect that $\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2$ is close to zero if the number of assets, n, is sufficiently large and the return equation is properly specified. However, the meaning of being "close to zero" depends on the desired significance level, which should be determined from the outset in order to avoid a selection (or publication) bias.

Now, Eq. 13 can be interpreted in the usual way: The cost of capital of Asset i, μ_i, is *essentially* an affine-linear function of its betas. The term $\lambda_j\beta_{ij}$ represents the risk premium of Asset i that can be attributed to the common risk X_j. The regression coefficient λ_j is referred to as the market price of risk that is related to X_j, whereas β_{ij} quantifies the exposure of Asset i to that common risk. Moreover, the intercept λ_0 is the time value of money, whereas the approximation error u_i is the part of μ_i that cannot be attributed to one of the aforementioned price determinants. Nonetheless, the market is in equilibrium and thus u_i cannot be considered a pricing error.

Each approximation error u_i is deterministic and, by construction, it holds that $\mathrm{E}(\varepsilon_i) = 0$. Thus, we have that

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} u_i\varepsilon_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(u_i\varepsilon_i) = \frac{1}{n}\sum_{i=1}^{n} u_i\underbrace{\mathrm{E}(\varepsilon_i)}_{=0} = 0.$$

This is a major result of our basic ergodicity assumption made in Section 2. On the one hand, it allows us to consider all expected returns and betas deterministic. On the other hand, it enables us to specify the market prices of risk, $\lambda_0, \lambda_1, \ldots, \lambda_m$, in the limiting case $n \to \infty$. That is, we do not have to fix any finite number, n, of assets. By contrast, Gagliardini et al. (2016) consider μ_i and $\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}$ realizations of a random vector and focus on a finite subset of the asset universe. The reader can verify that the ergodicity assumption maintained in this work essentially simplifies their proof of the APT equation.

Theorem 4 goes beyond the property $\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} u_i\varepsilon_i\right) = 0$. It requires the Weak Law of Large Numbers $\frac{1}{n}\sum_{i=1}^{n} u_i\varepsilon_i \xrightarrow{\mathrm{P}} 0$. Put another way, on average, the quantity $u_i\varepsilon_i$ must vanish if the number of assets that we take into consideration grows to infinity. Whether or not this condition is fulfilled depends essentially on the approximation errors and thus on the given set of common risks: Theorem 4 implies that the return equation is properly specified, but in Section 3.1.2 we have seen that this holds true only if the market prices of the omitted risks equal zero. This means that the Weak Law of Large Numbers might be violated if we choose the set of common risks in an arbitrary way.

The Weak Law of Large Numbers is satisfied if the idiosyncratic risks follow an approximate factor model (Chamberlain and Rothschild, 1983). Suppose that $u \neq \mathbf{0}$ for some sufficiently large number n.[17] Now, we have that

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} u_i\varepsilon_i\right) = \frac{u'\mathrm{Var}(\varepsilon)u}{n^2} = \left(\frac{u'u}{n}\right)\left(\frac{\tilde{u}'\mathrm{Var}(\varepsilon)\tilde{u}}{n}\right), \quad \tilde{u} := \frac{u}{\|u\|}, \tag{36}$$

and according to Theorem 2 it holds that $u'u/n \to \sigma_v^2 < \infty$. It is well-known that $x'\mathrm{Var}(\varepsilon)x$ cannot exceed the largest eigenvalue of $\mathrm{Var}(\varepsilon)$ for any vector $x \in \mathbb{R}^n$ with $\|x\| = 1$. Chamberlain and Rothschild (1983) assume that the sequence of the largest eigenvalues associated with $\{\mathrm{Var}(\varepsilon)\}$ is bounded above. Thus, $\tilde{u}'\mathrm{Var}(\varepsilon)\tilde{u}/n$ vanishes as n grows to infinity, and because $\frac{1}{n}\sum_{i=1}^{n} u_i^2$ converges to a finite number σ_v^2, it holds that $\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} u_i\varepsilon_i\right) \to 0$.[18] We conclude that the sequence $\left\{\frac{1}{n}\sum_{i=1}^{n} u_i\varepsilon_i\right\}$ converges in mean square to zero, which implies that $\frac{1}{n}\sum_{i=1}^{n} u_i\varepsilon_i \xrightarrow{\mathrm{P}} 0$, i.e., the Weak Law of Large Numbers is satisfied.

Theorem 4 does *not* presume that $\frac{1}{n}\sum_{i=1}^{n} w_{ni}\varepsilon_i \xrightarrow{\mathrm{P}} 0$ for any strategy $\{w_n/n\}$ other than $\{u/n\}$.

[17] Otherwise, the Weak Law of Large Numbers is trivially satisfied.

[18] Gagliardini et al. (2016) point out that one could make the weaker assumption $\tilde{u}'\mathrm{Var}(\varepsilon)\tilde{u} = o(n)$.

By contrast, Ross' (1976a) line of argument goes like this:

(i) Form a self-financing portfolio $w_n = (w_{n1}, w_{n2}, \ldots, w_{nn}) \in \mathbb{R}^n$, i.e., $\sum_{i=1}^n w_{ni} = 0$, such that $\sum_{i=1}^n w_{ni}\varepsilon_i \to 0$.

(ii) The profit of the portfolio w_n can be decomposed into three parts:

$$P_n = \sum_{i=1}^n w_{ni}\mu_i + \sum_{j=1}^m \left(\sum_{i=1}^n w_{ni}\beta_{ij}\right) X_j + \sum_{i=1}^n w_{ni}\varepsilon_i\,. \tag{37}$$

The first part is risk-free, the second part is systematic, and the third part is unsystematic. The latter vanishes as n grows to infinity. Hence, if n is a large number of assets, we have that

$$P_n \approx \sum_{i=1}^n w_{ni}\mu_i + \sum_{j=1}^m \left(\sum_{i=1}^n w_{ni}\beta_{ij}\right) X_j\,. \tag{38}$$

(iii) Moreover, choose the portfolio w_n such that $\sum_{i=1}^n w_{ni}\beta_{ij} = 0$. Hence, the systematic part of P_n can be ignored and thus $P_n \approx \sum_{i=1}^n w_{ni}\mu_i$.

