GRADUATE SCHOOL
OF DECISION SCIENCES

Universität
Konstanz

GSDS
Working Paper
No. 2017-09

The CAPM with Measurement Error: There's life in the old dog yet!

Anastasia Morozova
Winfried Pohlmeier

April 2017

Graduate School of Decision Sciences

All processes within our society are based on decisions – whether they are individual or collective decisions. Understanding how these decisions are made will provide the tools with which we can address the root causes of social science issues.

The GSDS offers an open and communicative academic environment for doctoral researchers who deal with issues of decision making and their application to important social science problems. It combines the perspectives of the various social science disciplines for a comprehensive understanding of human decision behavior and its economic and political consequences.

The GSDS primarily focuses on economics, political science and psychology, but also encompasses the complementary disciplines computer science, sociology and statistics. The GSDS is structured around four interdisciplinary research areas: (A) Behavioural Decision Making, (B) Intertemporal Choice and Markets, (C) Political Decisions and Institutions and (D) Information Processing and Statistical Analysis.

GSDS – Graduate School of Decision Sciences
University of Konstanz
Box 146
78457 Konstanz

Phone: +49 (0)7531 88 3633
Fax: +49 (0)7531 88 5193
E-mail: gsds.office@uni-konstanz.de

-gsds.uni-konstanz.de

ISSN: 2365-4120
April 2017

The CAPM with Measurement Error: There's life in the old dog yet!*

Anastasia Morozova[†]

GSDS, University of Konstanz

Winfried Pohlmeier[‡]

University of Konstanz, COFE, RCEA

April 24, 2017

Abstract

This paper takes a closer look on the consequences of using a market index as a proxy for the latent market return in the capital asset pricing model. In particular, the consequences of two major sources of misspecification are analyzed: (i) the use of inaccurate weights and (ii) the use of only a subset of the asset universe to construct the index. The consequences resulting from the use of a badly chosen market proxy reach from inconsistent parameter estimates to misinterpretation of tests indicating the existence of abnormal returns.

A new minimum distance approach of estimating the CAPM under measurement error is presented, which identifies the CAPM parameters by exploiting the cross-equation cross-sectional restrictions resulting from a common measurement error. The new approach allows for quantifying the impact of measurement error and for testing the presence of spurious abnormal returns. Practical guidelines are presented to mitigate potential biases in the estimated CAPM parameters.

Keywords: CAPM, Measurement Error, Roll's critique, Identification, Minimum Distance Estimation

JEL classification: G12, C58, C51, C36

*Paper to be presented at the 3^{rd} Konstanz-Lancaster (KoLa) workshop on Finance and Econometrics, Lancaster, May 2-3, 2017. Financial support of the first author by the Graduate School of Decision Science (GSDS) and the German Academic Exchange Service (DAAD) is gratefully acknowledged. The second author likes to thank Simon Benninga posthumously for a helpful discussion and for motivating him to extend his notes on measurement errors in linear systems to an application for the CAPM. All remaining errors are ours.

[†]Department of Economics, Universitätsstraße 1, D-78462 Konstanz, Germany. Phone: +49-7531-88-2204, fax: -4450, email: Anastasia.Morozova@uni-konstanz.de.

[‡]Department of Economics, Universitätsstraße 1, D-78462 Konstanz, Germany. Phone: +49-7531-88-2660, fax: -4450, email: winfried.pohlmeier@uni-konstanz.de.

Measure what is measurable, and make

measurable what is not so.

Galileo Galilei

1 Introduction

The empirical evidence supporting capital asset pricing model (CAPM) in the version of Sharpe (1964) and Lintner (1965) is far from being convincing. Nevertheless the CAPM is still a center piece of the asset pricing theory taught in MBA investment courses and it is still a widely used tool among practitioners. The reasons for its failure are manifold and have launched a large body of literature. Besides its theoretical simplicity leaving room for numerous generalizations based on more realistic settings (multifactor models, conditional CAPM, consumption CAPM etc.), the failure of producing convincing evidence in favor of the CAPM can also be attributed to the difficulties of its empirical implementation. This branch of the literature has generated numerous studies of using alternative estimation and testing procedures.[1] Fama and French (2004) offer a concise summary of the struggle to find empirical support for the CAPM.

This paper addresses the implications of using the return of a market index as a proxy for the return of the true (equilibrium) market portfolio for the estimation of the CAPM. In a certain sense the measurement problem can be regarded as a primary one, because the CAPM as such is not questioned, but relates to the problem of obtaining a workable empirical framework. The measurement problem is at the heart of the Roll's critique (1977) who argues that the true market portfolio includes a large range of investment opportunities including international securities, real estate, precious metals, etc., so that the true value weighted market portfolio is empirically elusive. In particular, he concludes (Roll (1977), p.130): *"The Theory is not testable unless the exact composition of the true market portfolio is known and used in the tests."* Moreover, by using market proxies due to data limitations empirical tests of the CAPM effectively test whether the market proxies are on the minimum variance frontier.

We take different approach to Roll's identification problem by assuming that the CAPM holds true and investigate the properties of the CAPM in the presence of measurement problem. By using a linear projection framework within the Sharpe-Lintner version of the CAPM we

[1] See Campbell et al. (1997), Chap. 5 and Jaganathan et al. (2010) for surveys on the econometric implementation.

pick up Roll's critique to account for the fact that the market index in empirical studies is only a proxy that correlates more or less strongly with the true market index. This leads to an observational model with a non-zero intercept which is observationally equivalent to a general factor model with non-zero excess returns. Consequently, in the presence of measurement error results of standard tests on the existence of abnormal returns, e.g. in the tradition of Gibbons et al. (1989), render to be spurious. Our projection framework can serve as an alternative approach by providing an indirect test of the viability of the conventional CAPM by testing the CAPM against a general factor model with measurement error.

The sources of misspecification of the market return can be manifold and lead to different biases in the estimated alphas and betas. Generally the weights in the market index used for a CAPM regression may simply differ from the true weights by including a subset of assets and/or by misspecification of the weights even if the asset universe is correctly defined. The choice of an index consisting of a limited number of large stocks such as the Dow Jones Industrial Average (DJIA) can serve as an example of the first type of misspecification, while, the choice between a volume weighted versus an equally weighted index is an example for the latter type of misspecification. What seems to be an econometric problem in the first place for both types of measurement error has its roots in false theoretical model assumptions, e.g. assumptions about the fixed supply side in terms of volume or quantities.

Measurement error in the market index leads to a systematic bias in the parameter estimates and may therefore lead to erroneous investment strategies. As we will show in this study, the direction and severity of the bias of the CAPM parameters strongly depend on the nature of the measurement error. The consequences for the least squares estimates in the classical errors-in-variables case, where the market index differs from the latent true market return only by an additive idiosyncratic error, are well-known: First, estimates of the beta coefficients suffer from an attenuation bias, i.e. the estimated risk premia are biased towards zero. Second, the estimated intercepts are biased upwards such that positive alphas occur even if the CAPM holds true.

This paper takes a closer look on the measurement error bias in the CAPM beyond the conventional measurement error bias. In particular, we derive the bias for the CAPM alphas and betas under different assumptions on the type of misspecification of the market index. In particular, we show that the typical attenuation bias occurring in linear regression models with additive measurement error generally does not hold. By means of Monte Carlo simulations, where the return process is generated from an artificial capital market, we assess the size of the bias and provide practical guidance for the choice of the market index in empirical work.

Finally this paper presents a novel approach of estimating the CAPM in the presence of measurement error. Contrary to general systems of linear regression equations with measurement error, the CAPM contains the same mis-measured explanatory variable in each equation. Using the property that the CAPM with measurement error is a system of regression equations with nonlinear cross-equation restrictions we present a new identification strategy which is superior to instrumental variables approaches that typically suffer from the weak instrument problem as market returns are only weakly auto-correlated. Our minimum-distance approach is easy to implement and allows to estimate different versions of the CAPM including the true CAPM without measurement error, the CAPM with measurement error and a factor model with measurement error and excess returns under rather general assumptions on the type of the measurement error.

While the vast majority of empirical studies simply ignores the measurement problem or implicitly assumes that its impact on the parameter estimates is negligible, only a few studies consider the impact of measurement error in the market return on the outcome of efficiency tests (e.g. Stambaugh (1982), Kandel and Stambaugh (1987) and Shanken (1987)). They show that a rejection of market efficiency by the market proxy implies also a rejection for the true portfolio if the true market portfolio is sufficiently correlated with the proxy (ca. 0.7 or larger). The projection framework considered in these studies, however, ignores that the measurement error maybe endogenous, i.e. the orthogonality between the (rational expectation) error in the CAPM and true market index is generally violated. Prono (2015) proposes a new measure of misspecification that not only accounts for the latency of the true market index and the resulting imperfect correlation between market proxy and the true market index but also for the effect of endogeneity on the CAPM estimates.

Jagannathan and Wang (1996) take a different perspective by trying to get closer to the theoretical concept of the market return. They use a broader market proxy which also takes into account the returns from human capital. In their empirical study of the conditional CAPM based on the broader concept the additional explanatory power of size and book-to-marker variables becomes negligible. Unlike previous studies analyzing the potential impact of measurement error in the market proxy on efficiency tests, the focus of this study is to assess its impact on the CAPM estimates with obvious consequences for performance measures, choices of investment strategies and outcomes of efficiency tests. Rather than defining correlation bounds we estimate the size of the attenuation bias (possibly the size of an amplification bias in some settings) directly. This yields new insights into the quality of different market proxies and provides evidence for the presence of spurious abnormal returns.

The paper is organized as follows. In Section 2 we introduce the theoretical framework under which the true return generation process of the true but infeasible market index is defined. Various types of misspecification of the market index and their consequences for estimation are considered as deviations from the true return generating process. Based on Monte-Carlo simulations we provide in Section 3 a quantitative assessment on the extend of the bias caused by different types of misspecification of the market index. In Section 4 we provide empirical evidence for the presence of measurement error in the market index using three different data sets and different definitions of the market returns. Section 5 concludes and gives an outlook on future research.