(iv) The term $\sum_{i=1}^n w_{ni}\mu_i$ is risk-free and in case $\sum_{i=1}^n w_{ni}\mu_i \neq 0$ the market cannot be in equilibrium, since any rational investor would try to make as much money as possible out of nothing. Thus, for *all* portfolios w_n with $\sum_{i=1}^n w_{ni} = 0$ and $\sum_{i=1}^n w_{ni}\beta_{ij} = 0$, we must have that $\sum_{i=1}^n w_{ni}\mu_i = 0$. This means that μ must belong to the column space of $[\mathbf{1}\ \mathrm{B}]$, i.e., there must exist some parameters $\lambda_0, \lambda_1, \ldots, \lambda_m$ such that $\mu_i = \lambda_0 + \sum_{j=1}^m \lambda_j\beta_{ij}$.

Argument iv requires that the Strong Law of Large Numbers $\sum_{i=1}^n w_{ni}\varepsilon_i \to 0$ is satisfied for *all* strategies $\{w_n\}$ that are such that w_n is orthogonal to the column space of $[\mathbf{1}\ \mathrm{B}]$. Theorem 4 shows that this is not necessary at all. Moreover, Argument iv is based on the idea that all vectors take place in the n-dimensional Euclidean space, which represents a finite-dimensional Hilbert space. However, in order to diversify away the idiosyncratic risks by the strategy $\{w_n\}$, the number of assets, n, must grow to infinity and thus Ross' whole line of argument is somewhat misleading. He observes that this is just a heuristic and shows that the APT equation is, at least, *essentially* true (Ross, 1976a, p. 347), i.e.,

$$\sum_{i=1}^\infty \left(\mu_i - \lambda_0 - \sum_{j=1}^m \lambda_j\beta_{ij}\right)^2 < \infty\,. \tag{39}$$

This result is slightly stronger than $\frac{1}{n}\sum_{i=1}^n \left(\mu_i - \lambda_0 - \sum_{j=1}^m \lambda_j\beta_{ij}\right)^2 \to 0$, i.e., the quintessence of Theorem 4, and requires some additional assumptions that go beyond the scope of this work.

3.2.2. Exact APT Equation

Theorem 4 does not say anything about the question of whether the APT equation holds true for some specific asset. For example, suppose that a riskless asset exists and let r be the risk-free interest rate. Consider the linear *regression* equation

$$r = \mu_0 + \sum_{j=1}^{m} \beta_{0j} X_j + \varepsilon_0 \,. \tag{40}$$

Here, the index "0" is just symbolic and shall only denote the riskless asset.

By definition, we have that $\mathrm{Var}(r) = 0$ and from $\mathrm{Var}(X) > 0$ it follows that $\beta_{0j} = 0$ as well as $\varepsilon_0 = 0$. Thus, we may conclude that $r = \mu_0 = \lambda_0 + u_0$, but the problem is that the approximation error u_0 can be arbitrary. In order to obtain the typical result "$\lambda_0 = r$" we must either *assume* that $u_0 = 0$ or provide an economic justification. For the latter purpose, we need the following definition.

Definition 3 (Orthogonal strategy)**.** *The strategy $\{w_n/n\}$ is said to be orthogonal if and only if*

$$\frac{1}{n}\sum_{i=1}^{n} w_{ni} u_i \longrightarrow 0 \,. \tag{41}$$

Now, we can simply rephrase Theorem 4 as follows: If the strategy $\{u/n\}$ is well-diversified, it must be orthogonal, provided Assumption **A** is satisfied and the market is in equilibrium.

Suppose that $\{w_n/n\}$ is *fully* diversified (Ingersoll, 1987, p. 177), i.e., the sequence $\left\{\frac{1}{n}\sum_{i=1}^{n} w_{ni}^2\right\}$ is bounded above. In this case, the Cauchy-Schwarz inequality reveals that

$$\left(\frac{1}{n}\sum_{i=1}^{n} w_{ni} u_i\right)^2 \leq \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} w_{ni}^2\right)}_{\leq\, x \,<\, \infty} \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} u_i^2\right)}_{\to\, 0} \longrightarrow 0 \,, \tag{42}$$

where x denotes the supremum of $\left\{\frac{1}{n}\sum_{i=1}^{n} w_{ni}^2\right\}$. This implies that $\{w_n/n\}$ is orthogonal. However, the assumption that $\{w_n/n\}$ is fully diversified is not based on the approximation errors and, as we will see later on, it is *essentially* stronger than the orthogonality condition expressed by Definition 3. Some typical examples of orthogonal strategies are $\{\mathbf{1}/n\}$ and $\{u/n\}$. It is clear that the strategy $\{a w_n/n\}$ with $a \in \mathbb{R}$ is orthogonal if $\{w_n/n\}$ is orthogonal, and each linear combination of orthogonal strategies is orthogonal, too. Hence, the set of all orthogonal strategies is a linear space. Further, any strategy $\{w_n/n\}$ is orthogonal whenever the sequence $\left\{\sum_{i=1}^{n} w_{ni} u_i\right\}$ is bounded. In particular, if the number of nonzero approximation errors is finite then *each* strategy $\{w_n/n\}$ is orthogonal.

Now, consider some orthogonal strategy $\{w_n/n\}$ and let $P_n = \frac{1}{n}\sum_{i=1}^{n} w_{ni} R_i$ be the corresponding profit. If $\frac{1}{n}\sum_{i=1}^{n} w_{ni} \to \gamma \neq 0$, then the normalized strategy $\{\gamma^{-1} w_n/n\}$ is orthogonal, too, and satisfies the budget constraint $\frac{1}{n}\sum_{i=1}^{n} w_{ni}/\gamma \to 1$. The following lemma describes the asymptotic distribution of the profit of a normalized orthogonal strategy.[19]

Lemma 1 (Normalized orthogonal strategy)**.** *Let $\{w_n/n\}$ be an orthogonal strategy with $\frac{1}{n}\sum_{i=1}^{n} w_{ni} \to$*

[19] In the case of $\gamma = 0$ we need no normalization at all.

$\gamma \in \{0,1\}$ and $\frac{1}{n}\sum_{i=1}^{n} w_{ni}\beta_{ij} \to \beta_{wj}$. We have that

$$P_n \rightsquigarrow P = \begin{cases} \sum_{j=1}^{m} \beta_{wj}(\lambda_j + X_j) + \xi, & \gamma = 0\,, \\ \lambda_0 + \sum_{j=1}^{m} \beta_{wj}(\lambda_j + X_j) + \xi, & \gamma = 1\,, \end{cases} \tag{43}$$

and it holds that

$$\mathrm{E}(P) = \begin{cases} \sum_{j=1}^{m} \lambda_j \beta_{wj}, & \gamma = 0\,, \\ \lambda_0 + \sum_{j=1}^{m} \lambda_j \beta_{wj}, & \gamma = 1\,, \end{cases} \tag{44}$$

if and only if the sequence $\{\xi_n\}$ is asymptotically uniformly integrable. Moreover, in this case, we have that $\mathrm{Var}(\xi_n) \to \mathrm{Var}(\xi) \geq 0$ if and only if also the sequence $\{\xi_n^2\}$ is asymptotically uniformly integrable.

Lemma 1 does *not* require a market equilibrium and it does not presume that the strategy $\{u/n\}$ is well-diversified. Put another way, the prerequisites of Theorem 4 need not be fulfilled. It tells us that we can create strategies for which the APT equation holds true *exactly*. In the case of $\gamma = 1$, $\{w_n/n\}$ represents a secondary asset possessing the betas $\beta_{w1}, \beta_{w2}, \ldots, \beta_{wm}$ and the idiosyncratic risk ξ.

Suppose that Eq. 6 represents an approximate factor model. Let $\{w_n/n\}$ be a fully-diversified strategy and assume, without loss of generality, that $w_n \neq \mathbf{0}$ for some sufficiently large number n. Then it holds that

$$\begin{aligned} \mathrm{Var}(\xi_n) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} w_{ni}\varepsilon_i\right) && (45) \\ &= \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} w_{ni}^2\right)}_{\leq x < \infty} \underbrace{\left(\frac{\tilde{w}_n' \mathrm{Var}(\varepsilon)\tilde{w}_n}{n}\right)}_{\to 0} \longrightarrow 0\,, \quad \tilde{w}_n := \frac{w_n}{\|w_n\|}, && (46) \end{aligned}$$

which implies that $\xi_n \xrightarrow{\mathrm{P}} 0$ and thus $\xi = 0$. Hence, the profit of a fully-diversified strategy satisfying the budget constraint, i.e., its return on investment, may contain a systematic but not an unsystematic part. We conclude that fully-diversified strategies essentially restrict the possibilities of the market participants to create secondary assets.

For the rest of this work, our main goal is to create secondary assets *possessing* idiosyncratic risk. Hence, it is not necessary to require $w_n' w_n = O(n)$, and (as we have seen above) this strong assumption can even be harmful. For our purposes, it only matters that the investors are able to *diversify away the approximation errors* by the strategy $\{w_n/n\}$, i.e., $w_n' u = o(n)$. This leads us to the following definition.

Definition 4 (Replication). *A strategy $\{w_n/n\}$ is said to replicate Asset k if and only if*

(i) $\frac{1}{n}\sum_{i=1}^{n} w_{ni} \to 1$,

(ii) $\frac{1}{n}\sum_{i=1}^{n} w_{ni}\beta_{ij} \to \beta_{kj}$, *and*

(iii) $\frac{1}{n}\sum_{i=1}^{n} w_{ni}\varepsilon_i \xrightarrow{\mathrm{P}} \varepsilon_k\,.$

It is worth emphasizing that this definition does not require the strategy to replicate the expected return on Asset k but only its betas and idiosyncratic risk.

The following proposition establishes the exact APT equation for a single asset.

Proposition 1 (Exact APT equation)**.** *Let Assumption* **A** *be satisfied and suppose that the market is in equilibrium. If Asset k can be replicated by an orthogonal strategy, we have that*

$$\mu_k = \lambda_0 + \sum_{j=1}^{m} \lambda_j \beta_{kj} \,. \tag{47}$$

This proposition does not require the asymptotic uniform integrability of $\{\xi_n\}$, which appears in Lemma 1. Moreover, it does not presume that $\frac{1}{n}\sum_{i=1}^{n} u_i \varepsilon_i \xrightarrow{\text{P}} 0$, i.e., $\{u/n\}$ need not be well-diversified, which is an essential requirement of Theorem 4. Nonetheless, it is plausible that the investors have more freedom to replicate any asset if Eq. 6 is properly specified, i.e., $\frac{1}{n}\sum_{i=1}^{n} u_i^2 \to 0$. In this case, the orthogonality condition imposed by Definition 4 appears to be least restrictive.

The following corollary is an immediate consequence of Proposition 1 and so its proof is skipped. It asserts that the risk-free interest rate, r, equals λ_0 if we can replicate the riskless asset by an orthogonal strategy $\{w_n/n\}$. We have that $\varepsilon_0 = 0$ and thus it holds that $\frac{1}{n}\sum_{i=1}^{n} w_{ni}\varepsilon_i \xrightarrow{\text{P}} 0$, i.e., the strategy $\{w_n/n\}$ is not only orthogonal but also well-diversified.

Corollary 2 (Riskless asset)**.** *Let Assumption* **A** *be satisfied and suppose that the market is in equilibrium. Assume that there exists a riskless asset and let r be the risk-free interest rate. If the riskless asset can be replicated by an orthogonal strategy, we have that $r = \lambda_0$.*

The next definition leads us to the exact APT equation.