2 CAPM and Measurement Error in the the Market Index

In the following we consider a well-defined CAPM where asset returns are equilibrium outcomes from security markets with rational investors. Therefore the data generating process for the returns and the true but unobservable market return is such that the CAPM holds by construction and identification of the model parameters is feasible. The initial set-up is based on common assumptions underlying the CAPM (e.g. Gourieroux and Jasiak (2001) and Fan and Yao (2017)) and is sufficiently flexible to allow for range of generalizations concerning the price process and assumptions on the investors' behavior. In a second step we deviate from the world of a perfect data generating process by replacing the true market return by different proxies and derive the conditions under which the true model parameters are feasible.

2.1 The Baseline Model

Let there be N risky assets and one risk-free asset. The portfolio is given by the vector of quantities $(q_0, q_1, \ldots, q_N)' = (q_0, q')'$. The price vector for the $N + 1$ assets at time t is given by $\left(1, P_{1,t}, \ldots, P_{N,t}\right)' = \left(1, P_t'\right)'$, where the price of the risk-free asset is taken as numeraire. Expected portfolio wealth for period $t+1$, W_{t+1}, given information up to t is given by $\mathrm{E}\left[W_{t+1}\middle| \mathcal{F}_t\right] \equiv \mathrm{E}_t[W_{t+1}] = q_0(1 + R_t^f) + q'\mathrm{E}_t[P_{t+1}]$. The allocation problem of investor i for period $t+1$ is given by

$$\max_{q_0,q} \quad \mathrm{E}_t[W_{t+1}] - \frac{\gamma_i}{2}\mathrm{V}_t[W_{t+1}] \qquad \text{s.t.} \qquad W_t = q_0 + q'P_t,$$

where W_t denotes the initial endowment in t and γ_i is the risk aversion parameter of investor i. Optimal allocation for investor i takes the well-known form:

$$q_{i,t}^D = \frac{1}{\gamma_i}\mathrm{V}_t[Y_{t+1}]^{-1}\mathrm{E}_t[Y_{t+1}],$$

where $Y_{t+1} = P_{t+1} - P_t(1 + R_{t+1}^f)$ is the vector of excess gains. Aggregate demand for a total number of I investors is given by

$$q_t^D = \sum_{i=1}^{I} q_{i,t}^D = \frac{1}{\gamma}\mathrm{V}_t[Y_{t+1}]^{-1}\mathrm{E}_t[Y_{t+1}],$$

with $\gamma = [\sum_{i=1}^{I}\frac{1}{\gamma_i}]^{-1}$ as the absolute risk aversion parameter of the market. Since most empirical studies use excess returns instead of excess gains, we reformulate excess gains in terms of excess returns such that $Y_{t+1} = \mathrm{diag}(P_t)r_{t+1}$, with the vector of excess returns defined as $r_{t+1} = R_{t+1} - \iota R_{t+1}^f$ with R_{t+1} as the $N \times 1$ vector of the return rates on the risky assets. In terms of the excess returns aggregate demand is given by

$$q_t^D = \frac{1}{\gamma}\mathrm{diag}(P_t)^{-1}\mathrm{V}_t[r_{t+1}]^{-1}\mathrm{E}_t[r_{t+1}]. \tag{1}$$

For the definition of the market index the assumption on the supply of assets is absolutely crucial. In the following we assume for the supply

$$q_t^S = \text{diag}(P_t)^{-1} b^*, \tag{2}$$

which is based on the underlying assumption that the value of each asset supplied is fixed. The market equilibrium $q_t^S = q_t^D$ yields a return process of the form

$$r_{t+1} = \gamma \text{V}_t[r_{t+1}] b^* + \varepsilon_{t+1}, \tag{3}$$

where the expectation error $\varepsilon_{t+1} = r_{t+1} - \text{E}_t[r_{t+1}]$ is a martingale difference sequence. The excess return of the market is defined as

$$r^*_{m,t+1} = b^{*\prime}\, r_{t+1}, \tag{4}$$

which is a value weighted index given the fixed value assumption from (2). The asterisk on $r^*_{m,t+1}$ indicates that the true market return is an unobservable random variable depending on the unknown parameter vector b^*. The process for the excess returns of the market takes on the form

$$r^*_{m,t+1} = \gamma\, b^{*\prime} \text{V}_t[r_{t+1}] b^* + b^{*\prime} \varepsilon_{t+1}\,. \tag{5}$$

For our simulations we assume for simplicity a homoskedastic process for the excess returns with $\text{V}_t[r_{t+1}] = \Omega$. This assumption is not crucial for the goal of our study and can easily be relaxed. Under homoskedasticity the processes for the returns and the market return simplifies to

$$r_{t+1} = \mu + \varepsilon_{t+1}, \tag{6}$$

$$r^*_{m,t+1} = \mu^*_m + b^{*\prime} \varepsilon_{t+1}, \tag{7}$$

with $\mu = \gamma \Omega b^*$ and $\mu^*_m = b^{*\prime} \mu = \gamma b^{*\prime} \Omega b^*$. Without loss of generality we can define the CAPM as the set of linear predictor equations of the excess returns r_{t+1} on the excess market return

$r^*_{m,t+1}$:

$$r_{t+1} = \alpha + \beta\, r^*_{m,t+1} + u_{t+1}. \tag{8}$$

By definition the vector of CAPM betas is given by

$$\beta = \frac{\mathrm{Cov}\left[r_{t+1}, r^*_{m,t+1}\right]}{\mathrm{V}\left[r^*_{m,t+1}\right]} = \frac{\Omega b^*}{b^{*\prime}\Omega b^*}\,, \tag{9}$$

where the second equality follows from the homoskedasticity assumption. Note, that the true CAPM betas are a function of the true, unobservable weighting scheme, b^*, and the variance-covariance matrix of the vector of returns of the entire asset universe Ω. The vector of intercepts of the linear predictor equations (CAPM alphas) vanishes, since $\alpha = \mathrm{E}\left[r_{t+1}\right] - \beta\,\mathrm{E}\left[r^*_{m,t+1}\right] = 0$. Moreover, since u_{t+1} and $r^*_{m,t+1}$ are orthogonal β can be consistently estimated by least squares.

2.2 Misspecified Index Weights

The true CAPM is basically given by the relationship between the two processes defined in (6) and (7). In the following we consider the linear relationships between the return process and alternative specifications of the index based on an observable weighting vector $b \neq b^*$. As shown below, the specific assumptions on b and its relationship to the true weighting index lead to different identification conditions concerning alpha and beta.

Weights with Random Measurement Errors

Consider first the case where the actual weights used to construct the market index deviate randomly from the true weights such that the actual weighting scheme b differs randomly from b^*:

$$b = b^* + \nu, \tag{10}$$

where ν is random vector of error terms with $\mathrm{E}\left[\nu\right] = 0$ and $\mathrm{V}\left[\nu\right] = \sigma^2_\nu\, I_N$.[2] The market index based on the mis-measured weighting scheme (10) yields a process for the observable proxy of

[2] The model may also be derivable under the general assumption $\mathrm{V}\left[\nu\right] = \Sigma^2_\nu$. This generalization can easily be introduced but does not add any additional new insights for identification problem.

the market return of the form

$$r_{m,t+1} = r^*_{m,t+1} + \nu' r_{t+1}, \tag{11}$$

such that the return of the market index varies randomly around the true market return. The CAPM equation based on the observable index (11) is given by

$$r_{t+1} = \beta\, r_{m,t+1} + \omega_{t+1}, \tag{12}$$

with $\omega_{t+1} = u_{t+1} - \beta\nu' r_{t+1}$. Contrary to the true market return its proxy does not satisfy the orthogonality condition with the error term, $\mathrm{E}\left[\omega_{t+1} r_{m,t+1}\right] = -\sigma^2_\nu\, tr(\mu\mu' + \Omega)\,\beta \neq 0$, so that least squares estimation of (12) yields inconsistent parameter estimates. Model (12) shares similar properties with a classical linear errors-in-variables (EIV) model (e.g. Fuller (1987)). However, contrary to the classical EIV model lagged market proxies $r_{m,t-j}$ cannot serve as instruments. They are orthogonal to ω_{t+1}, but they are not autocorrelated.

Although IV estimation is infeasible the parameters of the CAPM can nevertheless be identified by exploiting the information on the first and second moments of the market proxy. In order to detect the relationship between identifiable estimable parameters and the true model parameters consider first the linear projection of $r^*_{m,t+1}$ on $r_{m,t+1}$:

$$r^*_{m,t+1} = \lambda_0 + \lambda_1\, r_{m,t+1} + \zeta_{t+1},$$

with

$$\begin{aligned} \lambda_1 =& \frac{\mathrm{Cov}\left[r^*_{m,t+1}, r_{m,t+1}\right]}{\mathrm{V}\left[r_{m,t+1}\right]} = \frac{b^{*\prime}\Omega b^*}{\sigma^2_\nu\big(\mu'\mu + tr\Omega\big) + b^{*\prime}\Omega b^*}, \\ \lambda_0 =& \,\mathrm{E}\left[r^*_{m,t+1}\right] - \lambda_1\, \mathrm{E}\left[r_{m,t+1}\right] = (1-\lambda_1)\, \mathrm{E}\left[r_{m,t+1}\right] . \end{aligned}$$

Inserting the linear predictor function in the CAPM equation yields :

$$r_{t+1} = \alpha_\nu + \beta_\nu\, r_{m,t+1} + \tilde{u}_{t+1}\,, \tag{13}$$

with $\alpha_\nu = \beta(1-\lambda_1)\,\mu_m$ and $\beta_\nu = \beta\lambda_1$. The error term $\tilde{u}_{t+1} = u_{t+1} + \beta\zeta_{t+1}$ is orthogonal to

$r_{m,t+1}$ so that consistent estimates of α_ν and β_ν can be obtained.

Note that $0 < \lambda_1 < 1$, which is the usual reliability ratio in EIV models. The estimation of parameters leads to the well-known attenuation bias for the slope coefficient, i.e. the estimates of CAPM betas are driven towards zero, i.e. under the presence of measurement error the least squares estimates mimic a too small dependence on the market risk.

Moreover, the EIV-CAPM yields positive intercepts, $\alpha_\nu > 0$. Thus in the absence of abnormal returns ($\alpha = 0$), the EIV-CAPM mimics spurious abnormal returns, even if the CAPM holds true. Consequently in the presence of measurement error tests on the existence of abnormal returns ignoring the measurement error are jointly testing the validity of the CAPM and the absence of measurement error.

The sparse parametrization of the standard CAPM model and the parsimonious parametrization of the measurement error reflected only by the unknown parameter σ_ν^2 yields an overidentified model with only $N+1$ unknown parameters compared to $2N$ identifiable reduced form parameters given by α_ν and β_ν. The strong degree of overidentification simply results from the fact that i) the measurement error effects all N equations in the same way through the reliability ratio and ii) the asset specific intercepts are nonlinear functions of the reliability ratio and the true betas. Even without exploiting the nonlinear cross-equation restrictions the true beta and the reliability ratio can be identified by a single equation estimate provided the CAPM holds true.

Consider the nonlinear restriction for the model parameters for a single equation j with $\beta_{\nu j} = \lambda_1 \beta_j$ and $\alpha_{\nu j} = \beta_j(1-\lambda_1)\mu_m$. Then the reliability ratio is identified as the solution of the two equations as

$$\lambda_1 = \frac{\mu_m\left(\frac{\beta_{\nu j}}{\alpha_{\nu j}}\right)}{\mu_m\left(\frac{\beta_{\nu j}}{\alpha_{\nu j}}\right)+1}. \tag{14}$$

A simple estimate of the reliability ratio can be obtained by replacing the unknown parameters in (14) by the least squares estimates from (13), while μ_m can be estimated by the mean of the excess returns. Since this simple procedure generates estimates for every equation j, it seems meaningful to take the average over the single equation estimates in order to stabilize the results. Once λ_1 is determined β_j and α_j are identified.

Obviously, estimation of the system of equations (13) by ML, GMM or Minimum Distance estimation yield asymptotically more efficient estimates. In Section 4 we present an empirical application of this identification strategy based on the minimum distance estimation.

The EIV-CAPM can be generalized to the case where only $M < N$ assets define the market index and the remaining $N - M$ assets are ignored and do not enter the market proxy with a weighting vector given by:

$$b = \begin{bmatrix} b_M^* + \nu_M \\ 0 \end{bmatrix},$$

where b_M^* is the sub-vector of b^* of dimension M and ν_M is the corresponding vector of measurement errors. The relationship between the true market return and the market proxy becomes $r_{m,t+1} = r_{m,t+1}^* + w_{t+1}$ with $\mathrm{E}\left[w_{t+1}\right] \neq 0$. In this case the mean of the market proxy deviates from the mean of the true market return. A linear projection representation is feasible, but identification can only be obtained under additional assumptions. For instance, (i) the mean return of the assets ignored has to be equal to the mean of the market proxy's return and (ii) the overall measurement error w_{t+1} should be uncorrelated with u_{t+1}.

Weights with Fixed Measurement Error

Consider now the case where the differences between weights for the market index and the true market return are fixed such that

$$b = b^* + \Delta\,, \tag{15}$$

where Δ denotes the vector of fixed deviations from the true weights. This case includes a number of interesting special cases. For instance, if the equally weighted index, $b = \iota_N \frac{1}{N}$, is used instead of b^* the market proxy is simply the average over all return rates in the asset universe. If the proxy is based only on a subset of $M < N$ assets the deviation of the weights of the proxy from the true weights take the form $\Delta' = (\Delta_1, \ldots, \Delta_M, -b_{M+1}^*, \ldots, -b_N^*)'$, such that the first M assets receive some positive (most likely erroneous) weights, while the weights of the remaining assets are ignored. The return of the observable market index takes the form:

$$r_{m,t+1} = b' r_{t+1} = r_{m,t+1}^* + \Delta' r_{t+1}\,. \tag{16}$$

Replacing $r^*_{m,t+1}$ in (8) by $r_{m,t+1}$ gives a system of CAPM equations based on the fixed error market proxy:

$$r_{t+1} = \alpha_\Delta + \beta\, r_{m,t+1} + \omega_{t+1}, \tag{17}$$

with $\omega_{t+1} = u_{t+1} - \Delta'\varepsilon_{t+1}$ and $\alpha_\Delta = -\beta\Delta'\mu$. Contrary to the random error case given by (12) the observable system of CAPM equations contains nonzero intercepts, such that a test for the existence of abnormal returns would be misleading. Also contrary to the random error case the market proxy is no longer orthogonal to the error term, $\mathrm{E}\left[\omega_{t+1} r_{m,t+1}\right] = (I - \beta b')\Omega b \neq 0$. Therefore, estimation approaches based on the orthogonality assumption between the market proxy and the error term are inconsistent.

Replacing the latent market index by its linear projection on the observed index does not help, since the overall error term would also be correlated with the market proxy. Last but not least the size of the coefficient in the projection equation cannot be derived in the fixed error case, so that no ex-ante statements on the direction of the bias for the estimates of beta can be derived.

3 Monte Carlo Evidence

3.1 Simulation design

By means of Monte Carlo simulations for returns generated from an artificial capital market we illustrate in this section how and to what extend different proxies of the market index influence the quality of the CAPM parameter estimates. By simulating from a well-defined artificial capital market, for which the CAPM holds, we can define market proxies as the outcome of the true data generating process combined with misspecified weights rather than imposing arbitrary stochastic assumptions on the error process.

Our simulation study is based on $10,000$ Monte Carlo samples of monthly excess return series for an asset universe of $N = 205$ assets over 10 years. The data generating process for the excess returns is given by

$$r_{t+1} = \gamma\Omega b^* + \varepsilon_{t+1}, \quad \text{with} \quad \varepsilon_{t+1} \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \Omega\right).$$

The variance-covariance matrix Ω was chosen to be equal to the sample variance-covariance matrix calculated from monthly data on excess returns of 205 components of S&P500 index from January 1, 1974 till May 1, 2015. In order to use realistic values for the true weight vector b^* we use their empirical counter parts. More precisely, for each of the 205 stocks of the S&P500 index we compute the mean value of the market capitalization based on monthly data from January 1, 1974 till May 1, 2015 and define the true market weights as a proportion of the total market capitalization of the 205 stocks. Finally, the coefficient of risk-aversion γ was chosen to be equal to 0.04. Following our baseline model we assume a fixed supply of assets, so that $r^*_{m,t+1}$ is generated according to (4).

In a second step we generate proxies of the true market index under different types of measurement error. Table 1 summarizes the five different weighting strategies for the market proxy used in the Monte Carlo study.

Table 1: Functional form of misspecified weights b of market indices.

b	Weight	Description
b^{EW}	$\iota_N \frac{1}{N}$	equally weighted index, the asset universe is correctly defined
b^{RE}	$b^* + \nu, \nu \sim N(0, \sigma^2_\nu I_N)$	weights with random error, the asset universe is correctly defined
b^I_1	$b^* + \Delta$, $\Delta' = (0, \ldots, 0, -b^*_{M+1}, \ldots, -b^*_N)'$	index with true weights, subset of M assets is considered
b^I_2	$\frac{b^I_1}{\iota'_M b^I_1}$	normalized version of b^I_1 such that weights sum up to unity
b^I_3	$\iota_M \frac{1}{M}$	equally weighted index, subset of M assets is considered

For the case of random measurement errors we consider three different choices for the variance of the measurement error, $\sigma_\nu = 0.05, 0.025,$ and 0.01. For our analysis of the market proxies based on the subsets of the asset space we choose subsets covering 25% and 75% of the total number of assets. For these subsets only assets with the largest weights are selected, so that the indices are more comparable to real world market indices. We estimate the CAPM parameters for 15 randomly drawn assets by the seemingly unrelated regression (SUR) method. Table 2 summarizes our findings for market proxies based on the total asset space, while Table 3 contains the results for indices based on specific subsets. Both tables summarize our findings by reporting the means of the estimates for 15 selected assets. The detailed results for each of the 15 assets are given in Table 5 in Appendix A.2.

For reasons of comparison column 2 of Table 2 contains the results of the CAPM when the true market return is feasible. Since these estimates are obtained under the true data generating process they only differ from the true model parameters by the sampling error. Therefore these estimates can serve as a benchmark for the estimates using market proxies. Under the true data generating process the CAPM alphas are on average close to their theoretical value of zero. The empirical rejection rate for $\hat{\alpha}_j$ is close to the 5 % significance level, which indicates that the sample size of $T = 120$ chosen for the Monte-Carlo simulations is sufficiently large to produce estimates that come close to the true parameters given the distributional assumptions and the true market index.