Definition 5 (Exhaustive market)**.** *The market is said to be exhaustive if and only if each asset can be replicated by an orthogonal strategy.*

The attribute "exhaustive" shall indicate that one can replicate the betas and idiosyncratic risk of each single asset by combining all assets in the market such that their approximation errors are diversified away. However, an exhaustive market is not necessarily complete: In a complete market the investors must be able to replicate all contingent claims.[20]

The next theorem can be considered the main result of this work. It asserts that the exact APT equation holds true for *all* assets if and only if the market is exhaustive.

Theorem 5 (Fundamental Theorem)**.** *Let Assumption* **A** *be satisfied and suppose that the market is in equilibrium. It is exhaustive if and only if*

$$\mu_i = \lambda_0 + \sum_{j=1}^{m} \lambda_j \beta_{ij} \tag{48}$$

holds true for each asset in the market.

[20]More details on that topic can be found, e.g., in Biagini (2010), Frahm (2016) as well as Harrison and Pliska (1981, 1983).

Hence, in an exhaustive equilibrium market all approximation errors vanish. This implies that the mean square approximation error is zero, too, and thus $\sigma_v^2 = 0$. Put another way, the return equation $R_i = \mu_i + \sum_{j=1}^m \beta_{ij} X_j + \varepsilon_i$ must be properly specified, and it is clear that also the Weak Law of Large Numbers is fulfilled in this case. The return on a portfolio of a finite number of assets, satisfying the budget constraint, is a linear combination of single asset returns. Thus, if the market is exhaustive, also the expected return on the portfolio is an affine-linear function of its betas. The same holds true for any strategy $\{w_n/n\}$ with $\frac{1}{n}\sum_{i=1}^n w_{ni} \to 1$ and $\frac{1}{n}\sum_{i=1}^n w_{ni}\beta_{ij} \to \beta_{wj}$, provided its sequence of idiosyncratic risks, $\{\xi_n\}$, is asymptotically uniformly integrable.

4. Conclusion

Ergodicity of the expected returns and betas in the market is a natural assumption in APT. It enables us to define an inner product and thus to apply the theory of Hilbert spaces in a natural way. The Projection Theorem guarantees that the expected return on any asset can always be approximated by an affine-linear function of its betas. This general result does not require the market to be in equilibrium, and it need not even be arbitrage-free. We are able to estimate the relative number of assets that violate the APT equation by observing the given expected returns and betas. Irrespective of whether or not the market is in equilibrium, the APT equation is essentially satisfied *only* if we do not omit any common risk whose market price differs from zero, provided there exists an arbitrarily large number of common risks for which the return equation is properly specified. We are able to test for a proper specification of our return equation.

Further, in any equilibrium market, the APT equation holds true in its inexact form if the Weak Law of Large Numbers $\frac{1}{n}\sum_{i=1}^n u_i\varepsilon_i \xrightarrow{p} 0$ is satisfied. This simple condition just combines the approximation error, u_i, and the idiosyncratic risk, ε_i, of Asset i. We need not assume that the asset returns obey a strict or an approximate factor model, and the idiosyncratic risks need not even stem from a linear regression equation. Moreover, the APT equation holds true in its exact form if and only if the market is exhaustive. This means that the market participants must be able to replicate the betas and idiosyncratic risk of each asset by some strategy that diversifies away all approximation errors in the market. Once again, this requires the return equation to be properly specified, and also the Weak Law of Large Numbers is satisfied in an exhaustive equilibrium market.

A. Appendix

A.1. Positive Definiteness

A.1.1. Covariance Matrix of Common Risks

Suppose that $\mathrm{Var}(X)$ is *not* positive definite and let $\mathcal{N} \neq \emptyset$ be the set of all vectors $a \in \mathbb{R}^m$ with $a \neq \mathbf{0}$ such that $a'\mathrm{Var}(X)a = a'\mathrm{E}(XX')a = \mathrm{E}\big((a'X)^2\big) = 0$, i.e., $a'X = 0$. $\mathcal{N}$ represents a linear subspace of $\mathbb{R}^m$ with $k < m$ dimensions. Now, choose k linearly independent vectors $a_1, a_2, \ldots, a_k$ from $\mathcal{N}$ and write $X = (X_1, X_2)$, where we can assume that the $(m-k)$-dimensional subvector X_1 is such that $\mathrm{Var}(X_1) > 0$. Consider the $k \times m$ matrix $A = \begin{bmatrix} A_1 & A_2 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_k \end{bmatrix}'$. Hence, we have that $AX = A_1X_1 + A_2X_2 = \mathbf{0}$. Moreover, A_2 has full rank and thus $X_2 = -A_2^{-1}A_1X_1$, i.e.,

$$X = \underbrace{\begin{bmatrix} \mathbf{I} \\ -A_2^{-1}A_1 \end{bmatrix}}_{=:\,T} X_1 \,. \tag{49}$$

Now, in its matrix form, the return equation reads $R = \mu + \Gamma X_1 + \varepsilon$ with $\Gamma := \mathrm{B}T$ and $\mathrm{Var}(X_1) > 0$.

A.1.2. Covariance Matrix of Abstract Betas

The matrix $\frac{1}{n}\begin{bmatrix}\mathbf{1} & \mathrm{B}\end{bmatrix}'\begin{bmatrix}\mathbf{1} & \mathrm{B}\end{bmatrix}$ converges to

$$A = \begin{bmatrix} 1 & \mathbb{E}(\vartheta') \\ \mathbb{E}(\vartheta) & \mathbb{V}\mathrm{ar}(\vartheta) + \mathbb{E}(\vartheta)\mathbb{E}(\vartheta') \end{bmatrix} \in \mathbb{R}^{(m+1)\times(m+1)} \tag{50}$$

and we have that

$$A \begin{bmatrix} -\mathbb{E}(\vartheta') \\ \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{0}' \\ \mathbb{V}\mathrm{ar}(\vartheta) \end{bmatrix} \in \mathbb{R}^{(m+1)\times m}. \tag{51}$$

Consider any vector $x \in \mathbb{R}^m$ with $x \neq \mathbf{0}$. If A has full rank, it holds that

$$\begin{bmatrix} 0 \\ \mathbb{V}\mathrm{ar}(\vartheta)x \end{bmatrix} = \begin{bmatrix} \mathbf{0}' \\ \mathbb{V}\mathrm{ar}(\vartheta) \end{bmatrix} x = A \begin{bmatrix} -\mathbb{E}(\vartheta') \\ \mathbf{I} \end{bmatrix} x = A \begin{bmatrix} -\mathbb{E}(\vartheta')x \\ x \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{52}$$

with $y \neq \mathbf{0}$. This means that $\mathbb{V}\mathrm{ar}(\vartheta)$ has full rank and thus it is positive definite. Now, consider any vector $y \in \mathbb{R}^{m+1}$ with $y \neq \mathbf{0}$ and define $y := (y_1, y_2, \ldots, y_m)$. We have that

$$Ay = A \begin{bmatrix} y_0 + \mathbb{E}(\vartheta')y \\ \mathbb{V}\mathrm{ar}(\vartheta)y + \mathbb{E}(\vartheta)\big(y_0 + \mathbb{E}(\vartheta')y\big) \end{bmatrix}. \tag{53}$$

Suppose that $\mathbb{V}\mathrm{ar}(\vartheta)$ is positive definite, which means that it has full rank. If $Ay = \mathbf{0}$ holds true, we must have that $y_0 + \mathbb{E}(\vartheta')y = 0$ and thus $\mathbb{V}\mathrm{ar}(\vartheta)y = 0$, but this can happen only if $y = \mathbf{0}$ and thus $y_0 = 0$. This contradicts our initial assumption that $y \neq \mathbf{0}$. We conclude that $Ay \neq \mathbf{0}$ and so A has full rank.

A.2. Proofs

A.2.1. Proof of Theorem 1

(i)⇔(ii): We aim at minimizing the mean-square objective function

$$\begin{aligned}(b_0, b_1, \ldots, b_m) \quad &\longmapsto \quad \mathrm{E}\left(\big(Y - \sum_{j=0}^{m} b_j X_j\big)^2\right) &(54)\\ &= \mathrm{Var}\left(Y - \sum_{j=1}^{m} b_j X_j\right) + \mathrm{E}^2\left(Y - b_0 - \sum_{j=1}^{m} b_j X_j\right) &(55)\end{aligned}$$

with

$$\begin{aligned}\mathrm{Var}\left(Y - \sum_{j=1}^{m} b_j X_j\right) &= \mathrm{Var}(Y) - 2\sum_{j=1}^{m} b_j \mathrm{Cov}(X_j, Y) + \sum_{i,j=1}^{m} b_i b_j \mathrm{Cov}(X_i, X_j)\\ &= \mathrm{Var}(Y) - 2\,(b_1, b_2, \ldots, b_m)'\mathrm{Cov}(X, Y)\\ &\qquad + (b_1, b_2, \ldots, b_m)'\mathrm{Var}(X)(b_1, b_2, \ldots, b_m).\end{aligned}$$

The optimal value for b_0 must be such that $\mathrm{E}^2\big(Y - b_0 - \sum_{j=1}^{m} b_j X_j\big) = 0$, i.e., $\beta_0 = \mathrm{E}(Y) - \sum_{j=1}^{m} \beta_j \mathrm{E}(X_j)$. In the case of $m = 0$, we have accomplished our goal by setting $\beta_0 = \mathrm{E}(Y)$. By contrast, in the case of $m > 0$, we have to minimize also the quadratic objective function $(b_1, b_2, \ldots, b_m) \mapsto \mathrm{Var}\big(Y - \sum_{j=1}^{m} b_j X_j\big)$. The covariance matrix of X is positive semidefinite and so there exists a solution to this minimization problem. It must satisfy the first-order condition

$$\frac{\partial \mathrm{Var}\big(Y - \sum_{j=1}^{m} b_j X_j\big)}{\partial(b_1, b_2, \ldots, b_m)} = -2\,\mathrm{Cov}(X, Y) + 2\,\mathrm{Var}(X)(b_1, b_2, \ldots, b_m) = \mathbf{0} \tag{56}$$

and so we obtain $\mathrm{Var}(X)(\beta_1, \beta_2, \ldots, \beta_m) = \mathrm{Cov}(X, Y)$. Moreover, from $\mathrm{Var}(X) \geq 0$ we conclude that the vector $(\beta_1, \beta_2, \ldots, \beta_m)$ satisfies the second-order condition

$$\frac{\partial^2 \mathrm{Var}\big(Y - \sum_{j=1}^{m} b_j X_j\big)}{\partial(b_1, b_2, \ldots, b_m)^2} = 2\mathrm{Var}(X) \geq 0. \tag{57}$$

Further, it is also clear that $\mathrm{E}^2\big(Y - b_0 - \sum_{j=1}^{m} \beta_j X_j\big) > 0$ if b_0 departs from β_0. Hence, $(\beta_0, \beta_1, \ldots, \beta_m)$ with $\mathrm{Var}(X)(\beta_1, \beta_2, \ldots, \beta_m) = \mathrm{Cov}(X, Y)$ and $\beta_0 = \mathrm{E}(Y) - \sum_{j=1}^{m} \beta_j \mathrm{E}(X_j)$, in fact, minimizes the mean-square objective function. (ii)⇔(iii): The given equations imply that the residual $\epsilon = Y - \beta_0 - \sum_{j=1}^{m} \beta_j X_j$ is such that

$$\mathrm{E}(\epsilon) = \mathrm{E}\left(Y - \beta_0 - \sum_{j=1}^{m} \beta_j X_j\right) = \mathrm{E}(Y) - \beta_0 - \sum_{j=1}^{m} \beta_j \mathrm{E}(X_j) = 0 \tag{58}$$

and

$$
\begin{aligned}
\mathrm{Cov}(X,\epsilon) &= \mathrm{Cov}\left(X, Y-\beta_0-\sum_{j=1}^{m}\beta_j X_j\right) && (59)\\
&= \mathrm{Cov}(X,Y)-\mathrm{Var}(X)(\beta_1,\beta_2,\ldots,\beta_m)=\mathbf{0}\,. && (60)
\end{aligned}
$$