Only for the case of a small measurement error, $\sigma_\nu = 0.01$, the EIV-CAPM shows negligible distortions of the parameter estimates. The empirical correlation between the true market returns and the proxy is 0.96 and the attenuation bias reflected by $\overline{\widehat{\lambda_1}}$ is small. However, even for this mild case of measurement error, we find an empirical rejection rate of the null of no abnormal returns of 8%. The situation deteriorates for larger measurement errors. For the intermediate case with $\sigma_\nu = 0.025$ the correlation between true market return and the market proxy appears to be rather high with $\hat{\rho}(r^*_{m,t}, r_{m,t}) = 0.81$. However, for more than 47% of our estimates we find abnormal returns mimicking the existence of potential profits from trading. For the case of a large measurement error, the situation deteriorates even more, although the correlation between true market return and the proxy still remains 0.5. The three scenarios demonstrate that the correlation between the true market return and the market proxy provides insufficient information to make any conclusion about the bias in the CAPM estimates. After all what matters is the coefficient on the linear projection λ_1, which consists of the product of the square root of the reliability ratio and the correlation coefficient.[3]

[3]For the case of random errors this relationship is given by:

$$\lambda_1 = \frac{\text{Cov}\left[r_{m,t}, r^*_{m,t}\right]}{\text{V}\left[r_{m,t}\right]} = \Big(\frac{\text{V}\left[r^*_{m,t}\right]}{\text{V}\left[r_{m,t}\right]}\Big)^{1/2} \rho(r^*_{m,t}, r_{m,t}) = \Big(\frac{b^{*\prime}\Omega b^*}{(b^{*\prime}\Omega b^* + \sigma^2_\nu(\mu'\mu + tr\Omega)}\Big)^{1/2} \rho(r^*_{m,t}, r_{m,t}).$$

Table 2: MC-results for the CAPM with random measurement error (EIV-CAPM)

	$r^*_{m,t}$	$r_{m,t}$		
		$\sigma_\nu = 0.05$	$\sigma_\nu = 0.025$	$\sigma_\nu = 0.01$
$RR(r^*_{m,t}, r_{m,t})$		0.1997	0.4995	0.8618
$\hat{\rho}(r^*_{m,t}, r_{m,t})$		0.5357	0.8125	0.9694
$\overline{\widehat{\lambda_1}}$		0.3025	0.7002	0.9585
$\overline{\widehat{\alpha}}$	-0.0060	5.0250	2.6637	0.5024
$RMSE(\hat{\alpha})$	5.8412	23.0157	14.1939	6.4369
$ERR(.05)$	0.0513	0.8122	0.4755	0.0797
$\overline{\widehat{\beta}}$	0.9526	0.2834	0.6600	0.9110
$RMSE(\hat{\beta})$	0.5878	2.9080	1.5124	0.7997
$\overline{\widehat{SR}}$	0.1233	0.0491	0.0845	0.1166
$\overline{\widehat{TR}}$	9.0642	21.9288	12.5479	9.6945

Mean estimates for 10,000 replications and the parameter estimates for the different assets. The dimension of the asset universe is $N = 205$, the number of assets in the regression model equals 15. Column 2 contains the average estimates based on the true (excess) market return. Columns 3 - 5 contain estimation results for the EIV-CAPM with the proxies for market returns based on weights containing random errors. $RR(r^*_{m,t}, r_{m,t}) = \mathrm{V}\left[r^*_{m,t}\right] / \mathrm{V}\left[r_{m,t}\right]$ is the theoretical reliability ratio of the market proxy. $\overline{\widehat{\lambda_1}}$ is the mean estimate of the linear projection parameter. $ERR(.05)$ denotes the average empirical rejection rate for the null of no abnormal returns (single equation test) for a significance level of 5 %. $\overline{SR}$ and $\overline{TR}$ denote the mean Sharpe ratio and the mean Treynor ratio across all Monte Carlo estimates and assets. True values: $\overline{\beta} = 0.9520$, $\overline{SR} = 0.1175$, $\overline{TR} = 8.8070$.

Note also, the impact of measurement error on the precision of the estimates. Compared to the RMSE for the benchmark model, the RMSE for beta increases by 36% in the case of a small measurement error and almost triple for the medium size measurement error. Besides the attenuation bias affecting all betas uniformly the larger MSE due to the measurement error also increases the risk of a faulty sorting into defensive and aggressive stocks.

Table 3 contains the results for the case where the errors are fixed and the market proxies are based on a subset of the asset universe. Using a market proxy based on equal weights generates the expected distortions. We consider seven different scenarios with fixed measurement errors. The first two cases (columns 2 and 3) capture the case where the market proxy is based on the true weights for the 25 percent and the 75 percent largest assets and the corresponding smaller assets are ignored. Note, that in this case where the true weights of the smaller assets are ignored the active weights do not add up to one. We therefore also report in columns 3 and

4 the results based on the same index as before, but with normalized weights such that they satisfy the adding up constraint. This simply changes the dimensionality of the market proxy and has a scaling effect on the betas.

Table 3: MC-results for the CAPM with fixed measurement error

	true weights		normalized weights		equal weights		
M	0.25N	0.75N	0.25N	0.75N	0.25N	0.75N	M=N
$RR(r^*_{m,t}, r_{m,t})$	1.8371	1.0821	0.9745	1.0008	0.8895	0.9801	0.9497
$\hat{\rho}(r^*_{m,t}, r_{m,t})$	0.9934	0.9998	0.9934	0.9998	0.9908	0.9806	0.9730
$\overline{\widehat{\lambda_1}}$	1.3465	1.0401	0.9807	1.0002	0.9343	0.9709	0.9481
$\overline{\hat{\alpha}}$	0.2954	0.0273	0.2954	0.0273	0.2700	0.0205	0.0872
RMSE($\hat{\alpha}$)	6.1324	5.8448	6.1324	5.8448	6.0944	6.1449	6.2817
$ERR(.05)$	0.0620	0.0528	0.0620	0.0528	0.0638	0.0778	0.0903
$\overline{\hat{\beta}}$	1.2530	0.9872	0.9125	0.9494	0.8769	0.9587	0.9434
RMSE($\hat{\beta}$)	1.4764	0.6271	0.6248	0.5876	0.6439	0.6343	0.6454
$\overline{SR}$	0.1207	0.1237	0.1207	0.1237	0.1182	0.1214	0.1197
$\overline{TR}$	6.9287	8.7436	9.5136	9.0915	10.0315	9.0675	9.2609

Mean estimates for 10,000 replications and the parameter estimates for the different assets. The dimension of the asset universe is $N = 205$, the number of assets in the regression model equals 15. Columns 2 - 8 contain estimation results for the CAPM with the proxies for market returns based on weights containing fixed measurement errors. $ERR(.05)$ denotes the average empirical rejection rate for the absence of abnormal returns (single equation test) for a significance level of 5 %. $\overline{SR}$ and $\overline{TR}$ denote the mean Sharpe ratio and the mean Treynor ratio across all Monte Carlo estimates and assets. True values: $\overline{\beta} = 0.9520$, $\overline{SR} = 0.1175$, $\overline{TR} = 8.8070$.

The last three columns of Table 3 contain the estimates for models where the true weighting scheme is replaced by equal weights. In the last column the estimation results are given when the index is based on the all assets but equal weights of size 1/205 are used instead of b^*. These estimates come rather close to the ones obtained when the true market index is used. Both the reliability ratio as well as the correlation between true index and the proxy are close to one. However, this particular finding should not be overemphasized because it is strongly based on the underlying Monte-Carlo design. The true weights and the equal weights of size 1/205 do not differ very much, so that the fixed measurement error turns out to be rather small. More interesting is the non-monotonicity of the quality of the estimates in terms of the number of assets used. The estimates where the index represents the 75 percent largest assets are slightly superior compared to the two other scenarios. If the index represents only the capitalization of 25 percent largest assets the estimates reveal the largest biases. For all scenarios we find a slight over-rejection of the null of no abnormal returns.

3.2 Effects on Performance Measures

CAPM parameter estimates are frequently used to compute performance measures for single assets or portfolios. Obviously the bias in the parameter estimates directly passes on to biases in these performance measures. Consider, for instance, the Sharpe ratio and the Treynor ratio for asset $j = 1, \ldots, N$:

$$SR_j = \frac{\mathrm{E}\left[r_j\right]}{\sigma_j} = \frac{\beta_j \, \mathrm{E}\left[r_m\right]}{\sigma_j} \qquad \text{and} \qquad TR_j = \frac{\mathrm{E}\left[r_j\right]}{\beta_j} = \mathrm{E}\left[r_m\right].$$

In last two rows of Table 2 and Table 3 we report the estimates of average Sharpe ratios and Treynor ratios under the different regimes of misspecification. The attenuation bias for beta also leads to an attenuation bias for the Sharpe ratio, while for the Treynor ratio the attenuation bias leads to a strong upward bias because the too small estimates for beta enter the denominator of the Treynor ratio.

Figure 1 depicts the box plots for the estimated Treynor ratio for each of the 15 selected assets. As implied by the theory the mean estimates are the same for all 15 assets and are equal to the true value of the expected excess return of the market indicated by the blue horizontal line. Note that due to the large sampling variation the Treynor ratio can only be estimated with low precision, even in the absence of measurement error.

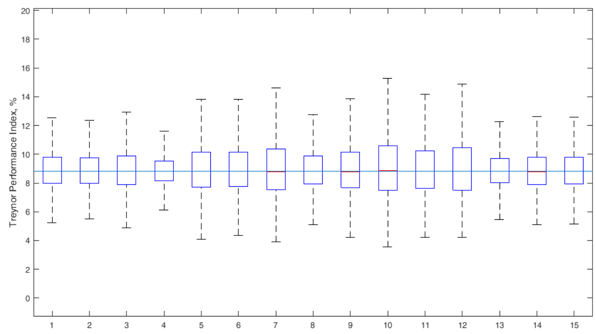


Figure 1: Estimates of the Treynor Ratio for 15 assets based on $r^*_{m,t}$

Figure 2 depicts the box plots for the Treynor ratio based on equal weights and weights with random measurement error. With measurement error in the market return we find an even stronger variation in the performance across assets although the CAPM holds true. The presence of measurement error mimics investment opportunities where none exist.