Conversely, from $Y=\beta_0+\sum_{j=1}^{m}\beta_j X_j+\epsilon$ with $\mathrm{E}(\epsilon)=0$ and $\mathrm{Cov}(X_j,\epsilon)=0$ we conclude that

$$
\mathrm{E}(\epsilon)=\mathrm{E}(Y)-\beta_0-\sum_{j=1}^{m}\beta_j\mathrm{E}(X_j)=0 \tag{61}
$$

and

$$
\begin{aligned}
\mathrm{Cov}(X,\varepsilon) &= \mathrm{Cov}(X,Y)-\mathrm{Var}(X)(\beta_1,\beta_2,\ldots,\beta_m) && (62)\\
&= \mathbf{0}\,. && (63)
\end{aligned}
$$

This implies that $\beta_0=\mathrm{E}(Y)-\sum_{j=1}^{m}\beta_j\mathrm{E}(X_j)$ and $\mathrm{Var}(X)(\beta_1,\beta_2,\ldots,\beta_m)=\mathrm{Cov}(X,Y)$. If the covariance matrix of X is positive definite, we can invert $\mathrm{Var}(X)$ and obtain the solution $(\beta_0,\beta_1,\ldots,\beta_m)$ with

$$
(\beta_1,\beta_2,\ldots,\beta_m)=\mathrm{Var}(X)^{-1}\mathrm{Cov}(X,Y) \tag{64}
$$

and

$$
\beta_0=\mathrm{E}(Y)-\sum_{j=1}^{m}\beta_j\mathrm{E}(X_j). \tag{65}
$$

Now, the quadratic objective function is strictly convex and so its solution $(\beta_1,\beta_2,\ldots,\beta_m)$ is unique, which implies that also β_0 is uniquely determined by $(\beta_1,\beta_2,\ldots,\beta_m)$. Finally, the second moments of $Z=(1,X,Y)$ are finite, and so the components of Z belong to a Hilbert space with inner product $\mathrm{E}(Z_iZ_j)$. The Projection Theorem guarantees that $\sum_{j=0}^{m}\beta_j X_j$ and thus $\epsilon=Y-\sum_{j=0}^{m}\beta_j X_j$ are unique.

A.2.2. Proof of Theorem 2

The regression parameters $\lambda_0,\lambda_1,\ldots,\lambda_m$ have already been derived at the end of Section 2 and note that $u_i=\mu_i-\lambda_0-\sum_{j=1}^{m}\lambda_j\beta_{ij}$. Hence, the residual u_i^2 is a measurable function of $\theta_i=(\mu_i,\beta_{i1},\beta_{i2},\ldots,\beta_{im})$. Since the market is ergodic, we have that

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(\mu_i-\lambda_0-\sum_{j=1}^{m}\lambda_j\beta_{ij}\right)^2 \quad\longrightarrow\quad & \mathbb{E}\left(\left(\vartheta_0-\lambda_0-\sum_{j=1}^{m}\lambda_j\vartheta_j\right)^2\right) && (66)\\
&= \mathbb{E}\big(v^2\big)<\infty\,, && (67)
\end{aligned}
$$

and from $\mathbb{E}(v)=0$ we conclude that $\mathbb{E}\big(v^2\big)=\sigma_v^2$. Moreover, the Linear Regression Theorem implies that

$$
\mathbb{V}\mathrm{ar}(\vartheta_0)=\lambda'\mathbb{V}\mathrm{ar}(\vartheta)\lambda+\mathbb{V}\mathrm{ar}(v)=\mathbb{C}\mathrm{ov}(\vartheta,\vartheta_0)'\mathbb{V}\mathrm{ar}(\vartheta)^{-1}\mathbb{C}\mathrm{ov}(\vartheta,\vartheta_0)+\sigma_v^2\,, \tag{68}
$$

i.e., $\sigma_v^2 = \mathbb{V}\mathrm{ar}(\vartheta_0) - \mathbb{C}\mathrm{ov}(\vartheta,\vartheta_0)'\mathbb{V}\mathrm{ar}(\vartheta)^{-1}\mathbb{C}\mathrm{ov}(\vartheta,\vartheta_0) \geq 0$.

A.2.3. Proof of Theorem 3

From $v = 0$ we conclude that

$$\vartheta_0 = \lambda_0 + \sum_{j=1}^{m} \lambda_j \vartheta_j = \lambda_0 + \sum_{j=1}^{m} \lambda_j \vartheta_j + \sum_{j=m+1}^{M} 0 \cdot \vartheta_j + 0 \tag{69}$$

is a linear regression equation. The covariance matrix of $(\vartheta_1, \vartheta_2, \ldots, \vartheta_M)$ is positive definite and thus Theorem 1 reveals that $\kappa_j = 0$ for all $j > m$ and $\kappa_j = \lambda_j$ for all $0 \leq j \leq m$. Conversely, due to the same arguments, the linear regression equation

$$\vartheta_0 = \kappa_0 + \sum_{j=1}^{m} \kappa_j \vartheta_j + \sum_{j=m+1}^{M} 0 \cdot \vartheta_j + 0 \tag{70}$$

implies that $\lambda_j = \kappa_j$ for all $0 \leq j \leq m$ and $v = 0$.