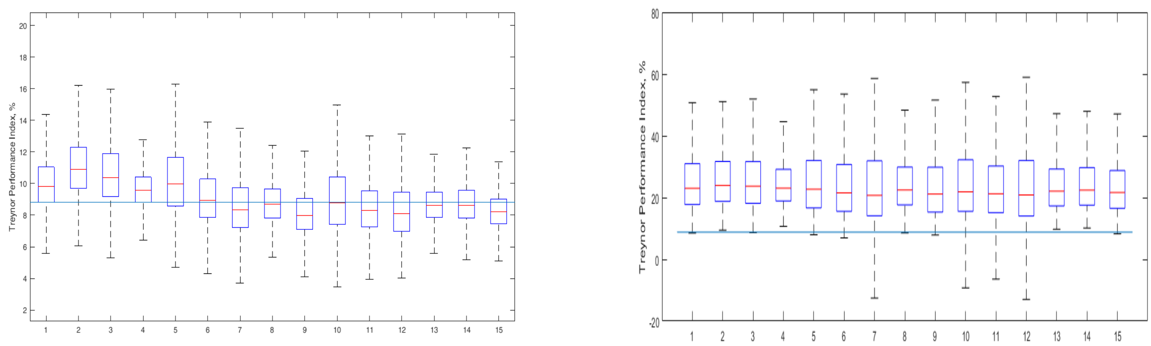


Figure 2: Estimates of the Treynor Ratio in the case of known asset universe and equally weighted market index (left) or with random measurement error with $\sigma_\nu = 0.05$ (right)

The consequences of measurement error in the market return for investment decisions based on CAPM estimates can also be seen for the security market line (SML). Figure 3 depicts the unbiased CAPM estimates based on true market return (green dots) and the biased estimates based on equal weights (blue dots). If the market return is measured correctly the estimates are scattered closely around the SML indicating no need for reshuffling the portfolio. However, with measurement error the estimates indicate investment opportunities due the spurious abnormal returns. Note, that the case depicted here is based on the true SLM, i.e its slope is given by μ^*. Since in the case of random measurement errors, $\mu^* = \mu$ the sample mean of the excess return $r_{m,t}$ can serve as an unbiased estimate for the slope of the SML. However, in all other cases considered $\mu^* \neq \mu$, and the mean excess return of the market proxy does not yield a consistent estimate of the slope. In this case the CAPM estimates deviate systematically from the true SML as can be seen from Figure 9 in Appendix A.3. The biased estimated SML and the attenuated CAPM beta estimates mimic investment opportunities with assets lying above and below the estimated SML.

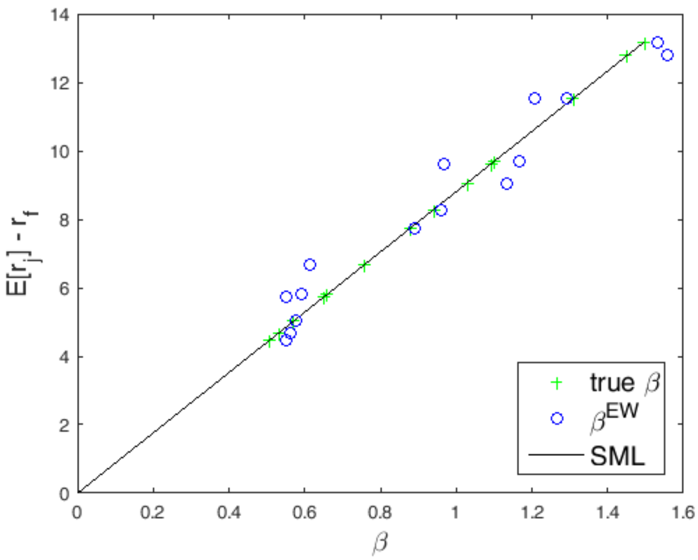


Figure 3: Security Market Line and CAPM estimates with and without Measurement Error. Green dots: estimates based on true market return. Blue dots: estimates based on equal weights.

4 Empirical Evidence in the Presence of Measurement Error

In the following we present empirical evidence on the relevance of measurement error in market returns using our cross-equation identification strategy presented in Section 2.2. We use minimum distance estimation (see Appendix A.1 for details) to estimate and test for attenuation bias and the presence of abnormal returns for various datasets and alternative measures of the market index. The effect of different market proxies on the beta estimates provides insights into the robustness of CAPM estimates and the quantitative relevance of the measurement error problem. In the first stage we estimate a CAPM system of regression equations by the seemingly unrelated regression (SUR) approach. We call this system of regression equations CAPM linear projection model (CAPM-LP) as it basically imposes no structure on the parameters implied by theory and can be taken as a pure statistical concept. Alternatively, one may regard this model as a general one-factor model with measurement error and/or abnormal returns, since non-zero intercepts as a consequence of measurement error and abnormal returns are not separately identifiable. The CAPM-LP, however, parametrically nests the standard CAPM (no abnormal returns, no measurement error, N true betas) and the CAPM with measurement error (CAPM-ME) ($\lambda_1 \neq 1$, N true betas and intercepts resulting solely from the presence of measurement error).

In the second estimation stage we impose the parametric structure implied either by the CAPM-ME or the standard CAPM. For all the second stage estimates we use the inverse of the first-stage variance covariance estimates as an optimal weighting matrix, so that the distance statistics at the minimum are asymptotically χ^2-distributed (see Appendix A.1). Tests for nested specifications can then be obtained by comparing the distance statistics of the restricted model against the less restricted specification. Thus the difference in the distance statistics of the CAPM-ME against the CAPM-LP is $\chi^2_{(N-1)}$-distributed and tests whether the intercepts can solely be explained by the intercepts of the CAPM-ME. This test therefore circumvents the problem of identifying the true from spurious alphas.

Assuming the CAPM-ME holds and imposing in addition the absence of any measurement error, $\lambda_1 = 1$, yields the minimum distance statistics for the standard CAPM. The difference between the two corresponding distance statistics yields a $\chi^2_{(1)}$-distributed test for the null hypothesis that the CAPM holds against the more general CAPM-ME.

Finally, the $\chi^2_{(N)}$-distributed Wald test on the zero intercepts tests in the tradition of the CAPM against any alternative which implies non-zero intercepts (CAPM-ME, abnormal returns, CAPM-ME including abnormal returns).

Table 4 summarizes the minimum distance estimates for three data sets consisting of different securities: (i) 25 Fama-French portfolios formed on size (market capitalization) and book-to-market ratio[4], (ii) a set of 20 randomly selected stocks from the S&P 500 (iii) 30 stocks of the Dow-Jones Industrial Average Index (DJIA). Our estimates are based on monthly data of size T = 60 and three different definitions of the market return. As market index we use the value-weighted return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ as provided on Kenneth French's website, the S&P 500 value weighted index and the Dow-Jones Industrial Average Index (DJIA). By construction the CRSP index is the broadest index in terms of the asset space covered, while the DJIA is the crudest proxy of

[4]see http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

the true market index. Therefore we would expect the lowest linear projection coefficient for the DJIA and the strongest attenuation bias for the beta estimates.

Table 4: Minimum Distance Estimates and Tests under Measurement Error

	French Fama	S&P500 stocks			DJ stocks		
Market Index	CRSP	S&P500	DJIA	CRSP	S&P500	DJIA	CRSP
CAPM vs. CAPM-LP	61.98 (0.00)	32.72 (0.04)	38.28 (0.01)	28.18 (0.11)	61.88 (0.00)	100.47 (0.00)	30.94 (0.42)
CAPM-ME vs CAPM-LP	39.16 (0.03)	26.23 (0.12)	26.12 (0.13)	26.95 (0.11)	32.03 (0.32)	31.99 (0.32)	30.90 (0.87)
CAPM vs CAPM-ME	22.82 (0.00)	6.5 (0.01)	12.16 (0.00)	1.24 (0.27)	29.85 (0.00)	68.49 (0.00)	0.043 (0.84)
λ_1	0.97	0.690	0.579	1.25	0.65	0.55	1.023
N	25	20	20	20	30	30	30
T	60	60	60	60	60	60	60

Minimum distance estimates of the linear projection parameter λ_1 and MD-based tests for different data sets and measures of the market proxies, p-values in brackets. N denotes here the number of assets in the regression model. Sampling period: 2010:06-2015:05 for all data sets.

Consider first the results of Wald test for the absence of intercepts (CAPM vs CAPM-LP). Our findings are in accordance with many previous empirical studies rejecting the null of no intercepts. However, note that in the presence of measurement error this test provides no information whether the non-zero intercepts result from true abnormal returns or are spurious due to measurement error. More interesting are the outcomes tor our test of the CAPM-ME against the CAPM-LP. Except for the FF-data we find that the measurement specification is sufficient to explain the presence of intercepts. Our test results indicate that the rejection of no abnormal returns often found for many data sets are likely to be the outcome of measurement error.

Note, that the linear projection coefficient λ_1 can only be interpreted as reliability ratio in the case of the EIV model. Only for this case the projection coefficient is bounded between by construction zero and one and implies an attenuation bias in the betas. Using the S&P500 and the DJIA as market proxies we find for all datasets significant evidence for an attenuation bias. The situation for the CRSP index as market proxy is somewhat different: The null of $\lambda_1 = 1$ for the Dow-Jones and the S&P500 stocks cannot be rejected, while it has to be rejected for the FF data. But for this case the attenuation bias is rather small.

5 Conclusions

In this paper we take a closer look at the consequences of a misspecified market index in the capital asset pricing model. Our focus is on two major sources of misspecification: (i) the use of inaccurate weights and (ii) the use of only a subset of the asset universe to construct the index. The consequences resulting from the use of badly chosen market proxy reach from inconsistent parameter estimates to a misinterpretation of tests on the existence of abnormal returns. High correlations between market proxies and the true market return are shown to be an insufficient to indicate that the choice of a particular market proxy is negligible. What matters is the predictive quality of the market proxy for the true market index in terms of a linear projection.

The estimation problems arising from measurement error in the market proxy deviate substantially from the ones typically found in errors-in-variable models for linear regressions. Unlike the true market index, market proxies are no longer orthogonal to the error in the CAPM. Instrumental variable estimation generally becomes infeasible, unless additional identifying assumptions are introduced.

For the the EIV-CAPM model, where the errors in the weights of the market index are assumed to be random, we present a new identification strategy. This strategy accounts for nonlinear cross-equation identifying restrictions which exploit the property that all CAPM equations are effected by the same attenuation bias for beta. Our estimation strategy allows us to test the linear projection based asset pricing model incorporating abnormal returns and measurement error against the EIV-CAPM as well as to test the CAPM with measurement error against the conventional Sharpe-Lintner CAPM without abnormal returns and measurement error. Our empirical findings indicate for three different datasets and different market proxies, that regardless of the assets under investigation the use of a more accurate proxy is rewarding and reduces the estimation bias. Moreover, our estimates indicate that the existence of abnormal returns in the conventional CAPM is spurious and can largely be explained by measurement error. In this sense Roll's (1977) fundamental critique can somewhat be mitigated.