A.2.4. Proof of Theorem 4

The asymptotic results at the beginning of Section 3.2.1 reveal that $P_n \xrightarrow{\mathrm{P}} \sigma_v^2 \geq 0$ if the strategy $\{u/n\}$ is well-diversified. In the case in which σ_v^2 is positive, each investor who satisfies Assumption **A** will try to implement an infinitely large amount $a > 0$ of the strategy $\{au/n\}$ in order to maximize his expected utility $u(c + a\sigma_v^2)$. Hence, the aggregated demand for some asset must exceed its aggregated supply in the market. In this case, the market cannot be in equilibrium, but this contradicts the prerequisites of the theorem. Thus, it holds that $\sigma_v^2 = 0$ and from Theorem 2 we conclude that

$$\frac{1}{n}\sum_{i=1}^{n}\left(\mu_i - \lambda_0 - \sum_{j=1}^{m} \lambda_j \beta_{ij}\right)^2 \longrightarrow 0\,. \tag{71}$$

A.2.5. Proof of Lemma 1

The profit of $\{w_n/n\}$ amounts to

$$P_n = \frac{1}{n}\sum_{i=1}^{n} w_{ni}\mu_i + \sum_{j=1}^{m} \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} w_{ni}\beta_{ij}\right)}_{\to \beta_{wj}} X_j + \underbrace{\frac{1}{n}\sum_{i=1}^{n} w_{ni}\varepsilon_i}_{\rightsquigarrow \xi} \tag{72}$$

with

$$\frac{1}{n}\sum_{i=1}^{n} w_{ni}\mu_i = \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} w_{ni}\right)}_{\to\gamma}\lambda_0 + \sum_{j=1}^{m}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} w_{ni}\beta_{ij}\right)}_{\to\beta_{wj}}\lambda_j \tag{73}$$

$$+ \underbrace{\frac{1}{n}\sum_{i=1}^{n} w_{ni}u_i}_{\to 0} \longrightarrow \gamma\lambda_0 + \sum_{j=1}^{m}\beta_{wj}\lambda_j\,. \tag{74}$$

From the Continuous Mapping Theorem we conclude that

$$P_n \rightsquigarrow P = \begin{cases} \sum_{j=1}^{m}\beta_{wj}(\lambda_j + X_j) + \xi, & \gamma = 0\,, \\ \lambda_0 + \sum_{j=1}^{m}\beta_{wj}(\lambda_j + X_j) + \xi, & \gamma = 1\,. \end{cases} \tag{75}$$

According to Theorem 2.20 in van der Vaart (1998, p. 17), we have that $\mathrm{E}(\xi_n) \to \mathrm{E}(\xi)$ if and only if the sequence $\{\xi_n\}$ is asymptotically uniformly integrable. We already know that $\mathrm{E}(\xi_n) = 0$ and thus, if $\{\xi_n\}$ is asymptotically uniformly integrable, we obtain $\mathrm{E}(\xi) = 0$. This leads to

$$\mathrm{E}(P) = \begin{cases} \sum_{j=1}^{m}\lambda_j\beta_{wj}, & \gamma = 0\,, \\ \lambda_0 + \sum_{j=1}^{m}\lambda_j\beta_{wj}, & \gamma = 1\,. \end{cases} \tag{76}$$

Conversely, this result implies that $\mathrm{E}(\xi) = 0$ and so it is clear that $\mathrm{E}(\xi_n) \to \mathrm{E}(\xi)$. Hence, the sequence $\{\xi_n\}$ must be asymptotically uniformly integrable. Further, suppose that also the sequence $\{\xi_n^2\}$ is asymptotically uniformly integrable. The Continuous Mapping Theorem guarantees that $\xi_n^2 \rightsquigarrow \xi^2$ and thus $\mathrm{Var}(\xi_n) = \mathrm{E}(\xi_n^2) \to \mathrm{E}(\xi^2) = \mathrm{Var}(\xi) \geq 0$. Conversely, since we have that $\mathrm{E}(\xi) = 0$, $\mathrm{Var}(\xi_n) = \mathrm{Var}(\xi)$ just means that $\mathrm{E}(\xi_n^2) \to \mathrm{E}(\xi^2)$ and so the sequence $\{\xi_n^2\}$ must be asymptotically uniformly integrable.

A.2.6. Proof of Proposition 1

Suppose that Asset k can be replicated by an orthogonal strategy $\{w_n/n\}$. Entering a long position into Asset k and a short position into the strategy $\{w_n/n\}$ is self-financing. The profit amounts to

$$R_k - P_n = \left(\mu_k - \frac{1}{n}\sum_{i=1}^{n} w_{ni}\mu_i\right) + \sum_{j=1}^{m}\left(\beta_{kj} - \frac{1}{n}\sum_{i=1}^{n} w_{ni}\beta_{ij}\right)X_j \tag{77}$$

$$+ \left(\varepsilon_k - \frac{1}{n}\sum_{i=1}^{n} w_{ni}\varepsilon_i\right). \tag{78}$$

We already know that $\beta_{kj} - \frac{1}{n}\sum_{i=1}^{n} w_{ni}\beta_{ij} \to 0$ and the Continuous Mapping Theorem reveals that $\varepsilon_k - \frac{1}{n}\sum_{i=1}^{n} w_{ni}\varepsilon_i \xrightarrow{\text{p}} 0$. Finally, we have that

$$\mu_k - \frac{1}{n}\sum_{i=1}^{n} w_{ni}\mu_i = \underbrace{\left(1 - \frac{1}{n}\sum_{i=1}^{n} w_{ni}\right)}_{\to 0}\lambda_0 \tag{79}$$

$$+ \sum_{j=1}^{m} \underbrace{\left(\beta_{kj} - \frac{1}{n}\sum_{i=1}^{n} w_{ni}\beta_{ij}\right)}_{\to 0}\lambda_j + u_k - \underbrace{\frac{1}{n}\sum_{i=1}^{n} w_{ni}u_i}_{\to 0} \longrightarrow u_k\,, \tag{80}$$

which means that $R_k - P_n \xrightarrow{\text{p}} u_k$. Due to Assumption **A** and the fact that the market is in equilibrium, we must have that $u_k = 0$ and thus

$$\mu_k = \lambda_0 + \sum_{j=1}^{m} \lambda_j\beta_{kj}\,. \tag{81}$$

A.2.7. Proof of Theorem 5

The "only-if part" is an immediate consequence of Proposition 1. The "if part" follows from the fact that each asset represents an orthogonal strategy replicating itself if we have that $\mu_i = \lambda_0 + \sum_{j=1}^{m} \lambda_j\beta_{ij}$ and thus $u_i = 0$ for each asset in the market.