However, the claim of this study is a rather moderate one. We do not intend to give an answer to the ongoing question of whether the CAPM is dead or alive. Our study simply points out that if we want the CAPM to give a chance to survive as a workhorse in academic finance and business, we should take it more seriously and try to account for the latency of the market return as best as possible. In particular, this includes taking on a more skeptical view on the evidence for the existence of abnormal returns in the presence of measurement error. Future research should consider other types of measurement error (e.g. multiplicative errors) as well as the consequences of measurement error in other multi-factor asset pricing models. In the light of our findings it seems also rewarding to reconsider the construction of popular market proxies which generally are characterized by a local bias contradicting the assumption that a large fraction of the investors follow a global investment strategy and ignore returns from other important tangible or non-tangible assets.

References

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997): *The Econometrics of Financial Markets*, Princeton University Press.

Fama, E. F. and K. R. French (2004): "The Capital Asset Pricing Model: Theory and Evidence," *Journal of Economic Perspectives*, 18, 25–46.

Fan, J. and Q. Yao (2017): *The Elements of Financial Econometrics*, Cambridge: Cambridge University Press.

Fuller, W. (1987): *Measurement Error Models*, Wiley.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989): "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57, 1121–1152.

Gourieroux, C. and J. Jasiak (2001): *Financial Econometrics*, Princeton: Princeton University Press.

Jaganathan, R., G. Skoulakis, and Z. Wang (2010): "The Analysis of the Cross-Section of Security Returns," in *Handbook of Financial Econometrics*, ed. by Y. Aït-Sahalia and L. P. Hansen, vol. Vol 2. - Applications, chap. 14.

Jagannathan, R. and Z. Wang (1996): "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance*, 51, 3–53.

Kandel, S. and R. F. Stambaugh (1987): "On correlations and inferences about mean-variance efficiency," *Journal of Financial Economics*, 18, 61 – 90.

Lintner, J. (1965): "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capitals Budgets," *Review of Economics and Statistics*, 47, 13–73.

Prono, T. (2015): "Market proxies as factors in linear asset pricing models: Still living with the roll critique," *Journal of Empirical Finance*, 31, 36 – 53.

Roll, R. (1977): "A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory," *Journal of Financial Economics*, 4, 129–176.

Shanken, J. (1987): "Multivariate proxies and asset pricing relations : Living with the Roll critique," *Journal of Financial Economics*, 18, 91 – 110.

Sharpe, W. F. (1964): "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 425–442.

Stambaugh, R. F. (1982): "On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis," *Journal of Financial Economics*, 10, 237–268.

A Appendix

A.1 Minimum Distance Estimation of the CAPM

Consider the stochastic form of the linear predictor equation for the j-th excess return on the (observable) excess return of the market $r_{m,t}$ including an intercept:

$$r_{j,t} = X_t' \pi_j + u_{j,t+1}, \qquad j = 1, \ldots, N,$$

where $\pi_j = (\pi_{j,1}, \pi_{j,2})'$ contains the reduced form parameters of a CAPM with measurement error and $X_t' = (1, r_{m,t})$. The collection of reduced form parameters into the so-called Π-matrix of dimension $N \times 2$ is given by

$$\Pi = \begin{bmatrix} \pi_1' \\ \pi_2' \\ \vdots \\ \pi_N' \end{bmatrix} = \begin{bmatrix} \mu_m(1-\lambda_1)\beta_1 & \lambda_1\beta_1 \\ \mu_m(1-\lambda_1)\beta_2 & \lambda_1\beta_2 \\ \vdots & \vdots \\ \mu_m(1-\lambda_1)\beta_N & \lambda_1\beta_N \end{bmatrix}.$$

With $Y_t = r_t = (r_{t1}, r_{t2}, \ldots, r_{tN})'$ as the vector of excess returns system of CAPM reduced for regressions is given by

$$Y_t = \Pi X_t + u_t,$$

where $\Pi = E[Y_t X_t']^{-1} E[X_t X_t']^{-1}$ is the matrix of linear predictor coefficients with its sample counterpart

$$\hat{\Pi} = \left(\sum_{t=1}^{T} Y_t X_t' \right) \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1}.$$

The Pi-matrix stacked into a row vector

$$\pi \equiv vec(\Pi') = g(\theta) = \begin{bmatrix} \mu_m(1-\lambda_1)\beta_1 \\ \lambda_1\beta_1 \\ \mu_m(1-\lambda_1)\beta_2 \\ \lambda_1\beta_2 \\ \vdots \\ \mu_m(1-\lambda_1)\beta_N \\ \lambda_1\beta_N \end{bmatrix}$$

relates the $2N$-vector of reduced form parameters to a vector $g(\cdot)$ of nonlinear functions of the structural parameter vector $\theta = (\beta', \lambda_1)$ of the same dimension, where π can be estimated consistently by single equation least squares without loss of efficiency:

$$\hat{\pi} = vec\left(\hat{\Pi}'\right) = \begin{pmatrix} \hat{\pi}_1 \\ \hat{\pi}_2 \\ \vdots \\ \hat{\pi}_N \end{pmatrix} = \begin{pmatrix} (X'X)^{-1} X'Y_1 \\ (X'X)^{-1} X'Y_2 \\ \vdots \\ (X'X)^{-1} X'Y_N \end{pmatrix}.$$

Under standard assumptions about the return process

$$\sqrt{T}(\hat{\pi} - \pi) \xrightarrow{d} \mathrm{N}(0, \Omega)$$

In the case of a homoskedastic error term vector with covariance matrix $\mathrm{V}[u_t] = \Sigma$, the asymptotic variance-covariance matrix of $\hat{\pi}$ takes the well-known form

$$\Omega = \mathrm{V}[\hat{\pi}] = \Sigma \otimes \mathrm{E}\left[X_t X_t'\right]^{-1}$$

assumed for seemingly unrelated regression models. Generalizations of Ω for the case of heteroskedasticity and autocorrelation can be easily derived.

The minimum distance estimator $\hat{\theta}$ based on the restriction $\pi = g(\theta)$ is defined by

$$\hat{\theta}(W_T) = \arg\min_{\theta \in \Theta} [\hat{\pi} - g(\theta)]' W_T [\hat{\pi} - g(\theta)],$$

where W_T converges asymptotically to the positive definite matrix weighting matrix W_0. The optimal feasible weighting is given by $W_T = [\frac{1}{T}\hat{\Omega}]^{-1}$, where $\hat{\Omega}$ is a consistent estimate of the asymptotic variance covariance matrix of $\hat{\pi}$. Note, that $g(\cdot)$ depends on the unknown parameter μ_m. In our empirical application, we replace μ_m by its sample mean.