References

Admati, A., Pfleiderer, P. (1985): "Interpreting the factor risk premia in the Arbitrage Pricing Theory," *Journal of Economic Theory* **35**, pp. 191–195.

Al-Najjar, N. (1998): "Factor analysis and arbitrage pricing in large asset economies," *Journal of Economic Theory* **78**, pp. 231–262.

Antoniou, A., Garrett, I., Priestley, R. (1998): "Macroeconomic variables as common pervasive risk factors and the empirical content of the arbitrage pricing theory," *Journal of Empirical Finance* **5**, pp. 221–240.

Biagini, F. (2010): "Second fundamental theorem of asset pricing," in R. Cont (editor), "Encyclopedia of Quantitative Finance," Wiley.

Burmeister, E., McElroy, M. (1988): "Joint estimation of factor sensitivities and risk premia for the Arbitrage Pricing Theory," *Journal of Finance* **43**, pp. 721–733.

Chamberlain, G., Rothschild, M. (1983): "Arbitrage, factor structure, and mean-variance analysis on large asset markets," *Econometrica* **51**, pp. 1281–1304.

Chen, S., Jordan, B. (1993): "Some empirical tests in the APT: macrovariables vs. derived factors," *Journal of Banking and Finance* **17**, pp. 65–89.

Clare, A., Thomas, S. (1994): "Macroeconomic factors, the APT and the UK stockmarket," *Journal of Business Finance & Accounting* **21**, pp. 309–330.

Connor, G., Korajczyk, R. (1995): "The Arbitrage Pricing Theory and multifactor models of asset returns," in R. Jarrow, V. Maksimovic, W. Ziemba (editors), "Handbooks in Operations Research and Management Science," , Volume 9Chapter 4, Elsevier, pp. 87–144.

Dhrymes, P., Friend, I., Gultekin, N. (1984): "A critical reexamination of the empirical evidence on the Arbitrage Pricing Theory," *Journal of Finance* **39**, pp. 323–346.

Fama, E. (1970): "Efficient capital markets: A review of theory and empirical work," *Journal of Finance* **25**, pp. 383–417.

Frahm, G. (2016): "Pricing and valuation under the real-world measure," *International Journal of Theoretical and Applied Finance* **19**, DOI: 10.1142/S0219024916500060.

Gagliardini, P., Ossola, E., Scaillet, O. (2016): "Time-varying risk premium in large cross-sectional equity data sets," *Econometrica* **84**, pp. 985–1046.

Geweke, J., Zhou, G. (1996): "Measuring the pricing error of the Arbitrage Pricing Theory," *The Review of Financial Studies* **9**, pp. 557–587.

Harrison, J., Pliska, S. (1981): "Martingales and stochastic integrals in the theory of continuous trading," *Stochastic Processes and their Applications* **11**, pp. 215–260.

Harrison, J., Pliska, S. (1983): "A stochastic calculus model of continuous trading: complete markets," *Stochastic Processes and their Applications* **15**, pp. 313–316.

Huberman, G. (1982): "A simple approach to Arbitrage Pricing Theory," *Journal of Economic Theory* **28**, pp. 183–191.

Ingersoll, J. (1984): "Some results in the theory of arbitrage pricing," *Journal of Finance* **39**, pp. 1021–1039.

Ingersoll, J. (1987): *Theory of Financial Decision Making*, Rowman & Littlefield Publishers.

Jarrow, R., Rudd, A. (1983): "A comparison of the APT and CAPM," *Journal of Banking and Finance* **7**, pp. 295–303.

Markowitz, H. (1952): "Portfolio selection," *Journal of Finance* **7**, pp. 77–91.

Markowitz, H. (2012): "The "great confusion" concerning MPT," *AESTIMATIO, The IEB International Journal of Finance* **4**, pp. 8–27.

McElroy, M., Burmeister, E., Wall, K. (1985): "Two estimators for the APT model when factors are measured," *Economics Letters* **19**, pp. 271–275.

Reisman, H. (1988): "A general approach to the Arbitrage Pricing Theory (APT)," *Econometrica* **56**, pp. 473–476.

Renault, E., van der Heijden, T., Werker, B. (2017): "Arbitrage Pricing Theory for idiosyncratic variance factors," Technical report, Social Science Research Network, http://dx.doi.org/10.2139/ssrn.3065854.

Roll, R., Ross, S. (1980): "An empirical investigation of the Arbitrage Pricing Theory," *Journal of Finance* **35**, pp. 1073–1103.

Roll, R., Ross, S. (1983): "Regulation, the Capital Asset Pricing Model, and the Arbitrage Pricing Theory," *Public Utilities Fortnightly* **3**, pp. 22–28.

Roll, R., Ross, S. (1984a): "The Arbitrage Pricing Theory approach to strategic portfolio planning," *Financial Analysts Journal* **40**, pp. 14–26.

Roll, R., Ross, S. (1984b): "A critical reexamination of the empirical evidence on the Arbitrage Pricing Theory: a reply." *Journal of Finance* **39**, pp. 347–350.

Ross, S. (1976a): "The arbitrage theory of capital asset pricing," *Journal of Economic Theory* **13**, pp. 341–360.

Ross, S. (1976b): "Return, risk, and arbitrage," in I. Friend, J. Bicksler (editors), "Risk and Return in Finance," Ballinger, pp. 189–218.

Ross, S. (1982): "The general validity of the mean-variance approach in large markets," in "Financial Economics Essays in Honor of Paul Cootner," Prentice Hall, pp. 52–84.

Ross, S., Walsh, M. (1983): "A simple approach to the pricing of risky assets with uncertain exchange rates," in C. Wihlborg, R. Levich, R. Hawkins (editors), "Research in International Business and Finance," JAI Press, pp. 39–54.

van der Vaart, A. (1998): *Asymptotic Statistics*, Cambridge University Press.