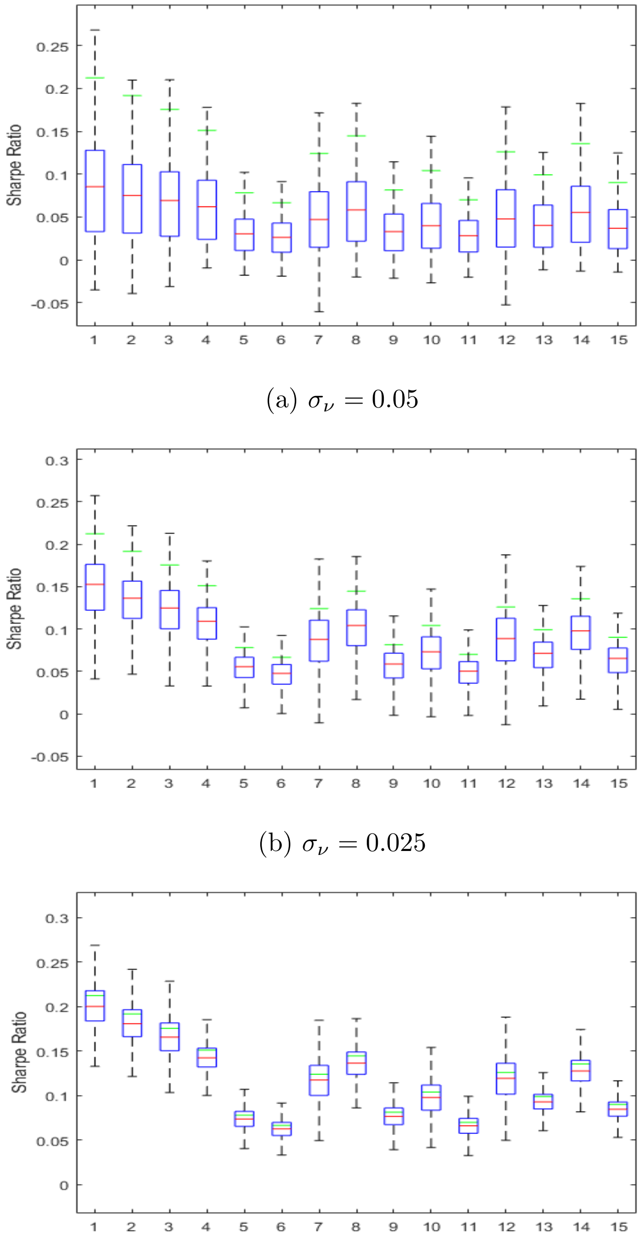
A.2 Detailed Monte Carlo Results

Table 5: MC-results for the CAPM with measurement error

			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean
CAPM	$r^*_{m,t}$	$\widehat{\alpha}$	0.008	0.001	-0.007	-0.005	0.008	0.007	-0.008	0.006	-0.030	0.010	-0.014	-0.003	-0.018	-0.028	-0.018	-0.006
		$ERR(.05)$	0.055	0.051	0.051	0.052	0.051	0.047	0.052	0.048	0.050	0.051	0.054	0.049	0.053	0.057	0.051	0.051
		$\widehat{\beta}$	0.659	0.756	0.651	1.310	1.092	1.309	0.532	0.878	1.030	0.570	1.100	0.508	1.497	0.943	1.452	0.953
		$\widehat{SR}$	0.212	0.192	0.176	0.151	0.078	0.067	0.124	0.145	0.081	0.104	0.070	0.126	0.099	0.136	0.090	0.123
		$\widehat{TR}$	8.975	8.947	8.995	8.885	9.127	9.107	9.244	8.986	9.091	9.349	9.159	9.252	8.934	8.952	8.961	9.064
CAPM-EIV	$\sigma_\nu = 0.05$	$\widehat{\alpha}$	3.548	4.136	3.532	6.981	5.874	6.877	2.821	4.632	5.316	3.030	5.688	2.684	7.827	4.926	7.502	5.025
		$ERR(.05)$	0.887	0.913	0.870	0.933	0.795	0.771	0.727	0.847	0.754	0.713	0.737	0.715	0.856	0.841	0.826	0.812
		$\widehat{\beta}$	0.207	0.243	0.205	0.398	0.338	0.388	0.159	0.259	0.294	0.169	0.320	0.149	0.432	0.278	0.412	0.283
		$\widehat{SR}$	0.083	0.073	0.067	0.060	0.031	0.027	0.049	0.058	0.034	0.042	0.029	0.051	0.040	0.055	0.037	0.049
		$\widehat{TR}$	14.687	-1.394	31.959	28.904	27.024	20.987	-9.751	13.205	-27.471	182.755	0.665	9.469	15.964	20.401	1.528	21.929
	$\sigma_\nu = 0.025$	$\widehat{\alpha}$	1.832	2.099	1.809	3.659	3.051	3.665	1.488	2.472	2.873	1.611	3.054	1.435	4.213	2.613	4.080	2.664
		$ERR(.05)$	0.549	0.578	0.494	0.672	0.416	0.426	0.373	0.522	0.412	0.348	0.392	0.372	0.553	0.512	0.515	0.476
		$\widehat{\beta}$	0.472	0.550	0.468	0.920	0.777	0.904	0.370	0.605	0.696	0.395	0.749	0.349	1.018	0.650	0.976	0.660
		$\widehat{SR}$	0.146	0.132	0.120	0.104	0.054	0.046	0.085	0.099	0.056	0.071	0.048	0.086	0.068	0.093	0.062	0.085
		$\widehat{TR}$	13.435	12.968	12.612	9.376	11.997	15.365	20.678	8.471	14.518	4.685	13.362	10.500	14.890	13.627	11.734	12.548
	$\sigma_\nu = 0.01$	$\widehat{\alpha}$	0.339	0.378	0.324	0.683	0.570	0.703	0.277	0.478	0.535	0.316	0.578	0.275	0.812	0.478	0.790	0.502
		$ERR(.05)$	0.087	0.078	0.067	0.098	0.072	0.074	0.075	0.077	0.071	0.066	0.077	0.076	0.099	0.083	0.095	0.080
		$\widehat{\beta}$	0.635	0.731	0.627	1.256	1.050	1.252	0.509	0.839	0.981	0.545	1.048	0.485	1.425	0.901	1.380	0.911
		$\widehat{SR}$	0.201	0.182	0.166	0.143	0.074	0.063	0.117	0.137	0.077	0.098	0.066	0.119	0.093	0.128	0.085	0.117
		$\widehat{TR}$	9.556	9.532	9.595	9.449	9.751	9.750	10.004	9.580	9.733	10.036	9.798	9.995	9.535	9.531	9.572	9.695
b_1^I	$M = 0.25 * N$	$\widehat{\alpha}$	-0.059	-0.330	-0.177	0.031	-0.062	0.199	0.370	0.341	0.757	0.269	0.718	0.419	0.689	0.402	0.864	0.295
		$ERR(.05)$	0.056	0.067	0.055	0.050	0.053	0.051	0.066	0.063	0.071	0.060	0.064	0.072	0.069	0.063	0.073	0.062
		$\widehat{\beta}$	0.908	1.084	0.916	1.783	1.499	1.753	0.667	1.146	1.284	0.739	1.387	0.627	1.936	1.221	1.845	1.253
		$\widehat{SR}$	0.216	0.202	0.182	0.152	0.079	0.066	0.114	0.139	0.075	0.099	0.065	0.114	0.095	0.129	0.084	0.121
		$\widehat{TR}$	6.492	6.228	6.388	6.527	6.629	6.794	7.474	6.898	7.340	7.276	7.321	7.629	6.928	6.932	7.074	6.929
	$M = 0.75 * N$	$\widehat{\alpha}$	-0.015	-0.054	-0.045	-0.009	0.041	0.015	0.023	0.018	0.073	0.043	0.073	0.036	0.074	0.020	0.117	0.027
		$ERR(.05)$	0.056	0.053	0.051	0.050	0.053	0.051	0.052	0.052	0.055	0.054	0.052	0.054	0.052	0.056	0.052	0.053
		$\widehat{\beta}$	0.687	0.794	0.683	1.364	1.131	1.359	0.550	0.912	1.060	0.590	1.134	0.523	1.549	0.976	1.495	0.987
		$\widehat{SR}$	0.214	0.194	0.178	0.152	0.078	0.067	0.124	0.145	0.081	0.103	0.070	0.125	0.099	0.136	0.090	0.124
		$\widehat{TR}$	8.583	8.520	8.583	8.531	8.802	8.759	8.960	8.646	8.866	9.029	8.883	9.004	8.641	8.648	8.700	8.744
b_2^I	$M = 0.25 * N$	$\widehat{\alpha}$	-0.059	-0.330	-0.177	0.031	-0.062	0.199	0.370	0.341	0.757	0.269	0.718	0.419	0.689	0.402	0.864	0.295
		$ERR(.05)$	0.056	0.067	0.055	0.050	0.053	0.051	0.066	0.063	0.071	0.060	0.064	0.072	0.069	0.063	0.073	0.062
		$\widehat{\beta}$	0.661	0.789	0.667	1.298	1.092	1.277	0.486	0.834	0.935	0.538	1.010	0.456	1.410	0.889	1.344	0.913
		$\widehat{SR}$	0.216	0.202	0.182	0.152	0.079	0.066	0.114	0.139	0.075	0.099	0.065	0.114	0.095	0.129	0.084	0.121
		$\widehat{TR}$	8.913	8.551	8.771	8.962	9.102	9.329	10.263	9.472	10.079	9.990	10.053	10.475	9.513	9.518	9.713	9.514
	$M = 0.75 * N$	$\widehat{\alpha}$	-0.015	-0.054	-0.045	-0.009	0.041	0.015	0.023	0.018	0.073	0.043	0.073	0.036	0.074	0.020	0.117	0.027
		$ERR(.05)$	0.056	0.053	0.051	0.050	0.053	0.051	0.052	0.052	0.055	0.054	0.052	0.054	0.052	0.056	0.052	0.053
		$\widehat{\beta}$	0.661	0.764	0.657	1.312	1.087	1.307	0.529	0.877	1.019	0.568	1.091	0.503	1.490	0.939	1.438	0.949
		$\widehat{SR}$	0.214	0.194	0.178	0.152	0.078	0.067	0.124	0.145	0.081	0.103	0.070	0.125	0.099	0.136	0.090	0.124
		$\widehat{TR}$	8.925	8.859	8.924	8.870	9.152	9.108	9.316	8.990	9.218	9.388	9.236	9.363	8.984	8.992	9.046	9.092
b_3^I	$M = 0.25 * N$	$\widehat{\alpha}$	0.305	0.320	0.218	0.218	1.016	-0.890	0.440	0.299	0.418	0.278	0.356	0.574	0.151	0.285	0.062	0.270
		$ERR(.05)$	0.071	0.065	0.057	0.055	0.083	0.070	0.070	0.061	0.060	0.058	0.054	0.088	0.055	0.059	0.052	0.064
		$\widehat{\beta}$	0.594	0.686	0.597	1.224	0.929	1.341	0.458	0.804	0.932	0.515	1.007	0.421	1.409	0.865	1.374	0.877
		$\widehat{SR}$	0.202	0.183	0.170	0.149	0.070	0.072	0.113	0.139	0.078	0.099	0.068	0.110	0.099	0.131	0.090	0.118
		$\widehat{TR}$	9.956	9.886	9.853	9.515	10.832	8.814	10.915	9.826	10.189	10.452	10.028	11.466	9.500	9.781	9.461	10.032
	$M = 0.75 * N$	$\widehat{\alpha}$	0.504	1.005	0.607	0.664	1.143	-0.034	-0.351	-0.159	-0.741	-0.081	-0.502	-0.377	-0.353	-0.221	-0.795	0.021
		$ERR(.05)$	0.100	0.184	0.097	0.086	0.091	0.052	0.066	0.053	0.075	0.053	0.057	0.067	0.055	0.058	0.074	0.078
		$\widehat{\beta}$	0.607	0.649	0.588	1.247	0.971	1.324	0.576	0.905	1.122	0.587	1.166	0.555	1.552	0.975	1.555	0.959
		$\widehat{SR}$	0.195	0.163	0.157	0.143	0.069	0.067	0.134	0.148	0.088	0.106	0.074	0.137	0.103	0.140	0.096	0.121
		$\widehat{TR}$	9.756	10.519	10.043	9.348	10.376	8.970	8.436	8.689	8.282	9.097	8.566	8.364	8.603	8.639	8.324	9.068
	$M = N$	$\widehat{\alpha}$	0.613	1.274	0.869	0.924	1.129	0.191	-0.254	-0.098	-0.931	-0.041	-0.581	-0.389	-0.301	-0.155	-0.941	0.087
		$ERR(.05)$	0.123	0.257	0.146	0.117	0.091	0.050	0.055	0.050	0.084	0.053	0.064	0.063	0.058	0.056	0.086	0.090
		$\widehat{\beta}$	0.591	0.613	0.552	1.207	0.967	1.290	0.561	0.891	1.134	0.577	1.166	0.553	1.531	0.959	1.559	0.943
		$\widehat{SR}$	0.191	0.156	0.149	0.140	0.069	0.066	0.131	0.148	0.090	0.105	0.074	0.138	0.102	0.139	0.097	0.120
		$\widehat{TR}$	10.052	11.177	10.720	9.668	10.436	9.232	8.678	8.834	8.174	9.185	8.568	8.389	8.713	8.782	8.304	9.261

Description: Average result over 10000 replications. First panel contains the average estimates based on the true (excess) market return. Further panels contain estimation results for the CAPM with the proxies of the market returns based on weights with fixed or random errors. $ERR(.05)$ denotes the average empirical rejection rate for the absence of abnormal returns (single equation test) for a significance level of 5 %. $\overline{SR}$ and $\overline{TR}$ denote the mean Sharpe ratio and the mean Treynor ratio across all Monte Carlo estimates. True values: $\overline{\beta} = 0.9520$, $\overline{SR} = 0.1175$, $\overline{TR} = 8.8070$.

A.3 Figures

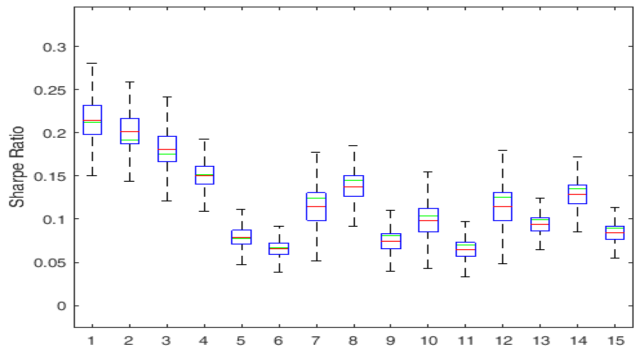


(a) $\sigma_\nu = 0.05$

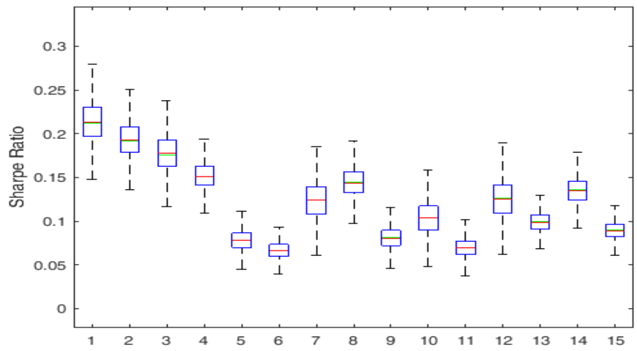
(b) $\sigma_\nu = 0.025$

(c) $\sigma_\nu = 0.01$

Figure 4: Sharpe Ratios of 15 assets based market indices with random measurement error

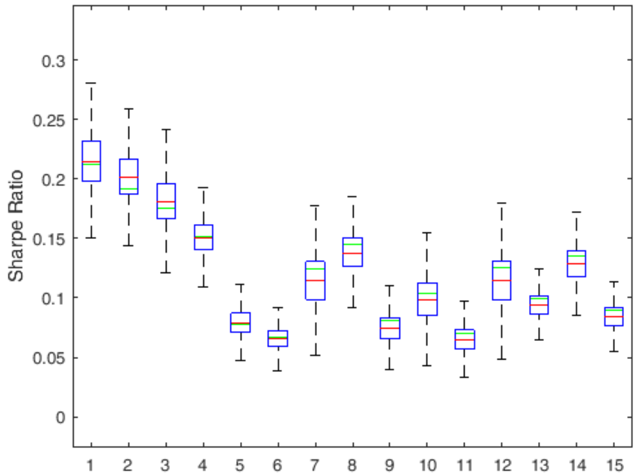


(a) 25%

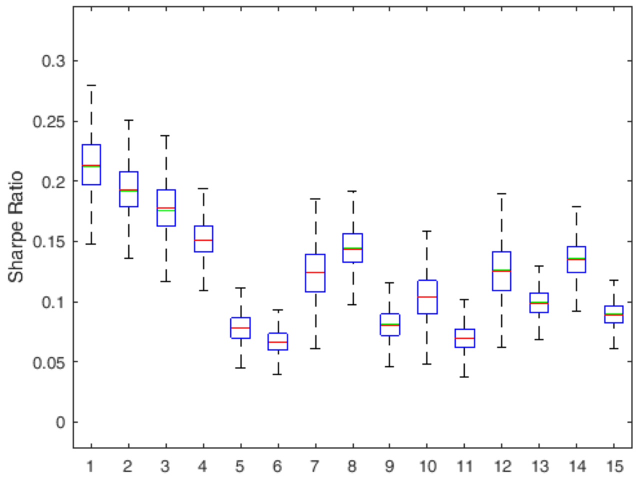


(b) 75%

Figure 5: Sharpe Ratios of 15 assets based on market indices based on the subset of assets with true weights

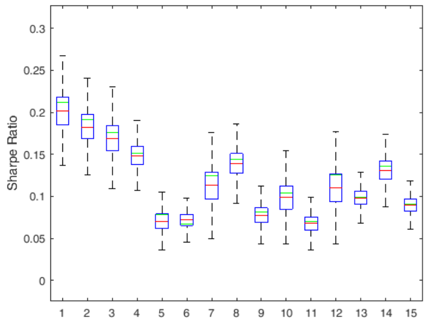


(a) 25%

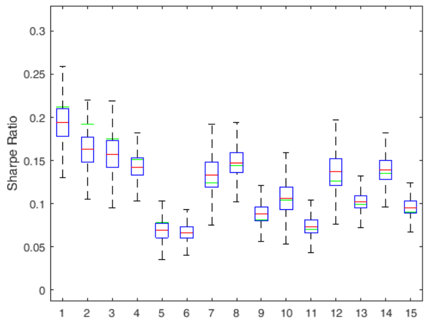


(b) 75%

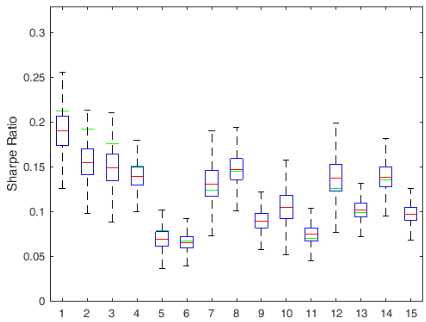
Figure 6: Sharpe Ratio of the assets from the CAPM with market proxies based on the subsets of assets with true normalized weights



(a) 25%



(b) 75%



(c) 100%

Figure 7: Sharpe Ratio of the assets from the CAPM with market proxies based on the subsets of assets with equal weights

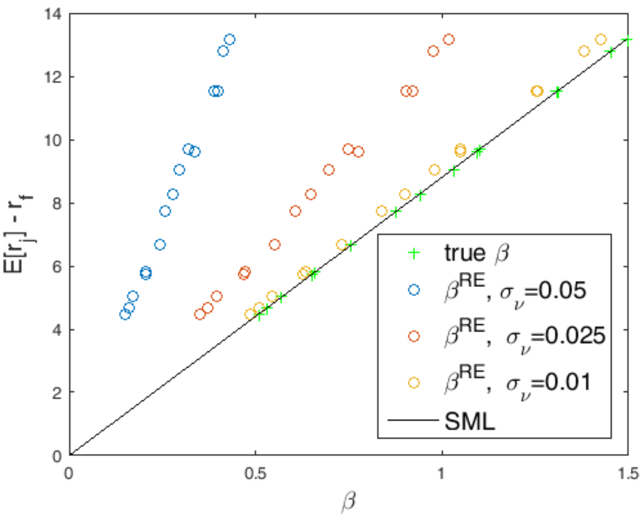
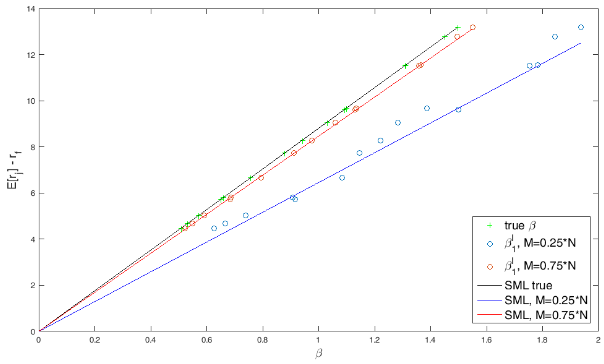
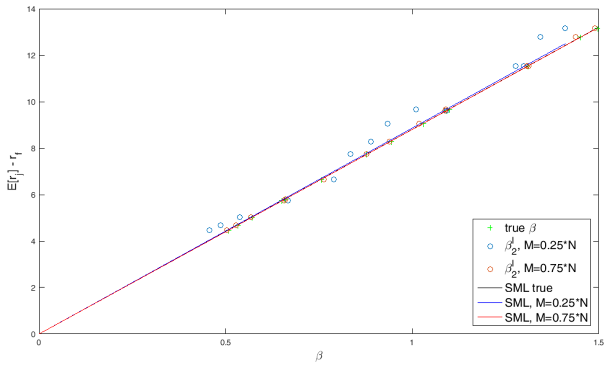


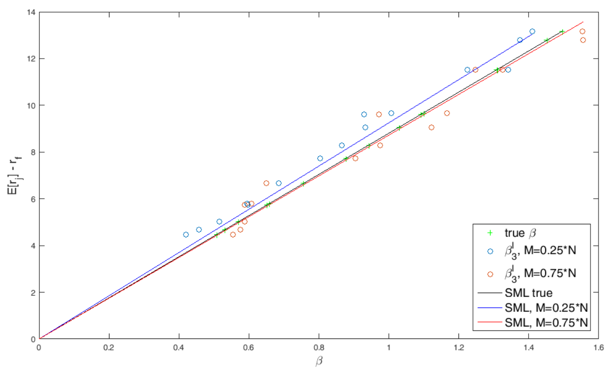
Figure 8: Security Market Line and estimated betas in the case of random measurement error in the weights of the market portfolio



(a) true weights for M stocks



(b) normalized true weights for M stocks



(c) equal weights for M stocks

Figure 9: Security Market Line and estimated betas in the case of misspecified index based on the subset of assets